

0

Distortion of Depth Perception in a Virtual Environment Application

by

Jonathan D. Pfautz

Submitted to the Department of Electrical Engineering and Computer Science in partial
fulfillment of the requirements for the degrees of

Bachelor of Science in Computer Science and Engineering
and Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1996

© Jonathan D. Pfautz, MCMXCVI. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and
electronic copies of this thesis document in whole or in part, and to grant others the right to
do so.

Author
Department of Electrical Engineering and Computer Science
May 28th, 1996

Certified by
Nathaniel I. Durlach
Senior Research Scientist, MIT Research Laboratory of Electronics
Thesis Supervisor

Accepted by
F. R. Morgenthaler
Chairman, Departmental Committee on Graduate Theses

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

Eng.

JUN 11 1996

Distortion of Depth Perception in a Virtual Environment Application

by

Jonathan D. Pfautz

Submitted to the Department of Electrical Engineering and Computer Science
on May 28th, 1996, in partial fulfillment of the requirements for the degree of
Master of Engineering Electrical Engineering and Computer Science

Abstract

Virtual Environment (VE) technology suffers from a variety of hardware and software problems that interfere with the ability to present the experience of being immersed in a synthetic world. Most notably, the poor quality of the viewing surfaces in head-mounted displays has led to many complications in the development of VE systems. These visual displays have a low spatial or pixel resolution that creates several side effects. Because the ability to present depth cues is hampered by the low resolution, the presentation of a sufficiently realistic three-dimensional world is very difficult. Depth perception is distorted by the lack of resolution in two ways. One, the threshold distance at which an object is just visible is much closer than in the real world. Two, the error and variance in estimating the depth of an object are significantly greater than in the real world. This thesis presents a series of mathematical analyses of these distortions, including a model of perspective geometry and the relevant human psychophysical biases. Algorithms are presented that improve both distorted threshold depth perception and distorted depth estimation. Experiments have been conducted to evaluate the usefulness of the algorithms in terms of both perceptual performance and training performance. The ability of the geometric models to predict human performance in the experiments serves as a measure of the models' effectiveness in devising solutions. The findings suggest that the visible range can be extended using the proposed algorithms without significantly affecting the ability to estimate depth.

Thesis Supervisor: Nathaniel I. Durlach

Title: Senior Research Scientist, M.I.T. Research Laboratory of Electronics

Acknowledgments

In an effort to streamline my acknowledgments section, I have used the concept of *Von Neumann Utility* to determine the subjective value of each person who influenced this thesis. Von Neumann Utility is used to determine the relative worth of an event or object to a decision maker or system user. For example, if we were attempting to determine the value of a dollar for a particular person, we would present them with a choice between $x\$$ and a 50-50 chance of receiving \$1000 or \$0, then ask for what value of x would they be indifferent between x and the lottery. By dividing the value x by 1000, we get a value for the utility of one dollar. Usually people are risk-averse and will choose a value of around \$350, yielding a normalized value of .35 utiles. By varying the reward in the lottery, a utility curve can be constructed to show the increase or decrease in utility as a function of the size of the objective (Sheridan, 1992b).

The idea of utility can be extended to determine the value of the people who have, in some manner, helped with my thesis. However, given that completing a thesis is essential to completing graduation requirements, merely valuing the assistance I received in terms of money seems meaningless. Thus, to evaluate the relative worth of the help I received, I devised a new scenario. For each individual, I posed the following question:

"Given a choice between a particular person's help on my thesis and a 50-50 chance of losing body part x , which would I choose?"

Then, to determine the value of their help, I found a body part (or parts) that I would be willing to risk. Body part refers to a limb or some other easily defined physical extension of my self including fluids.

The results of this self-inquiry are summarized in Table A. Given the difficulty in assessing a numerical value for the importance of a body part, the construction of a utility curve has been avoided. Future work may include modeling the importance of a body part in terms of degrees of freedom of movement in that limb or simply in terms of mass.

The list of people in the table is ordered roughly in terms of the value of their help, with the exception of Nat Durlach and Dick Pew, who are listed in alphabetical order for obvious political reasons.

Nat Durlach	Legs, below mid-shin
Dick Pew	Left arm, from above shoulder
Nora Luongo	Kneecap, right
Lewis Girod	Three fingers (not index), right hand
Thomas von Weigand	Eyelids
Sam Madden	Thumb, left hand
Dorrie Hall	Patch of skin, thigh, 5 cm ²
Greg O'Connor	Lesser 4 toes of right foot
Walt Aviles	Pinkie finger, 1 cm above knuckle
Yvette Tenney	Big toe, right foot
Bill Levison	One roll of belly flab
Wendelin Sachtler	Canine teeth (4)
Barbara Shinn-Cunningham	Lower lip, 0.67 cc
White (Blue) Kitty	All body hair
Fran Taylor	Bony protrusion on back of my skull
Evan Wies	Bellybutton
Melissa Poh	Nipple, left
Maroula Bratakos	Blood, 1 half-pint
David Getty	Small toe of left foot
Kathy Youngbear	Skin of chin, 1 cm ²
Alexis Black	Rear end, 1.5 cc
Carolyn Alsterberg	Molars
Aditya Matoo	Tip of nose, 1 mm
Michele Bierbaum	Earlobe, right
Jim Bandy	Eyebrows
Jeanne Krause	Eyelashes, upper
Johnny Yoon	Fingernails (6)
Clara Yang	Wisdom teeth
Doug Allen	Saliva, 15 ml
Alex Techet	Knuckle skin, left hand
Rakesh Gupta	Foot calluses
M.I.T. Women's Water Polo Team	Toenails

Table A: The utility value of the assistance I received in writing this thesis, broken down by individual. Relative value is given in terms of subject importance of body parts.

Finally, I'd like to acknowledge the United States of America as being a financially swell place to conduct research by extending my humble thanks to those folks at the Department of Defense, the U.S. Navy, the Office of Naval Research, the Naval Air Warfare Center Training Services Division, M.I.T. Research Lab for Electronics, and B.B.N. Corporation all of whom helped to keep me from starving.

Contents

1. Introduction	7
2. Background.....	7
2.1 Definition of Terms.....	8
2.2 Virtual Environment Equipment	13
2.3 Advantages of Virtual Environments.....	16
2.4 Virtual Environments Applications	17
2.5 Problems with Virtual Environments.....	18
2.6 Solving Problems with Virtual Environments.....	24
2.6.1 Example Solutions	24
2.6.2 Return to the Discussion of Realism.....	27
2.6.3 Task Dependence.....	27
2.6.4 Task Analysis.....	29
2.6.5 A Significant Example: The Visibility Problem.....	30
2.6.5.1 Background Geometry.....	31
2.6.5.2 Returning to the Big Picture.....	51
2.6.5.3 Implications	52
3. Experiment.....	53
3.1 Experiment Background	53
3.1.1 Geometry of the OOD model	56
3.1.1.1 Determining Real World Visibility.....	63
3.1.1.2 Assumptions	65
3.1.1.3 Previous Work on Visibility in the OOD Simulator	66
3.1.1.4 Finding a Solution.....	68

3.1.2 The Geometry of Optimal Solutions	70
3.2 Method.....	79
3.2.1 Subjects	80
3.2.2 Apparatus	80
3.2.3 Design	82
3.2.4 Procedure	89
3.3 Results	95
3.4 Discussion.....	112
4. Future Work.....	116
5. Conclusion.....	117
References	119
Appendix A: Hark Grammar.....	131
Appendix B: Subject Instructions.....	132

1. Introduction

"Virtual Reality" has been a media buzzword for a number of years. A rather distorted perception of what "VR" means has been integrated into popular culture, leading to some fantastical expectations for the application of the actual technology. One group of VR researchers plotted the frequency with which the term "virtual reality" appeared in the press from 1989 through 1993 and found that the usage of those words was growing exponentially (Biocca, Kim, & Levy 1995).

In reality, a large body of knowledge about virtual environment (VE) systems remains to be explored before any realistic simulated "world" can be developed. In particular, a program of rigorous experimentation should be based upon the study of human perception and psychophysics. In this manner, new and unusual phenomenon resulting from the development of VE systems can be understood.

One important component of a successful virtual environment is the perception of depth. Depth perception is needed so that a user may develop a sense of "being there" in a virtual world. Unfortunately, the deficiencies of the displays in VE systems present a number of problems for accurate presentation of depth. Combating the limitations of these displays is a formidable job, especially for situations in which completing a particular task requires a presentation with substantial realism.

2. Background

Virtual environments found their origin in teleoperator systems. Teleoperation refers to the remote manipulation of some device, where the operator must be presented with a sufficiently transparent human-machine interface in order to control the remote device. For an in-depth treatment of the history of VE, see the work of Sheridan (1992b), Kalawksy (1993), or Rheingold (1991).

Virtual environments is a very young field of study yet has received a large amount of attention from a variety of disciplines, including computer science, psychology, cognitive science, visual art, and industrial engineering. All of these disciplines seek to improve upon the human-centered paradigm exemplified by VE systems. However, the implementation of VE systems is a nontrivial task because of the technological limitations of the hardware. Before discussing these limitations and how best to approach them, a number of definitions should be provided.

2.1 Definition of Terms

Virtual Environment: We avoid the use of the term "virtual reality" both throughout this paper and throughout the research conducted at the Massachusetts Institute of Technology. Not only is the term "virtual reality" a product of the fertile imaginations of the popular media, it is linguistically self-contradictory and raises expectations of the capabilities of modern-day equipment. Perhaps the best definition of *virtual environment* is given by Stephen Ellis, one of the pioneers of the discipline:

"A virtual environment is that synthetic, interactive, illusory environment perceived when a user wears or inhabits appropriate apparatus providing a coordinated presentation of sensory information mimicking that of a physical environment" (1995a).

We can be more precise by giving meaning to a "virtual environment system." A VE system is multi-modal, interactive and adaptive, reconfigurable in software, and can generate supernormal situations. This system consists of the human, the human-machine interface and the computer-generated synthetic environment.

A virtual environment system consists of a visual display system, head tracker, and auditory displays (VETREC, 1992). Oftentimes, a glove-input device is included in the system, but such devices are no longer very popular due to a number of technical difficulties. Additional components of a VE system may include head-mounted displays (HMDs), automatic speech recognition, haptic input interfaces, force-feedback devices,

whole body movement sensing, and/or sensing of physiological responses, such as eye movement, blood pressure or heart rate (VETREC, 1992). The equipment and its properties will be discussed in further detail below.

In the broad sense, a virtual environment system is any system that attempts to fool people into accepting a computer-generated, synthesized world as real. The most important traits of a convincing VE are interactivity and adaptability. The ability to persuade a user that the VE is the real world leads us to define the concepts of *immersion* and *presence*.

Immersion and Presence: The definition of a virtual environment given above can be extended to describe the subjective sensation a user might feel:

"Sensory information generated only by and within a computer and associated display technology compels a feeling of being present in an environment other than the one a person is actually in" (Sheridan, 1992a).

Not only is a virtual environment supposed to fool the observer into perceiving different surroundings, but it is also supposed to imbue the user with the experience of "being there." This sense of "being there" is termed *immersion* or *presence*.

Without the feeling of immersion or presence, VE research would not have generated much excitement. The ability to provide a sense of immersion is very desirable in entertainment applications as well as in teleoperation or training where the tasks to be performed are wide-ranging, complex and uncertain (Held & Durlach, 1992). The desire for more immersive virtual environments motivates much of the research in the area. Scientists are trying to qualify and quantify what creates the experience of presence, how to measure it, how to improve it, and what effect it has on various VE tasks (Slater & Usoh, 1993).

Simulator: A knowledgeable reader might have noted that a number of the qualities associated with virtual environments have been present for many years in what are called *simulators*. However, virtual environments have some characteristics which distinguish

them from simulators. For example, a VE is more flexible than the typical simulator; a VE is reconfigurable for different levels of fidelity and/or various skill levels as well as for the characteristics of a particular task (VETREC, 1992). In addition, a simulator is generally trying to match the real world (the degree of which is measured as "simulator fidelity") while a VE need not present an exact copy of a real world environment. A simulator is closely tied to some physical situation, while a VE is most closely associated with the human user (Carr, 1995). This is encouraging; VE systems have much wider applicability than ordinary simulators since they focus on providing immersion, rather than merely a replica of the real world. Negroponte prognosticated in 1970 that human factors would eventually take a strong part in computer system and computer interface design. VE systems differ from the simulators that existed at that time because they embrace a human-centric view, encouraging designers to look to the human being to justify design decisions.

Kriloff, in 1976, presented another view, suggesting that the human should conform to the machine since machine programming is fixed and man is adaptable, with the caveat that unpleasant side effects might occur. The study of human factors engineering, which originated in the late 1940s, is concerned with how to design man-machine systems to reduce side effects and make them easy to use, safe, and efficient. The view that a VE adheres more to the human end of man-machine interaction accurately represents the relatively recent shift to more human-centric systems. Simulators are being replaced by VEs in situations where the human and their sensations are of the primary concern. Simulators act to produce a situation to which the human must adapt, rather than adapting the situation to human sensory capabilities. Both simulators and VEs share the common theme of presenting a synthetic task to the user; therefore, the concept of a *task* should be better defined.

Task: Human factors engineers recognized very early in the growth of their discipline that some formalization of human behavior was needed in order to examine the behavior

involved in man-machine interaction. Thus, the formal concept of a *task* was born. A task is an arbitrary unit of work, one or more specific related actions necessary to change or verify a system's state. A task may be mental, physical or a combination of these (Van Cott & Paramore, 1985). A task has a number of characteristics, including a set of conditions that require and initiate human performance, a specific purpose, a definite start and end defined by an initiating and terminating cue or stimulus, a relatively short life span, the ability to be interrupted by another task, and the ability to be performed by many people at once (Christensen, 1993). For the purpose of VE research, these definitions are more than sufficient.

Realism: When we mention the idea of a task, we are implying the existence of a real-world context for that task. In a simulator, we are trying to make the synthetic task as much like the real world task as possible. Furthermore, we could say that "realism" is the degree to which the sensory stimulation coming from the artificial environment matches that originating from an equivalent real environment (Christou & Parker, 1995). A serious issue of perception arises immediately: what is veridical and what is perceived? The common assumption is to believe that "what I see is real," a perspective philosophers call naive representationalism or naive realism (Weintraub, 1993). Research shows that the veridical world is not the same as the perceived world, yet in VE work, the main thrust is to utilize and exploit the inherent desire for naive realism to produce the experience of presence. We exploit this desire by studying the aspects of the human visual system that contribute to the perception of a world as real and using the results to synthesize a real situation. However, in VE systems, supernormal situations are possible, making comparison to the real world difficult, if not impossible. Therefore, realism is only really relevant to the discussion of VEs when the VE is attempting to simulate traits of the real world.

Could a virtual environment simulate a world that has no functional or logical connection to the real world? This is a difficult idea to grasp and seems far-fetched, yet the capability may exist. However, a human would be unlikely to understand such a world and would experience little or no immersion. Thus, a VE must embrace reality in some manner in order to present a sensible and usable environment. Characterizing the ways in which a VE should mimic the real world is an important part of creating an interactive and immersive environment.

Helmholtz, in 1882, proposed the Doctrine of Unconscious Inference, which states that people apply what they know about the world to achieve the percept, and that information from the world is oftentimes insufficient (Weintraub, 1993). His century-old doctrine suggests the basis for studying realism in VEs: the manner in which the world is perceived as real is based upon previously gathered information about the world. Realism, then, could be achieved through the study of the preconceptions that lead to perception of the world as real.

Realism in virtual environments has a connection to the feeling of immersion. Hendrix and Barfield report that, according to a subjective questionnaire, people react more to the realism of the interaction in a VE than to the realism of the objects in that VE (Hendrix & Barfield, 1995). The functional behavior of objects in a VE determines its subjective realism, while the actual appearance of the objects is less relevant. However, what determines functional behavior is not clear. This terminology should include perceptual functionality, such as how objects behave when a subject's head turns, or how objects move within the scene. Clearly, the exactitude of the human-machine interface is not the only determinant of immersiveness and/or realism. The significance of the behavior of the objects within the simulated environment should not be overlooked.

The computer graphics world has, for much of its short life, been pushing technology ever further in the pursuit of photorealistic images. New, higher-resolution displays and more powerful computational engines have been developed to aid the realism

of the images on a computer screen. The computer graphics field is finally acknowledging that the behavior of the objects in the scene is likely to be more important to realism than the quality of the image. A shift in focus is reflected by increased research into physically-based modeling and animation.

The young discipline of virtual environments is only beginning to absorb the large body of knowledge from psychophysics and perception that offers clues to the concept of realism. The perceived appearance and behavior of objects is being carefully quantified relative to some of the peculiar equipment used in VE systems. Therefore, a brief survey of standard VE equipment and its characteristics is warranted.

2.2 Virtual Environment Equipment

The study of Virtual Environments began with Ivan Sutherland's famous discussion of "The Ultimate Display" (Sutherland, 1965). Sutherland describes the ultimate display as one that is indistinguishable, to the user, from the real world. To this end, he constructed a head-mounted display (Sutherland, 1968). After Sutherland's landmark HMD in 1968, a number of other devices were built to display vision, sound, touch, and smell (Kalawsky, 1993). For the purposes of this thesis, only visual displays will be discussed. Readers are encouraged to read Deering (1993), Kalawsky (1993), Durlach and Mavor (1995), or Ellis (1995a) for more information about other displays.

The visual display gives the subject the most salient and detailed information about the synthetic world. A real-time display showing precise, continuous motion imagery, while maximizing normal visual sensory abilities is ideal. A visual display system incorporates the actual display surface, a system for monitoring the location and motion of the head and/or eyes, a system for generating the stimulus, and a positioning system for the displays (VETREC, 1992).

The type of visual display most commonly used in virtual environment work is a head-mounted display. An HMD is a helmet that sits on the user's head and presents an

image very close to the eyes. Usually, the helmet contains a tracking device which permits the graphics hardware to update the image to match the motion of the participant's head, allowing them to "look around" a scene. Often, to allow the presentation of stereoscopic images, an HMD will have a separate display for each eye.

A number of variations on the generic HMD have been developed. See-through displays using half-silvered mirrors have been developed to allow the superposition of virtual images with the real world (Kalawsky, 1993; Barfield, Rosenberg, & Lotens, 1995). Some HMDs use spinning color filters to present a color image from a grayscale display (Allen, 1993). Eye tracking has been introduced into a few HMDs so as to follow the user's gaze with high-resolution inserts (Kalawsky, 1993). New optical systems are always being developed that have wider field of views (FOVs) and better adaptability to the vision of different users. Significant effort is being devoted to designing and prototyping new HMDs by laboratories at the University of North Carolina and the University of Washington, while the commercial world continues to produce a variety of styles of HMDs (Barfield and Furness, 1995).

Despite the focus on the ubiquitous HMD, a number of other types of displays are worth mentioning since they lend insight into a few of the limitations of HMDs. The CrystalEyes field-sequential system uses glasses with liquid crystal shutters that are synched with a monitor to time-multiplex images to the eyes (Lipton, 1991). A high-resolution, narrow FOV, three-dimensional image can be presented with this system. The CAVE at the University of Illinois is a system designed to surround the user in a cube with back-projected images on the walls. These images are also presented in a time-multiplexed manner with shutter glasses so that a user perceives a large FOV, three-dimensional image. The shape of the CAVE presents some limitations; the projected images are lower resolution than those seen on a normal computer monitor, and a great deal of computation is required to present images that update at a sufficiently fast rate (Cruz-Neira, Sandin, & DeFanti, 1993). At the time of this writing, a number of other interesting alternatives to the

HMD are being developed. However, to present them all would be laborious. The reader is invited to consult Kalawsky (1993) and the journal *Presence: Teleoperators and Virtual Environment* for more information.

Head-mounted displays have a number of characteristics which determine their weight, comfort, durability, size, and price. Generally, HMDs weigh from 1 to 10 pounds and permit for a number of different adjustments to ensure a secure fit. Some HMDs allow inter-pupillary distance (IPD) and eye relief (the distance from the surface of the eye to surface of the display) adjustments to suit the individual. FOV ranges from 20° to 140° horizontally and 23° to 90° vertically, and most HMDs are capable of displaying around 50° horizontal by 40° vertical. Displays are usually active liquid-crystal displays (LCD), although electroluminescent, light-emitting diode, plasma, and cathode-ray tube (CRT) based displays have been successfully constructed. The pixel resolution of these displays hovers around 600 pixels by 400 pixels, although examples of both higher and lower resolution systems have been produced. The design of HMDs is a fairly complex and interesting field; the curious reader is referred to the work of Barfield, Hendrix, Bjorneseth, Kaczmarek, and Lotens (1995) for an introduction.

Ultimately, virtual environment designers would like a visual display system that has a high spatial resolution, high luminance and contrast, and high-fidelity color. The system should incorporate a wide FOV with a rapid stimulus generation and update rate. The system should also include a quality stereoscopic configuration, provide minimal interference with the user's motion and overall comfort, and eliminate noise and distortion. Safety is an important issue, as are reliability and durability (VETREC, 1992). Unfortunately, these characteristics are far from being achieved. Ideally, an HMD would weigh as little as a pair of glasses, fill the visual field, match the resolving power of the eye, and properly coordinate all visuo-spatial cues (Biocca & Delaney, 1995). However, the possibilities for VE technology are still substantial even if these constraints are only partially satisfied.

2.3 Advantages of Virtual Environments

Virtual environment technology has generated enthusiasm for several reasons. Obviously, tasks that require immersive visualization and understanding of complex, three-dimensional environments are perfectly suited to implementation in a VE (Travis, Watson, & Atyeo, 1994). VEs can enable multiple, simultaneous, coordinated, real-time foci of control in an environment, allowing simulation of tasks that involve the manipulation of objects in complex situation (Ellis, 1995b). In addition, VE technology provides for applications which require multiple viewpoints of a three-dimensional environment.

Another strength of virtual environment systems is that they are more cost-effective than traditional simulators. Because of a VE's reconfigurability in software, changing a simulator from, for example, an F-14 fighter to a Cessna 150, costs only the price of a programmer, not the price of a whole new cockpit mock-up. In this example, the VE is more cost effective by at least two orders of magnitude (a new simulator might cost millions of dollars, while reprogramming a VE may only cost tens of thousands of dollars). A VE could also be modified for each user based upon a variety of constraints from the user's visual acuity to his or her spatial reasoning abilities. The reconfigurability of VEs allow them to be supremely adaptable, requiring only standard equipment to provide a variety of services (VETREC, 1992).

In addition, virtual environment systems are networkable. Multiple users can be supported in a single environment, permitting cooperation and human-human interaction. Networking together VEs also allows for the geographical distribution of users and permits dispersed resources to be shared among users. Thus, VEs are well-matched to teleoperation tasks. Areas of applications for teleoperation include the exploration of space, undersea oil and science (geology, biology), nuclear power plants, toxic waste cleanup, construction, agriculture, mining, warehousing, mail delivery, firefighting, policing, military operations, telesurgery, and entertainment (Sheridan, 1992b). The ability to network VE systems allows for the possibility of achieving these applications.

Finally, VEs have the ability to present supernormal situations. The user can be given additional information in a synthetic world, gaining perspectives and opportunities for communication not possible in the real world. The potential for augmenting normal perception is staggering; adding information to a person's sense of the real world has hundreds of applications.

2.4 Virtual Environments Applications

In general, human factors engineering attempts to achieve maximum efficiency, productivity, dependability, reliability, safety, habitability, operator acceptance, while minimizing training and man-power costs, and maintaining a high degree of flexibility in the design and use of the man-machine interface (Pew, 1993). Virtual environments represent a significant step towards achieving these goals.

Before the lovefest with VEs goes too far, it must be noted that VEs are well-suited *to a particular set of tasks*. Tasks that require a tracked and transitional viewpoint, 3D spatial reasoning and visualization, and complex interactions with the environment utilize the advantages of VE technology (Stanney, 1995). Suggested applications include teleoperation (Sheridan, 1992b), entertainment, training (Ellis, 1995a; VETREC, 1992), education (Travis, Watson, & Atyeo, 1994), architecture (Slater & Usoh, 1993; Henry & Furness, 1993), scientific and medical visualization (Ellis, 1995b; Kalawsky, 1993), design, manufacturing and marketing, and telecommunication (Durlach & Mavor, 1995).

However, for each potential application, a careful analysis of both the task and the users is needed; failure to recognize the misapplication of VE technology can have disastrous effects. Some have suggested the formation of a virtual task taxonomy to direct design efforts for maximizing human performance in VEs. Classifying tasks according to types of displays and interactions which best improve efficiency in VEs would be extraordinarily helpful in determining for which tasks VE technology might be effective

(Stanney, 1995). However, the development of such a taxonomy is a formidable job, since the size and complexity of the application-space are imposing.

2.5 Problems with Virtual Environments

Why is task analysis so important? Because the high expectations for virtual environment systems have not been met by the advances facilitating new and novel hardware and software. The fundamental ideas underlying VE systems have been validated; however, the technologies needed to produce a sufficiently immersive virtual world for a large range of applications are not available. Again, we focus on the elements of the visual display system and their limitations rather than delving into the idiosyncrasies of other perceptual channels. Because HMDs are the current standard in VE systems, the problems with these devices are of the most interest.

Difficulties with Head-Mounted Displays

Many different display types have been introduced for use in virtual environment systems. Head-mounted displays are the most commonly used, and have become the de facto standard in the field despite their many drawbacks. One of the most redeeming characteristics of an HMD is that they provide a wider field of view (FOV) than typical computer displays. However, human vision has a FOV of about 200° horizontally and about 60° vertically, while most HMDs have a FOV less than 50° in either dimension (Ma, Hollerbach, & Hunter 1993; Barfield et al., 1995).

The vertical and horizontal field of views in an HMD are rarely equal since few displays are made with equal resolution along each axis. FOV is believed to have a strong effect on the feeling of presence (Robinett & Rolland, 1992). The wider an image appears, the more like the real world it seems. Studies have shown that a wider FOV improves performance on some tasks (Ellis, 1995a).

In order to achieve a wide FOV, the pixel resolution of the display is compromised. For example, normal computer monitors have a resolution of 1028 pixels by 1240 pixels and a diagonal measure of about 17 inches. When viewed from a normal sitting distance of 15 inches, the individual pixels are not detectable, and the FOV is about 20° horizontal. Viewed from a distance of 2 inches, the individual pixels are clearly discernible, and the FOV is about 100° horizontal. HMDs are analogous to the latter case, where displays with resolutions roughly 600 pixels by 400 pixels are placed about an inch from the surface of the eye.

Because of the placement of the display, an HMD has low pixel resolution. Text is difficult to present, and depth perception is seriously distorted. Users of typical HMDs qualify as legally blind (i.e. have 20/200 vision or worse) (Ellis, 1995a). In the domain of aviation, displays that had low resolution were shown to increase the root-mean-square deviation from an optimal descent path (Fadden, Browne, & Widemann, 1991). The pixel resolution of the display is important for a variety of applications.

The ability of a human eye to discriminate objects is called visual acuity. Acuity is determined by the physiology of the retina, which has varied density and accuracy of receptors. The eye has a main focus region, called the fovea, which has a resolution of about 30 seconds of arc (Buser & Imbert, 1991; Boff & Lincoln, 1988; Goldstein, 1989). HMDs are far from presenting images that are equivalent to this level of acuity.

As the eye moves, the high-acuity foveal region points to different regions of an image. Therefore, to present an image of sufficiently high resolution, the display must match foveal acuity. Most displays available today have inadequate resolution. In addition, eye movement may cause *vignetting*, which is the partial to total loss of light from the image. Vignetting occurs when the eye pupil is not at the intended exit pupil of the optic system of the HMD (Ma, Hollerbach, & Hunter, 1993). This problem can be partially repaired with an increased exit pupil.

A number of display systems have troublesome interfaces with the computational engines which produce the images. Often, the frame buffer of the graphics computer does not map precisely to the surface of the display; usually some image clipping occurs. Thus, the FOV of the display may differ from the published characteristics of an HMD. Erroneous computation of the perspective geometry can result unless the behavior of the computer-to-display interface is properly represented. In one system, neglecting the clipping of the frame buffer resulted in an error of 5° in the presented FOV, causing objects to look bigger and closer than intended (Rolland, Gibson, & Ariely, 1995). Clearly, the behavior of the interface between the graphics engine and the actual display should be well understood by the designer of a visual display system.

Furthermore, a designer should try to build a head-mounted display that has a large eye relief. A large eye relief would accommodate the 30-50% of the population aged 20-45 that use spectacles, more than half of which wear them when using optical devices (Ma, Hollerbach, and Hunter, 1993). However, increasing the distance between the eye and the display reduces the FOV. Many HMDs do not even allow any sort of eye relief adjustment to accommodate wearers of eyeglasses.

User variance, as illustrated by the issue of users with corrected vision, is one of the critical elements of display system design. A VE display system should be flexible enough to handle the variation between users yet robust enough to present the same image to each person. One of the major factors that differs from user to user is the interpupillary distance. IPD varies from about 53 mm to 73 mm between users, averaging out around 63 mm (The author's IPD measures 61 mm, 29 mm to the left, 32 mm to the right!). This discrepancy places the following constraint on the design: either the optics have to provide a wide enough exit pupil to accommodate both wide and narrow-eyed viewers, or a mechanical adjustment, like on binoculars, should be incorporated into the optical system (Robinett & Rolland, 1992).

User variance includes more than just physiological differences; psychological issues are also important. For example, familiarity with a particular display system has been shown to significantly affect some tasks performed in a VE (Stanney, 1995). For example, novice users of HMD systems often fail to take advantage of the multiple viewpoints presented or, even more critically, of the adjustments possible on the HMD itself.

Another problem with head-mounted displays stems from limitations of the display technology. Generally, the displays in HMDs lack luminance and contrast. In the real world, intensity in a particular scene might range from 1 to 1000 (i.e., a sunbeam in a dark room), while a typical CRT has an intensity range from only 1 to 100, and an LCD has even less. Contrast also aids visual acuity; as a display gets dimmer, acuity performance decreases (Christou & Parker, 1995). The relationship between contrast and visual acuity is given by the Contrast Sensitivity Function (CSF). The CSF relates the decrease in contrast to a decrease in visual acuity (Weintraub, 1993; Goldstein, 1989). Because HMDs are unable to produce realistic contrast and luminance values, visual acuity suffers.

The color capability of the display technology used in most head-mounted displays is also inadequate. The color range that can be presented simply does not match the colors that are perceivable and discriminable by the human eye. And, for each color, the brightness control produces less brightness levels than can be normally differentiated (Christou & Parker, 1995; Barfield et al., 1995). So, the use of color displays adds an additional layer of complexity to the limits of HMDs.

The optics that rest between the user and the display surface also contribute to the difficulty of designing an HMDs. Since most HMD optics are magnifiers to widen the FOV, a convex spherical distortion is introduced. Most code for presenting images on these displays fails to take into account this distortion. The optics, besides increasing the FOV, provide the user with an image that they can focus on, despite the fact that the display surface may be very close to the eye (Robinett & Rolland, 1992; Hodges & Davis, 1993).

The optics end up curving normally linear surfaces, and introduce a number of other aberrations. The main aberrations in HMD optics can be described as (in layman's terms) blurring, curvature, distortion, and color (Ma, Hollerbach, & Hunter, 1995). Humans have some ability to adapt to aberrations, and much work in optometry has been devoted to quantifying human tolerances to these distortions.

The optical system in an HMD presents another problem since its idiosyncracies are usually not modeled in the code used to display the image. Generally, graphics systems model the eye as a single point which is the center of the perspective projection. The pupil is not well-represented as a single point, nor do most displays account for the movement of the eye (Rolland, Gibson, & Ariely, 1995). Furthermore, the displacement of the "virtual eye" in some models results in significant spatial errors as well as a decrease in the speed of task completion (Rolland, Biocca, Barlow, & Kancheria, 1995). Furthermore, most HMD optical systems have large exit pupils which are also not accurately represented as a point. Clearly, the model of the geometry in the computation of the image should match the characteristics of both the optics in the display and the optics of the human eye.

Furthermore, human perceptual distortions can further complicate the precise modeling of the visual display. A distortion-free display and a precise formulation of the geometry will still result in some inaccurate perceptions. This is due, in part, to the psychology of self-location which states that accurate visual perception of an object requires a combination of perceived distance, perceived direction, and perceived location of the viewpoint (Pstoka, Lewis, & King, 1996). Because the visual system is an information-loss system, a number of filters from the physiological to psychological level act to extract relevant information from the stream of data being received from the real world (Weintraub, 1993). The physiological characteristics of these filters have been discussed, but human psychological biases are not well-modeled in most VE visual display systems.

Stereoscopic Image Presentation

Further complications arise if a stereo image pair is displayed. The role of stereoscopic vision in depth perception is often not important enough to warrant the use of stereovision in HMDs. In addition, a significant percentage of the population is stereoblind; a survey of 150 students at M.I.T. showed that 4% could not use stereopsis and that 10% had great difficulty in perceiving depth in a random dot stereogram. Somewhere between 1 in 10 and 1 in 100 people are stereoblind (Yeh, 1993). Other details and difficulties in presenting stereoscopic images are treated by Edgar and Bex (1995), Hodges and Davis (1993), Lipton (1991), Robinett and Rolland (1992), and Wann, Rushton, and Mon-Williams.

Simulator Sickness

The inadequate presentation of images to the user of a virtual environment has several implications. The phenomenon known as "simulator sickness" is one of the most distressing results of using insufficient realism in the visual display system. Simulator sickness refers to the malady experienced by some users during prolonged immersion in a sub-optimal virtual environment. Lag in head-tracking, misaccommodation, and HMD weight contribute to the feeling of nausea that may be experienced in a VE system. Simulator sickness has become such an area of concern that a system for rating the magnitude of simulator sickness has been proposed (Kennedy, Lane, Lilienthal, Berbaum, & Hettinger, 1992). In the domain of aviation, illness and nausea has been well-documented for years. Aircraft simulators have been built that are capable of reproducing some of the same symptoms as real airplanes (Leibowitz, 1988). Simulator sickness and traditional motion sickness differ in that simulator sickness has more oculomotor-related symptoms and less actual vomiting (Kennedy et al., 1992). The interested reader is referred to the special issue of *Presence: Teleoperators and Virtual Environments*, volume 1, issue 3, for more details.

2.6 Solving Problems with Virtual Environments

Because of the potential payoffs of highly immersive applications, many researchers are working hard to overcome the limitations of virtual environments. A number of solutions will be examined to provide insight into solving problems with VEs, and a rough philosophy will be presented. Finally, a previously unsolved problem is presented as a significant example of the complexity and intricacy of VE system design.

2.6.1 Example Solutions

Because related work in other fields has already been done, a number of problems with VEs have been successfully addressed by researchers. For example, motion sickness issues have been dealt with by groups working in aviation and teleoperation (Sheridan, 1992b). Of course, simulator sickness is a much larger issue in simulations where head motion is tracked. Since most VEs consider head-tracking to be integral to both interactivity and the experience of immersion, resolving the problem of simulator sickness is important.

Field of view of the display is another important design parameter for developing immersive simulations. Spatial resolution is generally compromised to provide a wider FOV; a narrow FOV with high resolution gives an unrealistic sense of tunnel-vision. Conversely, a low-resolution, wide-FOV gives a more primitive, yet more realistic image. Due to this trade-off, HMDs are simply inappropriate for certain tasks. One hardware solution follows the eye with a high-resolution patch of about 30° (VETREC, 1992; Travis, Watson, & Atyeo, 1994; Ellis, 1995a; Yoshidea, Rolland, & Reif, 1995).

The exponential growth of technology should not be ruled out as a solution to the problems with VE systems. Active matrix LCD displays have already surpassed good quality CRTs and are far ahead in size, weight, power consumption and operation voltage (Ma, Hollerbach, & Hunter, 1993). A recent development in LCD technology allows the placement of a 640 pixel by 480 pixel display on a single chip with pixel size measuring

only 30 microns by 30 microns. Not only is this chip small, but it also has low power consumption and a low production cost (MicroDisplay, 1996).

Another example of potential of technology is the CAE Fiber Optic HMD (FOHMD), considered to be one of the best visual displays currently available. The CAE FOHMD uses two 83.5° monocular FOVs with an adjustable binocular overlap up to 38°. It provides a horizontal FOV of 162°. The visual resolution is 5 minutes of visual arc, with a high-resolution insert (24° x 18°) with 1.5 arcminute resolution. In addition, the displays are bright, at 30 foot-Lamberts. The head tracker's performance is boosted by additional accelerometers to do predictive tracking, yielding an update rate of about 100 Hz (Ellis, 1995b; Kalawsky, 1993). Of course, the FOHMD is a fairly heavy piece of equipment and is prohibitively expensive.

Another interesting display is the Sparcchair, developed at Sun Microsystems. The Sparcchair trades off high resolution for a low FOV; it has a resolution of 1120 pixels by 900 pixels with a 20° by 25° FOV. The Sparcchair was developed for a specific task requiring high resolution, and thus its configuration seems reasonable (Reichlen, 1993). Yet, even with the arrival of new technologies and designs, some tradeoffs simply cannot be avoided.

The design of the optic system in an HMD also suffers from several unavoidable tradeoffs. The problem with the optics in HMDs has been given a fairly comprehensive treatment by Robinett and Rolland (1992). They attempt to quantify the problems associated with optical distortion and IPD variation by computing an extensive model of the image based upon the layout of the optics in an HMD. A comprehensive simulation should provide a consistent image by accounting for the properties of the HMD's geometry, including the relative positions of the display screens, optics, and eyes (Robinett & Rolland, 1992).

Once a computational model of the HMD geometry has been included in the code, IPD variation can be accounted for by using it as a parameter in the calculation and

presentation of the graphics. Measuring a user's IPD is a fairly trivial task and having adjustments on the HMD for IPD has become commonplace (Robinett & Holloway, 1995; Ma, Hollerbach, & Hunter, 1993). Further calculations have revealed ways to account for all the various transforms in the optics system (including some tracker transforms), as well as off-center perspective projection (Robinett & Holloway, 1995). Hodges and Davis have also contributed a description of the perspective geometry of a display system (1993). Their work, which describes the effects of pixels on stereo depth perception, has resulted in other solutions to display difficulties. Through extensive modeling and calculation, solutions to the optical distortions in HMDs can be resolved.

Watson and Hodges (1995), using Robinett and Rolland's model of the optics' geometry (1992), implemented pre-distortions in software to correct for optical distortion. Their work is particularly interesting because it represents a software solution to a hardware limitation – a methodology discussed in more detail below.

Inter-pupillary distance should not be the only parameter used to characterize user variance. A number of additional tests should be performed to assess other individual differences. Lampton, Knerr, Goldberg, Bliss, Moshell, and Blatt (1994) suggest a battery of tests to determine a subject's visual acuity, color and object recognition, size estimation, distance estimation, search and a number of other visual skills used in locomotion, object manipulation and target tracking. Such a battery seems more appropriate for rigorous experimentation in VE systems than for off-the-shelf VE systems. A good system should be able to accommodate population variance without seriously compromising performance. Thus, the job of VE designers is a difficult one; they must devise solutions that work around the limitations of the equipment and yet are capable of presenting a realistic environment.

2.6.2 Return to the Discussion of Realism

Given a sense of the limitations of the equipment, returning to the concept of realism is necessary. The goal of achieving realism is obstructed by the hardware and software limitations of the VE system. Sheridan (1991) discusses several factors that contribute to the sense of presence and realism in a VE. He claims that the extent of the sensory information, the spatial control of environmental sensors, and the ability to modify the environment all contribute to the experience of presence. Most sources agree that the sensations of presence and immersion are functions of wide FOV, high resolution, effective head-tracking, spatialized sound, and sufficiently rapid frame rate (Hendrix & Barfield, 1995). However, despite this intuition, no clear and logical method has emerged to link the physical characteristics of the VE system with the subjective sense of presence.

Thus, the level of realism is reduced by the low quality of the virtual world. Photorealism suffers from the low resolution of the display and the computational limitations of the graphics engine. Functional and logical realism suffer for the same reasons, as well as the others mentioned above. Clearly, the application of VE systems to simulating a real world task is warranted only if a suitable level of realism can be obtained.

2.6.3 Task Dependence

Obviously, certain issues will be more important in one application than in another. For instance, an assembly task performed in a VE might require a high fidelity haptic interface and a mediocre level of visual spatial resolution, whereas a car driving simulator may demand a higher level of visual resolution and only a relatively simple haptic interface. This kind of reasoning seems pedantic, yet a careful analysis of task requirements is necessary to determine which problems are most significant for a given VE application. As put forth in (VETREC, 1992):

"In designing a visual display system for virtual environments, it is important to remember the specific task to be undertaken and, in particular, the visual requirements inherent in these tasks. None of the available technologies is capable

of providing the operator with imagery that is in all important respects indistinguishable from a direct view of a complex, real-world scene. In other words, significant compromises must be made."

Miller (1976) takes the idea of task dependence much further. He states that simply cataloging the characteristics of the human and the computer is not the best approach to interface system design. Rather, the proper method is to examine what benefits the user. However, he argues that human psychophysics provides too artificial of a base for interface engineering since the tasks presented in psychophysical experimentation are often too divergent from actual human tasks. Thus, Miller argues for studying tasks in context, rather than in reductionist human perceptual experiments, and for matching attributes of the computer system to the human task.

Virtual environment systems should provide the sensory cues that are necessary for a particular task. A fully real physical world is too complex to simulate, so providing task-specific information in the best way possible is the only feasible solution (Zeltzer, 1991). Thus, some applications might benefit from a VE-type display, but the demands of many other tasks may be best met by more traditional (and cheaper) display types (Ellis, 1995b; Stanney, 1995).

For example, Smets and Overbeeke (1995) argue that spatial resolution is not important for some tasks, implying that low resolution HMDs may be tolerable in some situations. How much resolution is necessary is obviously a function of the type of task (Travis, Watson, & Atyeo, 1994). In summary, one might ask:

"For a defined task domain, how should one compromise between spatial resolution and field of view? Is color worth the added expense and loss of spatial resolution? Is a stereoscopic display worth the trouble? If so what are the appropriate parameters... for various types of tasks?" (VETREC, 1992).

Sheridan (1991) proposes two major properties that affect task performance: the difficulty of the task and the number of degrees of freedom in the task. These factors are fairly general, but help to clarify the kinds of tasks that might be performed effectively in a VE.

2.6.4 Task Analysis

A task analysis permits the designer to better understand both the task and its potential for implementation in a virtual environment. Formal task analysis is a large field. Theories have been proposed for analyzing tasks and the implications for training (Gopher, Weil, & Siegel, 1986; Frederiksen & White, 1989; Christensen, 1993).

Basically, a task analysis is the breakup of a task into behavioral components that can be further analyzed (VETREC, 1992). However, visual tasks are fairly complex. Researchers know the type of visual stimulation a user finds informative for particular tasks, but they have trouble linking the type of stimulus with the task type. For example, stereovision and motion parallax provide useful information about the relative distances of objects from the observer (Christou & Parker, 1995), but this result is hard to translate to a particular type of task.

Since we can derive the information to which the visual system is sensitive, the in-context (ecological) significance of this information, and the limitations on the use of the information in the visual system, we can design a VE to stimulate the visual system in a realistic manner. However, the limits of the human visual system be accommodated first, before other contributions to realism can be analyzed (Christou & Parker, 1995).

Realism in a VE can be improved by recognition of the redundancy in the human visual system. Tasks that provide a great deal of redundancy (i.e. multiple cues to the same piece of information) are well-suited to VE systems. Repeated information in the visual system reduces ambiguities, and improves the signal-to-noise ratio (England, 1995).

Further analysis reveals that spatial visualization, orientation, spatial memory, and spatial scanning skills are helpful in predicting the performance of a human-machine interface (Stanney, 1995). A task can be analyzed in terms of these component skills to determine its suitability for a given interface.

2.6.5 A Significant Example: The Visibility Problem

To further understand the constraints on realism imposed by the equipment used in virtual environment systems we look at a concrete example. A significant problem associated with the lack of spatial resolution in typical HMDs is the difficulty with detecting objects that are far away. The low resolution of the display causes the size of an object to change discretely rather than continuously as it moves from one range of depth to another. That is, an object displayed over some depth range will not appear to change in size until a pixel boundary is reached. Then, it will change its size by one pixel and remain that size until another depth range boundary is reached.

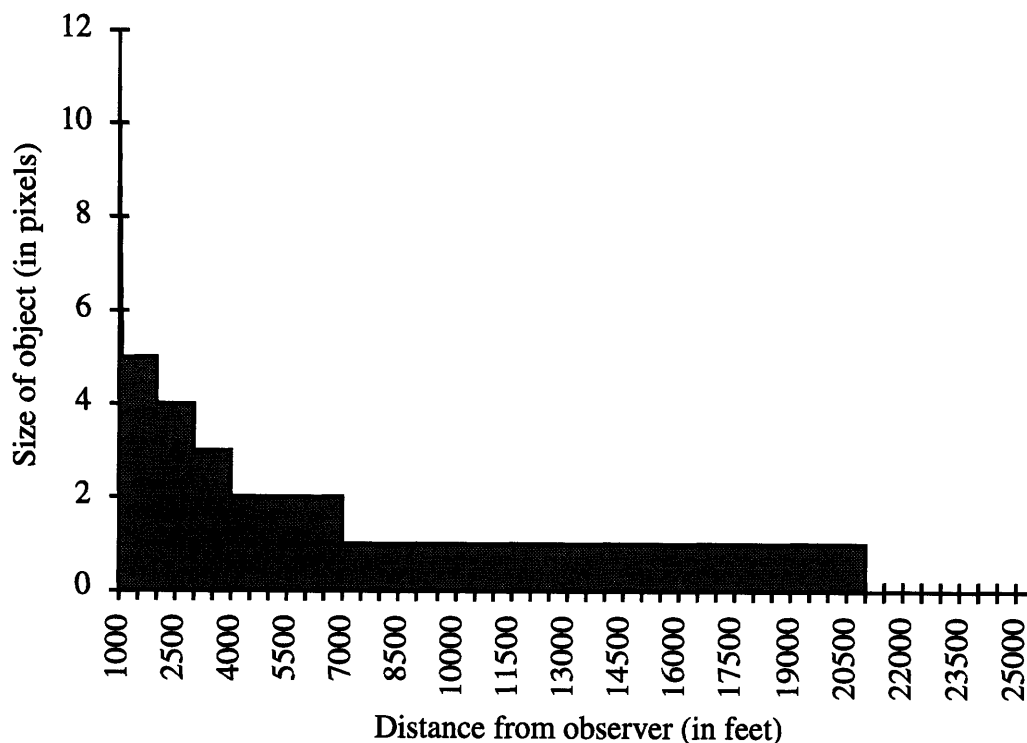


Figure 2.1: Assuming a square, 15 foot by 15 foot object and a display that is 600 pixels by 400 pixels, this plot shows the concept of pixellation of depth. The sample object remains the same size despite being at significantly different depths. For example, the object remains 2 pixels by 2 pixels from about 4000 feet to 7000 feet, a range which is much different from discriminability in the real world.

Pixellation of depth causes two major problems. One, the ability to judge depth is severely impaired; two, the range in which objects are visible is greatly reduced. Depth estimation is impaired and can be further exacerbated by improperly applied anti-aliasing techniques (Christou & Parker, 1995). Actual human depth judgment has an acuity of about 5 minutes of visual arc near the fovea, although lower values have been reported for special cases (Yeh, 1993).

Figure 2.1 shows the threshold problem caused by low resolution displays. As distance is increased, the jump from one pixel to no pixels occurs well before the human visual system would reach its threshold of detectability. The display assumed here is unable to match human abilities. This inadequacy lies at the heart of the problem with visibility in HMDs, and has received some acknowledgment in the literature (Christou & Parker, 1995; Pioch, 1995), but no reasonable solutions have been presented.

2.6.5.1 Background Geometry

The problems with depth perception at the threshold and over the visible range can be quantified by examining perspective geometry. *Visual acuity* can be defined as the smallest size at which two objects can be distinguished. Acuity is assessed in many ways, from discriminating frequency gratings to target detection tasks (Boff & Lincoln, 1988; Buser & Imbert, 1992; Graham, 1951).

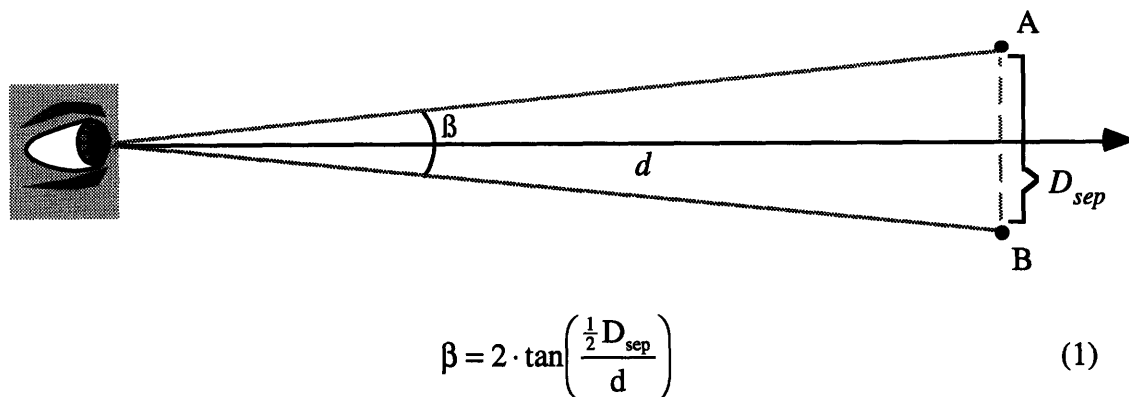


Figure 2.2: Basic model of visual acuity. The visual angle, β , increases as the tangent of the ratio of the separation of the two point-objects, A and B, to the distance from the cornea of the eye to the perpendicular bisector that intersects points A and B.

Formula (1) gives a fairly accurate representation of the angle subtended by the separation of two objects as a function of depth. Now, we can define visual acuity as the maximum value of β for which A and B can no longer be discriminated.

Depth acuity refers to the ability of a subject to discriminate between two objects positioned at different depths (Goldstein, 1989; Graham, 1951). Depth acuity is a particularly complex issue, since a depth percept is constructed from a number of cues. Depth cues can be classified into stereopsis cues and pictorial depth cues. Stereopsis refers to the production of a three-dimensional scene from the images acquired by each eye. Stereopsis cues also include accommodation and convergence, which help determine depth by noting the state of rotation of the eyes (convergence) and the focus of the lens (accommodation).

The main pictorial depth cues generally include:

- occlusion
- linear perspective
- size and familiar size
- distance to horizon
- color
- shading
- atmospheric effects
- texture gradient
- focus
- shadow
- motion parallax (Goldstein, 1991; Buser & Imbert, 1992; Graham, 1951; Boff & Lincoln, 1988)

Interaction among pictorial depth cues is very difficult to quantify. However, the influence of occlusion, linear perspective, and size constancy cues is known to be stronger, under most conditions, than most of the other cues. Linear perspective and size constancy are the cues used most frequently in VEs. This is due to the inability of most HMDs to produce a decent quality stereo image.

The other pictorial depth cues are generally more situation-dependent than the linear perspective and size constancy cues. For example, occlusion is useless unless two objects are placed so that one is at least partially in front of another. The color range available on most HMDs is not sufficient to produce a significant color depth effect. Plus, the "looking through binoculars" feeling of an HMD is not likely to produce an accurate familiar size cue. Most importantly, the deficiencies in color and resolution make blurring and defocusing cues nearly worthless, preventing the use of anti-aliasing techniques.

Because of the limitations of the visual displays some depth cues are simply unavailable, and the remaining cues generally lack the precision of the real world. Since size constancy and linear perspective are the main depth cues used in VE displays, the examination of these cues will provide insight into the depth perception problems that result from poor pixel resolution.

First, the size constancy cue is based on the observation that familiar objects become smaller as they move farther away. Prior knowledge of the size of the object is an important component of the size constancy cue. Size constancy was first noted in the literature in 1889 when evidence was given to match the virtual retina theory (as presented in Equation [1]) (Maurtius).

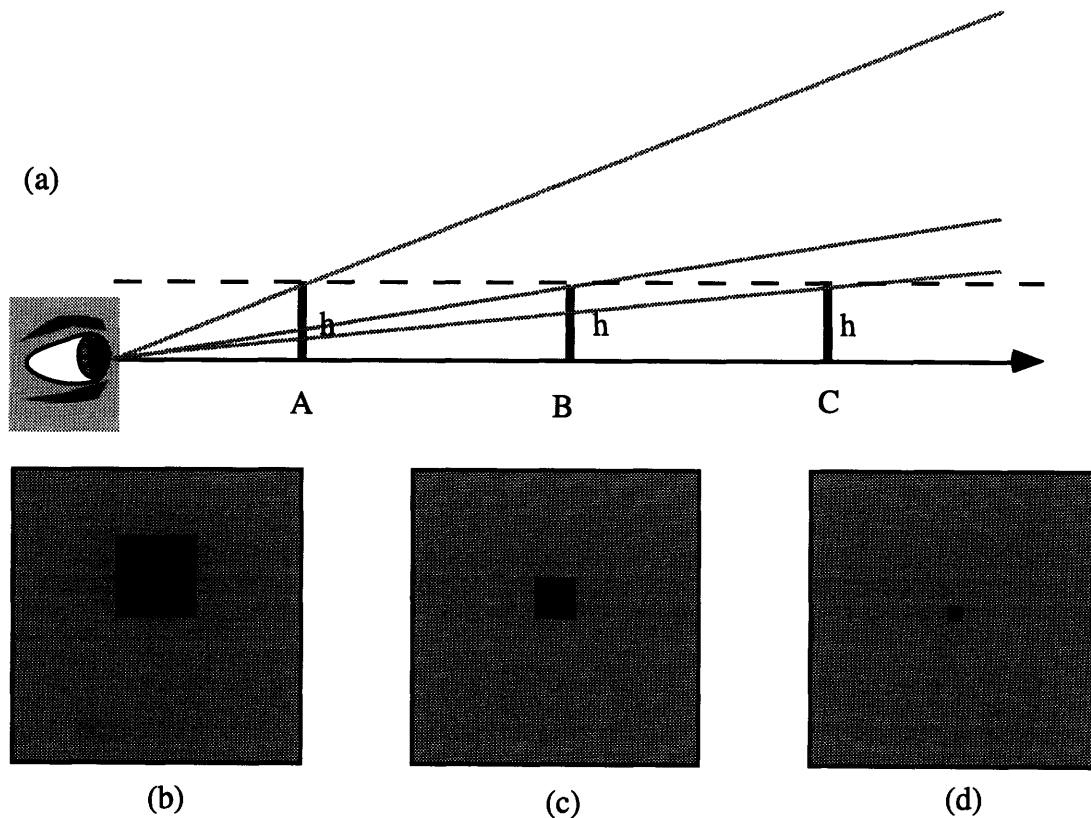


Figure 2.3: Size constancy. An object appears to shrink as the distance between it and the observer increases. (a) The size of the object is given as h at distances A, B, and C. (b) The object as seen at distance A. (c) The object as seen at distance B. (d) The object as seen at distance C.

Linear perspective cues generally require that the observer be some distance above the plane being viewed. Humans have their eyes conveniently located some distance above the ground which helps to provide this type of cue.

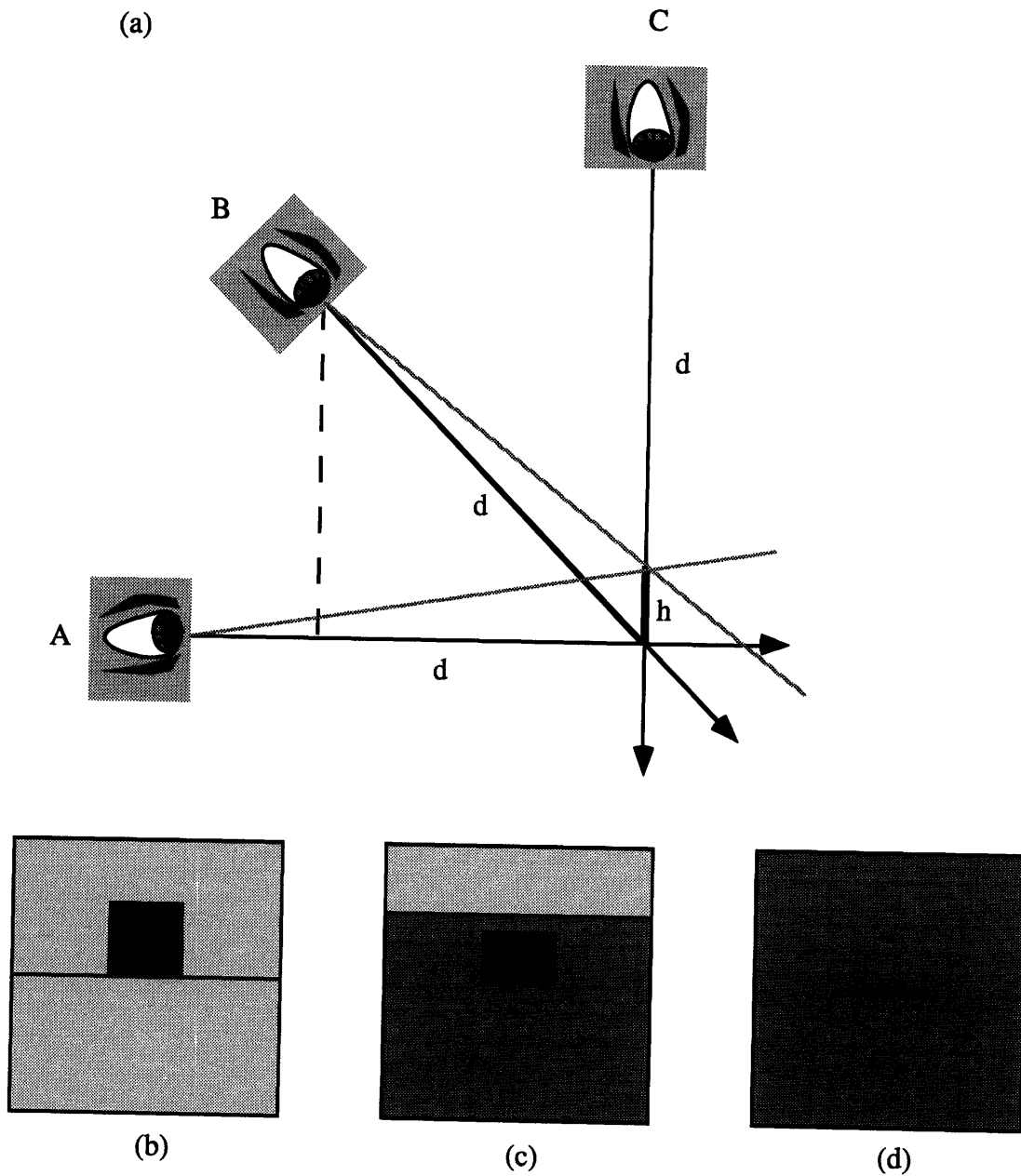


Figure 2.4: The effect of viewpoint height. For a constant viewing distance, d , and a consistent fixation point, increasing the observation height decreases the visual angle subtended by the object and moves the horizon line. (a) The size of the object is given as h and the object is viewed from locations A, B, and C. (b) The object as seen from location A; viewpoint height is zero. (c) The object as seen from location B; viewpoint height is $d/\sqrt{2}$. (d) The object as seen from location C; the viewpoint height is d .

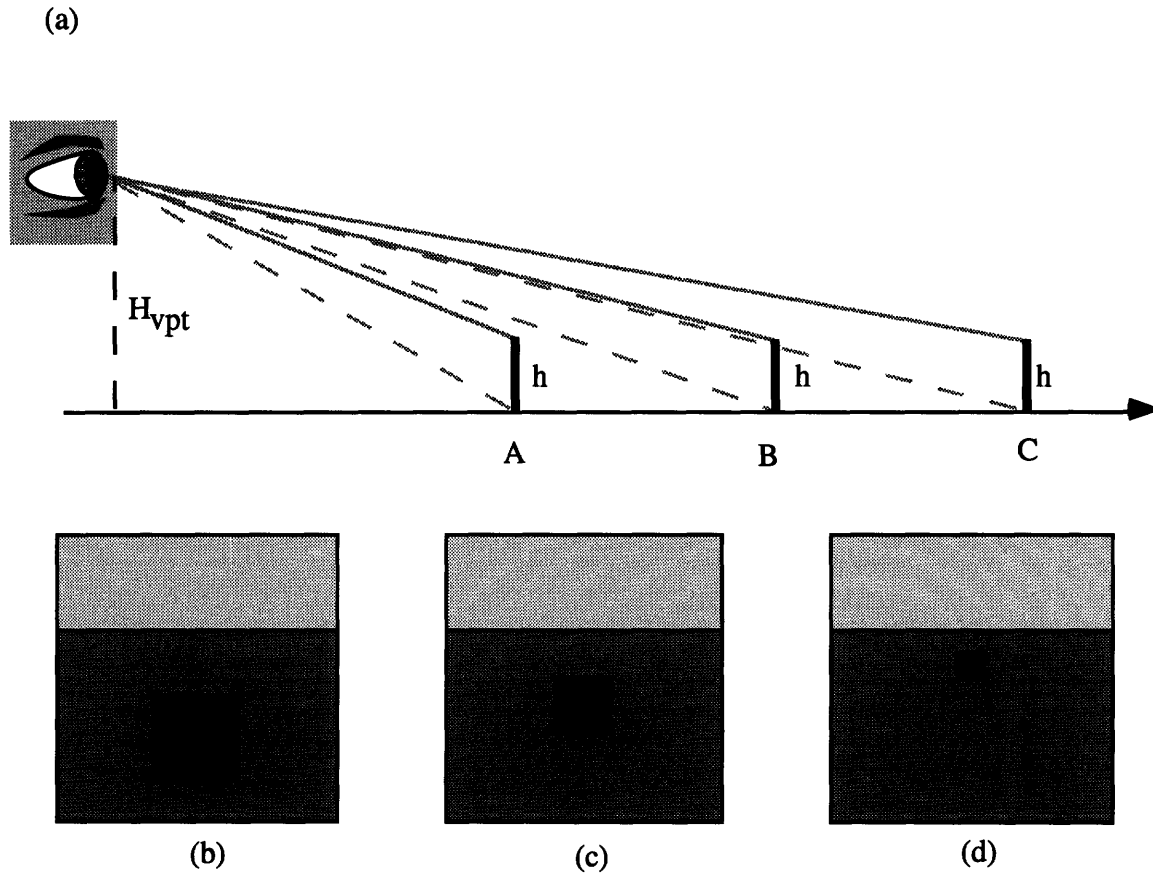


Figure 2.5: With the viewpoint located at a height H_{vpt} and fixed at a single point, the size of the object shrinks and it appears to move towards the horizon as the separation between the object and observer increases. (a) The size of the object is given as h and is viewed at distances A, B, C. (b) The object as seen at distance A. (c) The object as seen at distance B. (d) The object as seen from distance C

Given a particular viewpoint height, the linear perspective cue can be described as the motion of an object towards a center "infinity point" as it moves away from the observer. The following figure illustrates this idea:

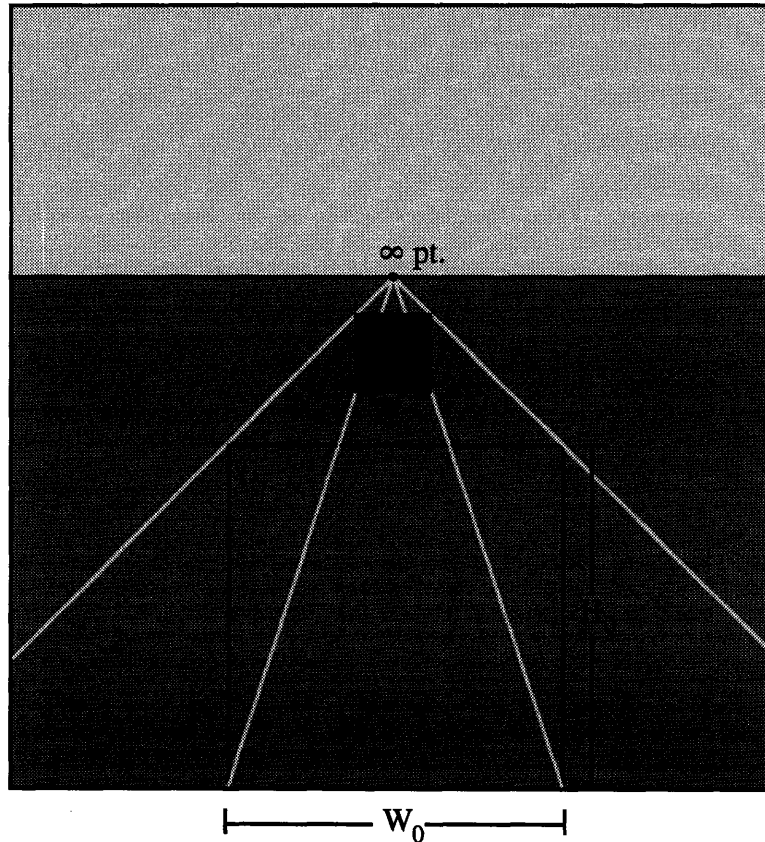


Figure 2.6: Result of tracing the corner points of a square with width W_0 and height H_0 as the separation between the observer and the object increases from zero to infinity.

Conveniently, both of these depth cues can be described by simple mathematics. A prediction of subject performance in a depth perception task can be based both on the perspective geometry and on the results of previous work in human visual performance. The development of a predictive model of visual depth perception in VEs will facilitate the quantification of threshold and depth estimation problems described above.

The first component of this model is a formula describing the visual angle subtended by an object as a function of viewing distance. For the following calculations, a simple model with no viewpoint height is assumed:

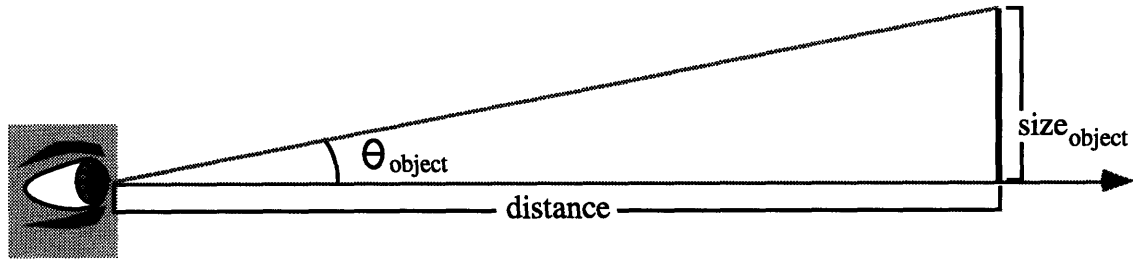


Figure 2.7: A simple model for the size constancy calculation. The angle subtended by the object, θ_{object} , decreases as distance increases according to the tangent function given in Equation (2).

Substituting the parameters of this model into Equation (1):

$$\theta_{\text{object}} = \tan^{-1} \left(\frac{\text{size}_{\text{object}}}{\text{distance}} \right) \quad (2)$$

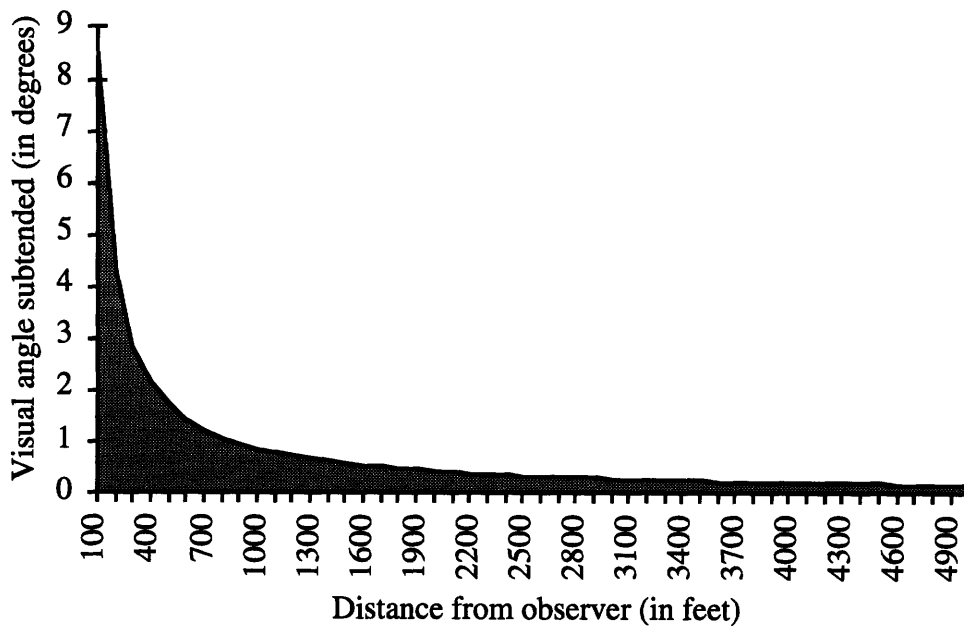


Figure 2.8: A plot of Equation (2). The size of the object is assumed to be 15 feet by 15 feet.

Given a value of 1 minute of visual angle for human spatial acuity, the greatest distance at which a 15 foot by 15 foot object can be detected is:

$$\text{far visibility distance} = \frac{\text{visual angle}_0}{\tan(\theta_{\text{visual acuity}})} \quad (3)$$

for visual acuity = 0.01667 degrees,

far visibility distance (of 15 foot object) = 51566 feet

However, visual acuity is not independent of the viewing distance (Boff & Lincoln, 1988; Geise, 1946) since environmental noise may further add to or detract from it. The actual visual acuity at such a great distance is difficult to determine. Nagata plotted the degradation of several cues as a function of distance, and found that the size constancy starts to become useless at about 1000 m (Nagata, 1991). An engineering approach to determining an actual visibility point will be discussed below.

Visibility in computer displays has been an issue since the late 1940s. Fitts (1951) describes a number of tests regarding visibility of CRT displays, and notes that object size, brightness, and contrast are the main contributing factors to visibility in a normal display. An HMD has certain characteristics which determine visibility, namely: field of view, pixel resolution, and display size. Contrast and brightness are also important in HMDs, but since the spatial resolution is so poor, visibility is not likely to be affected as significantly by those factors.

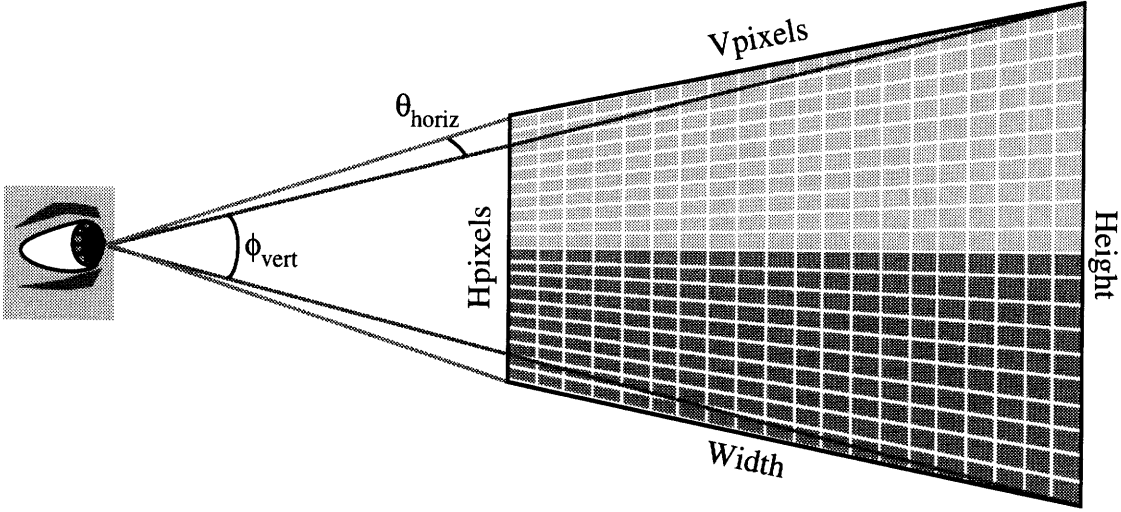


Figure 2.9: Parameters of a head-mounted display.

For some floating-point number $x.y$, we define:

$$\text{round}(x.y) = \begin{cases} x & \text{for } y < 5 \\ x + 1 & \text{for } y \geq 5 \end{cases}$$

Given the characteristics of an HMD presented in Figure 2.9, a formula for the actual number of pixels and displayed size of an object can be stated.

$$\text{pixels}_{\text{object}} = \text{round} \left(\text{pixels}_{\text{display}} \cdot \frac{\theta_{\text{object}}(\text{distance})}{\text{FOV}} \right) \quad (4)$$

$$\text{size}_{\text{object}} = \text{size}_{\text{display}} \cdot \frac{\text{pixels}_{\text{object}}}{\text{pixels}_{\text{display}}} \quad (5)$$

The problems caused by low display resolution are best illustrated with a particular example. The following list of constraints is typical of HMDs:

Vertical resolution = 400 pixels

Horizontal resolution = 600 pixels

Diagonal FOV = 60°

Vertical FOV = 48°

Horizontal FOV = 36°

These constraints are based roughly upon the current state-of-the-art (as described in the section entitled Virtual Environment Equipment). Given these values, we can compute the visual angle subtended by one pixel:

$$\theta_{\text{one pixel}} = \frac{\text{field of view}}{\text{Total number of pixels}} \quad (6)$$

$$\theta_{\text{one pixel horizontal}} = 0.08^\circ/\text{pixel}$$

$$\theta_{\text{one pixel vertical}} = 0.09^\circ/\text{pixel}$$

$$\theta_{\text{human visual acuity}} \leq 0.01667^\circ/\text{pixel}$$

Clearly, the visual angle subtended by one pixel in an average HMD is greater than the values for human visual acuity found in the literature. According to the visual angles given above, a 15' x 15' object in the real world would be barely visible at 51,566 feet, whereas in the display, the same object would be just visible at 10,743 feet.

The HMD characteristics needed to match a human visual acuity of 1 min of arc can be easily calculated. For an HMD with the typical FOV of 48° horizontal by 36° vertical, the display would have to have a resolution of 2,160 pixels by 2,880 pixels to match foveal acuity. For an HMD with a typical resolution of 400 pixels by 600 pixels, the display would have to have a FOV of 10.8° by 6.7°.

In addition to the desired resolution and the number of pixels per object, the *near complete visibility distance* can be calculated. The near complete visibility distance is defined as the point at which the object is first fully contained in the display (i.e. is not cutoff or bigger than the display). In this simple case:

$$\text{Near completely visible point} \Rightarrow \theta_{\text{object}}(\text{distance}) = \text{FOV} \quad (7)$$

Having calculated the limits on visibility imposed by a display, we can now examine the behavior of the object as it appears at different depths. A depth range is defined as the set of continuous distances over which an object stays the same size (i.e. number of pixels). Depth ranges are caused by the failure of the object to change by more than one pixel as it moves in depth.

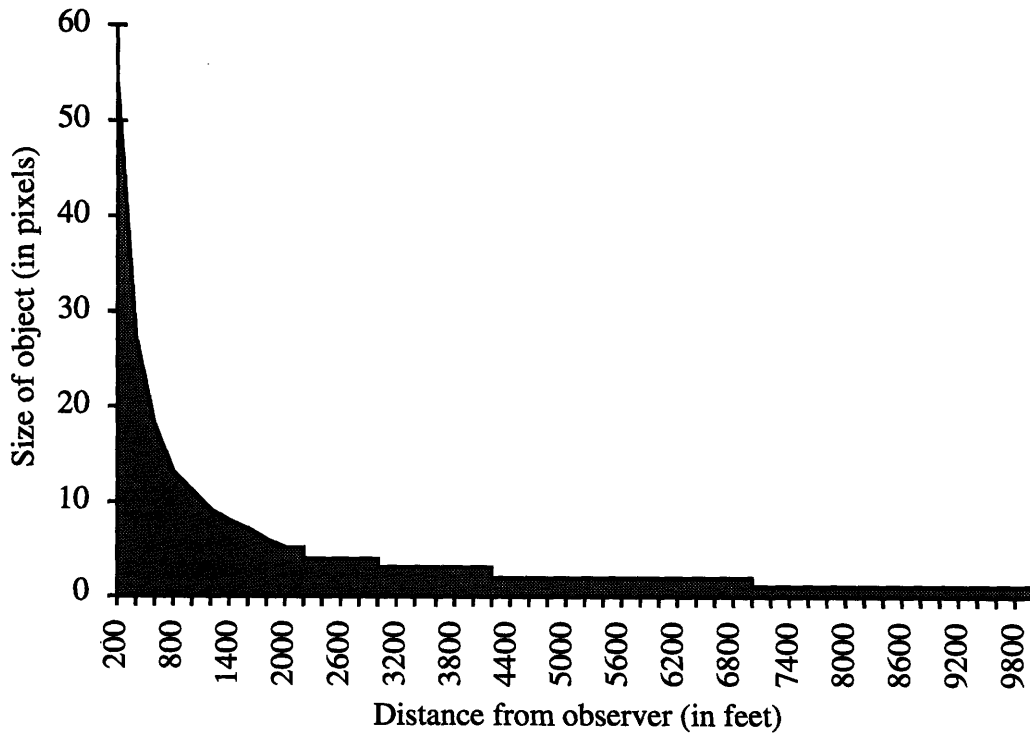


Figure 2.10: A plot of the discrete size steps caused by low pixel resolution. Assumed size of the target object is 15 feet by 15 feet, while FOV is taken to be 48° and pixel resolution to be 600 pixels by 400 pixels.

Not only does the pixellation of depth reduce depth resolution, but it also reduces the total range over which an object can be seen. Since the smallest visible unit is one pixel, and the visual angle subtended by one pixel is greater than the size that can be discriminated by the human eye, an object will disappear prematurely as it moves into the distance and reaches a size less than one pixel.

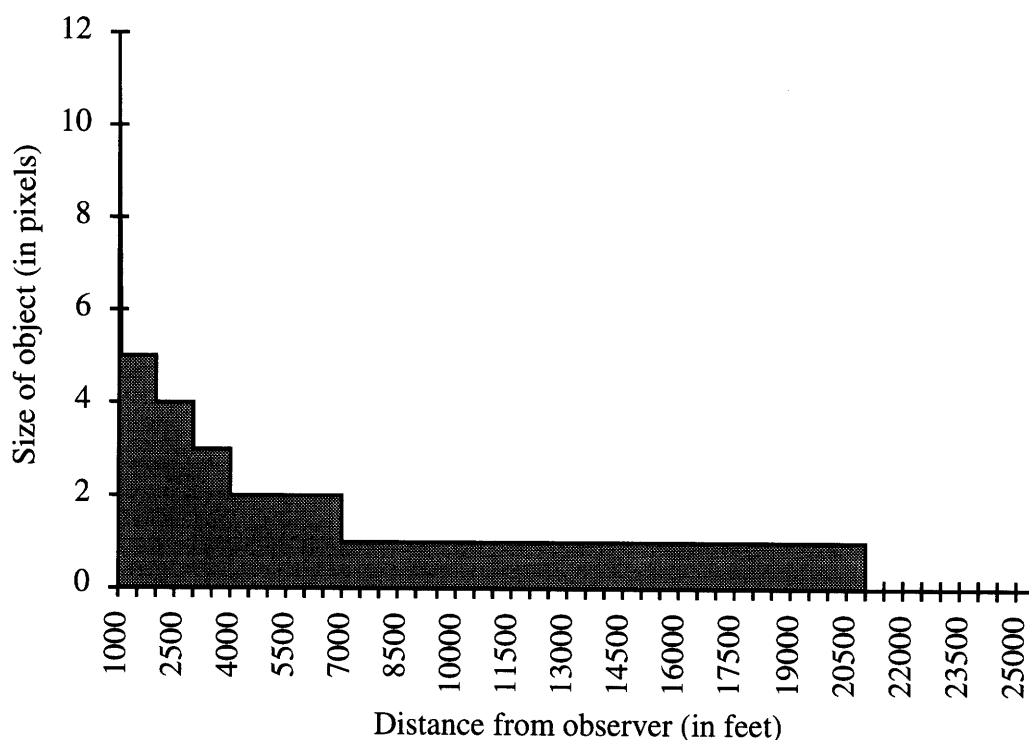


Figure 2.11: The discretization of object size as a function of pixel resolution and distance. The target object is assumed to be 15 feet by 15 feet, and the FOV is taken to be 48° and the resolution is assumed to be 600 pixels by 600 pixels.

Figure 2.11 dramatically illustrates the effects of pixellation on the appearance of an object at various depths. In this model, a viewer would be unable to discriminate between an object at 7,500 feet and an object at 21,000 feet. However, in some ways, the detection threshold issue is more of a concern than the distance discrimination issue. Because human

depth acuity at a great distance is considerably poorer than depth acuity at a close distance (Boff & Lincoln, 1988; Geise, 1946), the effect on distance discrimination is less important. From the calculations above, the predicted distance at which a human could spot a 15 foot tall object is about 51,000 feet, more than twice the distance at which the one-pixel cutoff occurs in this simple model. While in reality, the actual distance may be smaller, it is still significantly greater than can be seen with current displays.

The pixellation of depth cues also has a significant effect on linear perspective. One would expect an object to exhibit the same stepping problem when it moves towards the horizon as when it changes size. However, the model must include a non-zero viewpoint height to observe this effect. Since the appearance of the object as a function of distance is more simple when the viewpoint height is greater than the object height, the model will assume:

$$\text{height}_{\text{viewpoint}} > \text{height}_{\text{object}}$$

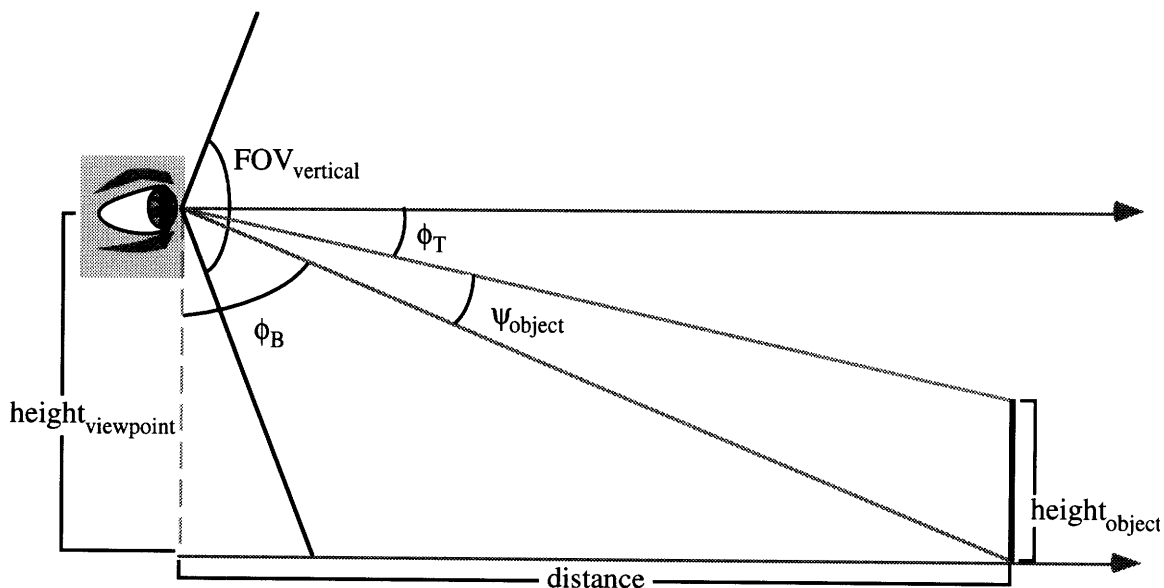


Figure 2.12: A side view of a model for calculating an object's visual angle. As distance increases, the visual angle subtended by the object, ψ_{object} , decreases according to a tangent function.

From the model in Figure 2.12, the following formulas can be derived:

$$\phi_B = \tan^{-1}\left(\frac{\text{distance}}{\text{height}_{\text{viewpoint}}}\right) \quad (8)$$

$$\phi_T = \tan^{-1}\left(\frac{\text{height}_{\text{viewpoint}} - \text{height}_{\text{object}}}{\text{distance}}\right) \quad (9)$$

$$\psi_{\text{object}} = \tan^{-1}\left(\frac{\text{height}_{\text{viewpoint}}}{\text{distance}}\right) - \phi_T \quad (10)$$

The formula describing the number of pixels composing the object is the same as before:

$$\text{Vpixels}_{\text{object}} = \text{round}\left[\text{Vpixels}_{\text{display}} \cdot \frac{\psi_{\text{object}}(\text{distance})}{\text{FOV}_{\text{vertical}}}\right] \quad (11)$$

Finally, the formulas determining the location of the end points of the object can be defined:

$$y_{\text{bottom}} = \text{round}\left[\text{Vpixels}_{\text{display}} \cdot \frac{\phi_B - 90^\circ + \frac{1}{2}\text{FOV}_{\text{vertical}}}{\text{FOV}_{\text{vertical}}}\right] \quad (12)$$

$$y_{\text{top}} = \text{round}\left[\text{Vpixels}_{\text{display}} \cdot \frac{\frac{1}{2}\text{FOV}_{\text{vertical}} - \phi_T}{\text{FOV}_{\text{vertical}}}\right]$$

These equations are used for the vertical dimension only. To fully understand the behavior of the object, the horizontal dimension should also be considered. The following figure shows the model of the object as viewed from above:

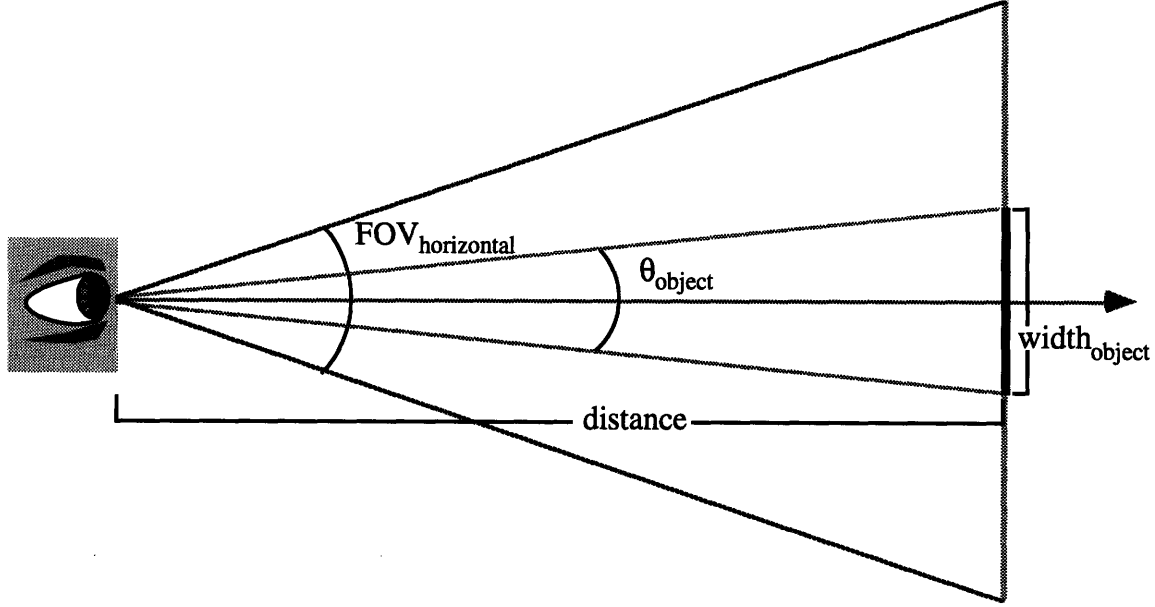


Figure 2.13: A top view of the same scene as depicted in the previous figure. The visual angle, θ_{object} , decreases according to Equation (13), for a fixed object width and increasing distance.

Repeating the previous derivations for the model shown in Figure 2.13, we have:

$$\theta_{\text{object}} = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} \text{width}_{\text{object}}}{\text{distance}} \right) \quad (13)$$

$$\text{Hpixels}_{\text{object}} = \text{round} \left[\text{Hpixels}_{\text{display}} \cdot \frac{\theta_{\text{object}}(\text{distance})}{\text{FOV}_{\text{horizontal}}} \right] \quad (14)$$

$$x_{\text{right}} = \text{round} \left[\text{Hpixels}_{\text{display}} \cdot \frac{\frac{1}{2}(\text{FOV}_{\text{horizontal}} - \theta_{\text{object}})}{\text{FOV}_{\text{horizontal}}} \right] \quad (15)$$

$$x_{\text{left}} = \text{round} \left[H_{\text{pixels}_{\text{display}}} \cdot \frac{\left(\frac{1}{2} \cdot \text{FOV}_{\text{horizontal}} + \theta_{\text{object}} \right)}{\text{FOV}_{\text{horizontal}}} \right] \quad (16)$$

The endpoints of the object will reflect both the effect of the size constancy and the effect of linear perspective since the endpoints are determined both by the location of the object and its size.

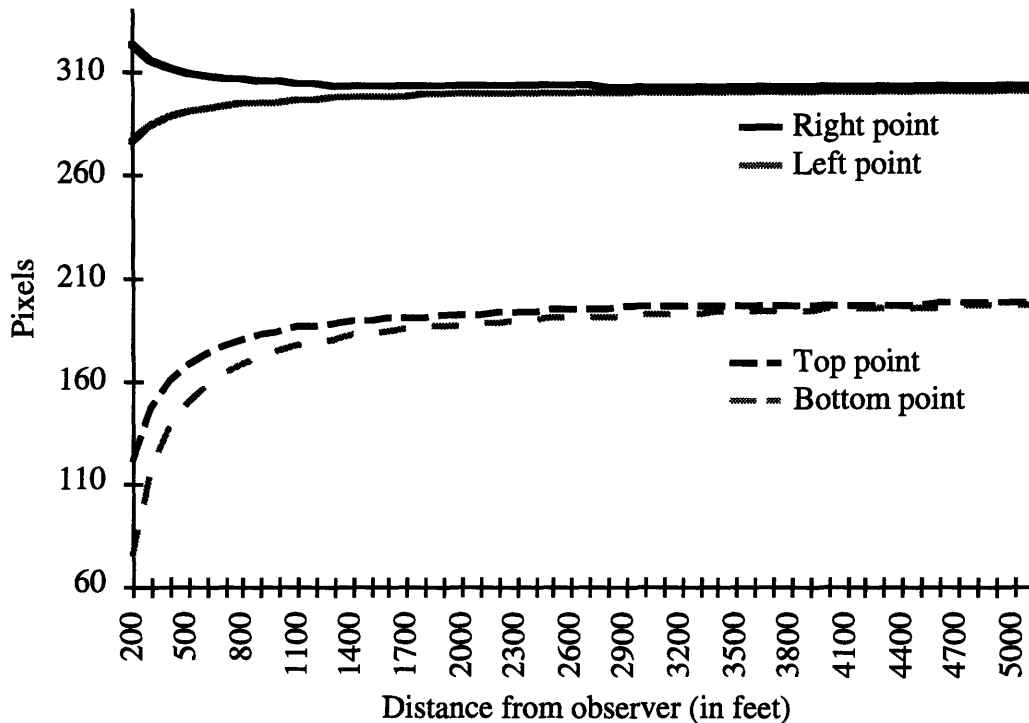


Figure 2.14: A plot showing the results of linear perspective and size constancy on object location. The space between the top and bottom lines (dotted) and the right and left line (solid) indicates the size of the object at various distances. The dimensions of the display are assumed to be 600 pixels by 400 pixels, and the object size is assumed to be 15 feet by 15 feet. In the right-left case, the object remains centered in the middle of the screen, at 300 pixels, while in the top-bottom case, the object moves towards 200 pixels.

Figure 2.14 shows a number of inconsistencies in the shape of the object as it recedes in depth. At a number of points the object is taller than it is wide, due to the 4 x 3 aspect ratio of the display. The interaction of size constancy and linear perspective is quite

apparent. Figure 2.15 shows, in more detail the behavior of the left and right points. In the horizontal case, everything seems to be appropriate; the size decreases consistently until the cutoff threshold point. Also, the cutoff point in this model (~5,400 feet) is closer than that in the simple model with a zero viewpoint height (~22,000 feet).

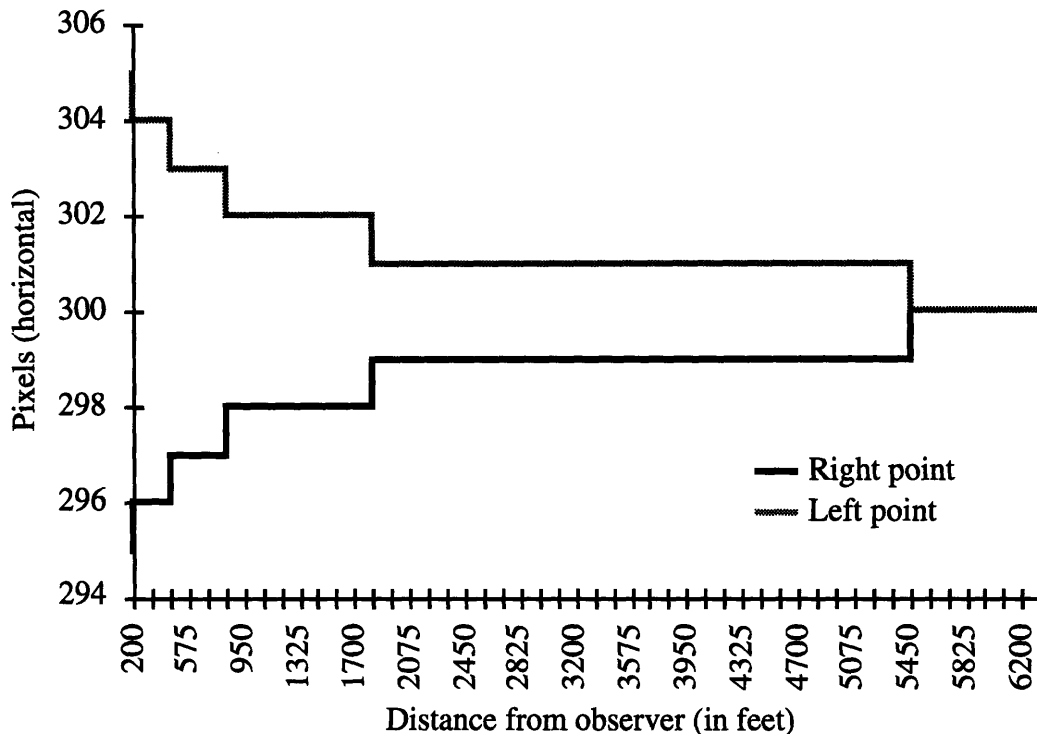


Figure 2.15: The predicted movement of the left and right edges of a 15' by 15' object as viewing distance increases. The size of the object at a particular distance is given by vertical distance between the plots for the left and right edges. The display is assumed to be 600 pixels wide.

The plot of the left and right points of the object shows no inconsistencies in the shape of the object. Again, the effect of pixel size on the appearance of the object is apparent. The movement of the top and bottom endpoints is more interesting since the observer is not viewing along the line to the center of the object. With the observer above the object being viewed, the object will move according to the equations that model linear perspective and will shrink according to the equations for size constancy. However, the

changes in the appearance of the object due to the two depth cues do not necessarily happen at the same time, as Figure 2.16 shows:

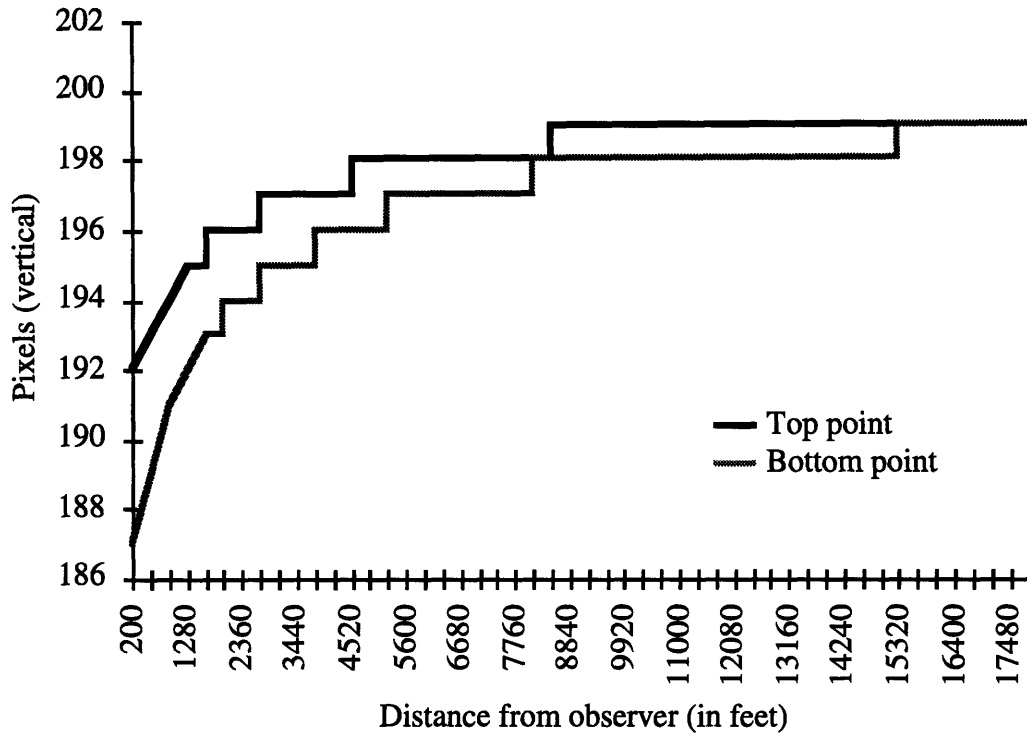


Figure 2.16: The predicted movement of the top and bottom edges of a 15' by 15' object as viewing distance increases. The size of the object at a particular distance is given by the vertical distance between the plots for the top and bottom points. The display is assumed to be 400 pixels tall.

Most notably, the object will disappear briefly at a distance of approximately 8,800 feet. The object, which is one pixel in size and moving towards the horizon, reaches a point where not enough of it is in either the pixel it is moving from or the pixel it is moving to. Thus, the object disappears until a sufficient portion of it moves into the new pixel.

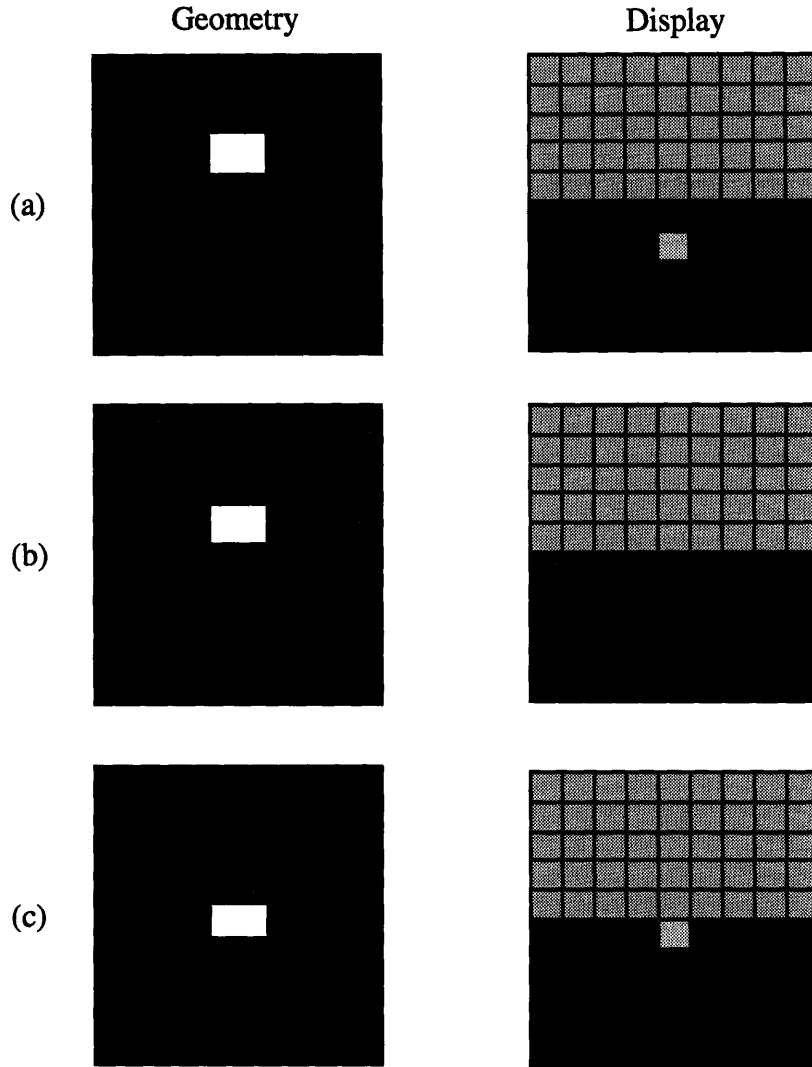


Figure 2.17: The disappearance-reappearance problem. The linear perspective and size constancy geometry predict the location and size of the object in the first column. Because of rounding in the graphics software and hardware, the object is actually displayed as in the right column. A traversal from (a) to (c) represents the result of increasing the viewing distance.

This disappearance-reappearance problem at the threshold distance has a parallel in the visible range. The size constancy and linear perspective steps do not occur at the same time, as shown in Figure 2.17. Thus, an object may shrink and grow intermittently. The object may be forced to move by linear perspective to a point where it overlaps more pixels and thus appears a pixel bigger than predicted by size constancy alone. So, the disappearance-reappearance problem implies a similar growth-shrinkage problem. Depth

estimation is clearly compromised by the disappearance-reappearance and growth-shrinkage problems.

The complexity of the problems in perspective geometry is proportional to the complexity of the model of the observer and the target stimulus. The problems associated with low resolution require more sophisticated analysis than is commonly thought. These problems deserve careful treatment since a carefully constructed solution has broad applications.

2.6.5.2 Returning to the Big Picture

As seen in the examples of other problems in virtual environments (see above, Solving Problems in Virtual Environments), a systems engineer often has to make difficult decisions about design tradeoffs. One way of compensating for the deficiency caused by the decision to make a hardware tradeoff (such as trading resolution for FOV) is to craft a software solution that makes a different compromise. As we have seen, the problems inherent in display systems are fairly complex.

A simple hardware solution to the problems caused by lack of spatial resolution is to simply make displays with more pixels per inch. However, the technology is not yet available to accomplish this, nor is it clear that additional pixels would be used to improve the spatial resolution of a display. The demand for improved FOV may outweigh the desire for better pixel resolution.

Thus, another kind of solution must be found. Perceptual tradeoffs are notoriously tricky and are best handled in a flexible way. Computer software is inherently adaptable and is a powerful tool for solving perception and display problems. Through careful measurement of human performance using the display with various software-controlled parameters, a reasonable solution can be achieved with relatively little effort.

The abstract idea of engineering software to match human perceptual performance is not a new one. Robinett and Rolland's model of the optical system in HMDs (1992) led to

Watson and Hodges' work (1995) involving the software predistortion of images to compensate for optical distortion.

The compromises made by VE systems designers should be based as much as possible on the best available evidence regarding the interaction between the human visual system and objective performance metrics (VETREC, 1992). An effective design results from trading off sets of variables, including economic and psychological cost factors, in order to optimize resources for reaching task goals (Miller, 1976). Determining operational parameters inevitably involves a number of tradeoffs among not only cost but also performance and efficiency. Zeltzer offers the throughput of geometric primitives, visual update rate, and display resolution as the major design parameters for a visual display (1991). Also, temporal sensitivity and resolution have a tradeoff (one cannot update a high-resolution image fast enough to show smooth motion), and image intensity and perceived color and brightness influence one another (Christou & Parker, 1995).

Given that any VE visual system design incorporates a significant number of tradeoffs between hardware limitations and human perceptual capabilities, providing software-based solutions seems to present an orthogonal domain in which to seek solutions. With the exception of the work done by Robinett and Rolland (1992), Watson and Hodges (1995), little effort has been made outside of traditional computer graphics to find the bridge between human visual perception and solutions found via the adaptability of software. Because of the flexibility of software and the ease and speed with which results can be tested, it seems an obvious direction to pursue solutions to some of the more daunting perceptual difficulties found in VE systems.

2.6.5.3 Implications

The previous statements about software solutions suggest that solutions in code are necessary elements in VE visual display system design. Furthermore, other capabilities of software have significant implications in the VE domain. Because of ability of VEs to

provide supernormal situations, exploiting tradeoffs in software could allow the transcendence of human visual capabilities. The psychological and perceptual biases mentioned above (in Problems with Virtual Environments) could be corrected by capitalizing on the flexibility of a software-driven system (Ellis, 1991). A solution that improves the range over which depth can be seen and does not significantly distort judgment could also be used to improve visibility to better-than-normal. This is exciting for potential enhanced-reality and instructional cueing applications.

The visibility-resolution problem itself has other implications. Not only would problems with visibility in VEs be solved, but other "smart" systems that suffer from the effects of poor resolution in depth judgment could also be improved. Most notably, night-vision goggles suffer from poor resolution which limits visibility and the overall effectiveness of the device. Thus, finding a solution for the effects of low resolution displays on visibility has other potentially useful ramifications.

3. Experiment

In order to validate the idea of using the aforementioned software-based perceptual manipulations, empirical evidence of the success of this methodology should be obtained. Because problems with VEs are so task-dependent, the best way to experiment with the software manipulation of visual cues is with a concrete example. The work in this thesis has been motivated by more than pure scientific interest, of course, and a practical application of the knowledge has been a driving force for this research.

3.1 Experiment Background

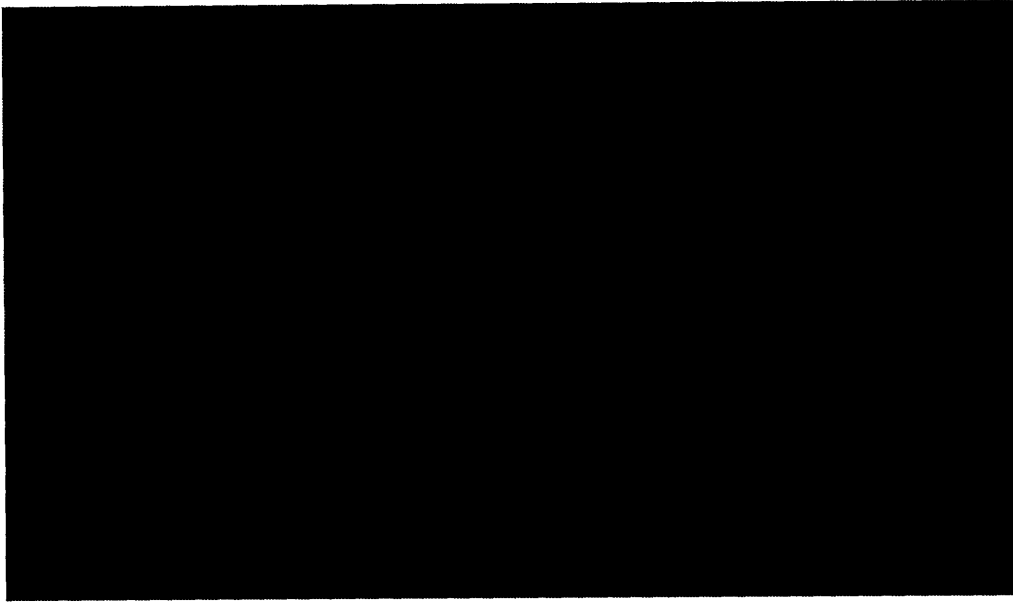
The mission of the Virtual Environment Technology for Training group at the Massachusetts Institute of Technology has been to investigate the manner in which newly-developed immersive interface techniques can be applied to the learning of complex tasks. One of the main projects of interest to both the group and the sponsor (the Naval Air

Warfare Training Systems Center) is the development of a submarine simulator that is capable of teaching Navy personnel the basics of boat navigation on the surface of a harbor or bay. The project aims to improve understanding of the advantages and disadvantages of VE training over more traditional simulation methods (VETREC, 1992).

In marine navigation, one officer is in charge of making all steering decisions for the boat. On a submarine, this officer stands on the "sail" and gives navigation commands to the rest of the navigation crew who are located below-deck. This position is known as the "Officer of the Deck" or "OOD" (Levison, Tenney, Getty, & Pew, 1995; Zeltzer, Aviles, Gupta, Lee, Nygren, Pfautz, & Reid, 1994).

The OOD task centers around the visual recognition of several cues: the motion of the water, the texture of the water, and, most importantly, the presence of channel buoys and range markers. The OOD guides the boat through the channel marked by the buoys, and uses the range markers to ensure that the boat is centered in the channel. For more information on the OOD task and how it was selected, see Levison, Pew, and Getty (1994) and Levison et al. (1995).

The ability to see the channel markers is extraordinarily important to the performance of the OOD task. However, the buoys and range markers are not visible in the simulator at the distance they would be visible in the real world. No data has been collected on the performance in the OOD simulator without the navigation aids. However, the following snapshots of the same scene illustrate the difficulty presented by the lack of visibility:



(a)



(b)

Figure 3.1: The effect of scaling on visibility in the OOD simulator. (a) shows the unscaled scene from a particular viewpoint, while (b) presents the result of scaling. The objects in the distance are much more visible in (b) than in (a).

Obviously, the ability to see the buoys is impaired when no scaling method is used. Trying to navigate a channel without being able to see more than a few buoys ahead is very difficult. In addition, the range markers are not clearly visible, inhibiting their use as navigational aids.

Basically, the development of a simulator like the OOD requires sufficient realism to allow the task to be trained. In this situation, critical information is eliminated, removing realism and making the virtual world too unlike the real world. The lack of resolution, in this task, is not acceptable.

3.1.1 Geometry of the OOD model

The OOD simulator had a specific model of the world that was used to determine the visual relationships between objects in the computation of the graphics. In particular, the submarine sail was said to be 34 feet off of the water, while the viewpoint was calculated to be 39 feet by adding 5 feet for the height of a human observer's eyes. This number was based on the size of the submarine model used in the simulator. In addition, the simulation designers consulted Navy personnel to ascertain the rectitude of the model's dimensions (Pioch, 1995).

These same individuals also validated the size of the buoys, whose dimensions were originally given by the U.S. Coast Guard. The following figure shows the appearance and dimensions of the buoys in the OOD simulator:

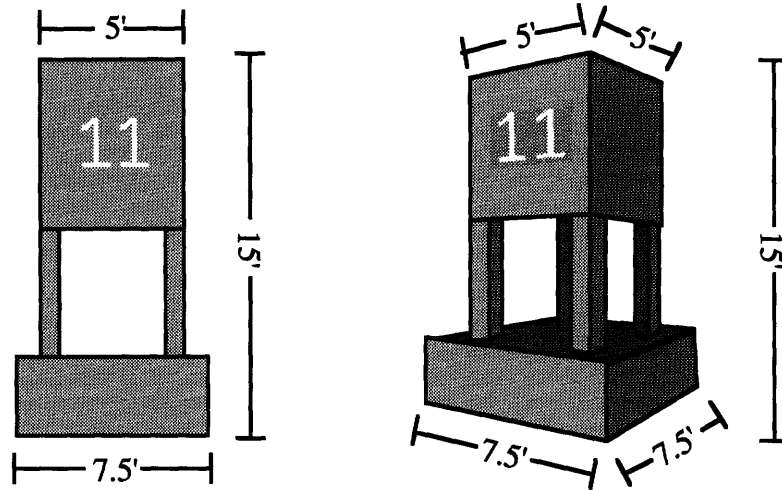


Figure 3.2: The dimensions of the buoy model used in the OOD simulator.

Given the dimensions of the submarine and buoy models, we can now extend the models presented above to accommodate objects with a three-dimensional shape. The following figure shows a three dimensional view of the perspective geometry in the OOD simulator:

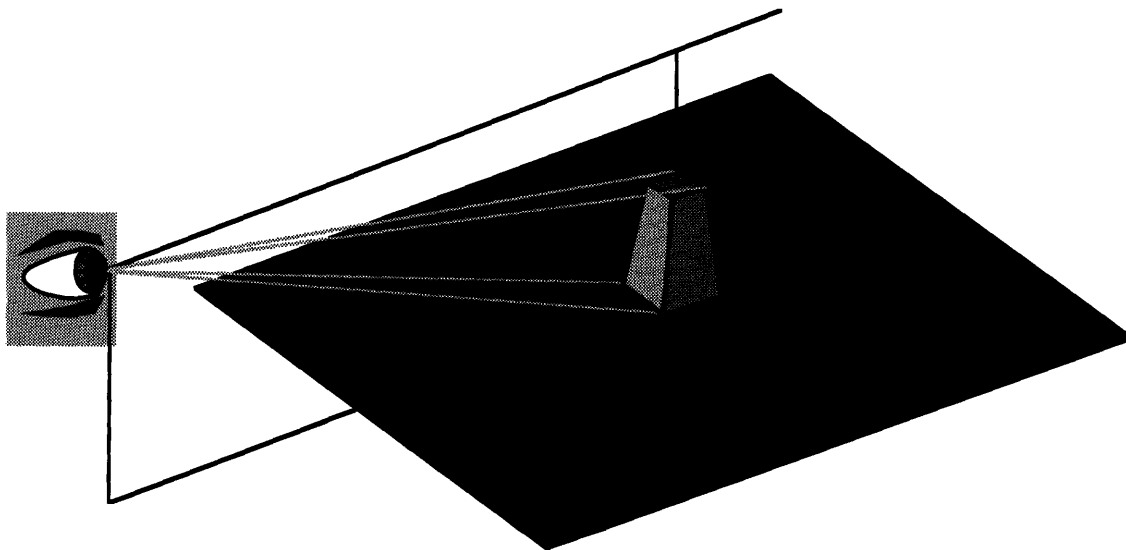


Figure 3.3: A representation of the perspective geometry model used to display a buoy object in the OOD simulator.

The relevance of the previous calculations of perspective geometry is readily apparent. We can proceed by repeating the calculations presented above and incorporating a three-dimensional object that is non-regular in one dimension (rather than the flat 15 foot by 15 foot square used before).

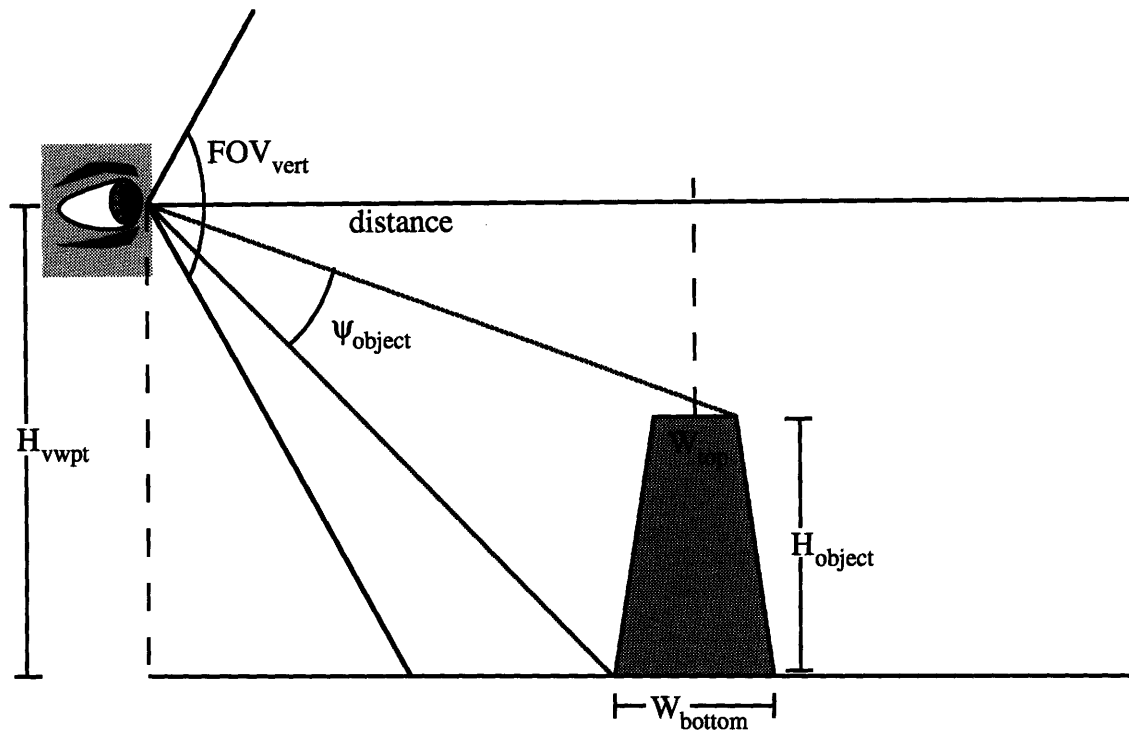


Figure 3.4: A side view of the perspective geometry used in the OOD simulator to view a buoy object.

From this model of the side view of the buoy object, the following formulas can be derived:

$$\phi_B = \tan^{-1} \left(\frac{\text{distance} - \frac{1}{2} W_{\text{bottom}}}{\text{height}_{\text{viewpoint}}} \right) \quad (17)$$

$$\phi_T = \tan^{-1} \left(\frac{\text{height}_{\text{viewpoint}} - \text{height}_{\text{object}}}{\text{distance} + \frac{1}{2} W_{\text{top}}} \right) \quad (18)$$

$$\psi_{\text{object}} = \tan^{-1} \left(\frac{H_{\text{vwpt}}}{\text{distance} - \frac{1}{2} W_{\text{bottom}}} \right) - \tan^{-1} \left(\frac{\text{height}_{\text{viewpoint}} - \text{height}_{\text{object}}}{\text{distance} + \frac{1}{2} W_{\text{top}}} \right) \quad (19)$$

The formula describing the number of pixels in the object is the same as before, as are the calculations to determine the location of the top and bottom points of the object since they depend entirely on the object's visual angle.

Proceeding to the top view of the figure, we are presented with the dilemma of deciding whether to model the object using the width of the base or the width of the top.

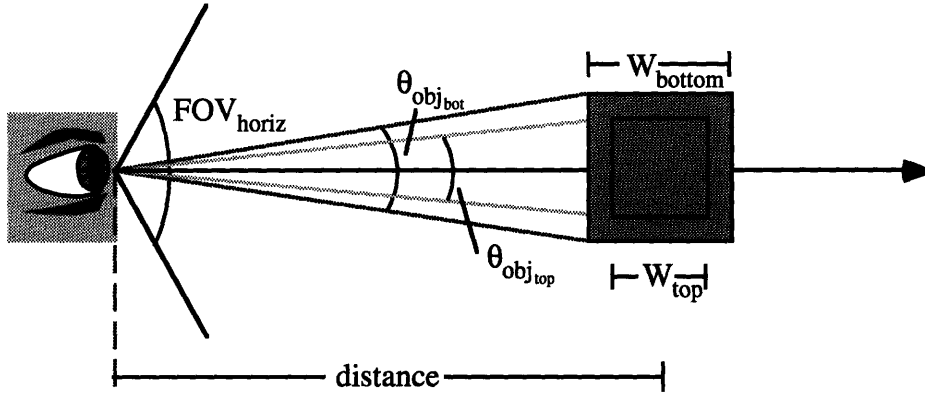


Figure 3.5: A top view of the perspective geometry used to present a buoy object in the OOD simulator.

Using formulas (13) through (16), and substituting the visual angle for either the top or bottom of the buoy for θ_{object} , the formulas for the horizontal appearance of the buoy object can be found. Before choosing either the top of the object or the bottom (or some combination) as a basis for modeling, the dominating dimension should be found. That is, because of the disparity in number of vertical and horizontal pixels and the disparity in the height of the buoy object versus its width, one of the sets of formulas will determine the

distance at which the object is last visible. Intuitively, the horizontal (θ_{object}) dimension would seem to dictate the cutoff point. This is easy to verify via substitution.

Substituting the parameters of the HMD and the OOD model (viewpoint height, buoy dimensions) used into formula (4), the visual angle subtended by half of a pixel can be determined. Half a pixel is used since the graphics system rounds half pixels to full pixels.

$$\text{pixels}_{\text{object}} = \text{round}\left(\text{pixels}_{\text{display}} \cdot \frac{\theta_{\text{object}}(\text{distance})}{\text{FOV}}\right) \quad (4)$$

$$\theta_{\text{object}} = 36^\circ \cdot \frac{0.5 \text{ pixel}}{486 \text{ pixels}}$$

$$\psi_{\text{object}} = 48^\circ \cdot \frac{0.5 \text{ pixel}}{648 \text{ pixels}}$$

$$\theta_{\text{object}} = 0.037^\circ$$

$$\psi_{\text{object}} = 0.037^\circ$$

Not surprisingly, since the aspect ratio matches the ratio of field of view and the number of pixels in either dimension, the half-pixel angle is the same. Solving equations (19) and (13) for $\theta_{\text{obj}_{\text{top}}}$, $\theta_{\text{obj}_{\text{top}}}$ and ψ_{object} equal to the half- or one-pixel angle, the threshold distance can be determined. First, the vertical dimension, using the half-pixel angle:

$$\psi_{\text{object}} = \tan^{-1}\left(\frac{H_{\text{vwpt}}}{\text{distance} - \frac{1}{2} W_{\text{bottom}}}\right) - \tan^{-1}\left(\frac{H_{\text{vwpt}} - H_{\text{object}}}{\text{distance} + \frac{1}{2} W_{\text{top}}}\right) \quad (19)$$

$$0.037 = \tan^{-1}\left(\frac{39 \text{ feet}}{d_{\text{vertical_cutoff}} - \frac{1}{2} \cdot 7.5 \text{ feet}}\right) - \tan^{-1}\left(\frac{39 \text{ feet} - 15 \text{ feet}}{d_{\text{vertical_cutoff}} + \frac{1}{2} \cdot 5 \text{ feet}}\right)$$

$$d_{\text{vertical_cutoff}} = 23218.49 \text{ feet}$$

Now, solving for the horizontal case, using the one-pixel angle:

$$\theta_{\text{object}} = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} \text{width}_{\text{object}}}{\text{distance}} \right) \quad (13)$$

$$\theta_{\text{obj}_{\text{top}}} = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} W_{\text{top}}}{d_{\text{horizontal_cutoff}_{\text{top}}}} \right) \quad \theta_{\text{obj}_{\text{bottom}}} = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} W_{\text{bottom}}}{d_{\text{horizontal_cutoff}_{\text{bottom}}}} \right)$$

$$0.074 = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} \cdot 5 \text{ feet}}{d_{\text{horizontal_cutoff}_{\text{top}}}} \right) \quad 0.074 = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} \cdot 7.5 \text{ feet}}{d_{\text{horizontal_cutoff}_{\text{bottom}}}} \right)$$

$$d_{\text{horizontal_cutoff}_{\text{top}}} = 3867.5 \text{ feet}$$

$$d_{\text{horizontal_cutoff}_{\text{bottom}}} = 5801.2 \text{ feet}$$

We use the one-pixel angle in the horizontal case since, in this model, the center line of the object would match up with the split between the middle two pixels in the display. Thus, to be visible, an object would have to cover half each of those two pixels. Given the values derived, our intuition that the horizontal aspect of the object will cutoff first is correct. Still, the final determination of the cutoff point should be determined empirically. By simply observing the point at which the object disappears (which is possible since 0.037° is significantly greater than human visual acuity), the actual threshold distance was found:

$$d_{\text{far_cutoff}} = 5347 \text{ feet}$$

The presented models could deviate from the data for several reasons. Most notably, the rounding procedure for displaying an object that covers half-pixels is not easily found. Most graphics software and hardware systems bury simple pixel-rounding

functions below many layers of other mechanisms. In addition, the answer may be very complex. Since a large number of graphics packages rely on blurring (anti-aliasing) to accommodate partial pixels, finding a simple answer for how a non-blurred pixel is rounded off in non-anti-aliased images is very difficult. However, since the empirical value for the cutoff distance falls between the values determined for the top and bottom widths, we can be assured that our models are sufficient for describing the visual behavior of the buoy object at different depths.

In addition to determining the threshold distance for visibility, the first point at which the object is fully visible in the display should be calculated. Because the object could subtend a visual angle greater than the FOV or could be cut off by the size of the FOV, its visual angle will not be properly represented in the display. The near limit is given in Equation (7), but both horizontal and vertical components need to be considered. Also, the more comprehensive models should be used in the calculation. Solving for distance in the vertical case, where the object's visual angle is equal to half of the FOV, we find:

$$\psi_{\text{object}} = \frac{1}{2} \cdot \text{FOV}_{\text{vert}} = \tan^{-1} \left(\frac{H_{\text{vwpt}}}{\text{distance} - \frac{1}{2} W_{\text{bottom}}} \right)$$

$$18^\circ = \tan^{-1} \left(\frac{39 \text{ feet}}{d_{\text{vertical_cutoff}} - \frac{1}{2} \cdot 7.5 \text{ feet}} \right)$$

$$d_{\text{vertical_cutoff}} = 123.78 \text{ feet}$$

In the horizontal case, the near cutoff occurs when the visual angle of the object reaches the full FOV. The difference in the solutions to the two different dimensions can

be elucidated by a quick examination of Figure 3.4 and Figure 3.5. Now, we solve for the near horizontal cutoff distance:

$$\theta_{\text{obj}_{\text{bottom}}} = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} W_{\text{bottom}}}{d_{\text{horizontal_cutoff}}} \right)$$

$$48^\circ = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} \cdot 7.5 \text{ feet}}{d_{\text{horizontal_cutoff}}} \right)$$

$$d_{\text{horizontal_cutoff}} = 8.42 \text{ feet}$$

Thus, the near fully visible point is determined by the vertical constraint. Given the previous calculations, we now are able to compute the range of visibility of any display. This presentation addresses a specific case, but extending the calculations presented here to accommodate other HMDs or other scene models is trivial.

3.1.1.1 Determining Real World Visibility

Having determined the range of visibility for buoy objects, the next step is to try and establish an approximate value for the real-world range of visibility. The problem of determining real-world visibility is extraordinarily difficult. A great deal of estimation must be done in order to find any sort of reasonable solution, and error in the answer is likely to be significant.

The problem of determining real-world visibility is difficult for a number of reasons. The most obvious explanation is that target detection is a form of visual acuity, and visual acuity varies significantly from person to person (Boff & Lincoln, 1988). Not only is population variance a factor in visual acuity, but there are a number of factors that

have been shown to strongly influence the detectability of an object. A short list of these factors includes:

- visual acuity increases with high illumination
- visual acuity decreases with target motion
- visual acuity decreases with increased distance-to-target
- visual acuity varies with the visual task
- visual acuity varies with the target object used (Boff and Lincoln, 1988)

Most of these elements receive treatment in the design of the experiment below. Human visual acuity, as treated in experimental psychology, is a sufficiently similar problem to the visibility trouble with HMDs that the methodology is the same.

Geise, in 1946, reported that visual acuity varies with distance. However, he noted that after a viewing distance greater than 5 m was reached the change in visual acuity was fairly minimal. He noted that visual acuity was about 1.5 times worse at 5 m than at 20 cm. These data suggest that visual acuity at a distance, while decreasing significantly, will remain close to acuity at near distances.

In addition to the constraints posed by human visual performance, the environment in which buoys are seen in the real world is highly variable. The time of day, the latitude and longitude, and the weather all determine the amount of illumination a buoy receives. The color and roughness of the water also play a part in the buoy's discriminability. Furthermore, not all buoys are seen with the water as a backdrop, some buoys are seen with a land mass behind them. The color of a land mass is also highly variable.

How can a reasonable estimate be derived if the variability in the real world is so great? One method is to solicit the experience of actual naval officers who have performed the OOD task in the real world. An experienced U.S. Navy Lieutenant explained the Navy's rules for buoy placement and distribution and claimed that, based upon his experience, buoys were visible at distances up to 3 miles. In addition, the officer pointed out the simple cases where "I normally could see that" in the simulation of a bay with

which he was familiar. This also provided data suggesting to a visibility threshold of about two or three miles (Pioch, 1995).

Independent of the variability of the real-world data, the need for better visibility in the OOD simulator is clear. The estimation of a threshold distance can be accomplished with some degree of accuracy. The thresholds based on pixellation in the simulator's display are much shorter than are needed to adequately represent estimated threshold found for the real world.

3.1.1.2 Assumptions

For simplicity, we will focus only on the buoys as significant examples of the visibility problem, disregarding the other objects that also suffer from reduced visibility. Solving the visibility problem with the buoys is tantamount to solving the visibility problem with the other navigational aids and may eventually be extensible to other visibility problems.

In addition, the curvature of the earth is ignored in all calculations. The simulator models the earth as a flat plane, not as an oblate spheroid, (ignoring the work of such notables as Christopher Columbus) in order to simplify the model and its dynamics. Thus, for the purposes of the following calculations, the Earth is flat.

Furthermore, the task in the OOD simulator is a training task (Levison, Pew, & Getty, 1994). While the focus thus far may seem to be directed towards human performance, the actual experimentation will attempt to assess not only the effects of poor spatial resolution on depth estimation performance, but also on the effects of resolution on the training of depth estimation.

Because the project involves a simulator, realism is a significant operating constraint. The solution to the difficulties in visibility should attempt to match real-world visibility and depth perception as well as possible. Bearing this in mind, we turn back to the issues of perspective geometry.

3.1.1.3 Previous Work on Visibility in the OOD Simulator

Problems with the visibility of the buoys and range markers were first reported in by Pioch (1995). However, his solution failed to account for a number of effects of perspective geometry and human performance. This work describes the implementation of a piecewise linear scaling algorithm which is used to make the buoys and other navigational aids visible (Pioch, 1995).

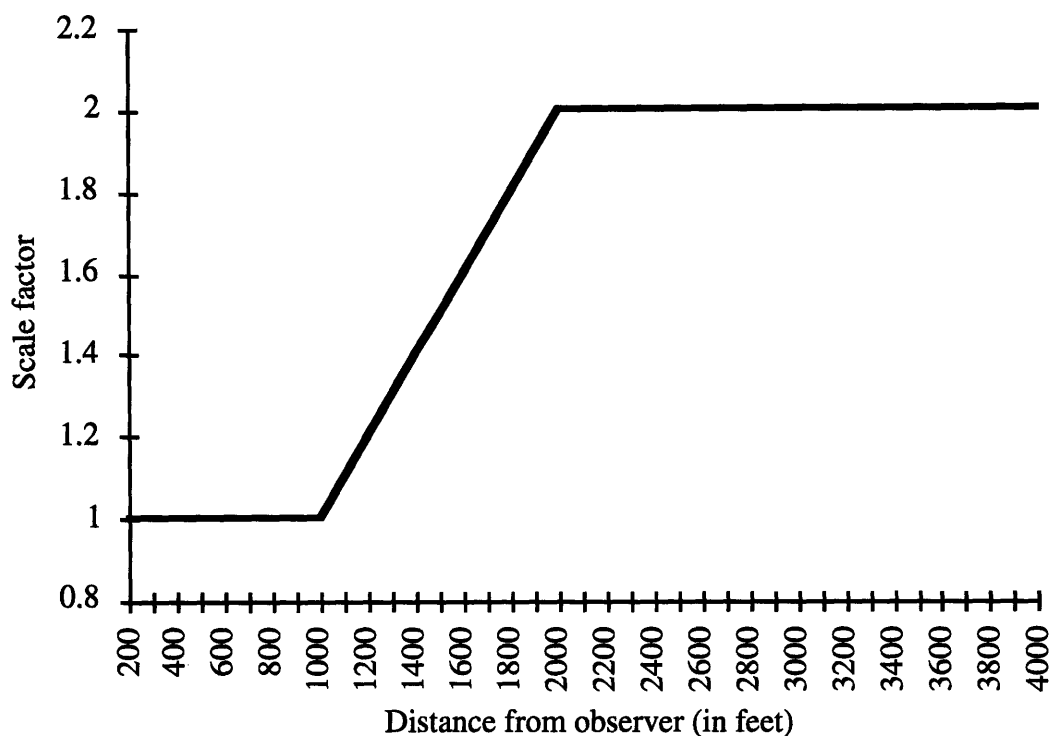


Figure 3.6: The behavior of a previous solution to the visibility problem in the OOD simulator. The object is gradually scaled to twice its original size over the range from 1000 to 2000 feet.

This algorithm fails to account for the effects on distance estimation that will occur between 1,000 and 2,000 yards when the buoy fails to shrink at the correct rate. In fact, an examination of Figure 3.7 shows that the object will remain the same size from 1,000 feet to 2,000 feet. Thus, distance estimation will be confused since users are normally able to discriminate a number of distinct depths in that range. Essentially, gaining additional

pixels at a distance in this manner sacrifices all discriminability in the 1000 to 2000 foot range.

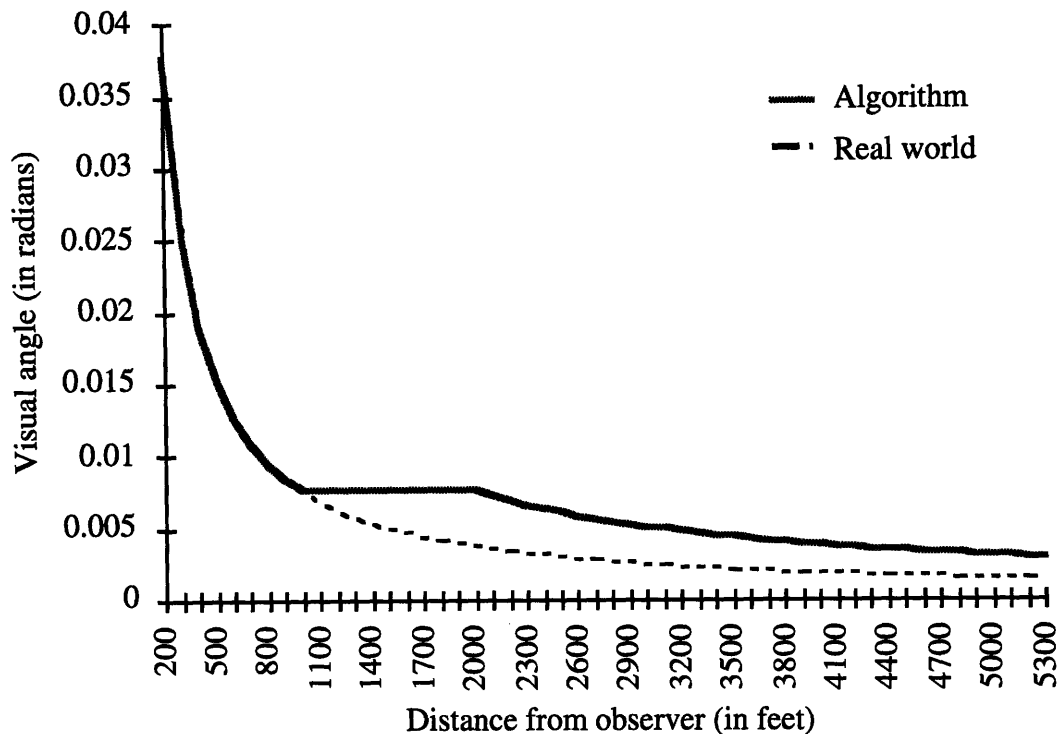


Figure 3.7: Graph showing that the previously-designed algorithm fails to minimize the distortion in distance estimation. It maintains the same visual angle for 1000 feet, eliminating depth discrimination in that range.

Clearly, the previously designed method proposed for solving the visibility problems in the OOD can only be classified as an "engineering solution" for the immediate improvement of the simulator since it was not based upon a robust investigation of the perspective geometry which determines the effects of size constancy and linear perspective.

The problem of visibility in a HMD can be described as a threshold perception problem; an object receding into the distance has a definitive point where it can no longer be seen by the user. Software solutions to the threshold problem must be careful not to introduce distance estimation errors in depth perception. Two kinds of errors can result: bias errors and resolution errors. Distortion of the ability to discriminate the depth of an

object is a bias effect. A change in the variability of a S's response at a particular depth is a resolution effect. That is, if an visibility-enhancing algorithm does not significantly increase the mean error in a Ss reply, it has a minimal bias. If an algorithm does not significantly increase the variability in the responses, it has a minimal effect on resolution.

Previous solutions have failed to account for both sorts of potential judgment problems caused by algorithmic distortions of the appearance of the object. A good solution should minimize the potential for distance estimation errors while extending the threshold visibility point.

3.1.1.4 Finding a Solution

An ideal solution to the visibility problem would maximize the distance over which an object is visible and minimize depth estimation errors; more simply, the best solution is the most realistic one in terms of bias and resolution. This constraint eliminates simple algorithms that might scale the target object by enough of a constant factor to make it visible at the distance required. Simply scaling the object introduces a significant distortion, especially when the distance between the observer and the object is small. In addition, extending the visibility of a 15 foot by 15 foot target object from two miles to three miles would require scaling the object to 22.5 feet, a distortion that would be clearly discernible at close distances.

Another possible solution would be to extend the range over which an object is one pixel in size.

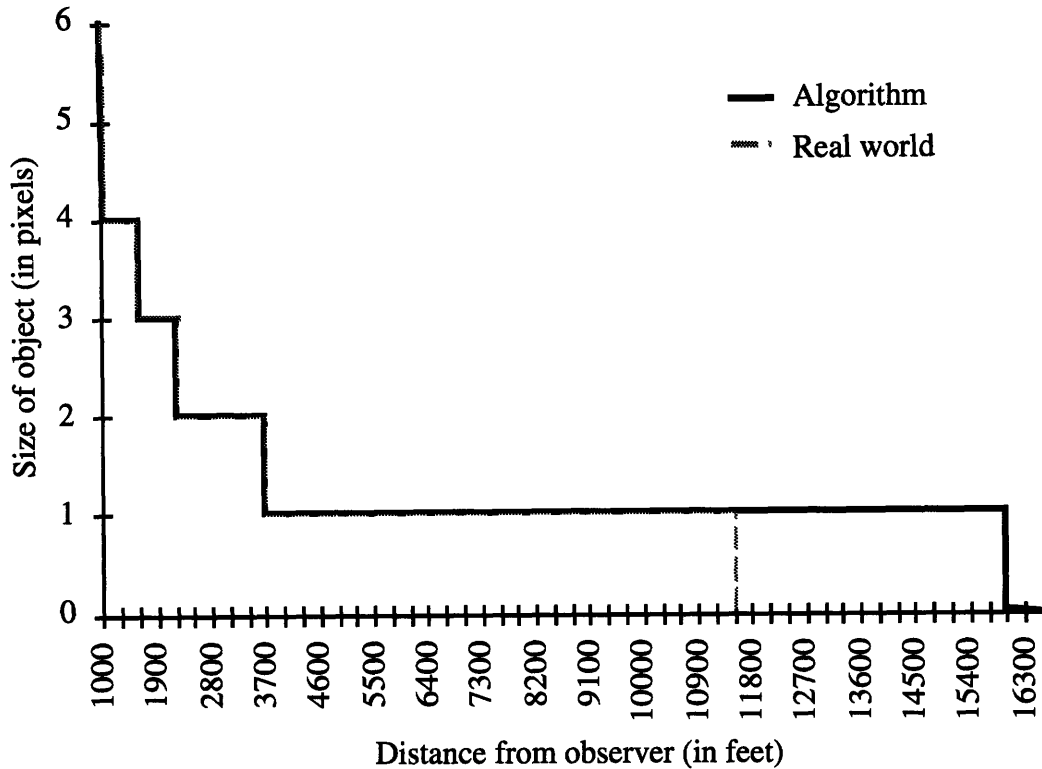


Figure 3.8: Extending the visibility by making the object stay one pixel in size until the desired visibility point is reached.

This would certainly extend the visibility but would eliminate the ability to discriminate depths over a large range of distances. According to Figure 3.8, an object would stay one pixel in size for about 13,000 feet.

Human perception utilizes more than just perspective cues in the perception of depth. A successful algorithm could utilize color and atmospheric cues to extend the visibility of an object. For example, a fraction of a pixel could be displayed by blending it with its neighboring pixels. This technique is known in computer graphics as *anti-aliasing*. Using anti-aliasing, an object could fade out into the background gradually as it recedes into the distance. Extending the distance could be achieved simply by reducing the rate at which it changes into the color of the background.

Recalling the discussion of realism above, this solution seems to be the best, upon first glance. It would minimize bias and resolution errors in distance estimation, while

providing a fairly close approximation of what occurs in human depth perception. However, quantifying the results of such an algorithm would be extraordinarily difficult. While HMDs are fairly consistent in their resolution and FOV, they vary greatly in their ability to produce color.

In our work at M.I.T., we have noted a significant number of color differences between displays in *the same brand of HMD*. As discussed above (see Difficulties with Head-Mounted Displays), most HMDs introduce a significant amount of color distortion. The optics may introduce chromatic aberration, or the color range of a particular display technology may be limited (Barfield et al., 1995). Furthermore, incorrect application of anti-aliasing techniques can actually exacerbate the pixellation of depth (Christou & Parker, 1995). Thus, the logical choice was to find an algorithm that optimized the perspective *geometry* rather than utilizing a color change across the visible range of distances.

A successful solution should minimize the deviation from the expected visual angle, especially at near distances, to minimize distance estimation error while also extending the visibility at far distances. These criterion match the desire for realism as well as providing sufficient improvement in task-specific performance. A realistic solution will deviate very little from the real-world visual angle and will present objects that are visible at real-world distances.

3.1.2 The Geometry of Optimal Solutions

The best way to avoid distorting depth judgment is to evenly distribute the pixels across the range which an object should be visible. An algorithm could let the number of pixels subtended by the object be normal at the closest distance and slowly increase the size the object as it moves away so that it subtends more pixels than it would otherwise.

The number of pixels subtended by an object at a distance can be increased in two ways. One, the size of the object can be scaled as a function of distance, where at the nearest distance the object is normal sized and is gradually scaled as it recedes so that it

reaches the disappearance point at the minimum size. This is best illustrated by observing the effect of the algorithm on visual angle as a function of distance.

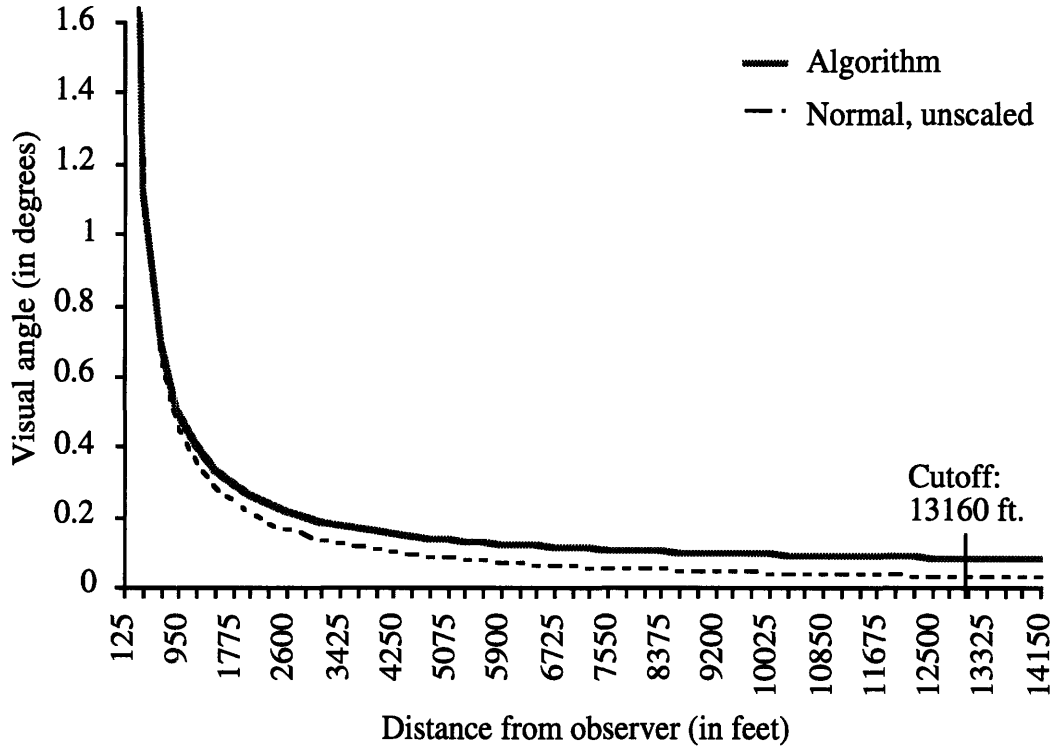


Figure 3.9: A plot showing the effects of the size scaling algorithm on visual angle as a function of distance. The deviation from the normal (dotted line) behavior of the visual angle is minimized.

To grasp this function completely, we recall that the empirical cutoff point for an unscaled object was 5,347 feet. Given this value for distance, we can determine the width of the object that serves as the actual determinant for cutoff:

$$\theta_{\text{object}} = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} \text{width}_{\text{object}}}{\text{distance}} \right) \quad (13)$$

$$0.074^{\circ} = 2 \cdot \tan^{-1} \left(\frac{\frac{1}{2} \text{width}_{\text{obj}_{\text{cutoff}}}}{5347 \text{ feet}} \right)$$

$$\text{width}_{\text{obj_cutoff}} = 6.91 \text{ feet}$$

This width serves as the basis for the scale factor at a particular distance. This is the width that is scaled to match the visual angle necessary for visibility at a particular distance, rather than just the top or bottom width. We can now present the formula for scaling the size of the object:

$$\text{size}_{\text{scaled_object}} = \frac{2 \cdot \text{distance} \cdot \tan\left(\frac{\theta_{\text{obj_cutoff}}}{2}\right)}{\text{width}_{\text{obj_cutoff}}} \cdot \frac{\text{distance}}{\text{distance}_{\text{desired_visibility}}} \cdot \text{size}_{\text{object}} \quad (20)$$

Despite appearances, this is a simple computation since the scale factor (the first fraction) can be computed before run time, so that the formula used is really:

$$\text{size}_{\text{scaled_object}} = \text{scalefactor} \cdot \frac{\text{distance}}{\text{distance}_{\text{desired_visibility}}} \cdot \text{size}_{\text{object}}$$

The result of this algorithm is to stretch the depth ranges caused by pixellation to accommodate the improved visibility while, as shown in Figure 3.9, the distortion of visual angle subtended is minimized.

Another algorithm utilizes a distortion of the field of view to improve visibility. By narrowing the FOV with increasing distance, an object appears normal at close distances but becomes larger when it is further away. This is best understood by imagining a pair of binoculars that dynamically increase magnification as an object gets further away. In this case, the increase in magnification is scaled so as to match the minimum size of the object with the desired visibility range.

The following graph shows how the distortion from the normal visual angle is minimized:

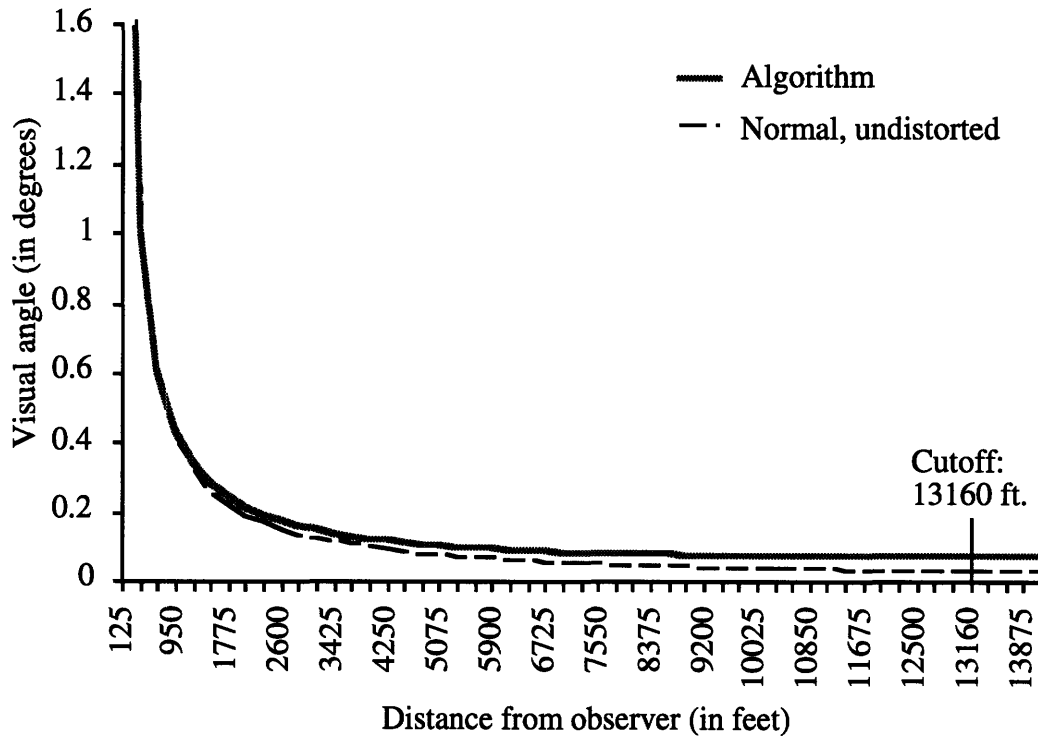


Figure 3.12: A plot showing the effects of the FOV distortion algorithm on visual angle as a function of distance. The deviation from the normal (dotted line) behavior of the visual angle is minimized.

The formula for the FOV distortion algorithm is given as:

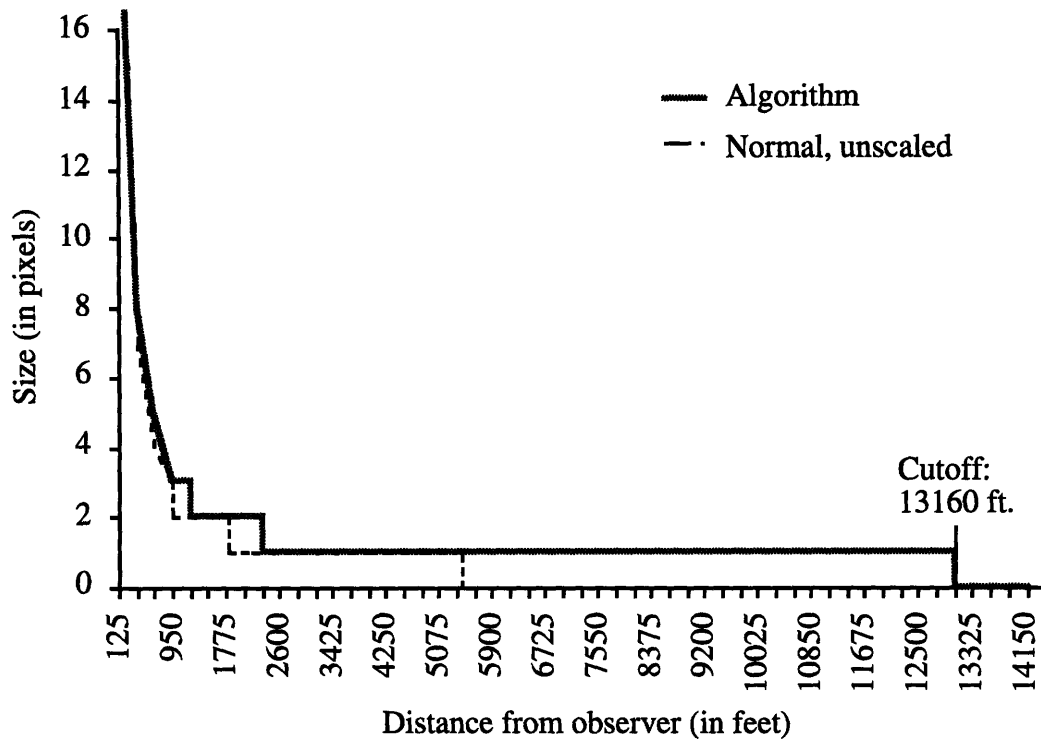


Figure 3.10: The stretching of depth ranges by the size scaling algorithm accommodates a large range of visible distances.

Finally, the effect of the algorithm can be best understood by comparing the scaled version of the buoy object with the unscaled version at a number of distances:

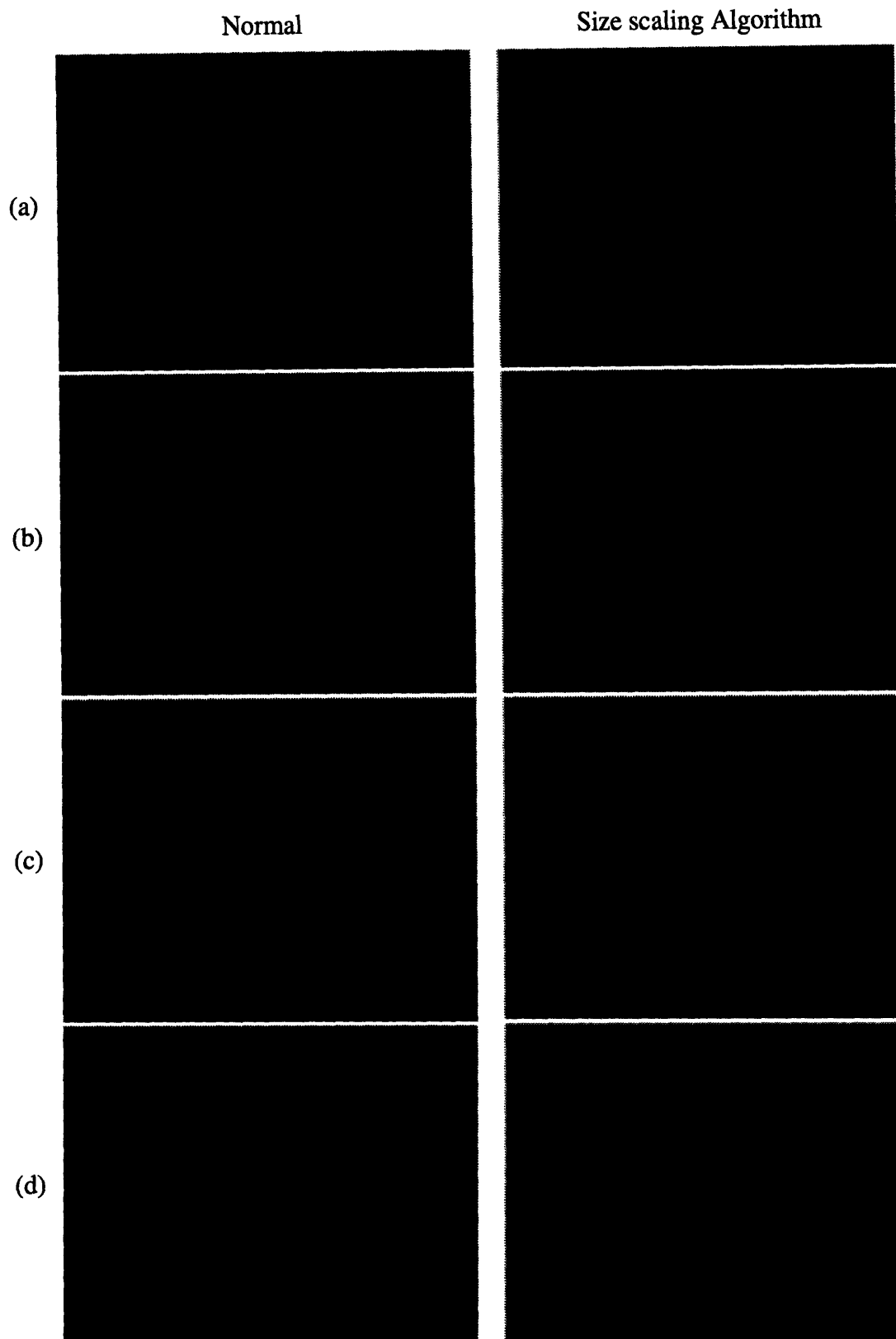


Figure 3.11: A comparison of the normal, unscaled buoy object to the buoy object altered by the size scaling algorithm. (a) At 125 feet. (b) At 1,100 feet. (c) At 3,000 feet. (d) At 13,000 feet. Note the motion of the buoy object towards the horizon, especially in (d).

$$FOV_{scaled} = \left(1 - \frac{2 \cdot H_{pixels_{display}} \cdot \tan^{-1} \left(\frac{\frac{1}{2} width_{obj_{cutoff}}}{distance} \right)}{FOV_{normal}} \right) \cdot \frac{distance}{distance_{desired_visibility}} \cdot FOV_{normal} \quad (21)$$

The width used in the calculation is the same as the size algorithm, thereby incorporating the empirical cutoff point into the visibility calculation. Formula (21) can be simplified by precalculating the scale factor:

$$FOV_{scaled} = scalefactor_{FOV} \cdot \frac{distance}{distance_{desired_visibility}} \cdot FOV_{normal}$$

As in the size-scaling algorithm case, the number of pixels and the visual angle subtended are distorted, while improving visibility to cover the desired range.

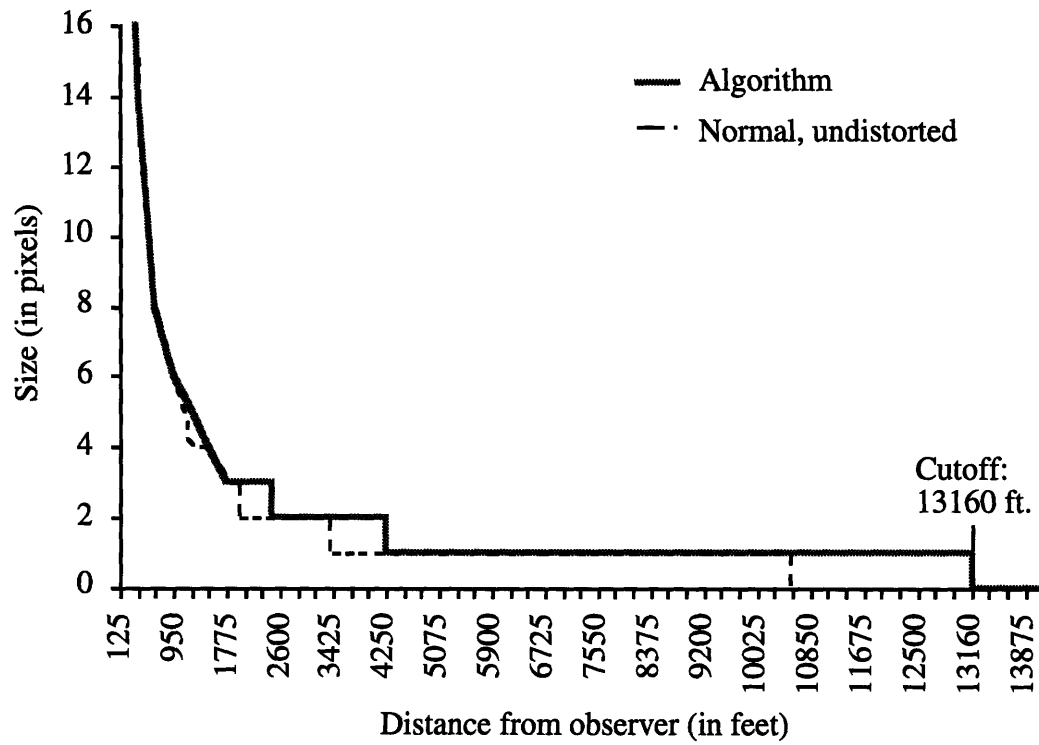


Figure 3.13: The stretching of depth ranges by the size scaling algorithm accommodates a large visible range.

Again, the properties of the FOV distortion algorithm are best understood in a visual comparison to the normal case:

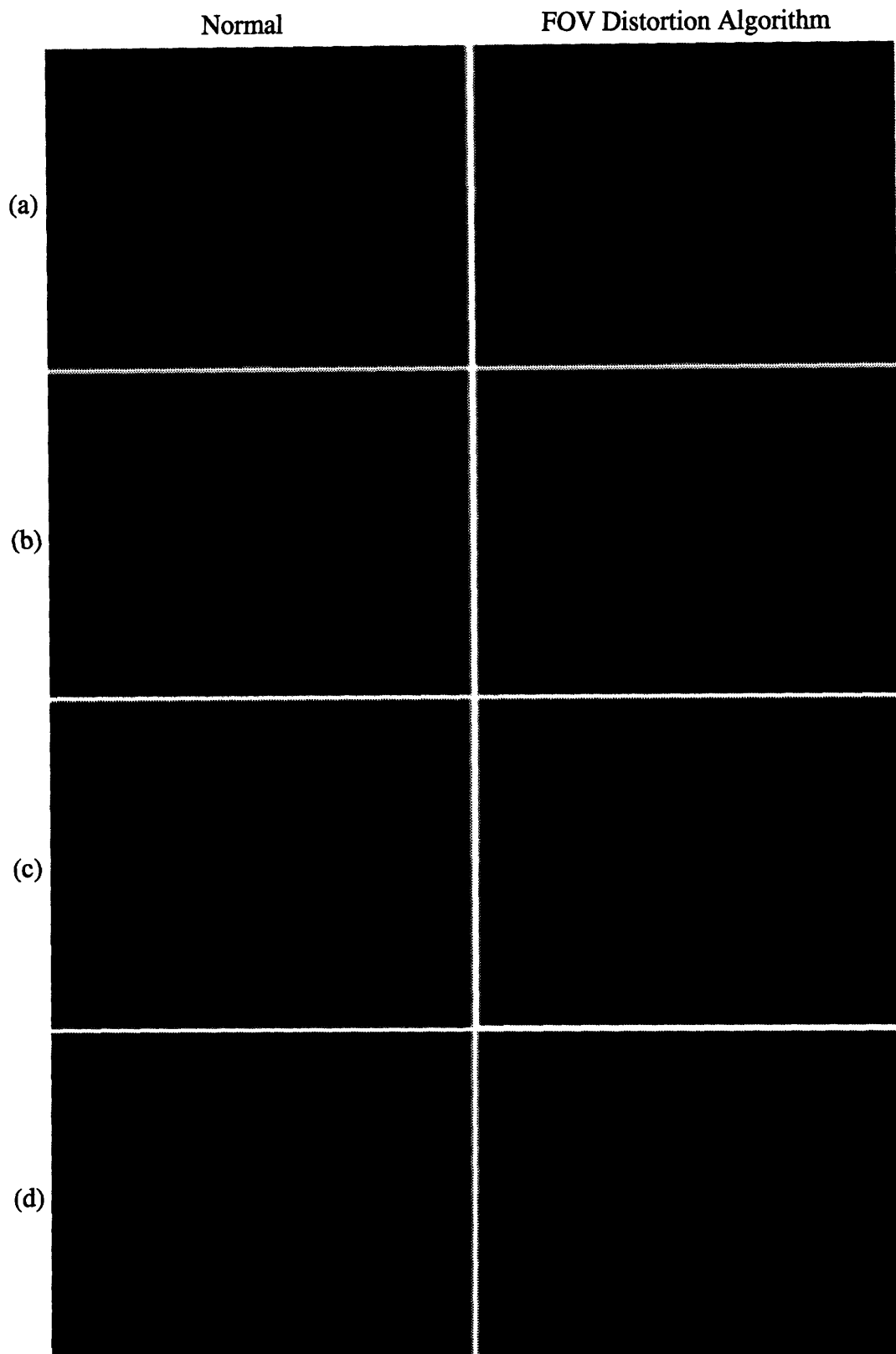


Figure 3.14: A comparison of the normal, unscaled buoy object to the buoy object altered by the FOV distortion algorithm. (a) At 125 feet. (b) At 1,100 feet. (c) At 3,000 feet. (d) at 13,000 feet.

The FOV distortion algorithm, unlike the size-scaling algorithm, distorts the entire scene, not just the buoy object. However, since the distortion of the surrounding scene is not critical to the visibility of the target object, it is ignored. The usefulness of this algorithm decreases when the appearance of other objects in the scene is also important. To solve the scene-warping problem, the FOV can be distorted only when the graphics engine is drawing certain objects. The location of the object, however, will still be distorted.

The clever reader will note that the pixel distributions of the size algorithm and FOV algorithms should be identical. That is, the perceived size of an object at any point is theoretically the same for both algorithms. However, the assumption that the two algorithms will result in identical performance is incorrect, since the complete scenes that result from the application of the algorithms are not the same. The two algorithms, while predicting identical object sizes, do not generate identical locations on the display surface. Figures 3.11 and 3.14, in part (d), show the locations of the buoy object at 13,000 feet for the different algorithms. The size scaling algorithm will push the object further to the horizon because it does not distort the linear perspective cues, while the FOV algorithm will manipulate both the size constancy cues and the linear perspective cues.

The solutions presented are sufficient since they provide an extended visible range to the observer while attempting to minimize depth judgment errors. In this manner, they improve the overall realism of the simulator and meet the task-specific requirements for visibility. However, this sufficiency is theoretical, and empirical evidence concerning the usefulness of these algorithms should be obtained before a full conclusion is reached.

3.2 Method

VE systems are prone to a variety of adjustment problems and other experimental noise (see section entitled Problems with Virtual Environments). Moreover, the association with a practical problem is necessary to provide a real, rather than academic, engineering

solution. The experimental method described below determines the effectiveness of algorithms for improving visibility.

3.2.1 Subjects

Six students at the Massachusetts Institute of Technology participated in the study. Subjects were required to fill out forms in compliance with the Committee on the Use of Humans as Experimental Subjects and were paid for their time. The subjects ranged in age from 17 to 22. Half the individuals had vision corrected by contact lenses, the other half had normal or nearly-normal vision. None of the subjects had any prior experience navigating boats or with other maritime activities which might give them knowledge of buoy location and identification. Three subjects were male and three subjects were female.

3.2.2 Apparatus

The visual stimuli for the task were presented using a Silicon Graphics Onyx (with a RealityEngine² graphics board). Software for generating the stimuli was developed using the Performer Library from SGI. Data-collection and experiment-control programs were developed in C.

The graphics were shown using a Virtual Research VR4 head-mounted display. The spatial resolution for each eye is given in product literature as 742 pixels by 230 pixels (Virtual Research, 1995). However, the HMD took as input an NTSC composite video signal with a resolution of 486 pixels by 648 pixels. As noted earlier, sometimes the frame buffer clips the image (Rolland et al., 1995), and this was tested empirically for the VR4 HMD. The HMD was revealed to be capable of displaying 486 pixels by 646 pixels; therefore, this was the resolution used in subsequent models. The displays measured 1.3 inches across the diagonal which means that a pixel subtended $.074^\circ$ of both horizontal and vertical visual angle.

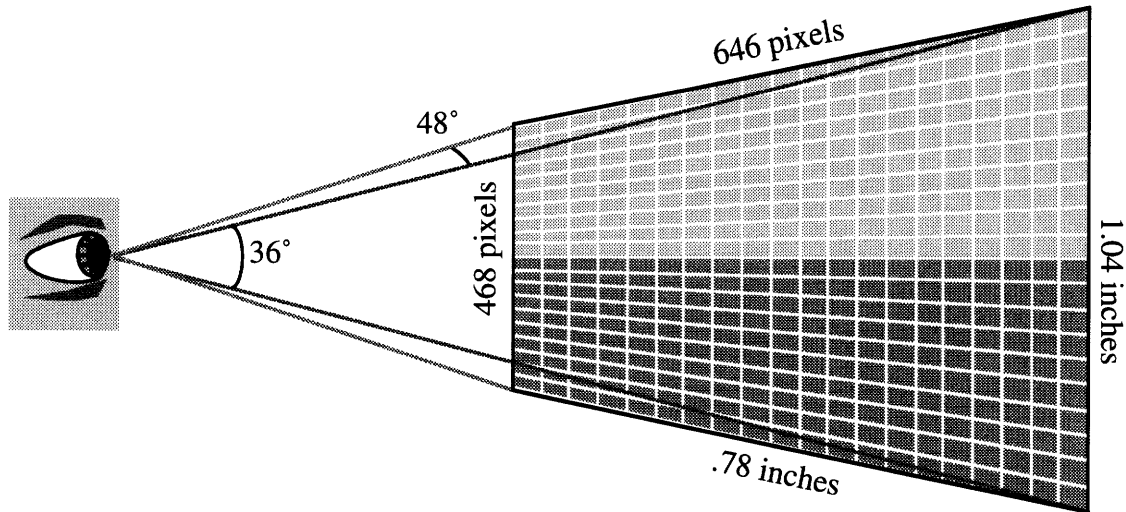


Figure 3.15: The characteristics of the HMD display used in the experiment.

Responses were collected using the BBN Hark Voice Recognition System. The system used a "press to talk" button and a Sennheiser microphone headset. A simple grammar was written to recognize spoken numbers from 1 to 99,999 (see Appendix A). The Hark system ran on a Silicon Graphics Indy computer. Responses were sent to data recording software running on the Onyx via typical ethernet connections. In addition, audio feedback was given to the subject via the Sennheiser headset and an identical headset mounted on the HMD. Audio feedback consisted of the playback of a recorded message asking for a repeat of a response that had confused the recognition system. A low-level test of the HARK system resulted in an average accuracy rate of 97% on a fairly simple grammar (Pioch, 1995). This was considered more than suitable for the needs of the experiment.

For an in-depth treatment of the design and implementation of the VETT core testbed hardware and software systems, the reader should consult Zeltzer et al. (1994).

3.2.3 Design

The primary experimental goal was to assess two visibility-enhancing algorithms in terms of human perceptual performance. The usefulness of these algorithms in improving the training of depth perception is also investigated.

The experiment design was influenced by two overall factors: performing experiments on far-field visibility and accommodating the constraints of the OOD simulation. The selection of only a part-task of the OOD navigation task allowed for careful simplification of problems of immersion and simulator fidelity to problems that could be resolved experimentally. The environment and target stimulus were presented in such a way as to be consistent with the OOD simulator. However, in the interest of reducing experimental noise, the OOD models were not followed precisely.

For instance, the landmasses and clouds were removed from the scene. This was done to avoid the introduction of conflicting depth cues. In addition, the buoy-object's position at a far distance would place some of the pixels next to that of a landmass, thus presenting a different background color. Having the background color at far distance be inconsistent could cause a different perceptions of depth at the same distance and was thus unacceptable. Furthermore, the color of the background could also influence target detection. The clouds were flat textures mapped into the sky. Because it was unknown how the clouds would be interpreted in depth, they were classified as noise and removed.

In addition, all other buoys in the model of the channel were removed, as were range markers and turning beacons. The presence of these other features would clearly influence the perception of depth of the target object. The submarine model that normally would be visible in a normal forward view was also removed. The remaining scene consisted of a flat plane that had a water texture mapped onto it and a sky that is lit from an overhead light that approximated the sun.



Figure 3.16: A number of elements were removed from the OOD simulation shown in (a) to get a noise-free environment for experimentation. (b) shows a sample scene from the experiment.

The target object was reduced to a frustum from a more complicated model. This was done to ensure that the underlying graphics software would only have six polygons to interpret and display rather than the twenty in the original model.

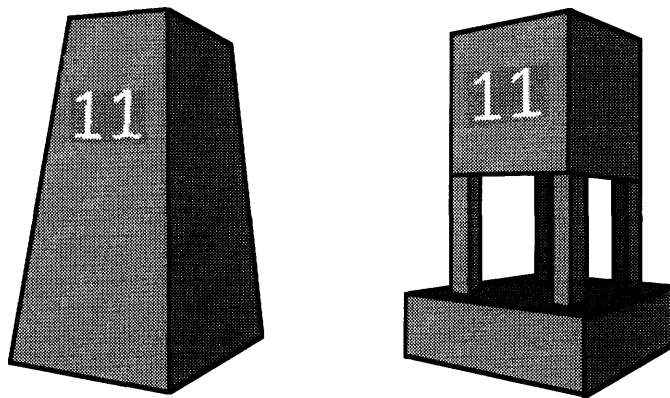


Figure 3.17: The buoy model used in the experiment (left) and the buoy model used in the OOD simulator (right).

The behavior of the graphics package at a far distance was unknown with respect to handling multiple polygons. If the more complex model is assumed, and the object has

been reduced to one pixel, we can hypothesize that a case exists where the model with fewer polygons will be visible and the other will not.

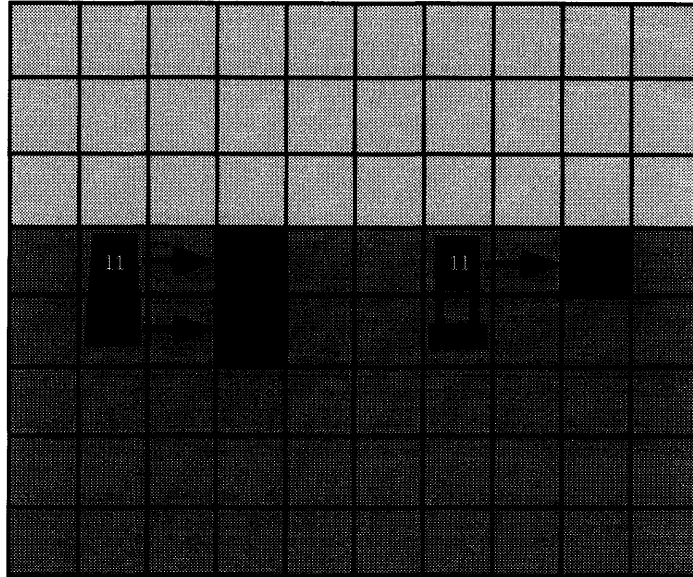


Figure 3.18: The predicted effect of using the simpler buoy model. The simple model (left), translates to two pixels because it fills more than half of the two pixels it covers, while the original OOD model (right), is displayed as only one pixel because of its shape.

Obviously, the advantage of using the model with fewer polygons outweighs the usefulness of adhering to the exact model used in the OOD simulator. Moreover, since the new model does not differ from the previous one in its dimensions and proportions, the geometry discussed above does not change

The target object was chosen to be red. In the OOD simulator, buoys are red, green and yellow, and have small white numbers labeling them. The decals were preserved since they did not interfere with the color of the object beyond a certain distance. Color, however, was determined to be a significant factor in determining depth in a pilot experiment. Thus, to reduce complexity, a single color was chosen.

Finally, the experiment differed from the simulator in that the point of view was fixed so that the buoy was directly straight ahead. The viewpoint was not based upon feedback from a head position tracker. Motion of viewpoint during acuity and target

detection tests significantly reduces accuracy (Boff & Lincoln, 1988). In addition, motion sickness associated with tracked-head motion was avoided (Kennedy et al., 1992).

The direction (heading) of the viewpoint was chosen randomly after pilot experimentation showed a significant effect of direction on accuracy. The water texture provides a important depth cue. Failure to randomize direction could result in the use of the texture as the main depth cue, rather than the size and shape of the object.

In order to fully address the issue of performance in the OOD simulator, the effects of training had to be considered. That is, since the OOD simulator was designed to train individuals at a task, performance of a sub-task in the OOD model should also be considered as a training task. Thus, the experimental goal was not only to assess the visibility-extending algorithms in terms of human perceptual performance, but also to examine their usefulness in improving the training of depth perception.

The subjects' task was to estimate the distance of the target object in feet. The units of measurement were chosen so that the subjects (hereafter: Ss) could give a sufficiently fine-grained response. Also, the units were influential in the accuracy scores on the initial assessment trial, since some transfer effects from real world-based expectations were observed. That is, the different subjects would form preconceptions based upon the units about how to judge depth in the experiment.

The target object was presented according to the perspective geometry discussed in detail above. Two identical control conditions and the two algorithms described above at two different distance thresholds (10,560 feet and 13,160 feet) constituted the six experimental conditions. Each S started on a different condition. The following Latin square was used to remove place-in-order effects:

Subject	Day						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	
1	CONTROL 5280	Size 10560	FOV 10560	CONTROL 5280	Size 13160	FOV 13160	Algorithm Visibility Dist.
2	Size 10560	FOV 10560	CONTROL 5280	Size 13160	FOV 13160	CONTROL 5280	Algorithm Visibility Dist.
3	FOV 10560	CONTROL 5280	Size 13160	FOV 13160	CONTROL 5280	Size 10560	Algorithm Visibility Dist.
4	CONTROL 5280	FOV 13160	Size 13160	CONTROL 5280	FOV 10560	Size 10560	Algorithm Visibility Dist.
5	FOV 13160	Size 13160	CONTROL 5280	FOV 10560	Size 10560	CONTROL 5280	Algorithm Visibility Dist.
6	Size 13160	CONTROL 5280	FOV 10560	Size 10560	CONTROL 5280	FOV 13160	Algorithm Visibility Dist.

Table 3.1: The Latin square distribution of conditions, days, and subjects used in the experiment.

Since six subjects were used, order effects between which algorithm was used could be counterbalanced. Algorithm-used was chosen over visibility distance since the interaction effect of the algorithms was deemed more important.

Two dependent variables were recorded. The first consisted of the verbal report received from the subject and ranged in value from 1 to 99,999, or -1, if the subject replied "I can't see that." The other dependent variable recorded was reaction time (RT). RT was measured as the time from the display of the stimulus to the receipt of the reply from the Hark system. The RT measurement did not subtract the time needed to speak different responses. That is, the amount of time needed to speak, "thirteen thousand, one hundred and twenty-four" is larger than the time need to say, "forty," and this discrepancy was unaccounted for.

The main measure of performance in the experiment was the deviation between the Ss' responses and the distance presented according to the perspective geometry. This

difference is referred to as "absolute error," which is not to be confused with "standard error" in later statistical calculations.

$$\text{absolute error} = |\text{response} - \text{distance presented}|$$

Distances were chosen from within ranges of depth. The depth ranges were selected so that thirty distances picked from within thirty depth ranges would constitute a set of trials. Depth ranges were chosen such that they would range from the near visibility point (calculated previously to be 123 feet) to just beyond 13,160 feet (2.5 miles). This was done to ensure that an equal number of trials would be presented for each experimental condition.

Distributing the depth "buckets" linearly across these distances made little sense. The geometrical models for the OOD simulator predict that the target object will be displayed at the same number of pixels over certain ranges, ranges whose size at the furthest distance is far larger than a linearly-chosen depth bucket. This implies that buckets at a distance should be larger. In addition, the number of invisible trials in the control and closer visibility point conditions were minimized to increase the number of useful data points.

Depth Range	From	To	Depth Range	From	To
1	120.00 ft.	431.25 ft.	16	4909.83 ft.	5256.33 ft.
2	431.25 ft.	742.82 ft.	17	5256.33 ft.	5612.04 ft.
3	742.82 ft.	1054.78 ft.	18	5612.04 ft.	5979.27 ft.
4	1054.78 ft.	1367.24 ft.	19	5979.27 ft.	6360.94 ft.
5	1367.24 ft.	1680.32 ft.	20	6360.94 ft.	6760.67 ft.
6	1680.32 ft.	1994.17 ft.	21	6760.67 ft.	7183.02 ft.
7	1994.17 ft.	2308.00 ft.	22	7183.02 ft.	7633.70 ft.
8	2308.00 ft.	2625.04 ft.	23	7633.70 ft.	8119.86 ft.
9	2625.04 ft.	2942.61 ft.	24	8119.86 ft.	8650.42 ft.
10	2942.61 ft.	3262.08 ft.	25	8650.42 ft.	9236.60 ft.
11	3262.08 ft.	3583.94 ft.	26	9236.60 ft.	9892.40 ft.
12	3583.94 ft.	3908.79 ft.	27	9892.40 ft.	10635.39 ft.
13	3908.79 ft.	4237.39 ft.	28	10635.39 ft.	11487.55 ft.
14	4237.39 ft.	4570.67 ft.	29	11487.55 ft.	12476.39 ft.
15	4570.67 ft.	4909.83 ft.	30	12476.39 ft.	13636.39 ft.

Table 3.2: The ranges of depth from which distances were selected in the experiment.

The depth buckets were determined according to the following equation:

number of trials = 30 trials

max increment = 850 feet

min increment = 300 feet

$$D_0 = 120 \text{ feet}$$

$$D_n = D_{n-1} + e^{\left(n \cdot \frac{\log(\text{max increment})}{\text{number of trials}} \right)} + \text{min increment} \quad (23)$$

This formula increases the depth bucket size at the larger viewing distances to account for the expectation of decreased depth acuity at those ranges. In addition, this calculation accounts for the desire to collect relatively similar numbers of data points across distances. Too many points presented in the far range would overtrain on those points,

while too few would not yield enough data to determine the effects of the algorithms at a far distance.

Trials in which the stimulus was not visible were presented in order to balance the total number of trials per condition. Keeping the number of trials constant and presenting only visible trials in a particular condition introduced the problem of training depth estimation on one condition more effectively on a particular range. Varying the number of trials and keeping the size of the depth ranges constant presented problems with the time needed to complete the trial and the amount of training for each condition. Presenting invisible trials seemed to be the best solution, even though the display of invisible trials between visible trials could interfere with the training of distance estimation. The issue of the number of trials is discussed in detail below.

The methods presented above represent a significant attempt at reducing noise inherent in the simulator. Certain problems were unavoidable (such as the color distortion in the HMD), but others were minimized or eliminated. Unfortunately, a major characteristic of experimentation in VEs is the difficulty of properly eliminating confounding factors.

3.2.4 Procedure

Subjects were solicited via ads sent out to electronic mailing lists. In addition, only Ss that could perform the experiment on six consecutive days were selected. Ss were scheduled to run over a seven-day period (four subjects on Days 1 through 6, 2 subjects on Days 2 through 7). Arranging times was a difficult task, but Ss were scheduled to do the experiment on the same time every day when possible.

Upon arrival on the first day, Ss completed forms in compliance with the Committee on the Use of Humans as Experimental Subjects. In addition, a set of instructions was presented (see Appendix B), and questionnaires on marine experience were completed. Finally, Ss filled out paperwork detailing their subjective physiological

state (e.g. did they feel nauseous, light-headed, etc.). On subsequent days, only a short set of instructions and the physiological surveys were required. The experimenter examined the responses regarding physiological state to see if there were any conditions that may interfere with the S's well-being during experimentation. Then, Ss were asked if they had any questions about what they were asked to do; short clarifications would be given if required.

Notably, Ss were asked to make fine-grained responses. Pilot tests showed that Ss had a strong tendency to estimate distance rather than guess. Performance improved when Ss made finer-grained responses (i.e. "4,435" vs. "4,500"). Therefore, Ss were encouraged to use more digits in their estimation (see Appendix B).

Because the HMD eliminated all vision except for that inside the helmet, recording Ss' responses became an issue. A keyboard could not be used to enter responses since it would require some typing training and mistyped responses would be difficult to catch. Instead, the Hark voice recognition system was chosen for its reliability and its speaker-independent recognition. The Hark system has a recognition rate that can approach 100%. Unfortunately, in practice, the recognition rate is about 95%. For an experiment with 17,280 total data points, this could represent a loss of 618 data points, clearly not an acceptable condition. With practice, however, individuals can become accustomed to the system and achieve much higher hit rates. In addition, having the Ss speak their responses allowed the experimenter to easily monitor the hit rate and make corrections as needed.

Therefore, after completing the paperwork, Ss donned the Sennheiser headset-microphone to perform a simple training regimen on the Hark voice recognition system. A number from 1 to 99,999 or the phrase "I can't see that" was presented on the screen of the workstation. The Ss would press the "push to talk" button and speak their response. If a response was not recognized, the subjects were informed of possible problems via information printed onscreen and a replay of the recorded verbal request, "Could you repeat that?" The experimenter was on hand throughout the process to monitor the S and to

provide assistance in case of any difficulty. In the case of a misrecognized response, the experimenter would record the trial and the correct response and later correct the data to reflect the given response. In this manner, near 100% accuracy of response recognition could be achieved. After the first day, Hark training was reduced to a much shorter set of trials.

Once the Hark training was complete, Ss were seated comfortably. The HMD had a foam seal that prevented most external light from interfering with vision. The experiments were conducted in a room with no windows and lighting was reduced to a single computer screen (approximately 12 cd/m²) used for experiment control and data collection monitoring.

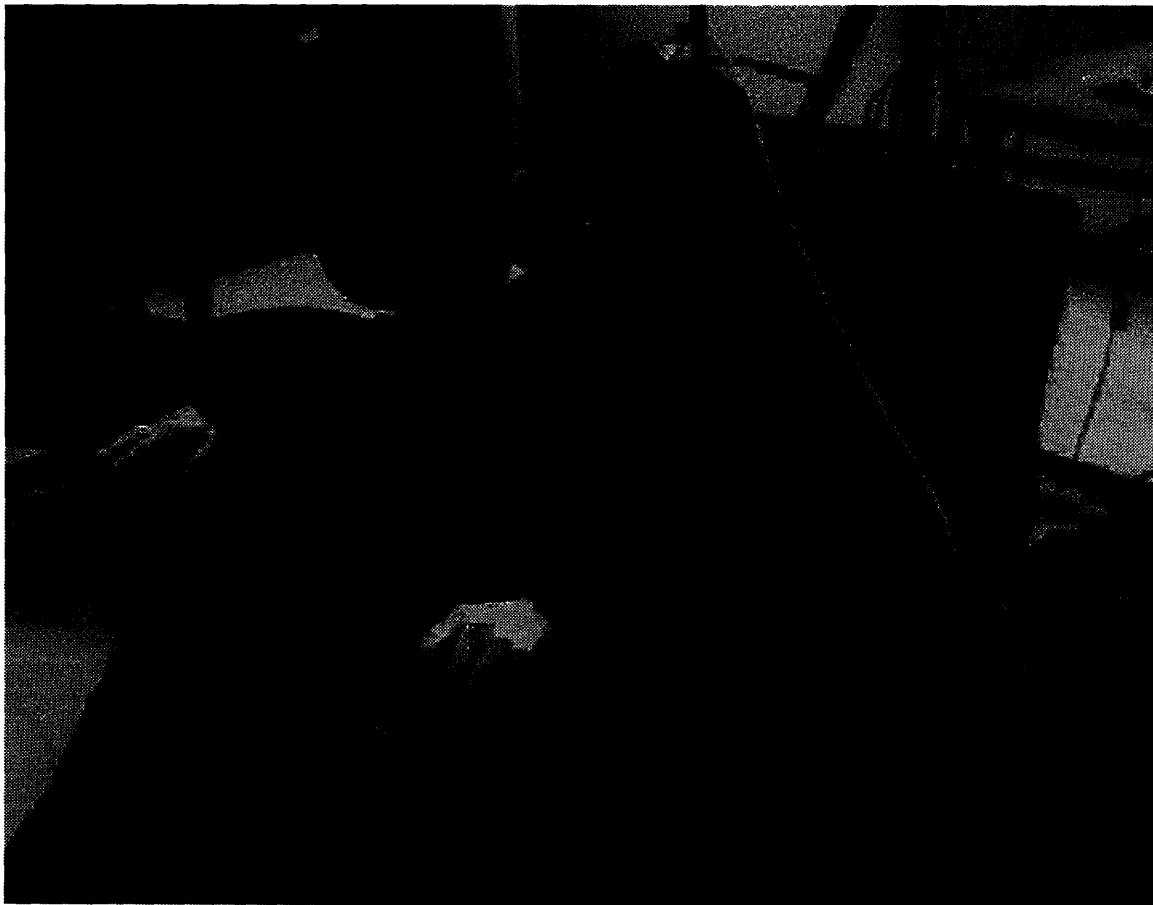


Figure 3.19: The experiment station. A subject sits comfortably, wearing the HMD and holding the "push to talk" button for the voice recognition system.

Ss were encouraged to keep their head positioned straight ahead so as to have the displayed horizon match what would be expected in the real world. Following the arrangement of the S at the experiment station, verbal instructions on the adjustment of the HMD were given. A test pattern was displayed during this adjustment so as to ensure a proper fit. The subject adjusted the headstraps, the interocular distance, and the eye relief of the HMD to obtain the clearest picture. Again, verbal clarifications on the experimental procedure were offered.

The stimulus was presented in two different ways. Feedback trials displayed the stimulus until a response was given then showed a number indicating the correct depth of the target object over top of the scene. The correct response was displayed for 1.25 seconds. On assessment trials, the correct distance was not displayed. Both feedback and assessment conditions distinguished separate trials by a blank screen shown for .75 sec.

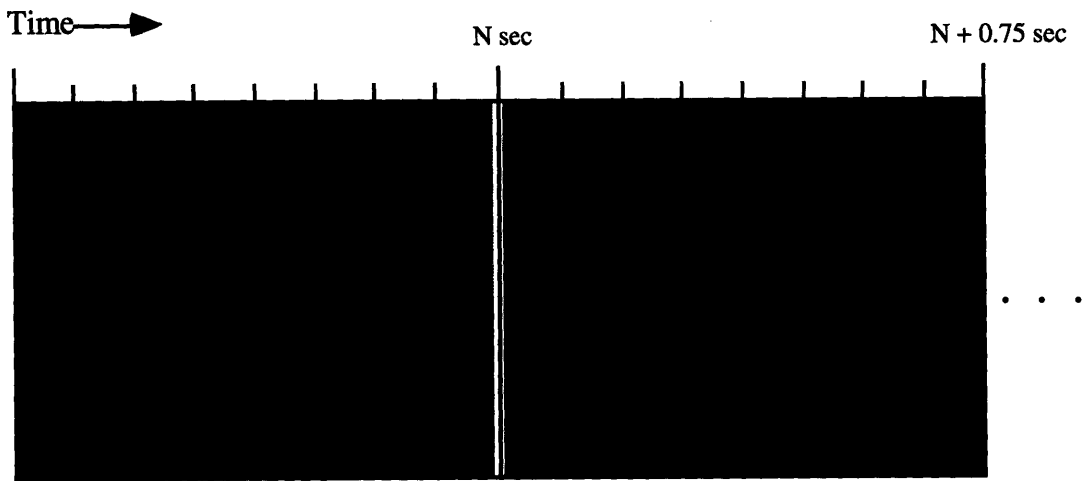


Figure 3.20: The timing of a typical assessment trial. After the stimulus was displayed and the subject gave a response at time N , the screen was blanked.

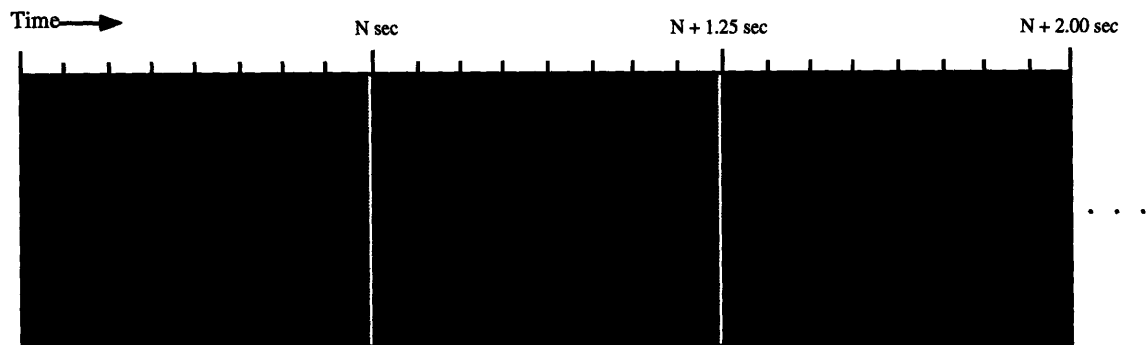


Figure 3.21: The timing of a typical feedback trial. After the stimulus was displayed and the subject gave a response at time N , the correct answer was displayed. Then, the screen was blanked.

A set of assessment trials was followed by four sets of feedback trials. Every four sets of trials, Ss received a break so as to reduce mental fatigue from repeating the task, physical fatigue from supporting the weight of the HMD, and visual fatigue from the optics of the display.

#	Type	# of Trials	Time
1	Assessment	30 trials	
2	Feedback	30 trials	
3	Feedback	30 trials	
4	Feedback	30 trials	
Break			2-3 min.
5	Feedback	30 trials	
6	Assessment	30 trials	
7	Feedback	30 trials	
8	Feedback	30 trials	
Break			10-12 min.
9	Feedback	30 trials	
10	Feedback	30 trials	
11	Assessment	30 trials	
12	Feedback	30 trials	
Break			2-3 min.
13	Feedback	30 trials	
14	Feedback	30 trials	
15	Feedback	30 trials	
16	Assessment	30 trials	

Figure 3.22: The ordering of breaks and trials for one subject during a typical day's run.

The first and third breaks were approximately three minutes long. During these breaks the Ss were told to remove the HMD but to remain seated. The midway break was ten to fifteen minutes long and Ss were encouraged to walk around outside the lab. After a break, the test pattern would be displayed and the Ss would readjust the HMD. Upon completion of that day's trials, Ss again filled out a physiological state form and were paid.

3.3 Results

On a given day of the experiment, a subject performed 16 sets of trials, 4 assessment and 12 feedback (see Figure 3.24), for a total of 480 data points. Over the six days of the experiment, each S performed a total of 2880 trials, some of which were discarded because of noise. Noise included skipped trials caused by improper use of the voice recognition system. A pause during the enunciation of a reply could confuse the system into thinking two replies had been given (i.e. "four thousand [pause] two hundred" was recognized as 4,000 and 200, not 4,200). Thus, the two were appended (by the experimenter) and the second trial discarded. Of a total of 17,280 points, 87 were discarded because of skips.

The performance of the subjects was only evaluated in the assessment sets of trials. Only trials where the S responded with an estimate of depth were considered. Those trials on which the S replied, "I can't see that" were discarded. Using these criterion, a total of 3,223 data points were considered. The data from a typical set of assessment trials is shown in Figure 3.23.

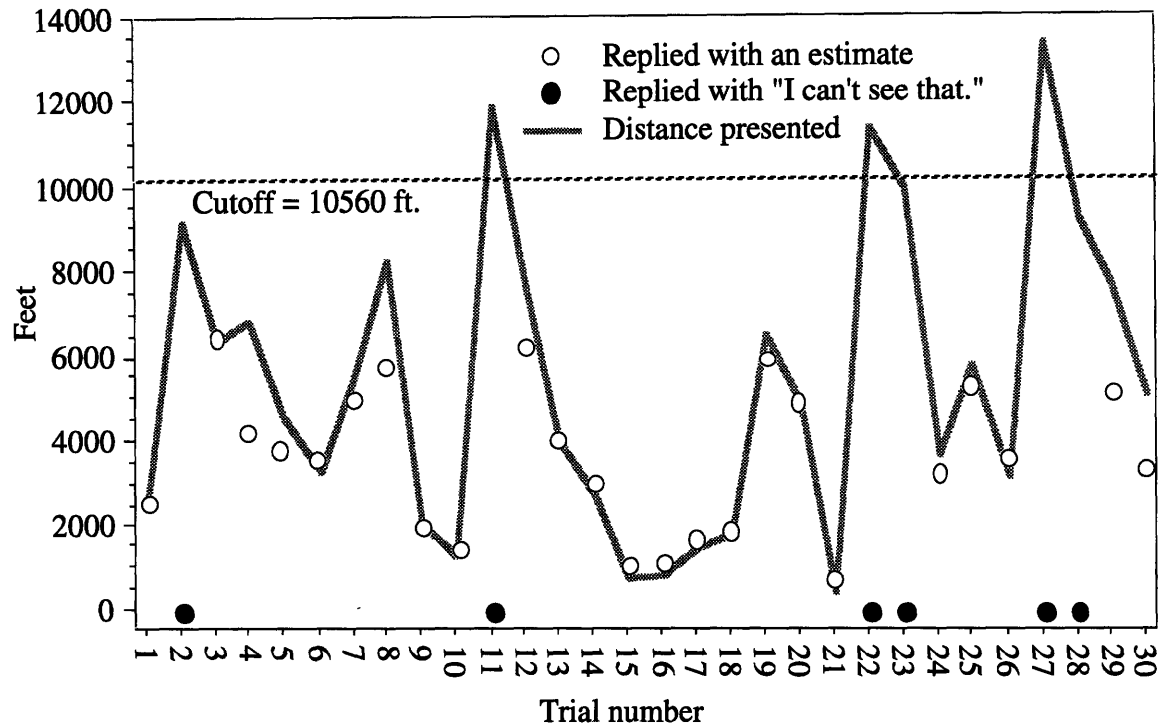


Figure 3.23: A typical assessment trial set. The condition represented here was the size scaling algorithm with a cutoff distance of 10,560 feet. The circles at 0 feet represent cases where the subject replied that he or she could not see the buoy object.

The control condition included 4 sets of assessment trials for each subject and was used on two separate days for a grand total of 1,440 data points. However, since about half of the distances displayed were invisible, the actual number of data points was 760. The following table provides a gross summary of the performance of the Ss on the control condition:

Subject	# of Trials	Error (feet)			RT (seconds)		
		Mean	Std. Dev.	Std. Err.	Mean	Std. Dev.	Std. Err.
1	128	551.325	665.945	74.219	8.301	5.610	.496
2	125	436.768	406.238	36.335	5.553	1.262	.113
3	124	604.566	777.809	116.322	5.867	1.218	.109
4	126	713.738	766.160	69.616	5.272	2.200	.196
5	130	417.240	478.764	285.208	5.865	.586	.051
6	127	806.063	1216.155	107.603	5.154	.542	.048
Mean over Subjects		588.283	718.512	114.884	6.002	1.903	.169

Table 3.3: The mean, standard deviation, and standard error for RT and absolute error for the control condition.

Again, the RT measure does not account for the time it took an S to speak a reply, thus those numbers should be considered to be much noisier than the error measure. Assuming that the data is normally distributed over subjects and that the subjects form a representative sample of the population, we can analyze the significance of population variance on performance in the assessment trials.

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Subject	5	70685841.130	14137168.226	7.447	<.0001
Residual	3217	6106843431.404	1898303.833		

Table 3.4: The by-subject data compiled for an analysis of variance. Subject is the independent variable and error is the dependent variable in the calculation.

According to the ANOVA presented in Table 3.4, the difference in performance between subjects over all visible assessment trials was statistically significant, $F = 7.447$, $p < .0001$

The performance of the subjects varied with the distance presented. Performance was assessed both by the mean error and the standard deviation. Absolute mean error corresponds to the effect of bias and standard deviation corresponds to the effect of resolution. By examining the effect of distance on both mean error and its standard

deviation in the control case, we can compare the effects of the visibility-enhancing algorithms on the bias and accuracy of depth estimation.

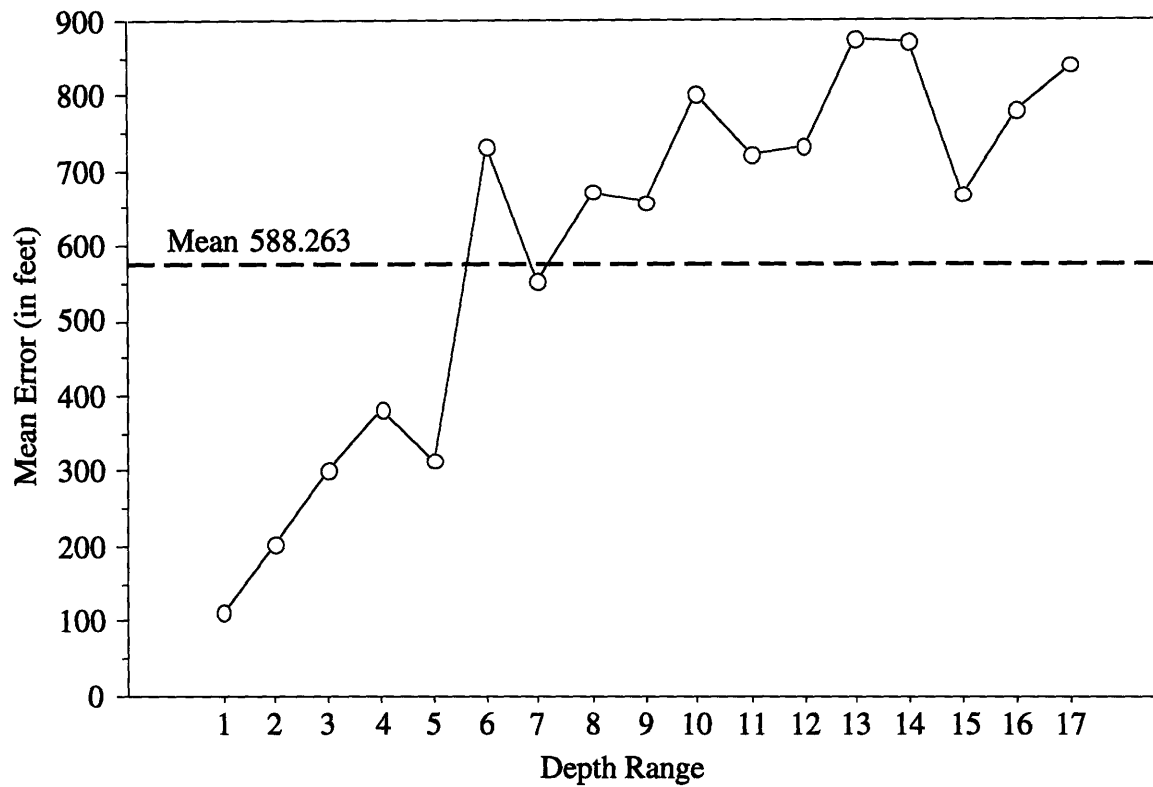


Figure 3.24: A plot of error (the absolute deviation between the presented distance and the subjects' response) versus depth range for the control condition. The error increased as a function of distance.

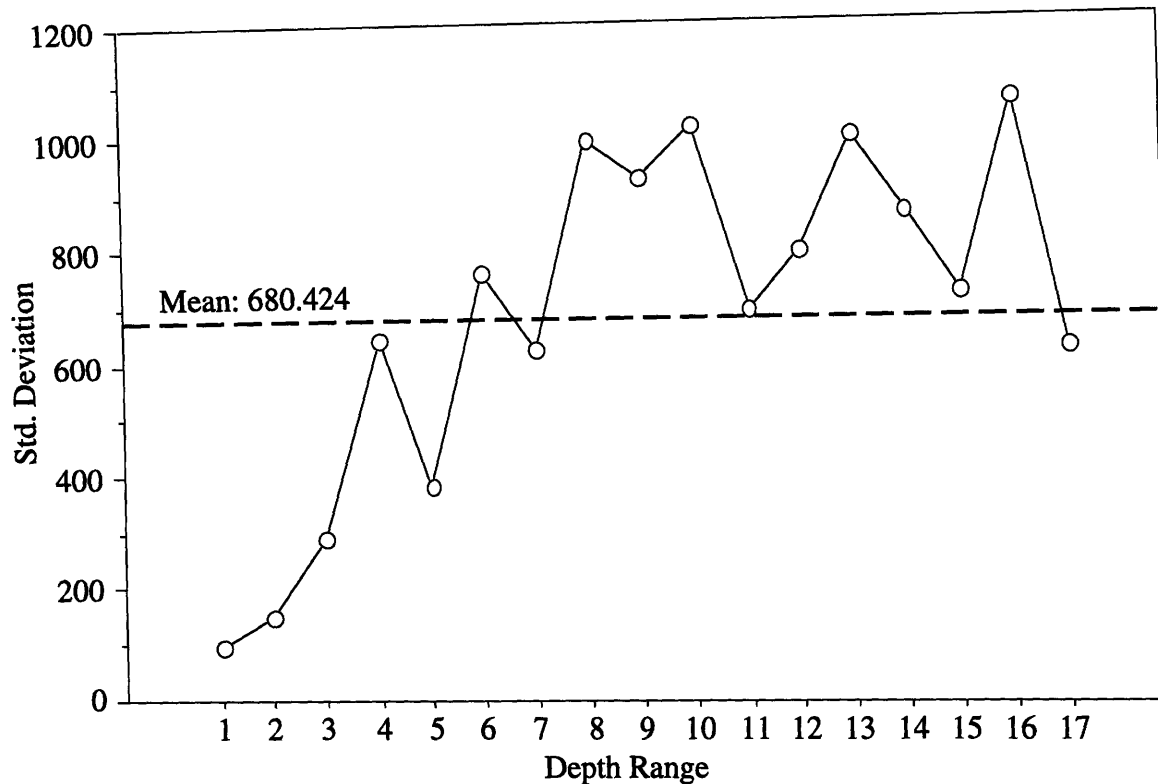


Figure 3.25: A graph of the standard deviation of mean error (averaged over subjects) versus depth range for the control condition. The variance of the responses increased as a function of distance.

Figures 3.25 and 3.26 show the effect of distance on absolute average error and accuracy in the control case. In the control condition, only images that were presented in depth ranges 1 to 17 (distances of 123 feet to 5612 feet) were visible. The cases utilizing the visibility-extending algorithms covered depth ranges 1 to 27 (with a cutoff distance of 10560 feet) and in depth ranges 1 to 30 (with a cutoff distance of 13160 feet).

We can see the results of changing the cutoff distance used in each algorithm by plotting the mean error and the standard deviation of error as a function of distance.

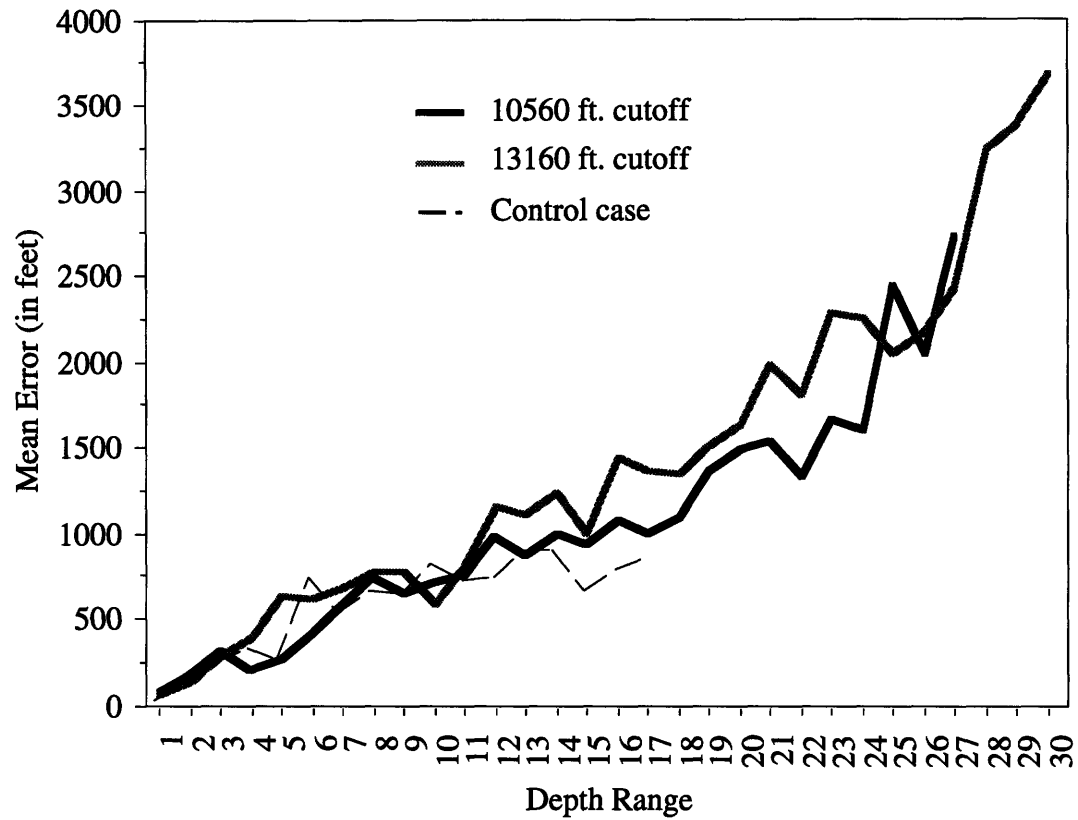


Figure 3.26: A plot of mean error as a function of depth range for the various cutoff distances. The control case has fairly equivalent accuracy to the algorithm-enhanced cases over the ranges it is visible (1 to 17). The 10,560 ft. case cuts off at depth range 27.

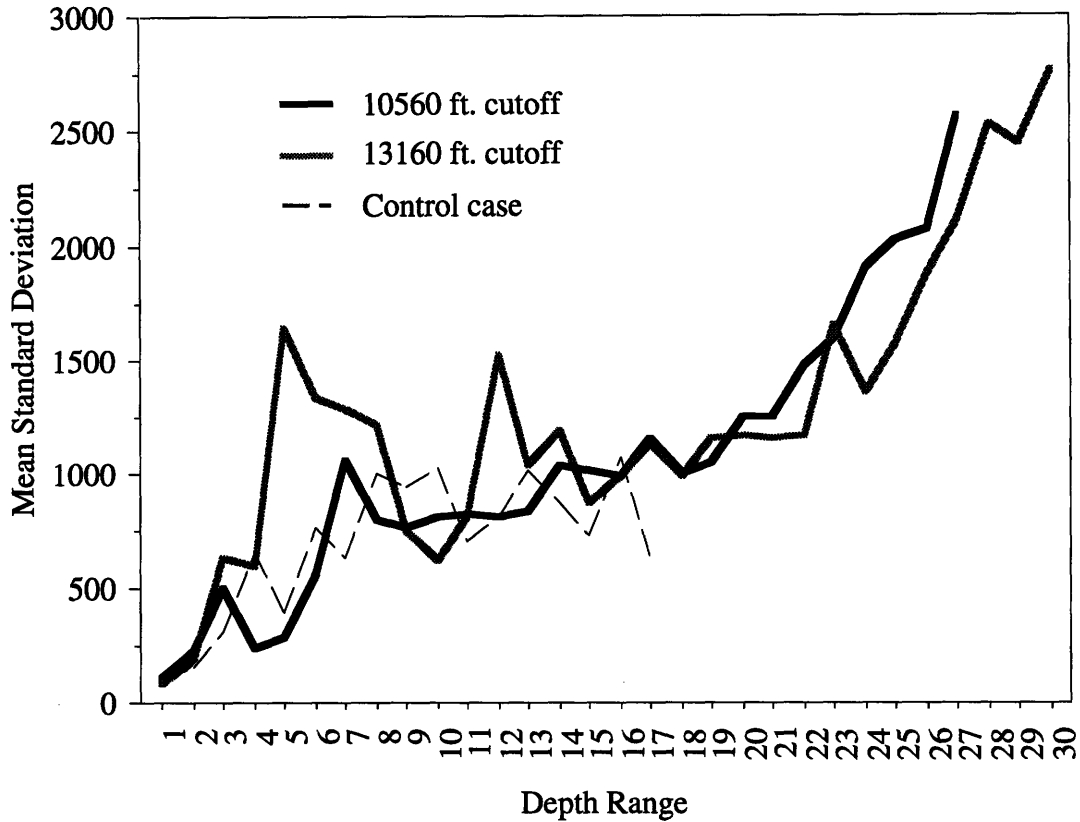


Figure 3.27: A plot of mean error as a function of depth range for the various cutoff distances. The control case has, as expected, better accuracy over the ranges it is visible (1 to 17). The 10,560 ft. condition is last visible in depth range 27.

Figure 3.26 shows that the mean error of the responses continues to increase with distance even in the extended visibility cases. Furthermore, the accuracy of the responses decreased with distance. The size scaling algorithm and the FOV distortion algorithm did not yield identical performance results.

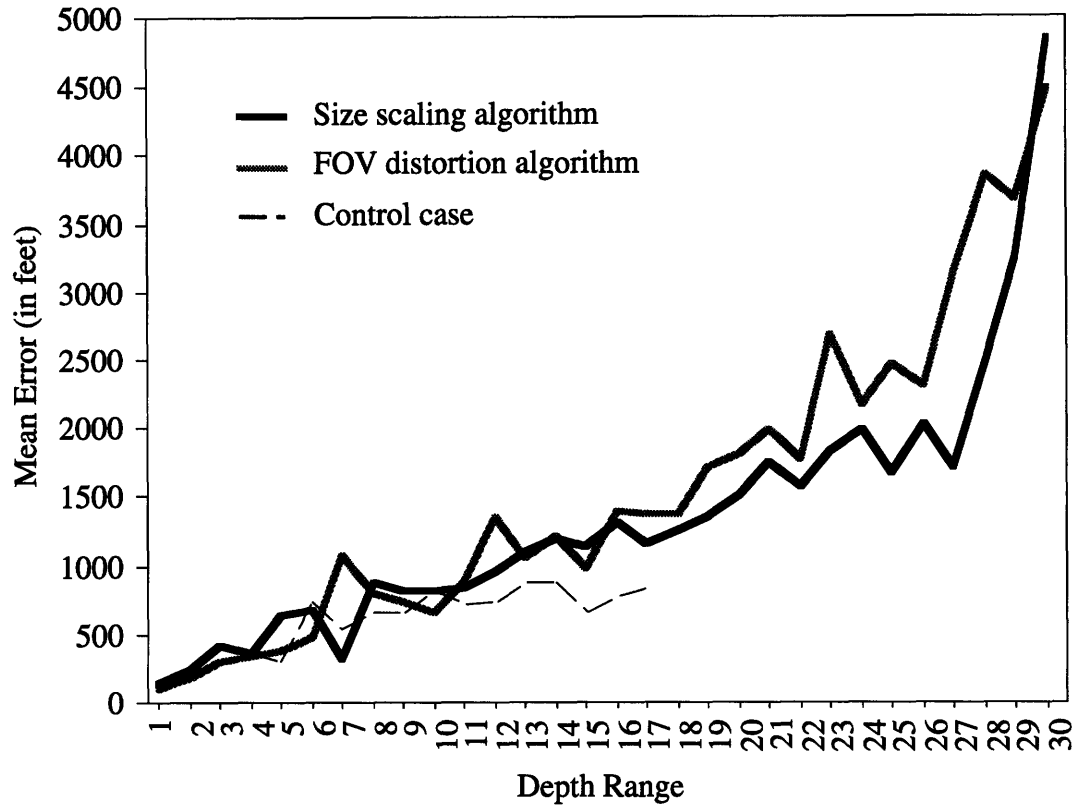


Figure 3.28: A plot of mean error as a function of depth range for the two algorithm conditions. The control case is only visible over depth ranges 1 to 17 (123 feet to 5612 feet). The FOV distortion algorithm is better than the size scaling algorithm for all depth ranges from 16 to 29.

The standard deviation of the of the error (Figure 3.29) also increased with distance, although it varied more in the near ranges.

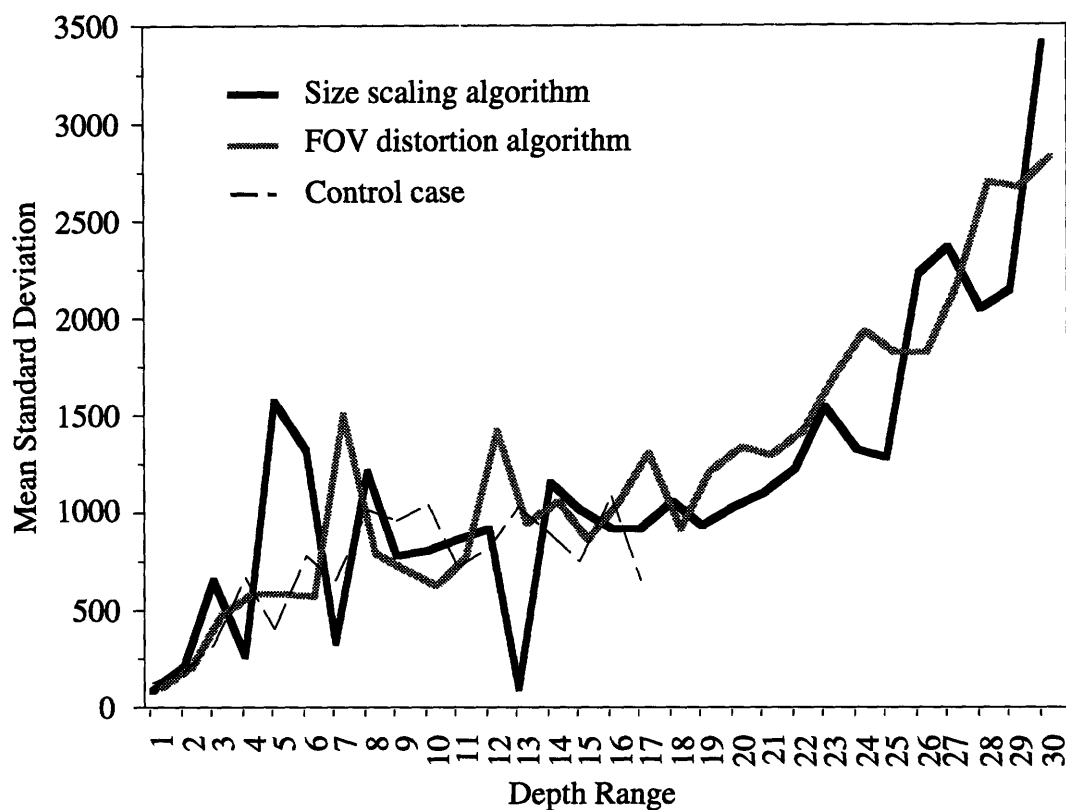


Figure 3.29: A graph of the variance in mean error as a function of depth range for the two algorithm cases. Mean standard deviation is calculated as the average for all six subjects. The control case is only visible over depth ranges 1 to 17 (123 feet to 5612 feet).

To summarize the effect of the different algorithms and cutoff distances, the mean error and standard deviation for all conditions and subjects were compiled. To properly compare the different conditions, only data from comparable distances can be evaluated. This complicates the analysis, but permits a more fine-grained investigation of the results. Table 3.5 summarizes the by-subject mean error.

Error		Control	SizeScaling		FOVDistortion		
Subject	# of Trials		at 10560 ft. at13160 ft.		at 10560 ft.	at13160 ft.	
		Mean	Mean	Mean	Mean	Mean	
Depth ranges 1 to 17	1	398	551.325	503.132	439.779	476.485	644.809
	2	379	436.768	1032.115	564.132	638.284	594.269
	3	388	604.566	533.682	558.631	1336.045	968.221
	4	395	713.738	664.294	783.299	881.074	665.424
	5	399	417.240	417.632	557.090	412.382	1088.716
	6	387	806.063	812.221	1966.930	921.147	709.882
	Mean over Subjects		588.283	660.5125	811.644	777.570	778.554
Depth ranges 1 to 27	1	410		858.887	757.282	782.779	1093.047
	2	342		1034.696	881.386	880.729	1114.020
	3	407		834.263	916.748	1924.206	1367.121
	4	405		938.716	1178.717	1214.583	1290.960
	5	413		647.163	955.368	815.686	1582.913
	6	386		997.032	2004.790	1485.760	1093.689
	Mean over Subjects			928.916	1115.715	1183.957	1256.958
Depth ranges 1 to 30	1	226			873.685		1263.835
	2	200			1008.633		1480.573
	3	225			1003.981		1609.975
	4	227			1297.817		1627.884
	5	230			1017.414		1752.395
	6	200			2007.059		1210.426
	Mean over Subjects				1201.432		1490.848

Table 3.5: The mean error (given in feet) for the various conditions and subjects.

The control algorithm showed better performance than either algorithm-enhanced case over the depth ranges 1 to 17. The average absolute error on the cases with a 10,560 foot cutoff distance was better than in the cases with a 13,160 foot cutoff. Also, the size scaling algorithm resulted in better performance than the FOV algorithm.

Before performing significance testing on these results, the accuracy for the various conditions should be examined.

Error		Control	SizeScaling		FOVDistortion	
Subject	# of Trials		at 10560 ft.	at13160 ft.	at 10560 ft.	at13160 ft.
		Std. Dev.	Std. Dev.	Std. Dev.	Std. Dev.	Std. Dev.
Depth ranges 1 to 17	1	398	665.945	555.458 480.133	661.006	985.640
	2	379	406.238	929.707 641.730	605.904	569.798
	3	388	777.809	498.501 629.579	1390.687	1169.171
	4	395	766.160	670.105 795.425	953.861	971.590
	5	399	478.764	447.301 778.636	468.180	1157.711
	6	387	1216.155	946.654 2232.205	961.117	898.535
	Mean over Subjects		718.512	674.621 926.285	840.126	958.741
Depth ranges 1 to 27	1	410		991.638 868.673	894.893	1206.950
	2	342		867.744 1064.285	871.672	1136.719
	3	407		1366.266 1008.128	2094.427	1361.698
	4	405		1002.534 1211.235	1258.139	1444.783
	5	413		646.843 1061.563	965.256	1740.283
	6	386		1069.881 2039.271	1462.575	1202.259
	Mean over Subjects			990.818 1208.859	1257.827	1132.115
Depth ranges 1 to 30	1	226		971.212		1347.176
	2	200		1357.263		1674.860
	3	225		1218.913		1730.080
	4	227		1338.597		1935.352
	5	230		1090.737		1968.133
	6	200		1997.105		1444.921
	Mean over Subjects			1328.971		1683.420

Table 3.6: The standard deviation of error for the various conditions and subjects.

The control algorithm showed the best depth estimation resolution over the depth ranges 1 to 17. Also, the resolution performance of the size scaling algorithm at both cutoff distances was much better than the performance of the FOV distortion algorithm at those distances. Increasing the cutoff distance decreased the accuracy of the responses.

By assuming a normal distribution of the data, we can perform a series of analyses of variance to determine the significance of the information in Table 3.5 and Table 3.6. Again, only data from similar depth ranges can be directly compared.

Depth Ranges 1 to 16, all conditions

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Cutoff Dist.	2	12208317.964	6104158.982	8.384	.0002
Depth Range	16	227810693.895	14238168.368	19.556	<.0001
Cutoff Dist. * Depth Range	32	24621306.850	769415.839	1.057	.3811
Residual	2295	1670932241.341	728075.051		

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Algorithm	2	11384548.843	5692274.421	7.887	.0004
Depth Range	16	226930383.315	14183148.957	19.652	<.0001
Algorithm * Depth Range	32	39961410.729	1248794.085	1.730	.0068
Residual	2295	1656306847.865	721702.330		

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Condition	4	15417211.604	3854302.901	5.337	.0003
Depth Range	16	243299061.215	15206191.326	21.057	<.0001
Condition * Depth Range	64	59507492.573	929804.571	1.288	.0639
Residual	2261	1632775505.143	722147.503		

Table 3.7: An analysis of variance treating data collected from assessment trials performed in depth ranges 1 to 17 over all conditions. The topmost table shows the effect of cutoff distance, depth range, and their interaction on average absolute error. The middle table displays the effect of the algorithm used, depth range, and their interaction on average absolute error. The bottom table displays the effect of condition, depth range, and their interaction on average absolute error.

The difference in performance between the cutoff distances was not statistically significant for depth ranges 1 to 17. Furthermore, the algorithm had no significant effect on performance. The condition, which represents the four combinations of the algorithms and cutoff distances as well as the control, was also not statistically significant.

As expected, the effect of depth range was strong, indicating that the influence of distance on performance is statistically meaningful. However, by analyzing only the algorithm-enhanced cases over depth ranges 1 to 27, more data points can be considered (although the control cannot be compared).

Depth Ranges 1 to 27, no control condition

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Cutoff Dist.	1	2845083.554	2845083.554	2.240	.1346
Depth Range	26	966665758.488	37179452.250	29.272	<.0001
Cutoff Dist. * Depth Range	26	28340626.882	1090024.111	.858	.6710
Residual	2309	2932751471.228	1270139.225		

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Algorithm	1	28454067.914	28454067.914	22.801	<.0001
Depth Range	26	854714987.902	32873653.381	26.343	<.0001
Algorithm* Depth Range	26	64847618.224	2494139.162	1.999	.0020
Residual	2309	2881448370.971	1247920.472		

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Condition	2	9751763.397	4875881.698	3.886	.0207
Depth Range	24	756270856.572	31511285.690	25.112	<.0001
Condition * Depth Range	76	108221891.590	1423972.258	1.135	.2025
Residual	2257	2832187891.892	1254846.208		

Table 3.8: An analysis of variance treating data collected from assessment trials performed in depth ranges 1 to 27 over the algorithm-enhanced conditions. The topmost table shows the effect of cutoff distance, depth range, and their interaction on average absolute error. The middle table displays the effect of the algorithm used, depth range, and their interaction on average absolute error. The bottom table displays the effect of condition, depth range, and their interaction on average absolute error.

Like the results for the depth ranges 1 to 16, the effect of cutoff distance and condition was not statistically significant and the effect of depth range is strong. No interaction effects was observed. Interestingly, the algorithm used was significant, $F = 22.801$, $p < .0001$. By referring to Table 3.5, we can draw the conclusion that the size scaling algorithm conditions had significantly better performance than the FOV distortion algorithm cases for depth ranges 1 to 27. Next, we compare the algorithms for the full range of depths.

Depth Ranges 1 to 30, conditions with 13160 ft. cutoff only

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Algorithm	1	31111631.494	31111631.494	18.589	<.0001
Depth Range	29	857980123.739	29585521.508	17.677	<.0001
Algorithm * Depth Range	29	108120488.589	3728292.710	2.228	.0002
Residual	1248	2088730614.354	1673662.351		

Table 3.9: An analysis of variance treating data collected from assessment trials performed in depth ranges 1 to 30 over the algorithm-enhanced conditions. The table displays the effect of the algorithm used, depth range, and their interaction on average absolute error.

To properly calculate the data in Table 3.9, only the conditions with a cutoff distance of 13,160 feet were examined. Comparing the performance of two algorithms with an F-test at 99% significance showed that the size scaling algorithm was better than the FOV distortion algorithm.

In summary, the performance in the algorithm-enhanced cases did not differ significantly from the control case for the depth ranges 1 to 17. The accuracy decreased with distance. The size scaling algorithm was significantly better than the FOV distortion algorithm over all distances.

Before examining the effects of the algorithms and various cutoff distance on training, we should clarify what is meant by training in this experiment. Training performance is given by the both the final trained performance and the rate at which that performance is achieved. Performance implies both mean error and variance of the responses; therefore, the analysis of training should include both bias and resolution effects. Because of the size of the experiment, the number of assessment trials was somewhat limited, so the learning curve for a particular condition has only four data points.

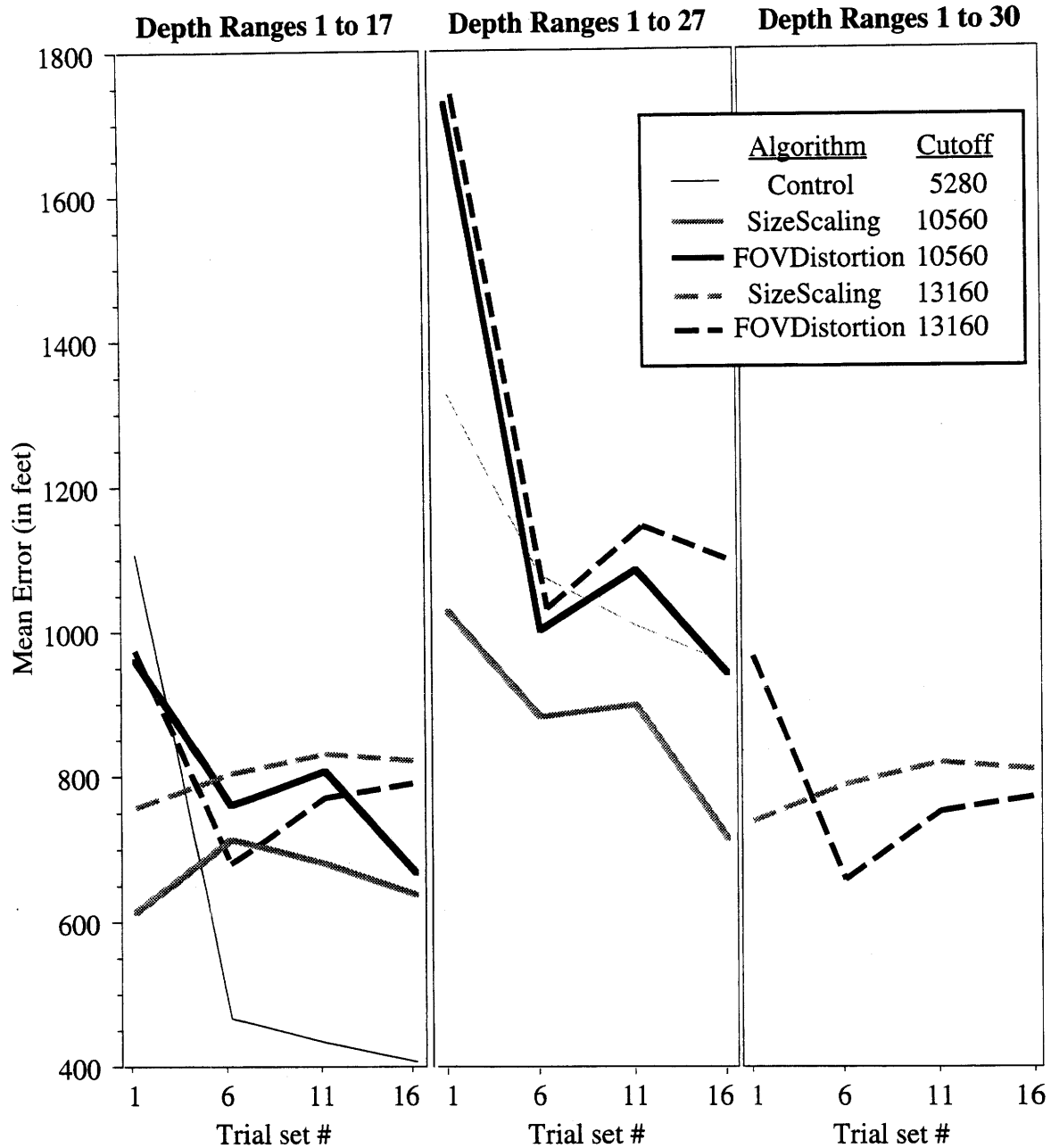


Figure 3.30: The learning curves of error for the various conditions. The control condition had the best final trained mean error. Only assessment trial means are plotted. All five conditions are shown in the leftmost plot, which shows training on the first 17 depth ranges. The middle plot compares the learning curves of the algorithm-enhanced cases for depth ranges 1 to 27. The rightmost plot shows the behavior of the two conditions that were visible over all 30 depth ranges.

Comparisons of the different conditions can only occur on the same sets of depth ranges. Because the control case is only visible on depth ranges 1 to 17, the performance of the algorithm-enhanced conditions can only be assessed versus the control on that range.

Figure 3.30 shows that most conditions had a learning curve with a negative slope implying that the performance of the subjects improved as they were trained longer. Notably, both size scaling algorithm cases show a positive learning curve slope on depth ranges 1 to 17.

The significance of the training effect for the various conditions is assessed by comparing the effect of the number of the trial on performance. Again, analysis can only be performed over comparable distances.

Depth Ranges 1 to 16, all conditions

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Trial set #	3	30681816.388	10227272.129	12.771	<.0001
Algorithm	2	14542159.940	7271079.970	9.080	.0001
Trial set # * Algorithm	6	35363332.584	5893888.764	7.360	<.0001
Residual	2334	1869057192.823	800795.712		

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Trial set #	3	32148447.469	10716149.156	13.352	<.0001
Cutoff Dist.	2	14978813.307	7489406.654	9.332	<.0001
Trial set # * Cutoff Dist.	6	30716722.326	5119453.721	6.379	<.0001
Residual	2334	1873211323.254	802575.546		

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Trial set #	3	14667064.397	4889021.466	6.100	.0004
Condition	4	18201870.218	4550467.555	5.678	.0002
Trial set # * Condition	12	36792739.141	3066061.595	3.826	<.0001
Residual	2326	1864178009.259	801452.283		

Table 3.10: An analysis of variance treating data collected from assessment trials performed in depth ranges 1 to 17 over all conditions. The topmost table shows the effect of algorithm, trial set number, and their interaction on average absolute error. The middle table displays the effect of the cutoff distance, trial set number, and their interaction on average absolute error. The bottom table displays the effect of condition, trial set number, and their interaction on average absolute error.

The significance of the training performance is revealed by the analysis of variance shown in Table 3.10. The significance of the trial set number, combined with the negative slopes in the learning curves shown in Figure 3.30 imply that training did take place. The data in

Table 3.10 also suggests that the difference in error between conditions at certain trial sets was statistically significant.

Depth Ranges 1 to 27, no control condition

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Trial set #	3	47116855.471	15705618.490	10.705	<.0001
Algorithm	1	27337350.330	27337350.330	18.634	<.0001
Trial set # * Algorithm	3	14559332.755	4853110.918	3.308	.0196
Residual	1144	1678321289.236	1467064.064		

Table 3.11: An analysis of variance treating data collected from assessment trials performed in depth ranges 1 to 27 over algorithm-enhanced conditions. The table shows the effect of algorithm, trial set number, and their interaction on average absolute error.

By looking at depth ranges 1 to 27, we can analyze the significance of the different algorithms on training. A training effect is quite apparent in Table 3.11, and the performance significantly changes with the different algorithms. By looking at Figure 3.30 we can conclude that the size scaling algorithm resulted in better performance than the FOV distortion algorithm. No interaction effect was found.

Depth Ranges 1 to 30, conditions with 13160 ft. cutoff only

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Trial set #	3	112183887.967	37394629.322	16.027	<.0001
Algorithm	1	29752942.052	29752942.052	12.752	.0004
Trial set # * Algorithm	3	10091976.297	3363992.099	1.442	.2289
Residual	1300	3033237046.934	2333259.267		

Table 3.12: An analysis of variance treating data collected from assessment trials performed in depth ranges 1 to 30. Only the two cases with a cutoff distance of 13,160 feet are considered. The table shows the effect of algorithm, trial set number, and their interaction on average absolute error.

An examination of the training effect on all 30 depth ranges reveals that training did take place, although no significant effect of algorithm was found.

In summary, significant training took place on all conditions. Training did not take place uniformly across all depth ranges. The algorithm used was shown to be significant for some depth ranges. Observation of the learning curves shows that positive learning took place for most conditions.

The effect of presenting invisible trials on training performance is difficult to determine. The control case had more invisible trials than the other conditions, and thus training performance may have been affected. However, as discussed earlier, the need to present a equal number of trials over all conditions outweighed the desire to have the same number of visible trials.

Presented	Replied	# of Trials	Control	SizeScaling		FOVDistortion	
				10560 ft.	13160 ft.	10560 ft.	13160 ft.
Invisible	Invisible	3284	2614	291	37	297	45
Visible	Visible	13098	3023	2273	2624	2434	2744
Invisible	Visible	19	15	4	0	0	0
Visible	Invisible	792	89	292	195	139	77

Table 3.13: The number normal, false-negative, false-positive cases for the various conditions.

The algorithms did have an effect on the false-visible case, as shown in Table 3.13. The number of trials where the subject claimed they could see an object when the model predicted it would be invisible was greater for the algorithm-enhanced cases. Furthermore, the control condition had more cases where the object was presented at a distance that was expected to be invisible and the subject responded that it was visible. Clearly, the effect of the algorithms on target detection should not be entirely discarded as trivial; although we will see that the sources of false-visible and false-invisible claims are easily discovered.

3.4 Discussion

We have shown that the error in estimating the location of an object in depth is a function of the viewing distance (see Table 3.7 and Table 3.8). This corroborates the

assertions made earlier regarding the number of pixels subtended for a given distance (see Equations 11 and 14). Because the number of pixels subtended is constant across particular ranges of depth, distances in those ranges can not be discriminated.

Furthermore, this assertion implies that if we were to map the depth ranges for a particular model, the accuracy in that range would depend entirely upon the size of the range. For example, if the object was two pixels from, say, 7200 feet to 10000 feet, and a number of trials were run to determine accuracy in that range, we would expect a mean error in the Ss' reply of 1400 feet. However, the linear perspective cue causes additional fragmentation of the depth ranges which results in better performance. An object may remain the same size, but move a pixel towards the horizon.

The effect of the linear perspective cue is difficult to assess because of the disappearance-reappearance and growth-shrinkage problems. Because the object may completely disappear at some depth ranges or growth in size with distance, the ability to predict the performance in particular depth range is too hard.

The control condition of the experiment reveals the effect of pixellation. An examination of Figure 3.24 shows a flattening of the curve at depth ranges 7 through 17. Looking at Table 3.2, we see that depth ranges 7 through 17 correspond to distances 1,994 feet to 5,612 feet. Then, observing the pixellation behavior shown in Figure 3.13, we see that, in the control case, the object is one pixel in size from about 2,000 feet to the cutoff distance. So, the mean error across depth ranges 7 through 17 should be about the same, which explains the plateau seen in Figure 3.24. A similar effect can be observed in the algorithm-enhanced cases, bearing in mind the logarithmic distribution of the depth ranges.

Pixellation explains the degradation of depth perception as a function of distance in a HMD and points to the effects of distance on acuity in the real world. If we are modeling the appearance of the real world in our VE, the lack of discriminability due to the spatial resolution of the HMD is only somewhat appropriate. Since the decrease in acuity with

distance occurs much faster than in the real world, the issues of visibility and depth judgment are still problematic.

As a solution to the discrepancies in the degradation of depth perception between the real world and the virtual world, two algorithms were proposed. The experiment showed that the algorithms could easily extend the visibility in the simulation, without significantly affecting performance. Only performance in depth ranges 1 to 17 could be compared to a control, but no significant difference was noted in that range.

The manner in which the algorithms extend visibility may have had an effect on training. However, the most relevant aspect of the algorithms is their ability to extend visibility without significantly affecting the error. The algorithms pushed the cutoff distance to 2 and 2.5 miles as two conditions of the experiment. As the size of a depth range was stretched to accommodate a greater range of distances, the mean error increased, but not significantly on the depth ranges 1 to 17.

The size scaling algorithm showed, for the full range of depths, a significantly better effect on performance than the FOV distortion algorithm. A possible explanation is that the size scaling algorithm took better advantage of linear perspective. Because the FOV distortion algorithm distorted the distance of the object to the horizon, it had the same number of linear perspective steps as the control. The size scaling algorithm caused the object to have more steps towards the horizon. This may have provided the additional discriminability.

The results of the experiment show that the final training performance of the control case for depth ranges 1 to 17 (see Figure 2.20) was better than any of the algorithm-enhanced conditions. This is reasonable since more pixel steps occurred in the control case over that range.

Given the significance of the number of the trial set on performance, the subjects clearly displayed a learning effect. In most cases, the learning was positive; the Ss performed with less error and better accuracy as they had more feedback trials and practice.

The cases in which the learning curve had a positive slope (negative learning) were caused by the size scaling algorithm. However, these cases were unusual in that training only was negative on depth ranges 1 to 17. This suggests that the training of the farther depth ranges had a poor influence on training at the closer ranges. However, this effect was noted only for the size scaling algorithm on depth ranges 1 to 17; positive training was observed when the range from 1 to 27 was considered.

The training implications of the experiment are not terribly serious. The experiment was subject to design constraints which prevented noise-free and easily comparable training across all conditions. Because of the higher goal of realism, the best way to assess training effectiveness is with a transfer of training experiment. This kind of test cannot be easily accomplished for the OOD task. Thus, the results regarding training performance should be taken as only guidelines, while the concrete geometrical and performance models should be considered as empirically and rigorously justified. The difficulties in performing and interpreting training experiments in virtual environments is discussed extensively by Lintern (1996).

The conclusions that can be drawn from the analysis of absolute mean error and accuracy are highly useful. Error and variance increase with distance, but may plateau over a depth range where the size of the object is constant. We have discovered that our algorithms extend visibility without introducing a significant change in depth estimation ability (based upon comparison with the control over depth ranges 1 to 17). The size scaling algorithm was shown to be significantly better than the FOV distortion algorithm.

The ramifications for the OOD simulator are obvious. The size scaling algorithm should be implemented to extend visibility in the simulator without introducing significant distortions in depth estimation. The implications of using the algorithms for training performance should be examined further if the ability to train depth perception is determined to be a critical component of the task. A solution to the visibility problem in this scenario has been presented and empirically justified.

4. Future Work

A number of changes can be made to the visibility-extending algorithms to improve their effectiveness. For example, the disappearance-reappearance (growth-shrinkage) problem can be factored into the algorithms so that the object never grows or shrinks inappropriately. In this way, we could produce a size scaling algorithm that is not subject to the interference between the size constancy and linear perspective cues. Furthermore, the FOV distortion algorithm could be geared to only distort the FOV when drawing particular objects. Although the object's position would still be distorted, the remainder of the scene would not be warped. Also, the FOV algorithm could be reworked to take better advantage of the linear perspective cues. The algorithms, as presented in this thesis, are adequate for the OOD application but could be improved further. Thus, they should be viewed as a foundation for future research into improving threshold depth perception.

In addition to improving the models and algorithms in this thesis, the effects of luminance and contrast in a typical HMD on anti-aliasing algorithms could be investigated. That is, the limited brightness and color ranges available in an HMD limits the additional resolution that can be displayed with an anti-aliasing or blurring algorithm. Determining empirical values for contrast and the resulting visual acuity predicted by the Contrast Sensitivity Function would yield a measure of the usability of anti-aliasing algorithms for improving acuity in HMDs.

The OOD simulation provides only one application of the algorithms and methodology presented here. Any VE system that requires an improvement in threshold depth perception or needs to combat the problems of low display resolution should find the models developed here to be very useful. Furthermore, applications outside the domain of VE may also find the information presented to be relevant; for example, night-vision goggles have also been plagued by low pixel resolution and decreased visibility.

5. Conclusion

While the ability to increase visibility in a virtual environment without significantly impairing depth estimation is an exciting outcome of these experiments, the successful application of software methods to hardware limitations validates our earlier assertion about the usefulness of in-code solutions. That is, the flexibility of software design permits VE system designers to account for the technological limitations of the hardware while minimizing the effect on human perception and task performance. This approach has not been widely used yet has proved to be a most effective solution.

Furthermore, the attempt to preserve realism via adherence to human characteristics has also been shown to be advantageous. Using human perceptual and performance data to design systems has long been the domain of human factors engineers, but the implementation of VE systems clearly needs extensive investigation in this area. The issue of realism in VE simulators is a nontrivial one, and an attempt should be made to better define what constitutes a good, immersive simulation.

Unfortunately, to answer the questions about realism and adherence to the real-world requires experimentation with transfer tasks. Transfer task experimentation would demonstrate how skills learned in a VE carry over to a real world environment. The nature of the problem in the OOD simulator prevented any such kind of experimentation (moving a buoy around in a bay and asking a person to estimate the distance is unreasonably difficult). If we were able to test both the VE and the real world performance on a task, we could gauge the effectiveness of the VE for simulating that task. We could also assess whether making the VE more "realistic" had an effect on performance in both the VE and the real world. We could, in this way, define what makes a VE realistic, and how we can go about attaining realism.

Furthermore, the implications of training in a VE could be examined more fully with a task that could be performed in both the synthetic world and the real world. We

could discover the ways in which a simulator needs to be realistic in order to train a given task. The best training methodology and the most realistic (in terms of human perception) may be totally different. While this seems unlikely, certainly the manner in which a VE simulator needs to be realistic should be examined both as a function of performance in the VE and as a function of transfer to the real world.

The issue of determining realism aside, using human performance as a guideline to design and implement VE systems is clearly important. VEs are the most human-centered systems ever developed by modern engineers and have incredible potential for a wide variety of applications. This potential can only be realized by continuing research into the particular eccentricities of virtual environment systems and their interaction with human beings.

References

- Allen, D. (1993). A 1" High Resolution Field Sequential Display for Head Mounted Applications. In *Proceedings of IEEE Virtual Reality Annual International Symposium '93*, (pp. 364-370).
- Barfield, W. & Furness, T. (Eds.). (1995). Virtual Environments and Advanced Interface Design. Oxford: Oxford University Press.
- Barfield, W., Hendrix, C., Bjorneseth, O., Kaczmarek, K., & Lotens, W. (1995). Comparison of Human Sensory Capabilities with Technical Specifications of Virtual Environment Equipment. *Presence*, 4(4), 329-356.
- Barfield, W., Rosenberg, C., & Lotens, W. (1995). Augmented Virtual Reality Displays. In Barfield, W. and Furness, T. (Eds.), Virtual Environments and Advanced Interface Design, (pp. 542-576). Oxford: Oxford University Press.
- Biocca, F. & Delaney, B. (1995). Immersive Virtual Reality Technology. In Biocca, F. and Levy, M. (Eds.), Communication in the Age of Virtual Reality, (pp. 57-126). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Biocca, F., Kim, T., and Levy, M. (1995). The Vision of Virtual Reality. In Biocca, F. and Levy, M. (Eds.), Communication in the Age of Virtual Reality, (pp. 3-14). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

- Boff, K. & Lincoln, J. (Eds.). (1988). *Engineering Data Compendium: Human Perception and Performance*. AAMRL, Wright-Patterson AFB, OH.
- Buser, P. & Imbert, M. (1992). Vision. Cambridge, MA: MIT Press.
- Carr, K. (1995). Introduction to Simulated and Virtual Realities. In K. Carr & R. England (Eds.), Simulated and Virtual Realities: Elements of Perception, (pp. 1-9). London: Taylor & Francis.
- Christensen, J. (1993). The Nature of Systems Development. In R. W. Pew & P. Green (Eds.), Human Factors Engineering Short Course Notes, (34th ed., Vol. 1). Ann Arbor, MI: The University of Michigan, Chrysler Center for Continuing Engineering Education.
- Christou, C., & Parker, A. (1995). Visual realism and Virtual Reality: A Psychological Perspective. In K. Carr & R. England (Eds.), Simulated and Virtual Realities: Elements of Perception, (pp. 53-84). London: Taylor & Francis.
- Cruz-Neira, C., Sandin, D., & DeFanti, T. (1993). Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In *Computer Graphics, Proceedings of SIGGRAPH '93*, (pp. 135-142). ACM SIGGRAPH, August 1993.
- Cunningham, H. & Pavel, M. (1991). Target Axis Effects under transformed Visual-Motor Mappings. In Ellis, S. R. (Ed.), Pictorial Communication in Virtual and Real Environments. (pp. 283-294). New York: Taylor and Francis.

Deering, M. (1993). Explorations of Display Interfaces for Virtual Reality. In *Proceedings of IEEE Virtual Reality Annual International Symposium '93*, (pp. 141-147).

Durlach, N. & Mavor, A. (1995). Virtual Reality and Technological Challenges.
Washington, D.C.: National Academy Press.

Edgar, G. K., & Bex, P. J. (1995). Vision and Displays. In K. Carr & R. England (Eds.), Simulated and Virtual Realities: Elements of Perception, (pp. 85-101). London: Taylor & Francis.

Ellis, S. R. (1991). Pictorial Communication: Pictures and the Synthetic Universe. In Ellis, S. R. (Ed.), Pictorial Communication in Virtual and Real Environments, (pp. 22-40). New York: Taylor and Francis.

Ellis, S. R. (1995a). Human Engineering in Virtual Environments. In *Proceedings of Virtual Reality World Conference*, (pp. 295-301). Stuttgart, Germany, February 21-23, 1995.

Ellis, S. R. (1995b). Virtual Environments and Environmental Instruments. In K. Carr & R. England (Eds.), Simulated and Virtual Realities: Elements of Perception, (pp. 11-51). London: Taylor & Francis.

England, R. (1995). Sensory-Motor Systems in Virtual Manipulation. In K. Carr & R. England (Eds.), Simulated and Virtual Realities: Elements of Perception, (pp. 131-164). London: Taylor & Francis.

- Fadden, D., Browne, R., & Widemann, J. (1991). Increasing Pilot Situational Awareness. In Ellis, S. R. (Ed.), Pictorial Communication in Virtual and Real Environments, (pp. 172-181). New York: Taylor and Francis.
- Fitts, P. (1951). Engineering Psychology and Equipment Design. In Stevens, S. (Ed.), Handbook of Experimental Psychology, (pp. 1287-1340). New York: John Wiley and Sons, Inc.
- Frederiksen, J. & White, B. (1989). An Approach to Training Based Upon Principled Task Decomposition. *Acta Psychologica*, 71, 89-146.
- Geise, W. (1946). The Interrelationship of Visual Acuity at Different Distances. *Journal of Applied Psychology*, 30, 91-106.
- Goldstein, E. (1989). Sensation and Perception. Belmont, CA: Wadsworth Publishing Company.
- Gopher, D., Weil, M., & Siegel, D. (1986). Is it only a game? Using Computer Games in the Training of Complex Skills (Final Technical Report HEIS - 86 - 8). Haifa, Israel: Technion - Israel Institute of Technology.
- Graham, C. (1951). Visual Perception. In Stevens, S. (Ed.), Handbook of Experimental Psychology, (pp. 868-920). New York: John Wiley and Sons, Inc.

- Held, R., & Durlach, N. (1992). Telepresence. *Presence*, 1(1), 109-112.
- Hendrix, C., & Barfield, W. (1995). Presence in Virtual Environments as a Function of Visual and Auditory Cues. In *Proceedings of IEEE Virtual Reality Annual International Symposium '95*, (pp. 74-82). Research Triangle Park, North Carolina, March 11-15, 1995.
- Henry, D. & Furness, T. (1993). Spatial Perception in Virtual Environments: Evaluating an Architectural Application. In *Proceedings of IEEE Virtual Reality Annual International Symposium '93*, (pp. 33-40).
- Hodges, L., & Davis, E. (1993). Geometric Considerations for Stereoscopic Virtual Environments. *Presence*, 2(1), 34-43.
- Kalawsky, R. S. (1993). The Science of Virtual Reality and Virtual Environments. New York: Addison-Wesley Publishing Company.
- Kennedy, R., Lane, N., Lilienthal, M., Berbaum, K., & Hettinger, L. (1992). Profile Analysis of Simulator Sickness Symptoms: Applications to Virtual Environment Systems. *Presence*, 1(3), 295-301.
- Kim, W., Tendick, F., and Stark, L. (1991). Visual Enhancements in Pick-and-Place Tasks. In Ellis, S. R. (Ed.), Pictorial Communication in Virtual and Real Environments, (pp. 265-282). New York: Taylor and Francis.

- Kriloff, H. (1976). Human Factor Considerations for Interactive Display Systems. In
Conference Proceedings of *The User-Oriented Design of Interactive Graphics
Systems: Application-Specific User Behavior and Cognition*. (pp. 45-52).
Pittsburgh, PA, October 14-15, 1976.
- Lampton, D., Knerr, B., Goldberg, S., Bliss, J., Moshell, J., & Blatt, B. (1994). The
Virtual Environment Performance Assessment Battery (VEPAB): Development and
Evaluation. *Presence*, 3(2), 145-157.
- Levison, W., Pew, R. W., & Getty, D. (1994). Application of Virtual Environments to
the Training of Naval Personnel (BBN Report no. 7988). Cambridge, MA: B.B.N.
Corporation.
- Levison, W., Tenney, Y., Getty, D., & Pew, R. W. (1995). Research for Virtual
Environment Training and Usability (BBN Report no. 8084). Cambridge, MA:
B.B.N. Corporation.
- Lintern, G. (1996). Human Performance Research for Virtual Training Environments. In
Proceedings of SimTecT '96, Advancing Simulation Technology and Training.
Melbourne, Australia, March 25-26, 1996.
- Lipton, L. (1991). The CrystalEyes Handbook. San Rafael, CA: StereoGraphics
Corporation.
- Leibowitz, H. (1988). Human Senses in Flight. In Weinder, E. and Nagel, D. (Eds.),
Human Factors in Aviation, (pp. 83-110). New York: Academic Press.

Ma, J., Hollerbach, J., & Hunter, I. (1993). Optical Design for a Head-Mounted Display. *Presence*, 2(3), 185-202.

Maurtius, G. (1889). Ueber die scheinbare Groesse der Gegen Staende und ihre Beziehung sur Groesse der Netzhautbilder. *Philos. Stud.* 5, 601-617.

MicroDisplay Corporation. (1996). Company world wide web pages,
<http://www.microdisplay.com>

Miller, R. B. (1976). The Human Task as Reference for System Interface Design. In Conference Proceedings of *The User-Oriented Design of Interactive Graphics Systems: Application-Specific User Behavior and Cognition*. (pp. 97-99). Pittsburgh, PA, October 14-15, 1976.

Nagata, S. (1991). Perception of Depth in Two-Dimensional Pictures. In Ellis, S. R. (Ed.), Pictorial Communication in Virtual and Real Environments, (pp. 527-545). New York: Taylor and Francis.

Negroponte, N. (1976). An Idiosyncratic Systems Approach to Interactive Graphics. In Conference Proceedings of *The User-Oriented Design of Interactive Graphics Systems: Application-Specific User Behavior and Cognition*. (pp. 53-60). Pittsburgh, PA, October 14-15, 1976.

Pew, R. W. (1993). Introduction: Human Factors Engineering. In R. W. Pew & P. Green (Eds.), Human Factors Engineering Short Course Notes, (34th ed., Vol. 1,). Ann Arbor, MI: The University of Michigan, Chrysler Center for Continuing Engineering Education.

Pioch, N. (1995). Officer of the Deck: Validation and Verification of a Virtual Environment for Training. Unpublished master's thesis, Massachusetts Institute for Technology, Cambridge, Ma.

Psotka, J., Lewis, S., & King, D. (1995). Effects of Field of View on Judgments of Self-Location: Distortions in Distance Estimations Even When the Image Geometry Exactly Fits the Field of View. (In review for publication).

Reichlen, B. (1993). Sparcchair: A One Hundred Million Pixel Display. In *Proceedings of IEEE Virtual Reality Annual International Symposium '93*, (pp. 300-307).

Rheingold, H. (1991). Virtual Reality, the Revolutionary Technology of Computer-Generated Artificial Worlds – and How it Promises and Threatens to Transform Business and Society. New York: Summit Books.

Robinett, W., & Rolland, J. (1992). A Computational Model for the Stereoscopic Optics of a Head-Mounted Display. *Presence*, 1(1), 45-62.

Robinett, W., & Holloway, R. (1995). The Visual Display Transformation for Virtual Reality. *Presence*, 4(1), 1-23.

Rolland, J., Biocca, F., Barlow, T., & Kancheria, A. (1995). Quantification of Adaptation to Virtual-Eye Location in See-Thru Head Mounted Displays. In *Proceedings of IEEE Virtual Reality Annual International Symposium '95*. Research Triangle Park, North Carolina, March 11-15, 1995.

Rolland, J., Gibson, W., & Ariely, D. (1995). Towards Quantifying Depth and Size Perception in Virtual Environments. *Presence*, 4(1), 24-48.

Sheridan, T. (1991). Musings on Telepresence and Virtual Presence. *Presence*, 1(1), 120-125.

Sheridan, T. (1992a). Defining our Terms. *Presence*, 1(2), 272-274.

Sheridan, T. (1992b). Telerobotics and Human Supervisory Control. Cambridge, MA: M.I.T. Press.

Slater, M. & Usoh, M. (1993). Presence in Immersive Virtual Environments. In *Proceedings of IEEE Virtual Reality Annual International Symposium '93*, (pp. 90-96).

Smets, G., & Overbeeke, K. (1995). Visual Resolution and Spatial Performance: The trade-off between resolution and interactivity. In *Proceedings of IEEE Virtual Reality Annual International Symposium '95*, (pp. 67-71). Research Triangle Park, North Carolina, March 11-15, 1995.

- Stanney, K. (1995). Realizing the Full Potential of Virtual Reality: Human Factors Issues That Could Stand in the Way. In *Proceedings of IEEE Virtual Reality Annual International Symposium '95*, (pp. 28-32). Research Triangle Park, North Carolina, March 11-15, 1995.
- Sutherland, I. (1965). The Ultimate Display. In *Proceedings of IFIP Congress*. Vol. 2, (pp. 506-508).
- Sutherland, I. (1968). A Head-Mounted Three-Dimensional Display. In *Proc. Fall Joint Computer Conference, AFIPS Conf. Proc.* Vol. 33, (pp. 757-764).
- Travis, D., Watson, T., & Atyeo, M. (1994). Human Psychology in Virtual Environments. In L. MacDonald & J. Vince (Eds.), Interacting with Virtual Environments, (pp. 43-59). New York: John Wiley & Sons, Ltd.
- U.S. Coast Guard. Aids to Navigation. U.S. Department of Transportation.
- VETREC (The Virtual Environment and Teleoperator Research Consortium) (1992). Virtual Environment Technology for Training (BBN Report no. 7661). Cambridge, MA: B.B.N. Corporation.
- Virtual Research Systems, Inc. (1995). VR4 head-mounted display data sheet.
- Van Cott & Paramore. (1988). In Gael, S. (Ed.), The Job Analysis Handbook for Business, Industry, and Government. New York: John Wiley and Sons.

- Wann, J., Rushton, S., & Mon-Williams, M. Natural Problems for Stereoscopic Depth Perception in Virtual Environments. *Vision Research*, (In Publication).
- Watson, B., & Hodges, L. (1995). Using Texture Maps to Correct for Optical Distortion in Head-Mounted Displays. In *Proceedings of IEEE Virtual Reality Annual International Symposium '95*, (pp. 172-177). Research Triangle Park, North Carolina, March 11-15, 1995.
- Weintraub, D. (1993). Human Visual Capacities. In R. W. Pew & P. Green (Eds.), Human Factors Engineering Short Course Notes, (34th ed., Vol. 1,). Ann Arbor, MI: The University of Michigan, Chrysler Center for Continuing Engineering Education.
- Yeh, Y. (1993). Visual and Perceptual Issues in Stereoscopic Color Displays. In D. McAllister (Ed.), Stereo Computer Graphics and Other True 3D Technologies, (pp. 50-70). Princeton, New Jersey: Princeton University Press.
- Yoshida, A., Rolland, J., & Reif, J. (1995). Design and Applications of a High-Resolution Insert Head-Mounted-Display. In *Proceedings of IEEE Virtual Reality Annual International Symposium '95*, (pp. 84-91). Research Triangle Park, North Carolina, March 11-15, 1995.
- Zeltzer, D. (1991). Autonomy, Interaction, and Presence. *Presence*, 1(1), 127-132.

Zeltzer, D., Aviles, W., Gupta, R., Lee, J., Nygren, E., Pfautz, J., & Reid, B. (1994).

Virtual Environment Technology for Training: Core Testbed. Technical Report.

Prepared for: Naval Air Warfare Center, Training Systems Division, Orlando,
Florida.

Appendix A: Hark Grammar

The grammar used in the voice recognition system is described below in standard Backus-Naur notation. Numbers in the range from 1 to 99,999 were recognized by this grammar. In addition, simple variations on the normal pronunciation were allowed, such as "forty-two thousand and one hundred" for "forty-two thousand, one hundred." Encompassing these variations improved flexibility and thus improved the accuracy of the recognition system.

NUMBER	=>	[(TENNUMBER DIGIT) thousand] and [DIGIT hundred] and [TENNUMBER]
TENNUMBER	=>	TENS DIGIT
TENS	=>	TENSXTY DIGIT TENSXTY TEENS
TENSXTY	=>	twenty thirty forty fifty sixty seventy eighty ninety
TEENS	=>	ten eleven twelve thirteen fourteen fifteen sixteen seventeen eighteen nineteen
DIGIT	=>	one two three four five six seven eight nine

Appendix B: Subject Instructions

Subjects were given the following instructions on the first day of testing. On subsequent days they received and signed an abbreviated version.

What's going on?

I am running an experiment that looks at how people see things that are far away (visual depth perception). I'd be happy to tell you more about the experiment and what I'm trying to do when you are done being a subject. For obvious reasons, I can't tell you anything now. I could also demo some of the stuff around the lab for you after your final trial, if you're interested.

What do I have to do?

A scene of (roughly) an ocean and sky will be presented to you. Your basic task is to guess how far away you think the stimulus object is in the scene you will see. You should make a guess based upon how many feet you think the object is.

It's important to realize where you are observing the scene from. Your viewpoint is 31 feet off the floor and the object you are looking at is 15 feet tall. Imagine standing on the 3rd floor of a building and looking out a small window at a lamppost some distance away. This is equivalent to the scene you will be presented with in the experiment.

What will happen?

You will be viewing the scene inside our head-mounted-display(HMD) This display is somewhat expensive and fragile, so please be careful. It is very important that you adjust the HMD to be as clear and comfortable as possible. There is a small area inside the HMD in which the image should be sharp and crisp, try to get the HMD adjusted so that this occurs. I'll help you to do this, don't worry.

You will be giving responses via a voice recognition system. After you see the object, you should press the push-to-talk button and say how far away you think the object is. After you make your guess, a brief pause will occur and another scene will be presented. On some of the trials, the computer will tell you what the right answer was. A number will appear after you make your guess; this was the correct answer for that trial. You will have a moment to absorb this number before a short pause and the next trial.

How should I guess?

Try to be as exact as you can with your guesses. Try to avoid generalizing or estimating; try to be as exact as you think you can be in your guesses (i.e. avoid "five thousand" for a set of trials where that's roughly correct; "four thousand five hundred and ten" would be better). Your responses should be numbers from 1 to 99,999. You will be trained to use the voice recognition system before running the actual experiment. A specific grammar is used with the system (you can't say "fifteen hundred" for 1500, you have to say "one thousand five hundred").

It might be really hard to see the stimulus object in some of the trials. Make the best guess you can. If you can't see the object, you can respond with "I can't see that." This stuff is elucidated in the voice recognition training.

What if I screw up?

You can't screw up. If you stumble with the response you are giving, don't worry, I'm listening too, and will catch errors. If you feel like you aren't doing very well, don't worry, it's fine.

How long do I have to do this for?

You will be given 3 breaks to help you remain focused. The first and third breaks are short, 2 minute, breaks. You can remove the HMD and relax, but you should stay in the lab. The middle break will be longer; you can wander around, get a soda, use the restroom, relax for 10 minutes. Please don't disappear for longer, it's important to only have a 10 minute break. The total time for running the experiment and filling out the forms should be a little under 2 hours.

It's a good idea to try to keep up a good pace of responding. Don't let your mind wander, but take the time you need to feel good about your response.

What if I start feeling bad or need to stop?

If you start feeling nauseous or super-uncomfortable, let me know immediately. I will be in the room during the test and can help you with any problems you might be experiencing. If for some reason you feel that you cannot continue the tests (nausea, bad headaches, etc.), that's okay and you will be paid for your time. Unfortunately, if you experience nausea or headaches, you will be excused from the experiment. Not many people experience "simulator sickness," so don't be worried, but do speak up if you start feeling really unpleasant.

I have read the subject information material and I have been familiarized with the experimental procedures. I am agreeing to participate in these experiments of my own free will. I understand that I am completely free to end my participation at any time for any reason.

Name (print): _____

Signed: _____

Date: _____