EGOCENTRIC DEPTH PERCEPTION IN OPTICAL SEE-THROUGH AUGMENTED REALITY

By

James Adam Jones

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 & Engineering in the Department of Computer Science & Engineering

Mississippi State, Mississippi
August 2007
EGOCENTRIC DEPTH PERCEPTION IN OPTICAL
SEE-THROUGH AUGMENTED REALITY

By

James Adam Jones

Approved:

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

T.J. Jankun-Kelly
Assistant Professor of Computer Science & Engineering
(Committee Member)

Carrick C. Williams
Assistant Professor of Psychology
(Committee Member)

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

Roger King
Associate Dean for Research and Graduate Studies of the Bagley College of Engineering
Augmented Reality (AR) is a method of mixing computer-generated graphics with real-world environments. In AR, observers retain the ability to see their physical surroundings while additional (augmented) information is depicted as simulated graphical objects matched to the real-world view. In the following experiments, optical see-through head-mounted displays (HMDs) were used to present observers with both Augmented and Virtual Reality environments.

Observers were presented with varied real, virtual, and combined stimuli with and without the addition of motion parallax. The apparent locations of the stimuli were then measured using quantitative methods of egocentric depth judgment. The data collected from these experiments were then used to determine how observers perceived egocentric depth with respect to both real-world and virtual objects.
DEDICATION

To my parents, James & Jane Jones, as well as all those without whom this would not have been possible.
ACKNOWLEDGMENTS

I would like to express sincere thanks to the following individuals and organizations for their contributions to the continuing of perceptual research in augmented reality through guidance, funding, equipment, and facilities:

J. Edward Swan, II
Carrick C. Williams
T.J. Jankun-Kelly
Stephanie Doane
Eric W. Kolstad
Gurjot Singh
InterSense, Inc.
Department of Computer Science & Engineering
Institute for Neurocognitive Science & Technology
Office of Naval Research
National Aeronautics and Space Administration

I would also like to express my gratitude to those who aided me in my endeavors:

Robert O. Bankston, Cybergate, Inc.
Tommy, Chris, & Marlo Dorsey, Cybergate, Inc.
Christopher B. Shedd, University of Mississippi
Kerri L. Wardlaw, John C. Longest Student Health Center
Patricia Mapp, Jones County Junior College
Tim Powell, Southeast Data Solutions, LLC
Ty Parker, TechEdge Consulting, LLC
Theodore Watson, creator of the videoInput library
# TABLE OF CONTENTS

DEDICATION ............................................................................................................. ii  

ACKNOWLEDGMENTS ........................................................................................... iii  

LIST OF TABLES ....................................................................................................... vi  

LIST OF FIGURES ..................................................................................................... vii  

ABBREVIATIONS ..................................................................................................... x  

CHAPTER  

I.  INTRODUCTION ........................................................................................................ 1  
   1.1 Augmented Reality .......................................................................................... 2  
   1.2 General Depth Perception .......................................................................... 5  
   1.3 Depth Perception in Augmented Reality ................................................... 15  
   1.4 Methods of Depth Perception Measurement ...................................... 17  
       1.4.1 Verbal Report ..................................................................................... 17  
       1.4.2 Visually Directed Walking ................................................................. 19  
       1.4.3 Visually Directed Imagined Walking ............................................... 21  
       1.4.4 Visually Directed Triangulated Walking ........................................... 22  

II. RELATED WORK .................................................................................................. 24  

III. USER STUDIES .................................................................................................... 29  
    3.1 Motivation ...................................................................................................... 29  
    3.2 Hypothesis ..................................................................................................... 30  
    3.3 Compliance and Screening ......................................................................... 31  
    3.4 Resources ...................................................................................................... 32  
       3.4.1 Display Devices .................................................................................. 32  
       3.4.2 Tracking System ................................................................................ 33  
       3.4.3 Third Party Applications .................................................................. 34  
    3.5 Experiment I. ................................................................................................. 36  
       3.5.1 Experimental Setup ........................................................................... 38  
       3.5.2 Experimental Task ............................................................................. 44  
       3.5.2.1 Visually Directed Walking ............................................................. 44  

iv
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Depth Cue Characteristics [3]</td>
<td>7</td>
</tr>
<tr>
<td>4.1</td>
<td>Normalized Errors from various Augmented and Virtual Reality studies expanded from Thompson et al. [17]</td>
<td>94</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.1</td>
<td>Video See-Through (a) and Optical See-Through (b) Augmented Reality...</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Milgram et al.’s Reality-Virtuality Continuum [9]</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Top-down View of a Stereoscopic Grid with One Voxel Colored</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Chauvet-Pont-d'Arc Cave Painting</td>
<td>9</td>
</tr>
<tr>
<td>1.5</td>
<td>15th Century Japanese Artwork</td>
<td>10</td>
</tr>
<tr>
<td>1.6</td>
<td>Gentile Da Fabriano’s Adoration of the Magi</td>
<td>11</td>
</tr>
<tr>
<td>1.7</td>
<td>Leonardo Da Vinci’s The Virgin and Child with Saint Anne</td>
<td>12</td>
</tr>
<tr>
<td>1.8</td>
<td>Stereogram</td>
<td>13</td>
</tr>
<tr>
<td>1.9</td>
<td>Photograph Exhibiting Focal Accommodation</td>
<td>14</td>
</tr>
<tr>
<td>1.10</td>
<td>Verbal Report</td>
<td>18</td>
</tr>
<tr>
<td>1.11</td>
<td>Visually Directed Walking</td>
<td>19</td>
</tr>
<tr>
<td>1.12</td>
<td>Comparison of Directed Walking Studies from Loomis and Knapp [8]</td>
<td>20</td>
</tr>
<tr>
<td>1.13</td>
<td>Visually Directed Imagined Walking</td>
<td>22</td>
</tr>
<tr>
<td>2.1</td>
<td>Depth Cue Saliency Graph from Cutting [3]</td>
<td>25</td>
</tr>
<tr>
<td>3.1</td>
<td>InterSense Constellation Tool</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>InterSense Fiducial Tool</td>
<td>36</td>
</tr>
<tr>
<td>3.3</td>
<td>Anaglyph Scene Generated for Experiment I</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>The Mounting Apparatus Constructed for Experiment I</td>
<td>39</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>3.5 Barrel Distortion and Correction</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>3.6 An Experimenter Acquiring Measurements</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>3.7 InterSense IS-1200 Fiducial</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>3.8 Three Panel Constellation</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>3.9 Camera/Tracker Mounting Apparatus</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>3.10 Measuring the Camera’s Vertical Field of View</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>3.11 Visually Directed Walking Task Choreography</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>3.12 Calibration Procedures</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>3.13 An Observer Performing the Calibration</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>3.14 The 3D Compass</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>3.15 Experimental Location (left) and Virtual Model (right)</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>3.16 Hall with Intersecting Corridor Occluded</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>4.1 Error per Observer in Experiment I – Directed Walking (a) and Verbal Report (b)</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>4.2 Experiment I Results</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>4.3 Experiment I Error</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>4.4 Error per Observer in Experiments I &amp; II</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>4.5 Experiment II Results</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>4.6 Normalized Error by Observer in Experiment II</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>4.7 Experiment II Results Minus Observers s01, s06, and s13</td>
<td>92</td>
<td></td>
</tr>
</tbody>
</table>
4.8 Experiment II Error 

ix
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>AR</td>
<td>Augmented Reality</td>
</tr>
<tr>
<td>AV</td>
<td>Augmented Virtuality</td>
</tr>
<tr>
<td>$\Delta d$</td>
<td>Change in Distance or Position</td>
</tr>
<tr>
<td>$d$</td>
<td>Original Distance or Position</td>
</tr>
<tr>
<td>DOF</td>
<td>Degrees of Freedom</td>
</tr>
<tr>
<td>GLUT</td>
<td>OpenGL Utility Toolkit</td>
</tr>
<tr>
<td>HMD</td>
<td>Head-Mounted Display</td>
</tr>
<tr>
<td>MR</td>
<td>Mixed Reality</td>
</tr>
<tr>
<td>OpenGL</td>
<td>Open Graphics Library</td>
</tr>
<tr>
<td>RV</td>
<td>Reality-Virtuality</td>
</tr>
<tr>
<td>$s$</td>
<td>Cue Saliency</td>
</tr>
<tr>
<td>VR</td>
<td>Virtual Reality</td>
</tr>
<tr>
<td>$\Delta z$</td>
<td>Change in position along the Z axis</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>Field of view in degrees</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

When viewing objects in Augmented Reality (AR), observers are presented with conflicting depth cues. Many of these inconsistencies are due to the engineering limitations present in the design of head-mounted displays. Before practical AR systems can be implemented, we must understand what limitations are present and if there are ways we can compensate for these shortcomings. A key factor in this process is determining how observers perceive egocentric depth relations in augmented environments. This introduction discusses these factors as well as the techniques and technologies that are applicable to this research.

![Figure 1.1](image)

**Figure 1.1**
Video See-Through (a) and Optical See-Through (b) Augmented Reality

1
1.1 Augmented Reality

There are two common types of augmented reality: video see-through and optical see-through. Video see-through AR provides observers with a completely digital view of the real-world by displaying streaming video of the surrounding environment. This takes a fundamentally different approach than optical see-through augmented reality. With optical see-through AR, the observer maintains a purely optical view of the real-world. This is done by overlaying a semitransparent computer-generated scene on top of the real-world view. This allows graphics to be painted on top of the observer's view of the surrounding environment, see Figure 112. By referring to Milgram et.al.'s reality-virtuality continuum, one can see that video see-through augmented reality falls more closely toward virtual reality than optical see-through augmented reality, See Figure 1.2 [1,9].

![Reality-Virtuality Continuum](image)

Figure 1.2

Milgram et al.'s Reality-Virtuality Continuum [9]
The first augmented reality display was developed in 1968 by Ivan Sutherland and was referred to as the "Sword of Damocles" due to its ceiling-suspended design [13]. Augmented reality displays have progressed greatly since then, from the widely popular Sony Glasstron of the late 90s to today's high-end displays, such as the NVIS nVisor. Though augmented reality head-mounted displays (HMDs) have improved since the 1960s, there are still many engineering limitations that prevent computer-generated AR scenes from being able to provide the same depth cues and sense of realism as encountered when viewing the real-world. These limitations include environment modeling, tracking, and visual inaccuracies.

Detecting the environment in which an augmented reality application is being used can be terribly difficult. This usually involves complex 3D modeling of the environment where the application will be used. Additionally, tracking the position and orientation of the observer is an interesting problem. This typically requires the AR application to be used in an environment prepared for tracking, thereby restricting where the observer can navigate [1]. Though such issues are substantial obstacles, greater still are the visual problems that exist with displays. One of these problems is that of focal accommodation. Due to the design of the optics in most HMDs, all graphics are displayed at a fixed focal depth. This arrangement typically requires that the display plane be at a fixed distance, and therefore fixed focal depth. This prevents virtual objects that are collocated with real world objects from exerting the same accommodative
demand, unless both objects appear at the exact focal depth of the HMD. Another issue
that arises from the design of HMDs is the fact that all the graphics are displayed in a
discrete manner that creates a certain amount of positional ambiguity. Each combined
pixel, when converged to by an observer, represents not a single point in space, as in
reality, but instead represents a discrete volume. This is caused by the pixilated nature of
the display itself. When looking at a standard computer monitor, the X,Y axes’
resolution can be no smaller than the dimension of a single pixel. Similarly, HMD based
applications, the pixels displayed in each eye, combine to form voxels that represent a
particular volume within the field of view, see Figure 1.3. These are only some of the
limitations that affect depth perception in augmented reality. It is the goal of this
research to develop a more thorough understanding of how such limitations affect the
perceived depth of both real and virtual objects in augmented reality [1,14, 15].
1.2 General Depth Perception

Without understanding how people perceive their surroundings, it would be impossible to have much of the art and technology that exists today. As research expands into the worlds of virtual and augmented reality the need to understand perception is even greater. Specifically, development of truly immersive environments requires a more thorough understanding of how depth perception operates when visual information is presented with emerging technologies, such as augmented reality. This section takes a broad look at the evolution of depth cues in man-made representations of our
surroundings. Integrating these psychological and technological aspects of this research can provide much needed insight into the creation of realistic immersive computer-generated environments [1,3,14].

Before the perceptual problems presented by creating realistic augmented and virtual environments can be understood, one must understand the phenomenon of human visual perception. The human visual system is a fascinating and complex collection of cleverly arranged cognitive and perceptual mechanisms, all of which are designed to provide keen awareness of our surroundings. If realistic virtual and mixed reality environments are to be created, it is necessary for researchers to be mindful of the eccentricities of the visual system. Specifically, such research must address the issues of depth perception [1,3,4,7,14,15].

Depth perception is an interesting aspect of human vision that allows us to understand the relative locations of objects within visual space. Depth perception is based on combinations of depth cues. Various researchers have proposed that there are anywhere from seven to fifteen cues that heavily influence how depth is perceived. Depth cues can be divided into several groups, including monocular, binocular, ordinal, and quantitative, see Table 1.1. Monocular cues are those that only require only one eye to provide depth information, as opposed to binocular cues that require both eyes. However, these cues are not always equal in their salience. Saliency will either remain constant or decrease with increasing distance. Certain cue combinations provide us with
information about object placement within a scene at varying depths. Cues can be arranged into one of three categories: ordinal, quantitative, or metrical. Ordinal cues only provide order-based information about the relative location of objects in a scene. Quantitative cues provide relative distance information about the position of objects with respect to each other. Metrical cues are a subset of quantitative cue. These cues can provide an observer with an exact depth of an object in space. These cues, in concert, provide observers with a highly detailed mental geometry of their surroundings [4,14,21].

To understand depth perception, one must understand the cues on which it relies. Cutting takes an interesting path when studying the development of our understanding of depth. He uses the evolution of artistic expression to map the awareness of higher order depth cues. The theory behind this approach is that the least complex depth cues are the most obvious and should therefore emerge earlier in the history of art [3].

Table 1.1

Depth Cue Characteristics [3]

<table>
<thead>
<tr>
<th>Depth Cue</th>
<th>Ocularity</th>
<th>Depth Information</th>
<th>Salience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accommodation</td>
<td>Monocular</td>
<td>Ordinal</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Aerial Perspective</td>
<td>Monocular</td>
<td>Quantitative</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Binocular Disparity</td>
<td>Binocular</td>
<td>Quantitative</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Convergence</td>
<td>Binocular</td>
<td>Metrical</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Height in the Visual Field</td>
<td>Monocular</td>
<td>Quantitative</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Motion Perspective</td>
<td>Monocular</td>
<td>Quantitative</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Occlusion</td>
<td>Monocular</td>
<td>Ordinal</td>
<td>Constant</td>
</tr>
<tr>
<td>Relative Density</td>
<td>Monocular</td>
<td>Quantitative</td>
<td>Constant</td>
</tr>
<tr>
<td>Relative Size</td>
<td>Monocular</td>
<td>Quantitative</td>
<td>Constant</td>
</tr>
</tbody>
</table>
The first known artistic expressions are those left by primitive humans in the caves of Europe. These cave paintings depict both people and animals in multitudes, see Figure 1.4. This is where we encounter the depth cue of occlusion. Occlusion means that an object obstructs the view of other objects along the same visual path that are further in depth from the viewer. In many cave paintings we can observe depictions of herds of animals with the foremost animals occluding the more distant animals. Occlusion is a monocular cue that provides us with ordinal information about object arrangement. This means that a quantitative measure of distance can not be determined by using this cue alone. Occlusion only provides the viewer with an ordinal sense of which object falls in front of the other. Though occlusion cannot provide us with the means to gauge distance, it does have an advantage over most other depth cues. Occlusion is one of the few cues that does not decrease in its salience as the distance from the observer to the object increases. This means that roughly the same amount of information can be received from occlusion at 5 meters as could at 500 meters. Relative size and relative density are the only two other cues that share the characteristic of constant saliency [3].
The next cue that we will discuss is known as height in the visual field, which first appears in 15th Century Japanese artwork, see Figure 1.5. This cue demonstrates the fact that objects that are below the horizon appear to rise in the visual field as they move further from the viewer. Conversely, objects above the horizon appear to descend in the visual field as their distance from the observer increases. Height in the visual field has no apparent informational value until objects are at least 5 meters from the viewer. The saliency of this cue decreases as distance increases until approximately 1,000 meters, at which point no additional depth information can be derived from this observation. This cue is also a monocular cue, but it is the first quantitative cue that we see emerge in art. This means that the depth of an object can be estimated by observing its position relative to the horizon [3,21].
In later Greek, Pre-Renaissance European, and Asian artwork we encounter relative size, see Figure 1.6. Relative size is the measure of the visual angle an object occupies in the field of view. This measure can then be compared to other objects of similar size and texture. Based on the ratio between multiple objects within a scene, the viewer can determine the relative distance between them. Like occlusion, relative size is a monocular cue with constant salience, but unlike occlusion it provides quantitative depth information. If the sizes of the objects are known, the viewer can estimate the objects’ distances. Even if the object sizes are not known, the viewer can still estimate proportional distances [3,21].
Around the 15th Century, European artists began utilizing the concept of relative density as a cue for depth, see Figure 1.6. Relative density provides depth information based on the number of similarly sized objects that occupy a visual angle. The further from the viewer a group of objects is placed, the more objects can be displayed within a given visual angle. This monocular cue provides an estimate of relative depth based on the number of objects within an angle of the field of view. This is not a strong cue, but it does provide a basis by which the viewer can compare groups of objects with equal spacing. Like occlusion and relative size, this cue has a constant salience across all depths [3,21].

Figure 1.6

Gentile Da Fabriano’s *Adoration of the Magi*
Aerial perspective is a cue that is an effect of the atmospheric conditions on vision. Due to the moisture, dust, and pollutants in the air, the light reflecting off of distant objects becomes more diffuse and appears hazy. This cue was not used in a systematic manner until the works of Leonardo Da Vinci in the late 15th and early 16th Centuries, see Figure 1.7. Though this cue provides quantitative information about distance, it is only effective at distances greater than 100 meters [3,21].

Figure 1.7

Leonardo Da Vinci’s *The Virgin and Child with Saint Anne*
The first binocular cue to be integrated into art was binocular disparity. Though artists were aware of binocular disparity as far back as Da Vinci, no one understood how it could be applied to creating more realistic images. It was applied in the 19th Century when Wheatstone created the first stereograms, see Figure 1.8. Wheatstone found that by providing a pair of images whose viewpoints were adjusted for each eye's position, a very compelling sense of depth was achieved. This cue exploits the fact that we process scenes binocularly by judging the relative changes of the images projected onto the retinas of both eyes [3,21].

![Figure 1.8](image_url)

**Stereogram**

It was not until the widespread use of photography that people became aware of accommodation as a depth cue. Accommodation is the focusing of objects by the adjustment of the lenses. In the eyes, this is achieved by expanding and contracting the
crystalline lenses, but in photography this is done by adjusting the glass lens of the camera. This cue is apparent when taking a photograph where there are objects in the scene at distances outside the focal length. Objects that are at the approximate focal length will appear more clearly with sharp, defined edges while objects at other depths will appear blurred and unfocused, see Figure 1.9 [3,21].

![Figure 1.9](image.png)

**Figure 1.9**

*Photograph Exhibiting Focal Accommodation*

Another depth cue that is closely connected to accommodation is convergence. Much like binocular disparity, convergence is a binocular cue that relies on the fact that our eyes have two differing views of the world. At close range the difference in our eye positions requires us to rotate our eyes independently of each other to fixate on objects at
a given depth. Typically this is done in exact correspondence with accommodation. Convergence is both a quantitative and metrical depth cue. This means that convergence can be directly measured in order to determine the exact distance at which an object is from the observer. Typically both convergence and accommodation are only effective depth cues until about 3 meters from the observer [3,21].

The final cue that will be discussed in this section is motion perspective, also known as motion parallax. Awareness of this cue came about with the advent of motion pictures. Motion parallax only becomes apparent when the observer is moving. While observing objects at different depths, closer objects in the foreground will appear to the moving more quickly than objects in the background. As objects approach the horizon, they will appear to move along with the observer. It is important to note that motion parallax can be induced by both motion of the observer relative to the scene or motion within the scene relative to the observer. With motion parallax the perceived movement within a scene can act as a strong indicator of relative depth. This can provide a monocular means of judging relative distances when in motion [3,21].

1.3 Depth Perception in Augmented Reality

One might ask how the evolution of depth cues in art applies to augmented reality. These cues provide observers with a strong sensation of depth when viewing the real world, and over the course of history artists have integrated them into their creations in order to provide more realistic representations of their surroundings. Such
representations are not strictly for aesthetic purposes, but can also be used as informational tools. The goal of computer visualization, in general, is to provide insight by conveying information through graphical representation. This is especially true in the case of computer-generated immersive environments. By optimizing the realism of such environments, the potential for correct interpretation and the capacity for insight is greatly increased.

In purely virtual environments, the computer-generated world must be represented in such a manner as to appear sufficiently realistic. This requires consistent and systematic placement and interaction of environmental elements with the observer. With regard to the evolution of artistic representations, augmented reality can be considered the next step in this series. Virtual and augmented environments can combine the aforementioned cues into interactive computer-generated scenes. However, such levels of interaction have never before been available. Now we must understand how observers will perceive the environment when combining these cues in an interactive virtual world [5,8].

In augmented reality such perceptual issues are of paramount importance. When dealing with a purely virtual environment, the observer has only the virtual world with which to compare relative realism. Additionally, if perceptual incongruities are present in a purely virtual environment, these incongruities are presented consistently throughout the entire virtual scene. However, in augmented reality, real-world scenes and virtual
scenes are intermingled. This produces unique and sometimes conflicting perceptual situations where inconsistencies between then real and virtual worlds exaggerate perceptual misinterpretations. One of the most important factors is understanding how observers perceive co-located virtual and real objects. In order to produce realistic and effective augmented reality applications, a thorough understanding must be developed of how depth is perceived within these mixed environments [1,3,14,15].

1.4 Methods of Depth Perception Measurement

How can you measure something that exists solely in the mind of an observer? This is an important question when studying how people perceive their surroundings. A researcher can not simply peer into the mind of observers and measure their sensation of depth. Such measurements must be done indirectly using measurable feedback from the observer. Such feedback can provide insight into the internalized spatial map developed by observers when viewing a scene. There are several different methods of obtaining a quantitative measurement of an observer’s depth perception.

1.4.1 Verbal Report

One of the most obvious methods is to simply ask the observer how far away an object appears, see Figure 1.10. This technique is known as verbal report. Even though this technique is very straight forward, it has a drawback. Verbal report can be a reliable and easy to implement protocol but, in some cases it has been found to be a noisy method
for determining observers' perceived depth judgments [6,7]. The main reason is that people typically have a very poor understanding of the units they use to report the distance to an object. This also makes it exceptionally difficult to compare judgments between observers since no normalized unit can be achieved without providing an external reference or getting an a priori measurement of the user's perception of the reported unit.

Figure 1.10

Verbal Report
1.4.2 Visually Directed Walking

Another protocol that is commonly used is known as visually directed walking, or blind walking. In visually directed walking, observers view an object and then are asked to close their eyes and walk toward it, stopping when they feel as though they are standing at the observed position of the object, see Figure 1.11.

Loomis and Knapp compiled the results of eight different studies comparing the accuracy of blind walking and found that it is an exceptionally stable and accurate way to measure an observer's perceived distance to a target object, see Figure 1.12. Since this method forces observers to rely solely on their internalized spatial map, it provides
experiments with a means of attaining a relatively pure measure of perceived depth. Though this method is exceptionally accurate, it does have one drawback. Observers must be able to walk the full distance to the target object. In some circumstances this is simply not possible, such as in restricted immersive environments, like CAVEs or display walls [6,8,10,16,17].

Figure 1.12

Comparison of Directed Walking Studies from Loomis and Knapp [8]
1.4.3 Visually Directed Imagined Walking

A further method that can be used is imagined walking, see Figure 1.13. Imagined walking is similar to blind walking, but the observer never actually moves. With this protocol, observers are holding a stop watch and view a target object. Once the observers are ready, they close their eyes, start the stopwatch, and imagine that they are walking to the target object. When they feel as though they have reached the imagined position of the object, they stop the stopwatch. This time is then recorded. After the experiment is over, the observers are asked to walk a certain distance while being timed. This time is used as a baseline to which the recorded imagined walking times are compared. Using this method the observers are never required to move and can therefore conduct the experiment in environments where movement is highly restricted [10]. Unfortunately, this method is not as well studied as visually directed walking and is not as widely accepted.
1.4.4 Visually Directed Triangulated Walking

Another method is triangulated visually directed blind walking, or triangulated walking. This method is very similar to blind walking but does not require that the observer be able to walk the entire distance to the target object. Instead, observers are asked to regard the referent object, then close their eyes, walk at an oblique angle to the object, then stop, point at the object, and drop a small marker from their pointing hand. An experimenter then marks the position between the observer's feet. Using the starting position, stopping position, and the pointing position, the distance from the observer's initial location to the observed location of the referent object can be determined by
trigonometric relationships. This method provides noisier results than blind walking, but allows observers to perform the task in much more spatially restricted environments [20].
CHAPTER II
RELATED WORK

Much work has been done in the field of virtual reality, especially concerning the area of depth perception. Cutting views virtual reality as an extension of other artistic expressions, such as paintings and photography [3]. By examining it in such a manner, he makes several key observations about the evolution of depth cues in art that are also relevant to virtual reality. Most notably, Cutting observes that as art evolved, so did the use and complexity of the depth cues. Keeping with this notion, the combination of cues involved in virtual reality should therefore be some of the most complex and compelling yet.

Cutting also discusses his classification of visual space [3,4]. Visual space is explained as being divided into three levels based on distance away from the observer: personal space, action space, and vista space. These can also be referred to as near, medium, and far fields respectively. Cutting defines personal space as the area within which an observer can directly manipulate an object, usually within arms’ reach or about 1.5 meters. Action space is then defined as the area around which an observer can accurately throw an object or hold a conversation, approximately 1.5 meters to 30 meters. Vista space is defined as all distance beyond 30 meters.
Over these distances different cues provide varying levels of saliency. The saliency of a cue is defined by the minimum change in position ($\Delta d$) a stimulus must undergo from its original position ($d$) that can be noticed by an observer. The saliency ($s$) of a cue can then be mathematically defined as $s=\Delta d/d$. This defines a just-noticeable distance threshold. Cutting proposed a theoretical model of depth cue saliency that provides a rough description the informational value of depth cues as distance increases, see Figure 2.1. Most depth cues have varying saliencies over a certain range of distances, but three cues are reported to have a constant saliency over all ranges of distances: occlusion, relative size, and relative density. These cues will provide the same amount of information at any distance [3].

![Depth Cue Saliency Graph from Cutting](image)

Figure 2.1
Depth Cue Saliency Graph from Cutting [3]
Cutting notes that the characteristics of these cues should be exploited in order to make virtual environments more representative of reality. Many virtual reality applications focus their utility on a particular region of space. For instance, most medical applications deal primarily with personal space, while navigation applications typically deal with either action or vista space. The same is true of many augmented reality applications as well. Therefore, the locus of activity needs to be considered when designing and testing virtual and augmented applications [3].

Loomis and Knapp studied the differences between the judged distances of reference objects in both virtual reality and the real world [8]. Several different means of perceptual measurement were used, including verbal, motoric, and aperture based techniques. These studies revealed a distinct tendency to underestimate the distance of objects displayed in virtual reality. Loomis and Knapp note that these perceptual issues are the result of a misinterpretation of egocentric distance and possibly the scale of visual space as well. These misinterpretations directly affect the observer’s spatial behavior, leading to inaccurate interactions in visual space.

There have also been several studies that have examined the effect of restricting the depth cues provided to the observer [5,8]. These studies have demonstrated that there is a distinct change in spatial behavior when the number of cues available to the observer is restricted. This would indicate that observers rely heavily on the presence of visual cues, combining them to form a more concrete understanding of the surrounding
environment. In a study conducted by Hu et al., observers were presented with various combinations of visual cues, including shadows, interreflections, and stereo vision [5]. Hu et al. then examined the observers’ ability to accurately place virtual objects on a virtual surface. The observers’ accuracy increased with the addition of each depth cue. The best performance was in a virtual environment with stereo vision, shadows, and interreflections, while the worst performance was in a virtual environment that only presented a monoscopic view of the scene [5].

Swan et al. conducted perceptual matching studies in augmented reality [7,15]. These studies have presented some exceptionally interesting finds. Similar to virtual reality depth studies, a tendency to underestimate object distances was observed; however, this bias changed to overestimation around 23 meters.

Other work, investigating the possibility that the quality of the images presented in virtual reality affect depth judgments, was conducted by Thompson, et al [17]. This research compared observer performance in directed walking tasks in virtual reality as the quality of the projected scene changed. The scenes included a real-world scene, a high-resolution panoramic image of the real-world scene, and a computer-generated wireframe representation of the real-world scene. This study found that there was little effect of scene quality on the resulting depth judgment.

Another study has considered the effect of prior exposure to an experimental environment and dissociation of presence on the effect of directed walking tasks in
virtual reality. The concern was that observers may not feel a sense of presence when immersing themselves into a virtual environment that they know does not match the real-world. Interrante, et al. studied this possibility by providing observers with a baseline condition where they had no prior exposure to the real-world environment that the virtual scene was modeled after and a post-exposure condition where the observers had previously seen the location modeled in the virtual scene [6]. It was discovered that there was only a very slight advantage to having previously seen the real-world location prior to being exposed to the virtual scene.

Another study has indicated that depth perception can be greatly affected by the height from which a scene is viewed. Observers are tuned to making depth judgments from their particular eye heights. Thompson et al. found that by altering the perceived height from which a scene was viewed underestimation would increase [16]. These results imply that ensuring that scenes are rendered from the proper vantage point is very important for proper perception of depth.

These studies reveal some interesting trends. Specifically, the addition of cues in virtual environments allows observers to build a more stable spatial map of their surroundings. However, in both virtual and augmented environments, some cues are almost always in conflict. The exact effect of these conflicts is still largely unexplored, but it has been widely seen that observers tend to underestimate egocentric distances to target objects placed along the ground plane.
CHAPTER III
USER STUDIES

For this research, two experiments were conducted. The first of which was intended to determine if the underestimation effects seen in virtual reality transfer to augmented reality. This study was also used to ascertain if depth estimation measurement techniques commonly used in virtual reality were also practical for use in augmented reality. The second study was intended to determine if the addition of motion parallax enabled observers to more accurately judge the location of objects in augmented reality. Both studies presented observers with real and virtual stimuli.

3.1 Motivation

It has often been found that observers underestimate the egocentric distance to targets displayed in a virtual environment. The work in Experiment I was geared towards finding if a similar phenomenon occurs when viewing targets in an augmented environment. Since virtual and augmented reality share many common traits, there seemed to be a high likelihood of finding underestimation effects in AR as well. If such effects did indeed exist, this research aimed to detect them by utilizing methods of deriving a quantitative measurement of egocentric depth judgments. Additionally, real world targets seen through the HMD, combined real and virtual targets, and purely real
targets were also tested during this experiment. These additional stimuli were tested in order to determine if they would also suffer from similar effects. If so, the underestimation found in VR may not be the result of viewing purely virtual environments but instead possibly from a combination of factors.

3.2 Hypothesis

The hypothesis of this research is that better, more ubiquitous augmented environments can be engineered by understanding which depth cues are present in augmented reality and how they affect an observer’s perception of visual space. Specifically, this research studies the idea that the addition of motion parallax may help resolve ambiguities introduced by conflicting depth cues in augmented reality. Many studies have revealed perceptual differences between virtual, augmented, and real-world scenes [5,7,8,14,15]. Once the sources of these inconsistencies are understood, a perceptual model of augmented reality can be developed to help facilitate future research.

It is the goal of this research to perform user studies to gather data on the effect of various cue combinations and virtual/real stimulus types on observer perception of visual space as well as their spatial behavior.

The hypothesis was tested by conducting two separate user studies, Experiments I and II. In these studies, observers were presented with varied real-world, real-world viewed through a HMD, virtual, and combined real-virtual stimuli. The apparent
locations of the stimuli were then measured using quantitative methods of egocentric depth judgment.

The results of these experiments revealed that observers tended to underestimate the distance to some stimuli under certain conditions. Some of these results imply that the HMD itself may be interfering with the observer’s perception of depth. One possible cause of this effect is the HMD’s limited field of view. This theory would be consistent with the findings of Wu et al., who found that when restricting the vertical field of view, observers began to consistently underestimate the depth of stimulus objects [22]. However, they also found that if the observers were allowed to move their heads, the underestimation tended to be resolved. The explanation of this effect is that the addition of cues related to motion enabled the observers to form a more accurate understanding of the object’s position in space by integrating patches of texture along the ground plane.

3.3 Compliance and Screening

For this research, it was necessary to directly study how observers perceive depth relations in augmented reality. Observers were recruited to participate in experiments to test the aforementioned hypothesis. However, before the experiments could take place, the procedures were reviewed and approved by the Institutional Review Board through the Office of Regulatory Compliance. This ensured that proper procedures were in place to protect the interests and privacy of the observers who participated in these studies.
All observers were screened for various medical conditions prior to participating in the studies. There are several major medical factors that must be considered when running studies of this nature. These include screening for a history of epilepsy, head or neck injury, and uncorrected vision. Exposure to any video device can potentially invoke an epileptic reaction that may result in a hazardous state. For this reason, volunteers with a past history of epilepsy were not permitted to participate in these studies. Also, volunteers who had head or neck injury were not allowed to participate, as the head-mounted display and subsequent head and neck movements may aggravate such conditions. Due to the nature of this study, normal or corrected-to-normal stereo vision was required. Normal vision is defined as 20/20 or better performance on a Snellen eye chart. Observers were asked to report their visual acuity prior to participating in the studies.

3.4 Resources

In order to carry out this research, certain resources were needed. These resources fall into three categories: display devices, tracking systems, and third party applications. Each of these is detailed as follows.

3.4.1 Display Devices

In order to present an observer with an augmented environment, there must be some form of display device on which to show the virtual components of the scene. For
the purposes of this research there were two display devices that are readily available. The first was a Sony Glasstron LDI-100b. This display is a binocular head-mounted system that has been equipped with aftermarket anaglyph filters to enable stereoscopic viewing. The Sony Glasstron is capable of displaying 3D anaglyph stereo scenes at a resolution of 800 × 600 pixels with a 33° diagonal field-of-view. The second display was an NVIS nVisor ST. The nVisor ST has a resolution of 1280 × 1024 pixels, 60° diagonal field-of-view, and 100% visible area and display area overlap. Another display option that will be available for future research is the display wall in the Empirical Visualization and Imaging Lab. The display wall is comprised of a large rear-projection screen illuminated by 15 high-resolution projectors. The total resolution of this display is 7000 × 3150 pixels.

3.4.2 Tracking System

One of the key requirements of an augmented reality system is knowing the observer’s position and gaze direction. In order to do this, the position and orientation of the observer’s head must be tracked in real-time. For this purpose, we used an InterSense IS-1200 position tracking system. The IS-1200 is a 6 degrees-of-freedom (6-DOF) tracker that uses a combination of optical and inertial tracking technology. The term “6 degrees-of-freedom” comes from the values that are reported by the tracker. These values include three rotational axes (roll, pitch, and yaw) and three positional axes (X, Y, and Z). These values are fundamental to creating an interactive virtual or augmented
environment, as it enables the view of the world to be adjusted to the observer’s current viewing position and direction.

3.4.3 Third Party Applications

The IS-1200 tracking system was shipped with several tools that were needed for configuring the system, as well as preparing the tracking environment. The constellation tool provides the user with a point and click interface for designing the constellation of fiducials. This tool also provides an interactive 3D view of the constellation for verifying and analyzing the fiducial arrangements. Once one has completed designing the constellation, the tool generates a tracking database and uploads it to the IS-1200 tracker, see Figure 3.1.
The IS-1200 fiducial tool was used to generate and print the fiducials needed to compose constellations, see Figure 3.2. This tool is necessary not only to print the fiducials, but to ensure that they are the optimal size. This tool calculates the optimal size based on the minimum and maximum tracking distances and the field of view of the lens on the optical sensor. This tool also has the ability to export images of the fiducials in various formats.
3.5 Experiment I

3.5.1 Experimental Setup

The head-mounted display used in Experiment I was a Sony Glasstron LDI-100b monoscopic, biocular optical see-through display. The oculi of the display were fitted with anaglyph filters in order to display fully stereoscopic scenes. The anaglyph filters were specially designed to filter the frequencies of red and blue typically produced by monitors. The HMD's resolution was 800 × 600 pixels. The optical see-through area of the display was larger than the actual display area. The field-of-view of the optical see-
through area was empirically measured to be $66^\circ \times 38^\circ$. The display window was approximately centered in this area with an empirically measured field-of-view of $27^\circ \times 20^\circ$.

A 3D anaglyph scene was presented through the HMD with the left eye view drawn in blue and the right eye view drawn in red, see Figure 3.3. The red and blue filters were attached to the right and left oculi of the HMD, respectively. The anaglyph filters, combined with the limited optical pass-through of the HMD, made the real-world scene very dark and difficult to see. To compensate for this, the experimental environment was lined with six 600-watt halogen lamps. The extra light allowed the real world scene to be visible through the filters. Additionally, the brightness of the virtual scene was adjusted so that the luminance of virtual objects approximated that of the real-world objects.
No head tracking system was available for this experiment, so the HMD was rigidly attached to a crossbar apparatus with an adjustable tripod attached to each end, see Figure 3.4. This allowed for the position of the HMD to be fixed in such as manner as to provide a particular point of view into both the real and virtual scenes. The height of the apparatus was adjusted on a per observer basis so that every observer could stand in a normal, comfortable position while viewing the experimental scenes. One end of the crossbar was hinged to one of the tripods. This allowed the experimenters to swing the crossbar out of the way during directed walking trials. The hinge of the apparatus was constructed out of a caster wheel barring assembly, which allowed the crossbar to have full 360° of rotation. The tripod at the opposite end of the apparatus had a wooden “L” shaped terminal on which the crossbar would rest when the observers were looking
through the HMD. The apparatus was engineered in such a way as to cause minimal
disruption of display alignment when swung away from the observer. For this reason
only small adjustments were needed during each block of trials.

Figure 3.4
The Mounting Apparatus Constructed
for Experiment I

The optics in the HMD caused a common form of visual anomaly known as barrel
distortion. This distortion causes pixels rendered toward the edge of the display area to
appear to move farther from the display’s center as they get closer to the middle of the
horizontal and vertical axes, see Figure 3.5. This distortion was corrected by using a 2D polygonal texture mapping technique [2]. This correction was calibrated by displaying a virtual grid of 16x12 cells of constant height at 1.2m, the focal depth of the Sony Glasstron LDI-100b, and placing a real grid of the same size at the same real-world depth. The real and virtual grids were viewed monocularly and a mouse was used to move the vertices of the virtual grid to match those of the real-world grid. This procedure was done separately for both the right and left eyes. Because the physical interpupillary distance of the HMD was fixed at 54mm, per observer adjustments were made in software to accommodate variations in interpupillary distance. The focal accommodation and ocular convergence were also fixed at 1.2m. Since these elements were constant, the barrel distortion correction procedure only needed to be performed once.
Prior to beginning the experiment, the experimenters measured the observers' interpupillary distance and eye height. An experimenter would first measure the observers’ eye height by using a landscape survey rod. Next, an experimenter would measure the observers' interpupillary distance to the closest millimeter using a small ruler, see Figure 3.6. Both the eye height and IPD were only measured once per observer and then entered into the software that generated the 3D scenes.
Observers took part in Experiment I in two groups. Each group consisted of exactly half of the total observers; 8 observers in each group. However, the experiment location for each of the groups was different. Location 1 was the third floor hallway of Etheridge Hall at Mississippi State University. Location 1 was 2.28m in width and 30.4m in length. Observers stood 8.83m from the north wall of the hallway facing south. Location 2 was a room 11.35m in length and 7.26m in width at the newly constructed Institute for Neurocognitive Science and Technology. At Location 2, observers stood 1.7m from the east wall facing west.

In order to prevent the observers from hearing the movement of the experimenters, the experimenters emptied their pockets of anything that could indicate their position, such as keys, loose change, and cellular phones. Also, to prevent the
observers from counting the steps of the experimenters while they placed and retrieved the target objects, the experimenters wore only socks on their feet. This was necessary since both locations had hard tiled floors on which shoe steps sounded loudly.

During this experiment observers were shown both virtual and real target objects. The virtual object was a wireframe pyramid that was 23.5 cm tall with a 23.5 cm square base. The real-world object was a small, wooden pyramid of the same dimensions as the virtual object. The real-world object was also painted white to give it an appearance that more closely resembled that of the virtual object. A fiberglass survey tape was attached to the floor in both locations. This allowed the experimenters to measure the distance walked by the observers when performing the directed walking task.

Objects were displayed at distances ranging from 3 to 7 meters. This gamut of distances was selected because it was the widest range of distances that could be seen through the fixed display by observers of a common span of heights, roughly 5 feet to 6.5 feet.

There were four environmental conditions: Real, Real+HMD, Virtual, and Real+Virtual. The Real condition was the real-world pyramid viewed without an HMD and with the crossbar apparatus moved to the side of the observer. The Real+HMD condition was the real-world pyramid viewed through the HMD mounted on the crossbar apparatus with no virtual augmentation. The Virtual condition was the virtual pyramid viewed in the real-world environment through the HMD mounted on the crossbar
The Real+Virtual condition was the virtual pyramid superimposed on top of the real-world pyramid viewed through the HMD mounted on the crossbar apparatus. The Real+Virtual condition required very precise alignment for the scene in order to appear correctly to the observer. For this reason, the HMD was realigned in between every trial within a block for this condition. For all other conditions, the HMD was realigned between each block of trials.

The data collection and scene generation for this experiment was handled by a Pentium M 1.80 Ghz laptop computer with an NVIDIA GeForce FX Go5200 graphics card. The software for this experiment was implemented in C++ using Cygwin, OpenGL, and Perl.

3.5.2 Experimental Task

Two depth judgment measurements were used in Experiment I, visually directed walking and verbal report. Observers were allowed to view stimuli for any length of time they required and were not timed. However, an average response time was observed to be approximately 2 seconds after stimulus onset.

3.5.2.1 Visually Direct Walking

In visually directed walking, also known as blind walking, observers view an object until they feel they have a good sense of where it is located in relation to their current position. When the observers feel confident in their judgment of the object's
position, they close their eyes and walk until they believe they are standing at the location of the object. The distance walked by the observers is then measured and recorded.

The protocol used for blind walking was consistent across all observers. Before beginning the experiment, an experimenter gave the observers the following explanation of visually directed walking:

Visually directed walking, or blind walking, is walking to an object with your eyes closed. We will show you an object and ask you too look at it until you feel like you have a very good sense of where it is located in space and feel as though you can walk to it with your eyes closed. For instance, when you wake up at night and need to get a drink of water, it is probably completely dark in your bedroom, but you can always walk to the light switch even though you can not see it. We want you to have the same sense of the target object's position. When you feel that you can walk to the object with your eyes closed, let me know you are ready, and I will ask you to walk forward. I want you to walk until you feel as though the tips of your toes are at the center of where the object was located. When you stop walking, you may open your eyes, but do not look down at your feet. The object will be removed from the scene, so there will be no chance of you stepping on or tripping over the object. We will also be walking near you during the experiment to help you in case you loose balance or become
disoriented. We will also stop you before you walk too close to the walls or any obstacle in the scene.

These instructions were orally recited to the observer in a conversational manner, so there were slight variations but all information listed above was conveyed. Observers seemed more comfortable and confident in the guidance of the experimenters when these instructions were given in a conversational manner. Instilling a sense of confidence in the observers was crucial to having them perform the directed walking task. Otherwise, observers seemed to be more nervous and distracted from the goal of the task when walking with their eyes closed.

Every trial, except for the real condition, began with the observer standing in an isolation area. In Location 1, this was a small room next to the display apparatus. In Location 2, this was a cubical-like area formed by a folding room divider. The observer returned to the isolation area after every block of trials, when the display apparatus needed realignment, or between every trial during the combined Real+Virtual condition. Once the display apparatus was ready, the observers were asked to come forward and stand with their feet centered on the metal clasp at the end of the survey tape and then close their eyes. At this point the experimenters would place the target object in the scene. The protocol for each directed walking trial went as follows:

**Experimenter 1:** Open your eyes and observe the object.
Look through the display. *(This instruction is omitted for the Real condition)*

Observe the object and tell me when you are ready.

**Observer:** *(The observer views the object and indicates readiness)*

**Experimenter 1:** Close your eyes...

*(Experimenter 1 swings the crossbar out of the way)*

...and walk forward.

**Experimenter 2:** *(Experimenter 2 removes the target object during the Real and Real+HMD conditions)*

**Observer:** *(The observer walks forward with eyes closed)*

**Experimenter 1:** *(Experimenter 1 walks behind the observer, watching for signs of disorientation)*

**Experimenter 2:** *(Experimenter 2 walks in front of the observer, watching for signs of disorientation)*

**Observer:** *(Stops when they believe they have reached the target distance)*

**Experimenter 1:** *(Experimenter 1 records the distance walked by the observer)*

Return to your starting position and close your eyes.

*(In the Real and Real+HMD conditions, Experimenter 1 indicates the position of the next target to Experimenter 2)*

**Experimenter 2:** *(In the Real and Real+HMD conditions, Experimenter 2 places...*
In the event that the experimenters felt that observers risked walking into an obstruction, the observers were asked to stop and open their eyes. The observers were then instructed to return to the starting point. For programatically displayed stimuli, such as the Virtual and Real+Virtual stimuli, the experimenter would move back two trials and start again. This was done in order to prevent observers from seeing the same stimulus position twice in a row. For the Real and Real+HMD stimuli, the experimenters would randomly select a position exclusive of the last trial distance and then resume the interrupted trial afterwards.

3.5.2.2 Verbal Report

In verbal report, observers were asked to view the object, attempt to estimate the distance from their current position to the target object, and verbally express the distance using any unit measure with which they were comfortable. The observers began each block of trials standing in the isolation area.

When the display apparatus was ready, the observers were asked to come forward and stand with their feet centered on the metal clasp at the end of the survey tape and then close their eyes. At this point the experimenters would place the target object in the scene. For every trial the observers received the following protocol:

**Experimenter 1:** Open your eyes, observe the object, and tell me how far away
you think it is.

**Observer:** *(The observer views the object then responds with the estimated distance)*

**Experimenter 1**: Close your eyes.

*(Experimenter 1 records the estimate. In the Real and Real+HMD conditions, Experimenter 1 indicates the position of the next target to Experimenter 2)*

**Experimenter 2**: *(In the Real and Real+HMD conditions, Experimenter 2 places the target object at the indicated position)*

### 3.5.3 Experimental Design

Experiment I utilized a within subjects design. The designed controlled for both stimulus type and estimation protocol. A 4x4 Latin Square for stimulus type and a 2x2 Latin Square for estimation protocol was used to counter-balance the exposure order of stimuli on a per observer basis. This ensured that for every group of 8 observers every presentation order of stimulus type and estimation protocol would be covered. This was done to minimize any asymmetric transfer effects. Observers were shown stimuli at distances of 3m, 5m, and 7m with four repetitions at each distance per block for each experimental trial. To help prevent observers from noticing the repeating distances, 25% of the stimuli were noise trials which were randomly selected in 0.25m increments from
the 3m to 7m range, exclusive of the 3m, 5m, and 7m distances. The noise trials were excluded from analysis.

A total of 16 observers participated in Experiment I. Observers 1 through 8 performed the tasks in Location 1, while observers 9 through 16 performed the tasks in Location 2. This allowed for proper counterbalancing with respect to location in addition to presentation order. A total of 1536 data points were collected for analysis: 16 observers, 4 stimulus types, 2 estimation protocols, 3 distances, and 4 repetitions.

3.6 Experiment II

3.6.1 Development

The setup for Experiment II was considerably more complicated. This was mostly due to the use of more advanced equipment, most notably the NVIS nVisor ST HMD and the InterSense IS-1200 tracking system. The nVisor required a more thorough calibration procedure than the Glasstron used in Experiment I. Also, since no tracking system was previously used, the software to calibrate and interface with the virtual scene had to be designed from the bottom-up. Developing the software and calibration procedures for Experiment II was a very complex and time consuming task and involved several months of design, testing, and revisions. The procedures used to prepare for this experiment are detailed in the following sections.
3.6.1.1 Tracking

One of the most important elements of creating a usable augmented reality system is ensuring that the position and viewing direction of an observer can be accurately determined. This is necessary to provide the observers with a correct view of virtual elements with respect to their corresponding locations in the real world. This is typically achieved by placing a head-tracking device on the HMD.

For the research discussed in Experiment II, a pre-production InterSense IS-1200 inertial/optical tracker was used. The IS-1200 tracker provides six-degrees of positional measurement, three rotational and three translational components. The translational values are provided by the tracker in the form of a 3 dimensional positional vector. The rotational values can be acquired from the tracker in one of three formats: Euler angles, 3x3 rotation matrices, or quaternions. Typically, when developing augmented or virtual reality applications, one would try to avoid using Euler angles in order to prevent a geometric anomaly known as gimbal lock. Gimbal lock is the effect of being caught in a rotational singularity when vertical rotations approach -90° or 90°. When this occurs, one of the rotational axes is canceled and the viewing position becomes locked around one axis. Rotation matrices and quaternions do not suffer from this vulnerability. However, the preproduction version of the IS-1200 suffers from geometric inconsistencies which result in translational values being presented in a right-handed coordinate system and rotational values being presented in a left-handed coordinate system. This inconsistency
rendered the 3x3 rotation matrix and quaternion values provided by the tracker unusable. However, by using Euler angles the problem can be resolved by multiplying the X-axis angle by -1. Fortunately, the head movements required for Experiment II presented little likelihood that an observer would ever experience gimbal lock.

The IS-1200 tracker uses a unique combination of inertial and optical tracking. The optical tracking component relies on its ability to see a user-defined constellation of tracking fiducials, see Figure 3.7. The constellation can be designed by the user and then programmed into the IS-1200's tracking database with the constellation tool provided by InterSense, see Section 3.4.3. The individual fiducials act as unique positional markers that correspond directly to a predefined 3 dimensional coordinate in the tracking database. The fiducials were generated and printed by the fiducial tool, also provided by InterSense, see Section 3.4.3. Ideally, the fiducials would be attached to the ceiling. The fiducial tool calculates the optimal diameter of the fiducials based on the field of view of the lens of the optical sensor, and the minimum and maximum tracking distances. In the case of a ceiling mounted constellation, the minimum and maximum tracking distance represent the tallest and shortest anticipated observer heights respectively. Once the fiducials are attached to the ceiling, their exact positions should be determined using a precision measuring device, such as a Total Station. Note that it is required that the down vector for the constellation be in the same direction as gravity.
The first attempt to create a trackable constellation was done by ceiling mounting the fiducials sans a Total Station. The ceiling in the building where both the lab and experimental environment was located was a standard suspended ceiling often found in office buildings. Though the 2 dimensional position of the fiducials along the general ceiling plane could be determined with relative accuracy, the suspended ceiling was not consistently level. This inconsistency caused small variations in the slope of the ceiling that were undetectable during a visual survey. The slope of the ceiling made it virtually impossible to guarantee that the constellation's down vector was in the same direction as
gravity. As previously noted, these two vectors must be in the same direction. Since the ceiling variations were very small, tracking errors were not immediately apparent, but would accumulate over time. The accumulated error would eventually result in a complete loss of tracking. The loss of tracking was caused by conflicts between the optical and inertial tracking data. The down vector observed by the optical sensor must match the down vector determined by the inertial sensor, which will always be in the direction of gravity. When they are inconsistent, the tracker attempts to resolve the error by integrating between the last known position and the conflicting optical and inertial data. Over multiple integrations, the error accumulates to the point that the data being reported by the tracker no longer reflects an approximation of its real-world position, resulting in a loss of tracking. However, it was found that obstructing the optical sensor's view of the constellation forced the tracker to rely solely on the data from the inertial sensor, which reset it to its last known stable position. However, once the optical sensor reacquired its view of the constellation, the error would again begin to accumulate and the process was repeated.

The next attempt at creating a usable constellation involved constructing a mounting board on which to affix the fiducials. The board was composed of an 8'x4' foam board, commonly found at arts and crafts stores, that was attached to a rigid frame made of 1”x1” wood supports that were connected and reinforced at the corners and midsection by metal 'L' braces that are typically used in furniture construction. The
fiducials were printed on 8.5”x11” sheets of paper and then placed on the mounting board one column at a time. The method used to arrange the fiducials on the board involved laying the board on the floor and affixing a measuring tape along its top and bottom edges. A laser level was then aligned at one end of the board so that the projected line intersected the same position on both the top and bottom measuring tapes. This line represented the center line for a column of fiducials. The fiducials were then placed at measured intervals along the projected line until the column was completed. This allowed for very accurate vertical alignment of the fiducials. To ensure accurate horizontal alignment a second laser level was used to align the fiducials along each row. Each fiducial was attached to the mounting board with a piece of clear tape at each corner. Once the constellation was completed, the distance between each fiducial and its horizontal and vertical neighbors was measured. If any measurement differed from the expected value by more than 2mm, then the fiducial's position would be re-measured and adjusted accordingly. The radial distance from the lower, rightmost fiducial to every other fiducial was also measured and compared to the calculated Pythagorean distance. Again, if the calculated distance varied by more than 2mm, the fiducial's position would be re-measured and adjusted. As a further precaution, any fiducial requiring a radial distance adjustment had its nearest neighbors’ distances checked again and adjusted if necessary.
Fortunately, this method removed that tracking error experienced with the previous attempt at a ceiling-mounted constellation. Though the method produced a constellation accurate enough to prevent loss of tracking, the error level was not low enough to prevent substantial jitter from occurring. Despite the jittering effects, this constellation allowed the research to move forward with general testing and development of the tracking and augmented reality applications. However, before more detailed refinement of these applications could take place, a more accurate tracking constellation needed to be developed.

The third attempt at attaining stable tracking involved attempting to reduce variability in the fiducials' positions. This was done by using the fiducial tool to export the fiducials to Encapsulated PostScript (EPS) formatted files. Once all the fiducials were exported, they were then imported into one of three 44”x36” Microsoft PowerPoint documents. Each of these documents represents one third of a single constellation and will be referred to as panels from this point forward. The fiducials in each panel could then be precisely positioned in software prior to being printed. Once each panel was prepared, the fiducials were printed en masse on a large format plotter. This removed the potential for per fiducial alignment error. The constellation could then be constructed by performing only three alignments, one for each panel. Once the panels were properly aligned, they were secured to the mounting board using tacks. For the most part, this method worked very well. However, there is one more potential source of error. Large
sections of paper seem more susceptible than small sheets of paper to changes in humidity and temperature. This caused very slight pinching and stretching of the panels around the areas where the tacks attached the panels to the mounting board. Fortunately, this did not seem to cause any noticeable tracking errors.

Figure 3.8

Three Panel Constellation

Once satisfied with the accuracy of the tracking constellation, attention was turned to placement of the mounting board. Horizontally hanging the mounting board parallel to the ceiling was considered, but rejected for the same reasons as the ceiling
mounted fiducials previously mentioned in this section. It was then attempted to vertically prop the mounting board against a wall or bookshelf. This proved troublesome as it was very difficult to assure that the down vector was exactly parallel with gravity. This resulted in unstable tracking. The method that proved most fruitful was a combination of these two methods, vertically hanging the mounting board from the ceiling like a pendulum. Hanging the board in this manner guarantees that the down vector is in the same direction of gravity since gravity itself is acting as the board's stabilizing force, see Figure 3.8.

3.6.1.2 Software

The software applications needed for this research were developed in a multi-level, incremental manner. The first step in making the software for this project was to develop an interface for the tracking system. The IS-1200 tracker comes with a software development kit that can be used to interface with the device and retrieve the translational and rotation information provided by the combined optical and inertial sensors. Sample code from InterSense was included and greatly facilitated rapid development of this aspect of the software.

Once the data could be collected from the tracker, it was then possible to empirically verify that the positional data reflected the tracker's real-world position relative to the tracking constellation. This was done by displaying a real-time stream of the tracker's translational and rotational values. These values were compared to the
observed position relative to the origin of the tracking constellation as the tracker was moved along the positive and negative directions of each axis. Once the translational values were confirmed, the same procedure was applied to checking the rotational values.

The next step was to integrate the tracker interface code into an application that applied the tracker's positional data to a 3D model in the form a geometric transformation. This involved making a simple OpenGL application that displayed the wireframe hallway model used in Experiment 1. As the positional data reported by the tracker changed, the model would rotate and translate accordingly. Applying this data to a geometric transform is a necessary step in implementing any virtual or augmented environment, as it allows for the observer to see a stable apparently solid virtual environment as the viewing direction changes.

The application was then extended to present the observer with a simple camera-in-hand virtual environment. This allows an observer to use the tracker as if it were a camera providing a view of a virtual world. The camera-in-hand model was implemented in two phases. The first phase was to display a single sphere at the tracking origin. It is important that this simple environment be tested first to ensure that the tracker is properly detecting the origin. Once this step was verified, a more complex model was created. This model consisted of an accurate 3 dimensional representation of the tracking constellation and a pyramid offset a specified distance from the constellation, both of which also existed in the real-world. The tracker could then be positioned relative to
these objects, and its viewing direction and position could be verified by comparing the observed real-world position with the virtual world position displayed on the monitor.

The next step in this process was to modify the application to support simple video see-through augmented reality. This step was considerably more complex than those previously mentioned and included several development phases. A Unibrain Fire-i digital firewire camera was used to provide streaming video of the real-world. This camera was selected because it is an inexpensive solution for acquiring video with a resolution of $640 \times 480$ pixels at 30 frames per second. High resolution and frame rate are important factors in producing a perceptually accurate video see-through augmented reality application.

Before any further work could be done, a mounting device was needed on which to place both the Fire-i camera and the IS-1200 tracker. The mount needed to be designed in such a manner as to allow for ease of handling yet still firmly hold the camera and tracking in exactly measured positions. The mounting device that was designed consisted of a small wooden block with the camera bolted to one end and the tracker attached to a small mounting bracket made of angled aluminum which was bolted to the opposite end of the wooden block, see Figure 3.9. The aluminum bracket and the wooden block were engineered in such a manner as to attach to each other in the same manner that the bracket would attach to the HMD. The area of the block between the
camera and tracker acted as a handle and allowed the experimenters to easily manipulate the entire apparatus.

![Figure 3.9](image)

**Figure 3.9**

Camera/Tracker Mounting Apparatus

For video see-through augmented reality to work properly, the field of view and aspect ratio of the camera must be accurately modeled in the virtual scene. According to the manufacturer provided specifications, the horizontal and vertical fields of view are 42° and 32° respectively [18]. These values were also empirically measured to be 38.9° and 31.1° respectively. The fields of view were determined by placing another block of wood with black tape marking both ends centered in front of the camera. The height of the camera was then adjusted so that the lens was aligned with the middle of the board. Then, while watching the video feed from the camera, the board was moved inward and outward until the edges of the black tape was no longer visible, see Figure 3.10. At this
point the distance from the camera lens to the board was measured. The field of view could then be calculated using the follow formula:

\[ \theta = 2 \cos \frac{y}{2 \Delta z} \]

Figure 3.10

Measuring the Camera’s Vertical Field of View

An OpenGL application was then developed that displayed tracked 3D geometry on top of a real-time video stream from the Fire-i Camera. The videoInput library, which is an open source library used to interface with DirectShow compatible video devices, was used to capture the video stream [19]. Additionally, a stack of transformation matrices was created in order to better handle the series of transforms that needed to take place from tracker space, to camera space, to virtual space. The approach used to handle
these transformations is similar to that suggested by Robinett and Holloway [11]. The main difference between their implementation and that discussed here is that all transformations are stored in the form 4x4 matrices as opposed to the Vector, Quaternion, Scalar data structures advocated by Robinett and Holloway.

After successfully implementing the video see-through augmented reality application, it was then time to attach the tracking device to the HMD and begin work on creating an optical see-through augmented reality system. The display used for Experiment II was an NVIS nVisor ST. The horizontal and vertical fields of view provided by the manufacturer, 48° and 40° respectively, were used to model the virtual scene as these values could not be empirically determined. This was due to the display area being slightly larger than the visible area, which prevented accurate measurements of the display edges. However, HMDs produced by nVis come with a build report that lists the exact specifications that were measured and verified for each particular display. The values provided by the build report were used in calibrating the virtual scene. The optics of this display are collimated, meaning that they are focused at optical infinity. The optics converge to near infinity with an inner rotation of 0.6°. No optical distortion was observed.

Since the software to perform this task had been developed incrementally, this step was relatively easy to implement. The exact position of the tracker relative to the midpoint between the HMD's exit pupils had to be measured and the chain of
transformation matrices modified accordingly. Also, a stereoscopic scene needed to be generated. For this both the physical IPD of the display elements and the virtual IPD had to be adjusted on a per observer basis. Then a 3D representation of the virtual scene was rendered based on the position of each eye and the convergence angle of the display optics.

The nVisor ST accepts dual channel DVI input which was provided by means of an dual head nVidia GeForce 6800. The 3D scene was displayed using a passive stereo technique. In passive stereo a single window twice the size of the desired scene is displayed with the right eye view displayed on the right half of the window and the left eye view displayed on the left half. This method also made debugging simpler since it could easily be displayed on two monitors with a single horizontal spanning desktop with the left and right inputs inverted. The developer could then get the gist of how the scene should look by cross-eye viewing the result on two standard desktop monitors. Once this step was completed and fully debugged, the optical see-through augmented reality system was ready for use in Experiment II.

The data collection and scene generation for this experiment was handled by a Pentium 4 3.2Ghz desktop computer with a NVIDIA GeForce 6800 graphics card for driving the HMD and an NVIDIA Quadro NVS 285 for displaying diagnostic information and control panels on a separate display. The software for this experiment
was implemented in C++ using Microsoft Visual Studio, Cygwin, OpenGL, GLUI, and Perl.

3.6.2 Experiment Setup

Observers in Experiment II were required to perform the same visually directed walking tasks used in Experiment I. The HMD was tethered to the display controller and host computer with only 4 meters of cable. Since observers would be required to walk distances greater than the length of the tether, the display controller and host computer were placed on a cart which was pushed behind the observers during the directed walking tasks. As an experimenter pushed the cart, an extension power cable unrolled behind them. When observers had to return to their starting position, a second experimenter pushed the cart back from the opposite side while the experimenter who originally pushed the cart rolled up the extension power cable, see Figure 3.11.
The nVisor ST plus IS-1200 tracker and cables weighed 2.45kg. A large portion of this weight was on the HMD's front, where the optics and display elements are located. For many observers, this was very uncomfortable. To help make the display more comfortable, observers were offered a headband which acted as additional padding to help distribute the weight of the display. Also, the weight of the cable would often make the display shift on the observer's head. In order to prevent this from happening, the observers were required to wear a backpack onto which the cables were clipped. This prevented the cables from moving and causing the display to shift.

Observers were also required to wear earphones during the experiment. The earphones were connected to a 6 gigabyte Apple iPod Mini MP3 player that played an MP3 of white noise through the duration of the experiment. This was done to prevent the
observer from hearing potential auditory cues that may provide an alternative depth measure, such as the foot steps of the experimenters as they are placing the object. The white noise also acted to help further isolate the observer from outside auditory distractions. The observers also wore a battery-powered wireless audio receiver similar to those typically used in theater and broadcast production. An experimenter wore a wireless microphone tuned to the same channel as the observers' receiver. It was through this microphone that the experimenters communicated instructions to the observers. As an additional precaution, the experimenters emptied their pockets of anything that may cause noise, such as keys, loose change, and cellular phones. The hallway was carpeted, so it was not necessary in this experiment for the experimenters to remove their shoes.

The earphones used for most of the observers in this experiment were small in-ear earphones. However, two observers were incapable of wearing them due to very small ears. These observers were given a pair of over-ear earphones to wear. It was preferred for observers to wear the in-ear earphones, as the over-ear earphones pressed against the HMD, making them less comfortable. Both sets of earphones were cleaned, disinfected, and had their covers replaced before each observer wore them.

As with Experiment I, prior to beginning the experiment, the experimenters would gather several measurements from the observers, specifically the experimenters would measure the observers' interpupillary distance and eye height. An experimenter would first measure the observers' eye height using a survey rod. Once the eye height had been
measured, an experimenter would place a cross of white masking tape on a dark door surface at the end of the hallway at the measured eye height, see Figure 3.12. Next, an experimenter would measure the observers' interpupillary distance to the closest millimeter using a small ruler. During this procedure the observers were asked to focus on the cross at the end of the hallway. It was observed with pilot observers that they tended to focus on the person measuring their IPD if not told to do otherwise. Observers' IPDs change as their gazes converge or diverge, so it was necessary to measure the observers' IPDs while they focused at some medium field distance. Both the eye height and IPD were only measured once per observer and then entered into the software that generated the 3D scenes.
The virtual scene was calibrated before every block of trials for the Virtual Reality and Augmented Reality conditions. This was done to make sure that ocular alignment, HMD placement, and tracker leveling were all corrected. First the ocular alignment and HMD placement were adjusted. This calibration procedure was adapted from that described by Rolland [12]. First, the observers placed the HMD on their heads in as comfortable a position as possible. Then the observers were asked to turn and face a blank wall and to close their right eye. At this point, a series of concentric circles and a
small crosshair were displayed through the HMD, see Figure 3.12a. The observers were then asked to use the mechanical knob on the left side of the HMD to horizontally adjust the view until the circles and crosshair appeared centered and they could see an equal amount of the outermost circle on both the left and right sides of the display area. They were then instructed to close their left eye, open their right eye, and perform the same adjustment with the knob on the right side of the HMD. Once this was done, the observers were instructed to use the knob on top of the HMD to vertically adjust the display area up and down until they could see an equal amount of the outermost circle on both the top and bottom of the display area. At this point, it was assumed that the HMD and the eyes were properly aligned along the optical axis.

Figure 3.13
An Observer Performing the Calibration
Every observer wears the HMD slightly differently on their heads. Because of this, the position reported by the tracker will be biased by these differences. For this reason it was necessary for the tracker's translational and rotational components to be adjusted. To do this, the observers were shown the same concentric circles and crosshair mentioned above. The observers were then asked to look down the hall and move their head until the crosshair was centered on the masking tape cross at the end of the hall, see Figure 3.12a. Then a yellow X was displayed in the HMD, see Figure 3.12b. The initial location of this X represents where the tracker was reporting the observers' gaze to be located. At this point, the observers were handed a game controller and instructed to use it to adjust the position of the yellow X until it was centered on the crosshair and the real-world cross, see Figures 3.12b & 3.13. This procedure simply adds offsets to the X and Y axis values reported by the tracker. The Z axis did not require adjustment as the design of the HMD assured that regardless of orientation of the observers' heads, the tracker was not translated anymore than roughly 2cm along the Z axis.
The next step in the calibration was to compensate for rotational offsets. For this procedure the crosshair displayed in the center of the concentric circles changes to a 3D crosshair that moves based on the roll, pitch, and yaw reported by the tracker. The experimenters referred to this 3D crosshair as the compass, see Figure 3.14. At the center of the compass was a white crosshair, the upper right quadrant of the compass consisted of red line segments extending from the white crosshair, and the lower left quadrant of the compass consisted of green line segments extending from the white crosshair. The observers were instructed to move their heads until the center of the compass was aligned.
with the masking tape cross at the end of the hall. The observers were then given a game controller and instructed to use it to adjust the compass until it turned into a +, see Figure 3.12c. If the tracker was reporting rotational values that were not consistent with the position of the observers' head, the compass would appear to have a star-like or polygonal shape, because points (a) and (b) in Figure 3.14 would not line up along the optical Z axis. However, when the tracker is correctly aligned, the compass looked like a + with red lines composing the upper right quadrant and green lines composing the lower left quadrant. This technique was effective for visualizing the rotational offsets reported by the tracker as simple two dimensional shapes produced by accidental views of the compass’ three dimensional form. To the knowledge of the experimenters, this compass method for calibrating rotational effects is novel to this experiment.

Experiment II took place in the central hallway at the Institute for Neurocognitive Science & Technology. The location used in Experiment I at this facility was no longer available for this experiment as a large portion of the room is now occupied by a high-resolution display wall which spans the entire width of the room. The hallway was 1.82m in width and 23.45m in length. Observers stood facing south at 6.92m from the north wall and 1.06m from the east wall. For the Virtual Reality condition a photorealistic model of this location was constructed. It was designed in such a manner as to accurately reflect the look and feel as the real-world location, see Figure 3.15. This
included texture mapping the scene with images taken from photographs of the actual location.

Figure 3.15
Experimental Location (left) and Virtual Model (right)

Another corridor intersected this hallway at 5.3m, spanning 1.67m, from the observers' standpoint. Pilot observers indicated that a significant luminance change was detectable when walking past this corridor, even with their eyes closed. This luminance change could act as a confound. To mitigate this risk, a bed sheet that was roughly the same color and texture as the wallpaper was hung from the ceiling and taped against the wall to simulate a smooth uninterrupted surface, see Figure 3.16.
All practice trials took place in an adjoining hall that runs east to west and measures 1.82m in width and 16.59m in length. Observers stood 2.2m from the east wall facing west.

The same real target object from Experiment I was used in Experiment II. However, for Experiment II, a polygonal, photorealistic virtual model was used instead of a wireframe model. Both the virtual and real objects were pyramids that were 23.5cm tall with a 23.5cm square base. A fiberglass survey tape was attached to the floor. This allowed the experimenters to measure the distance walked by the observers after performing each task.
Observers were given two depth cue conditions as well. Observers were asked to view objects while standing as still as possible (still condition) and while swaying back and forth from left to right (motion condition). The still condition was meant to approximate the viewing conditions in Experiment I. The motion condition was meant to provide the observer with the additional depth cue of motion parallax.

Observers were presented with four different environments: Real, Real+HMD, Virtual Reality, and Augmented Reality. The Augmented Reality environment corresponds to the virtual stimulus type in Experiment I. There is no equivalent environment in Experiment II to the Real+Virtual stimulus type from Experiment I. This is due to slight jitter in the position reported by the tracker. These small jitters made the virtual and real objects appear to be too disjoint to be recognized as a single object. Instead, a purely Virtual Reality environment was included in this experiment. This would allow for more direct comparison to the wide body of virtual reality depth perception research that currently exists. The target objects placed in these environments ranged from 2.5m to 8.5m. However, only responses at 3m, 5m, and 7m were used for analysis, so that the results of Experiment II could be more easily compared to those from Experiment I.

3.6.3 Experimental Task

Observers performed visually directed walking tasks under two cue conditions in four experimental environments. Observers were allowed to view the stimuli for any
length of time they required and were not timed. However, an average response time was empirically estimated to be approximately 3 to 7 seconds after stimulus onset.

The protocol used for Experiment II was very similar to that used in Experiment I and was consistent across all observers. Before beginning the experiment, an experimenter gave the observers the following instructions in a conversational manner:

Visually directed walking, or blind walking, is walking to an object with your eyes closed. We will show you an object and ask you to look at it until you feel like you have a very good sense of where it is located in space and feel as though you can walk to it with your eyes closed. For instance, when you wake up at night and need to get a drink of water, it’s probably completely dark in your bedroom, but you can always walk to the light switch even though you cannot see it. We want you to have the same kind of sense of the target object's position. When you feel that you can walk to the object with you eyes closed, let me know you are ready and I will ask you to walk forward. I want you to walk until you feel as though the tips of your toes are at the center of where the object is located. When you stop walking, you may open your eyes, but do not look down at your feet. The object will be removed from the scene, so there will be no chance of you stepping on or tripping over the object. We will also be walking near you during the experiment to help you in case you loose balance or become
disoriented. We will also stop you before you walk too close to the walls or any obstacle in the scene.

We will also be asking you to close your eyes very frequently. It is important that you do not “squint” your eyes and face when trying to close your eyes. This uses a lot of muscles in your face and they will get tired quickly. This might cause you to accidentally open your eyes. We suggest that you simply relax your face and close your eyelids and perhaps lower your head slightly. If you accidentally open your eyes, please let us know.

Now we will give you some practice with visually directed walking, so you can get comfortable walking with your eyes closed.

At this point observers were given five practice directed walking trials in a hallway that was separate from the hallway where the experiment took place. A small soccer ball patterned beanbag was used as the target object. The first three practice trials were done standing still with the target object placed with in ranges from 2m to 5m and 5m to 9m. One practice trial was given in the 2m to 5m range and two were given in the 5m to 9m range. It was empirically observed with pilot observers that they were most hesitant to walk longer distances. This is why two of the first three trials were in the 5m to 9m range.
After the first three practice trials, the observers were given additional instructions:

Now, we are going to ask you to sway back and forth while viewing the object. This means, just rock from side to side so that your heels leave the floor. Do not rock your head back and forth, but move your whole body from side to side. You should only sway when looking at the object. When you close your eyes, please stop swaying. Also, do not try to sway and walk at the same time. You only need to sway while looking at the object.

An experimenter would then demonstrate the swaying motion and have the observer sway with him. The final two practice trials were both done with the motion condition. One of the trials was in the 2m to 5m range and the other was in the 5m to 9m range.

At this point the observer was escorted to the hallway where the experiment would take place. Observers were then asked to stand with their feet centered on the metal clasp at the end of the survey tape. The floor was then marked with masking tape where the tips of the observers' toes were located. The observers were then told that these marks were their starting point and to center up with the survey tape with their toes behind the marks when they returned from a directed walking trial. Once the observers
were in place, they were asked to close their eyes. At this point the experimenters would place the target object in the scene. The protocol for each trial would then go as follows:

**Experimenter 1:** Open your eyes.

Start swaying back and forth. *(This instruction was omitted for the Still condition)*

Observe the object and tell me when you are ready.

**Observer:** *(The observer views the object and indicates readiness)*

**Experimenter 1:** Close your eyes and walk forward.

*(Experimenter 1 pushes the cart behind the observer, allowing the extension power cable to unroll behind him)*

**Experimenter 2:** *(Experimenter 2 removes the target object during the Real and Real+HMD conditions)*

**Observer:** *(The observer walks forward with eyes closed)*

**Experimenter 1:** *(Experimenter 1 walks behind the observer, watching for signs of disorientation)*

**Experimenter 2:** *(Experimenter 2 walks in front of the observer, watching for signs of disorientation)*

**Observer:** *(Stops when they believe they are at the target distance)*

**Experimenter 1:** *(Experimenter 1 records the distance walked by the observer)*
Turn to your right. (*This prevents the observer from getting tangled in the HMD and tracker cables*)

Return to your starting position, center up, and close your eyes.

**Experimenter 2:** (*Experimenter 2 pushes the cart back to the starting position*)

**Experimenter 1:** (*Experimenter 1 rolls up the extension power cable, ensuring that the cart does not roll over any cables*)

**Experimenter 1:** (*In the Real and Real+HMD conditions, Experimenter 1 indicates the position of the next target to Experimenter 2*)

In the event that the experimenters felt that observers risked walking into an obstruction, the observers were asked to stop walking and open their eyes. The observers were then instructed to return to the starting point. For programmatically displayed stimuli, such as in the Virtual Reality and Augmented Reality environments, the experimenter would move back two trials and start again. This was done in order to prevent observers from seeing the same stimulus twice in a row. For the Real and Real+HMD environments, the experimenters would randomly select a position exclusive of the last stimulus condition and then resume the interrupted trial afterwards.

### 3.6.4 Experimental Design

Experiment II utilized a within subjects design. The designed controlled for both environment type and motion condition. A 4x4 Latin Square for environment type and a
2x2 Latin Square for motion condition was used to counter-balance the exposure order of stimuli on a per observer basis. This ensured that for every group of 8 observers every presentation order and estimation protocol was covered. Observers were shown stimuli at distances of 3m, 5m, and 7m with two repetitions at each distance per block for each experimental trial. To help prevent observers from noticing the repeating distances, 25% of the stimuli were noise trials which were randomly selected in 0.5m increments from the 2.5m to 8.5m range, exclusive of the 3m, 5m, and 7m. The noise trials were excluded from analysis.

For Experiment II, a total of 19 observers participated in the experimental task. However, three observers were unable to complete the necessary tasks. One of these observers could not continue due to a software failure. One observer was prone to migraines and stopped the experiment for fear of inducing a headache. Another observer was unable to perform the calibration procedure.

A total of 16 observers completed the tasks for Experiment II. A total of 768 data points were collected for analysis: 16 observers, 4 environments, 2 parallax conditions, 3 distances, and 2 repetitions.
CHAPTER IV
RESULTS & ANALYSIS

Several dependant measures were considered for Experiments I and II. These were judged distance, error, and normalized error. Judged distance is the distance reported by the observer when using the verbal report technique and the distance walked by the observer when using the directed walking technique. Error is defined as the judged distance minus the target distance. Normalized error is a percentage of judged distance to actual distance. It is necessary to note that even though normalized error is similar to accuracy, they are not the same. For instance, 100% normalized error reflects an equal amount of both under- and overestimation and not that all judged distances were completely accurate.

4.1 Experiment I

One of the goals of Experiment I was to determine if the depth estimation techniques used in this study were suitable for Augmented Reality. For this reason, the individual protocols were examined on a per observer basis. When comparing the relative error measured between verbal report and directed walking, it becomes immediately apparent that the results obtained through verbal report are considerably less consistent with variations ranging from severe overestimation to severe underestimation,
see Figure 4.1b. This is not to say that verbal report is an inappropriate protocol, but that
the particular implementation used for this study may not have been properly executed.
Visually directed walking, on the other hand, proved to be a very stable protocol across
all observers, see Figure 4.1a. For this reason, all further analysis discussed in this
section will refer only to the results obtained from the directed walking trials.

When examining the data collected in Experiment I, all of the classic signs of
underestimation typically seen in Virtual Reality become apparent, see Figure 4.2. This
figure presents the environments and estimation techniques from Experiment I in a
separated manner. This is done to help more clearly present the data. This figure also
helps visualize the large degree of underestimation found with the verbal report
technique. When conducting a repeated measures analysis of variance of all conditions, including the control condition, a significant difference is found, $F(3, 15)=5.89, p=0.002$. However, when conducting an ANOVA on just the experimental conditions, no significant differences were found, $F(2, 15)=3.14, p=0.058$. This indicates that the previously found significance was due to differences between the experimental conditions and the control condition.

![Experiment I Results](image)

Figure 4.2

Experiment I Results
The first thing that becomes apparent when looking at the results from Experiment I, is that the underestimation effects seen in VR seem to be present in AR as well. This indicates that the mechanisms that contribute to spatial misinterpretation in VR may also affect AR. Considering that the technology used in AR and VR are very similar, this does not seem to be too abstract of a theory. However, when considering the fact that the experimental conditions did not significantly differ from each other, an interesting finding emerges. The Real+HMD condition was very similar to the control condition and consisted of a real-world object viewed through the HMD with no virtual augmentation. The only difference between this condition and the control condition is the presence of the HMD and crossbar apparatus. The observers’ foreknowledge of the crossbar in front of them could be the cause. On the other hand, the HMD itself may be causing perceptual interference.

Figure 4.3 depicts the mean error versus target distance by stimulus type. This figure makes clear the magnitude of underestimation experienced in the experimental conditions relative to the control condition. Also, it can be observed that the level of underestimation increases as distance increases.
4.2 Experiment II

Experiment II was intended to test the effect of motion parallax on depth perception in Augmented Reality. Since the results from Experiment I demonstrated that underestimation effects are present in AR, it was suspected that the addition of motion parallax may produce a more stable perception of depth relations. In Experiment I, observers performed both verbal report and directed walking estimation techniques; however, observers in Experiment II only performed directed walking. For this reason, the per observer error for directed walking in Experiment II was visually compared to that of Experiment I, see Figure 4.4. From this figure, it can be clearly seen that directed
The motivation behind Experiment II was to determine if the addition of motion parallax would allow observers to more accurately judge the egocentric depth to target objects. To test this, observers viewed objects in Still and Motion conditions. Figure 4.5
presents the environments and conditions from Experiment II in a separated manner. In this figure, very little difference can be seen between these two conditions. To further verify this observation, the error in these conditions was compared using a repeated measures analysis of variance. No significant differences were seen when comparing the Still and Motion conditions, \( F(1, 15)=3.20, p=0.94 \). Though no widespread effect of motion parallax was found across all environments, each environment was individually analyzed in order to detect any environmentally specific effects. The only environment exhibiting a significant effect of the Motion condition was the Real+HMD environment, \( F(1, 15)=9.23, p=0.008 \). As can be seen in Figure 4.5, the application of the Motion condition to the Real+HMD environment caused significant underestimation.
When doing a per observer analysis of the data from Experiment II, it became apparent that observers could be placed into one of three groups based on their overall normalized error: those who overestimated by 15% or less, those who underestimated by 15% or less, and those who underestimated by more than 15%. We can further generalize these groups as those who were within ±15% of the actual target distance and
those who were beyond ±15% of the target distance, see Figure 4.6. Three of the observers fell in the latter category. If those observers were removed from the analysis and the data were plotted again, it can be seen that the general trends in the data do not change, but the means are all shift towards the veridical. In particular, it can be seen that performance in the Real and Augmented conditions becomes almost entirely veridical, see Figure 4.7.

---

**Figure 4.6**

Normalized Error by Observer in Experiment II
The most remarkable result seen in the analysis of the data gathered from Experiment II was a marked lack of underestimation. When viewing the normalized error, it was found that none of the environments suffered less than 90%, which was in the Virtual Reality environment. The highest normalized error was in the Augmented Reality environment with 96%. The Real+HMD environment resulted in 92%
normalized error and the Real environment, the control condition, resulted in 94% normalized error. These values are truly remarkable when compared to those found in similar studies, as can be seen in Table 4.1. Only the results found by Interrante in 2004 reflect less underestimation in the Virtual Reality environment [6]. Thus far, the research discussed in this document is the only research where visually directed walking is used to measure depth judgments in Augmented Reality. For this reason, there can be no direct comparison made to other studies using this same depth judgment technique for the AR condition. However, Swan et al.'s perceptual matching task in Augmented Reality resulted in a normalized error of 99% [14]

Figure 4.8 depicts the mean error versus target distance by environment type. The level of underestimation seen in this figure is clearly less than that seen in Experiment I. It is also apparent that observers did not exhibit the same trend of increasing underestimation with increasing distance.
<table>
<thead>
<tr>
<th>Study</th>
<th>Distance (m)</th>
<th>Normalized Error</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wittmer &amp; Sadowski (1998)</td>
<td>4.6 - 32</td>
<td>92%</td>
<td>Treadmill Walking</td>
</tr>
<tr>
<td>Real Environment</td>
<td>4.6 - 32</td>
<td>85%</td>
<td>Treadmill Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knapp (1999)</td>
<td>5 - 15</td>
<td>100%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Real Environment</td>
<td>5 - 15</td>
<td>42%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>2 - 8</td>
<td>65%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Durgin, Fox, Lewis, &amp; Walley (2002)</td>
<td>2 - 5</td>
<td>100%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>2 - 5</td>
<td>81%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Willemsen &amp; Gooch (2002)</td>
<td>2 - 5</td>
<td>81%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Real Environment</td>
<td>2 - 5</td>
<td>95%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>2 - 5</td>
<td>44%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Thompson, et al. (2004)</td>
<td>5 - 15</td>
<td>95%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>5 - 15</td>
<td>44%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Interrante, et al. (2004)</td>
<td>3 - 9.2</td>
<td>100%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Real Environment</td>
<td>3 - 9.2</td>
<td>95%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Virtual Environment, Base</td>
<td>3 - 9.2</td>
<td>96%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Virtual Environment, Post</td>
<td>3 - 9.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willemsen, et al. (2004)</td>
<td>4 - 8</td>
<td>100%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Real Environment</td>
<td>4 - 8</td>
<td>85%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>4 - 8</td>
<td>62%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Real Environment</td>
<td>4 - 8</td>
<td>88%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Real+HMD Environment</td>
<td>4 - 8</td>
<td>67%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>4 - 8</td>
<td>47%</td>
<td>Triangulated Walking</td>
</tr>
<tr>
<td>Swan, et al. (2006)</td>
<td>5.25 - 44.31</td>
<td>99%</td>
<td>Perceptual Matching</td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Environment</td>
<td>3 - 7</td>
<td>88%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Real+HMD Environment</td>
<td>3 - 7</td>
<td>83%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Combined Environment</td>
<td>3 - 7</td>
<td>82%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Augmented Environment</td>
<td>3 - 7</td>
<td>79%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Environment</td>
<td>3 - 7</td>
<td>94%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Real+HMD Environment</td>
<td>3 - 7</td>
<td>92%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>3 - 7</td>
<td>90%</td>
<td>Direct Walking</td>
</tr>
<tr>
<td>Augmented Environment</td>
<td>3 - 7</td>
<td>96%</td>
<td>Direct Walking</td>
</tr>
</tbody>
</table>
4.3 Comparison and Discussion

A comparison of both experiments reveals significantly less underestimation in Experiment II than in Experiment I, $F(1,30)=4.95, p=0.034$. Since there were no widespread effects of motion parallax, what could be the cause of this enhanced performance? One possibility, proposed in Wu et al., is that a restricted vertical field of view could be a source of perceptual interference [22]. Wu et al. studied the depth judgments of stationary observers viewing real-world objects over a range of fields of view. This is very similar to the Real+HMD condition used in Experiments I and the Real+HMD, Still condition in Experiment II. Also, two of these fields of view closely
resemble the fields of view of the HMDs used in Experiments I and II. The HMD used in Experiment I has a vertical field of view of 20°. The mean underestimation found in Experiment I for the Real+HMD condition was 0.54m. Wu et al. found a mean underestimation of 0.41m when viewing objects through a vertical field of view of 21.2°. This value is within a 95% confidence interval of the mean underestimation found in Experiment I (0.54m ± 0.26m). The HMD used in Experiment II has a vertical field of view of 40°. The mean underestimation found in the Real+HMD, Still condition was 0.23m. Wu et al. found a mean underestimation of 0.04m when viewing objects through a vertical field of view of 38.6°. This value is also within a 95% confidence interval of the mean underestimation found in the Real+HMD, Still condition in Experiment II (0.23m ± 0.21m). This seems to provide strong evidence that the larger field of view available in Experiment II may be responsible for the improved performance.

The collimated optics of the nVisor ST used in Experiment II may also have played a large role in the outcome of the data. Many HMDs have a fixed optical focus only a few meters from the observer, including the Sony Glasstron used in Experiment I which has an optical focus of 1.2m. This means that regardless of where the observers are looking in the scene, they will be accommodating to 1.2m and all other depths will be in accommodative conflict. There is practically no change in ocular accommodation past roughly 5m. At this point the eyes are focused at optical infinity. Since the nVisor used in Experiment II is a collimated display, its optics are focused at optical infinity. This
means that objects viewed from roughly 5m outward would not be in accommodative conflict, but objects closer than 5m would be in conflict. If this is the case, then systematically varying the HMD’s optical accommodation may yield interesting results. However, that is beyond the scope of the current research.

The simple fact that views of the virtual elements were head-tracked in Experiment II may also explain the better performance and the lack of a difference between the Still and Motion conditions. In Experiment I, the view of the world was completely stationary since the HMD was rigidly mounted. Even though observers in Experiment II were asked to stand as still as possible, it is virtually impossible to hold one’s head perfectly still. For this reason, there was always a small degree of motion present when viewing stimuli in Experiment II. This small degree of movement may have been sufficient to enhance perception with motion parallax.

The unusual underestimation encountered when viewing objects in the Real+HMD environment in the Motion condition in Experiment II is rather perplexing. One possibility is that a combination of degraded luminance, HMD ergonomics, and other perceptual issues could have resulted in an interesting interaction that produced consistent underestimation. However, this explanation is purely speculative and additional research, that is beyond the scope of this document, would be necessary to shed further light on the subject.
CHAPTER V
CONCLUSIONS

This thesis presents new results from two experiments testing observers’ abilities to judge the apparent distances to objects in virtual, augmented, and real-world environments both with and without the aid of motion parallax. No significant, widespread effect of motion parallax was found, but marked improvement in depth judgments was observed in Experiment II as compared to Experiment I. Possible explanations for this improvement include HMD improvements, changes in accommodative demand, widened fields of view, and the addition of motion tracking. Each of these possibilities could constitute future research in the field of augmented reality.

Additional contributions were also made in this thesis. A detailed account of the calibration procedures used in this study was given, as well as a novel compass based method for calibrating an augmented reality display. A highly detailed step-by-step documentation of the procedure used to develop the augmented reality system used in this study was also provided. A more general contribution was a thorough analysis of the normalized error of a wide range of virtual and augmented reality depth perception
studies. These helped to gauge the results of the experiments against those found by the broader AR and VR community.
REFERENCES


