# Depth Discrimination in Cluttered Scenes Using Fishtank Virtual Reality

Shayan Rezvankhah

Master of Science

Department of Computer Science

McGill University

Montreal,Quebec

February, 2015

A thesis submitted to McGill University in partial fullfilment
of the requirements of the degree of Master of Science

# ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor, Michael Langer, without whom this thesis could never have been completed. His advice and guidence has been very valuable and the passionate conversations we have had about this thesis have motivated me.

It is also my pleasure to acknowledge the computer graphics group in University of Tehran who taught me a lot about the computer graphics fundamentals that is used in this thesis. Also special thanks to professor Moradi whom I learned the basics of academic research from.

I dedicate this work to my parents, Masoud Rezvankhah and Diana Safinazloo, who supported and encouraged me during my studies. I also thank my sister, Tara, and my friends Jamshid Rezai, Anahid Khajooei, Arman najafian, Amirali Bagherzadeh, and specially Hamidreza Fallah who have always been there for me.

## ABSTRACT

The accuracy of depth perception in virtual reality(VR) environments depends on several factors, including the scene geometry and the viewing position. Here we present one experiment to gain more insights on this relationship in general and two experiments that measured depth discrimination in a 3D cloud of small surface elements specifically.

The scenes were viewed in a fishtank VR setup. In the first experiment subjects adjusted a rectangular box until it looked like a cube and we compared the adjusted depth to width and height in different conditions. In the other two experiments, the task was to discriminate the depths of two target surfaces within the 3D clutter. Our main goal was to understand how performance in this task varies with the density of surfaces and to measure and compare the importance of cues such as the size, stereo, motion parallax and occlusion.

In these experiments we expected that, even though occlusions can be a cue to depth, performance should worsen as density is increased because occlusions reduce visibility and they make binocular stereo matching more difficult. We also expected that the decrease should be less in presence of head-tracking because fishtank VR affords the viewer a multi-view perspective and thus helps with visibility. Our results are consistent with these expectations.

We also examined whether occlusions per se led to the decrease in performance, or whether attentional aspect of clutter also play a role. We removed the effect of occlusions in one condition by digging virtual tunnels from the observer's eyes to the two targets.

Stereo gives more improvement in low density than in high density which is expected because the matching problem between the eyes is easier in low density as a result of better visibility. More interestingly, this superior performance in

low density is still present after we remove the effect of occlusions and limited visibility by tunneling in higher density scenes. This suggests that something else hinders depth discrimination in high density stereo conditions. Hypothetically there can be two reasons for this phenomenon. One hypothesis is that the tunneling condition makes some of the squares appear and disappear which is not a natural phenomenon to happen in scenes.

Another more remarkable reason could be that having dense clutter is distracting and we can not only focus on the parts of the image that we're looking for information there. So clutter reduces the performance as some type of peripheral distraction.

# ABRÉGÉ

La prcision de la perception de profondeur dans les environnements de ralit virtuelle (ERV) dpend de plusieurs facteurs, dont la gomtrie de la scne et la position de visionnement. Nous prsentons ici une exprience afin de gagner plus dintuition sur la relation gnrale entre ces deux paramtres et deux expriences afin de mesurer la diffrence de profondeur dans un nuage 3D de petits lments de surface spcifique.

Les scnes ont t visionnes dans un ERV de type aquarium. Durant la premire exprience, les sujets devaient ajuster une bote rectangulaire jusqu' ce qu'elle ressemble un cube et nous avons compar la profondeur ajuste la largeur et la hauteur dans diffrents conditions. Dans les deux autres expriences, la tche tait de diffrencier les profondeurs de deux surfaces cibles l'intrieur dun dsordre 3D. Notre objectif principal tait de comprendre comment la performance dans cette tche varie avec la densit de surfaces et de mesurer et de comparer l'importance de repres tels que la taille, la stroscopie, la parallaxe de mouvement et lobstruction visuelle.

Dans ces expriences, nous nous attendions ce que, mme si les obstructions peuvent tre des signaux la profondeur, la performance devrait diminuer lorsque la densit est accrue parce que les obstructions rduisent la visibilit et rendent la correspondance stroscopie binoculaire plus difficile. Nous nous attendions aussi ce que la diminution en performance devrait tre moindre en prsence dun suivi de la tte parce que le ERV en aquarium offre au spectateur une perspective multi-vue et donc aide la visibilit. Nos rsultats sont conformes ces attentes.

Nous avons galement examin si les obstructions en soi conduisent lune diminution de la performance, ou le type dencombrement joue galement un rle. Nous

avons enlev l'effet des obstructions dans une condition en creusant des tunnels virtuels des yeux de l'observateur jusquaux deux cibles.

La stroscopie donne plus d'amlioration  faible densit qu haute densit, ce qui est attendu parce que le problme de correspondance entre les yeux est plus facile  faible densit  cause de la meilleure visibilit. Plus intressant encore, cette performance suprieure  basse densit est toujours prsente aprs que nous enlevions l'effet des obstructions et dune visibilit limite par lajout de tunnels lorsque la densit est leve. Cela suggre que quelque chose d'autre entrave la diffrenciation de la profondeur dans des conditions stroscopie de haute densit. Hypothtiquement, il peut y avoir deux raisons  ce phnomne. Une hypothse est que lajout de tunnels fait apparaitre et disparaitre certains carrs, qui ne est pas un phnomne naturel se produisant dans les scnes.

Une autre raison plus remarquable pourrait tre que d'avoir un encombrement dense est distrayant et nous ne pouvons pas se concentrer uniquement sur les parties de l'image o nous cherchons de linformation. Donc lencombrement rduit la performance comme un type de distraction priphrique.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1
# Introduction

## 1.1 Introduction

The study of 3D visualization can help achieve better methods of visualization. Other than applications in education and training, entertainment, engineering and scientific visualization, visualization of 3D data is imperative in health care where misperception could potentially be critical.

One of the methods for visualization of 3D data is direct volume rendering (DVR). In this method every point (voxel) in the 3D data set is assigned a color and a transparency and is projected onto the 2D screen. One of the problems with this method is the need to define criteria for assigning different levels of transparency and choosing which parts need not be rendered. For example if our goal is to visualize arteries of the brain, we need to first come up with criteria to remove the non-artery parts of the brain as our scans would include all these parts. Even though DVR methods with semi transparent scenes are widely used for shape and depth judgment tasks specially in health care for visualization of MRI or other 3D imaging data, it's not clear how well humans can perceive such partially transparent scenes. In fact, there is some evidence that our perception of depth might be significantly handicapped in partially transparent scenes [30]. This thesis will study the effectiveness of an alternative method with no transparency.

Another important aspect is the environment for displaying the data. One of the developments in the last few decades is virtual reality (VR). Virtual reality, believed by some to be started in late 1950's [39], is a simulated three dimensional computer generated environment that creates the illusion of a user being present in that environment. There are many ways to have such an environment that will be discussed in Section 2.1. These methods could involve monoscopic as well as stereoscopic displays. In monoscopic displays, both eyes have the same view of the scene whereas in stereoscopic displays each eye sees the scene from a different viewpoint creating a 3D effect. The method we are going to use in this thesis is fishtank VR with stereopsis which will be discussed later in Section 2.1.

We address the perception of 3D layout using fishtank VR for the case of cluttered 3D scenes. Such scenes could include 3D data plots, volume rendered 3D data sets, or rendered worlds that contain cluttered spaces such as foliage. The scenes that we use consist of small oriented surface facets positioned randomly in a 3D volume. Such surfaces facets potentially carry more visual information than points that are commonly used in 3D data plots, since the surface facets have an orientation, in addition to a size and shape.

The disadvantage of using surface facets, however, is that they produce occlusions which limit the visibility of surfaces that are deeper within the volume. This raises the question of how well people can judge depth in such scenes. In particular, how much does a user benefit from the multiple views afforded by fish tank VR and how does this benefit trade off against the visibility problems that arise from occlusions? These are the main questions that we address in this thesis.

# CHAPTER 2
# Background and Previous Research

## 2.1 Types of VR systems

One of the many possible categorizations of VR systems is based on immersion. We have three levels of immersion for different VR systems which are immersive, semi-immersive and non-immersive [12]. Immersive VR systems encompass more senses than just the visual and involve more realistic auditory and haptic systems to make the interactions of the user in the virtual environment more natural. Semi-immersive VR systems tend to focus mostly on the visual and sometimes audio aspects and usually use projector screens around the user to provide a wide viewing angle. Lastly, non-immersive VR systems usually use monitors and have a smaller viewing angle. Our focus in this thesis is not on immersiveness and thus we will use one of the implementations of the non-immersive VR approach mentioned. Neither will we discuss the auditory, haptic or interactive aspects of VR systems.

There are many alternative technical approaches to VR systems. Some of these approaches which will be discussed below are: head mounted displays (HMD), computer assisted virtual environments (CAVE), dome projection systems, fish tank and chameleon. All aforementioned systems are visually coupled systems(VCS), "a special 'subsystem' which integrates the natural visual and motor skills of an operator into the system he is controlling." [7] [34] [22] These methods, which we will shortly introduce in more detail, use position and orientation tracking to dynamically change the displayed scene based on user's viewpoint.

Sutherland first proposed the idea of using head tracked displays to render virtual scenes depending on the observer's viewpoint[53]. A binocular head mounted display (binocular HMD) is a device worn on the head like a helmet with a display in front of each eye which allows users to be immersed in a virtual reality environment. In [54] Sutherland demonstrates a version of HMDs with crude head-tracking and emphasizes the importance of kinetic depth in depth perception. Typically, HMD devices use optical lenses between the eyes and the displays to achieve a wider viewing angle. However, these helmets tend to be uncomfortable with cables attached to it and heavy, e.g. Oculus Rift weighs almost half a kilogram without the cables. Additionally, the visual resolution is usually lower because of the size requirements for the displays.

CAVE and dome like systems rely on large displays or projector screens surrounding the user for creating an immersive virtual reality environment. In case of projector screens, multiple projectors are placed in strategic positions to seamlessly project the virtual scene on the screens. CAVE places the user inside a six sided cube where each side works as a display. Similarly, dome like displays project the virtual scene on a dome to avoid having sharp edges that can break the immersion because of inter-reflections. There are other variants of these systems, e.g. egg shaped displays that give a better resolution on the center as opposed to the periphery. These setups are not very portable and are usually expensive. With some compromises on having correct viewing angle, these systems can be used by multiple users [13].

Fishtank VR [60], which is the method we use in this thesis, is a method for rendering a small 3D scene though a desktop display window and adds motion parallax to stereo by using head-tracking or eye tracking [17] to change the displayed scene based on user's viewpoint. Although Fishtank VR is less immersive due to limited field of view, it can achieve a more accurate perception of 3D layout and position in a data visualization setting because the visual resolution is usually higher than other VR systems. Many studies of human performance using fishtank VR have been carried out showing that fishtank VR improves performance in various tasks [4] [60] [19] [50].

One of the less conventional approaches to VR systems is Chameleon. Chameleon is a hand held device with a screen that acts like a portable window to a virtual world [18] [11]. Since in this method only the position and orientation of the device is tracked, and not the user's, the user's viewpoint is not considered when showing the scene on the device. So, users are expected to hold the device close to virtual surfaces for this limitation to have minimal impact. Moreover, interactions with the virtual world can be done using the hand held device.

Darling and Ferwerda in [14] introduce a new approach, similar to Chameleon, called tangible display system. In this approach a mobile computer or laptop with accelerometer and camera is used to keep track of the computer's orientation and the viewer's position in relation to the computer.

## 2.2 Fish Tank Virtual Reality

In this thesis we will be using fish tank virtual reality which was briefly introduced in Section 2.1. Several studies have since examined how well humans perform various tasks when using fishtank VR displays. For example, Ware and colleagues looked at the problem of path tracing in graphs and compared performance under several conditions such as stereo and/or motion, where the motion could be due to the observer or to object rotation. [60, 4, 61]. Other studies have compared human performance using fishtank VR to CAVEs [16] or to volumetric displays [21] and have found that in many situations and types of tasks, fish tank VR performs as well or better than these methods.

The second and third experiments that we present in this thesis address a specific aspect of fishtank VR, namely, how depth discrimination performance depends on the 3D clutter in the scene. This issue has been addressed in the path tracing problem [62]. Here we will address a more basic perceptual problem of discriminating depth where subjects are asked to compare the depth of two red squares in a cluttered scene.

## 2.3 Issues Regarding VR Systems

One of the main issues of VR systems is the conflict of accommodation and vergence in stereoscopic displays [24]. Unlike the real world, in VR users need to maintain a fixed accommodation on the display screen regardless of the vergence angle as a cue embedded in stereoscopic disparity as otherwise the scene would be blurred and not in focus. Since as of yet there is no convenient method for tracking user's accommodation, virtual reality systems do not support simulating the correct blur effects for the displayed scene.

Another issue is ghosting or crosstalk present in the stereo display system which can reduce depth perception. Tsirlin et al. [57] measures the negative effects of crosstalk in both depth from disparity and depth from monocular occlusions. In [59] Wann discusses many other problems with stereo systems in virtual reality environments. One of these problems is with modeling the human stereoscopic system using a simple interocular distance which is a typical approach in virtual reality environments. The problem is that eye movements, or eye rotations, do not preserve interocular distance since the center of ocular rotation is different than the nodal point of the optical system [6].

Other problems include positional noise in the head-tracking sensor data as well as delay of sensors and scene rendering. These problems can cause discomfort and according to a theory proposed by Steele in the 1960's [12] this is because of the sensory conflict resulting from incompatible reports from different sensory data. One of these conflicts is the visual-vestibular conflict. The visual-vestibular conflict is caused in VR systems when the visual sensory data suggests that the observer is accelerating whereas the vestibular system located in the inner ear reports no acceleration because the user is stationary in reality [2]. Other than reducing depth perception, these problems and errors can cause eye strain, postural instability, motion sickness and nausea which are referred to as simulation sickness [12] [29].

Some of the problems mentioned above and the disadvantages of various virtual reality methods mentioned in Section 2.1 have been addressed in the technical and medical scene and methods have been developed to counteract them. Some of these issues include limited image resolution, accuracy in tracking the observer's position, handling temporal delays, limited field of view and simulation sickness [15, 38, 63, 42, 46].

Despite these problems, the fishtank VR setup produced a compelling qualitative sense of depth. It also allowed for remarkably high performance in the depth discrimination task, as the results will show. Next, we will discuss different depth cues that are combined to achieve this high performance.

## 2.4 Visual Cues for Depth Perception

The number of studies of depth discrimination from various visual cues is enormous[25]. Some of these cues include motion parallax, stereopsis, occlusions, and perspective. Studies about these cues and their interaction as well as other cues that can contribute to depth discrimination will be discussed in this section.

### 2.4.1 Clutter and Occlusions with Motion Parallax

In [51] Shimojo et al. show that humans can use occlusion-related geometric constraints to get depth information from moving objects and from whether an object appears first to the left eye or the right eye when in motion. We can assume that this is also true in our case with moving viewpoint as head-tracking has been especially helpful in determining depth of objects in cluttered scenes where many occlusion-related geometric constrains are present.

In [43] Nakayama and Shimojo prove that regions visible to only one eye, called unpaired regions, affect depth perception. They show that subjects perceive more depth when presented with stereo stimuli with valid unpaired regions than when presented with images with conflicting unpaired regions that are invalid.

In [65] Yoonessi and Baker compare the effects of two components of dynamic occlusion in motion parallax, expansion-compression comparable to shear motion or optical flow and accretion-deletion or covering and uncovering of parts of the farther object at different depths. They used random dot stereogram but in this thesis we don't use textures. However, the vertices of the objects in our experiments should provide features that are to some extent similar to features in textures as the colors of our objects are a random shade of gray. Note that in the papers mentioned so far in this section only limited number of occluders are present and dense clutter isn't studied.

In Section 1.1 we introduced direct volume rendering (DVR) as one of the visualization methods that assigns different transparency to the points in the 3D data set. In addition to making stereo matching more difficult [1], transparency removes the monocular occlusions cues and diminishes the accretion-deletion component of occlusion and thus can give less depth perception than methods that don't use transparency in many instances.

So far we have talked about cues that are useful in cluttered scenes, but the notion of clutter itself is also very important to our work, although typically that notion involves 2D images only [49]. In [36] Langer and Mannan suggest a model of visibility of surfaces in 3D clutter and verify it by comparing the model with actual visibility characteristics in a computer generated scene. Probabilities of stereoscopic and monoscopic surface visibility are calculated for both eyes separately. Additionally, probability of depth discontinuities is calculated which could potentially determine the significance of occlusion as a depth cue for the particular scene depending on depth and clutter properties.

The considerations in the model suggested by Langer and Mannan motivated the experiment of this thesis. A similar scene is used here for the second and third experiments and we will give further details on the model and compare our results with it in Section 3.6.4. We predict the clutter density to decrease visibility and thus depth discrimination performance even though occlusion, a cue quantified by depth discontinuities in the paper, and mono-occlusions cues could potentially increase with the density.

### 2.4.2 Stereo with Motion Parallax

Stereo is one of the strongest cues involved in depth perception. By comparing the features in left and right eye images the brain extracts disparity information which is then used to obtain depth [45]. In [48] Richards argued that stereoscopic depth perception is achieved by at least three different mechanisms in the brain by comparing perception of people with normal stereo vision and those that have problems in one of these mechanisms. He suggests these three mechanisms to be associated with uncrossed, near zero and crossed regions categorized based on the sign of disparity. These regions are defined to be roughly behind, on and in front of the fixation point respectively.

Depth from stereopsis works well in closer distances relative to the observer. On the other hand as the distance grows the disparity gets smaller and we have to rely more on cues other than stereo for depth perception. These cues will be discussed in Section 2.4.3. If the distance is too small (i.e. closer to the observer) the disparity between the left and right eye's images gets too large and they can not be fused together, causing diplopia or double vision [3].

The high accuracy of stereo in medium distance makes the high resolution per visual angle of Fish-tank VR suitable for applications that need accurate depth perception. The presence of other cues like texture can improve how well we can perceive the shape of objects and the accuracy of depth discrimination. An example of research done in these areas is [23] where Hillis et al. measure the just noticeable difference in the angle of stereoscopic textured slanted surfaces for different angles.

In Section 2.4.1 we discussed monocular occlusions resulting from stereoscopic viewing. Adding motion parallax to the stereo cue eliminates some of the distortions that will be mentioned in Section 2.5 and significantly improves depth perception. In [28], Johnston et al. use cylindrical surfaces in experiments meant to gain insight on how combining these cues affect depth perception. Later, Landy et al. [35] brings together this paper and others to suggest that cue combination in depth perception follows a model termed modified weak fusion(MWF) where the visual system does a weighted averaging between several cues and gives more weight for cues that are more informative.

Many studies have compared the importance of motion parallax and stereo [3] and some of the papers suggest that depth cues are combined differently in various tasks [9] [33]. Some of the tasks where this interaction is documented are motor control or hand eye coordination and depth perception or discrimination.

### 2.4.3 Other Cues

In [31] Kersten et al. compare perceptual cues such as stereo, kinetic depth, chromadepth and aerial perspective for vascular volume visualization.

Chromatic depth is a method introduced by Steenblik [52] where different depths in the scene are mapped to different colors, for example closer parts of the scene are colored red and as we move farther the color changes to orange, yellow, green and blue respectively. Similarly in aerial perspective or fog perceptual cue, closer objects have foreground colors and as we move farther away the color's saturation is decreased and it changes to a neutral background color, usually gray or blue. Aerial perspective is a natural cue resulting from scattering of light so arguably the usage of this cue or similar cues would be intuitive.

In Kersten's comparison which is done on both novice and expert subjects chromadepth and aerial perspective achieved more promising results than stereo showing that other cues can have a significant effect on depth perception. However, the stereo method used was toe in camera which is arguably geometrically wrong. The standard setup with parallel cameras might have provided better stereo results.

Madison et al. show that inter-reflection and shadows can affect perception of object contact in virtual environments in [41] but otherwise there is little evidence supporting that inter-reflections can affect depth judgments.

In Section 2.3 blur effect was discussed and it was mentioned that with current state of technology, tracking the point of accommodation is extremely difficult. One of the workarounds used is to assume a fixed accommodation or predict accommodation for different visual angles and use this value instead of the actual value for adding blur effects. Unfortunately this method is unreliable and introduces its own source of cue conflicts.

Lawson and Gulick in [37] show that the contours of objects in the absence of other cues specially shading when viewed in stereo can provide a perception of depth by themselves. Another cue that needs consideration is the lighting and shading. In [10] Bülthoff and Mallot show that the gradient of brightness when an object is viewed under diffuse lighting can also give a perception of depth. The lighting chosen for the experiments in this thesis is a single white point light from above but as the surfaces in this thesis are flat we don't have shading. Adding shadows or inter-reflections [41] to stereo would improve depth perception. In [26] Hu et al. study these improvements in motor control.

In this thesis, however, we will not use shadows, depth of field blur, fog and chromatic cues as we need to isolate only a select few cues in cluttered scenes in a VR environment in order to better study them. So we chose a few more significant cues, namely stereo, motion parallax and occlusion.

## 2.5 Distortions

One type of depth misperception is the egocentric distance [47] misperception where subjects misperceive their position after moving. Loomis et al. [40] compare the real and virtual world and have studied this phenomenon in his experiments that include subjects walking a distance with eyes closed and pointing to their previous position after seeing their new surrounding.

In [5] Baird and Biersdorf run distance estimation experiments by asking the subjects to verbally judge the ratio of a distance to some fixed standard distance. The subjects overestimated near distances when the standard distance was large but estimated well when the standard distance was small.

In the virtual world by using a 2D screen to show 3D scenes we generate cue conflicts, ambiguities and conditions different from those we are used to see in natural environments. These imperfections result in different types of distortions to occur in VR environments. In [55], Tittle et al. studies these distortions by comparing the physical and the perceptual spaces. In one of the experiments the subjects are asked to adjust the eccentricity of a cylindrical random dot stereogram so that they perceive a circular cylinder.

It was observed that the subjects adjust the eccentricity to a higher amount as the simulated viewing distance increases.

Another reason of erroneous perceptions is the ambiguity resulting from the projection from 3D to 2D. There are older studies, one of the more famous of which is Necker's cube introduced in [44] and dates back to 1832. In our first experiment we use a cube with all twelve edges visible. While using stereo prevents depth reversal for the most part, in mono conditions depth reversals may happen and affect the results. In order to reduce this depth reversal, we use perspective and brighten the front side of the cube while darkening the back side.

When using stereo, certain constraints should be applied for the characteristics of the camera and display system. In absence of these constraints or when one of the characteristics differs from the appropriate value, distortions occur [64]. These characteristics are the distance between the eyes and cameras, the convergence distance of the cameras which would be infinite for parallel cameras, the field of view of the cameras, the size of the screen and the distance of the observer to the display screen.

Some of the distortions mentioned have been shown to only be present when either stereo [27] or motion parallax are present but are resolved when both are used. As mentioned in mentioned in Section 2.4.2, this shows how combining the two cues can improve the overall depth perception by resolving conflicts.

Some of these distortions can be reduced or eliminated by applying various methods which might need calibration. For example, in [63] Wartell et al. present a method to eliminate distortions resulting from deliberate underestimation and overestimation of inter ocular distance for better fusing and more depth perception respectively. Liang et al. uses a Kalman filter to predict head tracking movement to reduce the delay and an anisotropic filter to reduce noise in [38]. Later, Kindratenko in [32] brings together the calibration methods for head tracking sensors.

# CHAPTER 3
# Experiments

Three experiments were carried out. The first one focused on depth perception in general in fish-tank VR and distortions in perception of the 3d scene. The second and third explored depth perception in cluttered scenes. In Experiment 1 subjects were asked to adjust the depth of a rectangular box until it looked like a cube. In Experiment 2 and 3 the task was to determine which one of two red squares in the clutter of gray squares was closer to the subject. In these experiments we study the interaction of size cue, stereo, motion parallax, density of the clutter, usefulness of occlusion as a cue and whether clutter can influence the accuracy of depth perception as a distraction.

## 3.1 Subjects

In all experiments, the participants were McGill University students, all naive to the purpose of the experiment. A stereo test was performed on all subjects before the experiment, using a random dot stereogram which consisted of five disjoint regions either in front of or behind the screen plane. Subjects had to sketch these regions on paper and identify their depth sign relative the screen.

Experiment 1 includes five subjects. All of the subjects correctly identified the rectangles and whether they were on the front or back. In Experiment 2, eleven participants performed the experiment, one of these failed the test and his results were removed. In Experiment 3, seven subjects participated, again one failed the test for whom the results are not included.

## 3.2 Apparatus

### 3.2.1 Stereo, Display and Rendering

In all three experiments the scene was rendered using an OpenGL program written in C++ with optimization enabled on the Visual Studio compiler on a Windows 7 operating system with display resolution of 1920 x 1080. Stereo was achieved by using NVIDIA 3D Vision shutter glasses synced with an IR emitter. The screen was refreshed at 120 Hz, so the frame rate for each eye was 60 fps.

In the three experiments we will use a ”mono” condition in addition to stereo. We define ”mono” to be a binocular viewing condition such that both eyes are shown the same frame rendered from the position between the two eyes. Although this means that neither of the eyes would have the correct perspective, the alternative would be to cover one eye or show a blank black background frame to one eye which would be annoying.

In Experiment 1 and 2 the images were presented on a 17” Dell XPS 702x laptop. In Experiment 3, a desktop equipped with an NVIDIA Quadro 4000K graphics card was used. Images were presented on a 1080p 52.5 cm in 29 cm 23.6 inches 120 Hz Acer GD235HZ monitor.

### 3.2.2 Head-tracking Sensor

Head-tracking was achieved using a mid-range 3D Guidance trakStar transmitter with magnetic sensors which were attached to the two handles of the 3D glasses. The tracking measurement rate was 80 Hz but the update rate was 240Hz. Eye position was set to be along the line segment connecting the two sensors. An interocular distance of 6.5 cm was assumed. The screen position and orientation were carefully measured relative to the trackStar coordinate system defined by the magnetic field transmitter to make sure the sensor positions relative to the screen were accurate.

For better accuracy many other factors such as warm up time, power grid frequency and avoiding proximity of the sensors or the transmitter to any metal object were taken into account as per the guidelines mentioned in the trakStar manual. We made sure there were no metal objects close to the tracker sensors and transmitter as metal objects would distort the magnetic field and produce errors in the reported sensor positions. The 3D glasses were not interfering with the magnetic field and the computer was far enough not to interfere.

### 3.2.3 Procedure

The subjects were instructed to move their heads in some conditions in Experiment 1 and in all head-tracking enabled conditions in Experiments 2 and 3 to take advantage of the head-tracking apparatus. In practice, most subjects moved their head horizontally only, and by roughly two times the interocular distance. A more accurate report of subject head motion is given in Experiment 3.

We limited the perspective information subjects could get from the scene by setting the screen to go blank if the subjects moved their heads to an extreme position, namely outside a $60 \times 60 \times 60$ cm$^3$ bounding box centered on the default viewpoint. The default viewpoint is the position of the viewpoint for conditions with no head-tracking and is different in each experiment. We also blanked the screen if the total head-motion was less than 1 cm during the last two seconds in the aforementioned conditions. Additionally, a blank screen was shown in Experiments 2 and 3 if the subject's head tilted too much from straight on viewing, which would create ghosting because of the stereo display technology used.

## 3.3 Experiment 1: Cube Distortion

Virtual reality is an imperfect simulation of reality. Some of the issues resulting in this imperfect simulation were mentioned in Section 2.3. It was also mentioned in 2.5 how Tittle et al. in [55] showed that in conditions with stereo and motion, subjects perceived more than actual depth for cylinder stimuli in close viewing distances and perceived less than actual depth for larger viewing distances. This suggests that our perception does not always match the reality. The goal of Experiment 1 was to study and measure how these distortions can influence depth perception.

With this goal in mind we have chosen a cube as a simple geometric object and proceeded to answer the question of how different conditions would affect the perceived depth of a cube. A problem we faced in designing Experiment 1 was having subjects report the perceived depth of a cube in a quantifiable manner. We used a point of subjective equality and set the task to be adjusting the depth of a rectangular box with fixed width and height until the subjects perceive a cube.

### 3.3.1 Stimulus

As seen in Figure 3–1 the stimulus was a simple white rectangular box with two rectangles as the front and back face positioned so that the center of the front size was at the center of the screen. The edges of the box were cylinders with a small radius of 1 mm. Edges that were farther from the viewpoint were darker in order to lessen subject reversals where subjects would wrongly perceive the farther face of the box closer, a well-known problem with perception of 3D objects in mono that was discussed in Section 2.5. Using perspective projection rather than orthographic projection also decreases the chances of reversals.

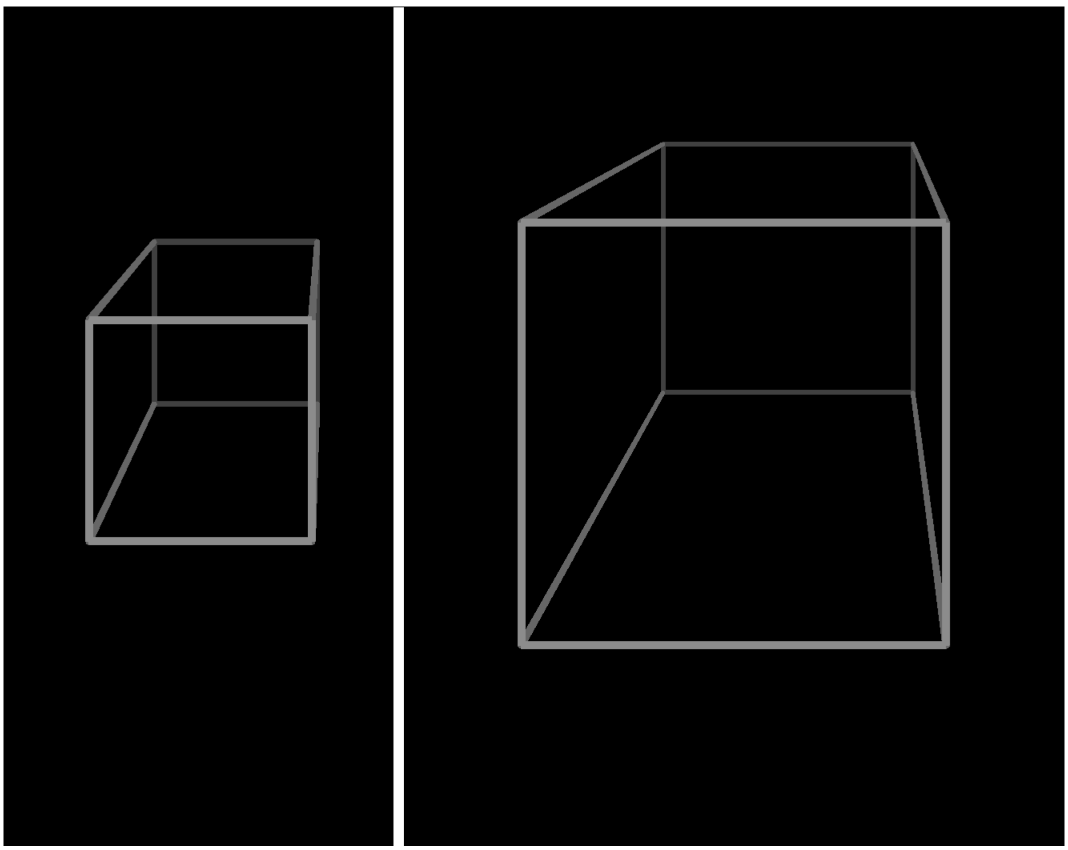

Figure 3–1: Experiment 1, the two different sizes for the stimuli

The task was to adjust the depth of the rectangular box by holding a keyboard's up or down arrow keys until it appears to be a cube. The size of the front and back rectangles were randomly chosen to be either 10.5 cm or 5.5 cm. The front face of the box was on the screen.

### 3.3.2 Procedure and design

There were a total of $72 = 6 \times 12$ test cases put into six blocks where each block contained twelve test cases with the same stereo, tracking and head-motion and combinations of the two different cube sizes, 10.5 cm or 5.5 cm, meaning six redundant test cases for each condition $72 = 6 \times (2 \times 6)$. We set the initial depth of the rectangular box to be either much longer or shorter than height and width. This was done to minimize the effect of initial depth and ensure that the subject would not just accept the test cases right away when depth was close to size from the beginning.

As mentioned before, we blanked the screen and showed a red X mark if the subject stopped moving his head in conditions with head motion whereas for conditions with no head motion, a chin-rest was used. For conditions where head motion was required the cube rotated between $\pm 25$ degrees with a constant angular velocity of 30 degrees per second. For conditions where head-motion was to be avoided, we asked the subject to put his head on the chin-rest 45 cm away from the monitor. This distance was generally greater, roughly 60 cm, for conditions with head-tracking, as subjects moved their heads away from the chin-rest to move.

This blocking was used because in our pilot experiments it was difficult for subjects to heed the directions about head-motion for each individual test case. Additionally, stereo was set to be different for each two consecutive blocks.

| Stereo | Head-tracking | Head-motion | Cube size |
|---|---|---|---|
| 0 | 0 | 0 | small |
| 0 | 0 | 1 | large |
| 0 | 1 | 1 | |
| 1 | 0 | 0 | |
| 1 | 0 | 0 | |
| 1 | 1 | 1 | |

Table 3–1: Conditions for the Experiment 1, in each condition starting depth was randomly set to be either much greater than size or much smaller.

As mentioned in Section 3.2.3, in conditions with head-motion we blanked the screen if the subject stopped moving his head. Out of the eight possible conditions for stereo, head-tracking and head-motion, the conditions with head-tracking but no head-motion enforcement were removed as no head-motion would defeat the purpose of head-tracking.

### 3.3.3 Results

As you can see in Figure 3–2, rectangular boxes appear to have less depth than they should. In stereo on the other hand, they appear deeper than they should. In other words, subjects underestimated depth for mono and overestimated depth in stereo. We saw in Section 2.5 how humans underestimate depth when there is no stereo cue [55]. We also showed studies that claim that with stereo, depth is overestimated in near field and underestimated when the object being viewed is farther from the observer. This can explain the overestimation in stereo as our viewing distances were relatively small in all conditions.

Additionally, enabling head-tracking instead of cube rotation only had a small effect where head tracking slightly increased the perceived depth. This made mono depth estimation more accurate but increased overestimation in stereo conditions thus making stereo less accurate. However these effects were too small to be statistically significant ($p = 0.2 > 0.05$).

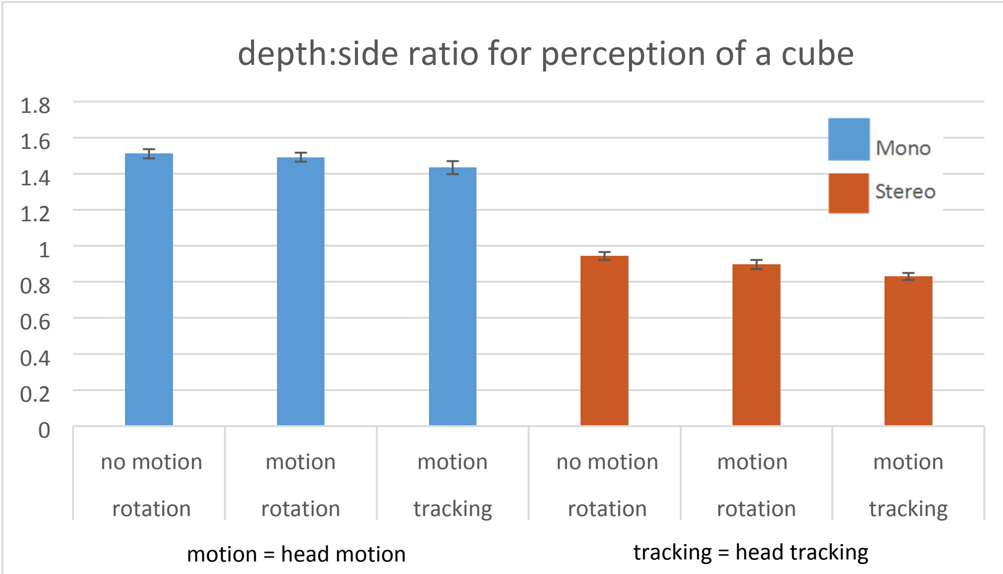


Figure 3–2: Here are the chosen depth to side length ratios pooled over cube size. Ratios more than one mean subjects perceived deeper rectangular boxes as cubes. Ratios less than one mean subjects perceived shallower rectangular boxes as cubes. Error bars show standard error of the mean.

As shown in Figure 3–3, the small and large cube sizes affected the results only slightly. We observed that in mono conditions when pooled over head-tracking and head-motion, depth to size ratio was roughly only 0.1 more when the cube was small but the effect was statistically significant with p = .004 where the comparison was between two groups of numbers each, consisting of 18 mono conditions with same cube-size for 5 subjects. The same was not the case for stereo. This suggests that in mono condition other than the usual depth underestimation, subjects underestimate depth slightly more for smaller sized cubes. This could be because of perspective cues being less useful in small cubes. Note that in our experiment the thickness of the sides of the cubes which were cylinders was the same for small and large cubes. This could have made ratio judgments harder for the smaller cubes.

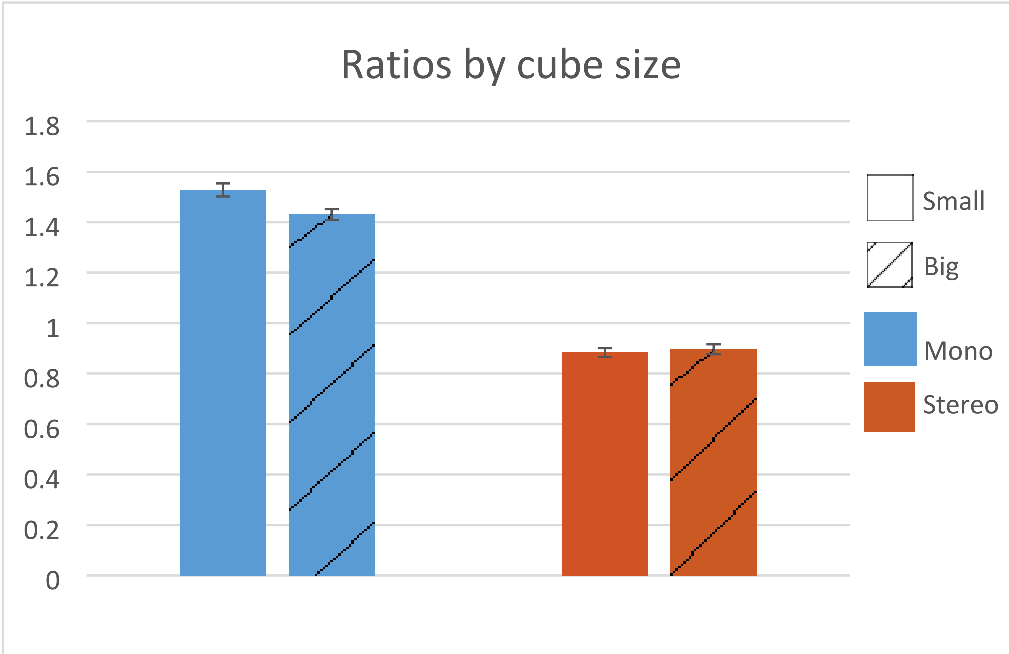


Figure 3–3: Experiment 1: In mono, subjects have underestimated the depth of the cube, whereas they have overestimated the depth in stereo conditions. Error bars show standard error of the mean.

A comparison of rotation vs. head-tracking is done in Figure 3–4. There's little difference between the two but head-tracking slightly increases depth perception. Note that this does not necessarily make the estimations more accurate.

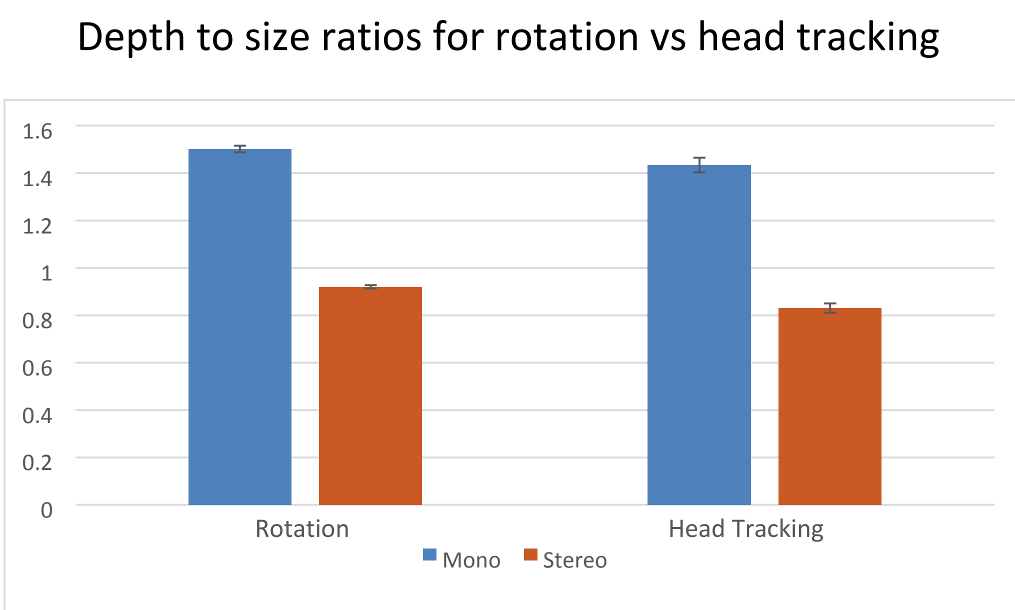


Figure 3–4: Experiment 1: The rotation is pooled over head-motion but head-tracking only contains head-motion=1. Error bars show standard error of the mean.

## 3.4 Experiment 2: Depth discrimination in cluttered scenes with size cue

In Chapters 1 and 2 we discussed how DVR methods use transparency whereas it is unknown how well humans perceive depth under this condition and we proposed an alternative approach by using opaque clutter instead. In Experiment 2 our goal is to gain insight on the accuracy of depth discrimination in cluttered scenes for different densities of clutter and the interactions of stereo, motion parallax, visibility and occlusions in these scenes. In particular, how much does a user benefit from the multiple views afforded by fish tank VR and how does this benefit trade off against the visibility problems that arise from occlusions? These are the main questions that we address in Experiment 2.

### 3.4.1 Stimuli

Subjects were presented cluttered scenes full of randomly oriented squares, each with a random shade of gray. Ambient lighting was used. Each scene was generated by placing the squares to be frontoparallel on a cubic lattice of size $22 \times 15 \times 22$ cm$^3$ and then jittering them by position and 3D orientation. The front center face of the lattice was 2 cm in front of the monitor. The total number of squares in each scene was $3^3, 6^3$ and $11^3$ for low, medium and high density respectively.

Among this clutter were two red squares facing the subject, one on the left side of the scene and one on the right. The subjects' task was depth discrimination, namely to indicate which red square was closer to the subject. Subjects responded by the pressing either the left or right arrow key on the keyboard. The average depth of the two red squares was 12 cm behind the screen, the average of the horizontal position was at the center of the screen and the average horizontal distance between the two was 11 cm. Both the horizontal and vertical positions of each of the squares were moved by a random value between 0 and 0.7 cm on every test case so that the subjects couldn't exploit the 2D position of the red squares on the monitor to judge the depth.

The conditions of the experiment including tunneling are explained below and also shown in Table 3–2: with or without stereo, head-tracking, and tunneling, and with three different densities of clutter. Since we do not use object rotations in Experiment 2 and 3, from this point on we will refer to the condition with head-tracking as the motion parallax condition. In conditions without motion parallax we fixed the viewing position to be 80 cm away from the monitor and 15 degrees above it which is a typical viewing position. Note that because the viewpoint is slightly from above, when the subject chooses the red square that is closer to the subject this does not necessarily mean that the red square is closer to subject along the Z axis of the scene. This provides a depth cue. We will control for this in Experiment 3. Unlike Experiment 1 we did not use the chin-rest so the distance for conditions with motion parallax was roughly the same as the other conditions.

| stereo | motion parallax | tunneling | density |
|---|---|---|---|
| true | true | true | low |
| false | false | false | medium |
| | | | high |

Table 3–2: $2 \times 2 \times 2 \times 3 = 24$ conditions in Experiment 2

For the tunneling condition, we removed from the scene any gray square that potentially occluded either of the red squares. We did so by setting a lower bound of 0.8 degrees on the visual angle between the center of that gray square and either of the red squares. For simplicity we didn't consider the orientation of the gray squares in the occlusion, so in very rare circumstances when the center of a gray square is just slightly behind the red square, the gray square could be randomly oriented in a way that its corner is actually in front of the red square and occludes it. If a gray occluding square was removed, then it was removed in both eyes. For conditions with both tunneling and motion parallax, gray squares would appear and disappear as the tunnel moved with the head. This was slightly distracting and may have been a factor in performance as we discuss later.

In order to prevent depth reversals for non-stereo conditions, as you can see in Figure 3–5, all scenes included two blue fronto-parallel planes that partially occluded the scene from the sides, coming in 3.8 cm from each side. These occluders were far enough to the side that they did not interfere with the depth discrimination task, and were sufficient to avoid depth reversals. We will discuss these planes again in Experiment 3 where we change the color to green.

As mentioned in Section 3.2, the subjects were instructed to move their heads to take advantage of the head-tracking apparatus. We set the screen to go blank if the subjects moved their heads to an extreme position, namely outside a $60 \times 60 \times 60$ cm$^3$ bounding box. We also blanked the screen if the total head-motion was less than 1 cm during the last two seconds or if the head rotated too much from straight on viewing, which would create ghosting. In practice, most subjects moved their head horizontally only, and by roughly two times the interocular distance as we will see in Section 3.6.4.

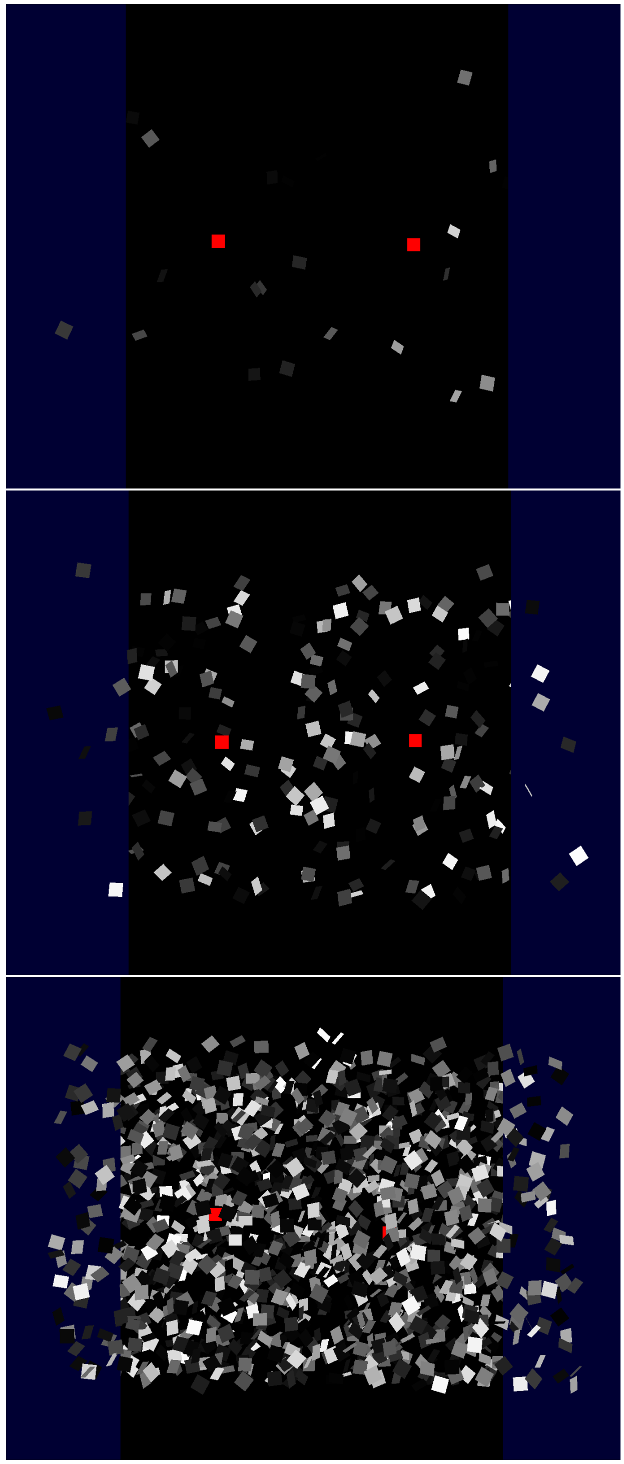

Figure 3–5: Three examples of the scene seen from a typical viewpoint. Density goes from low to high from top to bottom.

### 3.4.2 Procedure and Design

We measured depth discrimination thresholds for the red squares under the conditions shown in Table 3–2. A total of 24 interleaved 1-up 1-down staircases was used where the independent variable in these staircases was the depth difference of the two red squares.

The up and down step-size ratio was as suggested by Garcia in [20]. The step-sizes before the first reversal were $\Delta^+ = 0.5$ cm and $\Delta^- = 0.14$ cm. After the first reversal we divided these by four. The targeted proportion correct was approximately 78%.

In each trial, one of these 24 staircases was chosen at random. The termination condition for each staircase was ten reversals. In calculating the thresholds only the last six reversals were used.

Response time was limited to three seconds. If the subject didn't respond, then a random choice was made and a red X mark would show on the screen. We instructed the subjects before the experiment to try not to let this happen. Additionally, there was a rest period after each 300 trials for as long as the subject wanted. The average number of trials for the exit condition of ten reversals was 770 and the experiment typically lasted around forty minutes in total.

As mentioned earlier, we forced near continuous head-motion by blanking the screen if the subject stopped moving their head. The head-motion introduced perceptual deformations in conditions where there was stereo viewing but no motion parallax. We choose to enforce head-motion in all trials to simplify the subject's task. In a pilot study, we had tried to have subjects move their head only in motion parallax conditions, but this led to frequent blanking of the screen because the subjects stopped moving when they should have or moved when they shouldn't have which was confusing and distracting for the subjects.

### 3.4.3 Results

Figure 3–6 shows mean thresholds across subjects for each of the conditions, with the two tunneling conditions pooled since tunneling had little effect (see below). The thresholds ranged from 1.4 cm to 5 cm. Multiple linear regression was performed, yielding the following minimum least squares fit for a model of depth discrimination threshold $\tau$ in centimeters.

$$\tau = 4.28 - 1.52 * stereo - 0.96 * motionparallax + 0.7 * density - 0.1 * tunneling \quad (3.1)$$

where

$$stereo \in \{0, 1\}, \quad tracking \in \{0, 1\},$$

$$density \in \{0, 0.5, 1\}, \quad tunneling \in \{0, 1\}.$$

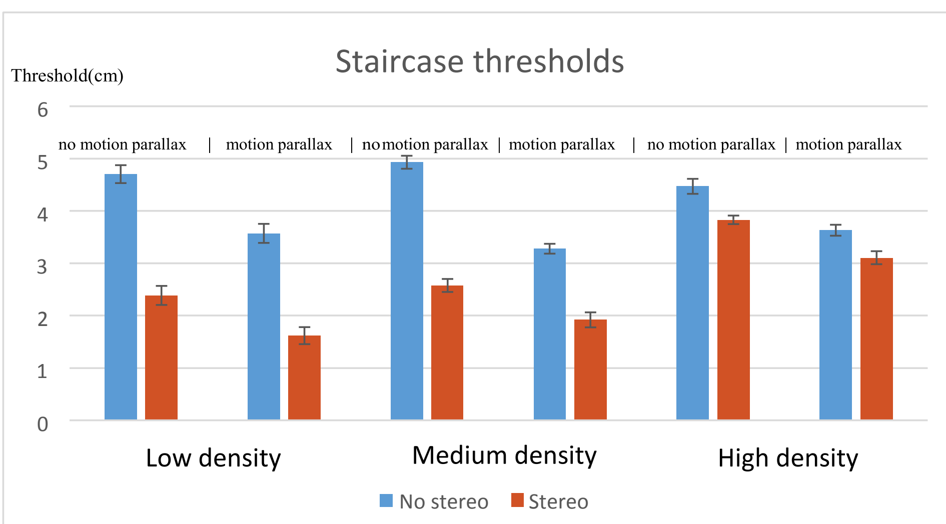


Figure 3–6: Thresholds for each condition (pooled over the two tunneling conditions) in centimeters. Error bars show the standard error of the mean.

Unsurprisingly, there were very significant differences between conditions in the ANOVA ($F_{4,235} = 32.0$ and $p = 2.7 \times 10^{-21}$). The effects of stereo, tracking and density were statistically significant with $p_{stereo} = 3.5 \times 10^{-17}$, $p_{tracking} = 2.7 \times 10^{-8}$ and $p_{density} = 8.0 \times 10^{-4}$ but the effect of tunneling was not, with $p_{tunneling} = 0.53 > 0.05$

Both stereo and motion parallax improve depth perception which is well known from other fishtank VR studies. We also found that depth discrimination is generally less accurate in higher density scenes meaning that the addition of occlusion cue does not make up for the reduction of visibility in the tested densities as will be explained below.

An interesting finding is that the monoscopic depth discrimination is independent of density. We hypothesize that this is because the occlusion cue in denser scenes makes up for the lack of visibility and the two cancel each other out resulting in the same performance. Stereo however gave less improvement in dense scenes than in sparse scenes. As seen in Figure 3–6, there is much less of a difference between each pair of orange-blue bars in the high density condition than in the medium and low density conditions. This suggests that there is an interaction effect between density and stereo effects ( $p < 0.01$ with single factor ANOVA using only low density and high density data). We hypothesize that this is because in mono conditions the depth information is scarce and occlusion cue plays an important role, so in more cluttered scenes the added occlusion cues can make up for the lack of visibility. However, we rely heavily on stereo as a strong cue so when the visibility is reduced the added occlusion cue can't make up for the decrease in stereo depth information.

In [36] Langer and Mannan propose a model that suggests that in denser scenes a larger portion of the visible surfaces are visible only to one eye. This is consistent with our results that stereo performance suffers more than mono from a loss of visibility from higher density. However, keep in mind that stereo also increases monocular occlusions which will have an effect on our comparison between mono and stereo.

For the motion parallax conditions, tunneling gives rise to spurious motion signals, namely when gray squares appear or disappear as the observer and tunnel moves. Eq. 3.1 indicates that the improvement from tunneling was very small. This was surprising, since in the medium and high density conditions occlusions were significant and tunneling allows both eyes to have a view of the red squares. The small effect of tunneling suggests that for both the tunneling and non-tunneling conditions, the reduction in performance in higher densities was not merely due to limited visibility. We hypothesize that this reduction might have been because the appearance and disappearance of the gray squares in head tracking conditions were distracting. Remember that we remove a gray square when it occludes a red square and in motion parallax conditions the subject's viewpoint changes which gray squares are visible in each frame. Additionally, tunneling removes occlusion cue which has useful depth information.

Another reason for this reduction of depth discrimination accuracy in stereo conditions with no head-tracking is that it is hard to fuse dense scenes. Tsirlin et al. show in [56] that as the number of pseudo transparent layers increase our depth perception degrades. We observe a similar pattern here in high density condition with stereo.

## 3.5 Conclusions

We have shown that, in cluttered scenes, accuracy of depth discrimination decreases with increased density of the clutter. In the general case, mono and stereo, we showed that the performance in the high density condition when removing occlusions with tunneling was the same as without tunneling. Therefore, we hypothesize that the reduced performance from the clutter wasn't merely due to occlusions. It was also likely due to extra processing needed by the stereo and motion tracking system.

We also found that stereo suffers more from lack of visibility from higher densities than mono and showed that this is consistent with the cluttered scene model introduced in [36]. In mono we don't see any difference between different densities. We hypothesized that this is because occlusion cue in denser scenes makes up for the lack of visibility and therefore some of the depth information in size cue.

### 3.6 Experiment 3: Depth perception in cluttered scenes without size cue

After analyzing the data for Experiment 2 we thought of areas where improvements could be made to the experiment to make the cues more independent or reduce the noise. Additionally, we were interested in how the size cue resulting from the perspective view specifically affects depth discrimination. We discuss the shortcomings of Experiment 2 and how we improved it in Section 3.6.1 and introduce a new condition for isolating size cue.

#### 3.6.1 Stimuli

The stimuli was similar to Experiment 2 with some improvements. In Experiment 2 we noticed that in conditions with head-motion, subjects moved away from the chin rest in order to more freely move their head. This resulted in a bias where viewing distance for head-motion conditions was usually greater than conditions with no forced head-motion. So, in Experiment 3 we did not use a chin rest and instead relied on blanking the screen on unwanted head movement to compensate. The default viewing position was changed to 62.4 cm away from the monitor to be the same as average viewing distance in Experiment 2. Remember that we blank the screen if subjects move too far from this position in conditions with motion parallax or without. The subjects were instructed to keep their distance to the display and not move closer or away from it.

Additionally, the height of the monitor was changed for each subject so that the subject would be roughly the same height as the center of the screen. This helped eliminate any perspective cues that might have been present in Experiment 2 from the horizontal position of the red rectangles. This is because in a view from front and slightly above, squares farther from the center of screen are more probable to be closer to the observer and vice versa for view slightly from below. To make sure we removed all the cues after adjusting the monitor's height, we implemented ideal observers based on horizontal position and compared the results with random choice. The observers will be discussed in Section 3.6.3.

We have two different densities instead of three in Experiment 2. Similar to Experiment 2, we manipulate the occlusion cue by tunneling and randomize the orientation of the gray squares. Similar to Experiment 2, each scene was generated by placing the squares to be frontoparallel on a cubic lattice of size $22 \times 15 \times 22$ cm$^3$ and then jittering them by position and 3D orientation. In Experiment 3 however, the front center face of the lattice was 3 cm instead of 10 cm in front of the monitor and total number of squares in each scene was $4^3$ and $11^3$ for low and high density instead of $3^3, 6^3$ and $11^3$ for low, medium and high density respectively. The slight increase in low density was because the new density was closer to our definition of a cluttered scene while maintaining the low density.

Lastly, the color of the blue planes on the sides of screen was changed to green to provide a more uniform coloring in the specific monitor as we noticed the blue color changed slightly depending on the viewpoint. These minor changes mean that while the comparison of Experiment 2 and 3 can give us valuable insight about cluttered scenes, not every difference in results should be generalized when making a direct comparison between the two.

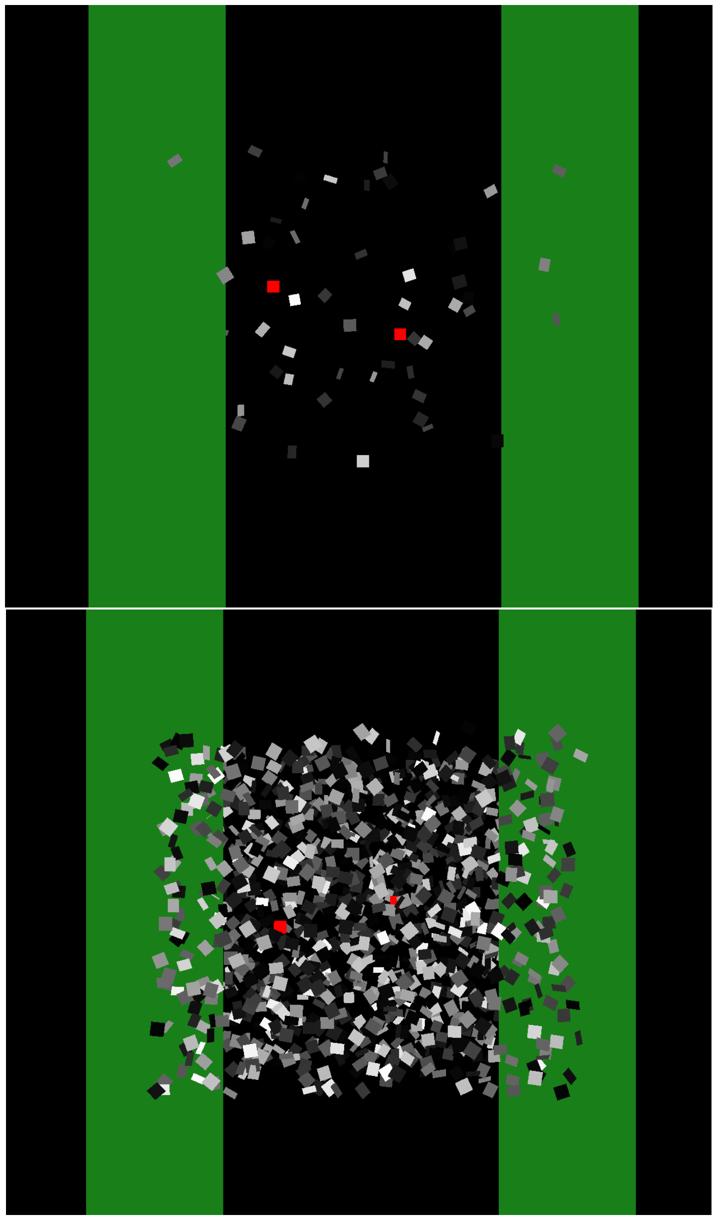

Figure 3–7: Two examples of low density and high density cluttered scenes, as seen by a typical viewing position in the experiment. The black area sides of the images are cropped.

### 3.6.2 Design and procedure

In the previous experiment we observed that interestingly, in conditions with no stereo, the density does not have a great effect on the depth discrimination accuracy of subjects. It could be that subjects were using the occlusion cue, which is related to the percentage of the red squares covered. Another relevant cue that must be considered here is size. Namely subjects may have compared the perceived 2D sizes of the two red squares in absence of any occlusion and perceived the larger one to be closer. We hypothesize that in lower densities subjects use the size cue and in higher densities the subjects switch to using the occlusion cue which could explain the results of Experiment 2 where depth discrimination was only slightly affected by density in conditions with no stereo. The goal of Experiment 3 is to confirm this hypothesis by showing that as the scene gets denser, the useful cue will change from size cue to occlusion cue.

In pilot experiments we saw that removing the size cue (explained below) has little effect except for mono low density condition with no motion parallax where the only usable cue is the size cue. This suggests that size cue is not used as much when there are stronger cues like stereo or motion parallax available. Consequently, the only condition we added to Experiment 3 was the condition where we hide the size cue in low density condition with no motion parallax. If subjects could perform better than chance in this condition, this would imply that there were other depth cues present in our stimuli than we had considered. We also removed low density conditions with tunneling as the red squares were almost never occluded. This reduced the number of staircases from 24 to 13, so we could increase the staircase termination condition from 10 to 14 reversals, increase the maximum time per test case from three to four seconds and slightly increase the minimum head-motion needed to keep the screen from going blank.

If in our results we observe that size cue is used in low density conditions when occlusion and other cues are not present, and on the other hand occlusion cue is used more in high density and not in low density then we would to some extent have explained the unchanging characteristic of depth discrimination in monoscopic conditions that we observed in Experiment 2.

We removed the downward tilt of the scene so that the typical viewpoint of the subject would be along the Z axis instead of being slightly above it. We also changed the depth of squares to start at 3 cm in front of the screen and end at 20 cm behind the screen. The new range appeared to be easier to fuse while maintaining the dynamic range of depth.

The conditions for Experiment 3 are shown in Table 3–3. Removing the size cue was implemented by having all the squares be the same 2D size regardless of distance to the viewer and perspective. In order to remove the size cue, the actual sizes of the squares was set to change dynamically based on distance to the observer.

| Size cue | Occlusion cue | Density | motion parallax | Stereo |
|---|---|---|---|---|
| 0 | 1 – No Tunneling | Low | 0 | 0 |
| 1 | 1 – No Tunneling | Low | All Combinations | All Combinations |
| 1 | 1 – No Tunneling | High | All Combinations | All Combinations |
| 1 | 0 – Tunneling | High | All Combinations | All Combinations |

Table 3–3: $1 + 3 \times 4 = 13$ conditions in Experiment 3

One of the problems with Experiment 2 was that some of the bigger thresholds were within one standard deviation of random choice threshold which could make these thresholds unreliable. Random choice threshold is the threshold we would get if a subject made every choice in the experiment at random. As we will explain later, in Experiment 3 there was a cap of 10 cm for depth difference between the red squares, otherwise the staircase would diverge with random choices. So in Experiment 3 we used a logarithmic scale for our staircases to remedy this and removed the initial phase where the step-size before the first reversal was different as this was no longer needed with logarithmic scale. The advantage of logarithmic scale is that in this scale the actual step sizes are small for the conditions where the depth difference of the red squares are small and grow larger as the depth difference of the red squares grows. So we have the accuracy of small step sizes and the dynamic range of large step sizes. The new step-size will be the same as initial step-size in last experiment except in logarithmic scale. Additionally, the maximum possible depth difference was 10 cm in Experiment 3 whereas there was no maximum defined in Experiment 2. The reason for this cap was to differentiate between the staircases that are bounded and those that are unbounded and don't home into any specific number.

We predict that stereo will be more useful in lower densities because there are more points from the red squares visible to both eyes. Existence of motion-parallax is expected to have significance in higher density conditions because it helps improve visibility much more than stereo does. Additionally, due to more visibility in low density conditions, the size cue is predicted to be more important in these scenes because subjects can more easily estimate the sizes of the red squares. Lastly, since there are fewer occlusions in low density condition, occlusion cue should perform better in higher density scenes because there will be more gray squares occluding the red squares and the amount of occlusion can give information about the depth of the red squares.

### 3.6.3 Ideal observers

Five Ideal observers were implemented for all conditions. The occlusion observer is based on number of red pixels visible for left and right red square.

This observer reports the square with the higher number of visible red pixels as closer. The size cue ideal observer calculates the maximum horizontal or vertical extents for each of the two red squares and then reports the farthest distance as the size of the red square. This observer reports the red square with larger size to be closer. In this method, we report the maximum vertical and horizontal distance between red pixels belonging to each red square as its size and choose the bigger square as closer.

Another two ideal observers are position observers and are based on the two dimensional positions of the red squares only. The position observers choose red squares farther from the center of the screen. One of these two observers solely considers vertical positions of the red squares and chooses the red square with larger vertical distance to the center of the screen. So, in other words, one reports the square farthest to the center of the screen to be the closest. The other only considers vertical position from the center of the screen and reports the farthest from the center as the closer square. Note that because of perspective the distances for the farther square will be smaller so the position observers should have greater than 50% accuracy.

The ideal observer thresholds were calculated at the same time the stimuli were being presented to the subjects. So we did not run staircases on the ideal observers since the intensity levels from trial to trial depend on the subject and not on the ideal observer. In the motion parallax conditions the observers average the number of red pixels for each side over the course of the subject's motion and thus consider changing viewpoint. For stereo, the same is done except that we perform averaging on red pixel number between the two viewpoints for the eyes as opposed to the range of all viewpoints in tracking. Again, the square with a higher number of red pixels is chosen to be closer.

The size cue ideal observer correctly answered 92% of the low density conditions, 83% of high density conditions but only 50% of the conditions where we removed the size cue. Subjects answered conditions with size cue with an average of 74% accuracy but had a 50% accuracy on the condition with removed size cues, see Table 3–4.

| | | |
|---|---|---|
| size cue ideal observer | low density | 92% |
| size cue ideal observer | high density | 83% |
| size cue ideal observer | no size cue | 50% |
| vertical position observer | all conditions | 50% |
| horizontal position observer | all conditions | 50% |
| subjects | size cue present | 74% |
| subjects | no size cue present | 50% |

Table 3–4: Optimal observer, position observer and subject accuracies in different conditions

Both position observers relying on 2D position only had approximately 50 % correct ratio which means that there are almost no 2D position cues left after changing the parameters, notably the downward slant of the whole scene. This means we have successfully isolated other cues.

Finally, a random observer was implemented that randomly chooses one of the red squares as closest regardless of any information. This random observer is to help compare subject staircases with chance staircase. We run the observer on hundreds of staircases to get a reliable threshold for the observer.

Recall that the maximum possible depth difference was 10 cm in Experiment 3. The reason for this cap was to differentiate between the bounded staircases with the unbounded staircases that don't home into any specific number. The random observer's threshold in 150 staircases was 8.5 cm with standard error of the mean being 1.5 cm.

### 3.6.4 Results for Human Subjects

As mentioned earlier in Section 3.6.2, in our pilot experiments we saw that removing size cue has little effect except for mono low density condition with no motion parallax where the only usable cue is the size cue. Size cue is not used as much when there are stronger cues like stereo or motion parallax available.

The thresholds for all conditions except the condition with hidden size cue in Experiment 3 ranged from 1 cm to 6.5 cm. By comparing these values to the random observer mean of 8.5 cm and standard error of 1.5 cm we can conclude that the results are fairly reliable.The staircase for the condition in which the sizes were hidden didn't home into any specific number and was 8 cm, close to the mean for random observer which was 8.5 cm. Which means that in this condition subjects could not detect the closer red square.

The subjects moved their head horizontally periodically with an average distance of 12 cm between changing directions. Each back and forth head movement took approximately 2 seconds. This falls well within acceptable range to perceive depth from motion parallax according to Ujike and Ono [58]. You can see the histograms for size of each motion without changing directions in Figure 3–8 and the histogram for speed of motion in Figure 3–9.

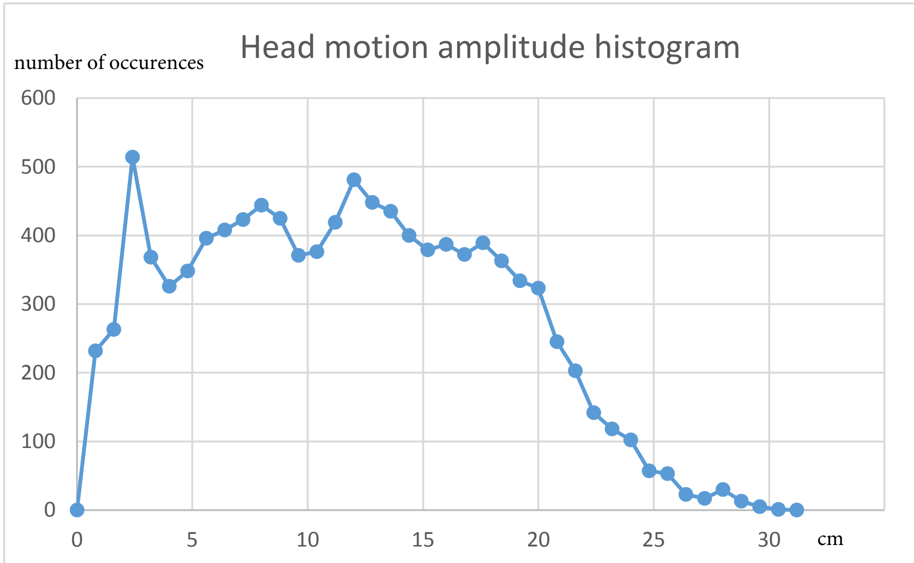


Figure 3–8: Experiment 3: Histogram for size of each motion without changing direction.

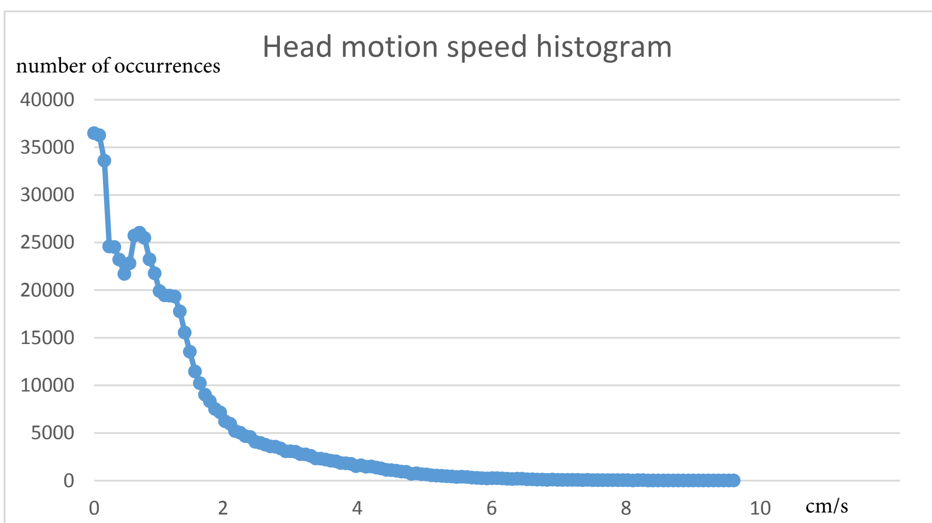


Figure 3–9: Experiment 3: Histogram for speed of head motion.

Since we removed some of the conditions, namely tunneling in low density condition, performing regression and pooling over some conditions to draw a bar chart as was done in Experiment 2 would result in a biased chart. This is because the bars in the chart would be pooled over different conditions and a direct comparison between the bars would not be possible. We can see an unbiased diagram for the conditions where size is not hidden and no tunneling is done in Figure 3–11 where every bar in any chart is pooled over the same conditions. You can compare these charts with Experiment 2 Figure 3–6. In Figure 3–10 you can see the effects of tunneling, stereo and motion-parallax only on high density conditions, as only these conditions included both tunneling and no-tunneling variants. It can be observed that performance improvement for tunneling in high density conditions was close to significant ($p = 0.06$) unlike Experiment 2 where the results were not close to significant. Additionally, although lowering the density improves performance in stereo, this is not true in monoscopic conditions. This is interesting because it shows how in mono condition occlusion can be a helpful cue and make up for lack of visibility.

Stereo gives more improvement in low density than in high density. In fact in high density condition with motion parallax stereo doesn't seem to offer any improvement at all. This is expected because the matching problem between the eyes is easier in low density as a result of better visibility. More interestingly, this superior performance in low density is still present after we remove the effect of occlusions and limited visibility by tunneling in higher density scenes. This suggests that something else hinders depth discrimination in high density stereo conditions. Hypothetically there can be two reasons for this phenomenon. One hypothesis is that in our experiment the tunneling

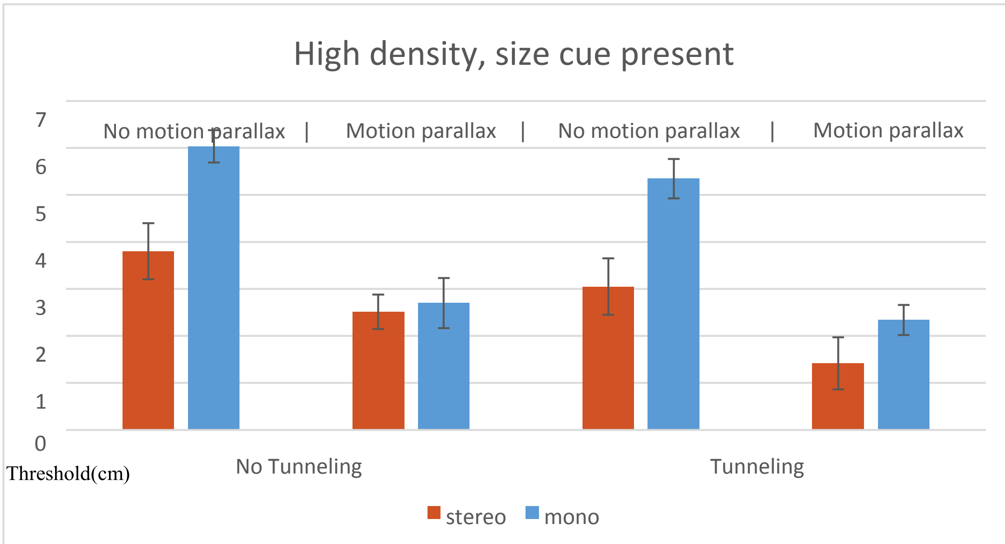


Figure 3–10: Experiment 3: Stereo and motion parallax improve depth perception while tunneling has more of an effect in stereo conditions. Error bars show standard errors of the mean.

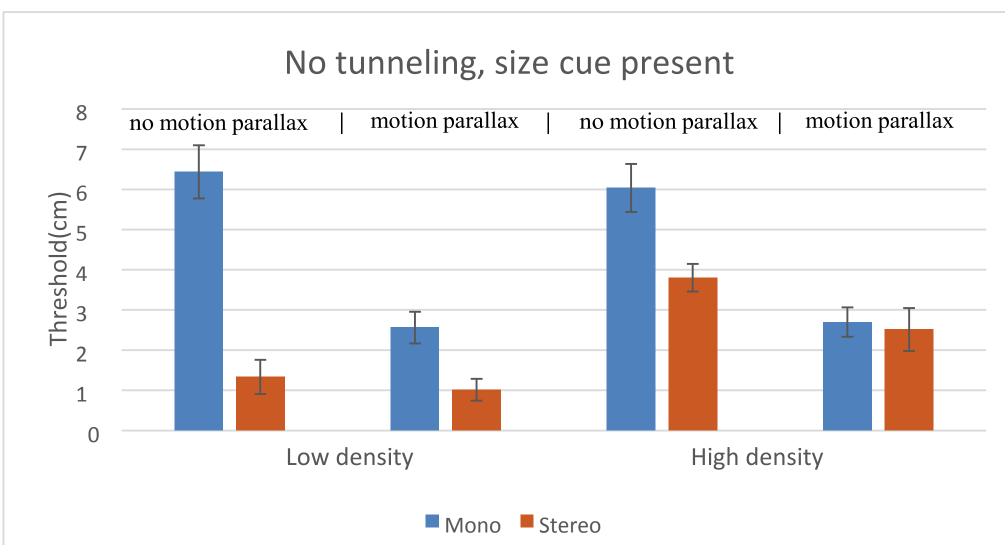


Figure 3–11: Experiment 3, error bars show the standard error of the mean.

condition makes some of the squares appear and disappear which is not a natural phenomenon to happen in scenes.

Another reason that was also mentioned in Experiment 2 is that having dense clutter is distracting and it's hard for the subject to only pay attention to fusing the parts of the image that we're looking for information. So in this case clutter reduces the performance as some type of peripheral distraction.

In contrast with stereo, we observed that motion parallax doesn't suffer in high density cases. This is because head tracking provides more size cue by offering more visibility and because more occlusions occur when the viewpoint changes.

#### 3.6.5 Conclusions

We improved the reliability of Experiment 2 by increasing the difference between subject staircases and random chance staircase so the data can rise above the noise. We also reduced the 2D position cues of the red squares and we confirmed in the experiment that this change successfully eliminated these cues. Additionally, we showed that the size cue isn't very useful except in conditions where other cues are nonexistent.

Contrary to the Experiment 2, in Experiment 3 the visibility increase resulting from tunneling in high density conditions made improvements on depth discrimination accuracy and were close to significant. This could be because the methods we used to reduce our errors in Experiment 3 resulted in the effect of tunneling to rise above the noise. Other results confirm what we saw in Experiment 2.

## 3.7 Conclusion and Future work

In Experiment 1 we studied depth distortions in order to gain a better insight on how depth is perceived in fish tank virtual reality. Subjects underestimated depth in monoscopic conditions and overestimated depth in stereo. The first is expected because in monoscopic conditions the absence of stereo cue can be perceived as having disparity of zero and thus zero depth difference between the front and back of the cube. The overestimation of depth in stereo conditions is less intuitive but confirms previous findings that depth is underestimated in near viewing distances and overestimated in far distances.

In the second and third experiments we focused on depth perception in cluttered scenes as the previous work in this area does not contain much studies about how clutter would affect our perception in virtual reality environments. We showed that stereo suffers from the clutter whereas motion parallax does not. One of the other more interesting findings was that unlike in stereo conditions where depth discrimination degrades with increased clutter, in monoscopic increasing clutter has little effect. The degradation of stereo was explained in Section 3.6.4 by using a probabilistic model of cluttered scenes in [36] while we suggested that motion parallax improves visibility more than stereo and thus doesn't degrade.

The behavior in monoscopic conditions suggests that there is an interaction effect between density and stereo and motion parallax effects. This behavior can be explained by considering that in monoscopic conditions the depth information is less than in stereo and occlusion cue plays an important role, so in denser cluttered scenes that are viewed monoscopically the occlusion cue partially makes up for the lack of visibility, whereas in stereo conditions we rely heavily on the stereo cue so when the visibility is reduced the added occlusion cue can't make up for the decrease in stereo depth information.

There were problems with the reliability of our data in some of the conditions in Experiment 2. We improved the reliability in Experiment 3 and also suggested that the size cue isn't very useful except in conditions where other stronger cues can not be easily used. Additionally, the visibility increase resulting from removing occlusions with tunneling in high density conditions made improvements on depth discrimination accuracy. This suggests that in high density conditions even though occlusion, as a cue, improves depth perception, it can not make up for the lack of visibility from occlusions.

We plan to follow up these experiments with other experiments that explore how people can perceive depth and spatial layout in cluttered scenes using fishtank VR. One important question is, what perceptual strategies are people using under cluttered scene conditions? One idea would be to train a neural network to combine the results of the ideal observers implemented in Experiment 3 and give a response close to subjects' response. Then we can compare the weights in the neural network to gain insight on which cues subjects are more likely to use in different conditions.

Another approach that we are currently exploring is to combine cluttered scenes with transparency. Traditional volume rendering uses transparency to indicate volume density and some perceptual studies have been done to explore how well people can judge depth from in such renderings e.g. [8]. It is possible that a hybrid approach that uses traditional "fog" combined with surface facets could be effective for illustrating 3D volume data.

Liang et al. in [38] propose a method to reduce head-tracking delay and noise in orientation and position data by using a Kalman filter and an anisotropic filter respectively for Isotrak sensors. Implementing a similar filter for our setup result in less temporal-spatial distortions and reduce lag as well as discomfort.

Comparison of this method of rendering cluttered scenes with other volume rendering methods, e.g. DVR, can give insight on which method would be best used in different conditions to achieve better depth perception.

Lastly, in this thesis we didn't use fog, proximity luminance, depth of field blur, chromatic, shadows and other cues which could be experimented on in future as including some of these cues may improve depth discrimination.

References


[1] Robin A Akerstrom and James T Todd. The perception of stereoscopic transparency. *Perception & Psychophysics*, 44(5):421–432, 1988.

[2] Hironori Akiduki, Suetaka Nishiike, Hiroshi Watanabe, Katsunori Matsuoka, Takeshi Kubo, and Noriaki Takeda. Visual-vestibular conflict induced by virtual reality in humans. *Neuroscience letters*, 340(3):197–200, 2003.

[3] Roland Arsenault and Colin Ware. The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank vr. *Presence: Teleoperators and Virtual Environments*, 13(5):549–559, 2004.

[4] Kevin W Arthur, Kellogg S Booth, and Colin Ware. Evaluating 3d task performance for fish tank virtual worlds. *ACM Transactions on Information Systems (TOIS)*, 11(3):239–265, 1993.

[5] John C Baird and William R Biersdorf. Quantitative functions for size and distance judgments. *Perception & Psychophysics*, 2(4):161–166, 1967.

[6] AG Bennett and RB Rabbetts. The eyes optical system. *Clinical visual optics*, pages 9–22, 1984.

[7] J.A. Birt, H.L. Task, AEROSPACE MEDICAL DIV BROOKS AFB TX., United States. Aerospace Medical Division, and Air Force Aerospace Medical Research Laboratory (U.S.). *Proceedings Of: A Symposium on Visually Coupled Systems: Development and Application*. Defense Technical Information Center, 1973.

[8] Christian Boucheny, Georges-Pierre Bonneau, Jacques Droulez, Guillaume Thibault, and Stephane Ploix. A perceptive evaluation of volume rendering techniques. *ACM Trans. Appl. Percept.*, 5(4):23:1–23:24, February 2009.

[9] Mark F Bradshaw, Andrew D Parton, and Andrew Glennerster. The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, 40(27):3725–3734, 2000.

[10] Heinrich H Bülthoff and Hanspeter A Mallot. Integration of depth modules: stereo and shading. *Journal of Optical Society of America A*, 5(10):1749–1758, 1988.

[11] Bill Buxton and George W Fitzmaurice. Hmds, caves & chameleon: a human-centric analysis of interaction in virtual space. *ACM Siggraph Computer Graphics*, 32(4):69–74, 1998.

[12] Patrick J Costello. *Health and safety issues associated with virtual reality: a review of current literature.* Citeseer, 1997.

[13] Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. The cave: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, 1992.

[14] Benjamin A Darling and James A Ferwerda. Tangible display systems: direct interfaces for computer-based studies of surface appearance. In *IS&T/SPIE Electronic Imaging*, pages 75270Q–75270Q. International Society for Optics and Photonics, 2010.

[15] Michael Deering. High resolution virtual reality. *ACM Siggraph Computer Graphics*, 26(2):195–202, 1992.

[16] Cagatay Demiralp, Cullen D Jackson, David B Karelitz, Song Zhang, and David H Laidlaw. Cave and fishtank virtual-reality displays: A qualitative and quantitative comparison. *Visualization and Computer Graphics, IEEE Transactions on*, 12(3):323–330, 2006.

[17] Andrew T Duchowski, Vinay Shivashankaraiah, Tim Rawls, Anand K Gramopadhye, Brian J Melloy, and Barbara Kanki. Binocular eye tracking in virtual reality for inspection training. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 89–96. ACM, 2000.

[18] George W Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 36(7):3949, 1993.

[19] Andrew Forsberg, Michael Katzourin, Kristi Wharton, Mel Slater, et al. A comparative study of desktop, fishtank, and cave systems for the exploration of volume rendered confocal data sets. *Visualization and Computer Graphics, IEEE Transactions on*, 14(3):551563, 2008.

[20] Miguel A. Garcia-Perez. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12):1861–1881, June 1998.

[21] Tovi Grossman and Ravin Balakrishnan. An evaluation of depth perception on volumetric displays. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, pages 193–200, 2006.

[22] Michael W Haas. Visually coupled systems as simulation devices. *Journal of Aircraft*, 21(8):639–640, 1984.

[23] James M Hillis, Simon J Watt, Michael S Landy, and Martin S Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12):1, 2004.

[24] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 2008.

[25] I.P. Howard and B.J. Rogers. *Perceiving in Depth, Volume 2: Stereoscopic Vision*. Oxford Psychology Series. Oxford University Press, USA, 2012.

[26] Helen H Hu, Amy A Gooch, William B Thompson, Brian E Smits, John J Rieser, and Peter Shirley. Visual cues for imminent object contact in realistic virtual environment. In *Proceedings of the conference on Visualization'00*, pages 179–185. IEEE Computer Society Press, 2000.

[27] Elizabeth B Johnston. Systematic distortions of shape from stereopsis. *Vision Research*, 31(7):1351–1360, 1991.

[28] Elizabeth B Johnston, Bruce G Cumming, and Michael S Landy. Integration of stereopsis and motion shape cues. *Vision Research*, 34(17):2259–2275, 1994.

[29] Robert S Kennedy, Julie Drexler, and Robert C Kennedy. Research in visually induced motion sickness. *Applied Ergonomics*, 41(4):494–503, 2010.

[30] Marta A Kersten, A James Stewart, Niko Troje, and Randy Ellis. Enhancing depth perception in translucent volumes. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):1117–1124, 2006.

[31] Marta Kersten-Oertel, SJ-S Chen, and D Louis Collins. An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. *Visualization and Computer Graphics, IEEE Transactions on*, 20(3):391–403, 2014.

[32] Volodymyr V Kindratenko. A survey of electromagnetic position tracker calibration techniques. *Virtual Reality*, 5(3):169–182, 2000.

[33] David C Knill. Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception. *Journal of Vision*, 5(2):2, 2005.

[34] Dean F Kocian and H Lee Task. Visually coupled systems hardware and the human interface. *Virtual environments and advanced interface design*, 6:175–257, 1995.

[35] Michael S Landy, Laurence T Maloney, Elizabeth B Johnston, and Mark Young. Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35(3):389–412, 1995.

[36] Michael S Langer and Fahim Mannan. Visibility in three-dimensional cluttered scenes. *Journal of the Optical Society of America A*, 29(9):1794–1807, 2012.

[37] RB Lawson and WL Gulick. Stereopsis and anomalous contour. *Vision Research*, 7(3):271–IN12, 1967.

[38] Jiandong Liang, Chris Shaw, and Mark Green. On temporal-spatial realism in the virtual reality environment. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, pages 19–25. ACM, 1991.

[39] JohnJ Lombardi. Virtual & augmented reality. *Communication Technology Update and Fundamentals*, page 182, 2008.

[40] Jack M Loomis, José A Da Silva, Naofumi Fujita, and Sergio S Fukusima. Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):906, 1992.

[41] Cindee Madison, Daniel J Kersten, William B Thompson, Peter Shirley, and Brian Smits. The use of subtle illumination cues for human judgement of spatial layout. *University of Utah Technical Report UUCS-99-001. Computer Science Dept., Univ. of Utah*, 1999.

[42] Jurriaan D Mulder and Robert Van Liere. Enhancing fish tank VR. In *Virtual Reality, 2000. Proceedings. IEEE*, pages 91–98. IEEE, 2000.

[43] Ken Nakayama and Shinsuke Shimojo. Da vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research*, 30(11):1811–1825, 1990.

[44] Louis Albert Necker. LXI. observations on some remarkable optical phænomena seen in switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid. 1832.

[45] Robert Patterson and Wayne L Martin. Human stereopsis. *Human Factors*, 1992.

[46] EC Regan and AD Ramsey. The efficacy of hyoscine hydrobromide in reducing side-effects induced during immersion in virtual reality. *Aviation, space, and environmental medicine*, 67(3):222–226, 1996.

[47] Rebekka S Renner, Boris M Velichkovsky, and Jens R Helmert. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)*, 46(2):23, 2013.

[48] Whitman Richards. Anomalous stereoscopic depth perception. *Journal of Optical Society of America*, 61(3):410414, 1971.

[49] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. Measuring visual clutter. *Journal of Vision*, 7(2):17, 2007.

[50] Beatriz Sousa Santos, Paulo Dias, Angela Pimentel, Jan-Willem Baggerman, Carlos Ferreira, Samuel Silva, and Joaquim Madeira. Head-mounted display versus desktop for 3d navigation in virtual reality: a user study. *Multimedia Tools and Applications*, 41(1):161181, 2009.

[51] Shinsuke Shimojo, Gerald H Silverman, and Ken Nakayama. Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29(5):619–626, 1989.

[52] Richard A Steenblik. The chromostereoscopic process: a novel single image stereoscopic process. In *OE LASE'87 and EO Imaging Symp (January 1987, Los Angeles)*, pages 27–34. International Society for Optics and Photonics, 1987.

[53] Ivan E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the AFIPS Fall Joint Computer Conference*, pages 757–764, 1968.

[54] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, part I*, pages 757–764. ACM, 1968.

[55] James S Tittle, James T Todd, Victor J Perotti, and J Farley Norman. Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3):663, 1995.

[56] Inna Tsirlin, Robert S Allison, and Laurie M Wilcox. Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision*, 8(5):5, 2008.

[57] Inna Tsirlin, Laurie M Wilcox, and Robert S Allison. The effect of crosstalk on the perceived depth from disparity and monocular occlusions. *Broadcasting, IEEE Transactions on*, 57(2):445453, 2011.

[58] Hiroyasu Ujike and Hiroshi Ono. Depth thresholds of motion parallax as a function of head movement velocity. *Vision Research*, 41(22):28352843, 2001.

[59] John P Wann, Simon Rushton, and Mark Mon-Williams. Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, 35(19):27312736, 1995.

[60] Colin Ware, Kevin Arthur, and Kellogg S Booth. Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, pages 37–42. ACM, 1993.

[61] Colin Ware and Glenn Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics (TOG)*, 15(2):121–140, 1996.

[62] Colin Ware and Peter Mitchell. Visualizing graphs in three dimensions. *ACM Transactions on Applied Perception*, 5(1), 2008.

[63] Zachary Wartell, Larry F. Hodges, and William Ribarsky. Balancing fusion, image depth and distortion in stereoscopic head-tracked displays. In *Siggraph*, pages 351–358, 1999.

[64] Andrew J Woods, Tom Docherty, and Rolf Koch. Image distortions in stereoscopic video systems. In *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, pages 36–48. International Society for Optics and Photonics, 1993.

[65] Ahmad Yoonessi and Curtis L Baker. Depth perception from dynamic occlusion in motion parallax: Roles of expansion-compression versus accretion-deletion. *Journal of Vision*, 13(12):10, 2013.