Depth Cues in Virtual Reality and Real World:

Understanding Individual Differences in Depth Perception by Studying Shape-from-shading and Motion Parallax

by danah boyd

An honors thesis submitted for a Bachelor of the Arts Degree
in the Department of Computer Science at Brown University

Providence, Rhode Island
2000

Abstract

In order to better understand the interaction between various depth cues, we analyzed two 3D depth cues that are found in both natural spaces and virtual reality environments – shape from shading and motion parallax. We wanted to see how well subjects used these cues as sole information for depth and how they weighted or integrated the two cues. In particular, we were interested in individual variance and potential differences along sex lines. Our results showed that subjects, without practice or experience, are virtually unable to acquire depth information when 2D cues such as luminance and speed are removed and only one 3D cue remains. Because of the irregular performance of our subjects, it is difficult to determine what individual variation exists and whether it was controlled by experience, social influences, or biological factors. We suggest further research be done to disentangle the 2D and 3D cues and to determine the impact of experience on performance. Once systems are in place to properly test for separate cues and give the appropriate level of training, it may be possible to study individual variance. At this juncture, we only know that learning the system improves ability as do the availability of 2D cues.

Acknowledgements

From helping me formulate ideas to suggesting appropriate research to helping develop an appropriate experiment, Leslie Welch has been integral in the completion of this project. I owe my greatest appreciation to her for helping me integrate ideas from three different disciplines in order to build a bridge strong enough to test. Her calm approach to life and work has motivated me immensely. I am in awe of her work and her desire to help me.

Without the support and dedication of Andy van Dam, I would never have realized the magic that technology holds and the power that I have to alter its future path. He has provided unwavering support, a few kicks in the head, and more advice than I thought an advisor was capable of giving. I am beyond grateful for his commitment to focusing my outlandish endeavors and overenthusiastic behavior into productive work.

This project could not have been completed without the ideas, work, direction and editing of: Benjamin White, Dan Gould, Matt Hutson, Margaret Klawunn, David Laidlaw, Jesse Chan-Norris, Jeff Potter, Hannah Rohde, and Vrije Universiteit Gender Clinic.

Finally, my eternal gratitude to those who support me with their absolute love and support: my mother, brother, grandparents, and Jon.

Contents

1 Introduction and background

There is a great deal of research about depth perception in both vision literature and computer graphics. For the most part, vision researchers are primarily focused on learning how people understand depth in a natural space, often by developing experiments that utilize the computer. In the case of graphics researchers and virtual environments' designers, the primary focus is on creating scenes where the information presented looks good enough for users to understand what is happening. Due to a great deal of computational expense, this sometimes means utilizing limited depth cues or modifying them based on probable assumptions. Interrante (1996) presents a comprehensive introduction to both vision and computational perceptions of depth.

1.1 How vision theory addresses depth

Developing a sense of three-dimensional space is something that most humans do not consciously think about, yet it happens constantly. Precisely how the brain computes this information is still disputed. In a natural space, there exist multiple cues that humans use to understand depth in three dimensions, including binocular disparity / stereopsis, pictorial cues (such as shading, texture, linear perspective, etc.) and moving cues (kinetic depth effect, motion parallax, etc.). How these different cues are used together is more of a mystery.

We understand that all of these cues give a certain amount of information to the visual system regarding depth. Individually, each cue gives indicative information for determining depth, and some cues appear to give stronger, or more convincing, information than other cues. Additional cues increase one's ability to accurately determine depth information, even when the additional cue is not nearly as strong as the original cue (Johnston et al., 1993). No one cue is necessary for an accurate portrayal of depth nor does any one cue dominate our depth perception in all scenarios (Cutting & Vishton, 1995). While this information amounts to a clearer understanding of how depth perception detection works, it does not create a convincing argument for how they are integrated in the visual system.

Four theories are commonly presented as possible explanations for understanding how the diverse cues interact with one another (Bülthoff, & Mallott, 1988; Johnston et al., 1993).

Accumulation or Weighted Linear Combination or ***Weak Fusion***. Each cue strengthens the estimate integrated by the visual system after each cue is processed separately. The following research indicated some form of linear combination as the explanation for various cue combinations:

- stereo, perspective, and proximity luminance (Dosher et al., 1986)
- motion parallax, occlusion, height in the picture plane, and familiar size (Bruno & Cutting, 1988)
- motion parallax, stereo (Rogers & Collett, 1989)

Cooperation or ***Strong Fusion***. Cues cooperate with each other prior to obtaining depth estimates. The relative importance of each cue varies depending on the reliability of the cue in a given situation. Reliability can be altered when visual information is noisy or when there is incomplete information for that cue (Maloney & Landy, 1989).

Disambiguation. One cue is used to locally disambiguate a representation derived by another, all of which provide inherently ambiguous information (i.e., stereo can disambiguate shading). Information from separate depth cues indicates which of two explanations is more accurate based on weighting of each module (Dosher et al., 1986; Blake & Bülthoff, 1990).

Veto. Cues veto one another when there is conflicting information, but because of weighting, there is a prioritization of which cues are ignored when in conflict. One cue conveys depth information that will not be challenged by other cues when the information is continuous; when it is not, cues that are less weighted are ignored. As an example, Bülthoff and Mallott (1988) found that when stereo indicates a flat surface but shading indicates an ellipsoid, no significant depth is perceived, indicating that stereo can veto shape from shading.

Various cues have been pitted against one another in order to determine what the relationship between them might be. Evidence is available for all of the potential explanations, with multiple combinations of cues and in differing scenarios. Research also indicates that observers are able to learn cue combination strategies and adapt them depending on the stimuli and the type of training (Jacobs & Fine, 1999). This raises the question of whether or not disentanglement is possible.

For the purpose of this experiment, we are using two of these cues – shape from shading (SFS) and motion parallax. These two cues are our primary cues because they are always available in both the real world and all artificial 3D worlds. In addition, while there is a great deal of information available about these common cues, no work has been done attempting to use them as conflicting separate cues. Information about the shape, including its relative depth, can be determined based on the shading of that object. Parts of an object with consistent absorption, reflection and matte coloring will have darker luminance when further away while the parts of the object closest to the observer will appear lighter[1]. Using luminance information to derive self-referential depth of an individual shape is called shape from shading. Motion parallax refers to the differences in relative speed between parts of an object when either the object is moving or when the observer moves hir head. Parts of the object that are further away will move slower while closer parts will appear to move faster. For many situations, relative comparisons of shading and motion construct a decent understanding of how the objects appear in relation to one another and within their own relative structure.

[1] This also assumes that properties of the lighting, such as the presence of a directional light from above. These assumptions address crucial differences between natural and virtual shape from shading. Regardless, for the clarification of SFS, we are assuming appropriate lighting.

Both SFS and motion parallax are available in real space as well as most 2D and 3D computationally created environments. While both cues have been studied individually, there is no published work putting these two cues in conflicting situations against one another. In addition, the anecdotes provided by individuals undergoing hormone treatment suggest that perception of these two cues might be undergoing change (Beard, 1999)[2].

As an attempt to disentangle these cues, much research has aimed to determine thresholds for individual cues or at least analyze them as singular information for depth. Rogers and Graham (1979, 1982) did a great deal of work attempting to determine the thresholds for motion parallax as the sole depth cue, using multi-curved surfaces and discontinuous blocks. Attempts to use shape from shading as the single depth cue have indicated that two assumptions must be guaranteed to get shape from shading results – 1) there is a single light source illuminating the whole scene and 2) the light is shining from "above" relative to the retinal coordinates of the user[3] (Kleffner & Ramachandran, 1992).

Although many combinations of cues have been tested in relationship to one another, there is no published information concerning the relationship between shape from shading and motion parallax.

1.2 How virtual reality environments address depth

"Virtual reality is the simulation of a real or imagined environment that can be experienced visually in the three dimensions of width, height, and depth and that may additionally provide an interactive experience visually in full real-time motion with sound and possibly with tactile and other forms of feedback" (whatis, 1999). Virtual reality environments[4] (VE) range from simple 3D worlds as portrayed on video game systems such as Nintendo64 and Playstation to fully immersive environments such as is used in the CA VE system. Some VE systems include haptic inputs, stereoscopic displays, or wearable devices while other systems only depend on the user and a simple input device such as a keyboard. All VE systems depend on 3D visual displays (which may or may not include stereo) while only some include audio and/or tactile interaction.

1.2.1 Virtual reality depth cues

Conveying depth information in a virtual environment is done using a wide variety of real-world cues. Most VE designers attempt to mimic world information as much as possible, but also factor in computational expense in determining what is essential to be portrayed and how precisely it needs to be displayed. For example, the depth-

[2] In 1999, the author recorded a series of anecdotes from transsexuals undergoing hormone treatment regarding vision changes. This work was an attempt to determine what research is needed to analyze sex differences in depth perception. While unpublished, this overview is available at School for International Training in Amsterdam and through the author at danah@danah.org.

[3] In natural space, most consistent light sources come from above. While the sun is the primary example, electronic lights also take on that property. Shading is also considered to be one of the most primitive cues for understanding 3D,due to the biological adaptation principle (i.e., countershading taken on by many animals to conceal themselves; pale bellies neutralize the effects of the sun shining from above). (Thayer, 1909).

[4] Many researchers prefer to use the term *Virtual Environments* as opposed to *Virtual Reality* because of the media blitz that virtual reality has received. Because of the audience of this paper, I have chosen to use *Virtual Reality* for the title since it is the more commonly understood term. Throughout the paper, I have used a combination of the two. It is important to realize that they are the same term with the same meaning and the only difference is the politicalization of the topic. VE is short for Virtual Environments/Virtual Reality.

accuracy versus computational expense demands of high-speed video games are very different than those of immersive surgical environments. In most VE systems, depth cue information can be derived using motion parallax, some shading cues, and linear perspective. Matte shading information is almost always available and specular shading information is usually available. In textured environments, it is extremely rare for the shading to take into account the information from the texture, but instead uses the information from the graphical surface to which the texture is attached. The texture does not protrude from the surface; thus, it is like wood grain, not fur. With available head-mounted displays, immersive glasses and other devices, stereoscopic depth information is becoming more and more available. Unfortunately, most of these devices do not account for differences in eye separation and thus individuals smaller than the expected wearer get depth information consistent with larger depth differences than would be found in the real world for a given depth difference while larger individuals get information consistent with a smaller depth difference than intended. In order to accurately focus on information at various depths, people adjust the angle each eye is directed, altering the vergence angle. None of the commonly used binocular devices account for altering vergence angles, information that is tied to stereoscopic vision in real environments. Due to practical constraints, only a limited number of VE systems attempt to create vestibular information[5].

Although quite a few cues are available to VE users, it is important to realize that these cues are not entirely realistic and that individuals adapt to the differences. Both single monitor and bi-monitor (i.e., VE goggles) displays construct a 3D scene assuming that the user is a certain distance from the monitor by creating a "camera" and constructing calculations dependent on that camera's position. This is identical to what is done when movies are made. As with computer monitors, if you sit at an odd angle to the movie screen, perspective appears a bit off. Most people learn to adapt to the information presented, even if they are not at perfect viewing distance or angle.

In the case of shape from shading, computational methods are not necessarily implemented in the way in which we actually believe the human vision system appears to address shape from shading. Computational systems invert the illumination equation to acquire the direction of the surface normal at each point in combination with assumptions about the nature of the surface and the source of light (Mingolla & Todd, 1986). While this certainly simplifies the computational expense, it is uncertain how this affects the users.

1.2.2 Simulator sickness

While head-mounted displays and goggles offer an additional and strong cue for depth detection, the disadvantages can be sickening; many users experience "simulator sickness" when working with VE systems, particularly when using head-mounted displays or goggles. Two proposed theories to explain simulator sickness are "Sensory Conflict Theory" and "Postural Instability" (LaViola, 2000).

Cue conflict is the primarily accepted theory, particularly because of a conflict between vestibular and visual cues (Kolaninski, 1995). Research shows that people are quite capable of adapting to worlds where vestibular and visual cues are in conflict. While they are able to adapt, this adaptation takes time and experience and the adaptation time may be dependent on the type of transformation (Welch, 1978; McCauley & Sharkely, 1992). For example, the

[5] Vestibular cues are those that the body feels when it knows it is moving.

Army report only addresses the conflict between vestibular and visual cues but does not address situations where individual visual cues are in conflict (Kolaninski, 1995).

The postural stability theory suggests that simulator illness occurs because the user has not learned enough to maintain the visual stamina necessary to interact with the system. Once the user learns to use the visual information available, s/he will not experience further simulator sickness (Riccio & Stoffregen, 1991).

It is also quite possible that both are accurate theories. One example is that most users attempt to focus by converging their eyes, as one would do normally. Since VE systems do not use eye-tracking mechanisms to refocus the information available, this may cause simulator sickness because the real-world correlation between depth and blur or depth and vergence needed for single vision is missing (Wann et al., 1993). When participants learn to "decouple accommodation and convergence," thus creating postural stability, simulator sickness appears to disappear (Robinett & Rolland, 1992).

Since not everyone experiences simulator sickness, there has also been an attempt to understand some of the factors that might be correlated with simulator sickness. In its report on simulator sickness, the Army suggests that the following are potential factors associated with simulator sickness (Kolaninski, 1995):

Potential Factors Associated with Simulator Sickness in Virtual Environments		
Individual	**Simulator**	**Task**
age	binocular viewing	altitude above terrain
concentration level	calibration	degree of control
ethnicity	color	duration
experience with real-world task	contrast	global visual flow
experience with simulator (adaptation)	field of view	head movements
flicker fusion frequency threshold	flicker	luminance level
gender[6]	inter-pupillary distance	unusual maneuvers
illness and personal characteristics	motion platform	method of movement
mental rotation ability	phosphor lag	rate of linear or rotational acceleration
perceptual style	position-tracking error	self-movement speed
postural stability	refresh rate	sitting vs. standing
	scene content	vection
	time lag /transport delay[7]	type of application
	update rate / frame rate	viewing region

This chart shows that a wide variety of factors potentially contribute to simulator sickness in virtual environments. This indicates that current systems are being designed to meet the needs of only a fraction of the population and that a wide variety of individual factors may make one unable to perform in current virtual environments. While this report does outline potential individual, simulator and task factors that contribute to

[6] The Army's notes on gender indicate that "females may be more susceptible to motion sickness" and that "females exhibit larger FOVs6 illness and personal characteristics many forms of illness may result in increased susceptibility to simulator sickness; possible effect of characteristics such as motivation, goals, or belief of susceptibility?" (Kolaninski, 1995, p. 39). While these are the exact words in the Army report, I believe that they imply that females are more likely to possess multiple individual factors that make someone susceptible to simulator sickness. For example, females are more likely to fall prey to the "illness and personal characteristics" that increase simulator sickness. These notes appear to be based on Biocca's report that women's increased susceptibility to motion sickness may be due to hormonal effects (1992). In addition, the Army's comments reflect Kennedy and Frank's research that women exhibit larger fields of view than men and that seems to result in increased susceptibility to simulator sickness (1983).

simulator sickness, it does not discuss what information in the tasks and or simulator are affecting individuals differently.

Due to the concern that simulator sickness will limit VE systems from becoming ubiquitous, there is quite a bit of research attempting to discover the source of the sickness as well as to test for correlations between vision strengths/weaknesses and simulator sickness. One such suggested correlation is that individuals who have difficulty with visuospatial tasks (i.e., mental rotation) are much more likely to experience simulator sickness (Parker & Harm, 1992). Interestingly, sex differences in visuospatial performance are well documented. Almost all of the psychophysical research indicates that males consistently outperform females (Harris, 1978; Maccoby & Jacklin, 1974; Newcombe, 1982). Sex hormone research indicates that presence of androgens (testosterone) increases one's ability to perform mental rotation tasks (Van Goozen et al., 1995).

Simulator sickness studies are the primary area of VE research that addresses visual human-computer interaction issues. It indicates that VE researchers are learning quite a bit about vision, depth, and the cue conflicts that are created, though most of the tests do not analyze individual cues, but rather infer that cues must be in conflict since people are getting sick.

---

[7] Time lag / transport delay describes the delay between information input to and visual information from the system. An example is the amount of time between when the user gives input to turn and the resultant image with turned visual display appears.

2 Specifically addressed problem and motivation

In order to begin understanding how depth information in virtual reality environments impacts individuals' performance, we analyzed how isolated visual cues impacted one's depth perception as well as how conflicting cues would be interpreted. The virtual environments literature indicated a general problem of cue conflict, even beyond stereo, while the vision literature had begun to approach testing individual depth cues and suggesting ways in which they may be integrated.

For the purposes of our experiment, we decided to use two real world 3D depth cues that are commonly available in all virtual reality systems – shape from shading and motion parallax. While both cues had been studied individually and in conjunction with other cues, the two had never been put into conflict in the available research. The literature suggested that both are strong cues to depth and experiments to test thresholds of the two were partially formulated.

The question of individual and sex-typed differences was suggested because of raising concerns over differences in performance in virtual reality games, sex-typed and individual differences in simulator sickness. In addition, as part of her work at School for International Training, the author acquired a great deal of anecdotal information from transsexuals undergoing hormone treatment that suggests hormonally induced depth prioritization (Beard, 1999). We wanted to determine if we could quantitatively show whether or not cue detection actually differs between individuals.

The goal of this research is to further our understanding of how multiple depth cues interact with the direction focused on a practical application of this information. As virtual environment designers have the explicit goal of mimicking real world to the best of their computational ability, it is crucial that we reach a better understanding of the importance of various cues so as to assure that systems are designed with the concerns of the majority of users in mind, as opposed to the personal experiences of the designers.

3 Methods

3.1 General methods

For this set of experiments, we chose two monocular depth cues – shape from shading and motion parallax – to analyze ways in which depth information is determined. In presenting these cues, we were attempting to eliminate all other potential cues.

3.2 Subjects

In Experiment 1, nine volunteer subjects (5 male and 4 female) and the author participated in the experiment, which took under 30 minutes. All ten subjects were between the ages of 18 and 23 and were college students who work in computer labs. All ten have been extensively exposed to 3D computer graphics through their work. Only the author had seen the experiment prior to participation. All participants had normal or corrected-to-normal vision.

Twenty paid subjects (10 male, 10 female) participated in Experiment 2, which took under 40 minutes to complete. All twenty subjects were between the ages of 18 and 23 and were college students. The background and field of study for these participants were varied with less than 1/3 of them being regularly exposed to 3D computer graphics. None of the subjects had seen the experiment prior to participation. All had normal or corrected-to-normal vision.

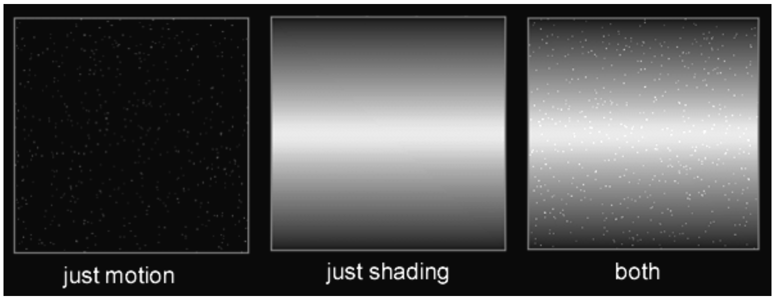
3.3 Experiment design

The images presented showed side-views of cylinders without intensity discontinuities (edges). We provided three display scenarios with similar setups. For each, the subject was presented with 500 milliseconds of a display with the appropriate depth cue. In each trial, two versions of the cue appeared - one 600 pixel by 600 pixel image on each side of the screen. We did this by placing a thick border around the edge of the window. The displays were consistent with shapes viewed through a window. The subject was wearing an eye patch, eliminating binocular disparity, and thus a potential cue conflict. The image is displayed on a 21” CRT display. After being presented with the images, the subjects were required to give a force choice response indicating which presentation appeared more curved in depth.

We provided three scenarios: motion parallax only, shading only, and a combination of motion parallax and shading. The image below represents a single view of all three scenarios. The far left shot is of just motion parallax information[8]; the center shot indicates just shading; and the far right represents a combination of shading and motion information.

---

[8] In the actual experiment, the individual dots are animated.



The range of curvature for Experiment 1 was 5-35 with 20 being the standard. For Experiment 2, the range was 5-55 with 30 being the standard. The curvature was used to determine how each image or animation is constructed. The number was based on how many times the "unit cylinder" it represented. 300 pixels (approximately 3¼ inches) is the unit radius for a unit circle on our monitors. Thus, the standard image for Experiment 2 of 30 represents a cylinder almost 100 inches in relative depth.

3.3.1 Motion parallax design

In the first scenario, a random-dot display is used to give the impression of movement. Six hundred dots were distributed randomly on each cylinder to construct a flow-field model pattern when animated (Julesz, 1971; Dosher et al., 1989). The animated flow fields move in opposite directions of one another. The dots move fasted in area near the center of the cylinder, which are the areas closest to the subject. The dots' speed was dependent on the curvature of the cylinder (i.e., the more curved the cylinder, the greater the difference in speed between dots at the edges and dots at the center of the image). In addition, a constant speed was added to all of the dots so that the subject would not just compare the relative speeds. Similar experiments using an oscilloscope indicated motion parallax is an adequate cue by itself (Rogers & Graham, 1979). While motion parallax through animated flow fields is not indicative of realistic motion parallax, it has shown to be an effective test of motion parallax (Adelson et al., 1990; Braunstein and Tittle, 1993; Mukai and Watanabe, 1999).

On each screen update, a dot is moved a number of pixels dependent on the curvature of the cylinder and the dot's vertical position in the cylinder. The equation for this movement is:

$$\textbf{\# of pixels to move when animating} = \textbf{sqrt (curvature} - \textbf{(curvature * ((relative y position)}^2\textbf{))}$$

where curvature is the aforementioned number ranging from 5-35 or 5-55

In Experiment 2, a constant between 0 and 1.6 was added to the number of pixels for every dot on the screen, as to give some alteration to the apparent movement without altering the apparent relative depth of the cylinder. The additional constant was consistent for each pixel on the screen but each image had a different constant. The fastest moving dots on the screen were always located at the absolute center of the screen, the closest position of the cylinder. For a curvature of 5, the center dots moved approximately 2.23 pixels per update (+ the random constant).

For a curvature of 20, they moved 4.47 pixels; for 30, 5.48; for 55, 7.42. This allowed for a wide range of apparent curvatures.

3.3.2 Shading design

In the second situation, the user was presented with pre-rendered images of shaded cylinders. For Experiment 1, *Radiance* cylinders were modeled based on the dimensions of the cylinder where the depth and vertical length ratios varied from 1:1 to 8:1 with the image rendered broadside facing the light source. In Experiment 2, the images were calculated manually. Based on tests with a photometer, we used images in line with the gamma correction of our system, determining the appropriate luminance based on the cylinder's properties and the curvature of the desired cylinder.

$$\textbf{luminance} = \mathbf{x / sqrt(x^2 + (curvature^2) * y^2)}$$

where $x = \mathrm{sqrt}(1-y^2)$ and y is the vertical axis

In Experiment 2, constants were added so that the images were of a reasonable darkness with some constant variation so as to eliminate one's ability to compare based on brightness at a standard point.

3.3.3 Cue conflict design

In the final scenario, we combined the two different cues in the same image/animation. In each case, the image on one side had the standard animation and the standard shading. Presented cues were in line if the non-standard image had shading and motion that both indicated more curvature or less curvature than the standard. In-line cues did not necessarily have to be representing the same magnitude of curvature, only a direction of curvature from the standard. Cues were in conflict when one cue indicated more curvature and one cue indicated less curvature than the standard. In Experiment 1, either the cues were in line or in conflict; in Experiment 2, we only used conflicting cues. This created an impossible scenario whereby the information given by the two clues was conflicting. By providing contradictory shading and motion parallax information simultaneously, the subject was forced to choose between the two. While their choice could be for a variety of reasons, research suggests that one would choose a cue because of stronger weighting or a cue dominance (Bülthoff, and Mallot, 1988).

In the first two scenarios, we gathered information about how accurate the user is at using the presented cue to estimate curvature and thus depth, using a staircase method. To determine the threshold of the individual, we displayed one image of the standard and one of the extreme in curvature. With two accurate answers, the extreme image moved closer to the standard, with the initial step size equaling four deviations from the current curvature. With a wrong answer, the step size was halved, with a minimum stepsize equal to one deviation of the cylinder. Given a wrong answer, the next image was of curvature a stepsize away from the standard, so as to narrow down what people were capable of distinguishing. When the subject registered eight changes in direction (from closer to further away from the standard or vice versa), the program stopped. The last five curvatures where a change in direction occurred were averaged to determine the threshold.

In the third trial series, we determined how important each cue was to the user depending on which cue they chose in a conflicting stimulus experiment. A preset series of conflicting numbers were used. For Experiment 1, the curvatures used were 8, 14, 26, 32 with a standard of 20. In other words, potential images could have been shown with a motion curvature of 8 and a shading curvature of 26 or any other conflicting curvatures where one is greater than the standard and one is less than the standard. In Experiment 1, only one of each combination was shown. For Experiment 2, curvatures used were 6, 14, 22, 32, 40, 48 where the standard is 30. In this version, 10 trials of each combination were scored.

4 Results and data analysis

There were two different types of analysis. First, it was important to determine the individual and average thresholds for single cue performance. The last five reversals were averaged to determine what the threshold was for that particular cue. For conflicting cues, we created a data matrix of cue choice frequency. By analyzing the data matrix along individual and sex lines, we determined how significant individual differences were as well as were able to analyze for sex-typed differences.

4.1 Experiment 1

In the threshold part of the experiment, individual variation was apparent, but no statistically significant differences were found based on either individual differences or sex differences. The chart below represents the average threshold for subjects in each area with the standard error information in parentheses. This data represents the average for 4 females and 5 males. For comparison, the standard in this section is 20 while the extremes are 5 and 35. The closer that the subjects' data is to the standard, the stronger the performance.

	Motion – Lower	**Motion – Upper**	**Shading – Lower**	**Shading – Upper**
Female	12.85 (0.74)	29.95 (0.50)	5.95 (1.03)	33.35 (0.82)
Male	13.56 (0.58)	30.70 (0.68)	6.16 (0.36)	27.44 (0.63)
Total	13.24 (0.46)	30.37 (0.43)	6.07 (0.45)	30.07 (0.50)
Author(db)	16.60 (1.34)	30.60 (1.52)	8.40 (0.55)	24.60 (0.55)

The author's data was removed from the group data because her results indicated that practice or other types of experience seemed to play a role in ability. In addition, these results indicate that the lower shading images are not strong cues; all other areas seemed to be reasonably strong.

When the cues were put into conflict, males chose shading 32% of the time while females opted for shading 55% of the time. The individual variation on this part of the study was extreme though not consistent. The following chart shows how frequently subjects chose the default (shading: 20, motion: 20). Each pair of cues was presented to each subject twice. For the most point[9], this meant that each combination's result was based on 18 total trials. In the upper left hand quadrant, the subject should have chosen the default; in the lower right hand quadrant, the subject should have chosen the non-default option. In the center cell, both sides showed the default, thus either answer is correct. These choices are expected since both options convey the same information. In the remaining quadrants, a cue conflict was presented. Thus, in the lower left hand quadrant, choosing the default suggests that shading was more dominating information. In the upper right hand quadrant, choosing the default indicated a

[9] An error in recording occurred where a total of 1 trial per user was not recorded, meaning that only 49/50 total trials actually recorded. Thus, some of the boxes only had 17 trials from which to average.

preference for the motion cue. All users combined are shown here to indicate the overall choice, which is not dissimilar to the typical responses of each individual.

		Motion Parallax Curvature				
		8	**14**	**20**	**26**	**32**
Shading Curvature	**8**	76.5%	38.9%	61.1%	50.0%	33.3%
	14	88.9%	88.9%	55.6%	44.4%	27.8%
	20	50.0%	64.7%	27.8%	38.9%	22.2%
	26	64.7%	70.6%	22.2%	29.4%	33.3%
	32	72.2%	43.8%	61.1%	33.3%	27.8%

The primary accomplishment in Experiment 1 was to show that the stimulus range chosen was slightly narrow, though observers were able to use both cues reliably. An important adjustment to the methodology was the need for adding constants to both the motion parallax and the shading to eliminate people's ability to judge depth based solely on speed or luminance of mapped positions. The other methodological alteration made apparent was the recognized need for multiple trials of identical conflicting scenarios, as opposed to the one of each combination used in Experiment 1. This adjustment is necessary to make it possible to evaluate individual performance. Because of this and the indication that users are usually able to perform well when given relatively related curvatures, we chose to eliminate that portion from our experiment. Based on the difference in capability between the inexperienced subjects and the author, we suspected that experience or background affected subjects' ability. Thus, we decided that a completely naïve subject pool was more appropriate.

4.2 Experiment 2

Based on knowledge gained from Experiment 1, we randomized the luminance and speed values by adding constants to both the shaded images and the motion animations, so as to eliminate the ability for individuals to determine information by directly comparing localized 2D cues. In addition, we used only naïve subjects who were not familiar with the experiment.

Most individuals' thresholds for the single cue segment were close to the extremes (30 is the standard comparison; 5 / 55 are the extremes), indicating poor sensitivity to the changes in curvature. The only portion where performance indicated capability was in the lower bounded shading cues. Since the curvatures used in this experiment were linear multipliers of the unit cylinder, the standard was six times the lower extreme curvature while the upper extreme was less than twice the curvature of the standard. This may explain the greater difficulty in working with upper extreme curvatures. Poor threshold information indicated that the cues were not strong enough for the subject to make an accurate judgment about depth.

The chart below represents the average threshold for subjects in each area with the standard error information in parentheses. This data represents the average for 10 females and 10 males. For comparison, the standard in this section is 30 while the extremes are 5 and 55. The closer that the subjects' data is to the standard, the stronger the performance.

	Motion – Lower	Motion – Upper	Shading – Lower	Shading – Upper
Female	9.48 (0.61)	52.95 (0.31)	22.37 (0.59)	52.97 (0.38)
Male	11.77 (0.35)	51.72 (0.53)	28.27 (0.37)	53.17 (0.24)
Total	10.65 (0.34)	52.33 (0.22)	25.32 (0.34)	53.07 (0.22)
Author (db)	16.00 (1.63)	46.00 (0.82)	24.33 (4.19)	51.67 (2.13)

The cue conflict data indicated that our subjects did not have a consistent idea of what appeared more curved. Each subject had ten trials of each combination of cues. This chart represents the average frequency in which users opted for the default image given each combination of cues. In the upper right quadrant, choosing the default suggests a preference for shading while preferring the default in the lower left quadrant would suggest a preference for the motion cue.

		Motion Parallax Curvature					
		6	**14**	**22**	**32**	**40**	**48**
Shading Curvature	**6**				91.0%	54.0%	80.5%
	14				86.5%	39.0%	79.5%
	22				83.5%	35.0%	78.5%
	32	91.0%	93.5%	93.5%			
	40	87.5%	95.0%	92.0%			
	48	88.5%	89.0%	87.5%			

Given this data, an actual preference for a given cue seems non-existent. Instead, it seems as though the subject regularly prefers the default combination. This may be a preference for the image where shading and curvature match. Since, in the first experiment, subjects indicated poor understanding of the upper threshold motion and shading information, this data may also mean that subject was constantly comparing the lower end information with the default image as opposed to comparing both sets of information (i.e. in the upper right quadrant, only comparing shading and only motion in the lower left quadrant). It was difficult to determine what the subjects were actually doing, but we know that they were not being consistent and that the cues were not strong enough for accurate judgment.

5 Discussion

5.1 Summary of results

In Experiment 1, we determined that subjects prefer to use available 2D cues such as luminance and speed over the intended 3D cues. In addition, we noticed that experience played a factor in ability, with the author outperforming the naïve subjects. When 2D cues were eliminated, as in Experiment 2, subjects performed atrociously, even with a wider range of curvatures. Experience with our system, experience with 3D computer graphics, and heavy weighting of 2D cues all appeared to play a role in performance. Due to the irregular results, analyzing this data for individual differences was not possible.

In literature with similar setups and experiments, subjects frequently were the authors or graduate students, most likely in the department doing the study. This is generally considered acceptable since the performance of those who designed the system is usually on par with individuals who are given learning tasks to adjust to the system. This suggests that little work has been done to eliminate the experience variable. In addition, it is uncertain but unlikely that 2D cues such as speed and luminance were randomized or otherwise removed.

The result of these studies indicates that much of the research in the literature may not be testing what they suggest they are testing or that tests with experienced participants do not accurately represent the general population.

5.2 Importance of practiced observers

In our experiment, the author participated in both experiments for comparison. (Her data is listed separately and not factored into the averages.) In addition, the participants in each experiment had different levels of background with 3D computer graphics. Those participating in Experiment 1 were all students in computer science, most with a focus on computer graphics, while subjects for Experiment 2 were randomly selected from the undergraduate community where less than ¼ of them were students in computer science. We do not know their level of exposure to 3D worlds in other contexts. Our results and the author comparison indicate that experience, practice, or both may impact performance.

The literature suggests that subjects are able to adjust to information presented, both in the current task and over an extended period of time. Jacobs & Fine (1999) show that subjects can learn to ignore individual cues as well as learn methods to combine the information. If subjects can learn to ignore individual or irrelevant cues, they can probably choose to ignore specific cues altogether when conflicting information arises. This means that practiced observers are more likely to extract the information that they are expected to use and successfully learn to eliminate the randomized 2D cues. In addition, based on personal experiences, an individual is most likely able to self-weight specific cues. If their experiences change or their weighting is no longer an accurate indicator, they are most likely to adjust to the new information presented by altering how they address and combine cue information.

The virtual environments literature on simulator sickness suggests that users appear to be able to avoid getting negative reactions through practice, yet they often have difficulty readjusting to the real world. This is probably due to their ability to learn to ignore conflicting cues presented or to adjust to ways in which the new information conveys the desired information. Over extended periods of time, users experience minimal simulator sickness, even on new systems with similar layouts, which suggests that experienced users adjust their method of weighting available information and can transfer this information to other similar systems. This explains why high performance at one video game can be transferred to strong ability at another when their graphics are similar in nature.

If one can alter one's weighting or can visually adjust to a system, virtual environment designers are not the most appropriate testers for new systems intended to address the general population. In these scenarios, those with the least amount of experience in computer graphics and VE are the most appropriate subjects to determine what level of learning is necessary to be able to adjust to a new system.

In our experiment, it is possible that both experience with the system and experience with similar graphical setups were aids in performance.

5.3 Importance of 2D cues

While it is impossible to remove 2D cues such as luminance and speed from our experiment, we failed to randomize those factors in Experiment 1 but corrected this problem in Experiment 2. Randomization does not eliminate the 2D cues but makes it virtually impossible to use this information and acquire accurate depth perception. The literature and our results indicate that shape from shading and motion parallax may not be nearly as strong without their associated 2D cues as we originally suspected.

Research suggests that computer simulated shape from shading gives minimal depth information and is confined by the assumption that there is a single light source illuminating the entire scene and that the light source comes from above relative to retinal coordinates (Kleffner & Ramachandran, 1992). In a natural environment, this is almost always true. Because this is what we are most accustomed to, we have learned to assume these conditions when using shape from shading, making it difficult to adjust otherwise. (For example, lower lit faces appear distorted and "scary.") Might we learn shape from shading in one context only? Situations where different elements in a natural environment are lit differently are practically non-existent. Because of these assumptions in the natural world, a direct luminance comparison usually results in accurate information regarding depth information. Conflicting cues such as those experienced in our experiment and virtual environments are not found in real space and thus our brain is probably not prepared to handle those situations with ease. Due to the simplicity of doing a luminance comparison, it is quite probable that SFS is not usually used, but luminance comparison instead. With life experience, knowing what luminance values regular objects are expected to have makes this process even easier.

Based on our findings regarding motion parallax and the differences in performance between Experiment 1 and Experiment 2 for both the "naïve" subjects and the author, we suspect that individuals performed a 2D speed comparison test as opposed to a 3D motion parallax depth test. While luminance is regularly compared in world space for depth comparisons, comparing relative speeds is more problematic and, at first, seems less likely.

The most common use of speed would be to determine the speed from self. We are quite able to detect the amount of speed produced in our visual system either by the movement of our eyes or the movement of our bodies. Determining the speed of a moving object is more difficult since objects in a space have little regularity both in speed and structure. In addition, most objects don't have a common speed for mental reference acquired through experience. Again, practice and familiarity with a given object does help, as in the case of a baseball player who can accurately detect distance of a moving ball at high speeds.

In the case of relative depth onto itself, differences in speed would seem to be even less likely to be a common tool for comparisons. Most objects to be compared are different, thus not having the same markers for comparison. More probably, as Watamuniuk and Duchon (1992) suggest, the mean speed of the stimulus is what is used for comparisons; other stimulus characteristics, such as mode, are ignored. In Experiment 2, the task necessary was a comparison of mode to min for each object. Using mean speed would have eliminated the depth information of the specific cylinder and would not be an accurate indicator of either the object's self-depth or the distance from the viewer.

6 Future work

Future work must be developed both within the vision and virtual environments communities. Rather than get at the core questions concerning the author in this project, a great deal of time was spent addressing problems within previous research and questioning the premises on which most theory in these communities rest. As a result, there is much left to do before the questions concerning this research can be adequately addressed.

6.1 Directions for the vision community

Currently, the role of learning or adapting visually is under-addressed or under-considered in cue conflict and cue threshold experiments. The four primary theories of cue interaction do not take into account one's ability to adapt to certain cues and potentially to reprioritize the incoming information. In order to better address how cue interaction occurs, it is necessary to factor in one's ability to learn cues. It would be quite advantageous to test the linear combination work with naïve and practiced observers and determine what level of control over prioritization or weighting the individual has. How does this relate to the real world? Is prioritization a conscious choice? Or biological? Based on social experience? Why or why not?

6.2 Directions for virtual environment designers

In order to actively make systems that are designed for the users, experiments must be conducted using the population for which the system is intended. Self-testing a system or testing based on whether or not it appears realistic to the designers does not constitute proper testing. We have learned that our visual system is capable of learning how to understand various cues and how to separate out information that is deemed unnecessary or conflicting. Those who actively work with a system are not able to detect the problems with that system since they have most likely successfully adapted. When designing a test, it is crucial to consider the population at hand, the types of cues that they will observe (and in particular, what might present conflicting information) and the amount of experience that they are expected to have. Only by considering these three factors can a properly designed experiment exist for these purposes.

Learning to dissect the information made available through computer graphics is not dissimilar to what humans do in real space. As children, learning to manipulate through the world is quite a chore. Until we learn what the information presented means, we run into things, fall down and are otherwise clumsy. This pattern repeats itself at puberty when our bodies adjust and our mental calculations need to readjust with the changes. Eye-hand coordination is a learned skill and tremendously associated with understanding depth information. It is quite reasonable to assume that users will perform better if they are accustomed to 3D computer graphics, but doing so may weaken individuals' ability to manipulate the natural world. Factors such as simulator sickness and need for adjustment indicate that virtual environments are not activating the same visual information as real world interaction. Having to adjust between real world and virtual world cue information is either problematic for the long-term feasibility of current virtual reality systems or dangerous for humans. It is our opinion that making the perception information in virtual environments as close to that in the real world as possible is far more crucial to the long-term success and health of virtual reality than training people to adjust because of computational reasons. In other words, having simpler systems with accurate visual information is better for the longevity of virtual reality.

6.3 Other questions and concerns

What role does individual variation play in depth perception? How can we explain simulator sickness differences based on sex, ethnicity, age and other personal characteristics? What biological differences exist within individuals to prompt such responses? Or is it pure social experience and practice?

If learning is crucial to success in estimating depth in VE, individual differences in performance may be heavily related to experience, which in turn is strongly affected by social issues. A wide variety of surveys and studies from Australia to the American Association of University Women have shown that adolescent girls have less access to computers and the web than their male peers. Class (and thus race) is also a factor in computer access. With performance linked to experience, this problem quickly spirals out of control.

There are limitations to the presentation of shape from shading, including the need for a single light source coming from above (Kleffner & Ramachandran, 1992). What are the differences between SFS on the computer screen and in real space? Do people just compare average luminance values in the real space?

Do people always average speed and use the average for comparisons as opposed to accounting for motion parallax? (Watamuniuk & Duchon, 1992).

7 Conclusion

While the initial goal of this project and research was to address what potential individual differences may exist, doing so within the realm of this project was impossible. The framework for this area of research has not been developed far enough to make this level of detailed analysis feasible. Instead, I have started to dismantle the current framework, showing faults with previous research in this are. In addition, I have given some suggestions as to how to rebuild a more stable version of this framework, so as the research suggested in this paper can actually be completed. By bridging two areas of research while they are still in their embryonic stages, I hope to indicate that this research cannot remain in a theoretical vision vacuum while computational communities are attempting to utilize it for use by the masses. As virtual environments become larger players in the computation community, it is vital that scientists working on these projects support and further the human-computer interaction research focused on depth perception, particularly as is utilized by the VE systems. Without that support or focus, the possibility of implementing systems that are potentially harmful, or create hierarchies in capability because of social or biological backgrounds, is much too great. This is only the beginning of research that must go much deeper in order to fully realize the implications of current systems and the direction that these fields appear to be going. Nonetheless, it is a step in the right direction and one that is crucial for both the vision and virtual environments communities.

8 Bibliography

Adelson, E.H., Hildreth, E.C., Inada, V.K., Grzywacz, N.M. (1990). The Perceptual Buildup of Three-Dimensional Structure From Motion. *Perception & Psychophysics*, 48(1), 19-36.

Beard, Danah (1999). Sex-based vision through the eyes of transsexuals undergoing hormone treatment. School for International Training, Amsterdam: unpublished. Copy available at http://www.danah.org/papers/transtech.pdf

Biocca, F. (1992). Will simulation sickness slow down the diffusion of virtual environment technology? *Presence*, 1(3), 334-343.

Blake, A. & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature*, 343, 165-168.

Braunstein, M.L. & Tittle, J.S. (1993). Recovery of 3-D Shape from Binocular Dispariy and Structure from Motion. *Perception & Psychophysics*, 54(2), 157-169.

Bruno, N. & Cutting, J.E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, 117, 161-170.

Bülthoff, H.H. & Mallot, H.A. (1988). Integration of Depth Modules: Stereo and Shading. *J. Opt. Soc. Am. A*, 5(10), 1749-1758.

Cutting, J.E. & Vishton, P.M. (1995). Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers, *Perception of space and motion*. San Diego, CA: Academic Press.

Dosher, B.A., Sperling, G., Wurst, S. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, 26, 873-990.

Dosher, B.A., Landy, M.S., Sperling, G. (1989). Kinetic Depth Effect and Optic Flow I. 3D Shape from Fourier Motion. *Vision Research*, 29(12), 1789-1813.

Harris, L.J. (1978). Sex Differences in spatial ability: possible environmental, genetic, and neurological factors. In M. Kinsbourne (Ed.), *Asymmetrical function of the brain*. Cambridge, UK: Cambridge University Press.

Interrante, V. (1996). Chapter Three: Perceiving and Representing Shape and Depth. *Illustrating Transparency: communicating the 3D shape of layered transparent surfaces via texture*. University of North Carolina, Chapel Hill: unpublished PhD thesis, 335-382.

Jacobs, R.A. & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research*, 39, 4062-4075.

Johnston, E.B., Cumming, B.G, Parker, A.J. (1993). Integration of Depth Modules: Stereopsis and Texture. *Vision Research*, 33(5/6), 813-826.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.

Kennedy, R.S. & Frank, L.H. (1983). A review of motion sickness with special reference to simulator sickness. Paper presented at the National Academy of Sciences/Nation Research Council Committee on Human Factors, Workshop on Simulator Sickness, 26-28 September 1983, Naval Postgraduate School, Monteray, CA.

Kleffner, D.A. & Ramachandran, V.S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, 52(1), 18-36.

Kolaninski, E. (1995). Simulator Sickness in Virtual Environments. *US Army Research Institute for Behavioral and Social Sciences Technical Report 1027*. http://www.cyberedge.com/4a7a.html.

LaViola, J. J. (2000). A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bulletin*, 32(1), 47-55.

Maccoby, E.E. & Jacklin, C.N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Maloney, L.T. & Landy, M.S. (1989). A statistical framework for robust fusion of depth information. *Proceedings of the SPIE: Visual Communications and Image Processing*, Part 2, 1154-1163.

McCauley, M.E. & Sharkey, T.J. (1992). Cybersickness: Perception of self-motion in virtual environments. *Presence*, 1(3), 311-318.

Mingolla, E. & Todd, J.T. (1986). Perception of Solid Shape from Shading. *Biological Cybernetics*, 53, 137-151.

Mukai, I. & Watanabe, T. (1999). The Influencec of Structure from Motion on Motion Correspondence. *Perception*, 28, 331-340.

Newcombe, N. (1982). Sex-related differences in spatial ability: problems and gaps in current approaches. In M. Potegal (Ed.), *Spatial Abilities: development and physiological foundations*. New York: Academic Press.

Parker, D.E. & Harm, D.L. (1992). Mental Rotation: A Key to Mitigation of Motion Sickness in the Virtual Environment? *Presence*, 1(3), 329-333.

Riccio, G.E. & Stoffregen, T.A. (1991). An ecological theory of motion sickness and postural instability. *Ecological Psychology*, 3(3), 195-240.

Robinett, W. & Rolland J.P. (1992). A computational model of the stereoscopic optics of a head-mounted display. *Presence*, 1(1), 45-62.

Rogers, B.J. & Collett, T.S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology*, 41A, 697-717.

Rogers, B. & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*. 8, 125-134.

Rogers, B. & Graham, M. (1992). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*. 22, 261-276.

Thayer, G.H. (1909). *Concealing coloration in the animal kingdom*. New York: MacMillan.

Van Goozen, S.H., Cohen-Kettenis, P.T., Gooren, L.J., Frijda, N.H., Van de Poll, N.E. (1995). Gender Differences in Behaviour: Activating Effects of Cross-Sex Hormones. *Psychoneuroendocrinology*, 20, 343-363.

Wann, J.P., Rushton, S.K., Mon-Williams, M. (1993). What is wrong with your head-mounted display? *CyberEdge Journal*, supplement.

Watamaniuk, S.N.J. & Duchon, A. (1992). The Human Visual System Averages Speed Information. *Vision Research*, 32(5), 931-941.

Welch, R.B. (1978). *Perceptual modification: Adapting to altered sensory environments*. New York: Academic Press.

whatis (1999). Virtual Reality. *http://www.whatis.com/*, http://www.whatis.com/virtualr.htm.