8-7-2010

Near-Field Depth Perception in See-Through Augmented Reality

Gurjot Singh

NEAR-FIELD DEPTH PERCEPTION IN SEE-THROUGH AUGMENTED REALITY

By

Gurjot Singh

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Computer Science
in the Department of Computer Science and Engineering

Mississippi State, Mississippi

August 2010

NEAR-FIELD DEPTH PERCEPTION IN SEE-THROUGH

AUGMENTED REALITY

By

Gurjot Singh

Approved:

J. Edward Swan II
Associate Professor of Computer Science and Engineering
(Major Professor)

T.J. Jankun-Kelly
Associate Professor of Computer Science and Engineering
(Committee Member)

Robert. J. Moorhead II
Professor of Electrical and Computer Engineering
(Committee Member)

Edward B. Allen
Associate Professor of Computer Science and Engineering,
and Graduate Coordinator

Sarah A. Rajala
Dean of the Bagley College
of Engineering

Name: Gurjot Singh

Date of Degree: August 7, 2010

Institution: Mississippi State University

Major Field: Computer Science

Major Professor: Dr. J. Edward Swan II

Title of Study: NEAR-FIELD DEPTH PERCEPTION IN SEE-THROUGH AUGMENTED REALITY

Pages in Study: 88

Candidate for Degree of Master of Science

This research studied egocentric depth perception in an augmented reality (AR) environment. Specifically, it involved measuring depth perception in the near visual field by using quantitative methods to measure the depth relationships between real and virtual objects.

This research involved two goals; first, engineering a depth perception measurement apparatus and related calibration and measuring techniques for collecting depth judgments, and second, testing its effectiveness by conducting an experiment. The experiment compared two complimentary depth judgment protocols: perceptual matching (a closed-loop task) and blind reaching (an open-loop task). It also studied the effect of a highly salient occluding surface; this surface appeared behind, coincident with, and in front of virtual objects. Finally, the experiment studied the relationship between dark vergence and depth perception.

DEDICATION

To my parents Gurdeep Kaur & Gurwinder Singh, my godfather Gurcharan Singh, as well as those without whom this would not have been possible.

ACKNOWLEDGMENTS

Dr. J Edward Swan II

Dr. T.J. Jankun-Kelly

Dr. Robert J. Moorhead II

J. Adam Jones

Dr. Stephen R. Ellis

Josh Frank

Lorraine Lin

Shauncey Hill

Andreas Werner, ART Gmbh.

Bertelkamp Automation

Vere, Inc.

InterSense, Inc.

Department of Computer Science & Engineering

Institute for Neurocognitive Sciences & Technology

National Aeronautics and Space Administration

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AR Augmented Reality

AV Augmented Virtuality

HMD Head Mounted Display

IPD Inter Pupillary Distance

MR Mixed Reality

NASA National Aeronautics and Space Administration

VR Virtual Reality

CHAPTER 1

INTRODUCTION

In augmented reality (AR) systems, the presentation of virtual objects in the real world causes perceptual problems. One of the perceptual problems in AR is the perception of inconsistent depth orders of various objects. In AR, virtual objects are rendered on the HMD monitors in front of the observer's eyes and will always be visible in front of real objects, even if the virtual objects are actually intended to be behind them. This conflict results in incorrect depth judgments of virtual objects. This and many other problems are attributed to limitations in head mounted display (HMD) designs. These problems must be understood and alternate ways to compensate for them must be found before AR can be used in many practical applications.

The goal of this thesis was to design and conduct an experiment that compared the use of closed-loop and open-loop reaching tasks for depth judgments. Both closed-loop and open-loop reaching tasks have been previously studied in perceptual psychology; however, only a very few of these studies have been conducted in virtual and augmented reality environments. While closed-loop reaching tasks have been previously studied in some AR experiments, there has been no similar study that used open-loop reaching tasks in AR that we are aware of. This experiment compared observer performance for depth judgments using these two protocols, at near-field distances of 35 to 50 cm. To our knowledge, it is

the first time that these two complimentary measures have been studied in AR as part of the same experiment.

1.1 Augmented Reality

The mixed reality continuum (Milgram et al. [16]) in Figure 1.1 explains the incorporation of real and virtual objects in an observer's perceptual environment. There are different subsets of mixed reality based upon the degree of virtuality and reality, such as Virtual Reality (VR), Augmented Virtuality (AV), Augmented Reality (AR), and the Real Environment. In VR and AV, the environment around an observer is primarily virtual, consisting of none or a few real objects. On the other hand, in AR, the surrounding environment is primarily real, while a few of the objects in it are virtual.

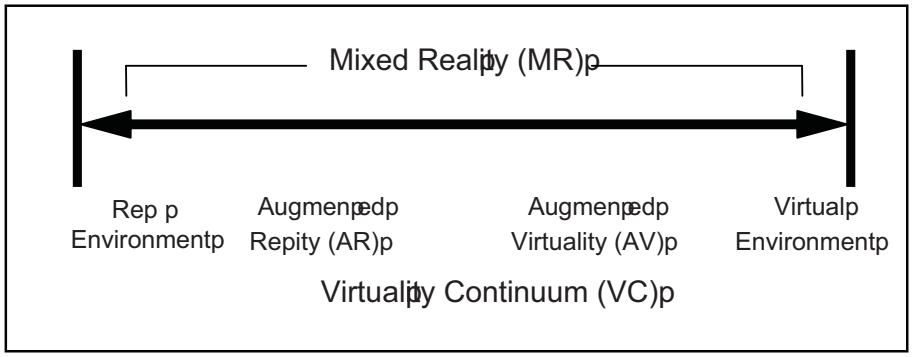

Figure 1.1

Milgram et al. [16] mixed reality continuum

In AR, the real world is supplemented with virtual, computer generated objects, which appear to be a part of the real environment. It is different from VR, where the observer is

completely immersed in a computer generated environment, and their perceptual environment is completely virtual.

An interesting application of AR is the ability to visualize hidden infrastructure occluded by real physical surfaces. This involves removing some real objects from or adding some virtual objects (Azuma et al. [2]) to the perceived environment. This application of AR is known as *x-ray vision*. It gives an observer the ability to see hidden objects, which are positioned behind opaque surfaces such as walls, etc.

In a complimentary categorization to Milgram et al. [16], Azuma [1] divides AR systems in two categories: *optical see-through* and *video see-through*. In optical see-through AR, the observer perceives the real and virtual objects using optical combiners (see Figure 1.2). These combiners are an optical system that is partially transmissive and partially reflective, and therefore the system can let light from the real world pass through and at the same time can reflect computer generated imagery. As a result the observer sees both real and virtual objects at the same time. One problem with optical see-through AR is that the optical combiners reduce the amount of incoming and reflected light. The combiner has to be reflective enough to reflect light and transmissive enough to transmit light. Therefore, there is a tradeoff involving reducing the incoming light in order to properly see the reflected graphics.

In video see-through AR, the real scene is captured using video cameras, combined with computer generated objects and then displayed onto monitors. The virtual objects are generated by a scene generator and then both real and virtual scenes are digitally blended. Since this technique uses image processing of the real world video, it does not have the

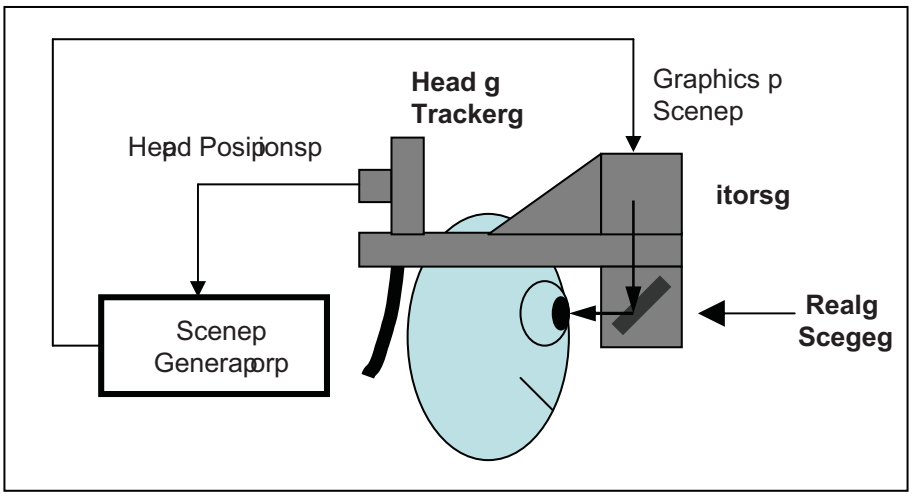

Figure 1.2

Optical see-through AR system (Azuma [1]).

light intensity tradeoff problem of optical see-through AR. It is also possible to completely remove real objects from the observer's view, and add virtual objects accurately in the scene by modifying the incoming video feed. Removing real-world objects is not possible in optical see-through AR because the observer sees the real world directly through the combiners.

Video AR encompasses two display modes: (1) through an HMD (see Figure 1.3) and (2) when viewed on a monitor (see Figure 1.4, Figure 1.5). In HMD video AR, the scene is displayed onto monitors in front of the observer's eyes. The cameras are mounted on the observer's head, where they capture live video of the real world. In the second case, when viewed on a monitor, the cameras are independent of the user's viewing position and can be positioned anywhere (see Figure 1.5). There is another video AR display modality

involving AR on a handheld device (see Figure 1.4). The camera on the back of the device captures the real scene, and then virtual objects are added to it.

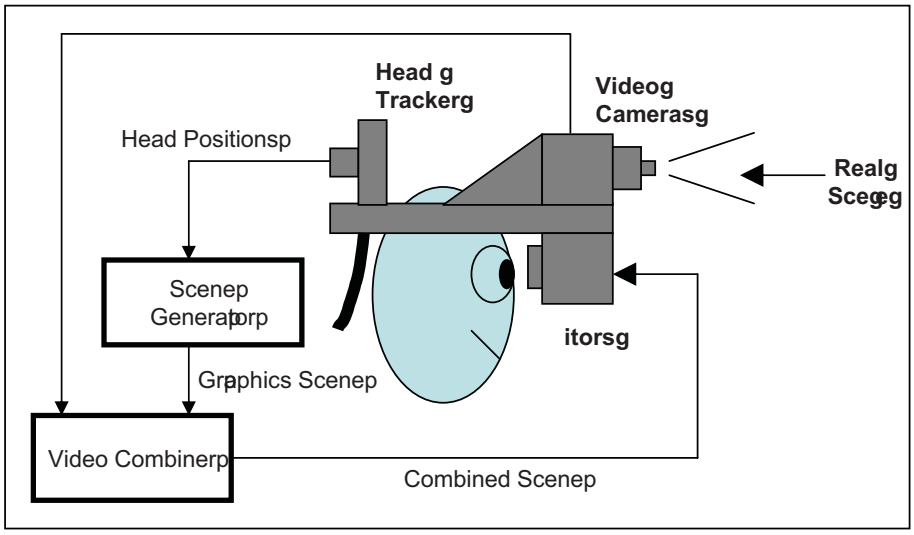

Figure 1.3

Video see-through AR system through an HMD (Azuma [1]).

Considering optical and video AR, video AR is much easier to implement and has been used in commercial products. The most common commercial implementation of video AR is sports broadcasting. For example, in a football game (see Figure 1.5), the ground and the players are real but the blue line of scrimmage is computer generated. Video image frames are modified so that the blue line appears to be stuck on the ground and under the players. This effect is extremely compelling. Another example was in the summer 2008 Olympics, where there were computer generated flags and other such markings in the lanes under the swimmers.

Figure 1.4

Video see-through AR on a handheld device

Figure 1.5

Video see-through AR for real-time sports broadcasting

1.2 General Depth Perception

The human visual system captures a pair of 2D retinal images of the scene, one from each eye. The brain then extracts a 3D perceptual world from these flat and ambiguous retinal images by identifying various depth cues and ascertaining depth relationship among different objects in the scene. At least nine or more sources of depth information have been identified (Cutting et al. [5, 6]) (see Table 1.1), which help an observer to understand the depth layout of the world. These depth cues can be categorized based upon whether they are monocular or binocular. The monocular cues need only one eye to give depth information, while binocular cues require both eyes. Similarly, some cues can provide only ordinal depth information while others can provide quantitative depth information. The ordinal depth information tells only about the arrangement of objects and does not give any information about the measurable distance between them. The quantitative depth information, on the other hand, tells both about the arrangement as well as the distances between objects, e.g. one object may be in front of another object and the distance between them may be 1 meter.

These depth cues have varying saliency levels at different distances. The saliency level of some depth cues decreases with distance, while for others it remains constant. Cutting [5, 6] defined saliency s as a ratio of the just-noticeable distance (Δd) and the distance d between object and observer:

$$s = \Delta d/d, \tag{1.1}$$

Table 1.1

Depth Cues

Depth cue	Information	Type	Saliency with Increasing Distance
Occlusion	Ordinal	Monocular	Constant
Height in Visual field	Quantitative	Monocular	Decreases
Relative Size	Quantitative	Monocular	Constant
Relative Density	Quantitative	Monocular	Constant
Aerial Perspective	Quantitative	Monocular	Increases
Binocular Disparity	Quantitative	Binocular	Decreases
Accommodation	Ordinal	Monocular	Decreases
Convergence	Quantitative	Binocular	Decreases
Motion Perspective	Quantitative	Monocular	Decreases

where Δd is the minimum distance of an object from its original position that is noticeable by the observer. Depth cues have different saliencies at different distances, as shown in Figure 1.6. Various depth cues defined by Cutting [5] are:

Occlusion (Figure 1.7) depth cue states that objects closer to the observer partially or completely occlude farther objects, along the same visual path. Occlusion is a monocular depth cue and can provide only ordinal information; it is not possible to measure any quantitative depth values with only occlusion. However, it is a very strong ordinal depth cue and is effective at all distances without any attenuation; it works as well at far distances as it does at closer distances.

Height in the visual field (Figure 1.8) is a monocular depth cue, which states that the distance to an object can be perceived by its vertical position in the visual field. The visual field of an observer is the image of the scene projected on the observer's retina. This depth cue is based on the fact that the horizon is the farthest point and objects which are

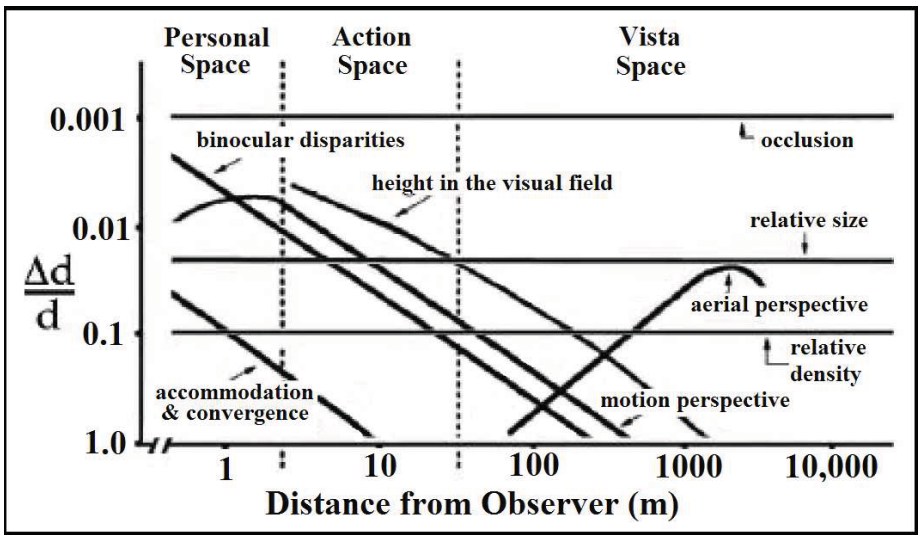

Figure 1.6

Saliency of depth cues by distance (Cutting [5])

Figure 1.7

Occlusion: The church occludes the buildings in the background

below the horizon will ascend in the visual field as they become more distant, and objects which are above the horizon will descend in the visual field as they become more distant. This cue works only when all the objects are placed on a common ground plane such as a terrain. As Figure 1.6 indicates, the depth information from this source attenuates with increasing distance.

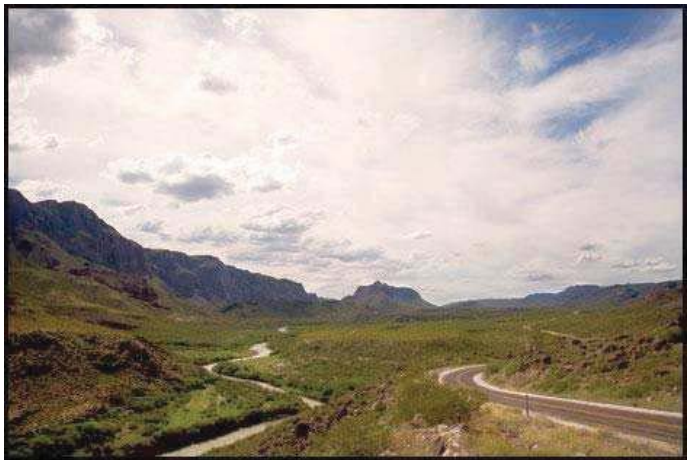

Figure 1.8

Height in the visual field

Relative size (Figure 1.9) depth cue establishes depth relationships among objects of the same size. An object projects a 2D visual image on the retina, and closer objects will extend a larger angular extent on the retinal image than a similar distant object. The more distant an object is, the smaller it will look as compared to a similarly-sized sized near object. By calculating the angular extent of different objects it is possible to calculate the depth ratio information. For example, if an observer sees two similarly-sized objects with

one subtending half of the visual angle of the other, the smaller object must be twice the distance as the larger one.

Figure 1.9

Daffodils image showing both relative size and density

Relative density (Figure 1.9) depth cue involves comparing the number of similarly-sized objects or textures per solid visual angle. When a group of objects is seen from a distance it subtends a smaller visual angle then if it is seen close by. A larger number of objects will be seen per solid angle when they are at a farther distance than when they are closer. Therefore, the more objects per solid angle, the more distant they are. Similarly, a farther texture will appear more dense then a closer texture. This cue can provide ratio information, as it is possible to calculate the distance ratio between objects by counting the number of objects per solid visual angle.

Aerial perspective (Figure 1.10) is the effect of atmospheric conditions on the appearance of objects. The atmosphere has suspended particles such as moisture, vapors, pollutants, dust, etc. Because of the presence of these particles, the contrast of an object decreases with distance, and its color desaturates and becomes shifted towards the atmospheric color. The more distant an object is from the observer, the more pronounced these effects become. Based on this contrast calculation, this cue can provide quantitative depth information in the scene; however, it is effective only at relatively large distances of 100 meters or more.

Figure 1.10

Aerial perspective

Binocular disparity (Figure 1.11) is the difference between the relative positions of an object's projection on the retinas of the two eyes. An object projects different images on

each retina because the eyes are in slightly different positions. This difference is called *disparity*. An object which is closer to the eyes will produce projections with more disparity than a more distant object. This disparity yields *stereopsis*, the impression of a solid space. For closer objects, this effect is very prominent.

Figure 1.11

Stereogram representing binocular disparity

Accommodation (Figure 1.12) is the change in the shape of the lens in the eye to bring the object of interest into focus. Objects which are in focus will look clear and sharp with well defined edges, while other objects which are closer or farther than the focused object will look blurred. The lens converges light from the focused object exactly on the retina. The light from the unfocused objects converge either in front of or behind the retina. The effective depth range for accommodation alone is about 2 meters.

Figure 1.12

Accommodation: The water drop is in focus

Convergence (Figure 1.13) is the angle between the lines of sight of the two eyes. Our eyes are situated at different positions and have different views of the world from each other. To fixate on an object at a given depth, the eyes need to rotate towards each other. In real world settings, for normal binocular vision, accommodation and convergence work together (Fincham et al. [8]). A reflex binds accommodation and convergence; as the convergence changes from a far object to a nearby object, accommodation also changes to bring the objects in focus. Similarly, if attention changes from one object to another, convergence and accommodation also change together to maintain correspondence between both retinal images. This reflex is called the *convergence-accommodation reflex*. While convergence and accommodation normally operate together in the real world, they are dissociated in most VR displays, because accommodation is typically a fixed, unadjustable value. This accommodation/convergence mismatch is a known source of depth perception

errors in VR displays (Mon-Williams et al. [18]). Convergence alone as a depth cue is effective only at close ranges within 2 meters. Convergence and accommodation together are effective to about 3 meters.

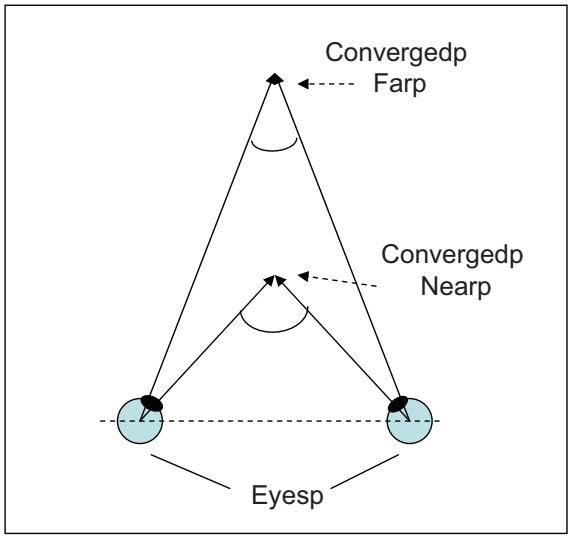

Figure 1.13

Eye convergence: the convergence angle is larger when converging at near objects than for far objects

A phenomenon related to convergence, known as *dark vergence*, is known to influence the perception of distance and size at near distances (Gogel et al. [9]). In total darkness, when there is no object to converge on and no depth cues present, the human oculomotor system adjusts itself and the eyeballs shift to a rest state. The angle of vergence adopted by eyes in this resting state is known as the dark vergence angle and the distance where the eyes are converged is known as the dark vergence distance. This dark vergence distance is unique for each individual and usually falls between 1-2m.

In AR, when virtual objects are presented along with real ones, the phenomenon of *proximal vergence* affects depth perception. Observers tend to converge on the most salient surface (typically a real surface) in a scene (Ellis et al. [7]). For example in Figure 1.14(a), when a virtual object is presented in front of a salient real background, observers tend to converge on the background, which in turn biases the depth perception of the virtual object towards the background. One possible explanation of this overestimation is that, in the process of converging on the background, the angle of convergence is less than is needed for the virtual object. Similarly, when an occluder is introduced in front of the virtual object as in Figure 1.14(b), observers tend to converge on the occluder itself. This biases the depth estimation of the virtual object towards the observer.

Motion perspective (Figure 1.15) depth cue is the difference in the apparent motion of objects located at different distances from the observer. It can be divided into two categories based on the movement of the observer. In the first category, the objects are fixed at certain distances and the observer is moving. When an observer fixates on an object and moves sideways, the fixated object does not seem to move. However, closer objects appear to move quickly in the opposite direction while farther objects appear to move slowly in the observer's direction of motion. In the second category, the observer is stationary while the objects are moving. For example, in a rain storm, the rain drops near the observer appear to fall more quickly than the drops farther from the observer.

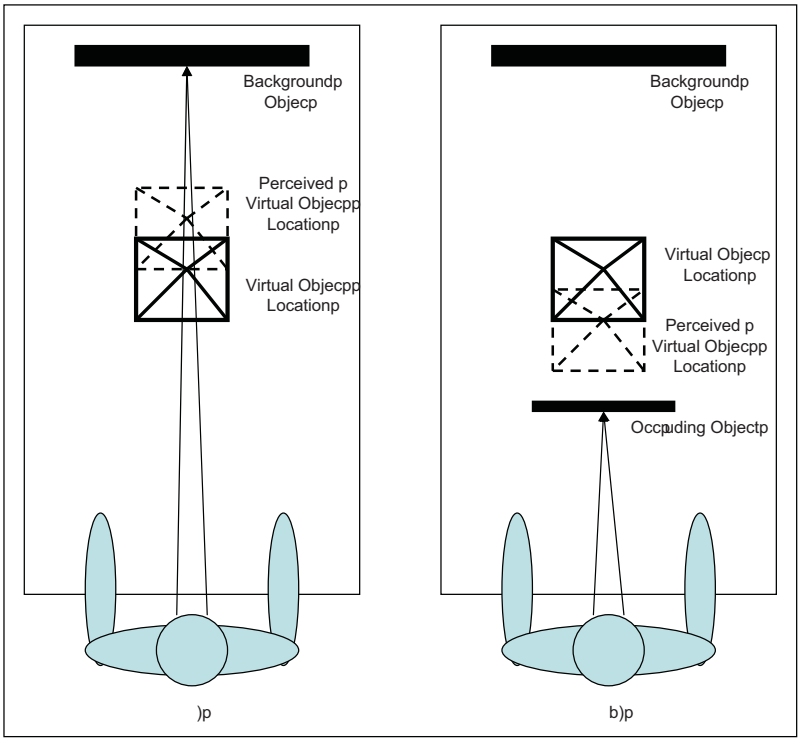

Figure 1.14

Proximal vergence: The depth of an object is biased towards the most salient surface. (a) The bias causes overestimation of distance, (b) the bias causes underestimation.

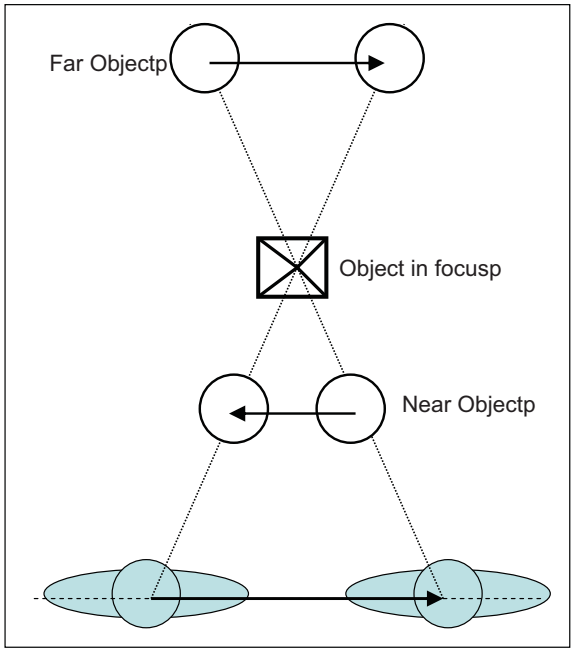

Figure 1.15

Motion perspective: The near object appears to move in the opposite direction and the far object appears to move in the same direction as the observer.

1.3 Depth Perception in Virtual & Augmented Reality

The presence of enough depth cues is necessary to understand depth relationships among various objects in an environment. This condition remains true for all environments, whether the environment is real, virtual, or augmented. In VR, only the virtual scenes are presented to the observer, which are more or less a replica of the real world. A combination of depth cues, which are consistent with the human visual system (such as occlusion), can be incorporated in these virtual scenes, which can allow the observer to intuitively interact with the environment. It is sufficient to make the virtual environment realistic enough with sufficient depth cues present to understand depth relationships among various objects. It can convince the observers that they are in some degraded version of a real environment.

In optical see-through AR, virtual objects are presented along with real objects. The presentation of virtual objects along with real ones presents some perceptual problems which are unique to AR environments. One major problem in AR is how the human visual system interprets the occlusion depth cue. Since the virtual objects are rendered on HMD monitors, which are located just in front of the observer's eyes, they will always be in front of the real world. This situation will remain true in all depth order scenarios; even if the virtual objects are intended to be located behind a real world object, they will still occlude it. This situation results in conflicting depth cues, because usually the human visual system interprets the occluded object as the farther one. Another problem occurs when more than one virtual object is presented at different distances along with a real object; if some virtual objects are in front of the real object and some virtual objects are behind the real object,

it is unclear how the human visual system will perceive depth relationships in this case. In order to design practical AR systems, we must understand how depth perception and depth ordering work in augmented reality.

One advantage of a virtual environment over an augmented one is that any undesirable inconsistency in depth cues will span the whole scene, keeping the spatial relationship among all the virtual objects constant. Nevertheless, depth perception in virtual environments does not operate correctly and is usually underestimated (Loomis et al. [14], Swan et al. [24]). Depth perception in video see-through AR is assumed to operate similarly to how it operates in VR, although it has been barely studied (Messing et al. [15]).

1.4 Depth Judgment Techniques

The measurement of depth perception is a problem in itself. Depth perception is an invisible cognitive state and is inside an observer's head. It is unique for every individual and can not be measured directly as it is observable only by the perceiver. Therefore, to measure depth perception, indirect methods are used which involve performing some action which results in a measurable behavior. These methods have been developed by scientists and assume that physical judgments of depth represent mental depth perception. Some examples of depth judgment techniques which are suitable for medium-field distances of about 1.5 to 30 meters are *blind walking*, *verbal report*, and *triangulated walking*. Another category of depth judgment techniques, which are suitable for near-field distances, are *perceptual matching techniques* (Ellis et al. [7]), and *blind reaching techniques* (Loomis et al. [14], Mon-Williams et al. [18]). These techniques involve adjusting one object in order

to match it to another object's depth. The techniques can be categorized as updating a response based on some feedback, and can therefore be analyzed using concepts from control theory.

The field of control theory (Levine [12]) deals with the behavior of dynamic systems. In control theory, some input is given to a system, and in turn it produces an output. In order to get the desired output, there is a control system which changes the behavior of the system. Control systems can be broadly classified as *closed-loop* or *open-loop* systems. A closed-loop system (see Figure 1.16) has a feedback loop. Here, the system receives an input signal, produces an output, gets feedback from the effects caused by the output, and then further adjusts the output. An open-loop system (see Figure 1.17) is controlled directly by the input; it is not corrected according to the output.

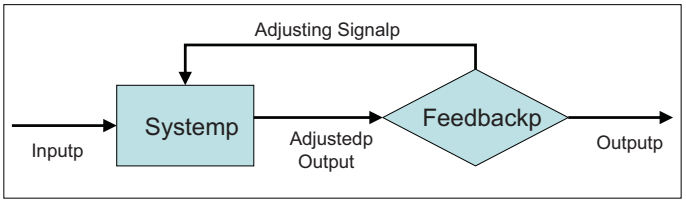

Figure 1.16

Closed-loop control system

In psychological research, both open-loop and closed-loop techniques are utilized to measure user performance. For closed-loop techniques, the participant performs some task with some kind of feedback, and hence they can alter their response based on this feedback. For open-loop techniques, the participant performs some task without any feedback. In our

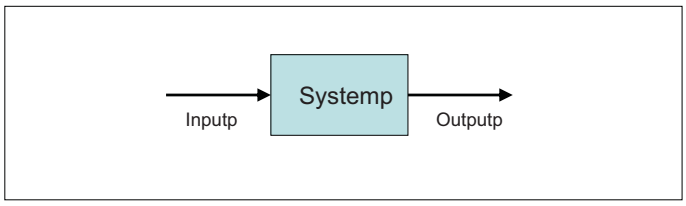

Figure 1.17

Open-loop control system

experiment we used depth judgment techniques, where the depth of a physical pointer (a referent object) was matched to the depth of a virtual pyramid (a target object), using both closed-loop (see Figure 1.18) and open-loop (see Figure 1.19) techniques.

In the real world, people typically manipulate objects using closed-loop perception; for example, keeping one's eyes on a target object while reaching for it. They execute the reaching action under the direct perceptual guidance of the information obtained from looking at the target object and their hand. In our experiment, participants moved a physical pointer above the table until it matched the distance to a virtual pyramid (see Figure 1.18); this is a *perceptual matching task*. Here both the pointer and pyramid were visible to the participant, and therefore visual feedback was available. Because of the natural perceptual-motor coupling in a closed-loop task, the visual feedback from looking at both the target object as well as the referent object acts as perceptual basis for the matching action.

An open-loop task involves seeing an object, and then reaching for it with closed eyes (while blind). Blind reaching is the near-field equivalent of blind walking: it is an action-based task that does not involve visual feedback. However, when blind reaching, observers

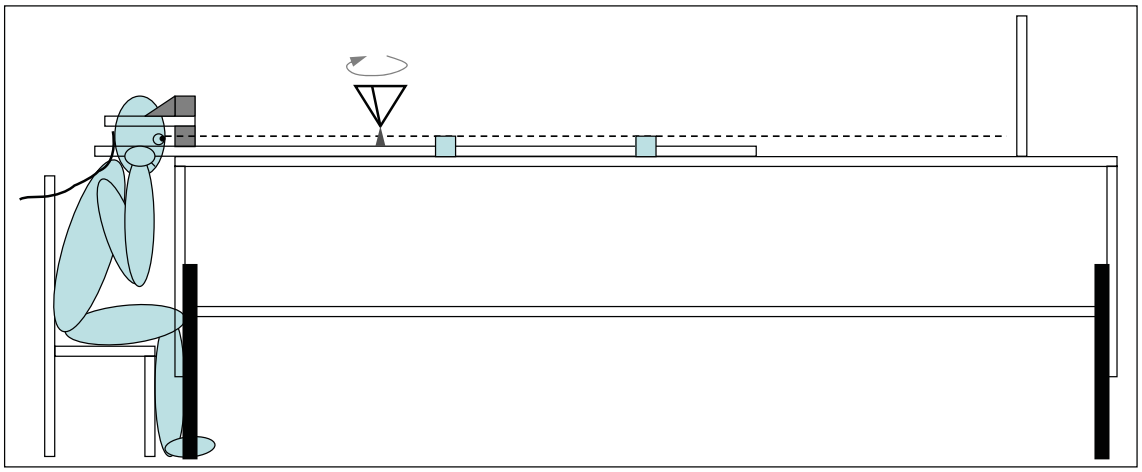

Figure 1.18

Closed-loop matching technique (with visual feedback).

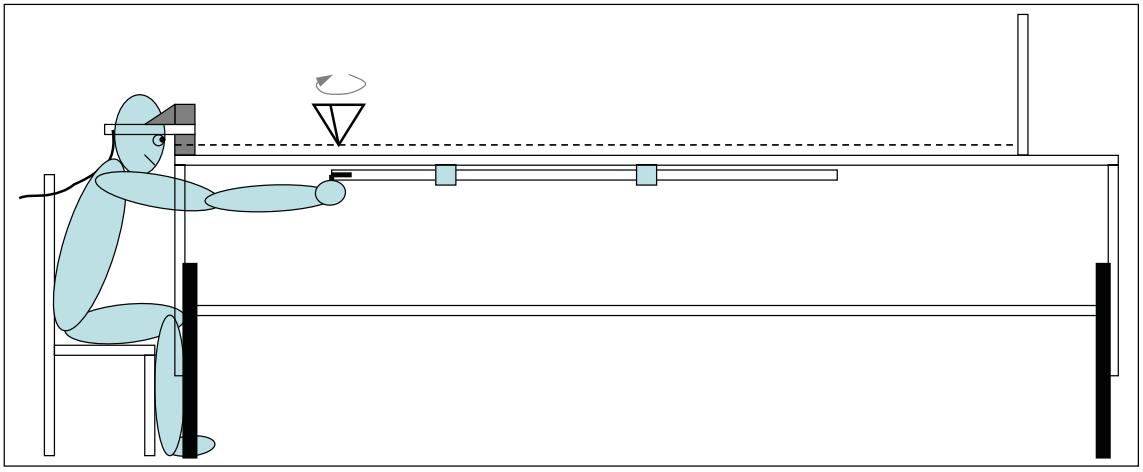

Figure 1.19

Open-loop blind-reaching technique (without visual feedback).

do have proprioceptive feedback about their hand's location, so it is not a fully open-loop task; it is more precisely a *visually* open-loop task. Nevertheless, we will use the term "open-loop" to refer to this task. By using an open-loop protocol, we can measure the perception of depth from that distance, which the human cognitive system stores, in order to allow the action to be performed while blind (without any visual feedback). With a closed-loop task, the human visual system continuously updates its estimate of the object's depth, which destroys our ability to measure it from a particular distance.

In our experiment, participants moved a physical pointer under the table until it matched the distance to the virtual pyramid (see Figure 1.19); this is a *blind reaching task*. Here only the pyramid was visible to the participant, and the pointer was hidden from view, and therefore no visual feedback was available. The open-loop blind reaching protocol divides the task into two sub-parts based on their cognitive requirement; it disconnects the perceptual basis of action from the action itself. The first part requires cognitive processing by forming a mental representation of an extension of the virtual pyramid location extending under the table, given the visual input from the location of pyramid above the table. The second part is reaching under the table to that mentally-represented location. This reaching process depends solely on proprioception and requires a mediating process of mentally visualizing and updating the target location while the blind reaching action is performed. This process is completely different from the closed-loop task, which is a perceptual matching task, with no cognitive processing and where the perceptual basis of the matching action is connected to the action by visual feedback. However, both protocols

are similar in their memory requirements as the stimuli are available and visible in both cases.

1.5 Motivating Applications

The applications of augmented reality span many fields, such as medical visualization, manufacturing, repair, maintenance, annotation, robot path planning, entertainment, military, and architecture (Azuma [1]). A promising application of near-field augmented reality is medical applications such as medical visualization, needle-guided biopsies, obstetrics, and cardiology (Azuma [1], Bichlmeier et al. [4], Soler et al. [22], State et al. [23]). To be successful, these applications require a correct perception of augmented graphics with respect to the real environment.

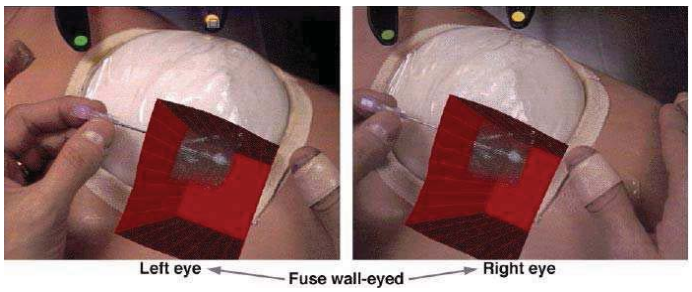

Figure 1.20

Ultrasound-guided needle biopsy

State et al. [23] used AR for ultrasound-guided needle biopsies of the human breast (See Figure 1.20). They developed a real-time stereoscopic video see-through AR system, which merged the live ultrasound data, the geometric elements, and the video feed as

stereo images in an HMD. They tested their system on a training phantom of a human breast, where physicians successfully guided a needle into an artificial tumor in the breast.

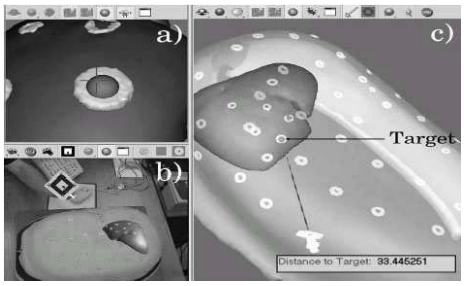

Figure 1.21

Augmented view of abdomen for abdominal surgery planning

Soler et al. [22] used a combination of VR and AR in the development of a surgical planning system (see Figure 1.21) for digestive surgery. They tested it on an abdominal phantom, where the task was to guide a real surgical tool to some modeled target tumors placed inside the abdominal phantom, while looking at the augmented view. The participants were successful in guiding the needle to the tumors.

Bichlmeier et al. [4] (see Figure 1.22) used AR to integrate 3D medical imaging data in-situ with the patient's own anatomy. They used video AR with the real video feed integrated with computer generated virtual anatomy. They used a new method for incorporating the data, which increases the user's depth perception by adjusting the transparency of parts of the video according to the shape of the patient's skin. Some additional depth cues were also provided for enhanced depth perception such as virtual shadows, extension of medical instruments and a keyhole port inside the patient's body.

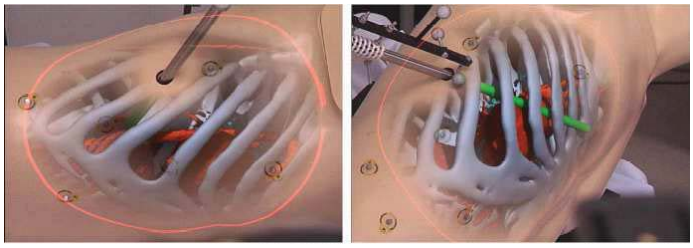

Figure 1.22

Virtual keyhole and instrument extension

As a manufacturing example, an AR prototype system (Sims [21]) was tested by the Boeing company, that helped technicians in building a wiring harness for the electrical system of an airplane. The technicians wore head-tracked HMDs, which showed computer-rendered circuit wiring. This allowed the technicians to route the wires without having to refer to complicated wiring diagrams located in reference books.

CHAPTER 2

RELATED WORK

2.1 Depth Perception in Virtual Reality

Depth perception has been an interesting area of psychological research for over 100 years. A large amount of literature is available on how depth perception works in the real world, and various depth cues for depth perception have been identified and classified. As discussed in chapter 1, Cutting [5] describes nine primary sources that provide the cues for depth perception in real and virtual worlds. He also examined the evolution of depth cues in drawings and painting over the centuries. He observed that depth cues have evolved from simple to a combination of complex cues, over the course of development of various art styles such as cave paintings from ancient times, artistic paintings from medieval times, photography, etc. He then extends this series of visual artistic expressions to virtual reality and concludes that, because the complexity of depth cues has increased as art has evolved, even more complex cues will be used in virtual reality.

As indicated in Figure 1.6, different depth cues have varying levels of saliency over different distances. Cutting [5] used these varying levels of saliency to classify distances around an observer into three distances. The first is the *near-field*, which extends to arm's length ($<$ 1.5 meters) around the observer. It is the distance where objects can be easily grabbed and manipulated. The second is the *medium-field*, which extends from arm's

length to almost 30 meters. It is the distance where objects can be accurately thrown and conversation is possible. The third is the *far-field*,which extends beyond 30 meters. It is the distance where the observer may be walking towards or scanning for game or predators. Our experiment took place at near-field distances, where binocular disparity, accommodation, and convergence are the most salient depth cues.

Over the past 15 years, more than 40 studies have examined egocentric depth perception in virtual reality environments at medium-field distances (Loomis and Knapp [14], Swan et al. [24]). Both of these references contain extensive literature surveys about depth perception in virtual reality. Almost every study has concluded that egocentric depth is underestimated in virtual environments. A number of reasons for this have been proposed and tested; however, the exact cause of this phenomenon is still unclear. It is most likely caused by a variety of interacting factors.

2.2 Depth Perception in Augmented Reality

As compared to virtual reality, augmented reality is a relatively new field. A major problem in AR is how an observer perceives the depth of virtual objects when presented along with real objects.

Ellis[1] et al. [7] examined this question at near-field distances, in order to develop design guidelines for augmented reality display systems. Their experiment involved measuring the effects of convergence, accommodation, observer age, viewing condition (monocular,

[1]Stephen R. Ellis, NASA Ames, has been a collaborator in the development of the experiment discussed in this thesis.

biocular[2] stereo, binocular[3] stereo) and the presence of an occluding surface (the x-ray vision condition) on depth judgments of a virtual object using a closed-loop perceptual matching technique. The task was to match the depth of a small light on an adjustable arm to the depth of a virtual pyramid (see Figure 1.18). They found that monocular viewing degraded the observer performance for depth judgments, and that most of the localization errors occurred when the physical surface was closer than the virtual object. They attributed these errors to the phenomenon of proximal vergence: they found that introducing a physical surface in front of the virtual object caused a relative forward movement in the localization of the virtual object. They proposed that this can be attributed to a small relative convergence of the eyes, caused by proximal vergence. Then, they cut a hole in the occluding surface so that the virtual object was visible through the occluder. In this condition, the depth judgment bias towards the observer was reduced by a significant amount. This phenomenon appears to strengthen the proximal vergence explanation.

Rolland et al. [20] studied egocentric depth perception for a pair of objects for real/real, real/virtual, and virtual/virtual conditions. The objects were placed side-by-side and participants were asked to give the position of the right side object. Their experiment concluded that virtual objects were perceived systematically farther away than real objects. However, this is one of the only studies to conclude this; almost all other studies have found that virtual objects are judged to be closer than real objects at the same distance.

---

[2]Biocular stereo: In this two identical images are presented to the eyes, but offset to indicate constant disparity.

[3]Binocular stereo: In this two stereoscopic images, which are generated from the perspective of each eye, are presented to the eyes.

Figure 2.1

Superman's x-ray vision problem

An application of AR is to visualize hidden infrastructure in multiple layers of virtual objects. However, this presents the problem known as *Superman's x-ray vision* (see Figure 2.1). If all of the objects situated at different depths are presented, there will be too much information for a viewer to make sense of it, and if only the objects of interest are presented, then the observer may not have enough contextual information to understand the depth relationships among them. The major question in AR is how to represent objects so that the observer can correctly identify the depth layers, and understand the depth relationships among them.

Livingston et al. [13] tested techniques to visualize multiple layers of objects. They tried to find display attributes and opacity settings that can correctly convey the depth relationship among various virtual objects. They found that *intensity* was the most powerful

graphical encoding for depth relationships, followed by drawing style and opacity. They also found that the presence of a ground plane strongly affected depth perception in AR.

2.3 Dark Vergence and Depth Perception

Gogel et al. [9] proposed that the perceived distance of an object will be biased towards an observer's dark vergence distance. This phenomenon is known as the *specific distance tendency*. If an object is in front of the dark vergence point, it will be perceived towards that point and hence farther away than it actually is (see Figure 2.2). Similarly, if an object is behind the dark vergence point, it will be perceived closer than it actually is.

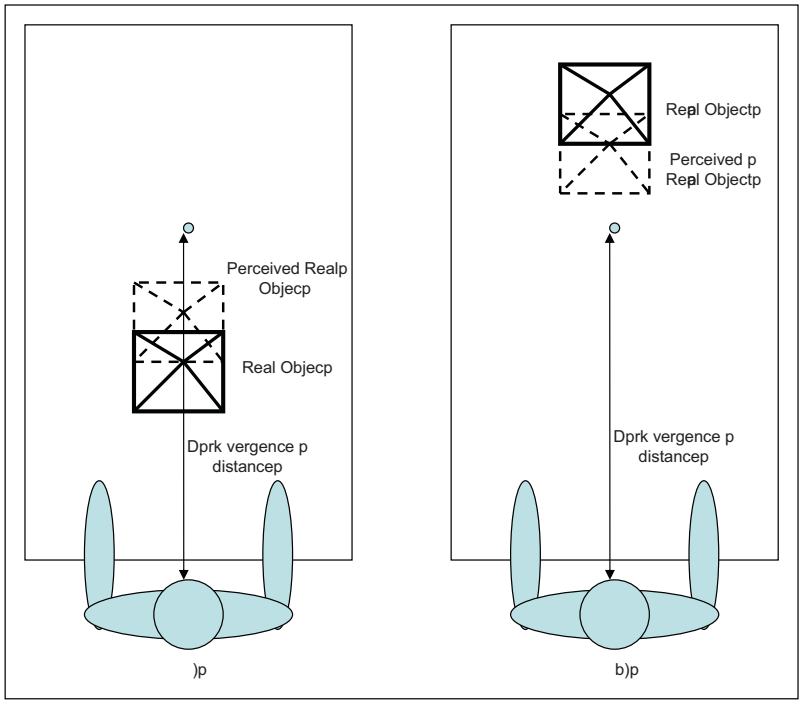

Figure 2.2

Dark vergence: The depth of an object is biased towards the dark vergence point. (a) The bias causes overestimation of distance, (b) the bias causes underestimation.

Owens et al. [19] found a positive correlation between dark vergence and perceived distance and suggested that perceived distance to an object is influenced by dark vergence. The farther away one's dark vergence point is from an object, the more biased the object's perceived location.

CHAPTER 3

EXPERIMENTAL SETUP

This experiment empirically studied how egocentric depth perception operates in near-field AR. It involved measuring depth judgments of virtual objects in an AR environment using perceptual matching and blind reaching tasks. There were two overall goals for the research; first, to engineer an apparatus that allowed us to accurately measure depth judgments in the near visual field, and second, to test this system and check its effectiveness by conducting a depth perception experiment. The main focus of this research was on depth perception in the near visual field, which is the domain of some of the motivating applications.

To run this experiment, we needed some apparatus to perform depth judgment tasks in the near-field, which extends to arm's length around the participant, where objects can be easily grabbed and manipulated. A table-top setup was found to be most suitable for performing the reaching tasks. In addition, we wanted to replicate Ellis's et al. [7] near-field depth perception study, which also involved a table-top based setup. We decided to use a setup similar to the one used by Ellis. However, this apparatus is not something commercially available; therefore, we engineered this apparatus from scratch, by assembling various parts from different manufacturers.

3.1 Resources

This experiment used a setup which required a combination of various resources as described below:

3.1.1 Hardware

The hardware resources used for this research were:

- Display devices
- Tracking systems
- Dark vergence measurement device.

3.1.1.1 Display Devices

Some sort of display device was needed to show virtual objects to the participant in the AR environment. For our research, the most suitable display device was a Head Mounted Display (HMD). One such device available for use in our user study was an NVIS nVisor ST (see Figure 3.1). It is capable of displaying 3D stereo scenes at a resolution of 1280 X 1024. It has a 60 degree diagonal field-of-view and all the display area in it can be overlapped with the real world scene. It supports variable inter-pupillary distance (IPD) by adjusting the left and right eye pieces independently.

3.1.1.2 Tracking System

In AR, when virtual scenes are presented on HMD screens, the viewer's head location must be tracked in real time to dynamically change the display properties of the

Figure 3.1

NVIS nVisor ST head mounted display device

virtual objects so that they correspond to the viewer's current viewpoint. For head tracking, we used an Advanced Realtime Tracking GmbH *ARTtrack* position tracking system (see Figure 3.2). It uses optical tracking to provide three position values (X, Y, Z) and three rotational values (Roll, Pitch, Yaw). These values were used to adjust the view of the virtual world with respect to the position and orientation of the participant.

The ARTtrack is a passive outside-in position tracking system. It uses a combination of retro-reflective spheres and cameras to track the position and orientation of the target object. The cameras are attached at a fixed position in the environment and a rigid configuration of the retro-reflective spheres are attached to the target to be tracked. Each camera consists of an infra-red light transmitter and receiver. During tracking, infra-red light is emitted and is reflected off these spheres. This reflected light is captured by the cameras and from this the position and orientation of the target object is calculated. The ARTtrack can track multiple target objects at the same time by attaching different configurations of spheres to different target objects. Preliminary testing showed the spatial accuracy and precision of the ARTtrack to be better than 1 mm.

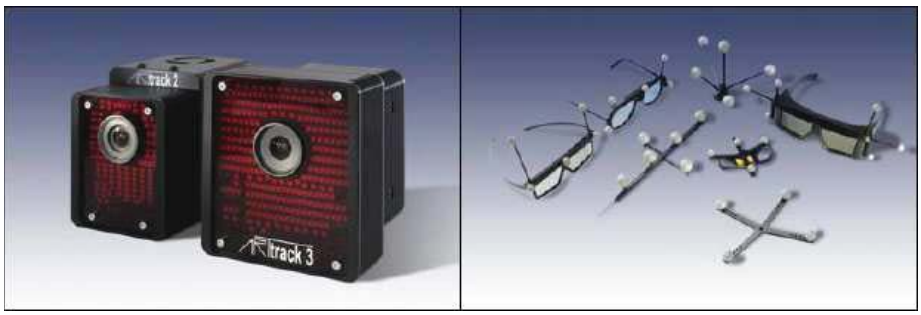

Figure 3.2

A.R.T. GmbH ARTtrack position tracking system.

3.1.1.3 Dark Vergence Measurement Device

To measure the dark vergence, we used a dark vergence measurement system (see Figure 3.3). This dark vergence apparatus is of our own design, and is based on a design described by Miller [17]. The apparatus displays two fast-flashing vertical lines, known as *Nonius* lines, on the monitor such that only one line is visible to each eye. In complete darkness, these lines are then adjusted until the participant perceives them to be vertically aligned, while keeping their head steady by placing it on a chin-rest at a certain distance from the monitor. In complete darkness, the eyes adopt the resting convergence state. Using the distance between the lines on screen, and the distance of the participant from the monitor, the dark vergence distance is calculated.

This system had been used in previous studies for collecting data from 50 subjects (Jones et al. [11]). We found the data collected by using this system to be accurate and to match the dark vergence values reported in the literature.

Figure 3.3

Dark vergence measurement system.

3.1.2 Software

The software to conduct this experiment consisted of several components, such as a program to generate stereoscopic 3D scenes, and a program to interface with the head tracker. The component to generate virtual scenes was developed in C++, OpenGL, and GLUT. These scenes were stereoscopic and rendered from two perspectives based upon the two eye positions, and then projected on the left and right screens of the nVisor HMD. The program that interfaces with the ARTtrack head tracker was developed using the API libraries provided by A.R.T. The data from the head tracker was read and the view of the virtual scenes was adjusted to match changes in the position and orientation of the participant.

3.2 First-Generation Depth Judgment Apparatus

The first version of table-top setup was engineered using two old tables scavenged from a computer lab (see Figure 3.4). One table surface was placed on top of the other table using four car-jacks at four corners. These car-jacks were used to adjust the height of the top table surface with respect to the lower table, according to the participant's height. A chin-rest was attached to the lower table.

Two sliders were attached to the table-top (see Figure 3.4); one above the table and used for the closed-loop perceptual matching task, while the other was under the table and was used for the open-loop blind reaching task. For the top slider, a white plastic PVC pipe was attached to the right side of the table surface. This pipe was attached to another L-shaped arm that extended to the middle of the table. At the tip of this arm a green LED was attached. While performing the matching task, participants slid the pipe in and out while viewing the target object as well as the LED at the same time. For the bottom slider, another PVC pipe was mounted to the center of the bottom of the table surface. A screw was attached at one end of this pipe, which was grabbed by participants to adjust the slider, while performing the blind-reaching task. On the left side of the table, an occluder was attached using a PVC pipe assembly. This occluder could also be slid in and out to appear at different depth positions from the participant.

Our first-generation table-top used a different tracking system from the one just described. It used an InterSense IS-1200 position tracking system with optical and inertial tracking to provide 6 degrees-of-freedom information. The IS-1200 uses a combination

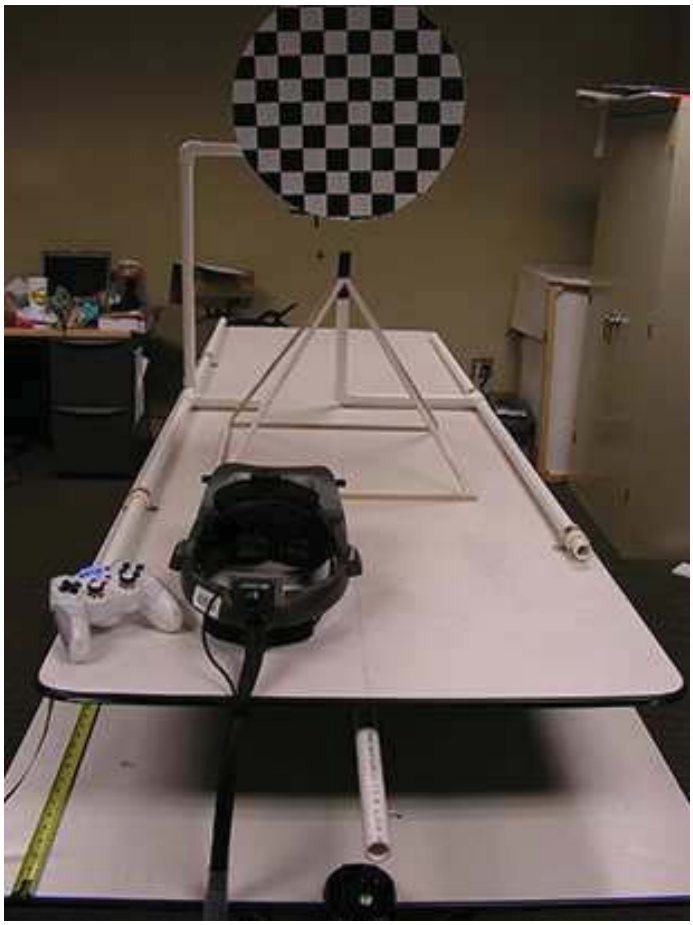

Figure 3.4

First-generation table-top apparatus.

of fiducials to track its position in space. We attached these fiducials on a wall behind the participant, when looking along the apparatus.

This first-generation table-top setup was inspired by the one used by Ellis et al. [7] and replicated their method for the closed-loop matching task. However, this design had many problems. The main problem was instability of the tables and problems resulting from it. The tables, though very heavy, were not stable enough and were prone to vibrations and misplacement caused by any accidental bumps. The car-jacks used to adjust the height of the top table surface were not mounted to the setup and hence they tended to shift any time the table height was adjusted. Another problem was manual height adjustment; whenever some participant wanted to adjust the height, the experimenter had to adjust the height of all four car-jacks separately, and it was very difficult to keep them equally aligned.

In addition, since the components such as the occluder and sliders were attached to the table using mounting screws, we needed to make new holes in the table surface everytime we decided to change the position of the equipment. This resulted in a series of unusable holes and markings on the table surface, which acted as a very salient depth cue while performing depth judgment tasks. We needed a better mounting mechanism along with reducing the amount of unwanted depth cues.

Another problem which resulted from the instability was tracking. Since the fiducials were mounted to a wall, they were in a fixed frame-of-reference. To keep a consistent view of the virtual world, it was necessary to keep the tables stable at one place, which took out any possibility of relocating the table. Every time the tables were adjusted, they tended to shift by a certain amount and hence the tracker needed to be recalibrated. Any

change in the table height also required entering an accurate height change value in the control software. A preliminary tracker data analysis showed an error of $\pm$ 2 cm, which was unacceptable for this experiment, since we expected significant effect sizes of $<$ 1 cm.

We wanted to study both the closed-loop perceptual matching task as described by Ellis et al. [7] as well as the open-loop blind reaching task, which needed to be performed under the table. Since this setup replicated Ellis's setup, it was suitable for the closed-loop task, but not good enough for the open-loop task. The participants were able to perform the closed-loop task without any problems, but could not perform the open-loop task as their arm movement was restricted by the top table and the chin-rest. Under the table, their range of motion for blind reaching was very limited as compared to the distance they could match above the table.

Because of these and many other problems, we decided to re-engineer our apparatus. This first-generation design helped us to understand and refine the requirements for our second-generation design. The first and most important requirement was that the new design accommodate both the closed-loop and open-loop tasks; it needed to allow an appropriate amount of movement for participants both above and below the table-top. The second requirement was that it needed to be stable and not allow any accidental shifting. It also needed to provide a better way of mounting and dismounting the sliders and occluder, and have a convenient way of adjusting the height. Another requirement was more accurate tracking, and it should not need to be recalibrated because of a change in the table's height and position. We incorporated all these changes in our second-generation design, as explained in the next section

3.3 Second-Generation Depth Judgment Table-top based Apparatus

Equipped with experience from the first-generation design, we engineered a new table-top setup (see Figure 3.5). Our first requirement was that the setup be stable and immune to accidental movements. We decided to use a custom-made *80/20 Industrial Erector Set* to build the table frame, which is very heavy ($\approx$ 300 lbs), stable, and provides vibration resistance. Our resulting table is 244 cm long, 92 cm wide and 100 cm high with caster wheels for movement. It has six legs, with four legs at the corners and two legs in the middle.

To mount the equipment on the table we decided to use an optical breadboard as the table top. It serves two purposes: first, it has a regular grid of bolt holes, which solves the problem of mounting and dismounting the equipment, and second, it added some more weight ($\approx$ 165 lbs) to the overall structure, making it more stable. The optical breadboard was custom-made, 244 cm long by 92 cm wide by 3.8 cm thick, made of two 1/4" thick aluminum skins laminated to a 1" nonmetallic honeycomb core for resonance damping. Both the top and the bottom of the breadboard have a 25 mm grid of M6 holes for mounting equipment. In order to prevent the grid from giving a strong perspective depth cue, we attached a matte black cloth to the top of the table, and equipment was mounted through the cloth (see Figure 3.5). We mounted a second particle-board table top (retained from our first-generation design) onto horizontal cross-beams 48 cm off the ground. This second table surface is used to store equipment such as the tracking system computer.

For height adjustment, we used a motorized hydraulic Dyna-Lift jack, where six hydraulic jacks are mounted to the six table legs, and the compressor is attached to the bottom

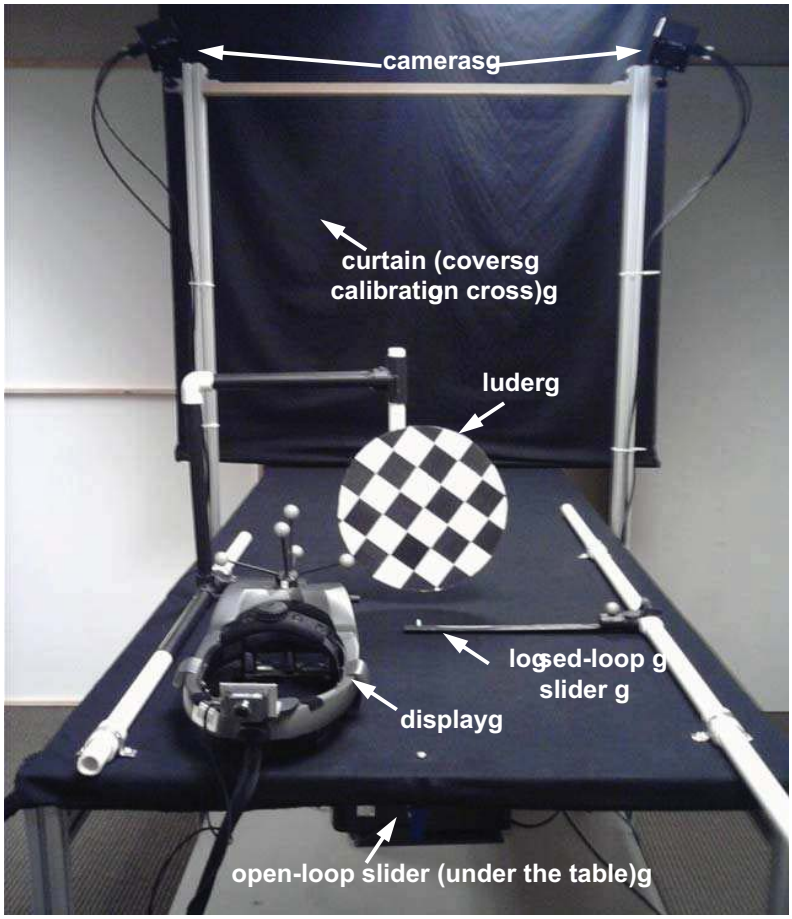

Figure 3.5

Second-generation table-top apparatus.

of the table top. This can lift the entire table apparatus, such that the table top is between 104 and 134 cm above the ground. The table height was adjusted so that sitting participants could comfortably rest the front portion of the AR display on the top of the table. This height adjustment and placement of the AR display on the table increased the range of movement of the participant's arm under the table, as now the table-top was at nose level and there was no need to use a chin-rest. This solved the problem of incorporating both the closed-loop perceptual matching as well as the open-loop blind reaching tasks in the experimental design.

The next problem to solve was tracking. Since the InterSense IS-1200 tracker used fiducials to track and these fiducials were fixed on a wall, even a slight movement of the table would throw off the tracking. We decided to solve this problem by two means: first, we used a new state-of-art ARTtrack tracker, and second, we decided to fix the tracker cameras to the table itself, so that there is a constant frame-of-reference transformation between tracker space and world space. We attached the tracker cameras to the two middle legs of our table; these legs extend 104 cm above the table surface (see Figure 3.5). By mounting the cameras to the table itself, the tracking does not have to be recalibrated when the table top is raised or lowered, or if the table is moved from one place to another.

To enable the perceptual matching and blind reaching tasks, we attached two sliders to the table-top: one above the table and one under the table. For the top slider, a length of white plastic PVC pipe was slid through two collars that were attached to the right side of the table surface. We used a T-joint to attach this slider to an arm that extends to the middle of the table, with a green LED light attached at its tip. While performing the perceptual

matching task, participants could slide the pipe in and out while seeing the target object as well as the LED at the same time. For the bottom slider, another length of PVC pipe was slid into two collars mounted to the center of the bottom of the table surface. At the front end of the pipe is a right angle bracket; this was grabbed by participants to adjust the slider while performing the blind-reaching task.

The position of the top slider was tracked using the ARTtrack; we attached a retro-reflective sphere to the top slider. Only one sphere was used because only the z-value (the distance of the LED from the participant) was required. This sphere was attached to the top slider at the same distance as the LED was from the participant. Being attached to the same slider and at the same distance, both the LED and the sphere moved to the same distance when the participant adjusted the slider. Any z-value for this sphere represented the distance of the LED from the participant.

Because the bottom slider is under the table and hidden from the tracker cameras, it was not possible to track it automatically. Instead, we measured the position of the bottom slider manually. We measured it using an assembly of a laser beam and a meter stick. A meter stick was mounted under the table at a distance of 75 cm from the front end of the table. A laser light was attached to the bottom slider at a distance of 78.2 cm from the front end of the slider (see Figure 3.6). It was attached so that whenever participants slid the pipe, the laser light projecting on the stick gives the position of the slider.

The stimulus presented in this experiment replicates that of Ellis et al. [7]; it was a white wireframe pyramid with a base of 10 cm and a height of 5 cm. The pyramid was displayed upside-down, and slowly rotating at a rate of approximately 4 revolutions per

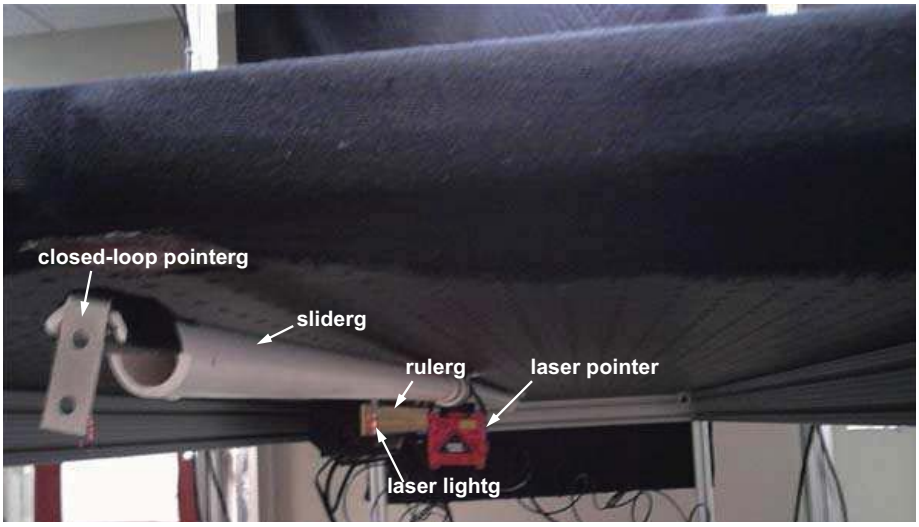

Figure 3.6

Open-loop slider setup

minutes. The size of the pyramid was randomly scaled from 70 % to 130 % of its actual size for each trial, to avoid possible usage of angular size as a depth cue by the participants. For each trial, the pyramid was presented at one of the five distances of 34, 38, 42, 46 and 50 cm from the participant using the head mounted see-through display. During the trials, participants were asked to match one of the pointers to the apex of this pyramid.

To study the effect of an occluding surface on depth perception, an occluder was introduced in the scene as described by Ellis et al. [7]. As shown in Figure 3.5, it was a circular foam disc with a diameter of 29 cm and had a checkerboard pattern of black and white 5 cm paper squares glued onto it. The center of the disc was mounted onto a small motor that rotated at a rate of 2 revolutions per minute. This disc and motor assembly was attached to a pipe that ran through two collars mounted on the left side of the table. With this mounting, the occluder could be positioned either in or out of the participant's field

of view. When the occluder was in the field of view, it could be positioned at a range of distances from the participants; however for this study, the occluder was always presented at a distance of 42 cm in front of the participant. This distance was selected so that two of the trial distances would be behind the occluder, two would be in front of the occluder, and one trial distance would be at the same as the occluder. The later two conditions replicated the *In-front* and *On* conditions of Ellis's experiment. This occluder was as salient as possible as the black and white checkerboard pattern contains many strong accommodative cues, and the slow rotation further enhances the salience.

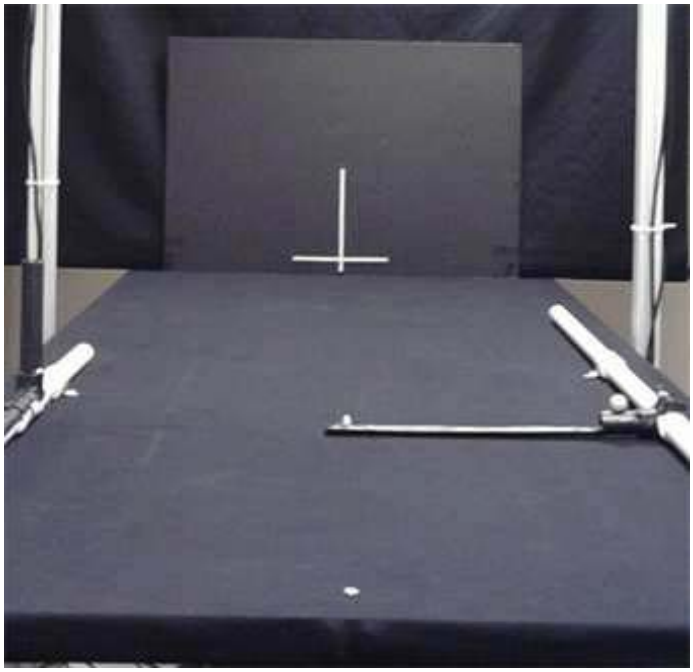

Figure 3.7

Calibration cross.

For calibration, a black cardboard rectangle was mounted on the table at a distance of 220 cm from the front end of the table (see Figure 3.7)). A white cross was drawn on this black cardboard with a center 3.5 cm above the table surface, and in the exact center of the table width. During calibration, this cross was matched to a similar looking virtual cross rendered on the HMD monitors. Once the calibration was completed, this board was hidden by a black curtain hanging from the ceiling and 220 cm away from the front end of the table (see Figure 3.5). All of the stimuli were presented against this black cloth background during the actual trials.

CHAPTER 4

USER STUDY

To test the usability of the system, a depth perception experiment was conducted using the apparatus described in the previous chapter. It involved gathering and comparing data for an open-loop blind-reaching task and a closed-loop perceptual matching task. This setup was inspired by the one used by Ellis et al. [7] and replicated their method for the closed-loop perceptual matching task. By replicating Ellis's method, we were able to compare our results to his results, which tested the repeatability of their results and also served as a control condition for us. Along with the closed-loop task, we tested an additional condition, the open-loop blind-reaching task. This allowed us to study depth perception using an open-loop task at near-field distances, which is similar to the open-loop blind walking task that has been widely studied at medium-field distances. No previous experiment that we know of has collected data for open-loop tasks at near-field distances in AR.

4.1 Hypotheses

The hypotheses which were tested by conducting this experiment are:

Hypothesis 1: The first hypothesis was to compare the participant's performance of depth judgments for open-loop and closed-loop tasks. We hypothesized that the open-loop measure would be similar to the closed-loop measure. This hypothesis was suggested by

the results from Mon-Williams et al. [18] and Ellis et al. [7]. Mon-Williams et al. found that in real environments, participants are very accurate in judging depth with an open-loop blind reaching task at near-field distances. Ellis et al. also found very accurate responses in near-field AR for closed-loop matching tasks.

Hypothesis 2: The second hypothesis tested the effect of an occluder on depth perception. We hypothesized that the presence of an occluder would cause participants to judge distances differently then when the occluder was not present. However, we believed that participant performance under open-loop and closed-loop measures would still be the same. Ellis et al. [7] suggested that the proximal vergence induced by the occluder would cause a change in depth perception by either underestimating or overestimating the actual distance.

Hypothesis 3: This hypothesis involved the relationship between dark vergence and the perceived depth of an object. Dark vergence has influenced the depth perception of real objects (Gogel et al. [9]) and we hypothesized that it would have a similar influence on virtual objects. We hypothesized that the participants would perceive the depth of the virtual objects to be biased towards their dark vergence distance; however, participant performance for open-loop and closed-loop measures would still be similar.

4.2 Variables and Design

Table 4.1 lists the independent and dependent variables and the design of this user study.

Table 4.1

Independent and Dependent Variables

INDEPENDENT VARIABLES		
participant	16	(random variable)
judgment	2	closed-loop, open-loop
occluder	2	present, absent
distance	5	34, 38, 42, 46, 50 cm
repetition	3	1, 2, 3
DEPENDENT VARIABLES		
judged distance		in cm
error		*judged distance - distance*

4.2.1 Independent Variables

The depth judgment conditions and occluder conditions have already been described. For experimental trials, participants were presented with the virtual object at distances of 34, 38, 42, 46, and 50 cm. Participants saw 3 repetitions of the trial distance for each combination of other independent variables.

4.2.2 Dependent Variables

The primary dependent variable was *judged distance*, which was recorded from the participant's responses. It is the distance where the participant placed the physical pointer. From judged distance, *error* = *judged distance* - *distance* was calculated. An *error* ≈ 0 cm indicates an accurately judged distance; an *error* > 0 cm indicates an overestimated distance; and an *error* < 0 cm indicates an underestimated distance. Judged distance and error have been widely studied by the depth perception community.

4.2.3 Experimental Design

In this experiment, a repeated-measures within-subject experimental design was used along with a factorial nesting of the independent variables. The variables varied in the order that they are listed in Table 4.1: the *judgment* condition varied the slowest; within each judgment condition, two *occluder* conditions were displayed. The presentation order of each *judgment* $\otimes$ *occluder* block was randomized using a 4×4 Latin square to counterbalance any temporally-based confounding effects. For every *judgment* $\otimes$ *occluder* block, 5 (*distance*) $\times$ 3 (*repetitions*) = 15 trials were generated. There were a total of 16 (*participants*) $\times$ 2 (*judgment*) $\times$ 2 (*occluder*) $\times$ 5 (*distance*) $\times$ 3 (*repetition*) = 960 data points.

4.3 Screening

All the procedures used in this experiment were reviewed and approved by the Internal Review Board (IRB) in the Office of Regulatory Compliance. Before beginning the experiment, each participant filled out a standard informed consent form, a simulator sickness survey, and then a brief demographic survey about their vision, computer usage, and experience with video games. They were checked for any medical condition, through self reporting, which could have interfered with their performance in the experiment. Participants with a history of epilepsy and head or neck injury were not permitted to participate in this study. Before the main experiment, participants performed a stereo acuity test. Only participants with normal or corrected-to-normal stereo vision performed the study. After the experiment, participants again filled out the simulator sickness exit survey. No participant had epilepsy or head or neck injuries, and all passed the stereo acuity test.

4.4 Measurements

Before the experiment, some measurements were collected from the participants. The first measurement was inter-pupillary distance (IPD), which is the distance between the participant's pupils. This distance varies due to changing eye vergence when looking at different distances. By using a digital pupilometer, two IPD values were collected from the participants looking at distances of 100 cm and 40 cm. The IPD for 100 cm was used for the dark vergence apparatus, and the IPD for 40 cm was used for the table-top apparatus. These values were collected for every participant and entered in the control software to render the scene for each participant.

4.5 Dark Vergence Measurement

After the stereo test, the participant's dark vergence was measured using the dark vergence measurement apparatus described in the previous chapter. Participants were trained for the nonius-line task in a different room than the dark room. After training, they were asked to wear polarized glasses and were taken to the dark room with their eyes closed and holding on a walker to direct them. They were told that both rooms were very different in size and we did not want their knowledge of the instruction room to influence their perception of the dark room. The dark room was a normal room with some modifications to avoid any incoming light. We covered up all the surfaces which were emanating any light. There was some light bleeding through the corners of the light fixtures on the ceiling and through the corners of the only door. We used aluminum foil to cover up these corners. We also covered up any glowing LEDs and indicator lights on the equipment using aluminum

foil. All these modifications made the room sufficiently dark to perform the nonius-lines task. After experiment, participants were asked to rate the darkness of room from a scale of 1 to 10; 10 being the darkest. Most of the participants rated it to be nearly 10.

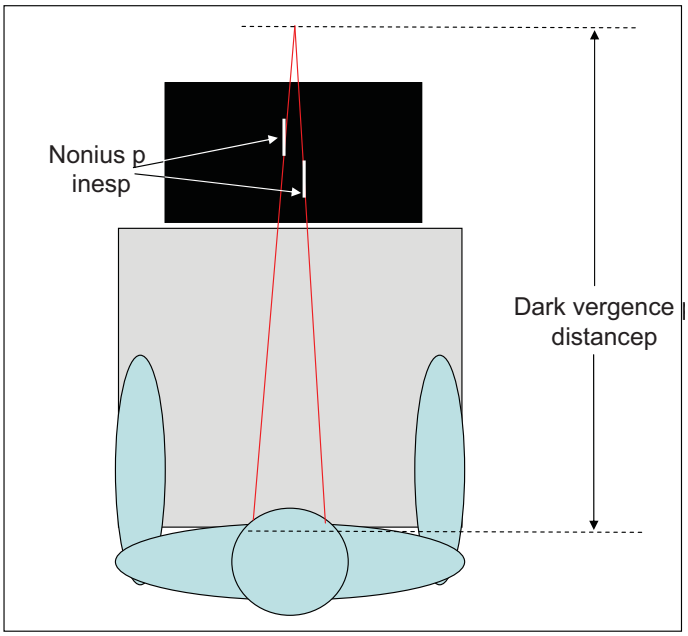

Figure 4.1

Nonius-lines task.

In the dark room, they sat on a chair, with their chin placed on a chin-rest. Though their eyes were closed, a black rectangular cardboard was present in between the participant and the monitor to avoid any knowledge of the location of the monitor. Once the control program started the experimenter asked the participant to open their eyes, let their eyes adjust to the darkness, and to try to not focus on anything. Then the experimenter took out the occluder, and participants could see the monitor screen. Participants were

presented with two flashing lines (see Figure 4.1); one on the top and other at the bottom of a computer monitor. A polarized film was fixed on the computer monitor in such a way that while looking through a pair of polarized glasses, participants would see the top line only in their left eye and the bottom line only in their right eye. The task was to align these lines vertically, where the top line was fixed and the bottom line was adjustable to the left and right. Participants were asked to look at the lines and tell the experimenter which way, left or right, to move the lower line. Once the lines were aligned vertically they would let the experimenter know and the next trial would begin. Participants performed 6 trials of the nonius-line alignment task.

4.6 Experiment

The experiment took place indoors in an artificially illuminated room. The main experiment started with participants sitting in a chair at one end of the table-top apparatus. They were asked to adjust the chair height to where it was comfortable for them. Next, they were asked to adjust the table height using a controller for the hydraulic lift, so that the table was at a level where their nose sat comfortably against a ridge on the edge of the table. A ridge made from thermocol was attached to the center of the front side of the table. The ridge was covered with a nosepad which was changed for every participant for sanitary reasons. Participants were told that they would be placing the front portion of the AR display on the table and their nose against the ridge, while performing the experiment, and that they could adjust the table or chair height at any time during the experiment.

4.6.1 Calibration

The next step was putting on and calibrating the HMD. The setup was calibrated based on the calibration procedure described by Jones et al. [10]. First, the participants put the HMD on their heads in a comfortable position. Then, the participant saw concentric circles and a white cross sign (see Figure 4.2), rendered according to the IPD of the participant. The participant was then asked to close their right eye and open their left eye and adjust the left eye-piece using a knob on the left side of the HMD. The participant adjusted the eye-piece until the outermost circle was cut off equally from both the left and right side. This was done to horizontally align the participant's left-eye optical axis with the HMD's left-eye optical axis. The same procedure was repeated for the right eye. Once both eye-pieces were adjusted, the optical axes of the HMD were horizontally in alignment with the optical axes of the participant's eyes. Next, participants were asked to adjust the HMD vertically, using a knob on top of their heads, which moved the entire display up and down. They adjusted the display until the outermost circle was cut off equally from top and bottom; after this the optical axes of the HMD were vertically aligned with the optical axes of the participant's eyes.

The participant was then asked to place the front portion of the display on the table so that their nose was exactly in the center of the table. The participant was asked to keep their heads steady and their nose in the ridge while performing the tasks. This was done to restrict any lateral head movement, and avoid using motion parallax resulting from the motion of their heads as a depth cue. The height of the center of the eye-pieces from the table was 3.5 cm when the front of the display was placed on the table. This distance was

used to draw the white cross on the calibration board, where its center was located 3.5 cm above the table surface.

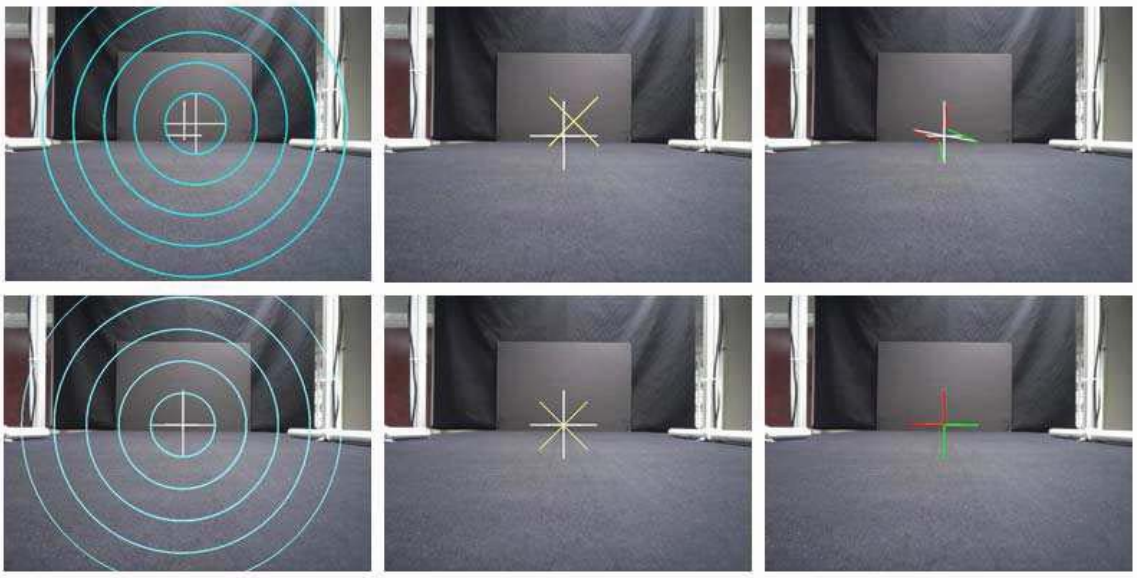

Figure 4.2

Calibration: (a) Aligning real and virtual world. (b) Translational error correction. (c) Rotational error correction.

The next step was to correct any translational or rotational errors caused by variation in the way the HMD fits on different participant's heads. The participant was asked to adjust their head so that the virtual white cross on the display aligned with the white cross on the calibration board (see Figure 4.2 (a)). This alignment provided a registration of the virtual world with a real world object. Next, translational errors were corrected along the x and y axes. This was done because every participant wears the HMD in a different way, which causes translations error horizontally or vertically, i.e. along the x and y axes.

Since the HMD monitors remain at a constant distance from the participant's eyes (along the z-axis), it was not necessary to consider z-axis errors. The participants were presented with a yellow 'X' sign, which reflected the x-axis and y-axis values of the participant's eye position in space. Then participants were asked to move the yellow X, using a game controller, until the center of the yellow X was in line with both the virtual white cross and the real cross on the calibration board (see Figure 4.2 (b)). This alignment removed any possible x and y axes translational errors between the real and virtual world.

Next the rotational tracker errors were corrected along yaw (left/right) and pitch (up/down) movements. The roll (twist) error was not considered as it was not important for the study, and its effects, if any, would have been negligible for this task. A 3D compass (Jones et al. [10]) was presented at a distance of 220 cm from the participant at the calibration cross. This compass was rotationally controlled by the tracker. The participant was asked to adjust it until it became a plus sign and covered both the white virtual cross and the real calibration cross (see Figure 4.2 (c)). This adjustment aligned the real and virtual world in the same direction by correcting any rotational tracker error. Once the calibration was done, the calibration board was covered with the black curtain and the actual task began.

4.6.2 Experimental Task

After calibration, participants were asked to perform a matching task, where they matched the distance to the stimulus (an upside-down virtual pyramid) by using a pointer; with either the open-loop or closed-loop task described earlier. We explained the task in a conversational manner by the experimenter using a real pyramid as a prop. They were told

that every trial would begin with the instruction *"observe the object and reach to it"*, and after hearing this, they would look at the pyramid and adjust the pointer until it was at the same distance from the participant as the apex of the pyramid. Participants were also told that they should let go of the pointer once they were done, and place their hand in their lap. This served two purposes: one, it let the experimenter know that the participant was done, and two, it gave participants a chance to rest their arm muscles. Three practice trials were given for each of the open-loop and closed-loop conditions before putting on the HMD.

4.6.2.1 Closed-loop Perceptual Matching Task

For the closed-loop perceptual matching task (see Figure 1.18), participants sat at one end of the table while keeping their nose in the ridge and the front portion of the AR display resting on the table-top. Then they were presented with the virtual pyramid. Participants adjusted the top pointer using their right hand until they perceived the LED to be exactly under the apex of the inverted pyramid (see Figure 4.4). The top pointer was tracked, therefore, once the participants acknowledged that they were finished, the experimenter took a reading of the pointer position using the tracker data. The above process was repeated for all closed-loop trials.

In the closed-loop perceptual matching task with the occluder (see Figure 4.3), the rotating occluder was presented at 42 cm in front of the participant. Its height was adjusted so that participants could see the apex of the pyramid against the lower rim and just above the physical pointer, while the rest of the pyramid overlapped the occluder (see Figure 4.4).

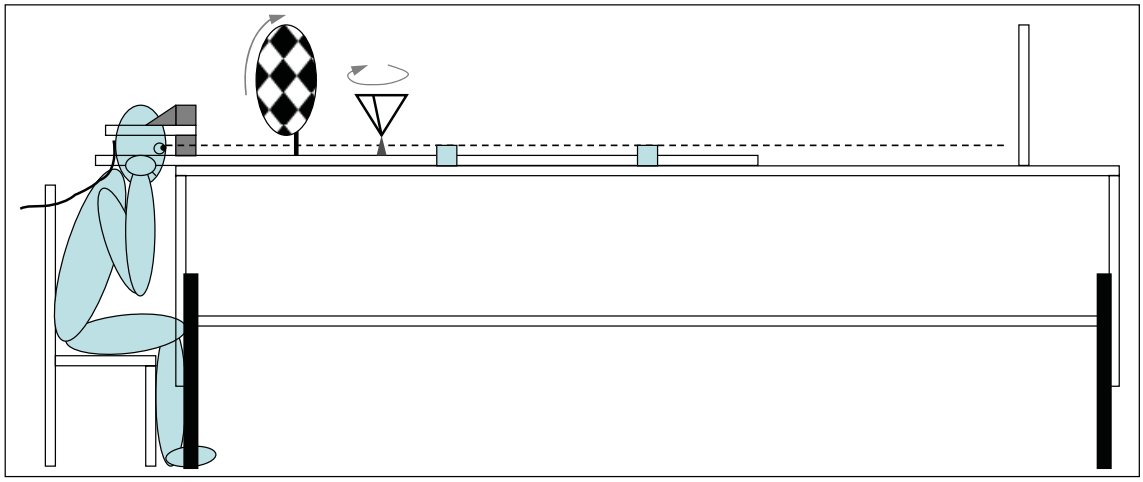

Figure 4.3

Closed-loop matching technique with occluder.

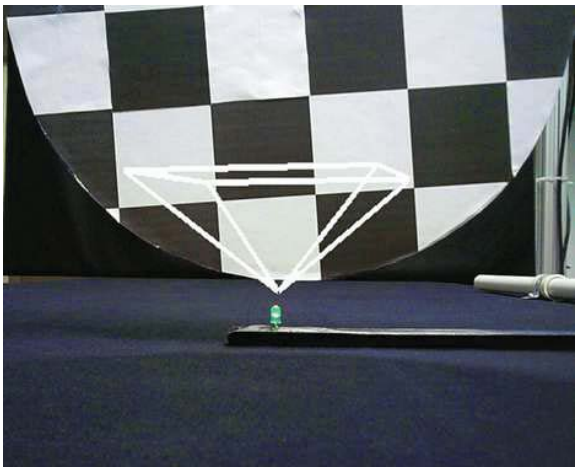

Figure 4.4

Occluder with pyramid and closed-loop pointer.

Also, the occluder was high enough so that the top pointer could easily move just below the lower rim of the occluder.

4.6.2.2 Open-loop Blind Reaching Task

For the open-loop blind reaching task (see Figure 1.19), the participant was presented with the virtual object above the table, similar to the closed-loop task. Then the participant was asked to grab the right-angle bracket attached to the lower pipe with their thumb pointing upwards. The participant next slid the bottom pipe until their thumb was directly underneath the apex of the pyramid. Once the participant acknowledged that they were done, the experimenter took a reading manually from the meter stick.

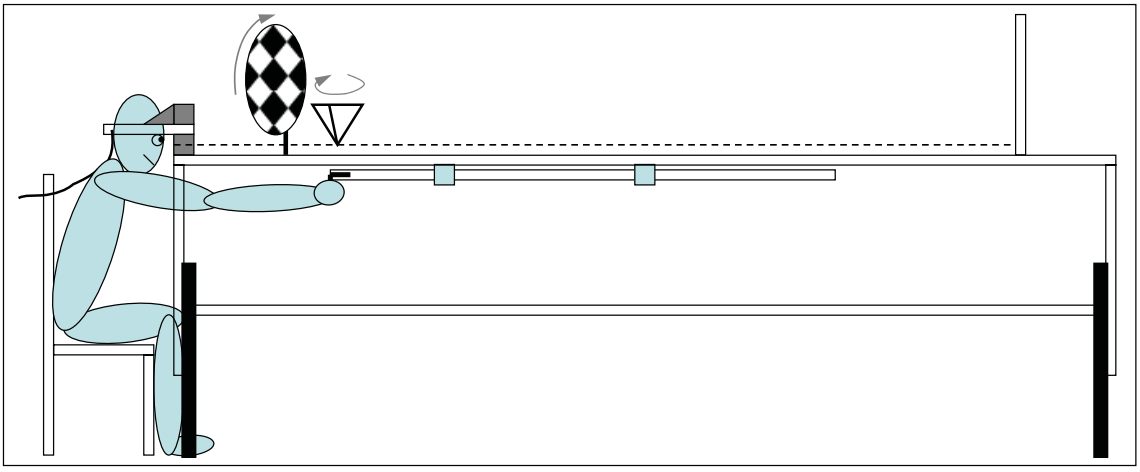

Figure 4.5

Open-loop blind-reaching technique with occluder.

In the open-loop blind-reaching task with the occluder (see Figure 4.5), participants performed the same task of reaching under the table until they felt that their thumb was directly underneath the apex of the pyramid.

4.7 Participants

A total of 18 participants performed the experiment, which were recruited from a population of university students, faculty and staff. The participants ranged in age from 19 to 30; the mean age was 21.6, and 7 were female and 11 male. Four of the participants were paid $12 an hour, and the rest received course credit. Data from 2 participants was not used because one participant did not perform the experiment seriously and the second participant told us after the experiment that she did not perform it correctly. Data from 16 participants was used, and a total of 960 data points were collected and analyzed.

CHAPTER 5

RESULTS AND ANALYSIS

The data analysis started with verification of the data, which involved looking for outliers which could affect the overall results and any specific participants who might have been an inordinate source of outliers. Then the remaining data (N = 960) was analyzed for depth judgment effects. We found that there was a learning effect present, and while performing the task participants were learning the task during the first repetition of the trial distances and using that knowledge in the second and third repetition. This finding resulted in culling out the first repetition, and re-analyzing the stable data just for the second and third repetition (N = 640). One interesting finding from this strategy was change in a three-way interaction of judgment, occluder, and distance. This interaction was not significant for all of the data, but became significant for the stable data. Lastly, we analyzed the dark vergence data.

5.1 Verification of Depth Judgment Data

To verify the data, standard histograms of data were plotted, which tested for normality of error distribution and any outlying data points. Figure 5.1 shows the histogram for the open-loop judgments. Here the distribution of error looks normal and there is no evidence of outliers. For preliminary analysis, we decided to consider all error values more than

±20 cm to be outliers. We also checked the results per participant, and found no evidence of any participant as a source of outliers (see Figure 5.2).

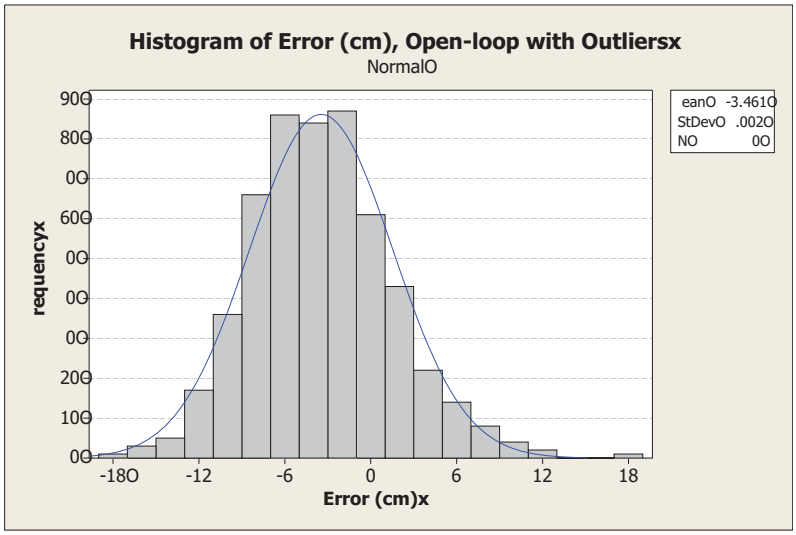

Figure 5.1

Histogram of open-loop data with outliers

Figure 5.3 shows a histogram for the closed-loop judgments. Here, although the distribution of error is normal, there are some overestimation outlying values (error $>$ 20cm) on the right side of the distribution. To find the source of these outliers, we examined the closed-loop judgments per participant to find out if there were any particular participants responsible for these outliers.

As shown in Figure 5.4, there were a total of 19 values greater than 20 cm. We concluded them to be outlying values. Also, examining histograms revealed that participants s4 and s12 were overestimating the distance and were responsible for 14 out of 19 out-

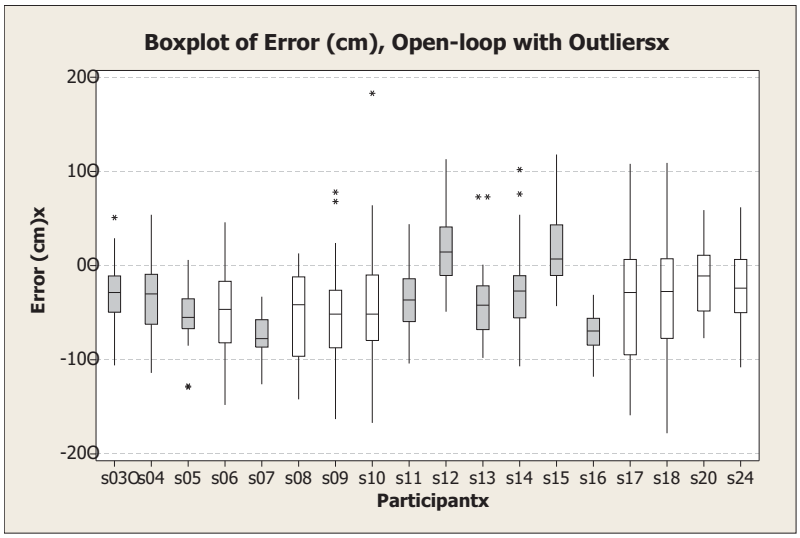

Figure 5.2

Box-plot of open-loop data per participant with outliers

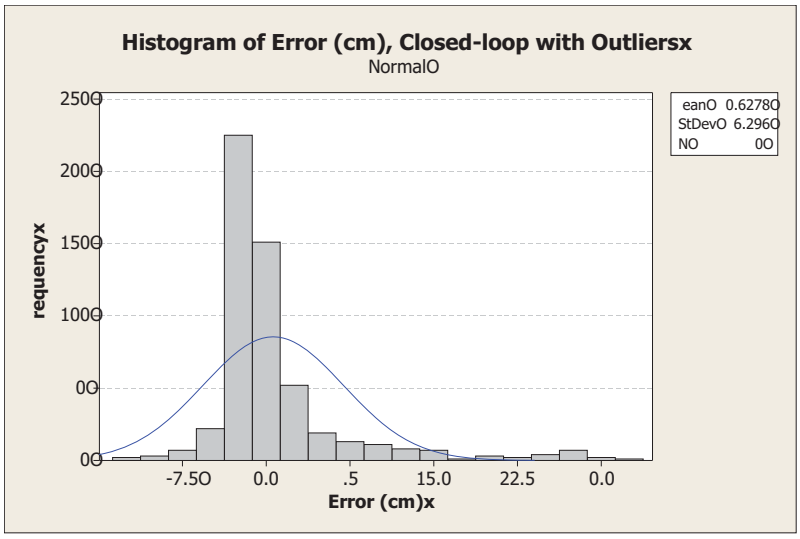

Figure 5.3

Histogram of closed-loop data with outliers

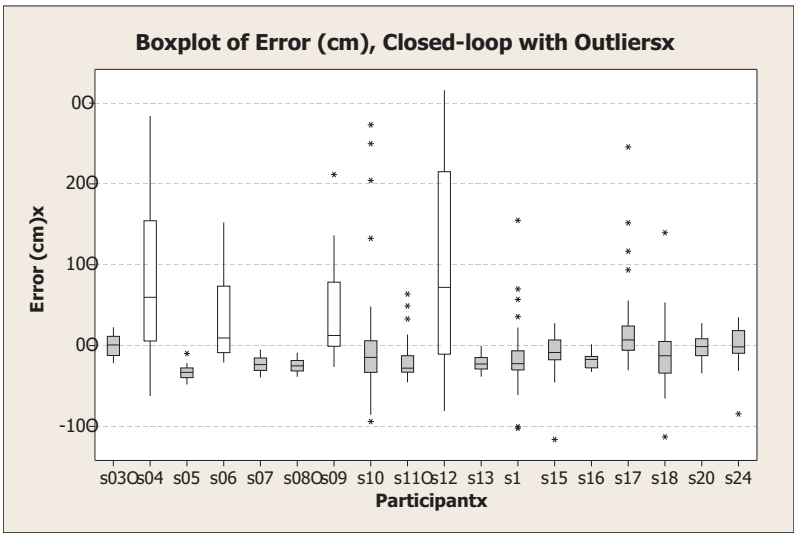

Figure 5.4

Box-plot of closed-loop data per participant with outliers

liers, while participants s9, s10, s17 have a total of 5 outlying values. In addition, one closed-loop data value was missing because of a tracker error.

After investigating the data, we found out that after performing the experiment, s4 reported that she was not judging the distance correctly. Also s12 was found to have not taken the experiment seriously. We decided to exclude data from these two participants and replaced s4 with s24 and s12 with s20 for further analysis. To fix 1 missing value and 5 outliers we used the procedure recommended by Barnett et al. [3]. These 6 values were replaced with the median of the remaining values in the *judgment* $\otimes$ *occluder* $\otimes$ *distance* experimental cell. The outliers were 5 out of 480 depth judgments = 1.04% of the values.

Note that after completing the ANOVA analysis described below, we re-ran the same ANOVA models with these 5 outliers included. This ANOVA model had exactly the same pattern of results as what is described below. There was however an additional main effect

of occluder on error ($F(1,15) = 5.03, p < .040$); we decided that this additional main effect was spurious because it was caused by the outlying data points. The F-value for the effect of occluder on error, after replacing the outliers, was ($F(1,15) = 4.36, p < .054$).

5.2 Depth Judgment Results over All Data

Figure 5.5 and Figure 5.6 show the histogram of open-loop data and closed-loop data after the replacement of the outliers and eliminating s4 and s12. There are no outliers present for both conditions. Although the open-loop data has a normal error distribution, the closed-loop data has a more positively skewed distribution.

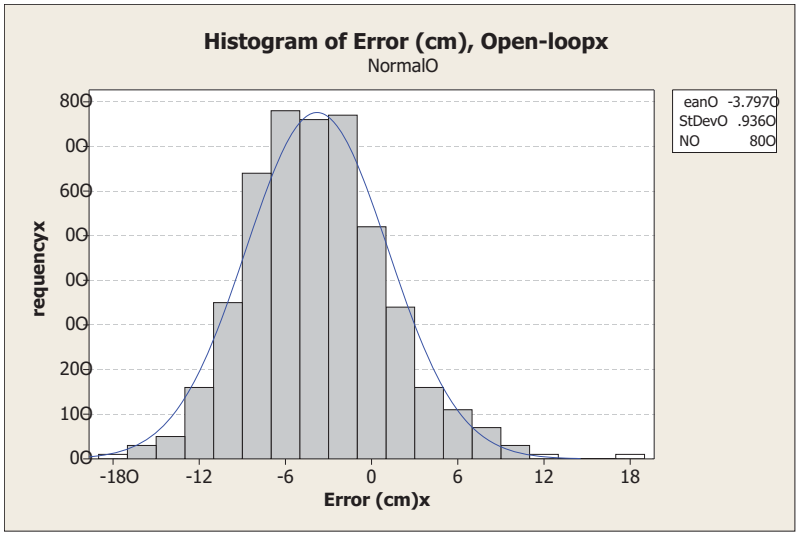

Figure 5.5

Histogram of open-loop data without outliers

Figure 5.7 shows the overall results for all $N = 960$ data values. This graph shows a lot of overestimated distances for the closed-loop judgments as compared to the open-loop

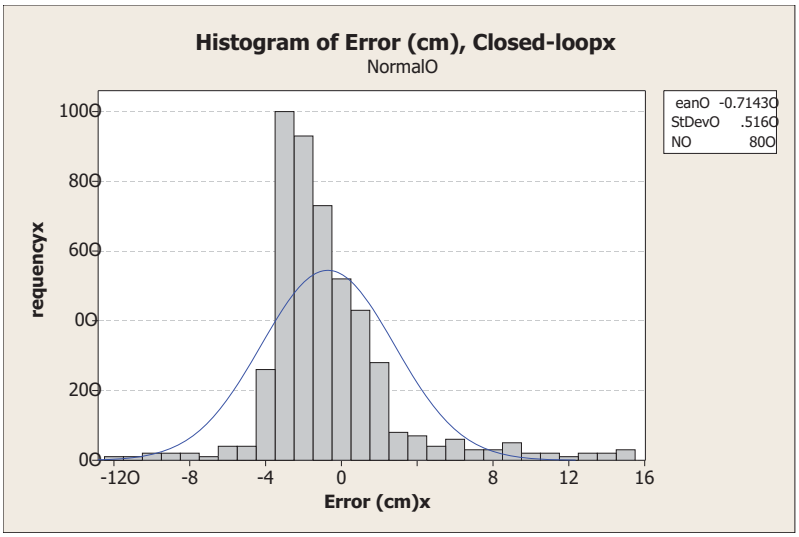

Figure 5.6

Histogram of closed-loop data without outliers

judgments. This explains the skewed distribution of the closed-loop histogram shown in Figure 5.6. However, this skew is understandable, as there was more distance available in which to overestimate than to underestimate.

Figure 5.8 shows the main results as mean judged distance versus actual distance. There is a trend of more underestimation for open-loop judgments as compared to closed-loop judgments. Also, depth judgments in the presence of the occluder have a different pattern than when it was not present.

Figure 5.9 shows the mean error for the overall depth judgments (N=960). The main result is a different pattern of means by distance for each of the four main *judgment* $\otimes$ *occluder* conditions. This is a 3-way interaction between judgment, occluder, and distance. The means are denoted with filled circles (●) and medians with hollow circles (○). As shown, the occluder was presented at a distance of 42 cm.

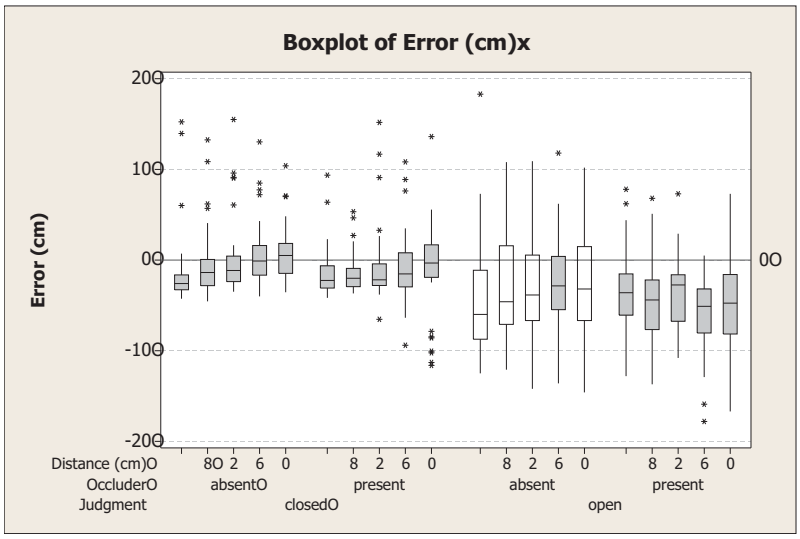

Figure 5.7

Box plot of all of the data.

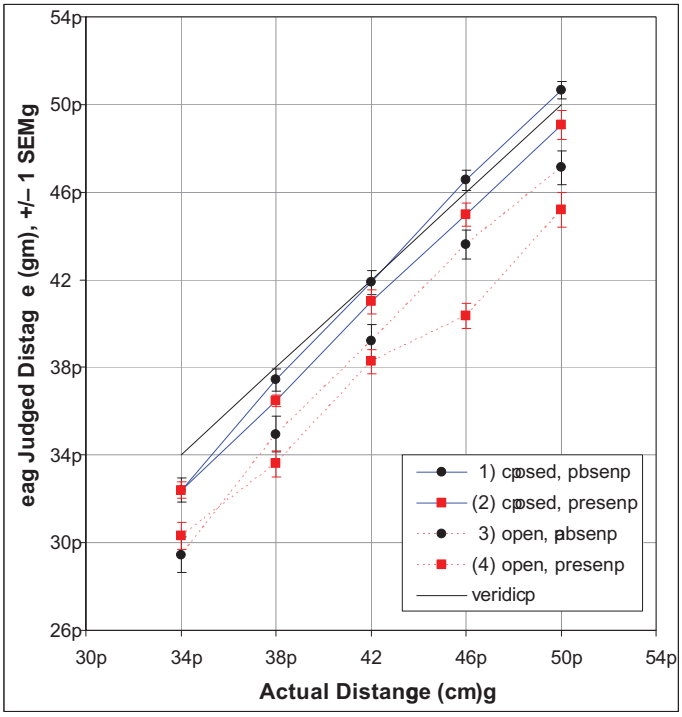

Figure 5.8

The mean judged distance versus the actual distance for all of the data.

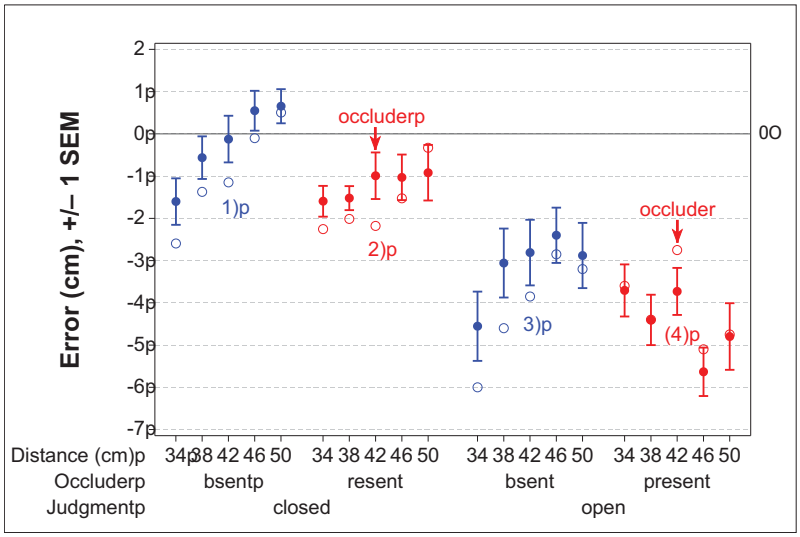

Figure 5.9

Interval plot of all data. means = (●); medians = (○)

5.2.1 General depth judgment effects on error

An ANOVA analysis was conducted on error for the main experimental design. Overall ANOVA results on error for all data are shown in Table 5.1, which include notable effects found during analysis. The shape of the data for the closed-loop data was positively skewed as shown in Figure 5.6 and Figure 5.7. Though the results in Table 5.1(a) are valid, there is a potential violation of one of the basic assumptions underlying ANOVA modeling: homogeneity of variance, which states that distribution of two samples from the same population should have similar variances. To test for this potential violation, we transformed the error with a base-10 log transformation, which made the distribution of the error normal, and then we analyzed the transformed data. Table 5.1(b) shows these results. The log-transformed error model displayed the same trends as displayed by the

error, which confirms the validity of our results on the untransformed data. Also, the ANOVA results for the transformed data generally revealed greater experimental power as compared to the untransformed data: the F values were generally larger, which would be expected from better meeting the ANOVA assumptions. We found the following notable effects:

Table 5.1

ANOVA results for all of the data with significant effects.

(a) ANOVA on ERROR						
SOURCE	SS	df	MS	F	p	
Judgment	2281.161	1,15	2281.161	20.988	0.000	***
Occluder	319.5726	1,15	319.5726	4.361	0.054	
Occluder*Distance	215.4476	4,60	53.8619	4.542	0.003	**
Judgment*Occluder*Distance	38.4604	4,60	9.6151	1.781	0.144	
Judgment*Repetition	98.53	2,30	49.265	3.447	0.045	*
Occluder*Repetition	165.7673	2,30	82.8836	5.361	0.010	*
(b) ANOVA on base-10 log-transformed ERROR						
Judgment	2.0052	1,15	2.0052	28.286	0.000	***
Occluder	0.2016	1,15	0.2016	4.075	0.062	
Occluder*Distance	0.2194	4,60	0.0548	6.930	0.000	***
Judgment*Occluder*Distance	0.0527	4,60	0.0132	3.551	0.011	*
Repetition	0.1183	2,30	0.0591	3.487	0.043	*
Judgment*Repetition	0.1308	2,30	0.0654	5.444	0.010	**
Occluder*Repetition	0.1296	2,30	0.0648	4.739	0.016	*

Lack of Judgment by Occluder by Distance interaction: The pattern in Figure 5.9 showed up as a three way interaction between judgment, occluder, and distance. This three way interaction was not statistically significant ($F(4,60) = 1.781$, $p < 0.14$), for all

data (N = 960); however, it became significant in our next analysis over the stable data, which is discussed below.

Main effect of Judgment: There was a main effect of judgment on error ($F(1,15)$ = 21.0, $p < .001$); closed-loop judgments are less underestimated than open-loop judgments. This effect rejected our first hypothesis, which assumed that closed-loop and open-loop judgments will have similar effects on error.

Occluder by distance interaction: There was an occluder by distance interaction on error ($F(4,60) = 4.5$, $p < .003$); in the occluder present and absent conditions, there is a difference in the shape of the patterns by distance in Figure 5.9.

5.2.2 Repetition effects and interaction

There were also some significant learning effects as shown in Figure 5.10.

Judgment by repetition interaction: Although we did not find a main effect of repetition on error, there was a significant judgment by repetition interaction on error ($F(2,30) = 3.4$, $p < .045$); the closed loop judgments had a relatively constant error of -0.71 cm, while open-loop judgments showed a learning effect of 1.2 cm between repetitions 1 and 2 (see Figure 5.10(a)).

Occluder by repetition interaction: There was also an occluder by repetition interaction on error ($F(2,30) = 5.4$, $p < .010$); when the occluder was absent there was a relatively constant error of -1.68 cm, while when the occluder was present the error decreased linearly from -3.74 cm to near equivalence with the occluder = absent judgments over the course of three repetitions (see Figure 5.10(b)).

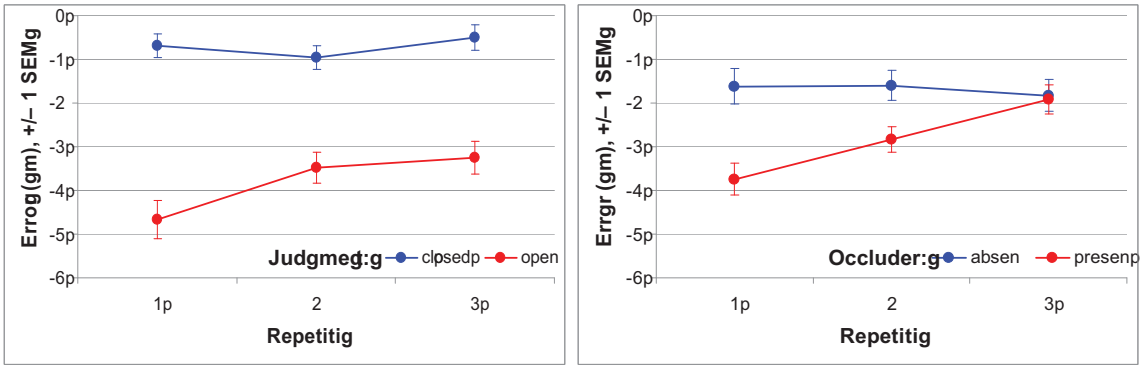

Figure 5.10

Interactions with repetition for all depth judgments. (a) Judgment by repetition interaction (b) Occluder by repetition interaction

5.2.3 Trial effects and interaction

The interactions by repetition indicated rapid learning by the participants. To understand these learning effects in detail and how the learning effect progressed over trials, we analyzed the 15 trials for each block. By performing another ANOVA analysis, we found the following effects:

Main effect of trial: There was a main effect of trial on error ($F(14,210) = 2.2$, $p < .010$)(see Figure 5.11). Participants became more accurate during the first 4 trials. After 4 trials, their performance remained relatively constant around error = -2 cm.

Judgment by trial interaction: There was a significant judgment by trial interaction ($F(14,210) = 2.0$, $p < .022$); the closed loop judgments had a relatively constant error, while in the open-loop judgments, participants showed a rapid learning effect over the first 5 trials (see Figure 5.12(a)).

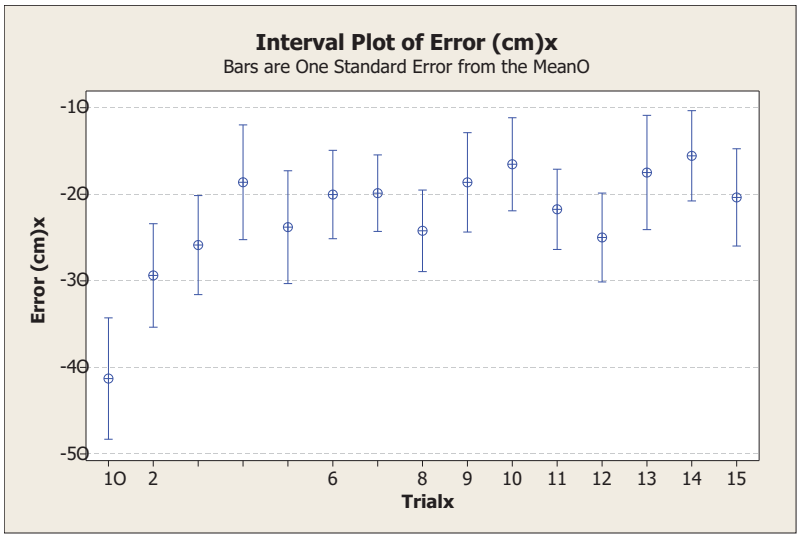

Figure 5.11

Effect of trial on error.

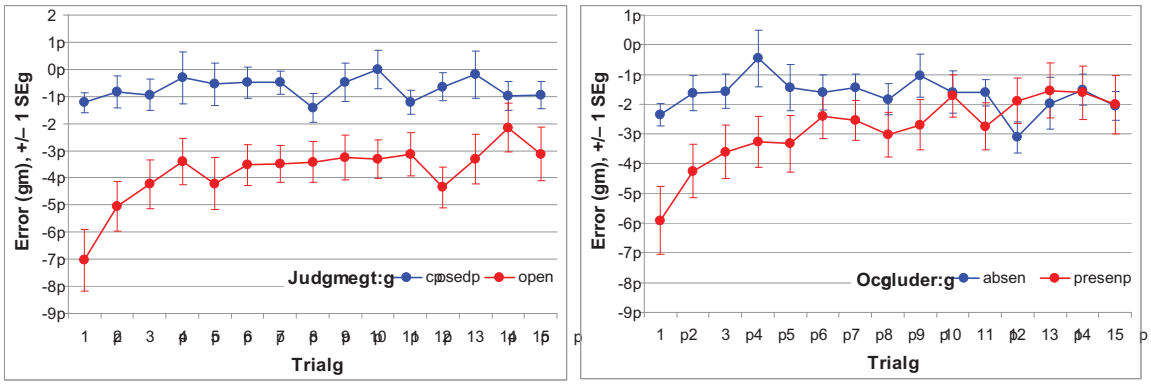

Figure 5.12

Interactions with trial for all depth judgments. (a) Judgment by trial interaction (b) Occluder by trial interaction

Occluder by trial interaction: There was also a significant occluder by trial interaction ($F(14,210) = 2.1, p < .012$); when the occluder was absent there was a relatively constant error, while when the occluder was present participants showed a rapid learning effect over the first 5 trials (see Figure 5.12(b)).

In both interactions, by trial 6 participants had more or less reached a steady state for the lower (red) responses (judgment = open-loop and occluder = present).

5.3 Depth Judgment Results over Stable Data

As discussed in previous section, there were learning effects present and a significant repetition interaction on error. Most of the learning was in the first repetition, and therefore we decided that the first repetition could not be considered a steady-state depth judgment, as participants were rapidly learning the task. Instead, we considered the second and third repetitions to be more accurate, and we analyzed them separately. We analyzed $N = 640$ data points after leaving out the data from repetition = 1; we refer to this as the *stable* data set.

In the histograms of the stable data, the distribution of error looks normal for the open-loop judgments, and the distribution is still positively skewed for the closed-loop judgments. However, both histograms looks similar to the overall data.

Figure 5.13 shows the main results as mean judged distance versus actual distance for the stable data. There is a trend of more underestimation in the case of open-loop as compared to closed-loop depth judgments. Also, depth judgments in the presence of the occluder have a different pattern than the case when the occluder is absent.

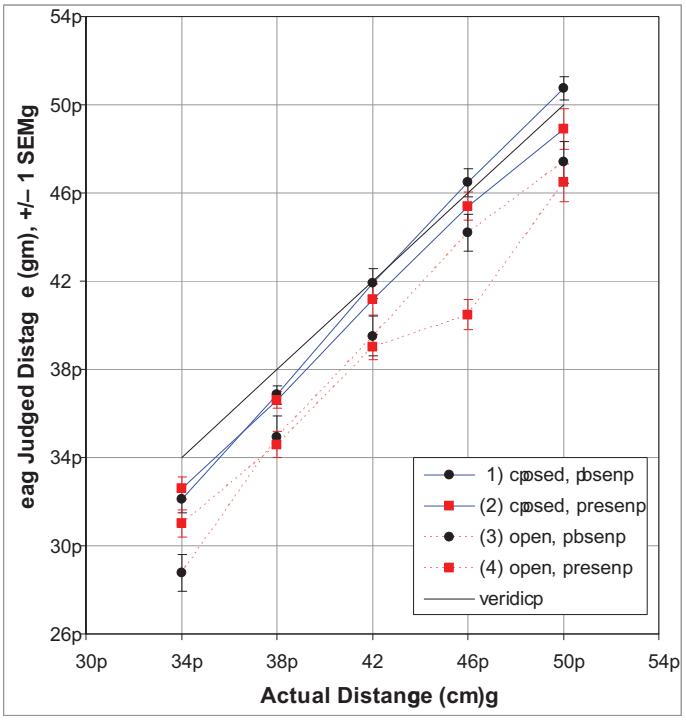

Figure 5.13

The mean judged distance versus the actual distance for the stable data.

Figure 5.14 shows the mean error for the overall depth judgments ($N = 640$). The main result is a different pattern of means by distance for each of the four main *judgment* ⊗ *occluder* conditions. This is a 3-way interaction between judgment, occluder, and distance. The means are denoted with filled circles (●) and medians with hollow circles (○). As shown, the occluder was presented at a distance of 42 cm.

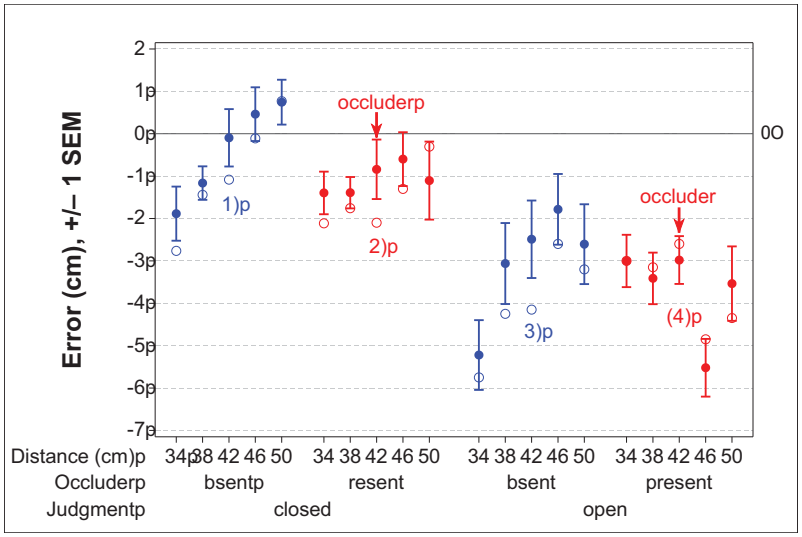

Figure 5.14

Interval plot of stable data.

Another ANOVA analysis was conducted on the stable data. Overall ANOVA results on error for stable data are shown in Table 5.2, which include notable effects found during analysis.

Judgment by Occluder by Distance interaction: Figure 5.13 and Figure 5.14 demonstrate a three-way interaction between judgment, occluder, and distance on error. This

Table 5.2

ANOVA results for stable data with significant effects.

(a) ANOVA on ERROR						
SOURCE	SS	df	MS	F	p	
Judgment	1110.9265	1,15	1110.9265	13.590	0.002	**
Occluder	71.4693	1,15	71.4693	1.364	0.261	
Distance	146.2858	4,60	36.5715	3.421	0.014	*
Judgment*Distance	50.6921	4,60	12.6730	0.958	0.437	
Occluder*Distance	248.2938	4,60	62.0735	4.642	0.002	**
Judgment*Occluder*Distance	88.5785	4,60	22.1446	4.295	0.004	**
Occluder*Repetition	52.5865	1.30	52.5865	4.167	0.059	

three way interaction is statistically significant here ($F(4,60) = 22.1, p < .004$) as opposed to over all of the data, where it was not significant.

This interaction of judgment, occluder and distance on error is a complex interaction as shown in Figure 5.14; there is a different pattern of means by distance for each of the four main conditions. In condition (1), the closed-loop condition with no occluder present, the judged distance shows a linear increase with increasing distance. The error varies from -1.89 cm to +0.77 cm, with almost accurate judgments at the distance of 42 cm (see Figure 5.14). In condition (2), the closed-loop condition with the occluder, the error remains relatively constant at -1.07 cm. Here, the presence of the occluder disrupts the linear pattern. A probable explanation of this phenomenon is proximal vergence induced by the occluder, which may have biased the convergence towards the occluding distance of 42 cm. This phenomenon also may have resulted in an increase in the underestimation of distances beyond the occluding distance. In condition (3), open-loop judgments with no occluder, the judged distance again shows a general pattern of increasing distance, with

relatively more underestimation for the distance of 50 cm. In condition (4), open-loop with the occluder present, the judgment is not symmetric as for the other conditions. Four of the five means cluster around an average error of -3.23 cm. When the target is at 46 cm, the distance to it was judged to be considerably closer to the participant, at an error of -5.52 cm. At this distance, the target was 4 cm behind the occluder. Here again it is possible that, similar to Ellis et al. [7], the incorrect occlusion cues which suggest that the target is in front of the occluder cause a change in convergence, which resulted in the underestimated distance judgment. However, when the target was at 50 cm, this effect is no longer operating, and the error is similar to the other open-loop depth judgments. This result is consistent with an observation made by both us and Ellis et al. [7]: when a virtual object is initially located in front of a physical object, and the physical object is slowly moved towards a participant, at first the virtual object appears to be pushed closer to the participant by the physical object. At some point, however, the virtual object suddenly appears to fall back behind the physical object, which imparts a strong sense of transparency to the physical object. This effect is easy to see in an AR system using one's hand.

Finally, note that there is a general trend of greater underestimation for the open-loop judgments (error = 5.52 to 1.78 cm) relative to the closed-loop judgments (error = 1.89 to +0.77 cm). This result rejects our first hypothesis, which stated that participant performance will be similar for open-loop and closed-loop conditions. We did not find any significant main effect of occluder on error; however, the significant 3-way interaction between judgment, occluder, and distance is a validation of our second hypothesis. Within

each judgment type, although the overall mean is similar whether the occluder is absent or present, the pattern of means by distance varies considerably. This is the main effect of the experiment.

5.4 Dark vergence analysis

In previous studies, it has been found that one's dark vergence affects the perceived distance of an object (Gogel et al. [9]), and also there is a positive correlation between dark vergence and perceived distance (Owens et al. [19]). These previous studies were done in the real environment with real objects. We expected to find similar effects in AR when the stimulus was a virtual object. We hypothesized that there would be a correlation between the dark vergence and the perceived distance, and the participants would perceive the depth of the virtual objects to be biased towards their dark vergence distance.

Figure 5.15 shows the dark vergence measurement per participant. Here, 7 out of 16 participants converged towards 1 meter distance, and 13 out of 16 participants converged inwards i.e. closer to themselves. While analyzing the dark vergence data, we looked for an effect of dark vergence on error but we could not find any.

One possible reason that we have not found any effect, even though Gogel et al. [9] did, is the difference between the stimuli of these two experiments. In their experiment, Gogel et al. [9] used a point of light as stimulus, while our stimulus was a pyramid with well-defined shape and boundaries. Also the pyramid was rotating, which made it even more salient. Therefore, our stimuli had relatively a richer set of depth cues, which could have caused participants to converge correctly on the stimuli, negating any effects of dark

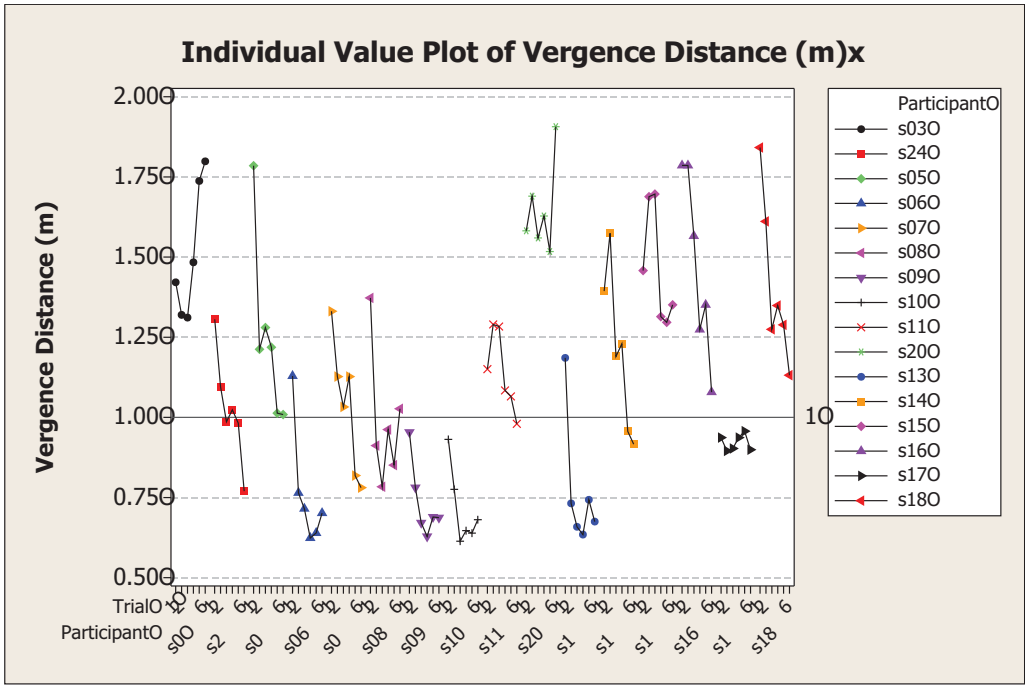

Figure 5.15

Dark vergence data for all participants.

vergence. Even though we could not find any effect of dark vergence on error in our analysis, some future analysis of this data may reveal some correlation between dark vergence and depth judgments.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

We have successfully developed an apparatus and related calibration and measuring techniques for collecting near-field depth judgments, using both closed-loop and open-loop tasks.

Our closed-loop results largely replicated the relevant results reported by Ellis et al. [7]. In addition, we directly compared closed-loop perceptual matching to visually open-loop blind reaching in the same experimental context. We found that blind reaching is significantly more underestimated than perceptual matching, with an average additional error of 3.1 cm for distances of 50 cm or less. This result rejected our first hypothesis, which stated that participant performance will be similar for open-loop and closed-loop conditions. This result also suggests that AR-presented virtual objects may be perceived as being somewhat closer than they are perceptually matched with closed-loop tasks. We also found that the presence of a highly-salient occluding surface has a complicated effect on depth judgments, but it does not lead to systematically larger or smaller errors. This result confirmed our second hypothesis, which stated the presence of an occluder would cause participants to judge distances differently then when the occluder was not present. Also we did not find any evidence to support our third hypothesis, which stated that the

participants would perceive the depth of the virtual objects to be biased towards their dark vergence distance.

In the future, we are planning to study additional issues in near-field depth perception in augmented reality, using the apparatus developed for this thesis. As suggested by this study, there was a significant amount of learning involved while performing the judgment task. A logical extension of current work would be studying the effects of practice on depth judgment accuracy. In most near-field AR applications, the users would interact with co-registered virtual and real objects for long periods of time, and through this interaction they may improve their initially-incorrect perception of the distance to virtual objects. Therefore, we imagine an experiment that uses a *pretest, intervention, posttest* design, where the pretest and posttest measure depth judgments using our apparatus, and the intervention is a period of time performing a task that involves manipulating virtual objects in the context of real objects, such as, for example, placing virtual blocks at specific locations on a real-world pattern. We hypothesize that the intervention would improve the accuracy of depth judgments.

Another experiment we can imagine involves adding other depth cues to the scene. An example would be studying the effect of motion parallax and relative size on depth judgments. In the current study, participants were required to keep their heads steady; however, motion parallax is considered a very salient depth cue. A future study could allow participants to move their heads and see the stimulus from different perspectives and with respect to other real or virtual objects. We hypothesize that adding motion parallax as a depth cue would result in more accurate depth judgments. Also in the current study,

the size of the virtual pyramid was randomly scaled; however, in real AR applications it is likely that most of the virtual objects would be of constant size. By keeping the stimulus size constant, another depth cue, relative size, would be available and could result in increased depth judgment accuracy.

In this experiment, we used an occluder that is as salient as possible. However, in real AR applications, it is likely that occluding surfaces would normally be, or could be designed to be, much less salient. For example, in a medical AR application supporting endoscopic surgery, where the doctor uses AR to see through the patient's skin, we could imagine the skin being covered by a plain white sheet (or black, if that color was found to be more helpful). We intend to perform future studies that vary occluder salience, where the occluder is monochromatic and does not rotate. Also, we found in the current study that there is a distance, possibly fixed for a specific user, where the occluder stops being salient and becomes transparent. In future studies we could examine different occluder distances to find where the occluder becomes transparent and possibly does not affect depth judgments.

In addition, in this experiment, we could not test the phenomenon of proximal vergence, which causes the "pushing" of a virtual object forward for a surface seen in front of a virtual object and backwards for a surface seen behind a virtual object. In future experiment, we plan to study this phenomenon by using an eye-tracker attached to the HMD to study how eye convergence affects depth judgments in the presence and absence of an occluder.

REFERENCES

[1] R. T. Azuma, "A Survey of Augmented Reality," *Teleoperators and Virtual Environment*, vol. 6, no. 4, Aug. 1997, pp. 355–385.

[2] R. T. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. Macintyre, "Recent Advances in Augmented Reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 7, 2001, pp. 34–47.

[3] V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3 edition, John Wiley & Sons., 1994.

[4] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual Anatomic Mimesis Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality," *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007, pp. 1–10.

[5] J. E. Cutting, "How the Eye Measures Reality and Virtual Reality," *Behavior Research Methods, Instruments, and Computers*, vol. 29, no. 1, 1997, pp. 29–36.

[6] J. E. Cutting and P. M. Vishton, "Perceiving Layout and Knowing Distance: The Integration, Relative Potency, and Contexual use of Different Information about Depth," *Handbook of Perception and Cognition*, vol. 5, 1995, pp. 69–117.

[7] S. R. Ellis and B. M. Menges, "Localization of Virtual Objects in the Near Visual Field," *Human Factors*, vol. 40, no. 3, Sept. 1998, pp. 415–431.

[8] E. F. Fincham and J. Walton, "The Reciprocal Actions of Accommodation and Convergence," *Journal of Physiology (London)*, vol. 137, 1957, pp. 488–508.

[9] W. C. Gogel and J. D. Tietz, "Absolute Motion Parallax and the Specific Distance Tendency," *Perception & Psychophysics*, vol. 13, no. 2, 1973, pp. 284–292.

[10] J. A. Jones, J. E. Swan II, G. Singh, E. Kolstad, and S. R. Ellis, "The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception," *APGV 2008*, Aug. 2008, pp. 9–14.

[11] J. A. Jones, J. E. Swan II, G. Singh, L. Lin, and S. R. Ellis, "The Effects of Augmented Reality, Virtual Reality, and Adaptation on Medium-Field Egocentric Depth Judgments," *manuscript in preparation*, 2010.

[12] W. S. Levine, *The Control Handbook*, 1 edition, CRC-Press, New York, 1996.

[13] M. A. Livingston, J. E. Swan II, J. L. Gabbard, T. H. Hollerer, D. Hix, S. Julier, Y. Baillot, and D. Brown, "Resolving Multiple Occluded Layers in Augmented Reality," *ISMAR'03*, Oct. 2003, pp. 56–65.

[14] J. M. Loomis and J. M. Knapp, "Visual Perception of Egocentric Distance in Real and Virtual Environments," *Virtual and Adaptive Environments: Applications, Implications and Human Performance*, Lawrence Erlbaum Associates, Mahwah, New Jersey, 2003, chapter 2, pp. 21–46.

[15] R. Messing and F. H. Durgin, "Distance Perception and the Visual Horizon in Head-Mounted Displays," *ACM Transactions on Applied Perception*, vol. 2, no. 9, July 2005, pp. 234–250.

[16] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Display," *IEICE Transactions on Information Systems*, vol. E77-D, no. 12, Dec. 1994, pp. 1321–1329.

[17] R. J. Miller, "Nonius Alignment Apparatus for Measuring Vergence," *American Journal of Optometry and Physiological Optics*, vol. 64, no. 6, 1987, pp. 458–466.

[18] M. Mon-Williams and J. R. Tresilian, "Ordinal Depth Information From Accommodation," *Ergonomics*, vol. 43, no. 3, 2000, pp. 391–404.

[19] D. A. Owens and H. W. Leibowitz, "Oculomotor adjustments in darkness and the specific distance tendency," *Perception & Psychophysics*, vol. 20, no. 1, 1976, pp. 2–9.

[20] J. P. Rolland, W. Gibson, and D. Ariely, "Towards Quantifying Depth and Size Perception in Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 4, no. 1, 1995, pp. 24–49.

[21] D. Sims, "New Realities in Aircraft Design and Manufacture," *IEEE Computer Graphics and Applications*, Mar. 1994, p. 91.

[22] L. Soler, S. Nicolau, J. Schmid, C. Koehl, J. Marescaux, X. Pennec, and N. Ayache, "Virtual Reality and Augmented Reality in Digestive Surgery," *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality(ISMAR)*, 2004, pp. 278–279.

[23] A. State, M. A. Livingston, W. F. Garrett, G. Hirota, M. C. Whitton, E. D. Pisano, and H. Fuchs, "Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies," *SIGGRAPH 1996*, Aug. 1996, pp. 439–446.

[24] J. E. Swan II, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, "Egocentric Depth Judgements in Optical, See-Through Augmented Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 3, May 2007, pp. 429–442.