UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
The role of focus cues in stereopsis, and the development of a novel volumetric display

Permalink
https://escholarship.org/uc/item/2nh189d9

Author
Hoffman, David Morris

Publication Date
2010

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The role of focus cues in stereopsis, and the development of a novel volumetric display

by

David Morris Hoffman

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Vision Science

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Martin S. Banks, Chair
Professor Austin Roorda
Professor John Canny

Spring 2010

Abstract

The role of focus cues in stereopsis, and the development of a novel volumetric display

by

David Morris Hoffman
Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor Martin S. Banks, Chair

Typical stereoscopic displays produce a vivid impression of depth by presenting each eye with its own image on a flat display screen. This technique produces many depth signals (including disparity) that are consistent with a 3-dimensional (3d) scene; however, by using a flat display panel, focus information will be incorrect. The accommodation distance to the simulated objects will be at the screen distance, and blur will be inconsistent with the simulated depth.

In this thesis I will described several studies investigating the importance of focus cues for binocular vision. These studies reveal that there are a number of benefits to presenting correct focus information in a stereoscopic display, such as making it easier to fuse a binocular image, reducing visual fatigue, mitigating disparity scaling errors, and helping to resolve the binocular correspondence problem. For each of these problems, I discuss the theory for how focus cues could be an important factor, and then present psychophysical data showing that indeed focus cues do make a difference.

Next, I describe a new approach to construct a stereoscopic display that presents these signals correctly. This new display uses a custom lens system to present correct focus information in a time-multiplexed manner. This approach has significant advantages over other volumetric display approaches, and is a significant advancement in stereoscopic display technology.

In addition to the technical development of the volumetric display, I have also developed a way of analyzing volumetric displays to assess the retinal image quality of objects in various positions within the display. This model of image formation in a volumetric display allows us to test how various properties of the viewer, display and stereoscopic content will change the image quality. The main conclusion of this analysis is that a volumetric display requires only a coarse depth resolution to create stereoscopic images that are a close approximation of real-world images.

Contents

List of Figures

List of Tables

Acknowledgments

I want to thank my committee members for their help preparing this thesis. Martin Banks offered excellent direction on selecting promising projects, and was full of insight for effective scientific writing. Austin Roorda guided me through the optical modeling. He was also generous in allowing me to base the wavefront analysis on his optical software tools, and aberration measurements collected using his aberrometer. John Canny offered some excellent suggestions on ways of unifying the individual projects into a comprehensive thesis.

I would also like to thank my many collaborators on the research composing this thesis. Kurt Akeley developed the original 3-mirrors volumetric display that was an important tool for much of the psychophysical parts of this work. I also had many conversations with him about effective ways for programming the software to drive the switchable lens display. The switchable lens display was conceptualized and constructed by Gordon Love with the help of Andrew Kirby from the University of Durham, UK and Phillip Hands at Cambridge, UK. I am also grateful for having the opportunity to work with Kevin MacKenzie and Simon Watt to validate the wavefront analysis with physiological measures collected at University of Wales, UK. I would also like to thank my past and current lab mates for their assistance in surmounting the many technical challenges I have faced in this work, Bjorn Vlaskamp, Robin Held, James O'Shea, Amanda Alvarez, Emily Cooper, James Gao, Joohwan Kim, Takashi Shibata, Johannes Burge and Ahna Girshick.

I am also grateful for the support of my family and friends.

Chapter 1

Focus information in stereoscopic displays

The principle of stereoscopic imagery seems deceptively simple. To produce a vivid and immersive depth experience, present slightly different images on a display screen, don the funny glasses, and voila—objects seem to leap from the screen. But, are the objects depicted in the scene perceived as they would in the real world? Usually not. Even if the disparity information is calculated carefully and the viewer is positioned at the correct distance from the display, the visual experience in the display can differ dramatically from viewing the equivalent real world scene. Compared to natural viewing, interpreting the 3d depth in a display is challenging: People often have trouble fusing the images, the scene frequently appears distorted, and prolonged viewing can lead to headaches and fatigue. What do these stereoscopic images have or lack that cause these problems? Is it possible to construct a display where these problems will be absent?

One artifact in typical stereo displays that could be related to these problems is incorrect focus information. The focal properties for the real world, and carefully constructed scenes displayed in a typical stereo display differ greatly. One of these differences is that in the real world, the lens of the eye focuses to a distance and objects at that distance form a sharp image on the retina. In stereoscopic images, this is typically not the case; one must focus to the display surface to see an image clearly, irrespective of its simulated distance. Furthermore, this means that normal differences in blur (blur gradients) for real-world images will be incorrect in a stereo display.

In this dissertation I will explore whether focus information contributes to the problems in viewing stereoscopic images, and will describe a project to construct and analyze a new type of stereoscopic display system that stimulates correct focus information.

1.1 The problem with traditional stereoscopic displays

Consider two viewing situations: a complex real scene viewed binocularly and a stereoscopic computer display of the same scene. The computer display is carefully constructed so all the standard depth cues–binocular disparity, texture gradients, occlusion, shading, etc.—are geometrically correct. Hence the geometric patterns of stimulation striking the two eyes from the scene and display are the same. Psychophysical research [6, 19, 36, 44, 137] and experience with virtual-reality displays [31, 108, 142] suggest that the perceived 3d structure will differ: the depth in the computer display will generally appear flattened relative to the real scene from which it is derived (there are also cases in which the perceived depth will be exaggerated). In computer displays, focus cues—accommodation and blur in the retinal image—specify the depth of the display rather than the depicted scene. Those cues to flatness may affect depth percepts. Conventional 3d displays also create unnatural conflicts between vergence and accommodation. Those conflicts may affect the ability to fuse binocularly and may cause visual fatigue.

These conflicts are seldom found in real-world stimuli where numerous cues all specify the same depth. Conversely, computer displays present images on one surface: e.g., the phosphor grid for cathode-ray displays (CRTs). As a consequence, stimuli presented on computer displays contain some cues that specify the depth intended by the graphics engineer (we call these simulated cues) and some cues created by the display screen (screen cues) that specify the relative depth of the screen (namely, that it is flat) rather than the intended depth. Screen cues include motion parallax due to the viewer's head movements relative to the screen, and visible pixelization due to the discrete nature of the screen, but we concentrate here on one class of screen cues—focus cues—because they are likely to cause perceptual distortions [19, 44, 137] and viewer fatigue [129, 134], and because they have proven difficult to eliminate[1]. Focus cues come in two forms.

1. Blur gradient in the retinal image. With real scenes, retinal blur varies consistently with changes in scene depth: the retinal image is sharpest for objects at the distance to which the eye is focused and blurred for nearer and farther objects (Figure 1.1, lower left). The correlation between blur and depth in real scenes aids depth perception [80, 86, 88, 89, 90, 100]. In computer displays, focal distance is constant so all retinal-image points are sharp if the viewer's eye is focused at that distance and blurred if the eye is focused elsewhere. Consequently, the blur gradient produced by viewing a conventional 3d display specifies flatness (Figure 1.1, lower right).

2. Accommodation. As the eye looks around a real scene, neural commands are sent to the lens muscles to change focal power and thereby minimize blur for the fixated part of the scene. This is a depth cue [94, 132]. As the eye looks around the simulated scene in a 3d display, the focal distance of the light does not vary, so the neural commands are constant and signal less depth (Figure 1.1).

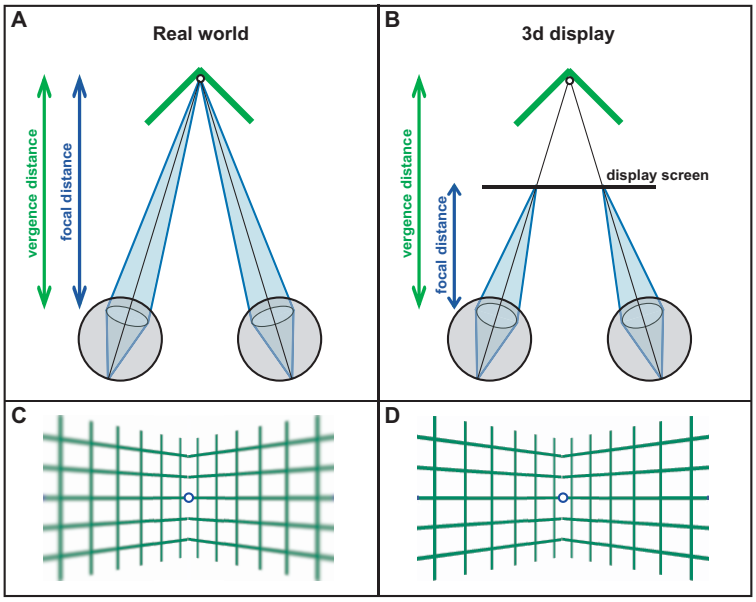


Figure 1.1: Vergence and focal distance with real stimuli and stimuli presented on conventional 3d displays. A) The viewer is fixated and focused on the vertex of a hinge. Vergence distance is the distance to the vertex. Vergence response is the distance to the intersection of the eyes' lines of sight. Focal distance is the distance to which the eye would have to focus to create a sharply focused image. Accommodative response is the distance to which the eye is accommodated. B) The viewer is fixated on the simulated hinge vertex on a computer display screen. Vergence distance is the same as in A. Focal distance is now the distance to the display. C) The appearance of the stimulus when the viewer is accommodated to the vertex of a real hinge. In the retinal image, the sides of the hinge are blurred relative to the vertex. D) Appearance when the viewer is accommodated to the vertex of a simulated hinge. The sides and vertex are equally sharp in the retinal image.

1.2 Displays with correct focus information

Producing a display that stimulates appropriate accommodation and retinal blur is challenging. With a conventional stereo display, it is possible to produce pictorial blur in the image that closely emulates retinal blur from defocus. However, computing this blur accurately requires making assumptions about where the observer is fixating, accommodating and knowing their pupil size. Even if all of this information was known the display remains incapable of presenting appropriate accommodative stimuli.

Alternatively, a class of displays that presents perfect focus information is the autostereoscopic volumetric display. It presents focus cues exactly correct by building a 3d model of an object out of 3d light units, either by sweeping a display surface through a volume or by using a number of light-scattering surfaces[39, 119]. In addition to displaying accurate focus cues, these displays support multiple viewers simultaneously. It has four main drawbacks:

1. Its resolution is measured in *voxels*, 3d pixels. As a result, instead of just needing the number of pixels to cover a display surface, it needs enough voxels to fill the entire volume, an exponential increase in bandwidth.

2. The display is additive; each voxel can add light but not occlude it. Therefore objects have a ghost-like appearance and do not appear realistic. Occlusion is a powerful depth cue, and incorrect occlusion can lead to severe spatial misperceptions. The failure of these displays to correctly handle occlusions makes them well suited for showing wireframe models, but poor at showing surfaces.

3. The display typically does not know the viewer position so it cannot compute view dependent effects such as reflections or specularities.

4. Because it is based on voxels, it can only present objects that fit inside the display's workspace volume; it cannot act like a window and show things at a distance.

An approach that combines the benefits of 2d displays with focus cues is the fixed-viewpoint volumetric display. A computer calculates a 2d projection of a 3d scene for a known camera position, like a conventional perspective rendering. Then, instead of presenting the image at a uniform focal distance, it modulates the focal length of the image in a spatially dependent manner such that pixels associated with objects at differing distances have different focal distances.

To determine the appropriate focal distance for each pixel, the display uses the *Z-buffer*, which specifies the depth of the geometry responsible for each pixel. In modern computer graphics the Z-buffer information is frequently calculated (as it is important for computing fog and other distance-varying effects), but this information is typically discarded once the 2d rendering is complete. Using some simple transformations, this Z-buffer can be converted

into the focal distance of each object in the scene. Thus the rendering engine can be modified to produce a 2d image projection with a depth map for each pixel in the image.

For a volumetric display, this type of parameterization of a scene is ideal. The visual system has exquisite spatial resolution approaching 50 cpd [24], and excellent color discrimination. Thus the 2d image projection should be presented with as high a resolution, and color fidelity as possible. Conversely, the visual system has quite poor resolution for focal depth (capable of focal discriminations of less than 0.33D [22, 26]). Thus the precision of the depth information can be drastically degraded with no perceptible change in image quality. The human visual system's coarse focal depth resolution dictates the minimum focal resolution that should be supported by the display.

The ability to construct a display with a different depth resolution than spatial dimensions offers us the opportunity to optimize the information conveyed by the displays to be most effective for the observer. The fixed-viewpoint volumetric display allows us to treat focal information as just another channel of information which we would like to present to the observer. This level of control is not possible in auto-stereoscopic-volumetric displays such as the Perspecta and DepthCube (Table 1.1) because they present a single projection of the image in the volume and create stereo signals through spatial modulation, which also creates the focal signals. This means that the spatial resolution must be high in three dimensions. Thus they will produce focus cues that are much more precise then they need to be at the expense of the spatial resolution, color depth and view dependent effects including occlusion. Conversely, the Samsung 2233RZ field sequential display retains excellent image quality and high resolution but presents the image at a single image plane. The fixed viewpoint volumetric display is a technological compromise that balances the need for accurate focus signals with the need for high quality images with lighting effects. I used one such display for many experiments, the 3-mirrors display (described in Section 1.2.1), and helped to develop a second display based on a the switchable lens element (described in Chapter 4). Both of these systems use a limited number of discrete image-planes to present high quality images with focus cues. In Table 1.1 I show the balance between resolution, color, frame rate and image planes used in various systems. The two autostereoscopic volumetric displays (DepthCube and Perspecta) trade off spatial resolution, refresh rate, and color depth to achieve a large number of voxels. The Samsung display omits focus signals to present a high resolution, full color image. The fixed-viewpoing volumetric displays trade off spatial resolution (3-mirrors) and refresh rate (switchable lens) to create a limited number of image planes on which distinct projections can be shown to each eye. Thus they can show fine depth intervals by modulating disparity, and present nearly correct focus information. Later in Chapter 5, I will justify why sparse image-planes may be adequate for an effective display.

Although only a limited focal resolution may be needed, achieving pixel-by-pixel control over focal distance remains challenging. At this time, these displays have been limited to research platforms employing various technologies to control each pixel's focal distance. All approaches decompose the image into depth regions and then present partial images at dif-

Display	Spatial Res.	Depth	Color	Refresh rate	bit rate
Samsung 2233RZ	$1680x1050$	2 proj.	3 x 8 bit	60 Hz/view	$5.1 * 10^9$
Perspecta	$768x768$	1 proj./ 198 slice	3 x 3 bit	24 Hz/vol.	$2.5 * 10^{10}$
DepthCube	$1024x768$	1 proj./ 20 planes	3 x 5 bit	50 Hz/ vol.	$1.2 * 10^{10}$
Switchable lens	$800x600$	2 proj./ 4 planes	3 x 8 bit	45 Hz/ vol.	$4.1 * 10^9$
3-mirrors	variable	2 proj./ 3 planes	3 x 8 bit	13 Hz	$2.9 * 10^9$

Table 1.1: Information tradeoff in stereoscopic displays. A selection of 3d displays and their specifications. Spatial resolution represents the resolution in the X-Y plane. Depth represents the number of depth planes (or volumetric slices) and the number of unique projections (proj.). Color represents the number of color primaries and the number of bits per primary. Refresh rate is the rate at which the full volume is updated. The bit rate is the bits/second (in Hz) at which information is passed to the display.

ferent focal planes. The sum of these partial images is equal to the 2D image projection with focus information. Most of the techniques use temporal multiplexing, where the full scene is constructed one image-plane at a time, but so quickly that the scene appears simultaneously. One approach uses a deformable mirror to change the focal length of the image[113]. Other labs have modulated focal distance by translating a microdisplay[117] or lens[116]. Still another promising method is by changing the power of the lens system itself either by deforming a liquid lens[82] or electronically changing the state of a liquid crystal lens[120].

These displays are difficult to implement and are limited in the frame rates that they can achieve. Possibly the most straightforward approach for creating different focal distances is to use a series of display surfaces that are at different distances. The prototypical display, developed by Kurt Akeley and colleagues[1], uses half silvered mirrors to superimpose images from multiple distances. Like the other fixed-viewpoint volumetric displays, it presents high-quality images rendered sharply and preserves specular lighting effects, occlusions and other view-dependent effects. After rendering, it distributes the 2d image across the image planes such that both accommodative and blur cues are approximately correct. I used this display extensively to investigate the perceptual importance of correct focus cues in the experiments described in Chapters 2 and 3, and therefore will give a brief overview of how this display works.

1.2.1 Three-mirrors volumetric display

The three-mirrors display is shown schematically in Figure 1.2A. In each eye's view, a mirror and two plate beam splitters create a light field that is the sum of aligned images drawn at three image planes, and a periscope assembly creates binocular overlap. This is a fixed viewpoint volumetric display because an image is rendered for each eye's position and the light originates from appropriate distances. The fixed viewing position is an advantage

of the system because it allows us to carefully calculate each eyes' view, display the correct disparities, and preserve viewpoint-specific lighting effects like occlusion, specularity, and reflection. The display presents all this informtion via two separate image paths. It presents focus cues with relatively low resolution because sensitivity to focus cues is far poorer than for spatial position [22, 107]. Figure 1.2A illustrates how we create three focal distances for each of the eyes by dividing the screen into six viewports.

Under typical viewing situations, depth of focus is ±0.25 to ±0.3D [22, 26], which corresponds to a range of 0.5–0.6D around fixation. (throughout this dissertation I will use D to represent *diopters*, which is the inverse of distance in meters) In this display, the image planes are placed slightly farther apart at 1.87 (far plane), 2.54 (mid plane), and 3.21D (near plane), separations of 0.67D (Figure 1.2B). The image-plane spacing is therefore just slightly greater than a standard observer's depth of focus, and the image planes make up a viewing frustum or workspace in which focus cues can be presented. The monitor, an IBM T221 liquid-crystal display (LCD), has a maximum resolution of 3840 x 2400. At that resolution, pixels subtend 1.38, 1.09, and 0.80arcmin at the near, mid, and far image planes. An Nvidia Quadro 4 900XGL graphics card drives the display at 12 Hz at full resolution or at 41Hz in half resolution, 1920 x 1200. The half resolution is useful in experiments where timing is critical. Because of the sample-and-hold manner in which pixels are illuminated in LCDs, the images do not flicker.

Alignment

Because the light rays approaching the eyes come from different distances, alignment of the eyes with the viewports is critical. We achieve this by first using a sighting device[58] to adjust the bite bar (and therefore the observer's eyes) relative to the apparatus so that we can place the rotational center of the eyes in the appropriate position. Once the observer is in position, the separation between the viewing apertures is set equal to the observer's inter-ocular distance; and we also set software parameters to the inter-ocular distance. We then fine-tune the alignment in software using a between-image-plane vernier alignment technique[1]. The alignment is accurate to within seconds of arc and ensures that light from the appropriate pixels on different image planes sums along lines of sight. (For more information on the alignment protocols see Appendix A.3)

No need to track accommodation

With this display, we can simulate the effects of differential focus without tracking the observer's accommodation. Consider a simple situation: we want to portray a small object at the distance of the far plane (53.6cm) and another small object at the distance of the near plane (31.1cm). For real objects positioned at those distances, accommodation to 53.6cm would make the far object in focus and the near one out of focus. With accommodation to the near object, the reverse occurs. Exactly the same applies for our display because the

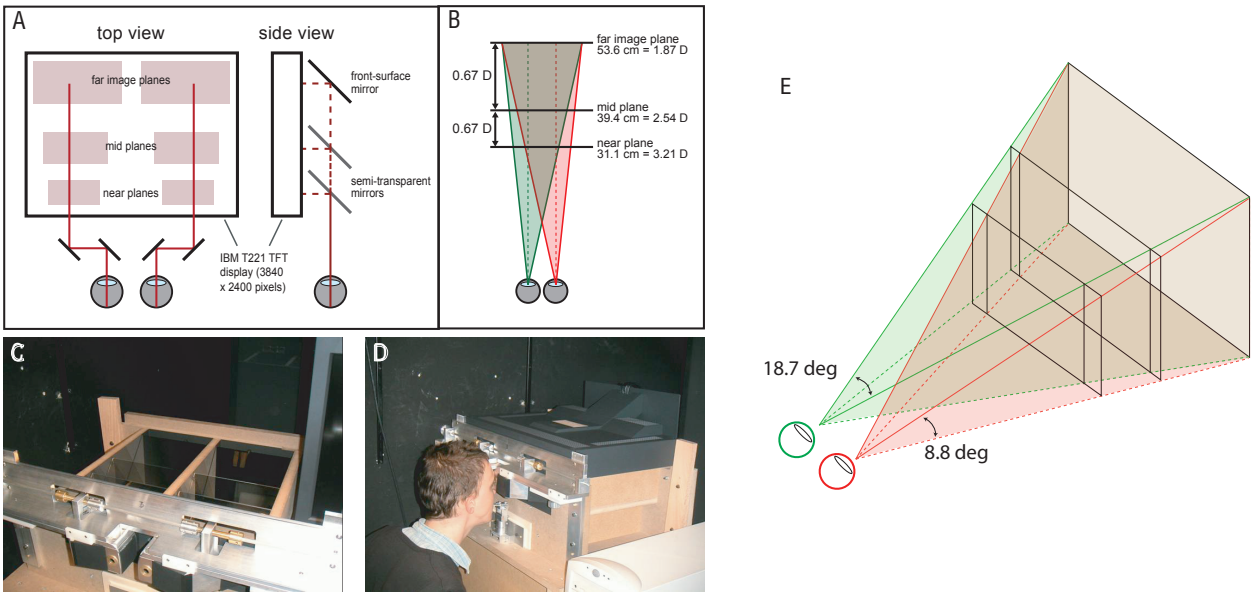


Figure 1.2: Fixed-viewpoint, volumetric display. The IBM T221 LCD panel is viewed through semi-transparent and front-surface mirrors such that each eye sees three superimposed images. A) Schematic. The left side is a top view. The T221 display is placed face down on top of the apparatus. Shaded rectangles represent the portions of the display surface that are the far, mid, and near image planes. Red lines represent left- and right-eyes' lines of sight if the eyes are in parallel gaze (vergence = 0). Periscope optics delivers the images to the eyes. The right side of Panel A is a side view. It shows the arrangement of the mirrors for creating each eye's view. The images from the three planes travel different distances before arriving at the eyes. The far image plane is reflected off the front-surface mirror and transmitted through two semi-transparent mirrors. The near image plane is reflected off one semi-transparent mirror, and the mid image plane is reflected by the mid semi-transparent mirror and transmitted through the front mirror. B) The focal distances of the image planes and the viewing frusta for the two eyes. The viewing frusta were skewed so that the binocular overlap was maximized at the far image plane. C) The display is removed from the top to expose the mirrors. The apertures for viewing the stimuli are in the lower left. A worm gear allows the aperture separation to be adjusted to the observer's interocular distance. D) An observer viewing stimuli with the display in place. E) The viewing frustum is shown in perspective. [1]

light comes from the appropriate distances. The situation is more complicated for simulated object positions in between the planes. We discuss this in the following section.

To create the retinal images that best approximate images formed in real-world viewing, we render objects unblurred and present them on the image planes. The observer's optics create the appropriate defocus blur in the retinal images, and thus we do not need to make assumptions about where the observer is looking or accommodating.

Depth-weighted blending

For all but the very unlikely case that the depth of a point in the scene coincides exactly with the depth of one of the image planes, a rule is required to assign image intensities to image planes. The simplest rule is the box filter (Figure 1.3, left): each point in the scene is drawn at the image plane to which it is closest. However, this approach produces blur discontinuities. For example, consider the situation in the left of Figure 1.3 in which a line extends from a near image plane to a farther one. For a given accommodative state, the retinal-image blur of the line will have one value for parts that are drawn on one plane and another value for parts that are drawn on the second plane. This produces a visible discontinuity in the retinal image of the line. To minimize this problem, we use a tent filter (Figure 1.3, right). With this rule, the image intensity at each image plane is weighted according to the dioptric distance of the point from that plane, determined along a line of sight. This approach, known as *depth-filtering* (or synonymously as *depth-weighted blending*), eliminates the discontinuity (by a mechanism similar to how anti-aliasing eliminates edge artifacts in regular computer graphics). This technique is a significant technical development that can be applied to all multi-plane displays.

Using the tent filter to simulated distance D_s (expressed in diopters) and a pixel intensity of I_s , we distribute the light to the nearer and farther planes respectively as I_n and I_f according to the formula:

$$I_n = (1 - \frac{D_n - D_s}{D_n - D_f})I_s;$$
$$I_f = (\frac{D_n - D_s}{D_n - D_f})I_s \tag{1.1}$$

where D_n and D_f are respectively the dioptric distances of the nearer and farther of the two bracketing focal planes. Thus, pixels representing an object at the dioptric midpoint between two focal planes are illuminated with half intensity on the two planes. The pixels for the two planes are aligned at the eye, so they sum in the retinal image to form an approximation of the image that would occur when viewing a real object at that distance. For an object closer to one of the display planes, intensities are suitably modulated to give more weight to the plane closest to the object. As Equations 1.1 show, the sum of the intensities remains constant at all simulated distances; i.e. $I_n + I_f = I_s$ for all D_s. Depth-weighted blending is crucial for

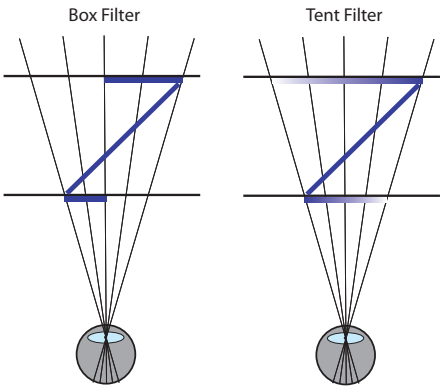


Figure 1.3: Pixel lighting with box and tent depth filters. The horizontal lines represent two image planes. The thick diagonal line is the surface we wish to draw, and the thick blue horizontal lines represent the pixel intensities on the image planes; dark blue represents high intensity and pale blue represents low intensity. The box filter illuminates the far image plane at high intensity for all regions in which the simulated surface is closer to the far than the near plane. It illuminates the near plane at high intensity for all regions in which the simulated surface is closer to that plane. The tent filter weights intensities according to the proportion of the dioptric distance between the two image planes. This occurs along each line of sight. The sum of the two intensities is the same as the intensity on one plane in a conventional display. [1].

avoiding discontinuous changes in the retinal image as simulated distance changes[1]. I will discuss the efficacy of this depth-filtering approach in Chapter 5.

1.3 Significance

The need for effective stereoscopic 3d visualization is growing. Consumers are interested in stereoscopic cinema[71] and display manufacturers are racing to bring stereo technology to the home for both movies and games[63]. There is also a growing interest in head mountable displays[21]. Of greatest importance, medical diagnosticians are finding value in stereoscopic displays, as their instruments increasingly improve at collecting 3d data; many new diagnostic machines can sample a volume of tissue and effective visualization of the data is gaining importance. We would like to display the data in a fashion similar to what a surgeon might experience during a surgery. This type of interaction is appealing only if the stereoscopic images produce a faithful percept, without undo effort. It is therefore important that we consider the problems that occur from viewing stereoscopic data with incorrect focus information.

There are a number of specific problems that could arise from incorrect focus signals in stereoscopic imagery. Some of the most serious problems are failure to fuse the stereo

images, seeing distorted depth relationships, and visual fatigue. These are problems that are frequently experienced by those attending stereo cinema and by 3d image analysts but are poorly understood. For example, film directors rely on rules of thumb and heuristics to produce stereoscopic cinema that will keep people from leaving the theaters with headaches and nausea.

Understanding these problems is challenging because conventional stereoscopic displays allow us to produce and observe these symptoms, but do not provide us with the control to identify causes. To test if the problems of stereoscopic displays can be attributed to incorrect focus cues, an experiment should compare identical viewing situations except for the appropriateness of the focus cues. A fixed viewpoint volumetric display offers this level of control, and is thus well suited for these experiments. It allows investigators to de-correlate focus information from other depth cues. Also, volumetric displays may be valuable outside of vision research. Not only may volumetric displays help us to understand some of the problems with traditional stereoscopic imagery, they may also offer us a solution.

The research described in this thesis is intended to answer three questions. 1) What is the role of focus information in stereoscopic vision, and how would its omission change the stereoscopic percepts? 2) Is it possible for us to construct a usable volumetric display that can present this information correctly? 3) Given what we know about human optics, how well can a low depth-resolution display present correct focus signals.

Chapter 2

Consequences of vergence-accommodation conflicts

This chapter describes work on the influence of incorrect accommodation stimuli in stereoscopic displays for binocular fusion, space perception, and visual fatigue. These experiments were originally published in the Journal of Vision in 2008. My coauthors were Ahna Girshick, Kurt Akeley, and Martin Banks.[59] Ahna, Kurt and Martin were predominately responsible for planning and carrying out experiments 1 and 2 of the four described in this chapter. I became involved afterwards in the analysis of this data, and planned and carried out experiments 3 and 4 from their onset.

2.1 Introduction

2.1.1 Focus cues and perceptual distortions

One concern of presenting incorrect focus cues in a 3d display is that accommodation and blur will conflict with the stereo specified depth, and could cause 3d shape misperceptions. To understand 3d percepts, we should consider the array of cues the visual system uses to estimate 3d layout. There is considerable evidence that depth cues are combined in optimal or nearly optimal fashion to produce minimum-variance depth estimates [57, 67, 75]. Even supposedly ordinal depth cues are combined this way[20]. Assuming that the noises associated with cue measurement are independent and Gaussian distributed and all depths are equally likely, one can derive from Bayes' Law a simple rule for producing the minimum-variance estimate [29, 49]:

$$\hat{D} = \sum_i w_i D_i; \quad w_i = \sigma_i^{-2} / \sum_j \sigma_j^{-2} \tag{2.1}$$

where D_i is the relative depth (difference in distances between positions in the scene) specified by cue i, and σ_i is the standard deviation of that cue's estimate. Because the weights are

proportional to the normalized inverse variances, more weight is assigned to less variable (more reliable) cues.

Stimuli in the real world contain numerous depth cues all specifying the same 3d layout. Computer displays present images on one surface. And thus some of the depth signals will be in conflict. In this analysis we consider conflicting focus cues because they have a strong potential to cause stereoscopic shaper misperception [19, 44, 137] and viewer fatigue [129, 134].

If the estimates from individual cues are unbiased from the true values, the minimum-variance estimate from Equation 2.1 is:

$$\hat{D} = w_{sim} D_{sim} + w_{foc} D_{foc} \tag{2.2}$$

where D refers to relative depth (the distance between one point and another), *foc* and *sim* to the relative depth specified by focus cues and by information other than focus cues (e.g., disparity, shading, perspective), and w_i are given by Equation 2.1. In the real world, relative depth specified by the two classes of cues would generally be the same, so the equation reduces to:

$$\hat{D} = D_{sim} \tag{2.3}$$

A plot of perceived relative depth as a function of the depth specified by the available cues would yield a line of slope one. As a consequence, we generally see the depth in real scenes correctly, at least at close range (Mon-Williams et. al.[94]; for counter-examples at long range, see Loomis, Da Silva, Fujita, & Fukusima,[84]). With a computer-displayed stimulus, simulated cues indicate depth variation while focus cues indicate a relative depth of zero:

$$\hat{D} = W_{sim} D_{sim} \tag{2.4}$$

Thus, a plot of perceived relative depth as a function of the relative depth specified by cues other than focus cues yields a line of slope w_{sim}, which is less than 1 for $w_{foc} > 0$. Consequently, we usually experience depth compression in 3d displays [19, 137, 142].

2.1.2 Focus cues, fusion, and stereopsis

For a stimulus to be sharply focused on the retina, the eye must be accommodated to a distance close to the focal distance of the object. The acceptable range is the depth of focus, which is roughly ± 0.3 diopters (D) under normal circumstances [22, 26]. Understandably, the blur caused by an accommodative error reduces the precision of stereopsis: With natural pupils, errors of 1 and 2 D cause nearly two- and ten-fold reductions in stereoacuity, respectively[97, 138, 143]. For a stimulus to be seen as single (i.e., fused) rather than double, the eyes must be converged to a distance close to the object distance. The tolerance range is Panum's fusion area, which is 15-30 arcmin[98, 111]. Thus, vergence errors larger than 15-30 arcmin cause a breakdown in binocular fusion and stereopsis is thereby disrupted [72].

Smaller vergence errors do not cause fusion to break down, but yield measurable reductions in stereoacuity [15]. Therefore, fine stereopsis requires reasonably accurate accommodation and vergence. Figure 2.1A shows the range of acceptable focal distances (the depth of focus) and the range of acceptable fusible disparities (Panum's area) when the viewer accommodates and converges to the same distance. Panel B illustrates the range of accommodative and vergence stimuli that a typical observer can view without excessive vergence or accommodative error which is known as the *zone of clear single binocular vision* ([45, 64, 95]; green region in Figure 2.1B).

Accommodation and vergence responses are normally coupled. Specifically, accommodative changes evoke vergence changes (accommodative vergence), and vergence changes evoke accommodative changes (vergence accommodation) [42, 87]. In the real world, accommodation-vergence coupling is helpful because focal and vergence distances are almost always the same no matter where the viewer looks (Figure 1.1, left). One benefit of the coupling is increased speed of accommodation and vergence. Accommodation is faster with binocular viewing—where blur and disparity signals specify the same change in distance—than with monocular viewing where only blur provides a useful signal [32, 77]. Similarly, vergence is faster when disparity and blur signals specify the same change in distance than when only disparity specifies a change [32, 114]. For these reasons, one expects that demanding stereoscopic tasks will require less time when the stimuli to accommodation and vergence are consistent with one another than when they are not.

In 3d displays, the normal correlation between focal and vergence distance is disrupted (Figure 1.1, right): focal distance is now fixed at the display while vergence distance varies depending on the part of the simulated scene the viewer fixates. For reasons stated above, we expect that stereoacuity will be reduced in 3d displays compared with viewing situations that maintain the normal correlation between focal distance and vergence distance. We also expect that the time required to fuse a stimulus binocularly will be increased in conventional 3d displays.

2.1.3 Focus cues and visual fatigue

Prolonged use of conventional 3d displays produces viewer fatigue and discomfort[37, 121, 122, 135, 146]. It has often been claimed that the symptoms are caused by the dissociation between vergence and accommodation that is required in such displays[37, 129, 134, 146]. Emoto and colleagues[37] observed symptoms of fatigue when observers viewed stimuli with a larger conflict between the vergence and focal distances. Wann and Mon-Williams[135] reported a pre-test vs post-test change in the cross-link functions (AC/A and CA/C ratios) after prolonged viewing of a 3d display; they also observed an increase in fatigue. As we will make clear in the Discussion (Section 2.6.3, and foreshadowed in the introduction to Experiment 4), factors other than the conflict in the vergence and accommodative stimuli could have been responsible for the reported fatigue in all of the above-mentioned references.

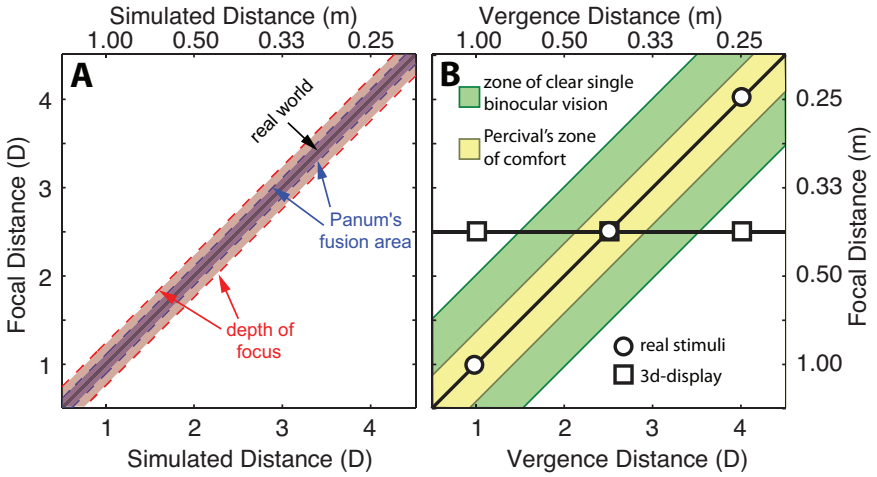


Figure 2.1: Consequences of vergence-accommodation coupling. The panels plot focal distance as a function of simulated (or vergence) distance. The bottom abscissa and left ordinate have units of diopters. The top abscissa and right ordinate have the corresponding values in meters. A) Depth of focus and Panum's fusion area in diopters. Here we simulate real objects, so the vergence and focal distances specified by the object are equal to one another. The diagonal line represents the viewer's vergence and accommodation assuming accurate responses to the object. Vertical cross sections of the red region represent the eye's depth of focus. Horizontal cross sections of the blue region represent Panum's fusion area (15 arcmin). Note that the sizes of the zones of clear vision and of single vision remain constant in diopters. Because of this, we will express simulated and vergence distances in diopters instead of in meter-angles, which are the conventional units in optometry and ophthalmology. B) Zones of clear single vision and of comfort. The green area represents the zone of clear single binocular vision: the range of vergence and accommodative responses that young adults can achieve without excessive effort and without exceeding depth of focus or Panum's area. The yellow area represents Percival's zone of comfort: the range of responses viewers can achieve without discomfort. Circles represent three real-world stimuli; squares represent the corresponding stimuli on a conventional 3d display.

Therefore, to our knowledge the link between the stimulus conflict and fatigue has never been directly tested.

The set of vergence and accommodative responses that can be achieved without discomfort is *Percival's zone of comfort*, which is about one-third the width of the zone of clear single binocular vision ([64, 95, 103]; yellow region in Figure 2B). Stimuli in the real world (the circles in Figure 2.1B) fall within the comfort zone while many stimuli on 3d displays (the squares) do not. To fuse and focus the latter stimuli, the viewer must counteract the normal accommodation-vergence coupling, and the effort involved is believed to cause viewer fatigue and discomfort[129, 134]. Accommodation and vergence cannot be set independently to arbitrary values because a change in vergence innervation affects accommodation and vice versa.

3d displays are used increasingly in several professions, particularly medicine, so it is important to determine if accommodative-vergence mismatches in fact cause the fatigue and discomfort. Determining whether or not such mismatches are the cause is a crucial first step toward figuring out how to minimize the problem.

2.1.4 General experimental methods

Many of our experimental questions concern whether incorrect focus cues cause observable problems with visual perception. To properly isolate focus cue conflicts, we need to directly compare conditions in which focus cues are correct, and conditions where focus cues are incorrect. We used the display described in Section 1.2.1 extensively in the following experiments.

2.1.5 Experiment preview

In the experiments reported here we used the unique properties of our display to investigate how the relationship between vergence and accommodation-specified distances affects visual performance and fatigue. Experiment 1 examined whether viewers can fuse a stereoscopic image more quickly when the conflict between simulated cues (in this case, disparity) and focus cues is minimized. Experiment 2 asked whether viewers achieve greater stereoacuity with brief stimulus presentations when the conflict is minimized. Experiment 3 examined distortions in perceived depth and how they are influenced by focus cues. Finally, Experiment 4 investigated the role of mismatches between the stimulus to vergence and the stimulus to accommodation in viewer fatigue and discomfort. We found in each case that minimizing the conflict between focus cues and disparity cues yielded a clear benefit.

2.2 Experiment 1: Time to fusion

When vergence and accommodative distances differ substantially, many viewers find it difficult to fuse a binocular stimulus. We conducted an experiment to determine how serious an effect this is for the vergence-accommodation conflicts that occur with conventional 3d displays. Specifically, we measured the effect of focus cues on the time needed to discriminate a cyclopean stimulus in a random-dot stereogram. We varied the conflict between vergence distance and focal distance, and for each combination of those distances, we found the shortest stimulus duration at which the observer could reliably determine the orientation of the cyclopean stimulus. Because of the coupling between accommodation and vergence, we expected that observers would perform better when vergence and focal distances were equal to one another.

We previously conducted an experiment similar to this[1], but that experiment was limited in its ability to reveal the consequences of vergence-focal conflicts on binocular fusion. First, the test stimulus always required that the observer diverge the eyes, and there are two problems with that: 1) the observer could anticipate that the test stimulus would jump to a greater distance and start making the divergence movement before the stimulus actually appeared; 2) we could not determine the consequences of vergence-focal conflicts when convergence was required instead of divergence. Second, the test stimulus consisted of two stereoscopically defined planes, a red one and a green one. The task was to indicate whether the red or green plane was nearer. Such a task does not require accurate accommodation so it is not well designed to reveal the consequences of vergence-focal conflicts. Experiment 1 was designed to overcome both of the problems with the experiment reported by Akeley et al.[1].

2.2.1 Methods

Observers

Three observers participated: ARG (29 years old), DS (23), and BGS (19). ARG was an author; DS and BGS were unaware of the experimental hypotheses. All had normal stereopsis as assessed by the Titmus Stereo Test. They wore their usual optical corrections. We avoided older observers because they are much more likely to have reduced accommodative range due to presbyopia.

Stimuli

The stimuli were random-dot stereograms depicting sinusoidal corrugations in depth. We wanted to make sure that vergence and accommodation would have to be reasonably accurate to perform the task, so the stereograms had a high dot density (45 dots/deg^2) and high corrugation frequency (1.35 cpd). With these values, relatively small errors in

accommodation or vergence affect performance[10, 97]. The corrugations were oriented ±15° from horizontal, and observers indicated which of the two orientations was presented on each trial. The peak-to-trough disparity was 4.9 arcmin at all viewing distances. The stimulus was presented in a virtual circular aperture with a diameter of 4.2°; the surround was black. We verified that the orientation-discrimination task could not be performed monocularly. We operated the display at half resolution, (1920x1200 pixels) to achieve a refresh rate of 41 Hz thereby allowing fine adjustment of the display time.

At the beginning of each trial, a fixation target was presented on the mid image plane (Figure 2.2A). The test stimulus was then presented at various vergence and focal distances. We presented two types of stimuli: Those in which the focal distances corresponded to image planes (non-filtered stimuli) and those in which the focal distances were between image planes (filtered stimuli). Testing with filtered stimuli provides an important check that the use of depth-filtering does not disrupt the fusion and interpretation of stereo stimuli. Specifically, we wanted to determine whether such filtering causes an unforeseen disruption of stereovision.

Non-filtered test stimuli were presented with the vergence distance at the near, mid, or far plane, and the focal distance at the near, mid, or far plane. Several combinations of vergence and focal distance were presented as shown in Figure 2.2B,C (conditions 1, 3, 4, 5, 6, 7, 9, A, C, D, and F). The conflicts in the stimuli to vergence and accommodation were 0, ±0.33, ±0.67, or ±1.33D.

Filtered test stimuli were presented with the vergence distance and focal distance at a position between the mid and near planes (near-mid distance; 34.8 cm) or between the mid and far planes (mid-far distance; 45.4 cm). We chose the dioptric midpoints between the image planes for the test of depth-filtereing because the focal approximations are poorest at those distances. Those stimuli are depicted in Figure 2.2B (rightmost column) and Figure 2.2C (conditions B and E).

Procedure

The observer initiated a trial with a button press upon which the fixation stimulus (Figure 2.2A) appeared for 634-1415 ms, the duration being drawn from a random uniform distribution; the random interval discouraged the observer from making anticipatory eye movements. The fixation stimulus was a cues-consistent cross that helped the observer converge and focus at the mid image plane. The fixation stimulus was then extinguished and the test stimulus appeared immediately at one of the various combinations of vergence and focal distance (1, 3, 4, 5, 6, 7, and 9 and A–F in Figure 2.2). At the end of the test stimulus presentation, a masking stimulus appeared to ensure that observers could not make judgments based on an after-image of the test stimulus. The mask was composed of dots randomly positioned in depth, and was presented at the same vergence distance and focal distance as the test stimulus. Observers indicated the orientation of the corrugation with a key press. They were not provided feedback about the correctness of their response. The

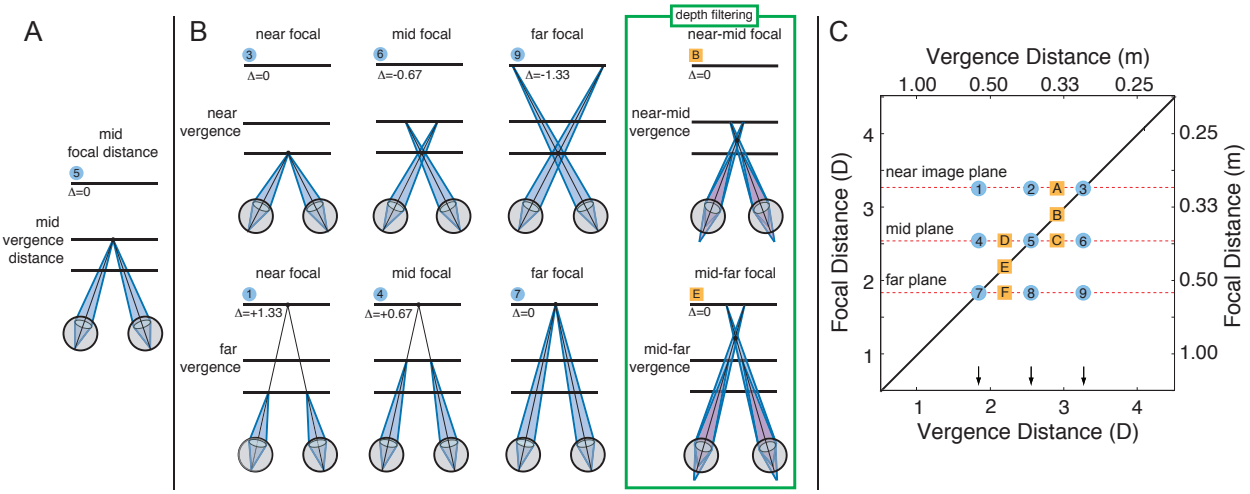


Figure 2.2: Stimuli locations in vergence-focal space. The horizontal lines represent the image planes in the apparatus. The thin black lines represent the visual axes of the eyes; those axes intersect at the vergence distance. The light blue regions represent the focal distance. Panel A represents the fixation stimulus before the presentation of the test stimulus. Δ is the difference in diopters between the vergence and focal distances. Panel B depicts eight of the test stimuli in the experiments. For the stimuli marked 3, 7, B and E, the focal and vergence distances were equal to one another (cues-consistent). The rightmost column, outlined in green, depicts the two stimuli with depth-filtering (see General Methods). In these stimuli, the vergence distance was the dioptric midpoint between two image planes; 50% of the light came from each of those planes. Panel C represents all the stimuli in the experiments. The horizontal dashed lines represent the three image planes. The vertical arrows represent the vergence distances. Stimuli at positions 1, 3, 4, 5, 6, 7, 9, A, B, C, D, E, and F were presented in Experiments 1 and 2; in B and E the focal stimuli are depth-filtered approximation. Experiment 3 used stimuli at 1-9. Experiment 4 used stimuli at 4, 5, and 6 for the cues-inconsistent session, and stimuli at 3, 5 and 7 in the cues-consistent session.

duration of the test stimulus was varied according to a 2-down/1-up adaptive staircase. The staircase was run until 12 reversals had occurred. To minimize fatigue, observers were prompted to take a short break after every eight trials. Each experimental session contained all the vergence-focal distance combinations presented in a random fashion so that the observer could not anticipate which condition would be presented at a given time. A session contained approximately 300 trials. Observers completed two sessions, so they were tested twice on each vergence-focal distance combination.

Cumulative Gaussians were fit to the combined psychometric data for each vergence-focal combination using a maximum-likelihood criterion[139, 140]. We found the 75% correct point on each function and defined that as the time required to fuse the stimulus in each condition.

2.2.2 Results

The results for the non-filtered stimuli are shown in Figure 2.3A, which plots the time the observers required to correctly identify stimulus orientation as a function of the difference between the vergence and focal distances. Different observers exhibited different effect sizes, but in every case the required time decreased monotonically with decreases in the vergence-focal conflict. The minimum time of 400-500 ms occurred when the conflict was zero. Observers were somewhat faster at fusing the stimulus when the vergence jumped from the mid to the near plane than when it jumped from the mid to far plane.

These data show that differences in the vergence and focal distances affect the ability to fuse stereoscopic stimuli. When vergence and focal distances change together, the stimulus can be fused and interpreted quickly. When vergence and focal distance do not change together, more time is required. These effects are surely caused by the cross-coupling between vergence and accommodation.

Figure 2.3B compares performance with the filtered stimuli and the cues-consistent subset of the non-filtered stimuli. It plots the duration required to correctly identify stimulus orientation as a function of the change in vergence and focal distance. The required time increased as the change in vergence and focal distance increased. We also measured the time to fusion for stimuli at locations A, C, D and F in Figure 2.3. The fusion times for the 0.33D conflicts did not differ from the cues-consistent blends; we omitted these data points for clarity. The results show that the depth-filtering technique used to simulate a cues-consistent stimulus does not hinder stereovision, at least in the difficult task of Experiment 1.

2.3 Experiment 2: Stereoacuity

We examined the consequences of inconsistency between focal and vergence stimuli using a second important criterion: the acuity of stereopsis. Specifically, we measured stereoacuity

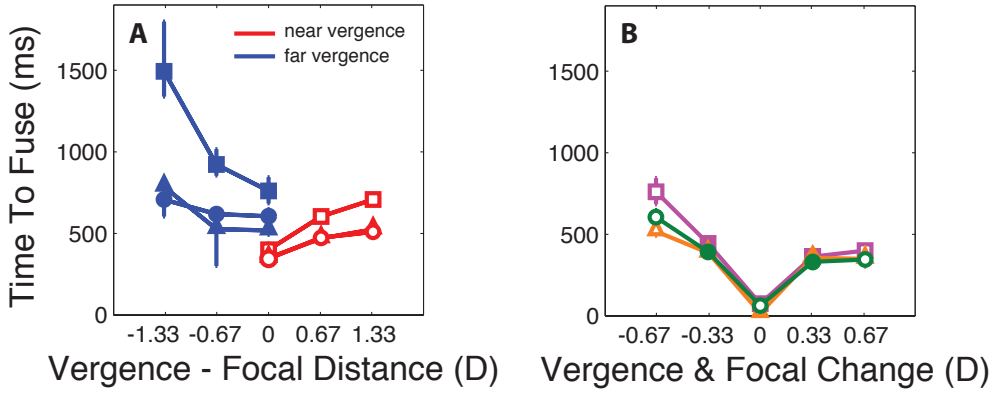


Figure 2.3: Results of Experiment 1. A) Results for the non-filtered stimuli. The abscissa represents the difference between the vergence and focal distances (vergence minus focal) in diopters. The ordinate represents the stimulus duration required for the observer to correctly identify the orientation of the cyclopean stimulus in 75% of the trials. The data for the ±0.33D conflicts are not shown. Different symbols represent the data from different observers: squares for ARG, triangles for DS, and circles for BGS. The blue filled symbols represent data when the vergence distance jumped from the mid to the far image plane; red unfilled symbols represent data when the vergence distance jumped from the mid to the near plane. Error bars are 95% confidence intervals. B) Results for the cues-consistent stimuli. The abscissa represents the change in the vergence and focal distance from the fixation to the test stimulus; positive values are changes in which the distance decreased and negative values are those in which the distance increased. The ordinate represents the stimulus duration required to correctly identify stimulus orientation in 75% of the trials. Different symbols represent data from different observers: purple squares for ARG, orange triangles for DS, and green circles for BGS. The unfilled symbols represent data with non-filtered stimuli in which the focal and vergence stimuli were at one of the image planes. The filled symbols are data with filtered stimuli in which focal and vergence distance were between image planes. Error bars are 95% confidence intervals.

thresholds for briefly presented stimuli for different vergence-focal conflicts.

2.3.1 Methods

The observers were the same as in Experiment 1. The stimuli, procedure, and task were also the same with two exceptions. First, stimuli were always presented for 1 sec rather than a variable duration. Second, the spatial frequency of the corrugations was varied in order to find the highest discriminable frequency. To enable the presentation of high corrugation frequencies, we set the display to its highest resolution (3840 x 2400 pixels), which required a reduction in refresh rate to 12Hz.

As in Experiment 1, the experiment had two kinds of stimuli. With the non-blended stimuli, focal distances corresponded with image planes; the combinations of vergence and focal distances were the same as with the non-filtered stimuli in Experiment 1 (1, 3, 4, 5, 6, 7, 9, A, C, D, and F in Figure 2.2C). We presented these stimuli to determine if vergence-focal conflicts adversely affect stereo acuity. With the blended stimuli, the focal stimuli were depth-weighted blends between image planes; the combinations of vergence and focal distances were the same as with the blended stimuli of Experiment 1 (B and E in Figure 2.2C). We presented these stimuli to determine if depth-weighted blending adversely affects stereopsis.

2.3.2 Results

Figure 2.4A shows the results with the non-filtered stimuli with near and far vergence distances. The highest corrugation frequency at which observers could reliably discriminate stimulus orientation is plotted as a function of the difference between the vergence and focal distances. The highest stereoacuity was obtained when the conflict was smallest. Thus, minimizing the vergence-focal conflict enables more precise stereopsis with relatively brief stimulus presentations. Stereoacuity was slightly higher for the near-vergence stimuli than for the far-vergence stimuli.

The differences in performance across the three observers are interesting. The effect of vergence-focal conflict was much larger in ARG than in DS and BGS. In Experiment 1, DS and BGS could fuse and interpret a 1.35 cpd corrugation with durations less than one second at all vergence-focal conflicts. ARG could not interpret the far-vergence stimuli unless the conflict was zero; she also exhibited a larger effect of conflict with the near-vergence stimuli. The stimulus duration in Experiment 2 was thus longer than DS and BRS required at 1.35cpd, but not necessarily longer than ARG required. For these reasons, it is not surprising that ARG exhibited the largest effect of varying the vergence-focal conflict in Experiment 2. Presumably, we would have observed larger effects of vergence-focal conflict in DS and BGS if we had used shorter durations.

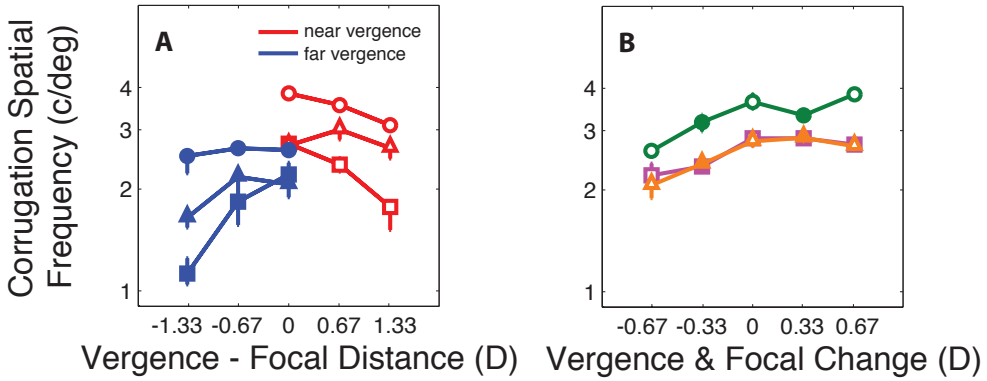


Figure 2.4: Results from Experiment 2. The format is the same as Figure 2.3 except the ordinate represents the highest spatial frequency at which observers could identify the corrugation orientation correctly 75% of the time. A) Results for non-blended stimuli with near or far vergence distances. The abscissa represents the conflict between the vergence and focal distances in diopters. The data for the ±0.33D conflicts are not shown. Different symbol types represent the data from different observers: squares for ARG, triangles for DS, and circles for BGS. The blue filled symbols represent the data for far vergence distances and the red unfilled symbols the data for near vergence distances. Error bars represent 95% confidence intervals. B) Results for cues-consistent stimuli. The abscissa represents the change in the vergence and focal distance from the fixation to the test stimulus; positive values are changes in which the distance decreased and negative values are those in which the distance increased. Purple squares are data for ARG, orange triangles for DS, and green circles for BGS. The unfilled symbols are data with non-blended stimuli presented at an image plane and the filled symbols are data with blended stimuli presented between image planes. Error bars are 95% confidence intervals.

The results for the cues-consistent non-filtered and filtered stimuli are shown in Figure 2.4B. Stereoacuity is plotted as a function of the change in the vergence and focal distances. The data at ±0.33D are from the filtered stimuli presented between image planes and the other data are from non-filtered stimuli presented at planes. The figure shows only results with cues-consistent stimuli (vergence distance equals focal distance). We also measured stereoacuity for stimuli at locations A, C, D and F in Figure 2.2 and found that performance with the 0.33D conflicts did not differ from the cues-consistent filtered images; we omitted those data points for clarity. Stereoacuity worsened as the magnitude of the change in vergence and focal distance increased. Depth filtering did not affect performance, suggesting that it does not adversely affect fine stereopsis.

2.4 Experiment 3: Distance estimation and distortions

We next examined how conflicts between the stimuli to vergence and accommodation affect depth perception: in particular, how such conflicts affect the perception of 3d shape. Horizontal disparity is an ambiguous determinant of 3d shape because a given pattern of horizontal disparity is consistent with an infinite set of shapes[6, 48, 106]. To uniquely determine shape, a distance estimate is also required. The process of using a distance estimate in the interpretation of the pattern of horizontal disparity is disparity scaling. The visual system uses both vertical-disparity and extra-retinal signals in disparity scaling; the weights assigned to those signals vary with viewing condition[6, 5, 106]. The extra-retinal signals are provided by the sensed horizontal vergence of the eyes and also possibly by the sensed accommodative state. Conventional 3d displays present appropriate vergence signals, but inappropriate accommodative signals because the focal distance is fixed at the distance to the display. Consequently, the distance estimate used in disparity scaling might be erroneous when conventional displays are used.

Watt and colleagues[137] found clear evidence for an influence of focal distance in disparity scaling. In their experiment, they varied focal distance by changing the physical distance between the observer and the display; they did so by moving the observer, so the observer could have been aware of the change in viewing distance by means other than blur and accommodative signals. In Experiment 3, we took advantage of the unique properties of our volumetric display to isolate accommodation and blur, and then reexamined how focus cues per se affect disparity scaling and thereby affect depth perception. Thus, this experiment was designed to determine how much of the effects observed by Watt and colleagues were due to vergence-focal conflicts as opposed to other distance cues.

2.4.1 Methods

Apparatus

The volumetric display was the one described in the Section 1.2.1. We will refer to the trials conducted with this display as the volumetric condition. We also used a conventional 3d display as in Watt et al.[137]. This was a CRT mounted on a translating platform that allowed us to vary viewing distance precisely. We will refer to trials conducted with this display as the conventional display condition. The viewing distances with the conventional display were very similar to the focal distances with the volumetric display: 30 (near; 3.33D), 40 (mid; 2.50D), and 55cm (far; 1.82D). In this condition, the room was dimly lit so that the frame of the CRT was visible, and observers were well aware of changes in viewing distance. Stereoscopic images were presented with liquid-crystal shutter glasses (Crystal Eyes, Stereographics, Inc.) synchronized to the CRT. We positioned observers with custom bite bars such that the midpoint of their interocular axis was on the surface normal from the middle of the CRT. The display was spatially calibrated[6, 137] so we could present disparities with an accuracy of better than 1/2 arcmin.

Observers

Eleven observers participated in preliminary testing. Three were not able to fuse the stimulus when the vergence-focal conflict was large, so we did not test them further. Two more could fuse all the stimuli, but experienced significant eyestrain and withdrew from the study. The six remaining observers were DEJ (20 years old), BGS (22), DMH (23), AAA (23), JYL (23), and BV (31). Five were emmetropes. JYL was an uncorrected anisometrope with 0.25D of myopia in the right eye and 1.5D in the left eye. None wore optical corrections during the experiment. All had normal stereoacuity as assessed by the Titmus Stereo Test. One observer (DMH) was an author; the others were unaware of the experimental hypotheses. We did not test older observers because they are much more likely to have reduced accommodative range due to presbyopia.

Stimuli

As in Watt et al.[137] , the test stimulus was a random-dot stereogram of a vertical hinge in an open-book configuration (Figure 2.5). Observers indicated on each stimulus presentation whether the perceived hinge angle was more or less than 90°. Dot density was 3.3dots/deg^2. The horizontal and vertical extents of the stimulus were determined by an elliptical clipping window with a height of 5° and an average width of 6°; a random component was added to the width to disrupt a correlation of hinge angle and stereogram width. The texture gradients of the sides of the hinge were consistent with the disparity gradients, but the texture was not a reliable slant cue because of its low density. We verified this by

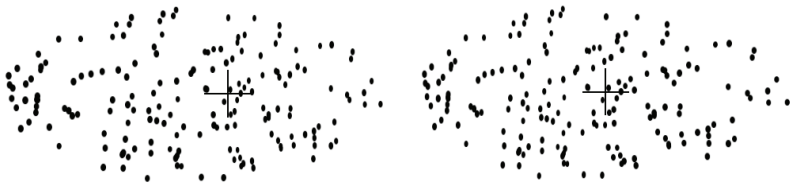


Figure 2.5: The stimulus in Experiment 3. A random-dot stereogram depicting a vertical hinge in an open-book configuration. Observers indicated whether the perceived hinge angle was greater or less than 90°. The task could not be performed monocularly. The reader can see the 3d hinge by cross-fusing the stimulus. The experiment used a red stimulus on a black background, but was reproduced here as black on white for clarity. The hinge is shown in plan view in Figure 1.1A.

presenting the stimuli monocularly to one observer and finding that he could not reliably perform the task. The background was black. The dots were produced by illuminating the red phosphor, which minimizes cross-talk in CRTs caused by phosphor decay. This artifact causes each eye to be able to see a dim ghost of the other eye's image.

We reduced the effectiveness of vertical-disparity signals so that observers would be sure to use extra-retinal signals for disparity scaling. We accomplished this by reducing the height of the stimulus, which decreases the weight given to vertical disparity[6, 106]. We randomly varied the slant of the angle bisector of the hinge by ±10° with a mean of 0°. Randomizing the base slant forced observers to base their judgments on the hinge angle rather than on the slant of one side of the hinge.

All of the stimuli in this experiment were presented on one image plane: i.e., depth-weighted blending was not used on the volumetric display. We did this to make the stimuli on the conventional display and volumetric display as similar as possible.

Procedure

Before every experimental session, observers completed a training session that was used to minimize bias in perceiving a right-angle hinge. During that session, the fixation cross and test stimulus were presented at the mid vergence and focal distance: 40cm on the CRT and 39.4cm on the volumetric display. On each training trial, the test stimulus was presented and observers indicated whether the hinge angle appeared more acute or obtuse than a right angle, which was followed by auditory feedback. An adaptive, 1-down/1-up staircase procedure varied the hinge angle from trial to trial. Training sessions lasted about 3 min.

An experimental session began after a short break. Trials in those sessions started with the 1500ms presentation of the fixation cross at the vergence and focal distance at which the hinge stimulus would appear. The hinge then appeared for 1500ms (the fixation cross remaining visible). Then the cross and hinge were extinguished, and the observer indicated with a key press whether the hinge angle appeared more or less than 90°. A 1-down/1-up

staircase adjusted the hinge angle over trials, terminating after 12 reversals. The initial step size of 24° was halved after each reversal until reaching a final step size of 3°. The resulting psychometric data were fit with a cumulative Gaussian using a maximum-likelihood criterion[139, 140]. The 50% point on the fitted function was the estimate of the hinge angle that appeared as a right angle. From those estimates, we calculated the equivalent distance, the distance at which the horizontal disparities in the stimulus specify a right angle. For a target straight ahead of the viewer, the equivalent distance is:

$$D = \frac{IPD(HSR+1)}{2(HSR-1)} \tan(-\pi/4) \tag{2.5}$$

where IPD is the interocular distance and HSR is the horizontal size ratio (a measure of relative horizontal disparity[64].

There were two kinds of experimental sessions.

1. The cues-inconsistent sessions were conducted on both the volumetric and conventional displays. In those sessions, the focal distance was constant and the vergence distance varied randomly from trial to trial. For some of those sessions, the focal distance was near and the vergence distances were near, mid, and far. These conditions are numbered 1, 2, and 3 in Figure 2.2C. For some of the cues-inconsistent sessions, the focal distance was at the mid distance and the vergence distances were at the near, mid, or far distance. These are numbered 4, 5, and 6 in Figure 2.2C. For the remainder of the cues-inconsistent sessions, the focal distance was at the far distance and the vergence distances were at the near, mid, or far distance. These are numbered 7, 8, and 9 in Figure 2.2C. Note that 1/3 of the trials in the cues-inconsistent session were actually no-conflict trials; that is, the vergence and focal distances were equal to one another 1/3 of the time.

2. The cues-consistent session was conducted on the volumetric display. In that session, the vergence and focal distances were always equal to one another. The distances were near, mid, and far (3, 5, and 7 in Figure 2.2C). By comparing observer behavior in the cues-consistent session with behavior for the three equivalent vergence-focal distance combinations in the cues-inconsistent sessions (the ones in which vergence distance equaled focal distance), we could determine the consequences of changing focal distance within a block of trials as opposed to between blocks of trials.

Each session consisted of six randomly interleaved staircases, two for each vergence distance. Every session was repeated at least three times, yielding at least 100 points per psychometric function.

2.4.2 Results

The results are shown in Figure 2.6, which plots equivalent distance as a function of vergence-specified distance. The left columns in each half of the figure show data from the conventional display and the right columns from the volumetric display. Different rows show data from different observers. The red, green, and blue points are the data from the near, mid, and far cues-inconsistent sessions, respectively; the color lines are regression fits to those data. The black points and lines in the right columns are the data from the cues-consistent session. If disparity scaling were based only on the vergence-specified distance and were done without error, the data would lie on the dashed diagonal lines. The slope of the data is always less than one, which reflects the well-known under-constancy in depth perception[70]. An effect of focal distance is evidenced by vertical separation of the data. There was an effect of focal distance in the conventional-display data, which replicates the findings of Watt and colleagues[137]. There was also an effect of focal distance in the volumetric-display data.

To assess the reliability of the effect of focal distance, we performed statistical analyses on the slopes of the regression lines in Figure 2.6. Recall that a slope of one indicates that the observer was taking the change in distance into account without error (i.e., had perfect depth constancy), and that a slope of zero indicates that the observer was not taking changes in distance into account at all (i.e., he/she was accepting the same pattern of horizontal disparities as specifying a right-angle hinge no matter what the distance was). The cues-inconsistent entries to the analysis were the average slopes of the three colored regression lines in Figure 2.6. We used two kinds of cues-consistent data for the analysis: 1) data from the volumetric display in which vergence and focal distance were equal to one another and changed form trial to trial; we call these cues-consistent, within-session data; 2) a subset of the data from the cues-inconsistent sessions on the conventional and volumetric displays, the subset containing only those trials in which vergence and focal distance were equal to one another; we call these cues-consistent, between-session data. The entries for cues-consistent, between-session were the slopes of the regression lines fit to the appropriate data subset. The entries for cues-consistent, within-session were the slopes of the black lines in Figure 2.6. On the volumetric display, the slopes were greater for the cues-consistent, between-session data than for the cues-inconsistent data; this result approached statistical significance (paired t-test, $p = 0.13$, one tailed). Also on the volumetric display, the slopes were significantly greater for the cues-consistent, within-session data than for the cues-inconsistent data (paired t-test, $p = 0.02$, one tailed). On the conventional display, the slopes were marginally significantly greater for the cues-consistent, between-session data than for the cues-inconsistent data (paired t-test, $p = 0.09$, one tailed). Thus, observers exhibited more depth constancy on both displays when focal distance was consistent with vergence distance than when it was not. Said another way, presenting stimuli with appropriate focal distance increased the accuracy of depth perception.

The results are summarized in Table 2.1, which shows the slopes of the regression lines

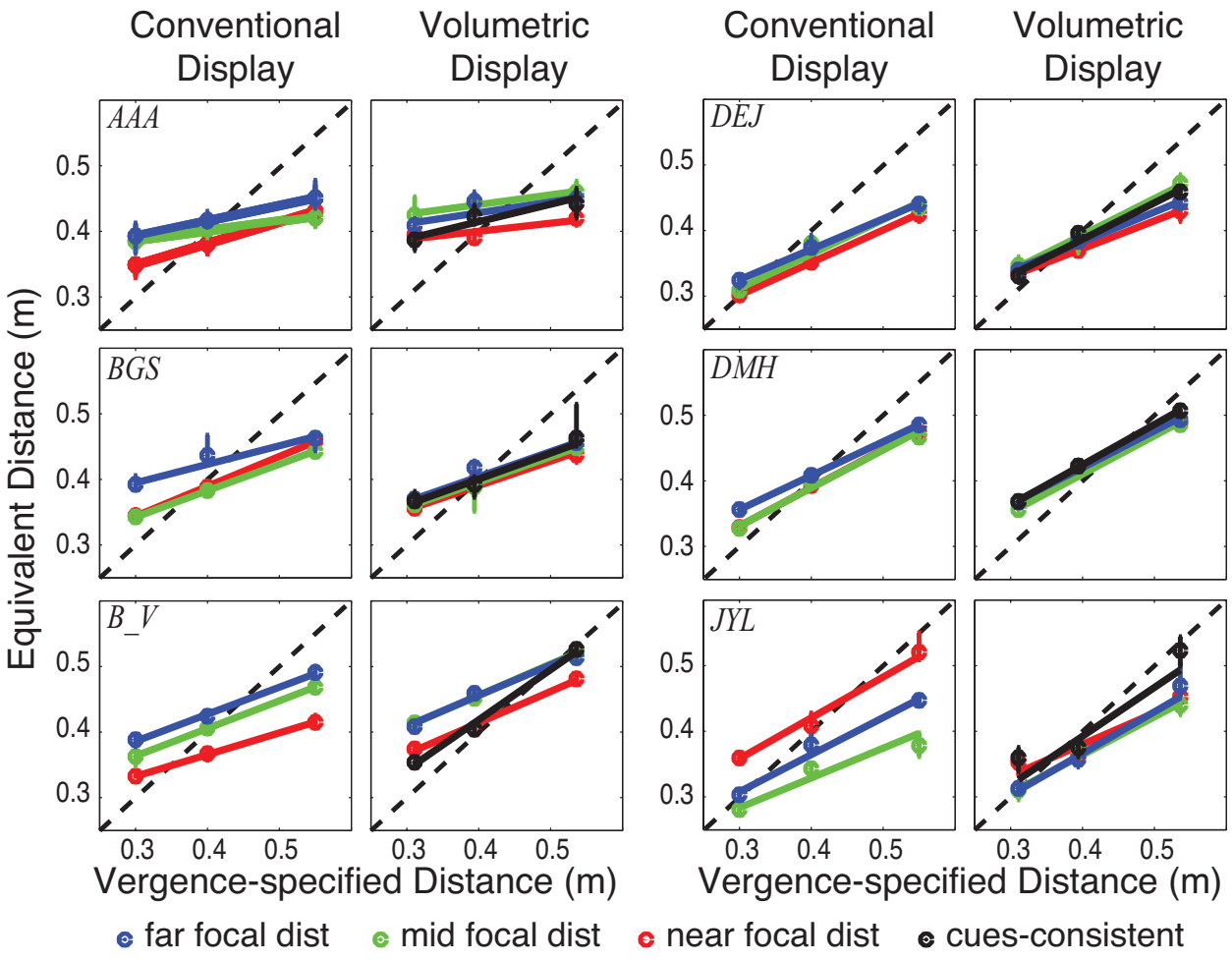


Figure 2.6: Results from Experiment 3. Equivalent distance—the distance at which the disparity setting would correspond to a right angle—is plotted as a function of vergence distance. The left columns on each half of the figure show the data obtained with the conventional display. The right columns show the data obtained with the volumetric display. Each row on each half of the figure shows the data from a different observer. The abscissa in each plot is the vergence distance. The ordinate is the equivalent distance (Equation 2.5). The diagonal dashed lines represent the predicted data if equivalent distance were based solely on the vergence-specified distance. The red, green, and blue points represent the cues-inconsistent data from the near, mid, and far focal distances, respectively (note that 1/3 of those points are in fact cues-consistent). The colored lines are regression fits to those data points. Error bars are 95% confidence intervals. The black lines in the right column represent the data from the cues-consistent session.

for each subject and condition. The cues-inconsistent values are the average slopes of the three line fits in Figure 2.6. The cues-consistent, between-session values are the slopes of the lines fitted through the cues-consistent trials in the different cues-inconsistent sessions. There were no statistically significant differences between the cues-inconsistent data collected with the conventional display and the volumetric display (paired t-test, $p = 0.50$). There were also no statistically significant differences between the cues-consistent, between-session data collected with the conventional display and the volumetric display (paired t-test, $p = 0.87$). Thus, the additional information available with the translating conventional display did not systematically affect perceived depth. This result indicates that focus cues per se were the key determinant of the improvement in depth constancy observed by Watt and colleagues[137] when focal distance was the same as vergence distance.

There was significantly greater depth constancy in the volumetric display when vergence and focal distance were equal to one another and changed within a session than when they were equal to one another, but changed between sessions (paired t-test, $p = 0.05$). Thus, trial-to-trial changes in vergence and focal distance appear to improve the accuracy of disparity scaling.

In an attempt to understand why some observers responded differently to changes in focal and vergence distance than others did, we assessed vergence-accommodation coupling in every observer using standard clinical tests. AC/A (accommodative convergence over accommodation) is a measure of how much the eyes converge when the accommodative state of one eye is driven to different values. CA/C (convergence-accommodation over convergence) is a measure of how much the eyes accommodate when vergence is driven to different values. We found no correlation between these measurements and the perceptual results in Figure 2.6 and Table 2.1. Watt and colleagues[137] made similar measurements with the same outcome.

It is possible that the between-subject variability was due to the use of a spectrally narrowband stimulus (red only). Narrowband stimuli are not an effective accommodative stimulus for many observers[41], so we might have observed less between-subject variability if we had used a white stimulus.

2.5 Experiment 4: Visual fatigue

As discussed earlier, conventional 3d displays are believed to cause fatigue and discomfort. Many researchers and engineers have assumed that the symptoms are caused, at least in part, by differences between the stimuli to vergence and accommodation because such differences require the viewer to uncouple vergence and accommodation[37, 46, 54, 65, 92, 93, 135, 146]. The evidence offered in support of this hypothesis is that viewers report more fatigue and discomfort when viewing 3d displays than when viewing 2d displays[37, 54, 69, 145, 147]. This observation, however, does not prove that vergence-accommodation conflicts cause fatigue

Subject	Experimental Condition	Conventional Display	Volumetric Display
AAA	cues-inconsistent sessions	0.24	0.15
	cues-consistent between-sessions	0.44	0.23
	cues-consistent witihn-session		0.26
BGS	cues-inconsistent sessions	0.38	0.38
	cues-consistent between-sessions	0.45	0.45
	cues-consistent witihn-session		0.4
B_V	cues-inconsistent sessions	0.39	0.46
	cues-consistent between-sessions	0.65	0.6
	cues-consistent witihn-session		0.77
DEJ	cues-inconsistent sessions	0.5	0.48
	cues-consistent between-sessions	0.62	0.49
	cues-consistent witihn-session		0.56
DMH	cues-inconsistent sessions	0.55	0.57
	cues-consistent between-sessions	0.63	0.57
	cues-consistent witihn-session		0.61
JYL	cues-inconsistent sessions	0.54	0.56
	cues-consistent between-sessions	0.37	0.48
	cues-consistent witihn-session		0.7
Average	cues-inconsistent sessions	0.43	0.43
	cues-consistent between-sessions	0.53	0.47
	cues-consistent witihn-session		0.55

Table 2.1: Results from Experiment 3. Each row shows the data for a different observer. The first column provides the observers' initials. The second indicates the experimental condition. The third shows the data from the conventional display condition, and the fourth column shows the data from the volumetric display condition. The entries in the third and fourth columns are the slopes of the linear-regression fits to the data in Figure 2.6. The entries in the cues-inconsistent condition are the averages of the three regression-line slopes for each observer. The entries in the cues-consistent, between-sessions condition are the fits to the cues-consistent data across the three inconsistent sessions (near, mid, and far vergence and focal distance). The entries in the cues-consistent, within-session condition are the fits to the data obtained with the volumetric display when focal and vergence distance were changed from trial to trial, but were always equal to one another. The entries at the bottom of the table are the slopes averaged across observers.

and discomfort because there are several other important differences between viewing 2d and 3d displays; these include the eyewear required with 3d displays to separate the two eyes' images, the ghosting or cross-talk from one eye's image to the other's[76], vertical vergence eye movements resulting from head roll and the perceptual distortions that occur with 3d displays[13] and not with 2d displays[131]. To our knowledge, no one has shown that vergence-accommodation decoupling per se causes fatigue and discomfort. Two papers—Emoto et al. (2005)[37] and Yano et al. (2004)[146]—came closest. We review those papers in the Discussion (Section 2.6.3) and make clear why they were unable to show that conflicts in the vergence and accommodative stimuli cause visual fatigue.

Our display provides a unique opportunity to test the conflict hypothesis because with this display we can independently vary vergence and focal distances without any other changes in the stimulus or task. Experiment 4 tests the hypothesis directly.

2.5.1 Methods

Observers

Eleven observers, ages 23 to 31 years, participated. We measured their stereo vision using the Titmus Stereo Test and all were normal. They all wore their usual optical corrections during the experiment. All were unaware of the experimental hypotheses, but were warned that they might experience fatigue and discomfort. We did not include older observers because they were likely to have decreased accommodative ability due to presbyopia.

Apparatus & stimuli

The experiment was conducted on the volumetric display shown in Figure 1.2. The test stimuli were disparity-defined sinusoidal corrugations as in Experiments 1 and 2. They were presented within a circular aperture with a diameter of 6.5°. Peak-to-trough disparity was 15 arcmin (three times greater than in Experiments 1 and 2) and spatial frequency ranged from 1 to 3 cpd with dot density roughly proportional to the square of spatial frequency. Observers could not perceive the corrugation waveform if they accommodated or converged inaccurately.

Procedure

We used a three-interval, forced-choice oddity task. A trial began with the presentation of a 600 ms fixation cross on the mid image plane (vergence distance = focal distance = 39.4 cm). Then three 1500 ms presentations of the test stimulus occurred separated by inter-stimulus intervals of 250 ms. Each presentation was marked with a brief audible beep. The spatial frequency of the corrugation was the same in the three intervals, but the orientation (±15° from horizontal) was different in one. The task was to identify the odd interval.

We provided auditory feedback to indicate the correctness of the response. The next trial began immediately after the observer responded or 4.5 sec after the end of the previous trial, whichever came first. Average trial duration was 6 sec. No breaks were allowed once the experimental session had started. We used this task because good performance required that the observer accommodate and converge reasonably accurately in each of the three stimulus intervals, and this was very demanding visually.

Experimental sessions had 1230 stimuli presentations and lasted about 45 minutes; each observer went through two sessions on consecutive days. During the cues-consistent session, the focal and vergence distances of the test stimulus were equal to one another at the near, mid, and far image planes (3, 5, and 7 in Figure 2.2C); those distances were randomly selected for each of the three stimulus intervals in a given trial. During the cues-inconsistent session, focal distance was fixed at the mid plane throughout, and vergence distance was randomly assigned in each interval to the near, mid, or far plane (4, 5, and 6 in Figure 2.2C). Thus, 2/3 of the intervals in the cues-inconsistent session contained vergence-focal conflicts of ± 0.67D while none of the intervals in the cues-consistent session contained conflicts. The cues-consistent and cues-inconsistent sessions were otherwise identical: same stimulus waveforms, same task, same timing, etc. Therefore, the only difference between the two types of sessions was the relationship between the vergence and focal distances. By isolating the vergence-focal conflict, we were able to directly test whether differences between vergence and focal distances contribute to fatigue and discomfort. The order of sessions for each observer was assigned randomly. The observer and experimenter were unaware of which session was being run on a given day, so the experimental design was double blind.

At the beginning of each session, the experimenter read a script explaining that we were examining visual fatigue associated with viewing 3d displays. The script also described the procedure without revealing how the experimental sessions differed. In keeping with the human-subjects protocol, the experimenter informed subjects of their right to withdraw from the study at any time. A session began with five minutes of training in the task. Training trials were the same as experimental trials except there was no conflict in the training trials (vergence and focal distances were both at the mid image plane). After training, observers were given a short break and then an experimental session began.

We wanted to assess the subjective experience of fatigue and discomfort. To do this, we had observers complete two questionnaires. They completed a symptom questionnaire at the end of each session (Figure 2.7, top). The questionnaire was modeled after one developed by Sheedy & Bergstrom[115]. It had five questions: 1) How tired are your eyes? 2) How clear is your vision? 3) How tired and sore are your neck and back? 4) How do your eyes feel? 5) How does your head feel? In each case, observers indicated the severity of their symptoms at that moment. Questions 1, 2, 4, and 5 concerned symptoms that are believed to be affected by the vergence-focal conflict in 3d displays. Question 3 concerned the neck and back, which should be unaffected by the conflict. We added that question to check whether participants were responding specifically to the queried symptom in each question or

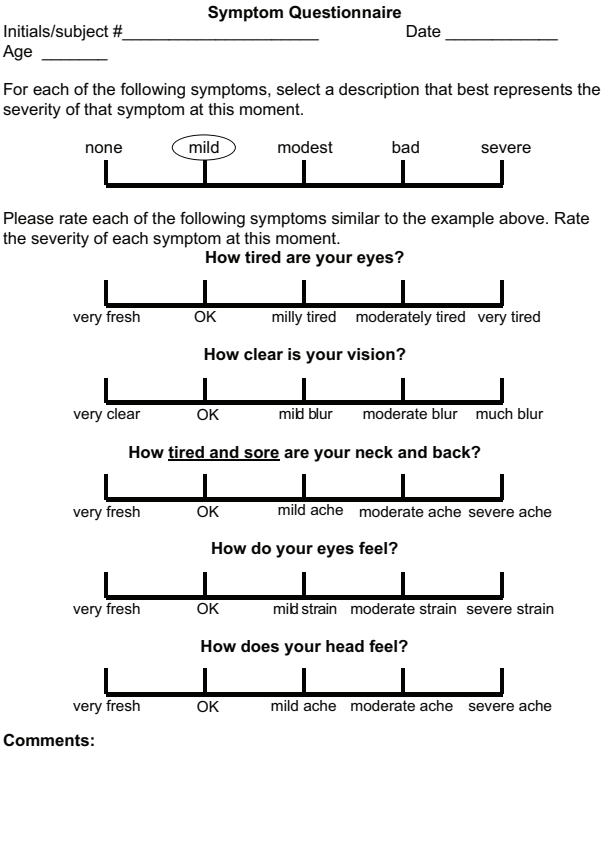
more generally. Observers also completed a display-evaluation questionnaire after the second session (Figure 2.7, bottom). It had four questions. 1) Which session was most fatiguing? 2) Which session irritated your eye the most? 3) If you felt headache, which session was worse? 4) Which session did you prefer? After observers completed the two sessions and the questionnaires, we invited them to return for another two sessions. If they agreed, the cues-consistent and cues-inconsistent sessions were presented again on consecutive days but in the opposite order from what they encountered before. Of the 11 observers, 9 returned for a second set of sessions with the ordering reversed. Three of these subjects were participants in a pilot experiment and only their second set of data is reported. Two observers declined to return for the second set of sessions. Our results are thus drawn from a total of 17 sets of questionnaires. We analyzed the percent-correct data for the orientation-discrimination task. Performance was nearly always at 80-100% correct, which means that observers were generally attentive and able to do the task. There was no systematic decline in performance over time.

2.5.2 Results

All subjects reported fatigue and/or discomfort at the end of every session, so the experiment was effective in producing the symptoms of interest. Figure 2.8 shows the average reported symptoms on the symptom questionnaire: The more severe the reported symptom, the higher the plotted bar. Orange and blue bars represent the reported symptoms for the cues-consistent and cues-inconsistent sessions, respectively. For questions 1, 2, 4, and 5, the cues-inconsistent symptoms were significantly more severe than the cues-consistent symptoms (Wilcoxon signed-rank test, $p < 0.025$, one tailed). The severity of reported symptoms did not differ significantly for question 3. Thus, when vergence and focal distance were not always the same, subjects experienced more symptoms associated with the eyes and head; as expected, they did not experience more symptoms associated with the neck and back.

Figure 2.9 shows the results for the display-evaluation questionnaire in which subjects were asked to compare their symptoms in cues-consistent and cues-inconsistent sessions. Higher values indicate more favorable ratings for the cues-consistent session in which vergence and focal distance maintained equal values even as they changed. For all four questions, the cues-consistent session received more favorable ratings: the difference was statistically significant for questions 2, 3, and 4 (Wilcoxon signed-rank test, $p < 0.025$, one tailed) and question 1 ($p < 0.05$, one tailed). Subjects experienced the two sessions in random order, so the observed difference reflects preferences for the type of session and not the order in which they were presented. The results from the display-evaluation questionnaire make clear that subjects had more favorable experiences when vergence and focal distance changed together than when vergence changed and focal distance was fixed as in conventional 3d displays.

The fatigue and discomfort results in Figure 2.8 and Figure 2.9 may actually underrepresent the size of the difference between the two experimental sessions. The two subjects



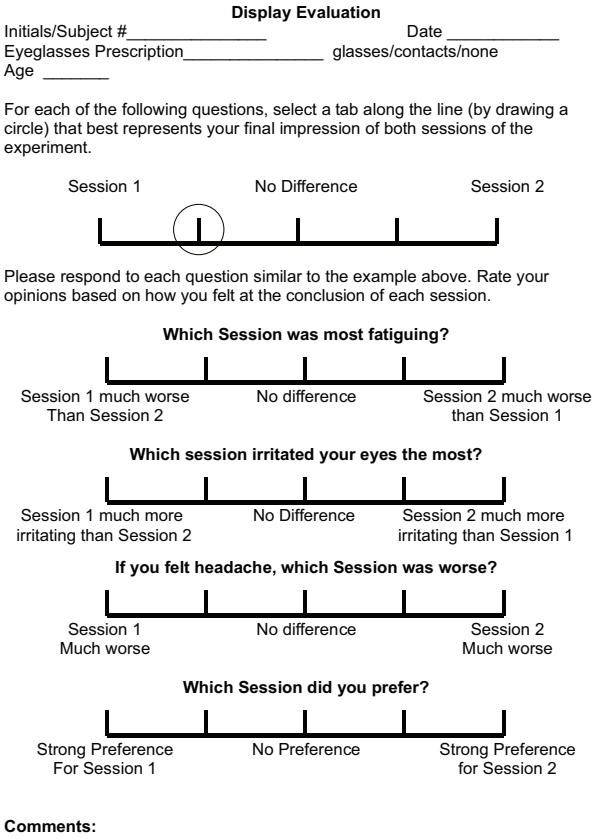


Figure 2.7: Left: Symptom questionnaire. Participants completed this questionnaire after each session. For each of the five questions, they indicated the severity of their symptoms at that moment. Right: Display-evaluation questionnaire. Participants completed this questionnaire after the second session. For each of four questions, they indicated which session was better (or worse).

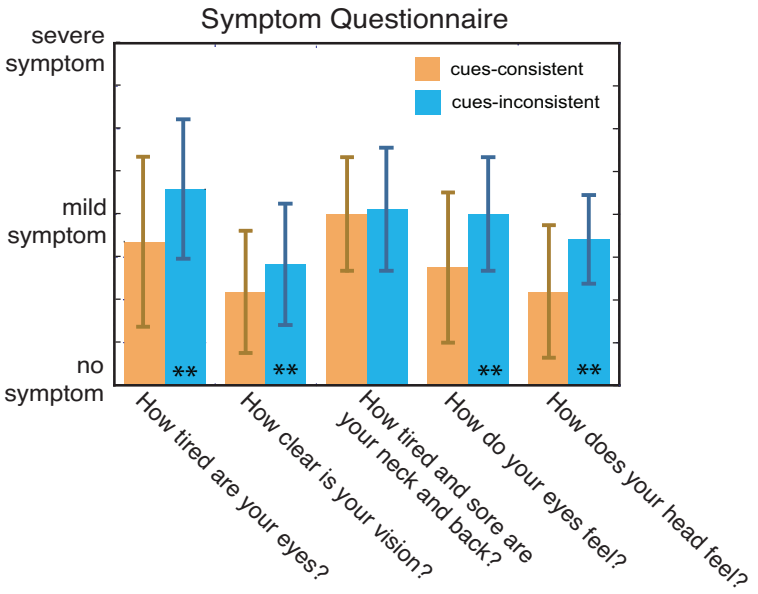


Figure 2.8: Results from the symptom questionnaire (Figure 2.7, left). The average severity of the reported symptom is plotted for each of the five questionnaire items. Blue bars are the average reported symptoms after the cues-inconsistent session, and orange bars the average symptoms after the cues-consistent session. Larger values are associated with more severe symptoms. Error bars represent ± 1 standard deviation. This graph shows data from 11 subjects, with six of them contributing twice. The double asterisks denote significantly worse symptoms reported in the cues-inconsistent, than the cues-consistent session ($p < 0.025$, Wilcoxon signed-rank test, one tailed).

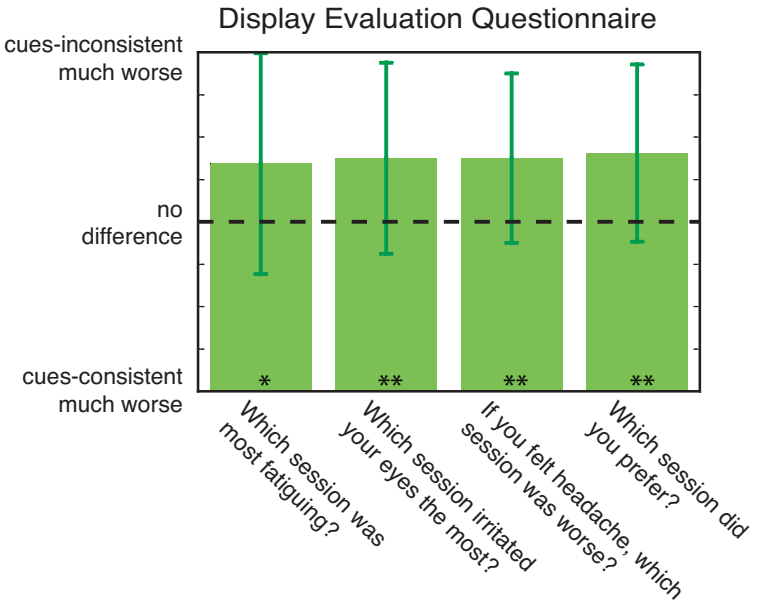


Figure 2.9: Results from the display-evaluation questionnaire (Figure 2.7, right). The average comparative rating of a pair of sessions (one cues-consistent and one cues-inconsistent) is plotted for each of the four questionnaire items. The dashed horizontal line indicates a response of no difference. Values above that line indicate more favorable responses for the cues-consistent session. Error bars represent ± 1 standard deviation. The graph shows data from 11 subjects, five of them contributing twice. (One subject misunderstood the instructions, and her first display-evaluation questionnaire was omitted.) Asterisks mean that the cues-consistent rating was significantly more favorable than the cues-inconsistent: double asterisks indicate $p < 0.025$; single asterisk indicates $p < 0.05$ (Wilcoxon signed-rank test, one tailed).

who reported the most severe symptoms during the cues-inconsistent session declined to return for the voluntary second pair of sessions. Thus, unlike the other subjects, they only contributed data from two sessions as opposed to four. These results are, we believe, the first demonstration that mismatches in the stimuli to vergence and accommodation cause visual fatigue and discomfort. As such, our findings are important for a variety of applications of 3d technology.

2.6 Discussion

2.6.1 Vergence-focal compatibility and visual performance

In Experiment 1, we found that the time required to discern the cyclopean stimulus in a random-dot stereogram was minimized when the vergence and focal distances were equal to one another. In Experiment 2, we found that finer depth corrugations could be discriminated when vergence and focal distances were the same. Both of these effects are almost certainly byproducts of the cross coupling between vergence and accommodation[42]. When the demand in the vergence stimulus is equal to the demand in the accommodative stimulus, the cross links facilitate rapid and accurate responses, so accommodation and vergence attain their end states more rapidly than when the demands are unequal[32, 77, 114].

Accommodative responses to changes in object distance consist of two components: a slow component driven by blur in the retinal image and a fast component driven by binocular disparity, the stimulus to vergence[73]. In natural viewing, the changes in vergence and focal distances are the same (meaning that the vergence and accommodative demands are equal), and the slow and fast components work together to produce a relatively rapid and accurate accommodative response. When viewing conventional 3d displays, the vergence and accommodative demands generally differ, so the slow and fast components attempt to drive accommodation to different values. The vergence-driven fast component produces a rapid response to an accommodative state that does not minimize blur. The blur-driven slow component senses the error and feeds the cross-coupled system to correct the overshoot or undershoot produced by the fast component. As a consequence, the response is slower than when the vergence and focal stimuli are consistent; indeed, when the vergence and focal stimuli differ, accommodation occasionally oscillates[32, 114]. These effects have been observed in viewers of conventional 3d displays[126].

The vergence response is also affected by the consistency of the vergence and focal stimuli, but the effects are smaller than with the accommodative response[32, 126]. In a study of 3d displays, Emoto and colleagues[37] presented two kinds of stimuli: ones in which the vergence and focal distances were consistent with one another, and ones in which vergence distance was varied and focal distance was not. Observers reported many more fusion failures (i.e., diplopia) with the latter stimuli, which indicates that vergence was less accurate when

vergence-focal conflicts were present.

The stimulus in Experiments 1 and 2 required accurate accommodation because retinal-image blur impedes the ability to discern high-frequency depth corrugations[10, 97]. Thus, conflicts between the stimulus to vergence and the stimulus to accommodation undoubtedly affected the speed and accuracy of the accommodative response thereby producing some, if not all, of the effects we observed.

It is interesting to note that the accommodation response produced by the vergence-accommodation system is quite dependent on the spatial-frequency content of the stimulus. When the stimulus contains high frequencies, changes in retinal-image blur are readily detected, so the blur-driven slow component drives the cross links yielding corrections, even oscillations, for hundreds of milliseconds[99]. In contrast, when the stimulus contains only low spatial frequencies, changes in retinal-image blur are hard or impossible to detect, so the blur-driven component does not drive the system to correct accommodative errors; rather accommodation settles relatively quickly near the value specified by the vergence stimulus[99]. Because of this, the need to minimize the vergence-focal conflict becomes greater and greater for stimuli with increased detail.

Many viewers cannot fuse a binocular stimulus with a vergence-focal conflict. We documented this in the preliminary testing for Experiment 3. We presented our stimuli to 11 young subjects. Three could not fuse most of the stimuli. Another two could fuse the stimuli, but complained that doing so was too fatiguing. Only six subjects could fuse all the stimuli without significant fatigue and discomfort and they were chosen for the main experiment. All of the disqualified subjects had normal binocular vision and could also fuse binocular stimuli in the natural environment across a wide range of distances. They only had problems with the experimental stimuli in which vergence-focal conflicts were present. The fact that nearly half of observers, all of whom were young adults with normal binocular vision, could not readily fuse stimuli with vergence-focal conflicts is a reminder that researchers and 3d-display engineers should minimize the conflict in their displays to increase the number of people who can participate.

2.6.2 Focus cues and depth perception

In Experiment 3, we investigated the consequences of vergence-focal conflicts on the perception of 3d shape. Although the effect was small, perceived depth was consistently more accurate when the vergence and focal distances were the same as opposed to when they differed. This effect was presumably caused by changes in the distance estimate the visual system uses to scale horizontal disparities[48, 137].

The fact that focus cues can affect 3d percepts has significant implications for psychophysical research on depth perception. The great majority of experiments are conducted with conventional 3d displays in which focus cues specify the flat surface of the display rather than the simulated depth of the experimental stimulus. Consider, for example, research on

the perception of structure from motion. The retinal-image motion caused by relative movement between the viewer and object creates a compelling 3d impression, but in laboratory experiments, the judged depth is often smaller than the simulated depth[18, 25, 35, 83, 125]. For instance, Hogervorst[61] presented monocular structure-from-motion stimuli that simulated hinges similar to the ones in our Experiment 3. As in our experiment, observers judged the angle between the sides of the hinge. They consistently overestimated the angle (underestimated the depth). The results were consistent with a Bayesian model incorporating optic flow measurements and their associated noises. However, depth underestimation would also be expected if focus cues affected observers' percepts because such cues specify the flat surface of the display rather than the varying distance of the sides of the simulated hinge.

Hogervorst and Eagle[62] tested for effects of focus cues by redoing part of the experiment with a pinhole pupil. They reasoned that a pinhole would render focus cues uninformative, so if focus cues had contributed to the depth underestimation in their main experiment, they would observe less underestimation with a pinhole than with a natural pupil. However, pinholes do not necessarily render focus cues uninformative; instead they may cause the retinal blur and accommodative signals to be interpreted as specifying flatness[43, 137]. For this reason using pinholes with conventional displays is not an adequate test for the influence of focus cues. By not including signals that may well have affected observers' percepts, Hogervorst and Eagle may have misinterpreted their data yielding an erroneous theoretical account.

Todd and colleagues have repeatedly observed non-veridical depth estimation when 3d structure is specified by disparity, motion, pictorial information, and combinations of those cues (reviewed by Todd, 2004[124]). Their stimuli are also generally presented on computer displays, so focus cues may have contributed to the compression in perceived depth. Todd and colleagues have discounted this concern by noting that non-veridicality is frequently observed with real stimuli as well (e.g., Baird & Biersdorf, 1967[8]). But to gain a full understanding of depth perception, we need to quantify what the depth percepts are, not just whether or not they are veridical. Quantification could be improved by evaluating the contribution of focus cues in such experiments.

2.6.3 Focus cues and visual fatigue

As we noted in the Introduction, many authors have stated that conflicts in the stimuli to vergence and accommodation are the cause of the well-known visual fatigue and discomfort associated with conventional 3d displays[65, 135]. Despite this widespread belief, no one has proven (with the appropriate control conditions) that the stimulus conflict causes fatigue and discomfort. In our opinion, the two reports that come closest are Emoto, Niida and Okano (2005)[37] and Yano, Emoto and Mitsuhashi (2004)[146].

In a study involving the manipulation of vergence and focal demands while viewing displayed images, Emoto and colleagues (2005)[37] had subjects view a 3d display under

different viewing conditions. There was a range of disparities in the scene. The absolute disparity of the entire scene was changed with variable-power prisms. In one condition, the absolute vergence stimulus was varied periodically while the focal stimulus was fixed. In a second condition, the vergence and focal stimuli were periodically changed in a fashion that made them somewhat similar to one another. Subjects reported the most fatigue when the prism power was changing, but the differences across conditions were not statistically significant. The authors concluded that vergence-accommodation conflicts cause visual fatigue. As we implied earlier, this conclusion is not warranted for three reasons. 1) The 3d movie that served as their stimulus presented multiple potential vergence distances, so Emoto and colleagues could not know the size of the conflict between the stimulus to vergence and the stimulus to accommodation for any of their experimental conditions: it would have depended on where the subject was looking in the virtual scene. 2) The task—counting the instances of a particular Japanese character—could have been performed monocularly and therefore did not require that the observer resolve the vergence-accommodation conflict. 3) The vergence stimulus was varied by changing the power of prisms in front of the observers' eyes and such prisms produce spatial distortions that can cause fatigue[76].

Yano and colleagues[146] presented stereo images of a page of text on a conventional 3d display. The subjects were instructed to read text. The focal distance of the text was constant, but the disparity-specified distance of the text differed from one experimental session to another. At the end of each session, they assessed subjects' fatigue and discomfort. The severity of the reported symptoms grew monotonically with the magnitude of the conflict between the focal and disparity-defined distances of the text. Yano and colleagues concluded that vergence-accommodation conflicts cause visual fatigue and discomfort in conventional 3d displays. This conclusion is not justified for two reasons. First and most important, the vergence-focal conflict co-varied with the vergence demand, so one cannot determine from the results whether the conflict or vergence demand caused the fatigue. Second, the subjects' task did not require stereopsis, it could have been performed monocularly.

Thus, to our knowledge, it has never been shown in an experiment with appropriate control conditions that differences in the stimuli to vergence and accommodation actually cause viewer fatigue and discomfort. We performed a direct test of the conflict hypothesis in Experiment 4. We isolated the variable of interest—the vergence-focal conflict—while keeping all other aspects of the stimulus and procedure the same. Furthermore, the task—discriminating the orientation of a corrugation in a random-element stereogram—could only be performed binocularly and thus required stereopsis. We believe this is the first demonstration that this conflict is a cause of visual fatigue and discomfort.

The fatigue and discomfort produced by vergence-focal conflicts are presumably byproducts of the cross coupling between vergence and accommodation[42]. When disparity is present, commands are sent to the extra-ocular muscles to modify vergence and minimize the disparity. When blur is present, commands are sent to the ciliary muscles to adjust accommodation and minimize the blur. However, vergence and accommodation are ad-

ditionally complicated because of the vergence-accommodation cross coupling: changes in disparity affect accommodation via vergence commands and changes in blur affect vergence via accommodative commands. When the vergence and focal stimuli are consistent with one another, the commands are compatible and the two systems can adjust to appropriate states relatively quickly. When the vergence and focal stimuli differ, the commands are no longer compatible. Presumably, the difficulty in resolving the incompatibility is an important contributor to fatigue and discomfort. Consistent with this argument, Wann & Mon-Williams[135] observed a change in the coupling between vergence and accommodation (i.e., changes in the AC/A and CA/C ratios) after prolonged viewing of a stereoscopic display.

2.6.4 Transient versus sustained changes in focal distance

In Experiment 3, there were two kinds of experimental sessions presented with the volumetric display: 1) sessions in which the focal distance was the same for the whole block of trials and vergence distance changed from trial to trial (in 1/3 of those trials, the focal and vergence distances were the same), and 2) sessions in which the focal distance and vergence distance were always equal to one another and changed from trial to trial. We compared performance in the cues-consistent trials from those two sessions (Table 2.1, cues-consistent, between-session trials and cues-consistent, within-session trials). When we did so, we found that shape perception was more accurate in the cues-consistent, within-session trials than in the cues-consistent, between-session trials. This seems surprising because with trial-to-trial changes in focal distance, the system has less time to accommodate to the new distance and to estimate accommodative state than it has when focal distance is fixed within a block. We offer two hypotheses to explain this observation. First, the cues-consistent trials in the first session were presented among cues-inconsistent trials in which the vergence and focal distances differed. Perhaps the presence of those inconsistent trials caused changes in the vergence-accommodation system (i.e., adaptation of the cross links) that affected performance in the cues-consistent trials[112].

Second, perhaps the visual system can take transient changes in accommodative state into account better than sustained changes. There are examples of similar phenomena in the literature. For example, consider the problem of judging the perceived direction of a target relative to the head when the eyes are in eccentric gaze. To estimate direction correctly, the visual system must measure the position of the target on the retina and the position of the eyes relative to the head. Direction judgments are more accurate shortly after the eye movement is made than after the eyes are held in the eccentric position for a while[102]. Thus, transient changes in eye position are taken into account better than sustained changes, and the same may apply to accommodation.

Chapter 3

Focus information is used to interpret binocular images

In addition to producing incorrect stimuli for accommodation, conventional 3d displays also produce retinal blur signals that are inconsistent with the scene's depth. I alluded to this with Figure 1.1 showing that the blur gradient signal for the faces of the hinge was absent in the display. Often, blur is described as a qualitative depth cue, or an artistic tool rather than an important depth signal. In this chapter I describe the work was recently accepted for publication at the Journal of Vision. In this paper we examined the use of blur in solving the binocular correspondence problem and in interpreting monocular occlusions.

3.1 Introduction

To measure 3D scene layout from disparity, the visual system must correctly match points in the two retinal images, matching points that come from the same object in space and not points that come from different objects. This is the binocular correspondence problem. In solving binocular correspondence, the visual system favors matches of image regions with similar motion, color, size, and shape [78, 130]. Matching by similarity helps because image regions with similar properties are more likely to derive from the same object than image regions with dissimilar properties. The preference for matching similar regions is a byproduct of estimating disparity by correlating the two eyes images[10].

Focus information—blur and accommodation—also provides potentially useful information for binocular correspondence. The accommodative responses of the two eyes are yoked[9]. If one eye is focused on an object, the other eye will generally be as well, and the two retinal images will be sharp. Likewise, if one eye is not focused on an object, the other eye will also generally not be focused on that object and the two retinal images will be blurred. Thus, accommodating affects blur in the retinal images, and blur similarity could distinguish otherwise similar objects and thereby simplify binocular matching.

In conventional stereo displays, the relationship between scene layout and focus information is greatly disrupted. Conventional displays have one image plane, so all the pixels have the same focal distance. If objects in the simulated scene are rendered sharply and the eye accommodates to the screen, all objects produce sharp retinal images; when the eye accommodates elsewhere, they produce equally blurred images. In those cases, blur and accommodation provide no useful information about the 3D layout of the simulated scene. If some objects are rendered as sharp and others as blurred, the differential blur in the retinal images provides useful 3D information, but changes in accommodation do not affect retinal blur as they would in natural viewing. Again the natural relationship between accommodation and blur is disrupted.

The discrepancy between real-world focus cues and those in conventional stereo displays is greatest in scenes containing large depth discontinuities. In this paper, we analyze situations in which focus cues in real and simulated worlds differ. We begin with a theoretical analysis that outlines the potential usefulness of focus information for binocular viewing of objects in the natural environment. We then describe a psychophysical experiment in which focus information was manipulated; this manipulation allowed us to assess the use of blur and accommodation in binocular matching. Finally, we describe how focus information could aid the interpretation of images in which monocular occlusion occurs.

3.1.1 Is focus information a useful cue for binocular matching?

The creation of blur and binocular disparity derives from the same viewing geometry[109, 56]. To show the fundamental similarity, we first consider blur. On the left side of Figure 3.1, a lens of focal length$f$ focuses an object at distance d_0 onto the image plane. Another object is closer to the lens at distance d_1 and its image is formed behind the image plane creating a blur circle of diameter b_1. Using the thin-lens equation and ray tracing, Held et al.[56] showed that:

$$b_1 = A\frac{s_0}{d_0}\left|1 - \frac{d_0}{d_1}\right|$$

where A is the pupil diameter and s_0 is the posterior nodal distance (roughly the distance from the lens to the retina). This equation does not incorporate all of the eyes aberrations (i.e., diffraction, higher-order aberrations), but provides an accurate approximation of blur when the eye is defocused[28]. Rearranging, we obtain:

$$b_1 = As_0\left|\frac{1}{d_0} - \frac{1}{d_1}\right| \tag{3.1}$$

Using the small-angle approximation, we can express the blur-circle diameter in radians:

$$\beta_1 = 2\tan^{-1}\left(\frac{b_1}{2s_0}\right) \approx \frac{b_1}{s_0} \approx A\left|\frac{1}{d_0} - \frac{1}{d_1}\right|. \tag{3.2}$$

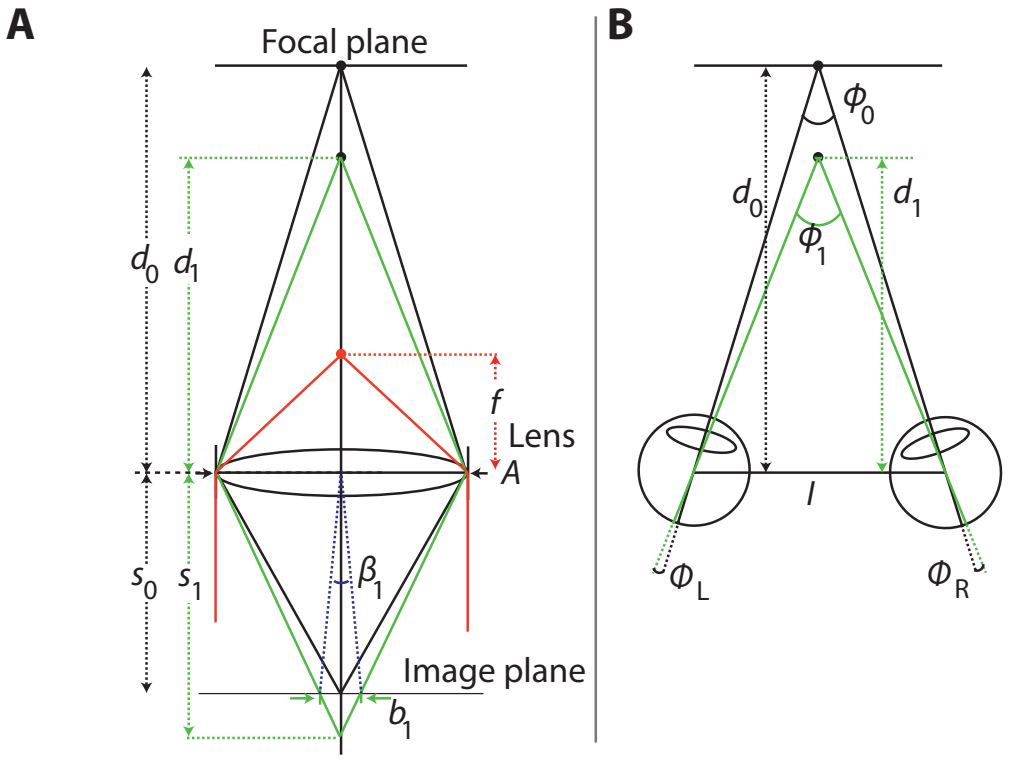


Figure 3.1: The viewing geometries underlying blur and disparity. A) Formation of blur. Lens of focal length f is focused on an object at distance d_0. Another object is presented at distance d_1. The aperture diameter is A. The images are formed on the image plane at distance s_0 from the lens. The image of the object at d_1 is at best focus at distance s_1 and thus forms a blur circle of diameter b_1. That blur circle can be expressed as the angle β_1. B) Formation of disparity. The visual axes of the two eyes are converged at distance d_0. The eyes are separated by I. A second object at distance d_1 creates horizontal disparity which is shown as the retinal angles ϕ_L and ϕ_R.

Thus, the angular size of the blur circle is directly proportional to the pupil diameter and the dioptric separation of the two object points (Eqn 3.1). We will capitalize on the fact that blur is proportional to depth separation in diopters throughout the remainder of this paper.

Now consider the creation of disparity with the eyes separated by distance I and converged on the object at distance d_0. The object at distance d_1 creates images at positions ϕ_L and ϕ_R in the two eyes. Using the small-angle approximation, we obtain:

$$\begin{aligned}\delta_1 = \phi_L - \phi_R &= 2\left[\tan^{-1}\left(\frac{I}{2d_0}\right) - \tan^{-1}\left(\frac{I}{2d_0}\right)\right] \\ &\approx I\left(\frac{1}{d_0} - \frac{1}{d_1}\right). \end{aligned} \tag{3.3}$$

Combining Eqns 2 and 3, we obtain:

$$\beta_1 = \frac{A}{I}\left|\delta_1\right| \tag{3.4}$$

Thus, the magnitudes of blur and disparity are proportional to one another: specifically, the blur-circle diameter is A/I times the disparity. In humans, the inter-ocular separation I is roughly 12 times greater than the steady-state pupil diameter A[33], so blur magnitude should be roughly 1/12th of the disparity magnitude. The correlation between blur and disparity suggests that focus information could be useful in correlating the two eyes image, an essential part of disparity estimation.

This relationship of blur and disparity could be a useful signal for binocular processes, but only if it satisfies two requirements: 1) The same object should produce similar blur in the two retinal images, and 2) different objects should produce dissimilar blur in the two images. We consider those requirements in the next two sections.

3.1.2 Inter-ocular focal differences for one object

Accommodation is yoked between the two eyes; objects that are equidistant from the eyes will yield the same blur in the retinal images (assuming that differences in refractive error between the eyes—anisometropia—have been corrected with spectacles or contact lenses). However, objects at eccentric viewing positions (i.e., not in the heads sagittal plane) are closer to one eye than the other, so they yield inter-ocular focal differences (Figure 3.2A). We estimated the viewing situations in which those differences would be perceptible.

The distances from a point object to the eyes are given by:

$$d_R = \sqrt{\left(\frac{I}{2}\right)^2 + D^2 - DI\sin(\gamma)}$$
$$d_L = \sqrt{\left(\frac{I}{2}\right)^2 + D^2 + DI\sin(\gamma)} \tag{3.5}$$

where I is inter-ocular distance, D is distance from the cyclopean eye to the object point, and γ is the azimuth of the object (Figure 3.2A). It is useful to express the differences in distances to the two eyes in diopters (D, the reciprocal of distance in meters) because blur is proportional to distance in diopters (Eqn 3.2[68]). Thus, inter-ocular focal difference is the dioptric difference of the object to the two eyes:

$$\Delta D = \left|\frac{1}{d_L} - \frac{1}{d_R}\right| \tag{3.6}$$

Figure 3.2 shows inter-ocular focal difference as a function of object position. The abscissa and ordinate represent respectively the horizontal and in-depth coordinates of the object relative to the cyclopean eye. Each contour indicates the locations in space that correspond to a given inter-ocular focal difference. We assume that the eyes depth of focus under most viewing situations is 0.33D[22, 133], so any difference greater should be perceptible (shaded in blue). These regions are both eccentric and near. They are rarely fixated binocularly for two reasons. First, people usually make head movements to fixate objects that are more than 15° from straight ahead[7]; that is, they do not typically make version eye movements greater than 15°. In Figure 3.2B, the typical version zone is represented by the dashed lines, and version movements outside this zone are shaded gray. Second, people have a tendency to step back from objects that are too close. A typical accommodative amplitude is 8D and thus we have shaded in red the region closer than this. Figure 3.2B shows that the region in which focal differences would be perceptible and would correspond to azimuths and distances that people would normally fixate is vanishingly small. We conclude that single objects viewed naturally would almost never create perceptually distinguishable retinal blurs. By inference, perceptually distinct blurs are almost always created by different objects.

3.1.3 Inter-ocular focal differences for different objects

We next estimate the likelihood that two objects, A and B, will produce perceptually distinct retinal blurs (Figure 3.3A). Assuming that the eye is focused on A, the defocus for B is given by:

$$\Delta D = \frac{1}{d_A} - \frac{1}{d_B}. \tag{3.7}$$

Thus, the focal difference depends on the distance to A and the separation between the objects. Figure 3.3B plots focal difference as a function of accommodative distance (d_A)

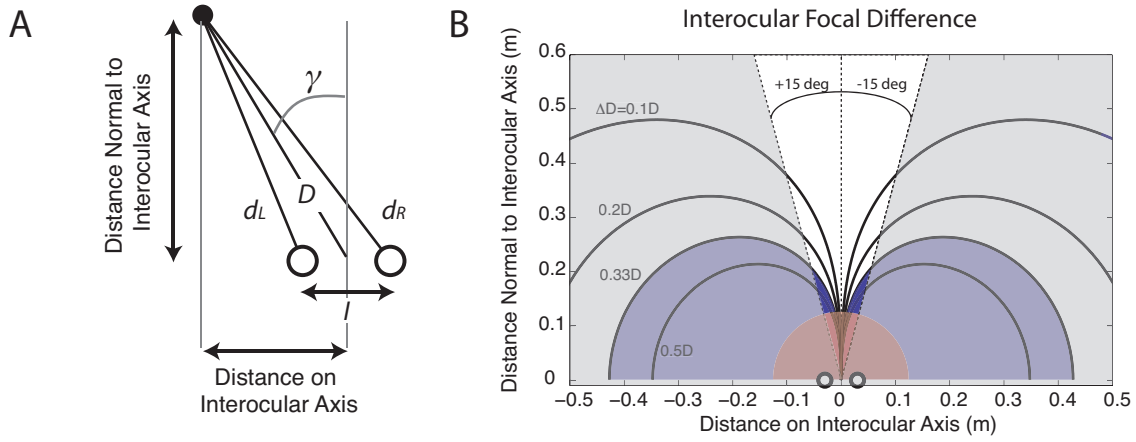


Figure 3.2: Inter-ocular differences in focal distance. A) Object viewed binocularly. An object is presented at distance D and azimuth γ from the eyes, which are separated by I. The distances from the object to the two eyes are d_L and d_R. B) Interocular focal difference as a function of object position. The abscissa is object position parallel to the inter-ocular axis and the ordinate is object position perpendicular to that axis. Interocular focal difference is the difference between the reciprocals of the distances to the two eyes in meters (Eqn 3.6). The black contours depict the locations in space where the inter-ocular focal difference is constant. The blue region indicates the locations where the focal difference is greater than the eyes depth of focus and therefore the focal difference should be perceptible. The gray region shows the range of positions that an observer would typically make a head movement to fixate. The red zone marks the region where objects are closer than 8D, and an observer would typically step back to fixate. The dark blue zone is the region where fixation is possible and there is a perceptible interocular focal difference.

and depth separation ($d_B - d_A$). The translucent plane represents the eyes depth of focus of 0.33D, a reasonable value for a wide range of viewing conditions. Any focal differences above that plane should be perceptually distinguishable and any differences below the plane should not be distinguishable. There is a high probability of observing perceptually distinct focal differences when the target distance is short; in that case, even small depth separations create distinguishable focal differences. The figure also shows that the probability of observing distinct differences diminishes as target distance increases; when the target distance exceeds 3m (0.33D), no physically realizable separation can create distinguishable focal differences.

The height of the threshold focal difference (the green translucent plane) depends on the quality of the eyes optics, the pupil size, and the properties of the stimulus[133]. We assumed a value of 0.33D because that is representative of empirical measurements with young observers and photopic light levels[22]. Other values have been reported. For example, Atchison, Fisher, Pedersen, and Ridall [4] observed a depth of focus of 0.28D with letter stimuli and large pupils. Walsh and Charman [133] found that focal changes as small as 0.10D can be detected when the eye is slightly out of focus. Kotulak and Schor[66] reported that the accommodative system responds to focal changes as small as 0.12D, even though such small changes are not perceptible. We have ignored these special cases for simplicity. If we had included them, the threshold plane would be slightly lower in some circumstances, but our main conclusions would not differ.

The figure shows that there is a wide variety of viewing situations—particularly near viewing—in which two objects will generate perceptually distinguishable blurs in the retina. An interesting target distance is 50 cm (dark contour in Figure 3.3B), which corresponds to the typical viewing distance for a desktop display and to the display used in many experiments in vision science. When the eye is focused at 50 cm, another object would have to be only 10 cm farther to produce perceptibly different blur. In this case, the differences in blur would be potentially useful for interpreting the retinal images. This amount of blur could only occur with multiple objects because at 50 cm, there is no azimuth at which one object could produce the same difference in retinal blur.

In a conventional stereo display, such differences would be either not present or would not be properly correlated with accommodation.

Having established that blur can provide useful information for interpreting binocular images, we next describe psychophysical experiments that examined whether humans actually capitalize on this information.

3.2 Experiment 1A: Binocular matching with and without appropriate focus information

We used stereo-transparent stimuli to investigate the visual systems ability to solve the binocular matching problem with and without appropriate focus information. We used

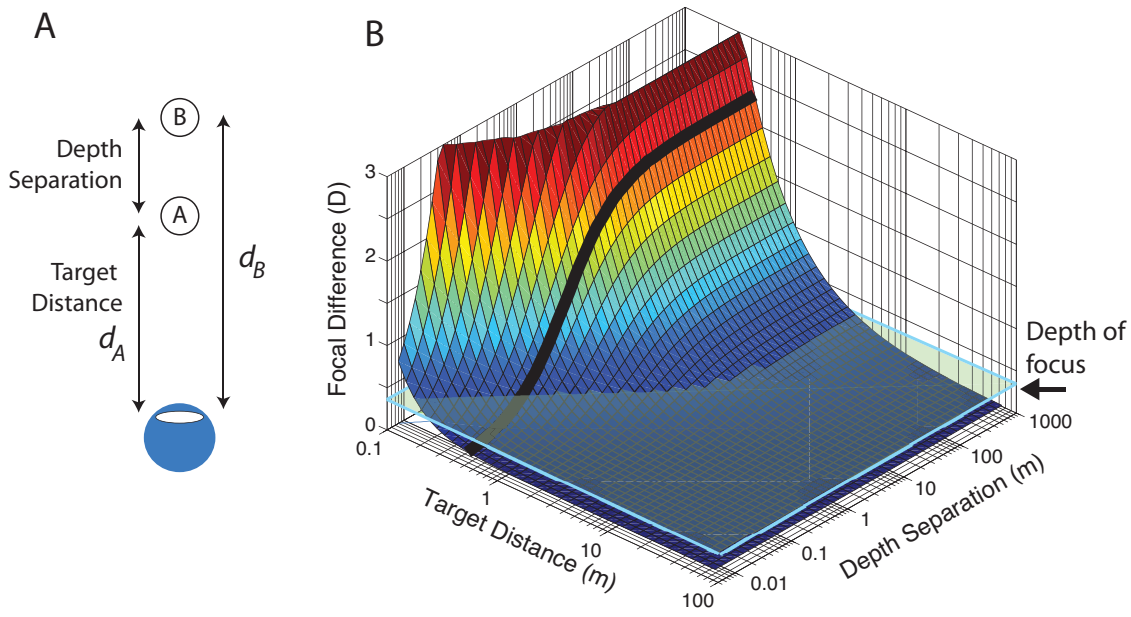


Figure 3.3: Difference in focus for two objects. A) Objects A and B at distances d_A and d_B. The eye is focused on A. B) Focal difference (Eqn 3.7) plotted as a function of target distance d_A and the depth separation between objects: $d_B - d_A$. The translucent green plane represents depth of focus, the just-noticeable change in focus. We assumed a value of 0.33D[22, 133]. The heavy curve represents a fixed viewing distance of $d_A = 50$ cm.

transparent stimuli because matching is particularly difficult with such stimuli[2, 130, 127]. When the surfaces have different colors, matching performance is significantly improved[2]. We wanted to know if blur, like color, can help the visual system distinguish the surfaces of a stereo-transparent stimulus.

3.2.1 Methods

Apparatus

We needed a means of presenting stimuli in which we could show geometrically equivalent stimuli with and without appropriate focus information. We achieved this by using the fixed-viewpoint, volumetric display described in Section 1.2.1. By using this display, we could present geometrically identical retinal images with or without correct focus cues.

Observers

We screened observers by first measuring their stereoacuity using the Titmus Stereo test. We then had them practice the task. Nine observers who did not have normal stereoacuity or could not consistently fuse the stimulus in the experiment were dismissed. Two experienced psychophysical observers, BV (age 34) and KY (22) participated. Both were emmetropic and naïve to the experimental hypothesis.

Stimulus

The two planes in the stimulus were each textured with 10 white lines of various orientations on black backgrounds. The lines were 8.5 arcmin thick and their ends were not visible. The average orientation of the lines in one eyes image was 0° (i.e., vertical), but we presented five orientation ranges (uniform distribution): 0, ±3, ±6, ±12, and ±24°. Previous research had shown that solving the correspondence problem is relatively easy when the range of orientations is large and is quite difficult when the range is small[130, 34]. Figure 3.4 is a schematic of the stimulus configuration. The near stimulus plane was fronto-parallel and the farther plane was slanted about a vertical axis (tilt = 0°). Both planes were 14° wide and 8.8° tall.

The stimulus manipulation of primary interest was the appropriateness of focus information. On half the trials, both stimulus planes were rendered sharply and displayed on the mid image plane at 39.4 cm (Figure 3.5A); thus, blur in the retinal images was the same for both planes despite the difference in the disparity-specified depth. If the observer accommodated to the image plane at 39.4 cm, both planes were sharply focused in the retinal images. On the other half of the trials, the stimulus planes were presented on multiple image planes so the blur in the retinal images was consistent with the disparity-specified depth (Eqn 3.4; Figure 3.5B). On those trials, the near stimulus plane was presented on the near image plane

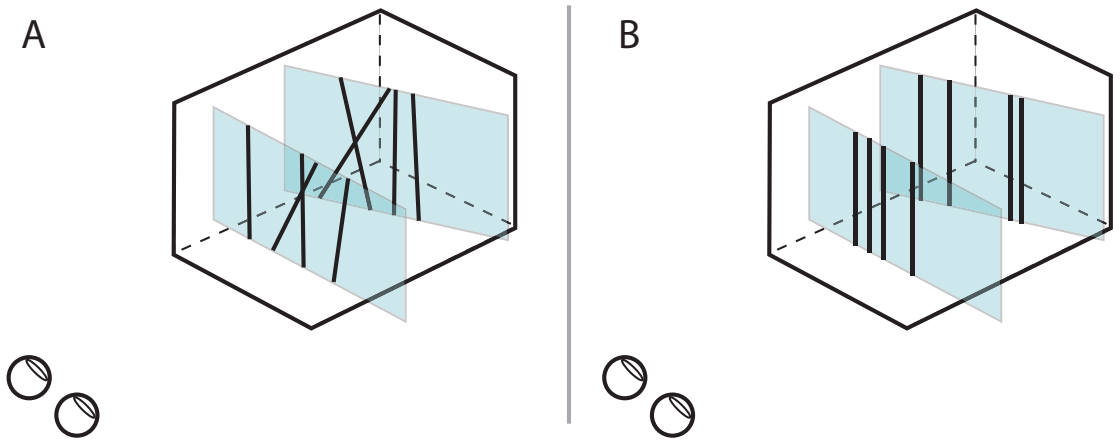


Figure 3.4: Schematic of the experimental stimuli. The stimuli consisted of a near fronto-parallel plane and a more distant plane that was slanted about the vertical axis (tilt = 0°). The planes were each textured with 10 lines. The ends of the lines were not visible. A) Stimulus when the orientation range is large. B) Stimulus when the orientation range is 0°.

at 31.1 cm and the far stimulus plane was presented mostly on the mid image plane at 39.4 cm. Because the far stimulus plane was generally slanted with respect to fronto-parallel, that plane was actually rendered and displayed with depth-weighted blending of light intensities for the near, mid, and far image planes[1]. If the observer accommodated to the mid image plane, the far stimulus plane would be sharply focused in the retinal images, and the near plane would be appropriately blurred. The blur of the front plane corresponded to a focal difference of 0.66D. The blur would of course also depend on the observers pupil size. The space-average luminance was 5.5 cd/m^2, which should yield an average pupil diameter of 4.2mm[33].

Some example stimuli are shown in Figure 3.6. Cross fuse to see in depth. A and B have orientation variation, while C and D do not. The lines in A and C are rendered sharply while in B and D the lines in the near plane are rendered blurred. Most people find it easier to perceive the front and back planes when blur is appropriate and when the orientation range of the lines is large.

Procedure

Stimuli were presented for 1500 msec in a 2-alternative, forced-choice procedure. Observers indicated after each stimulus presentation whether the back plane was positively or negatively slanted (right side back or left side back, respectively). A 2-down/1-up staircase adjusted the magnitude of the slant to find the value associated with 71% correct. Step size decreased from 4 to 2 to 1° after each staircase reversal and remained at 1° until the staircase was terminated after the ninth reversal.

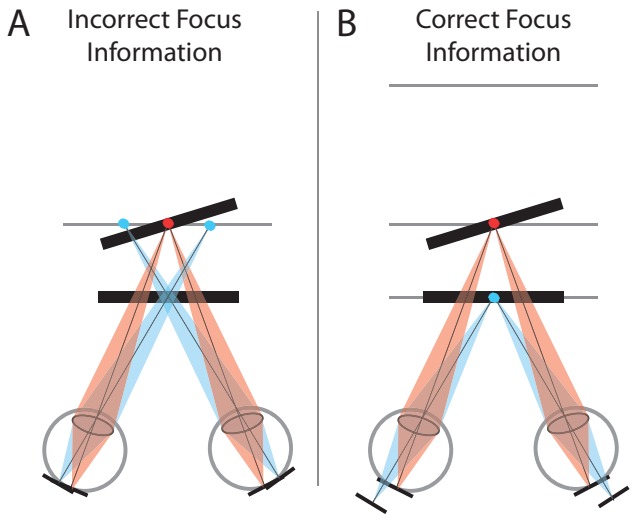


Figure 3.5: Focus information for the experimental stimuli. The dark line segments represent the disparity-specified planes. The gray lines represent the image planes. The blue and red dots represent the focal stimuli. A) Incorrect focus information. Both stimulus planes are presented on the mid image plane. When the eyes are accommodated to that plane, the retinal images of both stimulus planes are sharp. B) Correct focus information. The near and far stimulus planes are displayed on the near and mid image planes, respectively. The small dark lines at the bottom indicate the locations of the images of the stimulus planes.

All five orientation ranges and both focal conditions were presented in an experimental session. Thus, there were 10 interleaved staircases in a block of trials. The block was repeated eight times so there were approximately 180 stimulus presentations for each stimulus condition. We plotted the data for each condition as a psychometric function and fit the data with a cumulative Gaussian using a maximum-likelihood criterion[139, 139]. The mean of the best-fitting Gaussian was an estimate of the 75% correct point, which we define as discrimination threshold.

3.2.2 Results

Figure 3.7 plots discrimination threshold for each stimulus condition. The just-discriminable change in slant is plotted as a function of line-orientation range. The black circles represent data from the incorrect focus cues condition in which both stimulus planes were presented on the mid image plane. Discrimination thresholds were large when the range of line orientations was smaller than 6° and improved when the range was greater than 6°. The green triangles represent the correct focus cues data in which the front stimulus plane was presented on the near image plane and the back stimulus plane was presented on the mid image plane. In this condition, observers could discriminate slant equally well for all orientation ranges. We subjected the data to statistical test and found no significant independent effects of orientation range or focus cues; however, the range x focus interaction approached significance (ANOVA (F=3.77, d.f. =4) p=0.055, one tailed).

The results of Experiment 1A are consistent with the use of focus information in solving the correspondence problem, but there is a possible alternative explanation. In the correct-focus-information condition, the back stimulus plane was displayed on three image planes with depth-weighted blending and hence it had a blur gradient proportional to its slant. In other words, there was a large difference in the focal distances of the near and far image planes, and a small blur gradient on the far stimulus plane. In the incorrect-focus-information condition, both of the stimulus planes were presented on the same image plane, which eliminated both the small blur gradient and the large focal difference between the stimulus planes. Thus, the blur gradient could have provided slant information when focus cues were correct but not when those cues were incorrect, and observers could have used that monocular information to make responses without solving the binocular matching problem.

3.3 Experiment 1B

The slant-discrimination thresholds in Experiment 1A were as large as 8°. At that slant, the dioptric difference between the near and far edges of the plane were 0.1D, which is probably sub-threshold[22]. But just to make sure, we reduced the informativeness of the

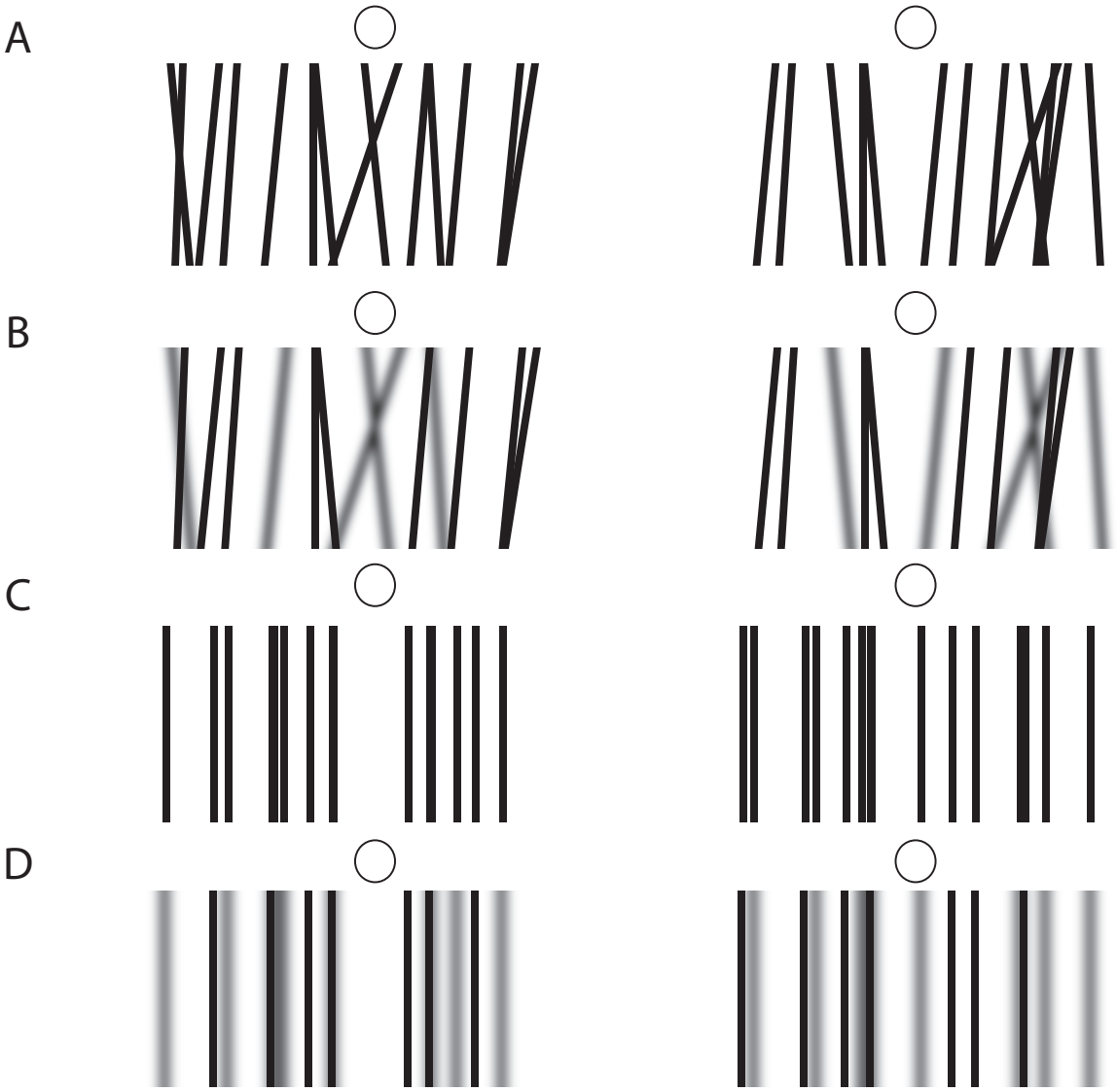


Figure 3.6: Example stimuli. Cross fuse to see in depth. A) Line orientation range is 24° and all lines are sharp. B) Orientation range is 24°; lines on the front plane are blurred and those on the back plane are sharp. C) Orientation range is 0° and all lines are sharp. D) Orientation range is 0°; lines on the front plane are blurred and those on the back plane are sharp. Most people find it easier to perceive the front and back planes when the front plane is blurred.

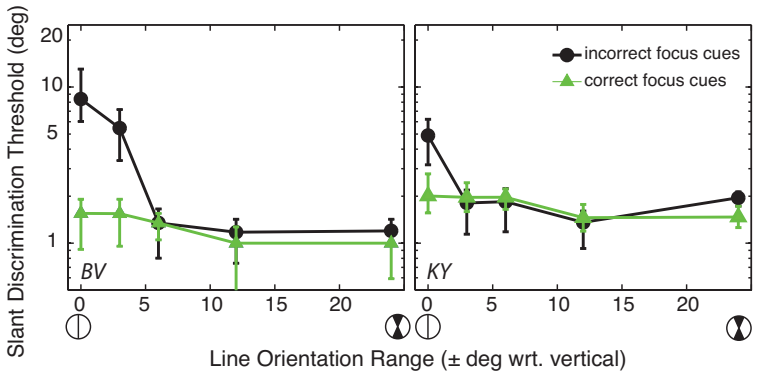


Figure 3.7: The abscissa represents the half range of line orientations in the stimulus (uniform distribution). The icons beneath the abscissa depict the orientation range. The ordinate is the slant discrimination threshold. The black circles represent the data when focus cues were incorrect (front and back planes presented on one image plane) and the green triangles represent the data when focus cues were correct (front and back planes presented on different image planes). Error bars represent 95% confidence intervals.

blur gradient in Experiment 1B by decreasing its magnitude to even smaller values.

3.3.1 Methods

Subjects

Three observers participated: BV (age 34), DMH (25), and MST (22). BV and MST were naïve to the experimental hypothesis. DMH is the first author. MST is a 2D myope and was corrected with contact lenses. Five observers were excluded because they could not fuse the stimulus planes.

Apparatus

The apparatus was the same as in Experiment 1A.

Stimulus

The stimulus was the same as in Experiment 1A with the following exceptions. The fronto-parallel near stimulus plane was presented on the mid image plane at 39.4 cm and the slanted back stimulus plane was presented on the far image plane at 53.6 cm. At the distance of the far image plane, the back stimulus plane was at the edge of the workspace. Thus the farther half of the stimulus plane was presented on the far image plane only. The increase in the distance to the far plane coupled with placing it at the edge of the workspace yielded a three-fold decrease in the blur gradient associated with a given slant. If the blur

gradient had provided useful information in Experiment 1A, this information would not be available in Experiment 1B, and thresholds would be the same in both focal conditions. The stimulus planes were slightly narrower with a width of 11°, and were textured with slightly thinner 7-arcmin lines. The ranges of line orientations were 0, ±2, ±4, ±8 and ±16°.

Stimulus duration was 1sec for observers BV and DMH and 1.5sec for observer MST. The fixation target was presented at the back stimulus plane.

Procedure

The procedure was identical to the one in Experiment 1A.

3.3.2 Results

The results are shown in Figure 3.8. Thresholds were again lower in the focus-correct condition than in the focus-incorrect condition when the orientation range was less than 6°. The interaction effect for focal condition and orientation range was significant (ANOVA ($F=5.269$, d.f. $=4$) $p<0.025$, one tailed). Thresholds were also not systematically higher in Experiment 1B than in 1A, which supports the conclusion that the blur gradient was not a useful cue to slant.

The results of Experiments 1A and 1B therefore provide strong evidence that the visual system can utilize focus information in solving the binocular correspondence problem.

3.4 Experiment 2

3.4.1 Background

When an opaque surface occludes a more distant surface, one eye can usually see more of the background surface than the other eye can. This is a monocular occlusion, and the region that is visible to only one eye is the monocular zone. Such occlusions are commonplace in natural viewing. In Figure 3.9A, the observer is looking at a background surface behind an occluder. The eyes are converged on the background, so part of it (the monocular zone) is imaged on the fovea of the left eye, while part of the occluder is imaged on the fovea of the right eye. The eyes are both accommodated to the background, so the retinal image is sharp on the left fovea, and the occluders image is blurred on the right fovea. In Figure 3.9B, the same situation is illustrated on a conventional stereo display. Again the eyes are converged at the (simulated) distance of the background. They are, however, accommodated to the distance of the display surface because the light comes from that surface. Now the retinal images on the foveae are both sharp even though the left eyes image is of the background and the right eyes image is of the occluder.

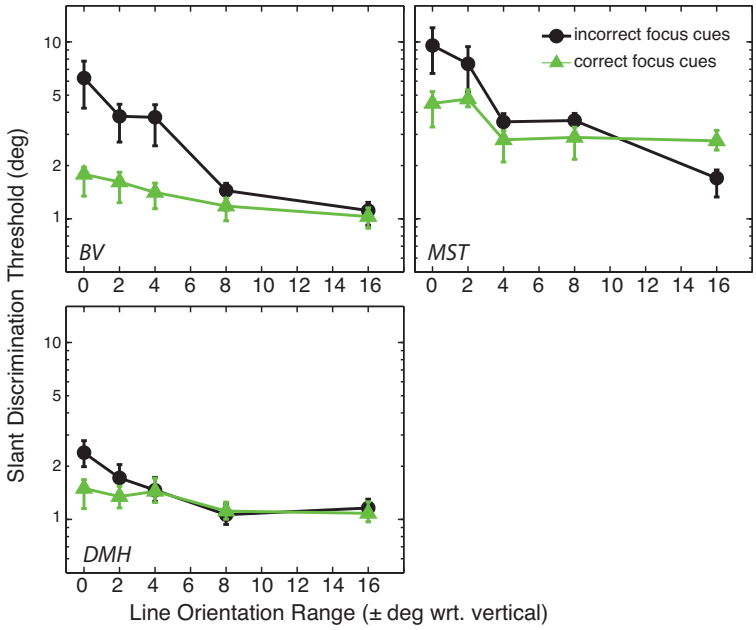


Figure 3.8: Results from Experiment 1B. The individual panels plot data from the three observers. The abscissas represent the half-range for line orientation (uniform distribution). The ordinates represent the slant discrimination thresholds. The black circles are the data from the incorrect-focus-cues condition and the green triangles are the data from the correct-focus-cues condition. Error bars are 95% confidence intervals.

When different images are presented to the foveae, the viewer typically experiences binocular rivalry: alternating dominance and suppression of the two images. A well-focused image is more likely to dominate a defocused image than a well-focused image[38, 96]. We hypothesize, therefore, that rivalry is more prevalent with stereo displays than natural viewing.

Figure 3.10 demonstrates our hypothesis. In Figure 3.10A, the background is sharp and the occluder is blurred. Notice that it is relatively easy to read the text on the background. In Figure 3.10B, the background and occluder are both sharp. Notice that it is now more difficult to read the background text. Thus the background region near a foreground object might appear more stable under conditions with correct focus cues than with incorrect focus cues. Figure 3.10A demonstrates the effect of pictorial blur on rivalry, but the effect of defocus blur is similar, and can be demonstrated with the cutouts in Figure 3.10C. Experiment 2 tested if pictorial blur can help the background region to appear more stable.

Stimuli

The stimuli were presented in a stereoscope. They were stereograms depicting a background and occluder (Figure 3.11). The background had a random texture and the circular occluder had an oriented random texture. The left side of the occluder blocks more of the background for the right eye than for the left eye: a crescent of background texture is therefore visible to the left, but not the right eye. The texture was oriented in the crescent region of the background and the orientation was perpendicular to the orientation of the texture in the occluder. The background was always rendered sharply. The occluder had four levels of defocus generated by using the Photoshop lens-blur function. The defocus levels corresponded to ratios of pupil diameter divided by inter-ocular separation of 0 (pinhole; Figure 3.11A), 1/13 (3.11B), 1/7 (3.11C), and 1/3.5 (3.11D). The edge blur was consistent with the texture blur. The order of presentation of the different types of stimuli was randomized. The stimuli had a space-average luminance of 46 cd/m^2. At this luminance, the average steady-state pupil diameter is 3.5 mm/citedegroot1952. For real-world stimuli, this would correspond most closely with stereogram B.

Protocol

Observers were instructed to fixate the crescent region of the stereogram: the part of the background just to the left of the occluder. While viewing a stimulus for 3 min, they indicated the perceived orientation of the texture in the crescent region by holding down the appropriate button. When the left-eyes image was dominant, the perceived orientation was up and to the left; when the right-eyes image was dominant, perceived orientation was up and to the right.

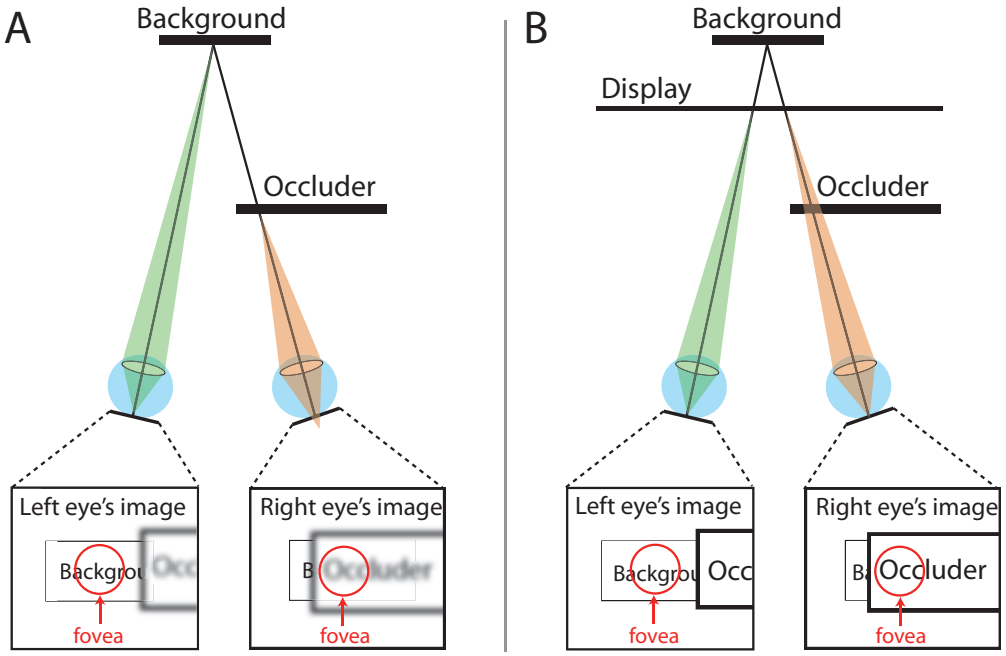


Figure 3.9: Monocular occlusions in natural viewing and with conventional stereo displays. A) Natural viewing. An opaque surface monocularly occludes a more distant background surface. The eyes are converged and focused on the background. The insets represent the retinal images for this viewing situation. The background is visible on the fovea of the left eye but not on the fovea of the right eye. The retinal image of the background is well focused and the image of the occluder is not. B) The same viewing situation presented in a conventional stereo display. The eyes are converged at the simulated distance of the background, but they are focused at the surface of the display because the light comes from there. The insets again represent the retinal images for this viewing situation. The retinal images of the background and occluder are well focused.

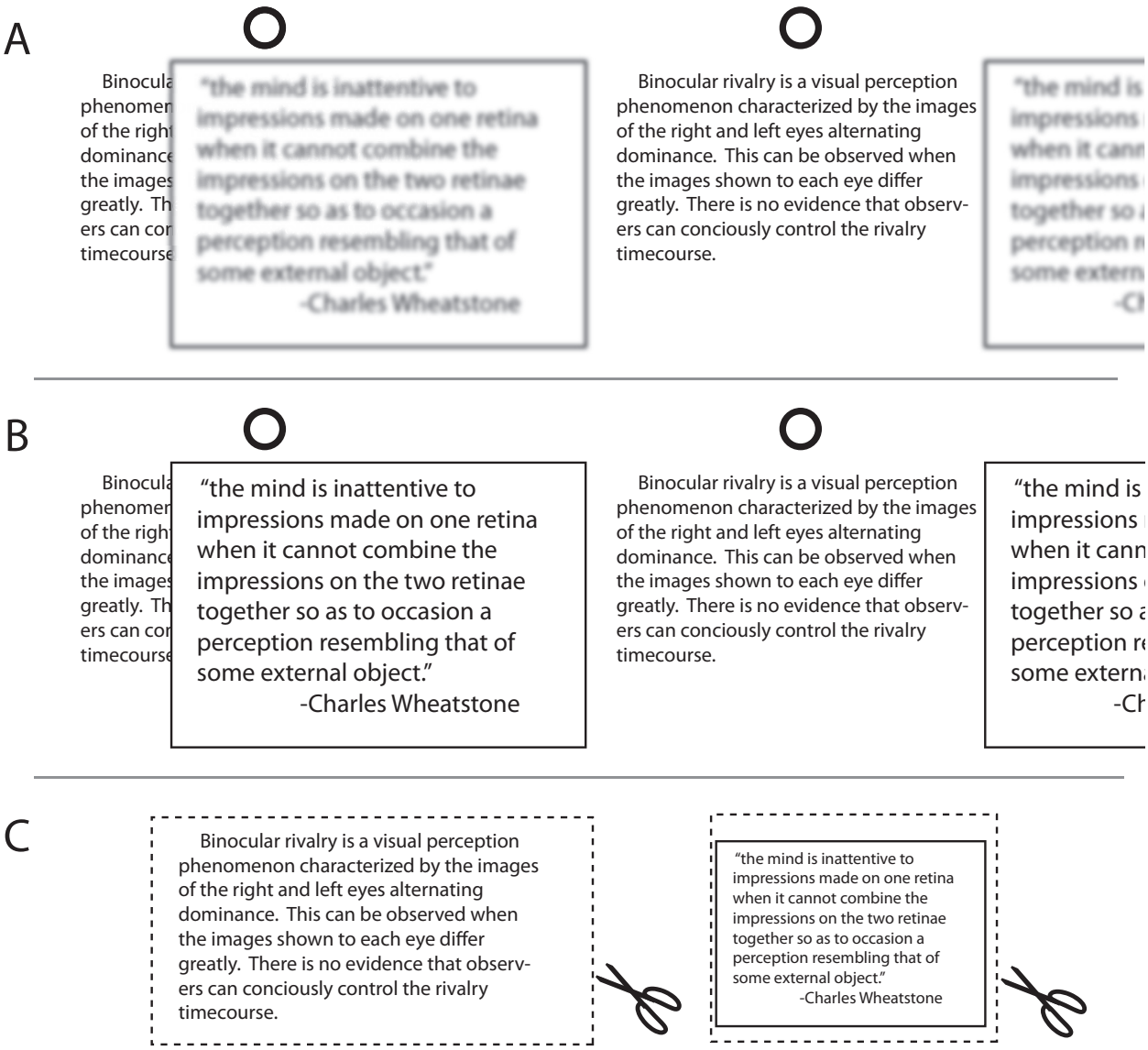


Figure 3.10: Demonstration that blur affects binocular rivalry with monocular occlusions. A) A stereogram depicting a well-focused textured background monocularly occluded by a blurred textured surface. Cross fuse to see the image in depth. Use the fixation circles to guide fixation. Direct your gaze and attention to the background near the edge of the occluder. Most people experience little rivalry in this case, and can easily read the background text. B) The same stereogram, but with the occluding surface equally focused. Again cross fuse and direct your gaze and attention to the background near the occluders edge. Most people experience rivalry near the foveas making it somewhat difficult to read the text on the background. C) Text for making cutouts. Cut on the dotted lines to make a background (left) and occluder (right). Hold them such that the occluder is near and on the right partially occluding the background.

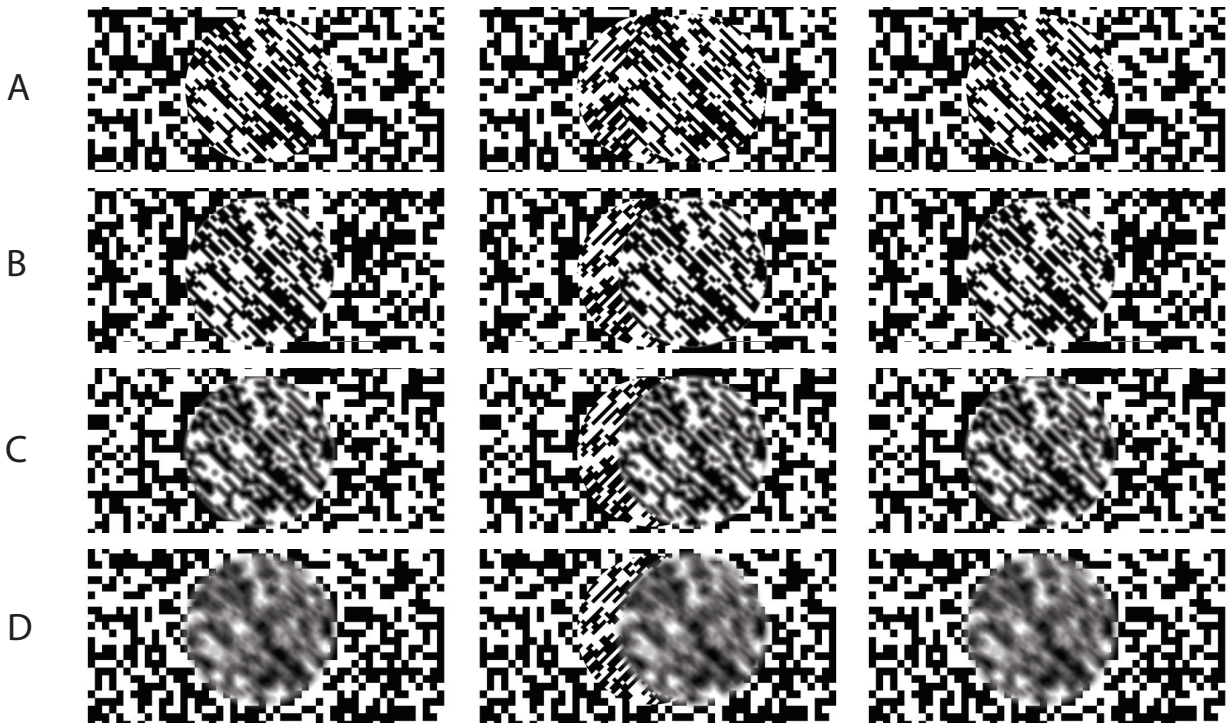


Figure 3.11: Stimuli used in Experiment 2. Cross-fuse the images in the first and second columns or divergently fuse the images in the second and third columns. A) Occluder has well-focused random texture. B) Occluder has blurred random texture and contours produced by simulating an aperture that is 1/13 the inter-ocular separation. C) Occluder with blurred random texture and contours produced with aperture/inter-ocular ratio of 1/7. D) Occluder with blurred random texture and contours produced with ratio of 1/3.5.

3.4.2 Results

We computed the percentage of time the background texture was dominant over the occluder texture. Those percentages are plotted in Figure 13 for each of the stimulus types. Each observer is represented by a different color. The black points and lines represent the averages. In the unblurred condition in which the background and occluder were well-focused (Figure 12A), the background was dominant roughly 50% of the time. Adding blur to the occluder significantly increased the dominance of the background (repeated-measures ANOVA ($F=11.2$, d.f.$=3$), $p<0.001$). Recall that the most realistic blur was presented in stereogram B. The difference in dominance between stereograms A and B was significant (paired t test: $p<0.05$). The results clearly show that the addition of blur affects binocular rivalry in the presence of monocular occlusion. This is consistent with our hypothesis that rivalry near occlusion boundaries is more noticeable in stereo displays than in natural viewing.

3.5 Discussion

Focus information can have a striking influence on binocular percepts[19, 43, 137, 59]. We showed that blur could in principle provide a useful cue for guiding appropriate matches when solving the correspondence problem. The results of Experiment 1 showed that human observers do in fact use such information in solving binocular correspondence. We also pointed out that appropriate focus information might minimize the occurrence of binocular rivalry near occluding contours. The results of Experiment 2 demonstrated that this occurs: rivalry is less prominent when focus information is appropriate as opposed to inappropriate. In the remainder of the discussion, we consider other perceptual phenomena that are affected by focus information.

3.5.1 Blur and stereoacuity

Stereoacuity is markedly lower when the half images are blurred. Westheimer and McKee[138] found that with increasing defocus, stereo thresholds degraded faster than visual acuity. For 3.0 D of defocus, stereothresholds rose by a factor of 10, whereas visual acuity thresholds rose by a factor of 2. Thus, to achieve optimal stereoacuity, both eyes must be well focused on the stimulus. Surprisingly, if the eyes are both initially defocused relative to the stimulus, but one eyes image is then improved, stereoacuity worsens. In other words, having one blurred and one sharp image is more detrimental to stereoacuity than having two blurred images. This phenomenon is known as the stereo blur paradox[138, 50]. Stereo thresholds are also adversely affected when the contrasts in the two eyes differ. This effect is called the stereo contrast paradox[55, 81, 110, 118, 30]. The blur and contrast paradoxes both manifest the importance of image similarity in binocular matching. Our analysis of retinal-image blur in natural viewing suggests that for a single object, perceptually discrim-

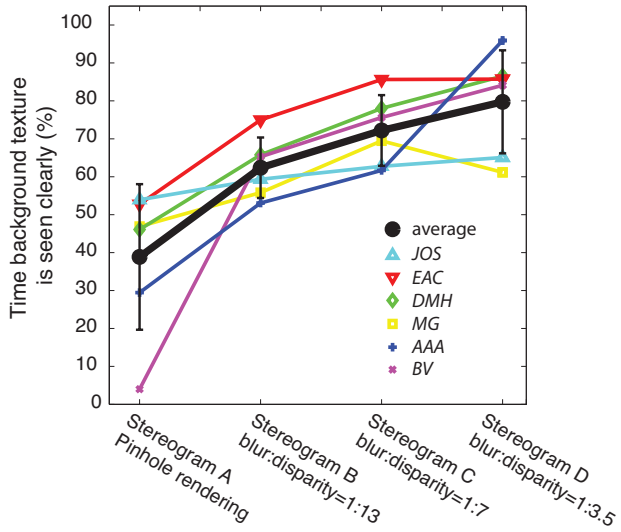


Figure 3.12: Results for Experiment 2. The percentage of time the background texture was perceived as dominant is plotted on the ordinate. The different stimulus types are plotted on the abscissa. The colored symbols represent the data from each observer. The black symbols represent the across-observer averages. Error bars are standard deviations.

inable image blur between the two eyes rarely occurs (Figure 3.2). Thus, the blur paradox is unlikely to occur in natural viewing.

3.5.2 Estimating depth from blur with monocular occlusions

With monocular occlusions, like the situation in Figure 3.9, disparity cannot be measured because one eye cannot see the object. As a consequence, the objects distance cannot be estimated from disparity[12, 3]. Here we consider the possibility that one can estimate the distance from blur when monocular occlusion occurs.

We assume that the observer is converged and focused on the edge of the occluder at distance d_0. A small object at distance d_1 is visible to the left eye only. The object is myopically focused creating a blurred image at the retina. (Figure 3.13) The relationship between the dioptric difference in distance and blur magnitude is the same as the hyperopic situation in Figure 3.1 (except that the sign of defocus is opposite). From Eqn 3.2, the object at d_1 creates a blur circle of diameter β_1 in radians. We can then solve for d_1:

$$d_1 = d_0 \frac{A}{A - \beta_1 d_0}$$

The observer can estimate d_0 from vergence and vertical disparity[48, 6]. The pupil diameter A is presumably unknown to the observer, but steady-state diameter does not vary significantly. For luminances of 5-2500 cd/m^2, a range that encompasses typical indoor and

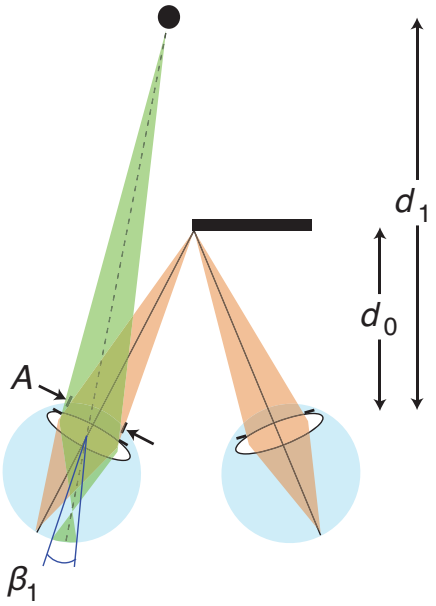


Figure 3.13: Depth from blur in monocular occlusion. The viewer fixates the edge of an occluder at distance d0, so the eyes are converged and accommodated to that distance. Pupil diameter is A. A small object at distance d_1 is visible to the left eye only. That object creates a blur circle of diameter β_1.

outdoor scenes, steady-state diameter varies by only ± 1.5 mm within an individual[33]. Thus it is reasonable to assume that the mean of A is 3 mm with a standard deviation of 0.6 mm. The blur-circle diameter b_1 also has to be estimated and it too will be subject to error[90]. We can then treat the estimates of d_0, A, and β_1 as probability distributions and find the distribution of d_1. In principle then, one can estimate the distance of a partially occluded object from its blur. Can this be done in practice?

Figure 3.14 demonstrates that blur can indeed affect perceived distance in the monocular occlusion situation. The three stereograms in panel A depict a textured surface that is partially occluding a more distant small object. The blur of the small object increases from the top to bottom row. Most viewers perceive the object as more distant in the middle row than in the top row and more distant in the bottom than in the middle row. Thus blur can aid the estimation of distance in a situation where disparity cannot provide an estimate.

Panel B depicts the same 3D layout but the small object is farther to the left such that it is visible to both eyes. In this case, the effect of blur on perceived distance is reduced, presumably because disparity now specifies object distance.

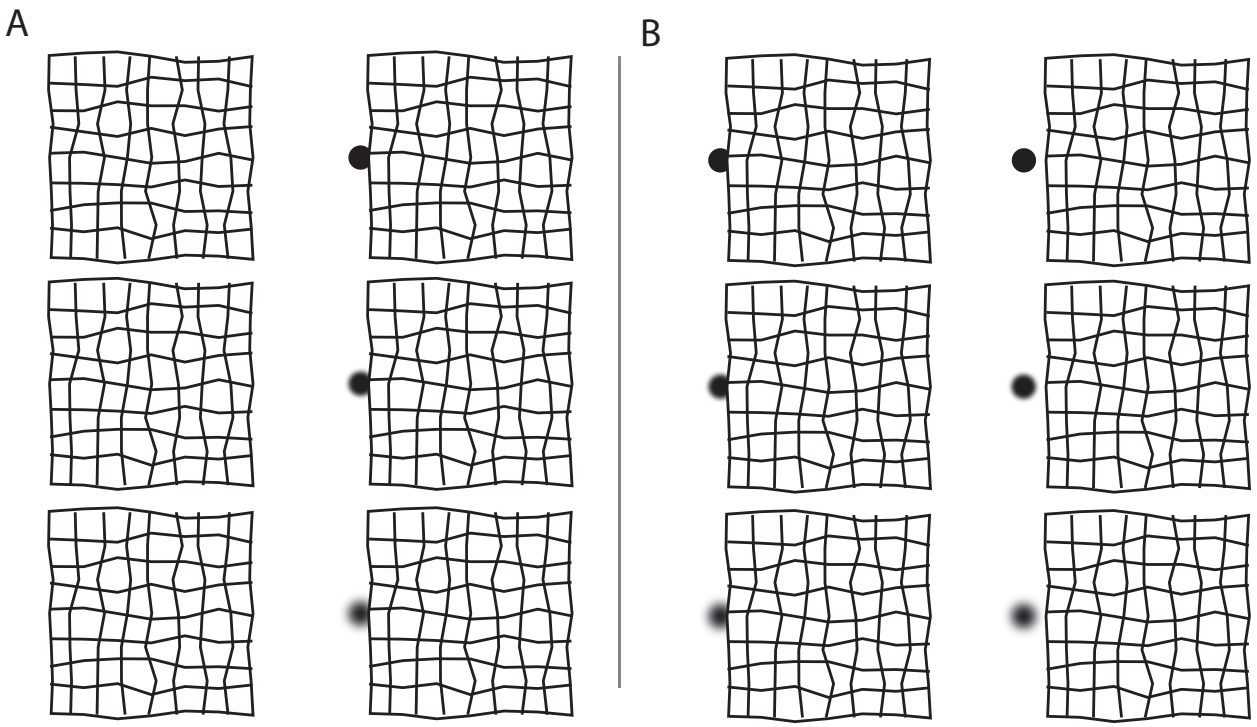


Figure 3.14: Demonstration of an effect of blur in monocular occlusion. The blur of the small disk increases from the top to bottom row. Cross fuse. A) Fixate the left edge of the textured square and estimate the distance to the disk. Most people report that the disk appears more distant in the middle than top row and more distant in the bottom than middle row. B) The object is visible to both eyes, so disparity can be computed. Blur now has a reduced effect on perceived distance of the disk.

3.5.3 The natural relationship between blur and depth

In natural viewing, the blur-circle diameter expressed in angular units is determined only by the pupil diameter A, the distance to which the eye is focused, d_0, and the distance of the object generating the blur circle, d_1. The pupil diameter is relatively constant in typical viewing situations, so there must be a commonly experienced relationship between blur, focal distance, and the 3D layout of the scene. We can recreate this relationship in a picture provided that the viewer focuses on the distance of the picture surface. Do changes in this relationship affect perceived 3D layout? Figure 3.15 shows that they do. The figure contains three stereo pictures of the same 3D scene. They were rendered with different ratios of aperture divided by inter-ocular distance (A/I). Specifically, the upper, middle, and bottom rows were rendered with ratios A/I of $\sim$0, 1/12, and 1/4, respectively. For an inter-ocular distance of 6 cm, these correspond to aperture diameters of: 0, 5, and 15mm. As we said earlier, the steady-state pupil diameter for a typical viewing situation is 3mm ($\pm$0.6mm). To most viewers, the second row looks more realistic than the other two provided that one maintains focus on the red dot. The upper row appears somewhat flattened and artificial. The lower row appears miniaturized because it is an example of the tilt-shift miniaturization effect[56]. This demonstration shows that obeying the natural relationship between blur and depth yields the most realistic percept. This implies that the human visual system has incorporated the depth information in blur and uses it to interpret 3D imagery.

The fact that the blur pattern affects depth perception does not necessarily mean that one should always render images with a camera aperture the size of the viewers pupil. Consider, for example, using medical images for diagnosis. With stereo images of chest X–rays, the doctor will undoubtedly prefer having the entire volume in sharp focus so that he/she can detect anomalies whenever they occur. Alternatively, in some medical applications, such as using an endoscope, the camera separation may be very different from a persons inter-ocular separation. Scaling the aperture diameter appropriately may be impractical in this situation due to diffraction effects, and light-gathering limitations. In this situation, the reduced depth of focus may help the doctor to view a surface that is partially occluded or obscured by a semitransparent membrane.

3.6 Conclusions

The work presented here demonstrates that retinal-image blur influences the interpretation of binocular images. Some of the mechanisms involved are low-level: the binocular-matching process and the processes underlying image suppression in binocular rivalry. Other mechanisms seem to be more complex, such as how disparity in combination with the appropriate magnitude of blur makes stereoscopic images appear more natural.

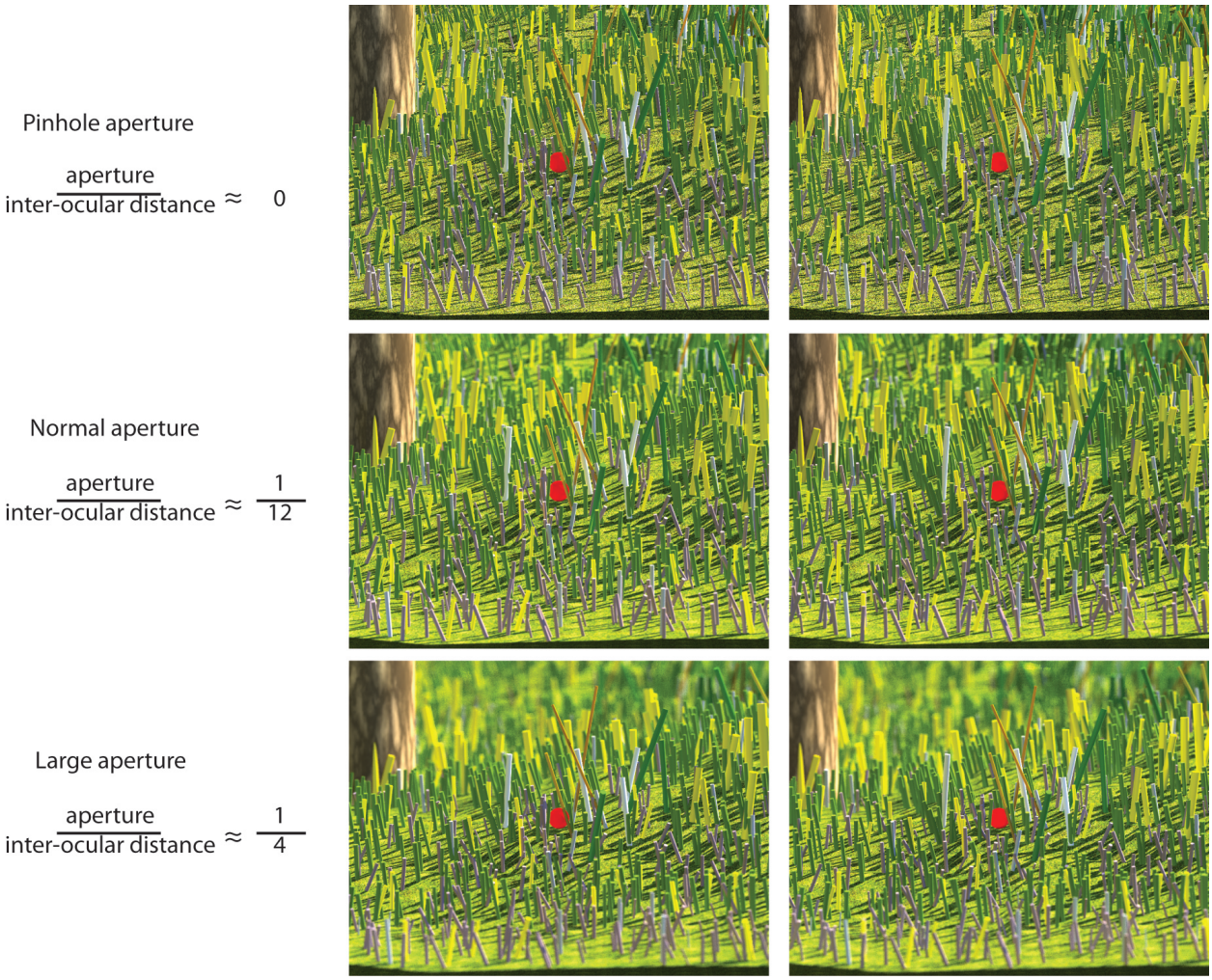


Figure 3.15: Demonstration of the appropriate relationship between blur and 3D layout. Cross fuse. The appropriate viewing distance is 10 picture-widths. In each case, the virtual camera is focused on the red dot. Upper row: The cameras aperture was a pinhole. Middle row: The cameras aperture was 1/12 times the inter-camera separation. Bottom row: The aperture was 1/4 times the camera separation. These stereograms do not recreate the vergence eye position that would correspond to converging on the red dot in the corresponding real scene. One would have to present the images in a stereoscope to recreate the correct vergence. The demonstration works reasonably well nonetheless. (Computer model provided by Norbert Kern.)

Chapter 4

Stereo display with time-multiplexed focal adjustment

In this chapter I describe a project to develop a time-multiplexed volumetric display. This work constructing a high-speed-switchable-lens display was originally published in SPIE in 2009. My coauthors were Gordon Love, Andrew Kirby, Phillip Hands and Martin Banks[60, 85]. Gordon and his team at the University of Durham, UK, were responsible for the optical engineering, and fabrication of the lens assemblies. At Berkeley, I was responsible for integrating the switchable lens assemblies with the displays and developing the software to drive the displays and hardware.

4.1 Introduction

The "three-mirrors" volumetric display used throughout Chapter 2 and Chapter 3 was important in demonstrating the value of focus information for stereoscopic imagery. It also demonstrated how modern computer graphics techniques could use the Z-buffer information to modulate light intensity across multiple image planes. However, the technique used—dividing a LCD monitor into a number of viewports and viewing them through beam splitters—is inherently limited. One of the biggest limitations is that it depends on the spatial dimensions of the display. It achieves different focal distance by using viewports that are at different distances from the observer. The system may be able to achieve a larger workspace through clever optical path folding, or by using Badal optics but the display size ultimately remains a limiting factor. Thus spatially-multiplexed fixed-viewpoint volumetric displays will be inherrently large and bulky that could make them unsuitable for consumer displays. A more promising tactic to achieve correct focus cues in a stereoscopic display is to use temporal multiplexing. This uses a single display to generate imagery and a high speed optical device to modulate the focal length. The key development making this display possible was Gordon Love's concept for constructing a lens that can change focal power at a

rate of 1KHz or higher. By placing the variable-power lens in front of the eye, we can adjust the lens to make the focal distance consistent with the simulated distance of an object even though the display screen is a fixed distance from the viewer. We synchronize the lens to the graphic display such that each depth region in the simulated scene is presented when the lens is in the appropriate state, and in this fashion construct a system with correct or nearly correct focus cues.

4.2 A time multiplexed volumetric stereo display

4.2.1 Overview of the approach

We generate a volumetric image by producing a geometrically correct pinhole rendering for each eye and then use a switchable lens to adjust the focal distance associated with each part of the simulated scene appropriately so that the focus information is consistent or nearly consistent with the simulated depth. The human visual system has very high spatial resolution and stereo acuity, so it is critically important to present high-resolution images to each eye[23]. The resolution requirements for changes in focus cues are fortunately relatively low because the depth of field of human vision is roughly ± 0.3 diopters[22]. We built a display with 0.6 D spacing between four focal planes; the workspace therefore encompassed 1.8 D.

We divide the simulated scene into different distance ranges and illuminate those regions on the display only when the lens is at the appropriate focal power. By cycling through the four focal states, we sweep out the full scene every four frames. The fast, switchable lens is the key innovation for this project that makes this display possible.

4.2.2 Correct or nearly correct focus cues generated with switchable lens system

The switchable lens combines the principle of a fixed birefringence optic with the adaptive control properties of liquid crystals. A lens-shaped object made of transparent birefringent material has two focal lengths depending on the orientation of the polarization of the incoming light: one corresponds to the ordinary refractive index and the other to the extra-ordinary index[74]. Our birefringent material is calcite, which is transparent and can be machined into a desired shape. It has refractive indices of 1.658 and 1.486. The incident unpolarized light from a display first passes through a linear polarizer, and then through a ferroelectric liquid-crystal (FLC) polarization rotator. The FLC acts as a switchable half-wave plate and, when oriented 45° to the plane of polarization, enables the incident polarization state to be rapidly modulated (240 μs switching time) between 0 and 90°. Thus, we can select one of two refractive indices (1.658 and 1.486), and thereby one of two focal powers, by applying the

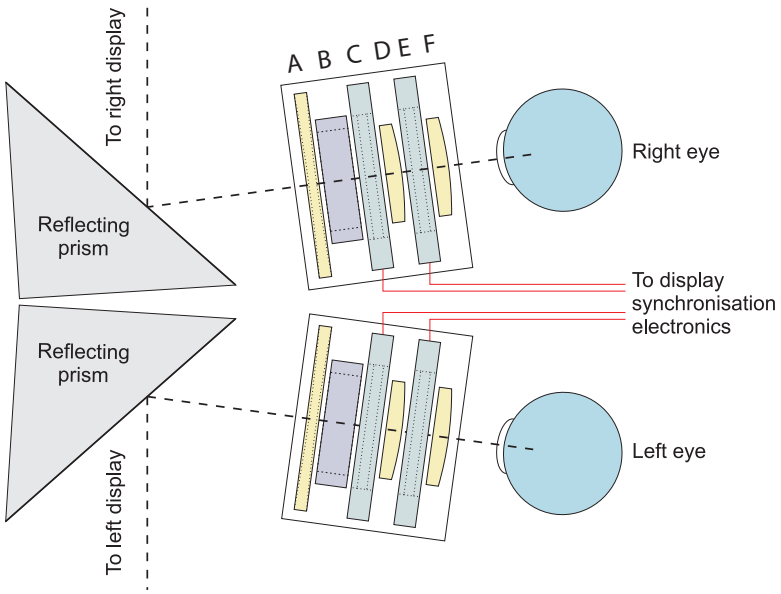


Figure 4.1: Schematic of the switchable lens assembly. Between each eye and its respective display, there is a lens system consisting of a static glass lens (A) to adjust overall focal power, a linear polarizer (B), ferroelectric liquid crystal 1(C) to modulate the polarization angle of the light approaching the first birefringent lens (D), and ferroelectric liquid crystal 2 (E) to modulate the polarization angle of the light approaching the second birefringent lens (F).

appropriate voltage to the FLC. We place lenses and FLCs in series to produce additional focal states; with N lenses, 2^N states are achievable.

Figure 4.1 illustrates a system of two lenses and FLCs for each eye. FLC1 is modulated at twice the frequency of FLC2, which enables the sequential addressing of all four combinations of polarization to the two lenses. The result is a four-state switchable lens. In our display, there is one four-state lens for each eye, thereby producing a stereo display with correct or nearly correct focus cues. The FLC states, the optical power and the resulting image plane distance are presented in Table 4.1.

Halbo Optics, Ltd. manufactured two pairs of plano-convex calcite lenses 20mm in diameter with an anti-reflection coating. The convex surfaces have radii of curvature of 143.3 and 286.7 mm, so the four focal powers are nominally 5.09, 5.69, 6.29, and 6.89D. Note that those values are separated by 0.60 D. An additional pair of static glass lenses can be inserted in the light path to move the workspace to a different range of distances. The optics were mounted in custom assemblies (Figure 4.2B) along with a pair of linear polarizers (Meadowlark Optics) and two pairs of FLC polarization rotators (LV2500, DisplayTech). Front-silvered prisms were placed in the light path to reflect light from two CRT displays

Focal state	FLC1 /FLC2	Lens power	Image distance	Magnification
Near	– / –	5.09 D	2.50 D, 0.40 m	6.2%
Mid-Near	– / +	5.69 D	1.96 D, 0.51 m	4.3%
Far-Mid	+ / –	6.29 D	1.28 D, 0.78 m	2.7%
Far	+ / +	6.89 D	0.75 D, 1.33 m	—

Table 4.1: Image plane specifications. The focal power of the lens system changed depending on the state of each of the FLCs. This table shows how the FLC settings adjusted the lens power, the corresponding change in the image plane distance and their magnification. The empirical image plane distances for the four FLC states with a -7D static ophthalmic lens are shown in the table as well as the empirical magnification of the image planes with respect to the far image plane. The magnification is expected because the changes in focal power occur in the spectacle plane.

through the lens systems and into the viewer's eyes (Figure 4.2A).

4.2.3 Synchronization with CRT displays

The FLCs are driven sequentially with appropriate electronic signals, and synchronized to the vertical refresh of the CRTs. The switching time of the FLCs is very short so the transition between focal powers can be made during the vertical retrace period (i.e. the period between a vertical synch pulse and the start of a visible frame trace). On each display refresh, the lens is switched to a new focal state; with four refreshes, the lens has gone through its four states. To register the lens system with the on-screen images, we presented a small rectangle in the corner of the display on the first frame of the sequence and detected it with a photodiode. This photodiode triggered a reset of the custom electronics driving the lenses. (Figure 4.2B) This also gave the system the ability to recover from a dropped frame, which would otherwise lead to a register shift between the lens state and the on-screen image.

As the lenses cycle through the four focal states, the CRTs present the parts of the simulated scene whose distances are consistent with the lenses' current focal state (Figure 4.3). When the distance of a simulated object corresponds to one of the available focal distances, pixels representing the object are illuminated only on frames in which the lens is set to the appropriate focal distance. When the distance of a simulated object falls in-between available focal distances, we use the depth-filtering technique described in Section 1.2.1.

4.2.4 Display and computer hardware

We performed the graphics computations in Matlab using Psychtoolbox OpenGL function calls[17]. The software runs on a Quad Core Mac Pro (2.8 GHz) with an NVIDIA 8800GT dual-head graphics card. The displays are Iiyama HM204DT 22-inch CRTs. We

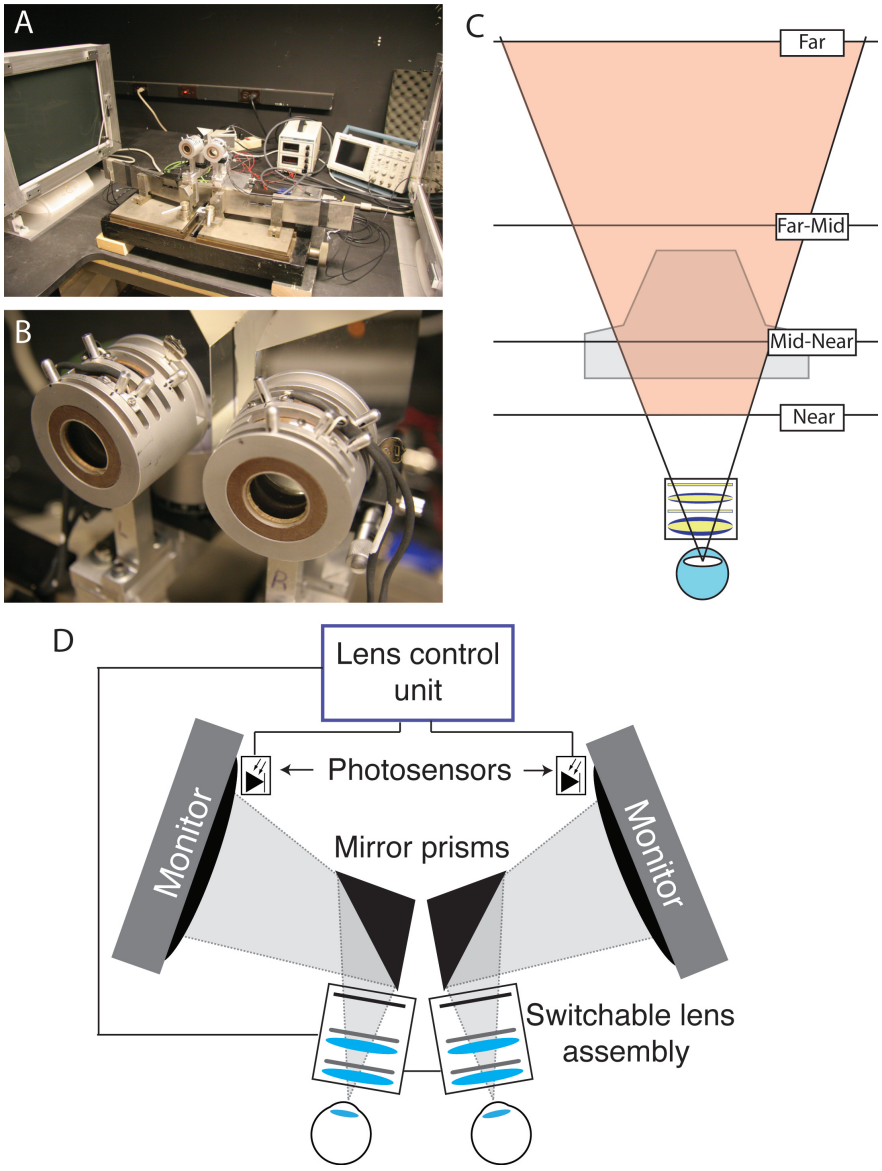


Figure 4.2: Switchable lens stereo display. A) Photograph of the apparatus showing the haploscopic arrangement of the lens system, mirror prisms, and CRTs. B) Close up of the lens systems. C) Top view of the light path from the CRT through the lens system to one eye. The shaded area is the workspace in which we can present approximately correct focus cues. D) A schematic layout of the system with synchronization electronics.

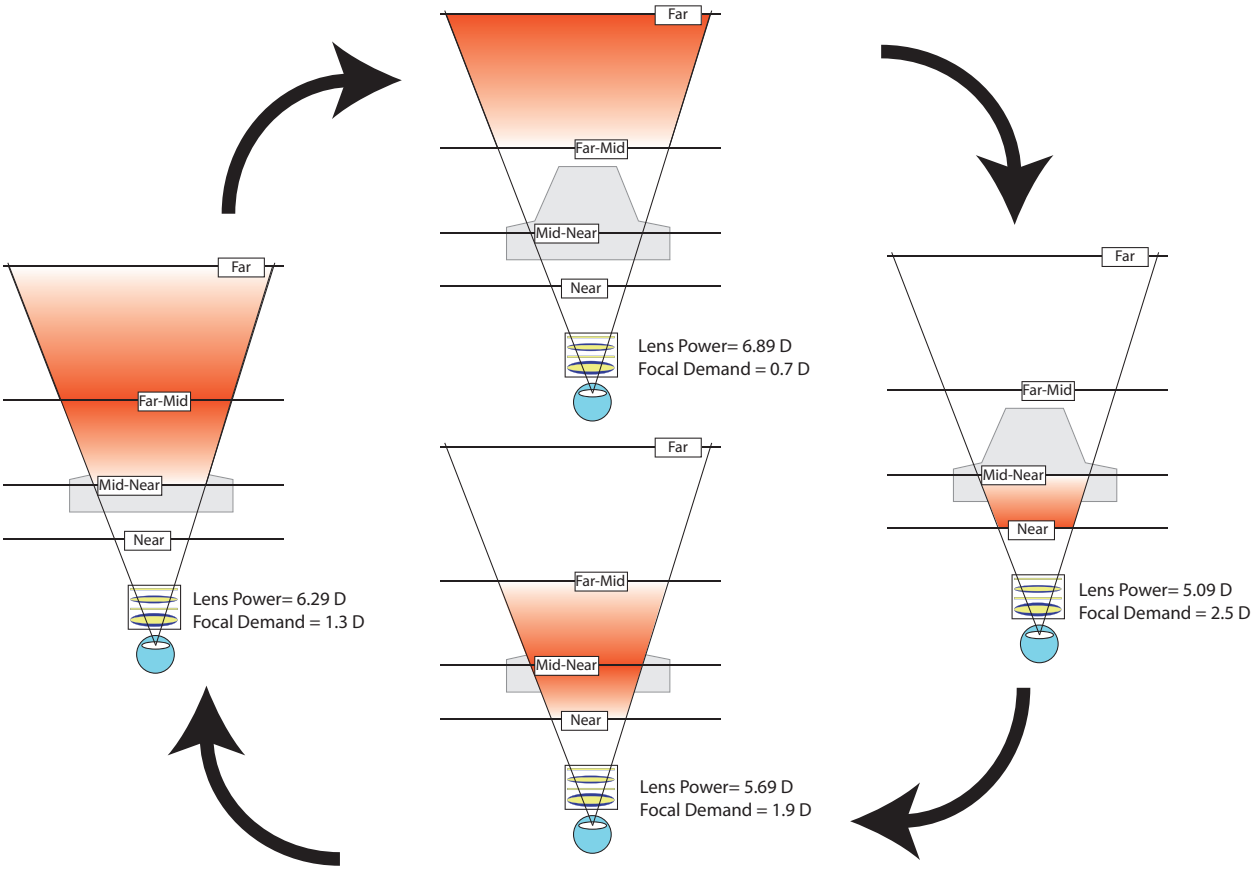


Figure 4.3: Lens state and pixel illumination over time. The workspace is the trapezoid bounded by the Near and Far planes and the diagonals on the left and right. For each simulated distance, pixels are illuminated according to their distance in diopters from the focal planes. When the simulated distance corresponds to a focal plane, pixels are illuminated solely at that plane (corresponding to a single image of the overall frame of the display); this situation is represented by dark red. When the simulated distance falls in-between available planes, pixels are illuminated on both planes (i.e. partially illuminated in two images of the overall frame) with their intensity corresponding to the relative dioptric distance of the simulated object from the planes (Equations 1.1); represented by the gradient of red.

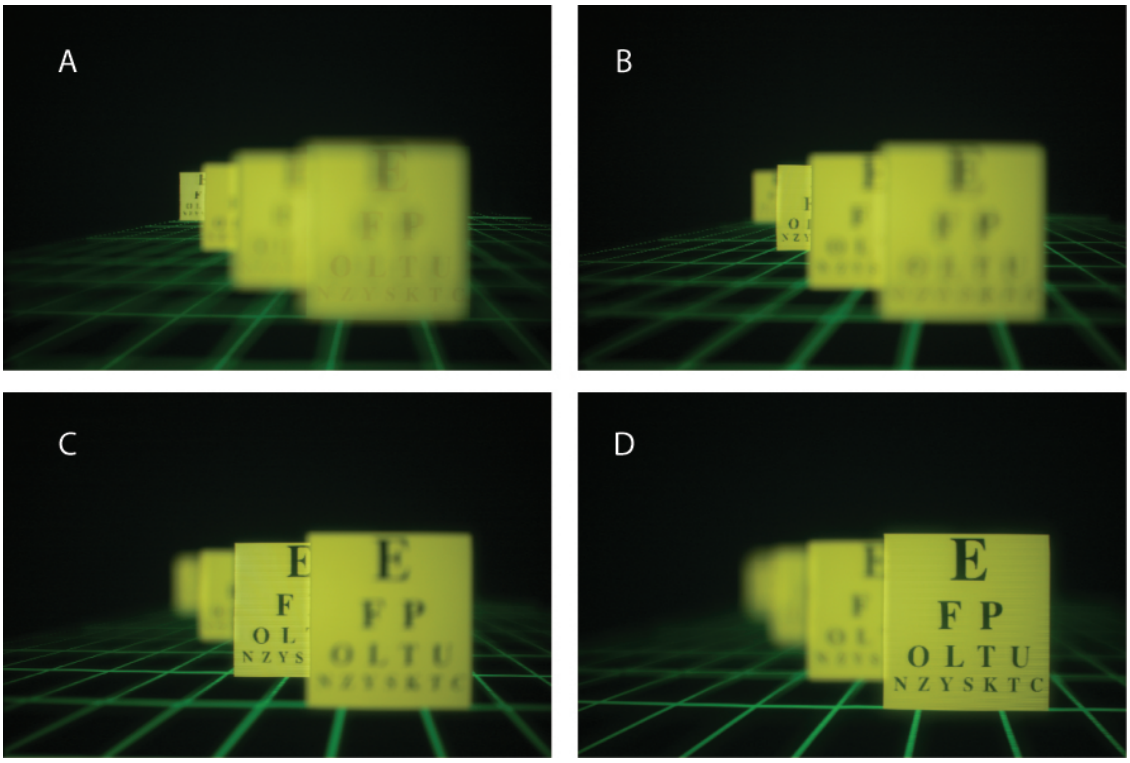


Figure 4.4: Snapshots from display demo video. The full video with audio narration is available at: http://www.opticsinfobase.org/oe/viewmedia.cfm?URI=oe-17-18-15716-1. Panels A-D show stillshots taken by a camera focused to each of the display's image planes. Notice that only a single object at a time is in sharp focus, and that the blur of the groundplane increases in a continuous fashion.

run the displays at 180 Hz, which allows us to cycle through the four focal states at 45 Hz. At that frame rate, the display resolution is 800x600, which yields a pixel width of 2.4 arcmin (Nyquist frequency of 12.5 cycles/deg). The refresh rate is essentially as high as possible with CRTs because at higher rates phosphor persistence introduces ghosting between focal states. LCOS, DLP, and OLED displays could provide much higher frame rates and improved resolution. See Appendix B for an overview of the appeal of several of these leading candidates for use as the display system.

4.3 Discussion

4.3.1 Perceptual image quality

The switchable-lens stereo display produces correct or nearly correct focus cues across its 1.75 D workspace. When viewing computer generated imagery through the display, the

focus information is quite noticeable. To demonstrate the effectiveness of this approach for producing nearly-correct focus cues, we produced a stop-action video of the display as we focused a SLR camera through the display. The photographs were taken with a slow shutter speed to ensure that each exposure captured several full volumetric refresh cycles. The scene contains a series of letter charts located at positions corresponding to the image planes of the display. In the first segment of the video, the entire image is presented on a single image plane. As the camera focuses through the volume, the entire scene comes into focus when the camera focuses to the image plane distance and becomes blurred when the camera is focused elsewhere. In the next segment of the video, the focus cues are presented correctly and each letter chart appears on its corresponding image plane. In this video, each individual letter chart comes into best focus when the camera is focused to it's simulated distance and goes out of focus as the camera focuses to the other charts. Snapshots from this segment of the video are shown in Figure 4.4, and the caption contains the link to the full video.

In the third segment of the video the scene is again drawn on the near image plane, and the camera is focused to this image plane. Now an object slowly translates from the front of the display volume to the back and then forward again. With the full scene drawn on a single image plane, the object remains in best focus throughout its motion. In the final video segment we present the same motion with the switchable lens active. Now the moving figure begins in sharp focus and rapidly blurs as its distance recedes from the camera. This segment demonstrates that with depth-filtering, there is no discontinuity when an object transitions between image planes.

4.3.2 Display artifacts

From the video, several display artifacts are visible. Other problems are less noticeable in the video and are difficult to fully characterize but are issues that could be limiting the image quality.

1. Chromatic aberration. We presented yellow images in this demonstration because the optics are not as effective for short-wavelength light. There are several reasons for this artifact. One minor reason is that linear polarizers are not completely effective at short wavelengths. A more significant cause is that the FLC elements are optimized at a specific wavelength. With polychromatic light, they will not be as effective at very long and very short wavelengths. Thus the polarization of long and short wavelength light will be misaligned with the crystalline axes of the lens. In addition to these birefringent color artifacts, there will also be conventional chromatic aberration of the lenses themselves, where the lens has a slightly different focal power for different wavelengths.

2. Off axis aberrations. The lens system has the best optical quality for axial object. For off-axis objects, the optical quality can suffer significantly. In some cases, eccentric

objects can appear blurred or doubled when viewed through the lens system.

3. Phosphor decay. The phosphors of the monitor have persistence. This means that an image may not be fully extinguished from one frame to the next. This effect can lead to ghosting. Because each focal state has a different calibration, the ghost image can have not only an incorrect focal state, but also an incorrect size and position.

4. Edge occlusions. The images presented in the volumetric display are pinhole renderings. This means that they do not sample the image from an area equivalent to a pupil. There will be information at an occluding edge which is not rendered for a conventional point projection. There is also an optical problem, that the imagery in the display can not optically self occlude for the different parts of the human pupil. This can create hot spots on edges where light from a far surface shines through part of a near object. This is a problem which is not unique to multi-plane displays, but is a problem for all volumetric displays. (The only exception are the integral displays designed with supersampling, so the pupil captures multiple discrete views of the scene) An interesting experiment would be to assess the perceptual importance of the effects occurring near hard occlusions, and the perceptibility of these artifacts.

5. Flicker. The CRTs can operate effectively at no faster than 180 Hz. With four image planes, this creates a volumetric refresh rate of 45 Hz. The images are fairly dim, and often foveal, so flicker is not particularly noticeable, but nevertheless, a 60 Hz volumetric refresh rate would be desirable for flicker-free operation.

Despite these problems, the images presented in this display are perceptually excellent quality. This display has demonstrated that switchable lens technology is applicable to solving this problem of incorrect focus cues in stereo displays, and has the potential to resolve many of the perceptual problems caused by incorrect blur and vergence-accommodation conflicts[137, 59, 1]. The design is extensible so more veridical image formation will be possible as higher-resolution and faster display technologies become available.

4.3.3 Modulation transfer analysis

To assess the display objectively we measured the modulation transfer of the lens system in each of its focal states. These measurements are shown in Figure 4.5. This plot shows the modulation transfer function (MTF) for on-axis viewing of black and white square wave test images. (Details of how we made the MTF measurements are included in the figure caption.) The MTFs show excellent transfer extending beyond 30cpd. For example, at 26cpd the modulation transfer is between 0.6 and 0.8 for the different focal states. We have managed to achieve this level of optical quality without optimizing the lens system to minimize chromatic or spherical aberration. The MTF plot also shows that the high power

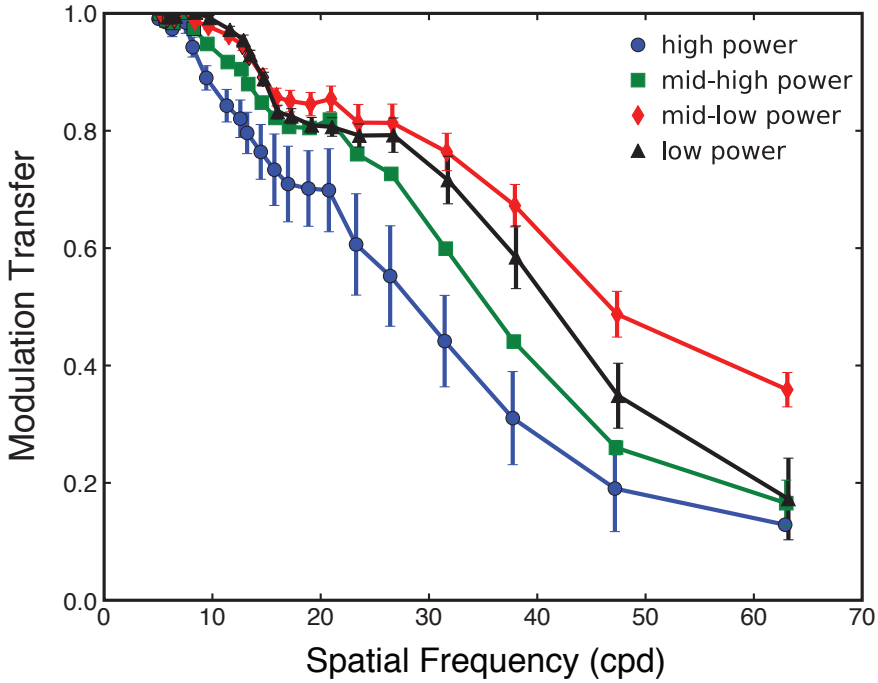


Figure 4.5: Modulation transfer of the switchable lens system. Modulation transfer is plotted as a function of spatial frequency for the four focal states of the system. We used a Canon 20D dSLR camera with a Canon 50mm prime lens (set to f/4.0 apperture) to image a range of sine-wave grating test-images. These images had spatial frequencies ranging from 5 to 63 cpd. We used direct camera measurements (without the switchable lens) to measure the attenuation from the camera optics, sensor and printing process. We then captured images of the test patterns through the lens system. The plotted MTFs are the measured MTF from the camera focused at each lens state divided by the MTF of images without the lens system. There are some differences in the modulation transfer for the different focal states, but the overall image quality is quite good.

state is inferior to the lower power states. In future designs of the system we hope to utilize optical engineering tools to design a system that achieves balanced optical quality across all the states, and minimizes the system's optical aberrations.

4.3.4 Future refinements

One of the projects we would like to pursue to improve the display is utilizing a higher frame-rate display technology. If based on a projector, this could allow us to present brighter images and a higher flicker rate. If a display were available that could operate at 360 Hz, then a single display could present imagery to both eyes via shutter glasses. The main advantage of a single display solution would be facilitating the calibration processes and giving the

display a more compact form factor.

Another avenue for improvement for the lens system is optimizing the optics. The current design used predominately off-the-shelf optical hardware (with the exception of the calcite lenses). In future iterations, it would be advantageous to design the system to be more compact. A more compact optical system would increase the field of view of the system, and make the optics more light-weight. This could facilitate the switchable lens eventually becoming part of a wearable system. Moving the lenses closer to the eyes and bonding the lens elements together would reduce the magnification effects, alignment errors and could expand reduce some of the off-axis aberrations.

A third opportunity for advancing this type of display is revisiting the choice for image plane separation. Earlier in this section I described that we chose a 0.6D plane separation based roughly on the depth of focus of the human eye. With the theoretical analysis in Chapter 5 and the empirical results discussed in Chapter 6, we could revise the choice for image spacing so that it is small enough to produce images that are perceptually indistinguishable from real-world imagery while simultaneously maximizing the size of our volumetric workspace.

Chapter 5

Retinal-image formation in multi-plane displays

As we design new display systems it is important to consider the quality of the retinal images they produce. The ultimate goal is to produce displays that create retinal images that are indistinguishable from those of real objects. In a fixed-viewpoint volumetric display, the image quality for a position simulated at the dioptric midpoint of two image planes will be the most difficult case, and will dictate the maximum permissible image plane spacing. The midpoint is particularly challenging because hyperopic, and myopic light does not sum to create an in-focus image; the summed image will also be out of focus. Fortunately, the depth of field of the human visual system makes the eye relatively tolerant to moderate amount of defocus. In this chapter, I will discuss how various factors of the eye and of the display will contribute to the retinal image quality.

There are several metrics that can describe the quality of an image, but I have chosen to focus on one, the retinal contrast ratio. This specifies the contrast attenuation of a stimulus with a specific spatial frequency. A single spatial frequency stimulus is a two dimensional luminance sine wave pattern. When this type of pattern is defocused, or undergoes modulation, the only change in the stimulus is a change in its contrast. The magnitude of the contrast attenuation differs with spatial frequency, so I often report the percent transfer at multiple spatial frequencies.

The modulation transfer is related to blur, which is a general term describing a loss in detail. It is likely that blur would be perceived in cases where there is significant contrast attenuation at high spatial frequencies relative to the normal transfer. This chapter does not discuss any perceptual metrics of blur, but examines what information reaches the retina. Parts of this analysis have previously been published in the Journal of Vision[59], SPIE[60] and Optics Express[85].

5.1 The eye's optics and the quality of retinal images

Contrast is most affected by defocus when the eyes' optics are best. The optical aberrations of a typical eye prevent a well-focused object from achieving its optimal contrast, but these aberrations expand the depth of field. The larger depth of field makes the visual system less sensitive to errors introduced from depth filtering. In this next section I will describe how spatial frequency, pupil size, and aberrations influence the retinal quality for simulated and real imagery.

5.1.1 Characterizing optical aberrations

I have based much of the analysis of image formation in the volumetric display on my own eyes (primarily my left eye) as well as several of my lab-mates. To characterize the optics of my eye, Professor Austin Roorda measured my aberrations using a Shack-Hartmann wavefront sensor. The aberrometer uses a narrow laser aimed through the pupil to create a point source of light on the retina. It then collects the scattered light from retina that exits the pupil and analyzes it using a lenslet array. An optically perfect eye will fully collimate the light reflecting from the retina and the lenslet array will reveal that at every position of the pupil the exiting light is parallel; this would correspond to a flat *wavefront*– the surface orthogonal to the rays leaving the pupil. However, aberrations in the lens and cornea cause the wavefront leaving the corneal plane to be an irregular surface and this shape can be characterized with *Zernike polynomials*, a basis set to describe the shape of the wavefront. An advantage of the Zernike parameterization is that one of the Zernike terms (Z_2^0) represents the object vergence. By manipulating this defocus term, we can simulate the wavefront for objects of various distances and can compute the point-spread function (PSF; the 2d retinal image created by a point source) for various stimuli. By summing multiple defocused PSFs with appropriate weightings, we can model the PSFs of stimuli presented by a volumetric display. This analysis allows us to explore how plane spacing, object position and eye properties influence image quality.

This type of analysis is an approximation of image formation in a real eye. During accommodation, many of the aberrations change, however the only aberrations which change consistently across observers are defocus and spherical aberration[28, 105]. The other aberrations fluctuate in a different ways for each observer. Despite these random-like changes in many of the high order terms and the predictable change in spherical aberration, the overall optical quality undergoes very little change. I have compared the retinal image quality including and excluding the spherical aberration changes, and while there are subtle changes in the shape of the point spread function, there is very little difference in the resulting MTF, or the conclusions that I would draw from the analysis. Thus, I have simplified accommodation as a simple change in the defocus term. (A more robust model including high order aberration changes with defocus is under development in Hong Hua's lab at the University

of Arizona.)

In the following analyses, I examine a position at the dioptric midpoint of two image planes. The object's intensity is evenly split between these two image planes, and there is no light originating from the position that we are attempting to simulate.

5.1.2 Optical aberrations and the modulation transfer function

First, consider a diffraction-limited eye that is focused to a real object. This ideal eye has no aberrations, and only diffraction-effects limit the image. The quality of the retinal image can be expressed as the modulation transfer function (MTF), which specifies the fraction of the stimulus contrast imaged at the retina at each spatial frequency. This MTF function is shown as the dark line in the left panel of Figure 5.1. Notice that there is excellent modulation transfer at low spatial frequencies (a high contrast ratio) and it falls off gradually for increasing spatial frequencies. Now, consider the situation where the diffraction-limited eye is focused to an object simulated at the dioptric midpoint between two of the image planes in the volumetric display (separated by 0.6D). In this case the contrast ratio, shown in red, declines rapidly with increasing spatial frequency and by 20 c/deg it is nearly fully attenuated. Furthermore, accommodating to the simulated distance does not necessarily maximize image quality. If the eye accommodates to one of the image planes, green line, there is a rise in the MTF for spatial frequencies greater than about 10 c/deg. Depending on which spatial frequencies drive accommodation, the visual system might not accommodate to the simulated distance.

The huge difference between these MTFs for real and simulated objects indicates that this display would likely be inadequate for a diffraction-limited eye. Subjectively, such an image would be blurry.

I repeated the same analysis using the aberrations of my left eye (which has typical aberrations) and show these results in the right panel of Figure 5.1. The black line shows the MTF for a real object. The transfer drops quickly with increasing spatial frequency. The MTF looks qualitatively similar when the eye accommodates to an object simulated at the dioptric midpoint of two image planes (red line). The main difference is that modulation transfer begins to drop slightly sooner. Also accommodating to the simulated distance produces better or comparable modulation transfer compared to the situation in which the eye accommodates to one of the image planes (green line). From this analysis, it is not obvious if the degraded image quality from depth-filtering would be unacceptable, and to what distance the eye might accommodate.

The large difference between an aberrated and diffraction-limited eye (illustrated in Figure 5.1) reveals how dramatically optical aberrations limit transfer of high spatial frequencies. The plot also shows that under most circumstances, images in the display will be lower contrast than their real world analogs, but that there are some spatial frequencies that will have a higher contrast in the display. Depending on the importance of these spatial frequencies to

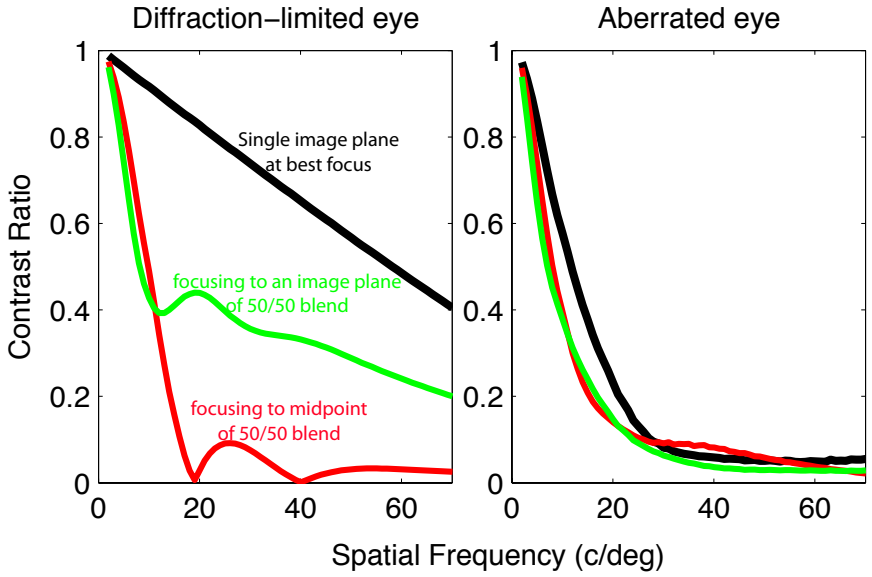


Figure 5.1: Modulation transfer functions (MTFs) for a monochromatic stimulus (550nm). The left panel shows several MTFs for a diffraction-limited eye with a 4.5 mm pupil size. The black line is the MTF for an eye that is focused to a real image. The red line is for the case when the eye is focused to an object that is simulated at the center of two image planes separated by 0.6D. The green line is the modulation transfer when the eye is focused to one of the image planes. The right panel was computed using the aberrations from my left eye with a 4.5 mm pupil size.

the visual system, this could have ramifications for blur perception or accommodative control. To better analyze the accommodative efficacy of these stimuli, I calculated the MTFs for real stimuli and stimuli in multi-plane displays for a range of accommodative responses.

Spatial frequency dependence

In Figure 5.2 I have selected 5 spatial frequencies, 2, 4, 6, 10, and 18 c/deg, and plotted their contrast ratio as a function of accommodative state. The dashed lines represent contrast ratio for a real object and the solid lines are for a depth-filtered object simulated at the midpoint between image planes (spaced 0.6D apart, and marked with the solid vertical lines). The difference between the real-world contrast ratio and the contrast ratio from the depth-filtered images is shaded. At low spatial frequencies, accommodation has only a minor influence on the contrast ratios, and there is only a small decrement in the contrast for the multi-plane images. At higher spatial frequencies, three main properties of multi-plane images become prominent:

1. With changes in accommodation, the contrast ratio does not change as much as it does for real images.

2. The peak contrast-ratio at a given spatial frequency is much lower for a multi-plane image than for the real image.

3. Beyond a certain spatial frequency (in this figure, between 10 and 18 c/deg) depth filtered images cease to provide accurate accommodative stimuli. Retinal contrast is maximized by accommodating to one of the image planes instead of to the simulated position.

The real stimuli in Figure 5.2 reveals some interesting features about signals for accommodation and blur detection. At low spatial frequency, accommodative errors will not produce significant changes in contrast, and at high spatial frequencies, there is little information reaching the retina irrespective of defocus. Thus, we might expect that the mid spatial frequencies, where a small change in defocus produces a large change in contrast, are most important for blur detection and accommodative control. Indeed, this hypothesis has been supported by a number of empirical studies. (4 to 8 c/deg [51, 133, 91, 101, 104, 128, 136])

Optical quality dependance

Figure 5.1, depicting the MTFs for a diffraction-limited and a real eye showcase the profound difference that eye quality can make on image contrast. Within the population, there is a wide range of eye quality[123, 27]. Especially interesting is the effectiveness of depth filtering for people with different quality optics. In Figure 5.3 I consider the modulation

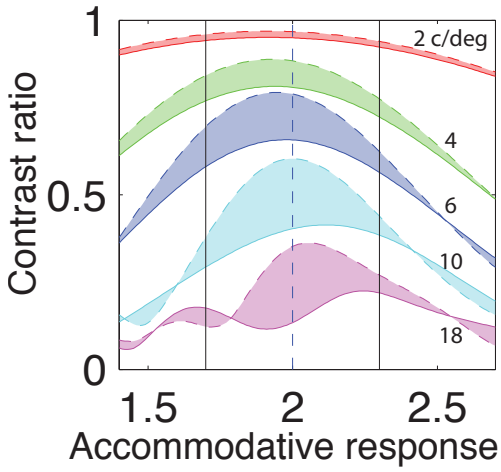


Figure 5.2: Contrast ratio at different spatial frequencies for different accommodative responses. The dashed lines are the contrast ratio for a real object, and the solid lines are for the simulated focal distance at the midpoint the image planes (separated by 0.6 D). 2, 4, 6, 10 and 18 c/deg are shown respectively in red, green, blue, cyan, magenta. The vertical solid lines represent the location of the image planes for the depth-filtered stimulus, and the dashed line represents the focal distance of the real object.

transfer of a 6c/deg grating for someone with a perfect eye, a normal eye, and a below-average eye, and plot the retinal contrast for these three different eyes as the red, green and blue curves respectively. The dashed lines represent the contrast ratio for a real image, and the solid lines are for a depth-filtered image, and their difference is shaded. One of the most prominent effects of the quality of the eye is the peak contrast ratio. The diffraction-limited eye produces a much higher contrast image than the badly aberrated eye. However, the diffraction-limited eye also exhibits the greatest difference between the real-world image and the volumetric display image. Both of these effects are visible in Figure 5.1, however Figure 5.3 shows the rate at which image quality falls off with errors in accommodation. An eye with lower optical quality can tolerate a far greater defocus with minimal attenuation compared to the diffraction-limited eye. This also reduces the contrast differences between real and depth-filtered stimuli. Furthermore, irrespective of the eye quality, the 0.6D image plane separation produces a well behaved depth-filtered image at 6 c/deg. There are no erroneous contrast peaks. Thus, if 6 c/deg drives accommodation, we would expect depth filtered images to be effective accommodative stimuli.

5.1.3 Pupil size

In addition to the eyes' aberrations, the size of the pupil also governs retinal image quality. The visual system's depth of focus is roughly inversely proportional to pupil diameter [26, 53]

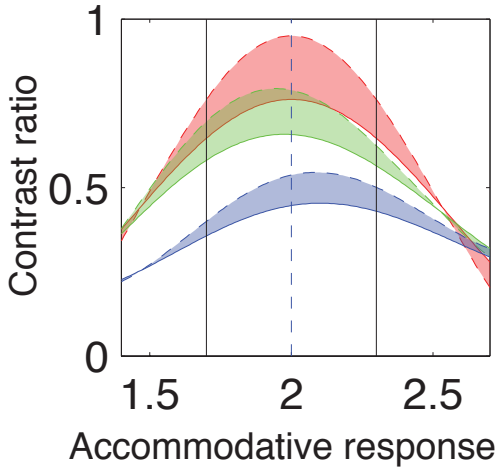


Figure 5.3: The influence of the aberration level on the quality of depth-filtered images at a spatial frequency of 6 c/deg. The solid lines show the contrast ratio as a function of accommodative state for depth-filtered stimuli, and the dashed lines show the contrast ratio for real objects. The red lines are for a diffraction limited eye, the green shows a normal eye, and blue a less-than-average eye. All the eyes had a pupil size of 4.5mm. The vertical solid lines represent the location of the image planes for the depth-filtered stimulus, and the dashed line represents the position of the real object.

and as the pupil constricts, the optics increasingly resemble a diffraction-limited system. To examine how pupil diameter would affect image quality, I use the aberrations from my left eye (the eye I have used for the normal aberration in Figure 5.3) to compute the contrast ratio for real and depth-filtered images for several different pupil apertures. I plotted the contrast ratio for a 6 c/deg stimulus in Figure 5.4 for a 3mm pupil (in red), 4.5mm (in green) and a 6mm pupil (in blue). The contrast ratios for real-world stimuli are marked with dashed lines, and the depth-filtered stimuli are shown with a solid lines. With a smaller pupil, the aberrations had a reduced attenuation on the contrast.

With a reduced pupil size, depth filtering becomes a better approximation of real-stimuli but the usefulness of focus information rapidly decreases. A smaller pupil expands the depth-of-field and could elevate depth-from-blur thresholds. However, pupil size remains fairly stable across a large ranges of luminance. de Groot and Gebhard[33] have reported that pupil size, d, in millimeters varies with the field luminance, B, (in cd/m^2) according to the function: $\log_{10} d = 0.8558 - 0.000401 \left(\log_{10}(\frac{10}{\pi} B) + 8.1\right)^3$. This means that for luminance ranging from 0.1 to 2000 cd/m^2, the pupil size changes from 4.8 to 1.5mm.

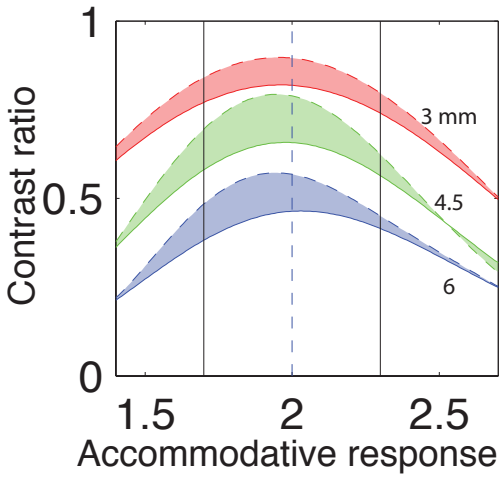


Figure 5.4: The influence of pupil size. The contrast ratio for a 6 c/deg grating is shown for the normal aberrated eye with a pupil size of 3 (red), 4.5 (green), and 6mm (blue). The solid lines show the contrast ratio as a function of accommodative state for depth-filtered stimuli, and the dashed lines show the contrast ratio for real objects. The vertical solid lines represent the location of the image planes for the depth-filtered stimulus, and the dashed line represents the focal distance of the real object.

5.1.4 Chromatic aberration

An optical attribute of the eye that we have not yet considered is *chromatic aberration*—an aberration causing short-wavelength light to have a shorter focal distance than long-wavelength light. These differences in focal length can affect image quality. Any aberration that degrades image quality will improve the liklihood that depth-filtered stimuli will be acceptable, and thus we might expect volumetric displays to be more effective at presenting polychromatic than monochromatic stimuli.

Chromatic aberration cannot be measured using the Shack-Hartmann wavefront sensor because the laser is monochromatic. Furthermore, Austin Roorda's aberration analysis software is designed to compute monochromatic point spread functions. Thus, in order to model these chromatic effects, I used the chromatic aberration data collected by Bedford and Wyszecki (1957)[11]. They asked an observer to adjust the focal distance to a stigma until it was perceptually best focused. This data is shown in the top of Figure 5.5. To compute a broadband point spread function, I added the chromatic defocus from the Bedford and Wyszecki data to the Zernike defocus term, and computed PSFs for wavelengths spanning the visible spectrum. I combined these PSFs by weighting them according to the spectral sensitivity function for the human observer (1924 CIE V(λ), lower panel of Figure 5.5) to compute a luminance PSF. Thus, the combined retinal image is the luminance PSF for an equal energy white stimulus.

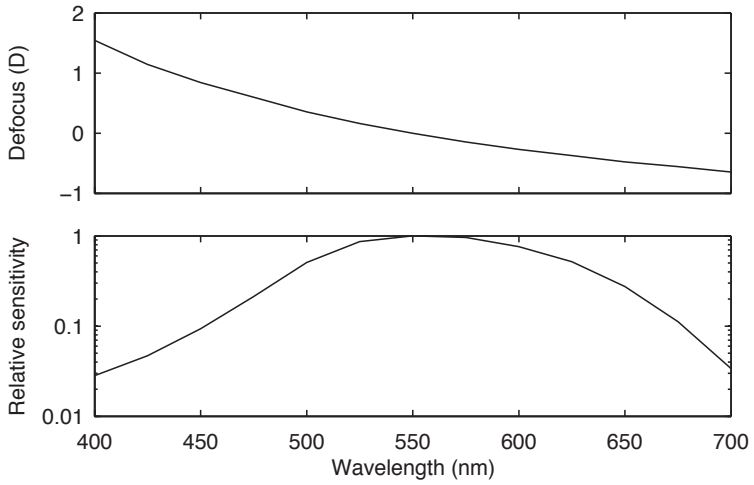


Figure 5.5: Chromatic aberration and weighting data. The top panel shows Bedford and Wyszecki's (1957)[11] data from a psychophysical experiment where observers adjusted the defocus of a series of monochromatic stimuli so that they appeared sharp. These defocus setting are plotted on the ordinate, and the wavelength is plotted on the abscissa. The lower panel shows the 1924 CIE Standard Observer V(λ) function. This is the normalized sensitivity of a standard observer to each wavelength.

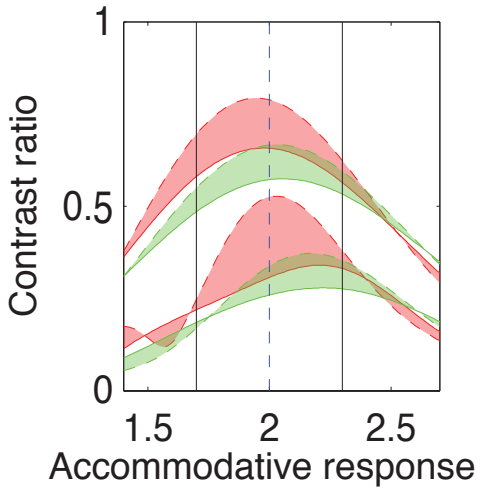


Figure 5.6: The influence of the chromatic aberration on image quality. The solid lines show the contrast ratio as a function of accommodative state for depth-filtered stimuli, and the dashed lines show the contrast ratio for real objects. The lower set of curves is for 12 c/deg, and the upper set of curves is for 6 c/deg. The contrast ratios with broadband stimuli are shown in green, and monochromatic stimuli in red. The vertical solid lines represent the location of the image planes for the depth-filtered stimulus, and the dashed line represents the focal distance of the real object. The pupil size is 4.5mm.

Introducing chromatic aberration, as with other oberrations, lowered the peak contrast ratio, but expanded the depth of field. In Figure 5.6 I have plotted the contrast ratios for a real (dotted line) and multi-plane image (solid line) both with and without chromatic aberration. The upper sets of curves show 6 c/deg and the lower curves show 12 c/deg. The monochromatic contrast ratios are shown in red, and the broadband contrast ratios are shown in green. Chromatic aberration reduces the difference between the real and depth-filtered images. This suggests that a multi-plane display showing broad-spectrum images is likely to be more forgiving than a monochromatic image.

This discussion of chromatic aberration has simplified chromatic aberration to its impact on the luminance of the PSF. This oversimplification is only accurate to a first approximation. The human visual system has trichromatic vision, and estimates luminance by weighting light at all wavelengths, but the three cone types preserve some of the color information. There is evidence that the visual system can use chromatic aberration to disambiguate the sign of defocus for accommodative control[79, 41]. There are also studies showing that chromatic aberration could play a role in emmetropization[141]. The chromatic-aberration signal for depth-filtered stimuli will be different than for real objects. Furthermore, the lens system itself contains chromatic aberration. This is a shortcoming of the display that may restrict its usefulness in some vision science experiments.

5.2 Quality of depth-filtered images throughout the display workspace

In addition to the special case of an object simulated at the center of two image planes, we would also like to know the quality of images throughout the workspace, and for a variety of hypothetical accommodative responses. And, as in previous analysis, we would like to compare the quality of the approximated focus cues to those of the real world.

Consider a real stimulus of 6 c/deg (top center of Figure 5.7) at a distance of 2D. As the observer accommodates from far to near (i.e. from 1 to 3 D, a vertical slice through the diagram), the retinal image increases to its peak near 2.0D, and reaches a contrast ratio of 0.78 before decreasing. (Note that the accommodation producing a maximum contrast ratio may differ slightly for each spatial frequency[52].) At the lower spatial frequency, the rise and fall is gentler and it peaks at a contrast ratio of 0.97. At 18 cycles per degree, the rise and fall is much steeper but the peak contrast only reaches 0.34. Note that these vertical cross-sections correspond with the dashed lines from Figure 5.2, and represent the normal relationship between object distance, accommodative response and contrast ratio.

Now consider our multi-plane display (Figure 5.7, middle row). The four image planes are positioned at distances of 1.1, 1.675, 2.38, and 2.88 D, so the workspace encompasses a 1.8D volume. When the simulated distance is at the distance of an image plane, the retinal contrast produced by viewing our display is identical to the contrast produced by viewing the

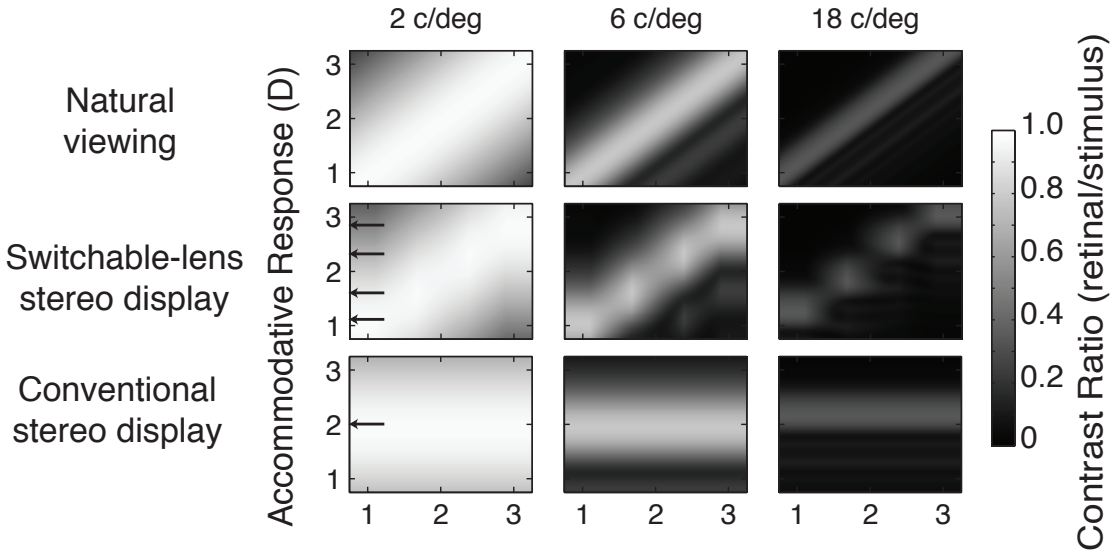


Figure 5.7: Image quality for arbitrary distance and hypothetical accommodative response. The top, middle, and bottom rows represent stimuli in the real world, in the switchable lens display, and in a conventional display. The left middle and right columns represent spatial frequencies of 2, 6, and 18 c/deg respectively. The abscissas are the actual (for natural viewing) or the simulated distance (for the displays) to the stimulus in diopters. The ordinates are the accommodative response in diopters. The grayscale represents contrast ratio: retinal-image contrast divided by stimulus contrast. The panels were based on my left eye with a 4.5 mm pupil size. In the volumetric display, the four image planes are at distances of 1.1, 1.675, 2.38, and 2.88 D.

real world. When the simulated distance is between planes, the retinal image is formed by a depth-weighted blend (Figure 1.3) of intensities from the two nearest planes. At 2cpd, the blended image within the displays workspace is a nearly perfect approximation to the image produced by a real target. Importantly, retinal-image contrast is maximized by focusing at the simulated distance rather than at one of the image planes. At 6cpd, the blended image is still a good approximation, and retinal contrast is again maximized by focusing at the simulated distance rather than at one of the image planes. At 18cpd, the blended image is a poorer approximation to the real world: the peak contrast occurs near the image planes rather than at the simulated distance. Furthermore, at the intermediate positions there is an appreciable drop in contrast. This figure demonstrates some of the limitations of the volumetric display and motivated us to conduct the following analysis to examine the maximum image plane spacing to achieve high quality imagery even at the higher spatial frequencies.

The lower row of Figure 5.7 shows the retinal contrast for stimuli in a conventional display. In this display, the screen distance is at 2.0D. Accommodating to approximately this distance maximizes retinal contrast. The vertical cross-section of this plot is identical to a vertical cross-section for real stimuli, but with a phase shift when the simulated distance does not equal the display distance. The conventional display has the advantage that there is an accommodative response for each simulated distance where the observer could see a high quality image. However, this accommodative response may not be possible (see zone of clear single binocular vision, Section 2.1.2), in which case the stimulus could appear badly blurred.

5.3 Image plane spacing and the effect on image quality

The previous analysis (Figure 5.7) shows that depth-filtered images have notably degraded image quality at higher spatial frequencies. What image plane spacing is needed to produce high quality images? To begin answering this question, we must first assume that the observer accommodates to the simulated distance. In Figure 5.8, we compute the retinal contrast for the eye accommodating to simulated distances across the gap between image planes; I looked specifically at how the contrast ratio changes for these intermediate positions as a function of the image-plane spacing. The dioptric separation between image planes is plotted on the abscissa of each panel. The simulated distance of the object is plotted on the ordinate as a proportion of the distance between focal planes. The left, middle, and right columns represent the results for sinusoidal gratings of 2, 6, and 18 cpd, respectively. In each panel, color represents contrast ratio, red representing a transfer of 1 and dark blue a transfer of 0. The contrast ratio is maximized when object position is at 0 or 1 because those distances correspond to cases in which the image is on one plane only and

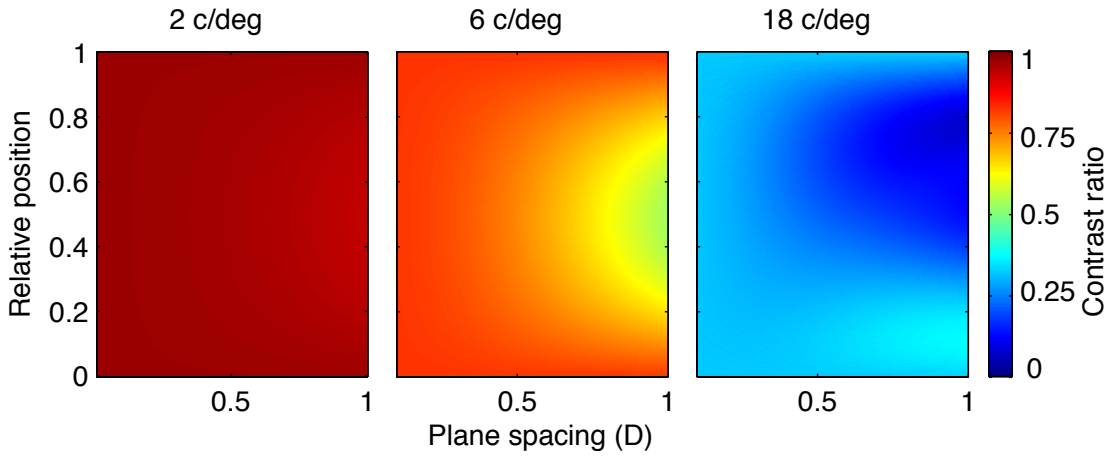


Figure 5.8: Retinal-image contrast with the multi-focal display for different spatial frequencies, and separations of the focal planes. Each panel plots plane separation in diopters on the abscissa and the position of the simulated object on the ordinate. Color (see color bar) represents modulation transfer (contrast in the retinal image divided by incident contrast). In constructing the figure, I have constrained the eye to accommodate to the simulated distance. The left, middle, and right columns show the results for spatial frequencies of 2, 6, and 18 cpd, respectively. The contrast ratios are computed for my left eye with a 4.5 mm pupil size and monochromatic stimuli.

our simulation is perfect. When the object position is at the midpoint (0.5), image intensity is distributed equally between the bracketing near and far planes. This analysis reveals that spatial frequency is a dominant factor in determining the minimum spacing between image planes. At higher spatial frequencies, a much smaller image-plane separation is required. Another interesting feature of the plot for 18 c/deg is that it is highly asymmetrical between the image planes. This is because the best accommodative response at one spatial frequency is not necessarily the best response at other spatial frequencies[52]. In this case, the visual system maximizes contrast at 18 c/deg by accommodating at a slightly different distance. Because spatial frequency has such an important role in determining the acceptable image-plane spacing, I have replotted the data differently in Figure 5.9 showing a continuous range of spatial frequencies.

In this follow up analysis, I calculate the maximum decrement in the contrast ratio that occurs for objects between two image planes (the max-min from a column of Figure 5.8) and plot the isodifference positions as contours as a function of spatial frequency (abscissa) and plane spacing (ordinate). Each contour line represents the display conditions where the contrast decrement is uniform. Interestingly, the difference in retinal image quality does not grow monotonically worse with increasing spatial frequency. At higher spatial frequencies, there is a lower nominal contrast ratio, and thus there is a smaller difference in contrast ratio irrespective of the simulated distance.

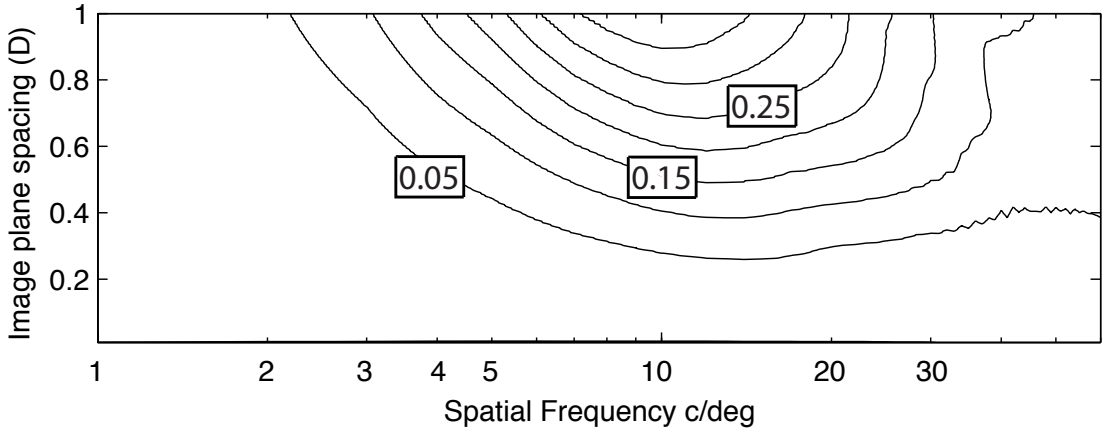


Figure 5.9: Effect of image-plane spacing and spatial frequency on quality of intermediate position simulations. Assuming the eye is focused to the simulated distance, these contours show the change in the contrast ratio that the observer will experience for objects at different simulated positions between the image planes. From bottom to top, the contours represent a change in contrast ratio of 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, and 0.35. The calculations were based on the aberrations of my left eye with a 4.5 mm pupil size.

5.4 Conclusions

I have shown that in many situations, accommodating to the simulated position at the midpoint of the image planes maximizes the retinal contrast. It is thus a good hypothesis that these images will stimulate appropriate accommodation. I will discuss an experiment to test this hypothesis in the upcoming Chapter (Section 6.1).

Another issue with this data, is that so far I have expressed image quality only in units of retinal contrast ratio. Of greater importance for a display application is whether these differences in contrast ratio are perceptible, and if they make subjectively high quality images. This problem will be addressed in Section 6.2.

Chapter 6

Perception of depth-filtered images

The analysis of retinal image formation has shown that there are a number of viewing situations and display configurations that produce high-contrast retinal images. However, until now, the analysis has considered exclusively optics. In this chapter we add factors of the brain to this analysis to estimate the quality of images in the volumetric display. Primarily, we are concerned with two properties of depth-filtered images. 1) They should stimulate an appropriate accommodative response. 2) The images should have perceptually similar contrast to real-world images.

6.1 Accommodation to depth-filtered images

The precision to which depth-filtered images drives accommodation is ultimately an empirical question. At Berkeley, we lacked a display capable of presenting volumetric images while simultaneously tracking accommodation. Fortunately, Simon Watt and Kevin Mackenzie at Bangor University in Wales, have recently constructed a display that is ideal for this experiment. I collaborated with Dr. Watt and MacKenzie to make predictions about how observers should accommodate to images presented in their volumetric display. As of March 2010, this work is in review at the Journal of Vision.

Watt and MacKenzie have constructed a multi-plane volumetric display similar to the 3-mirrors display (Section 1.2.1), but with two key differences. First, it uses Badal optics which allows image-plane spacing to be adjusted without needing to magnify the images to correct for linear perspective. This robustness to image plane position allows the image plane spacing to be adjustable. Second, an autorefractor (Grand Seiko WV-500) is integrated into the apparatus so that the display can present stimuli while simultaneously tracking accommodation, pupil diameter, version and vergence.

I used the model described throughout Chapter 5 to predict how an observer's accommodative response might become biased towards an image plane for positions that were not at the display's midpoint. In the spatial frequency analysis shown in Figure 5.2 there are

two local maxima in the accommodative response to the 16c/deg stimulus. This pull to the image planes can also be seen for the volumetric display in the third column of Figure 5.7, which shows the retinal contrast for arbitrary accommodative responses and simulated positions within the workspace. Conversely, at low spatial frequencies, throughout the display's volume there is a single accommodative response that is optimal for each simulated position, and this optimal position is highly correlated with the simulated position.

To predict how accommodation might respond to depth filtered images, MacKenzie, Watt and I computed the accommodative response that maximizes retinal contrast at various spatial frequencies and for several different image-plane spacings. We used the model of retinal image formation described in Chapter 5. The results from these calculations are shown in Figure 6.1. Each plot represents a different image-plane spacing. The abscissa represents the simulated position, and the ordinate represents the accommodative response where the contrast ratio is maximized. The colors, red, green, blue and magenta represent 2, 4, 8, and 16c/deg respectively. In the first panel, the images planes (represented in dark green) are separated by 4/9D. At this spacing, the optimal accommodative response closely tracks the simulated distance as it transitions from one image plane to the next. In the last plot with 12/9D spacing, we predict accommodation would continuously track a 2c/deg stimulus, but could not track the 8 or 16c/deg stimulus at all. At these spatial frequencies, contrast is maximized by accommodating to one of the image planes. At 4c/deg accommodation is just beginning to become biased towards the closer image plane, and we predict that accommodation should be slightly sigmoidal.

With intermediate plane spacings of 6/9, 8/9 and 10/9D, there is a shift in the spatial frequency at which we would predict the accommodative response becomes sigmoidal. At the smallest plane spacing we analyzed, (4/9D), we predict accommodation would not resemble a step function until beyond 25 c/deg.

Given this strong spatial frequency dependence on how we expect accommodation to behave, it becomes important to understand which spatial frequencies drive accommodation. Empirical research with sinusoid stimuli has revealed that accommodation seems to respond most effectively to spatial frequencies between 4-8 c/deg[91, 101, 104, 128, 136]. At these spatial frequencies, there is still reasonably good modulation transfer, and relatively minor defocus could lead to a significant change in contrast ratio. Thus we would expect that accommodation would begin to adopt a sigmoidal response around 4-8c/deg range.

To test this hypothesis MacKenzie and Watt measured accommodative responses directly. They used image-plane spacings of 4/9, 6/9, 8/9, 10/9, and 12/9D to present monocular (to avoid convergence-driven accommodation) white Maltese crosses to the observers. The observers' accommodative responses only became sigmoidal for the condition with 12/9D spacing. Across all the image plane spacings, the accommodative response was consistent with the visual system using 4c/deg to control accommodation. (One caveat of MacKenzie and Watt's results is that the gain for accommodation is not 1. Observers tended to under-accommodate to near stimuli, even when the stimulus was presented entirely on a single

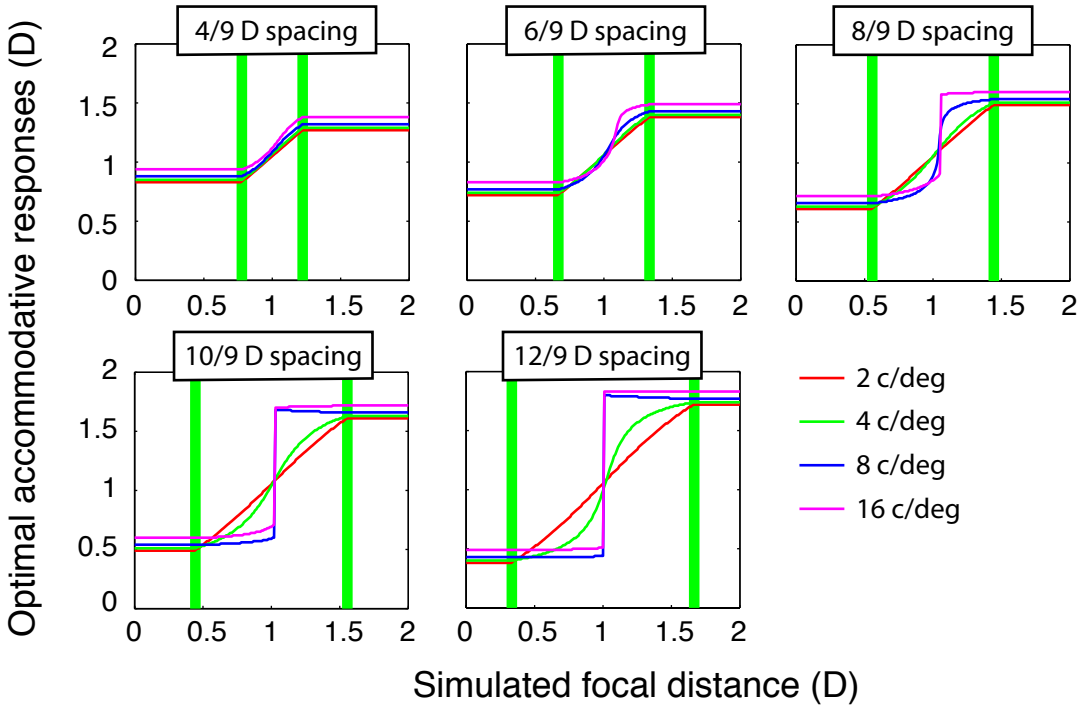


Figure 6.1: Optimal accommodative responses. In each panel, the accommodative response that produces that maximum contrast ratio is plotted as a function of the simulated position. Each panel has a different image plane separation, and 4 spatial frequencies, 2, 4, 8, and 16 c/deg in red, green, blue and magenta respectively. The location of the image planes are marked by the heavy green lines.

image plane.) The results of this experiment offer evidence that a relatively large image plane spacing can be used to stimulate correct accommodation.

6.2 The visual sensitivity to retinal contrast differences

The data collected by Mackenzie and Watt suggest that image plane spacing as large as 12/9D could be used to stimulate appropriate accommodation with a broadband stimulus. However, even if the display stimulates appropriate accommodation, the image may still be visually unacceptable. The acceptability of an image is stimulus and application dependent, and ultimately is a subjective question. However, it is possible to objectively predict viewing conditions and stimuli that will be perceptibly indistinguishable, based on contrast discrimination.

Bradley and Ohzawa (1986)[16] measured the sensitivity of human observers to a contrast increment. This is a measure of the minimum contrast increase needed for the visual system to perceive a change, and this threshold varies with spatial frequency, baseline contrast and the average luminance. Generally, the visual system becomes more sensitive to contrast changes at higher luminance, and since we can constrain the display to have the same luminance as its natural image counterpart, we ignore the effect of luminance in this analysis.

For a given spatial frequency, contrast sensitivity varies with the baseline contrast like a dipper function. In the nonlinear region at sub-threshold contrasts, the contrast increment is simply the detection threshold. The observer is discriminating the stimulus from an empty field. But, as the baseline contrast increases, a curious phenomenon occurs where, although sub-threshold, the baseline contrast will sum with the contrast increment and the visual system will be hypersensitive to the contrast increment. Beyond this dip in the contrast sensitivity function, the visual system nearly follows *Weber's law* which states that a detectable change in contrast will be a constant fraction of the baseline contrast. This means that if we double the baseline contrast, we can expect the detectable contrast change will also double. Because the ratio of signal to baseline remains constant, it is called a Weber fraction.

With other spatial frequencies, the contrast discrimination function maintains the same dipper shape but shifts so that the constant for the Weber portion of the contrast discrimination function remains constant. The magnitude of the shift corresponds with the contrast sensitivity threshold at that spatial frequency. We can use this data from Bradley and Ohzawa[16] to predict whether the difference between a depth filtered image and a real image will be perceptible for different image plane spacing, spatial frequencies, and accommodative responses.

Figure 6.2 uses grayscale to represent the retinal contrast ratio for a range of simulated positions (abscissa) and accommodative responses (ordinate). The columns represent spatial frequencies of 2, 6 and 18 c/deg, and each row is for a different display. The top row shows

the expected retinal contrasts for a real world image. The lower panels simulate real world focus information with image plane spacing at 4/9, 6/9, 8/9, and 10/9 D which are marked by the dark vertical lines. For this calculation, objects beyond the image planes clamp to the nearest image plane. Overlaid with this analysis, I have included contours marking the zones where there will be a perceptible contrast difference between the display and the real world image. The green contours show the zones where the retinal contrast in the display will be perceptibly lower than for the real world. The red contours mark the boundary of zones where contrast would be perceptibly higher in the multi-plane display.

The green zones are especially troubling because at these stimuli positions, an object with reduced contrast in the display could appear blurry. This problem occurs most severely at the midpoint of the two image planes in cases in which the eye has accommodated to the simulated distance. Because the eye has accommodated to a distance in which no light originates, the image is slightly out of focus. The analysis in Figure 6.2 shows that this problem is severe at 16 c/deg for all the image plane spacings, but at 6 c/deg it is not a problem at 4/9D plane spacing and only has a limited zone where it is a problem 6/9D. Accommodation typically fluctuates and may lead or lag the optimal response, and therefore the slight contrast decrement at 6/9 c/deg may be imperceptible.

In addition to zones where we are concerned that imagery may appear blurred, we are also concerned with zones where images may appear too clear. In scenes depicting multiple distances, as discussed in Chapter 3, blur can be a useful cue for monocular occlusion and binocular correspondence. An image that is unnaturally clear could disrupt this signal. Fortunately, in the volume bracketed by the image planes, unnaturally sharp objects only occur when image plane spacing approaches 1D. One approach to compensate for unnatually sharp objects is to use a filter that attenuates high spatial frequencies as an object gets closer to an image plane.

6.3 Acceptable image quality

Contrast discrimination threshold is possibly the most challenging test for a display. It encourages observers to identify a difference. There are few displays of any type that could pass such an objective measure and produce images that are truly indistinguishable from the real world. Also, this analysis has assumed that each of these narrowband sinusoids was independent. With broadband stimuli, it is highly unlikely that the visual system would be as sensitive to a high frequency image component when there are strong low and medium components.

These stimuli may exist in real world viewing situations, such as viewing a pin-stripe suit from across the room, or newspaper print (30 c/deg), but they are certainly not natural viewing situations. Natural scenes contain a spectrum of spatial frequencies, and typically the contrast or amplitude of the higher spatial frequencies tends to be small relative to the

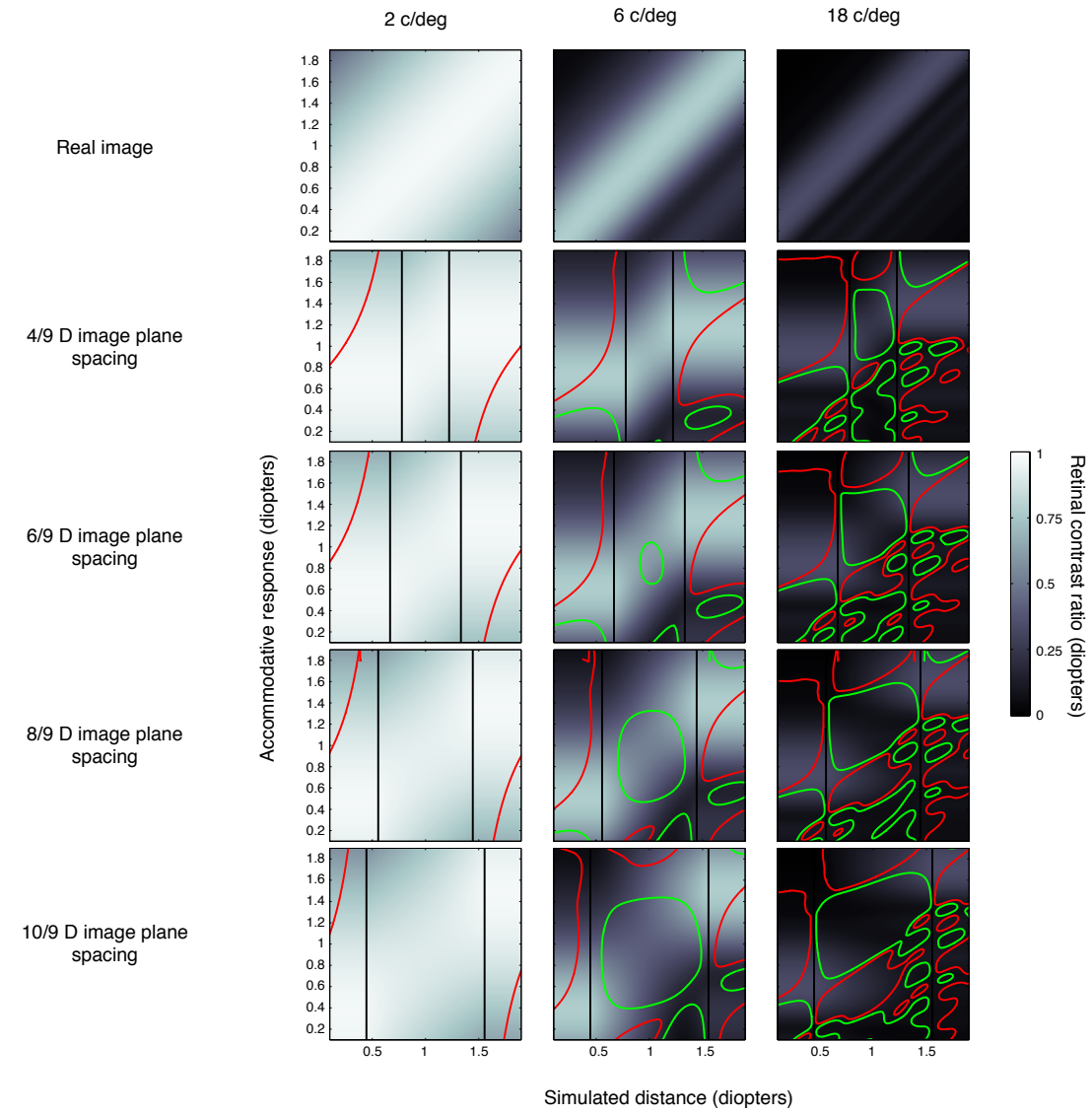


Figure 6.2: Perceptible differences in contrast. Each plot shows simulated distance on the abscissa and hypothetical accommodative responses on the ordinate. Each column represents a different spatial frequency: 2, 6, and 18 c/deg. Each row contains a different display–from top to bottom: real world, and 4/9, 6/9, 8/9 and 10/9 D image plane spacing volumetric displays. The image planes are marked with the black lines. Conditions in which the display has perceptibly less contrast are marked with green outlines and conditions where contrast is too high are marked with red.

lower spatial frequencies. People often describe this distribution as the $1/f$ frequency distribution, where the amplitude of each spatial frequency component of an image is inversely proportional to its spatial frequency, f. All of the analyses that I have included in this dissertation have dealt with contrast ratio, and were thus independent of the stimulus contrast. A theoretical analysis of a complex stimulus including a $1/f$ spectrum would inevitably be more forgiving than the analysis that I have reported here. Blur detection is often described as a problem of determining if the amplitude spectrum of an image is consistent with natural scene statistics[40]. And thus instead of determining the contrast of high spatial frequencies in isolation, they would be part of a broad range of information, and the visual system would examine the image holistically to judge its blurriness.

Chapter 7

Discussion and ongoing work

7.1 General interest in stereo "3D"

I have demonstrated that the subtle difference between correct focus signals in natural viewing and the incorrect focus signals in a stereo display produce objective and subjective differences in the viewing experience. This work has garnered international attention in the last two years as 3D cinema has staged a resurgence. Reporters have asked us why some viewers of 3D movies, such as Avatar, leave the theater with blurred vision, headaches and eyestrain. As a lab, we have explained the visual fatigue experiment and how it is truly the only experiment with the critical comparison–testing identical images except for focus signals. This experiment frames the problem in an unambiguous way, but can only offer limited information on how to correct the problem.

Directors of stereoscopic cinema have an immense amount of control over the disparities they present in their film. They often have control over the camera separation, shooting distance, the convergence of the cameras and the magnification of the images. They use these parameters to achieve certain story-telling goals, such as ramping up the depth during moments of excitement, and flattening scenes during dull moments. Other stereoscopic directors consider stereoscopic depth as a finite resource and attempt to budget their use of depth throughout the movie to alleviate fatigue. At this time, there is no underlying science that offers clear guidelines for what disparity configurations will cause visual fatigue and for what fraction of the population. We can only suggest ways to minimize the size of the conflicts that are created, which unfortunately often will minimize the stereoscopic effects that the directors would like to achieve.

Even some of the simple techniques for stereo imaging remain controversial. One such question is whether it is more comfortable to view crossed or uncrossed disparity in a cinema setting. And a related question is whether the same disparities presented at a near display such as a TV, 2 meters away, or on a mobile display, 50 cm away will generate more or less fatigue. These questions are part of a larger problem dealing with how stereoscopic images,

which are designed for being viewed from a specific position, are perceived when viewed from incorrect positions. The position one sits in a movie theater, or home theater, will greatly change both the retinal disparities and the accommodative distance, and if the content is not adapted carefully, could change the depth percepts.

7.2 Age related issues

What is the influence of observer age on stereoscopic viewing comfort? Children are flexible enough to deal with conflicting focus signals, and have an impressive accommodative amplitude, but their brains are also plastic enough that stereoscopic display viewing could change the way their visual system develops. An important question is if chronic viewing of stereoscopic displays could permanently change a child's accommodative-convergence cross-coupling, and if such a change is of concern [14].

During middle adulthood, the visual system undergoes another change–the crystalline lens hardens and the visual system loses the ability to accommodate. When this happens, people require reading glasses, or use either bifocal or progressive lenses. These adults have essentially learned how to use glasses to replace their accommodation response. This learned response is not needed in typical stereo displays. This suggests that the problems of visual fatigue due to this conflict between vergence and accommodation may be less of a problem for presbyopes than for people who can still accommodate. If we had test this group of observers in the visual fatigue experiment (Section 2.5), perhaps we would have found the opposite result.

These age-related issues are only one part of a thorny puzzle of predicting the differing severity of this problem for different individuals. The strength of the cross coupled signals for the accommodation and convergence varies between observers, and the literature is mixed on whether these differences can be used to anticipate problems in viewing stereoscopic imagery[47, 137].

7.3 The switchable lens display as a research tool

As vision scientists, we can speculate on how age, content and viewing configuration will affect percepts, but, with the switchable lens display, we can begin to design and conduct definitive tests to answer these questions. This research tool provides a superb platform for conducting the experiments needed to provide scientific advice on the least objectionable techniques for presenting stereo imagery. The display hardware allows for computerized control of focus signals within the display workspace and by inserting ophthalmic lenses and adjusting the prism angles, the workspace can be translated from infinity to 30cm. I have designed the software so that it is quick and easy to produce a new experiment.

Already, this display is gaining traction as an important experimental tool for the lab. I am collaborating with postdocs Takashi Shibatta and Joohwan Kim to extend the visual fatigue experiment (Section 2.5) to map out the zone of comfort. In the first experiment, we are examining whether the zone of comfort has a uniform width in dioptric units or whether it fans out at near viewing distances. In a followup experiment we are investigating whether changing the sign of the disparity (presenting either crossed or uncrossed stimuli) at a given focal distance will cause greater or less fatigue. We hope to compile this data from a series of experiments to provide clear advice for makers of stereo displays on some comfortable limits on disparity magnitudes and signs for viewing distances.

Two other colleagues, Robin Held and Emily Cooper, are currently using the display to investigate the importance of defocus and pictorial blur as a cue to relative depth. This research, as well as the research discussed in Chapter 3, may be pertinent to the way in which pictorial blur and depth of field is used in stereoscopic movies. Currently, the depth of field computed in the image is a stylistic choice that is made independently of the stereoscopic camera separation and shooting distance. Some of our experiments could demonstrate that there are significant advantages to coupling these two system parameters.

7.4 Image quality

The analysis of retinal image formation provides some insight into how the display creates images, but it is an approximation. Ultimately, it breaks down a retinal image into several contrast ratios. The important question remains–how do these depth-filtered images look? So far, I have been unable to design an experiment which adequately answers this question. The demonstrations that we use to showcase the display are not natural images, and it is unclear whether the known display artifacts would be objectionable, or whether the display would be preferable to a conventional stereoscopic displays.

An important project that relates to this topic is assessing if the tent depth filter is indeed optimal. The accommodation experiment has demonstrated that the tent filter drives accommodation accurately, but our analysis shows that its stimulation at high spatial frequencies is incorrect. It may be possible to design a more complex, spatial frequency dependent, depth filter that better emulates real focus cues. A related improvement could be to perform post-processing operations on the images before they are presented in the display. Such operations could mitigate the edge boundary problem, or perform contrast enhancement to reduce blurriness in difficult to render locations, such as the midpoint of two image planes.

7.5 Putting this research to practice

In the near future, I hope to see some of the Vision Science stereo research become a more influential part of content preparation and display. There are a number of choices that

display engineers and film directors have available to them that could change the viewing experience. Basing these choices on solid theory from the academic community could help to make the viewing experience more comfortable and thus more fulfilling. I recommend that the industry proceed with measured caution. Stereo content is not a thrill ride, and bigger disparities do not necessarily yield a stronger sense of depth; and the thresholds where images transition from exciting to fatiguing, and from immersive to gimmicky are largely unmapped. Some of the tenets of this thesis could prove valuable for determining these thresholds and making effective choices during stereo image production.

Bibliography

[1] Kurt Akeley, Simon J. Watt, Ahna Reza Girshick, and Martin S. Banks. A stereo display prototype with multiple focal distances. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 804–813, New York, NY, USA, 2004. ACM.

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

[3] B.L. Anderson and K. Nakayama. Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychological Review*, 101(3):414–445, 1994.

[4] D.A. Atchison, S.W. Fisher, C.A. Pedersen, and P.G. Ridall. Noticeable, troublesome and objectionable limits of blur. *Vision Research*, 45(15):1967–1974, 2005.

[5] B T Backus and M S Banks. Estimator reliability and distance scaling in stereoscopic slant perception. *Perception*, 28(2):217–242, 1999.

[6] B.T. Backus, M.S. Banks, R. van Ee, and J.A. Crowell. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39(6):1143–1170, 1999.

[7] AT Bahill, D. Adler, and L. Stark. Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science*, 14(6):468, 1975.

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

[9] EA Ball. A study of consensual accommodation. *American journal of optometry and archives of American Academy of Optometry*, 29(11):561, 1952.

[10] M.S. Banks, S. Gepshtein, and M.S. Landy. Why is spatial stereoresolution so low? *Journal of Neuroscience*, 24(9):2077–2089, 2004.

[11] RE Bedford and G. Wyszecki. Axial chromatic aberration of the human eye. *Journal of the Optical Society of America*, 47(6):564–565, 1957.

[12] P.N. Belhumeur and D. Mumford. A bayesian treatment of the stereo correspondence problem using half-occluded regions. In *Proc. Computer Vision and Pattern Recognition Conf*, pages 506–512. Citeseer, 1992.

[13] Y. Bereby-Meyer, D. Leiser, and J. Meyer. Perception of artificial stereoscopic stimuli from an incorrect viewing point. *Perception and Psychophysics*, 61(8):1555–1563, 1999.

[14] SR Bharadwaj and TR Candy. Accommodative and vergence responses to conflicting blur and disparity stimuli during development. *Journal of Vision*, 9(11):4, 2009.

[15] C. Blakemore. The range and scope of binocular depth discrimination in man. *The Journal of Physiology*, 211(3):599–622, 1970.

[16] A. Bradley and I. Ohzawa. A comparison of contrast detection and discrimination. *Vision Research*, 26(6):991–997, 1986.

[17] D.H. Brainard. The psychophysics toolbox. *Spatial vision*, 10(4):433–436, 1996.

[18] M.L. Braunstein, J.C. Liter, and J.S. Tittle. Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of experimental psychology, human perception and performance*, 19:598–598, 1993.

[19] D. Buckley and JP Frisby. Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision research(Oxford)*, 33(7):919–933, 1993.

[20] J. Burge, M.A. Peterson, and S.E. Palmer. Ordinal configural cues combine with metric disparity in depth perception. *Journal of Vision*, 5(6):5, 2005.

[21] O. Cakmakci and J. Rolland. Head-worn displays: A review. *Journal of Display Technology*, 2(3):199–216, 2006.

[22] FW Campbell. The depth of field of the human eye. *Journal of Modern Optics*, 4(4):157–164, 1957.

[23] FW Campbell and DG Green. Optical and retinal factors affecting visual resolution. *The Journal of Physiology*, 181(3):576–593, 1965.

[24] FW Campbell and JG Robson. Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197(3):551–566, 1968.

[25] C. Caudek and D.R. Proffitt. Depth perception in motion parallax and stereokinesis. *Journal of experimental psychology, human perception and performance*, 19:32–32, 1993.

[26] WN Charman and H. Whitefoot. Pupil diameter and the depth-of-field of the human eye as measured by laser speckle. *Journal of Modern Optics*, 24(12):1211–1216, 1977.

[27] L. Chen, P.B. Kruger, H. Hofer, B. Singer, and D.R. Williams. Accommodation with higher-order monochromatic aberrations corrected with adaptive optics. *Journal of the Optical Society of America A*, 23(1):1–8, 2006.

[28] H. Cheng, J.K. Barnett, A.S. Vilupuru, J.D. Marsack, S. Kasthurirangan, R.A. Applegate, and A. Roorda. A population study on changes in wave aberrations with accommodation. *Journal of Vision*, 4(4):272–280, 2004.

[29] WG Cochran. Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, pages 102–118, 1937.

[30] L.K. Cormack, S.B. Stevenson, and D.D. Landers. Interactions of spatial frequency and unequal monocular contrasts in stereopsis. *PERCEPTION-LONDON-*, 26:1121–1136, 1997.

[31] S.H. Creem-Regehr, P. Willemsen, A.A. Gooch, and W.B. Thompson. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual environments. *Perception*, 34(2):191–204, 2005.

[32] BG Cumming and SJ Judge. Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey. *Journal of Neurophysiology*, 55(5):896–914, 1986.

[33] SG De Groot and JW Gebhard. Pupil size as determined by adapting luminance. *Journal of the Optical Society of America*, 42(7):492–495, 1952.

[34] HEM den Ouden, R. van Ee, and EHF de Haan. Colour helps to solve the binocular matching problem. *The Journal of Physiology*, 567(2):665, 2005.

[35] F. Domini and C. Caudek. Perceiving surface slant from deformation of optic flow. *Journal of experimental psychology, human perception and performance*, 25:426–444, 1999.

[36] S.R. Ellis, S. Smith, A. Grunwald, and M.W. McGreevy. Direction judgement error in computer generated displays and actual scenes. *Pictorial communication in virtual and real environments*, pages 504–526, 1991.

[37] M. Emoto, T. Niida, and F. Okano. Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. *Display Technology, Journal of*, 1(2):328–340, 2005.

[38] M. Fahle. Binocular rivalry: suppression depends on orientation and spatial frequency. *Vision research*, 22(7):787, 1982.

[39] G.E. Favalora, J. Napoli, D.M. Hall, R.K. Dorval, M.G. Giovinco, M.J. Richmond, and W.S. Chun. 100 million-voxel volumetric display. In *Proc. SPIE*, volume 4712, pages 300–312, 2002.

[40] D.J. Field and N. Brady. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37(23):3367–3383, 1997.

[41] EF Fincham. The accommodation reflex and its stimulus. *British Medical Journal*, 35(7):381, 1951.

[42] EF Fincham and J. Walton. The reciprocal actions of accommodation and convergence. *The Journal of physiology*, 137(3):488, 1957.

[43] J.P. Frisby, D. Buckley, and P.A. Duke. Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *PERCEPTION-LONDON-*, 25:129–154, 1996.

[44] J.P. Frisby, D. Buckley, and J.M. Horsman. Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception-London*, 24(1):181–198, 1995.

[45] G.A. Fry. Further experiments on the accommodation convergence relationship. *Am J Optom*, 16:125, 1939.

[46] GA Fry and PR Kent. The effects of base-in and base-out prisms on stereo-acuity. *Am J Optom*, 21:492–507, 1944.

[47] T. Fukushima, M. Torii, K. Ukai, JS Wolffsohn, and B. Gilmartin. The relationship between ca/c ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images. *Journal of Vision*, 9:13–21, 2009.

[48] J. Garding, J. Porrill, JEW Mayhew, and JP Frisby. Stereopsis, vertical disparity and relief transformations. *Vision Research*, 35(5):703–722, 1995.

[49] Z. Ghahramani, D.M. Wolpert, and M.I. Jordan. Computational models of sensorimotor integration. In P.G. Morasso and V. Sanguineti, editors, *Self-organization, computational maps, and motor control*, pages 117–147. Amsterdam: North-Holland, 1997.

[50] RT Goodwin and PE Romano. Stereoacuity degradation by experimental and real monocular and binocular amblyopia. *Investigative ophthalmology & visual science*, 26(7):917, 1985.

[51] EM Granger and KN Cupery. An optical merit function (sqf) which correlates with subjective image judgements. *Photographic science and engineering*, 16(3):221–230, 1972.

[52] DG Green and FW Campbell. Effect of focus on the visual response to a sinusoidally modulated spatial stimulus. *J. Opt. Soc. Am*, 55(9):1154–1157, 1965.

[53] DG Green, MK Powers, and MS Banks. Depth of focus, eye size and visual acuity. *Vision Research*, 20(10):827, 1980.

[54] J. Häkkinen, M. Pölönen, J. Takatalo, and G. Nyman. Simulator sickness in virtual display gaming: A comparison of stereoscopic and non-stereoscopic situations. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services*, pages 227–230, 2006.

[55] D.L. Halpern and R.R. Blake. How contrast affects stereoacuity. *Perception*, 17(4):483–495, 1988.

[56] R Held, Cooper EA, J O'Brien, and MS Banks. Using blur to affect perceived distance and size. *ACM Transactions on Graphics*, in press, 2010.

[57] J. Hillis, S. Watt, M. Landy, and M.S. Banks. Slant from texture and disparity cues: optimal cue combination. *Journal of vision*, 4(12):967–992, 2004.

[58] J.M. Hillis and M.S. Banks. Are corresponding points fixed? *Vision Research*, 41:2457–2473, 2001.

[59] DM Hoffman, AR Girshick, K. Akeley, and MS Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 2008.

[60] D.M. Hoffman, P.J.W. Hands, A.K. Kirby, G.D. Love, and M.S. Banks. Stereo display with time-multiplexed focal adjustment. In *Proceedings of SPIE*, volume 7237, page 72370R, 2009.

[61] MA Hogervorst and RA Eagle. Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings: Biological Sciences*, 265(1406):1587–1593, 1998.

[62] M.A. Hogervorst and R.A. Eagle. The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *Journal of experimental psychology, human perception and performance*, 26(3):934–955, 2000.

[63] K. Hopf, P. Chojecki, F. Neumannn, and D. Przewozny. Novel autostereoscopic single-user displays with user interaction. In *Proceedings of SPIE*, volume 6392, page 639207, 2006.

[64] I.P. Howard and B.J. Rogers. *Seeing in depth*. I. Porteous, 2002.

[65] PA Howarth and PJ Costello. The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, 18(2):107–116, 1997.

[66] J.C.K. IT and C.M. Schor. The accommodative response to subthreshold blur and to perceptual fading during the troxler phenomenon. *Perception*, 15:7–15, 1986.

[67] R.A. Jacobs. Optimal integration of texture and motion cues to depth. *Vision Research*, 39(21):3621–3629, 1999.

[68] RJ Jacobs, G. Smith, and CDC Chan. Effect of defocus on blur thresholds and on thresholds of perceived change in blur: comparison of source and observer methods. *Optometry & Vision Science*, 66(8):545, 1989.

[69] Z.X. Jin, Y.J. Zhang, X. Wang, and T. Plocher. Evaluating the usability of an autostereoscopic display. *Lecture Notes in Computer Science*, 4551:605, 2007.

[70] EB Johnston. Systematic distortions of shape from stereopsis. *Vision Research*, 31(7-8):1351, 1991.

[71] H. Jorke, A. Simon, and M. Fritz. Advanced stereo projection using interference filters. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008*, pages 177–180, 2008.

[72] B. Julesz. *Foundations of cyclopean perception*. University of Chicago Press Chicago:, 1971.

[73] M. Khosroyani and G.K. Hung. A dual-mode dynamic model of the human accommodation system. *Bulletin of mathematical biology*, 64(2):285–299, 2002.

[74] A.K. Kirby, P.J.W. Hands, and G. D. Love. Adaptive lenses based on polarization modulation. In *Adaptive Optics for Industry and medicine*, volume 14. Proc. of SPIE 6018, 2005.

[75] D.C. Knill and J.A. Saunders. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43(24):2539–2558, 2003.

[76] F.L. Kooi and A. Toet. Visual comfort of binocular and 3D displays. *Displays*, 25(2-3):99–108, 2004.

[77] VV Krishnan, D. Shirachi, and L. Stark. Dynamic measures of vergence accommodation. *American journal of optometry and physiological optics*, 54(7):470, 1977.

[78] J.D. Krol and W.A. van de Grind. The double-nail illusion: experiments on binocular vision with nails, needles, and pins. *Perception*, 9(6):651–669, 1980.

[79] PB Kruger and J. Pola. Stimuli for accommodation: blur, chromatic aberration and size. *Vision research*, 26(6):957, 1986.

[80] H. Kuribayashi, Y. Ishigure, S. Suyama, H. Takada, M. Date, K. Ishikawa, and T. HATADA. Effect on depth perception by a blur in a depth-fused 3-D display. *Journal of the Institute of Image Information and Television Engineers*, 60(3):431–438, 2006.

[81] G.E. Legge and Y. Gu. Stereopsis and contrast. *Vision Research*, 29(8):989–1004, 1989.

[82] S. Liu and H. Hua. Time-multiplexed dual-focal plane head-mounted display with a liquid lens. *Optics Letters*, 34(11):1642–1644, 2009.

[83] JM Loomis and DW Eby. Relative motion parallax and the perception of structure from motion. In *Proceedings from the workshop on visual motion*, pages 204–211, 1989.

[84] J.M. Loomis, N. Fujita, J.A. Da Silva, and S.S. Fukusima. Visual space perception and visually directed action. *Journal; i Exp m mental Psychology: Human I'crccptiori unci Performance*, 1(4):921, 1992.

[85] G.D. Love, D.M. Hoffman, P.J.W. Hands, J. Gao, A.K. Kirby, and M.S. Banks. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express*, 17(18):15716–15725, 2009.

[86] J.A. Marshall, C.A. Burbeck, D. Ariely, J.P. Rolland, and K.E. Martin. Occlusion edge blur: a cue to relative visual depth. *Journal of the Optical Society of America A*, 13(4):681–688, 1996.

[87] TG Martens and KN Ogle. Observations on accommodative convergence; especially its nonlinear relationships. *American journal of ophthalmology*, 47(1 Part 2):455, 1959.

[88] G. Mather. The use of image blur as a depth cue. *PERCEPTION-LONDON-*, 26:1147–1158, 1997.

[89] G. Mather and D.R.R. Smith. Depth cue integration: stereopsis and image blur. *Vision Research*, 40(25):3501–3506, 2000.

[90] G. Mather and D.R.R. Smith. Blur discrimination and its relation to blur-mediated depth perception. *Perception*, 31(10):1211–1220, 2002.

[91] S. Mathews and PB Kruger. Spatiotemporal transfer function of human accommodation. *Vision research*, 34(15):1965, 1994.

[92] M. Menozzi. Visual ergonomics of head-mounted displays. *Japanese Psychological Research*, 42(4):213–221, 2000.

[93] FA Miles, SJ Judge, and LM Optican. Optically induced changes in the couplings between vergence and accommodation. *Journal of Neuroscience*, 7(8):2576–2589, 1987.

[94] M. Mon-Williams, J.R. Tresilian, and A. Roberts. Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research*, 133(3):407–413, 2000.

[95] MW Morgan. The clinical aspects of accommodation and convergence. *Am J Optom Arch Am Acad Optom*, 21(8):301–13, 1944.

[96] T.J. Mueller and R. Blake. A fresh look at the temporal dynamics of binocular rivalry. *Biological cybernetics*, 61(3):223–232, 1989.

[97] JV Odom, G. Chao, and M. Leys. Symmetrical refractive error elevates stereo thresholds. In *Engineering in Medicine and Biology Society, 1992. Vol. 14. Proceedings of the Annual International Conference of the IEEE*, volume 4, 1992.

[98] K.N. Ogle. An analytical treatment of the longitudinal horopter, its measurement and application to related phenomena, especially to the relative size and shape of the ocular images. *Journal of the Optical Society of America*, 22(12):665, 1932.

[99] Y. Okada, K. Ukai, J.S. Wolffsohn, B. Gilmartin, A. Iijima, and T. Bando. Target spatial frequency determines the response to conflicting defocus-and convergence-driven accommodative stimuli. *Vision Research*, 46(4):475–484, 2006.

[100] R.P. O'Shea, D.G. Govan, and R. Sekuler. Blur and contrast as pictorial depth cues. *Perception*, 26:599–612, 1997.

[101] DA Owens. A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings. *Vision research*, 20(2):159, 1980.

[102] K.R. Paap and S.M. Ebenholtz. Perceptual consequences of potentiation in the extraocular muscles: an alternative explanation for adaptation to wedge prisms. *Journal of experimental psychology: Human Perception and Performance*, 2(4):457–68, 1976.

[103] A.S. Percival. *The prescribing of spectacles*. Bristol: J. Wright., 1920.

[104] S.R. Phillips. *Ocular neurological control systems: Accommodation and the near response triad*. PhD thesis, University of California, Berkeley, 1974.

[105] S. Plainis, H.S. Ginis, and A. Pallikaris. The effect of ocular aberrations on steady-state errors of accommodative response. *Journal of Vision*, 5(5):7, 2005.

[106] BJ Rogers and MF Bradshaw. Binocular judgements of depth, size, shape and absolute distance: is the same 'd'used for all judgements. *Invest Ophthalmol Vis Sci*, 36:230, 1995.

[107] J.P. Rolland, M.W. Krueger, and A.A. Goon. Dynamic focusing in head-mounted displays. In *Proceedings of SPIE*, volume 3639, page 463, 1999.

[108] C.S. Sahm, S.H. Creem-Regehr, W.B. Thompson, and P. Willemsen. Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Transactions on Applied Perception (TAP)*, 2(1):35–45, 2005.

[109] Y.Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 39(2):141–162, 2000.

[110] C. Schor and T. Heckmann. Interocular differences in contrast and spatial frequency: effects on stereopsis and fusion. *Vision Research*, 29(7):837, 1989.

[111] C. Schor, I. Wood, and J. Ogawa. Binocular sensory fusion is limited by spatial resolution. *Vision research(Oxford)*, 24(7):661–665, 1984.

[112] CM Schor and TK Tsuetaki. Fatigue of accommodation and vergence modifies their mutual interactions. *Investigative ophthalmology & visual science*, 28(8):1250–1259, 1987.

[113] B.T. Schowengerdt. True 3-D scanned voxel displays using single or multiple light sources. *Journal of the SID*, page 135, 2006.

[114] J. Semmlow and P. Wetzel. Dynamic contributions of the components of binocular vergence. *Journal of the Optical Society of America*, 69(5):639, 1979.

[115] J. Sheedy, P.D. OD, and N. BERGSTROM. Performance and Comfort on Near-Eye Computer Displays. *Optometry and Vision Science*, 79(5):306, 2002.

[116] T. Shibata. Stereoscopic 3-d display with optical correction for the reduction of the discrepancy between accommodation and convergence. *Journal of the SID*, page 665, 2005.

[117] S. Shiwa, K. Omura, and F. Kishino. Proposal for a 3-d display with accommodative compensation: 3ddac. *Journal of the Society for Information Display*, 4:255, 1996.

[118] H.S. Smallman and S.P. Mckee. A contrast ratio constraint on stereo matching. *Proceedings: Biological Sciences*, pages 265–271, 1995.

[119] A. Sullivan. DepthCube solid-state 3D volumetric display. In *Proceedings of SPIE*, volume 5291, page 279, 2004.

[120] S. Suyama, M. Date, and H. Takada. Three-dimensional display system with dual-frequency liquid-crystal varifocal lens. *Japanese Journal of Applied Physics*, 39(part 1):480–484, 2000.

[121] H. TAKADA. The progress of high presence and 3D display technology. The Depth-Fused 3-D display for the eye sweetly. *Optical and Electro-Optical Engineering Contact*, 44(6):316–323, 2006.

[122] Y. Takaki. Novel 3D display using an array of LCD panels. In *Proceedings of SPIE*, volume 5003, page 1, 2003.

[123] L.N. Thibos, X. Hong, A. Bradley, and X. Cheng. Statistical variation of aberration structure and image quality in a normal population of healthy eyes. *Journal of the Optical Society of America A*, 19(12):2329–2348, 2002.

[124] J.T. Todd. The visual perception of 3D shape. *Trends in Cognitive Sciences*, 8(3):115–121, 2004.

[125] JT Todd and P. Bressan. The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, 48(5):419, 1990.

[126] M. Torii, Y. Okada, K. Ukai, J.S. Wolffsohn, and B. Gilmartin. Dynamic measurement of accommodative responses while viewing stereoscopic images. *Journal of Modern Optics*, 55(4):557–567, 2008.

[127] I. Tsirlin, RS Allison, and LM Wilcox. Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision*, 8(5):5, 2008.

[128] J. Tucker, WN Charman, and PA Ward. Modulation dependence of the accomodation response to sinusoidal gratings. *Vision research(Oxford)*, 26(10):1693–1707, 1986.

[129] K. Ukai. Visual fatigue caused by viewing stereoscopic images and mechanism of accommodation. In *Proceedings of the First International Symposium on University Communication*, pages 176–179, 2007.

[130] R. van Ee and B.L. Anderson. Motion direction, speed and orientation in binocular matching. *Nature*, 410(6829):690–694, 2001.

[131] D. Vishwanath, A.R. Girshick, and M.S. Banks. Why pictures look right when viewed from the wrong place. *Nature neuroscience*, 8(10):1401–1410, 2005.

[132] H. Wallach and C.M. Norris. Accommodation as a distance-cue. *The American Journal of Psychology*, pages 659–664, 1963.

[133] G. Walsh and WN Charman. Visual sensitivity to temporal change in focus and its relevance to the accommodation response. *Vision research*, 28(11):1207, 1988.

[134] J.P. Wann and M. Mon-Williams. Health issues with virtual reality displays: What we do know and what we don't. *ACM SIGGRAPH Computer Graphics*, 31(2):53–57, 1997.

[135] JP Wann and M. Mon-Williams. *Measurement of visual aftereffects following virtual environment exposure*. Lawrence Erlbaum Associates, 2002.

[136] PA Ward. The effect of spatial frequency on steady-state accommodation. *Ophthalmic and Physiological Optics*, 7(3):211–217, 1987.

[137] S.J. Watt, K. Akeley, M.O. Ernst, and M.S. Banks. Focus cues affect perceived depth. *Journal of vision*, 5(10):834, 2005.

[138] G. Westheimer and S.P. McKee. Stereoscopic acuity with defocused and spatially filtered retinal images. *Journal of the Optical Society of America*, 70(7):772–778, 1980.

[139] F.A. Wichmann and N.J. Hill. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63(8):1293–1313, 2001.

[140] F.A. Wichmann and N.J. Hill. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, 63(8):1314–1329, 2001.

[141] C.F. Wildsoet, H.C. Howland, S. Falconer, and K. Dick. Chromatic aberration and accommodation: their role in emmetropization in the chick. *Vision research*, 33(12):1593–1603, 1993.

[142] P. Willemsen, A.A. Gooch, W.B. Thompson, and S.H. Creem-Regehr. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(1):91–101, 2008.

[143] ICJ Wood. Stereopsis with spatially-degraded images. *Ophthalmic and Physiological Optics*, 3(3):337–340, 1983.

[144] A.J. Woods and K.L. Yuen. Compatibility of lcd monitors with frame-sequential stereoscopic 3d visualisation. *IMID/IDMC*, 6:98–102, 2006.

[145] T. Yamazaki, K. Kamijo, and S. Fukuzumi. Quantitative evaluation of visual fatigue encountered in viewing stereoscopic 3 D displays: near-point distance and visual evoked potential study. *Proceedings of the society for information display*, 31(3):245–247, 1990.

[146] S. Yano, M. Emoto, and T. Mitsuhashi. Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays*, 25(4):141–150, 2004.

[147] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites. A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays*, 23(4):191–201, 2002.

Appendix A

Calibration and alignment issues with the switchable lens display

One of the biggest challenges of producing high quality stereoscopic images using the switchable lens was developing a series of procedures to register the time multiplexed images to one another and to ensure that the vergence and accommodative stimuli were presented as intended. I have documented the detailed calibration routines on a lab wiki page, but I will provide a high level overview of the various calibration challenges and how each was addressed.

A.1 Optics calibration

A.1.1 Lens-element alignment

The switchable lens' optics are encased in aluminum housing. The housing and the thickness of the lens elements restricts their axial position, but their rotational position is adjustable. The orientation of the FLCs, lenses, and polarizer are of great importance. Angular misalignment of these elements, where the light polarization is not orthoganal to the axes of the calcite lenses, causes focal ghosting (where light bleeds into an incorrect focal plane).

The goal of the optical component calibration is to align the polarizer, calcite lenses, and FLCs so that the light incident on the calcite lenses has a polarization that is aligned with the primary refractive indices. The vector component of the light polarization that is improperly aligned will have an incorrect focal distance and it will also have a magnification error.

A.1.2 Hardware direction alignment

The lens system is mounted to a converted amblyoscope base which allows us to rotate the lens assemblies about a vertical axis that passes through the eyes' center of rotation. The amblyoscope base allows for translation of the observer forward and backwards as well as adjusting for variations in interocular separation.

Once the observer is aligned, the lens assemblies can be adjusted so that they are centered in front of the eyes, and are as vertical as possible. The prisms are attached to the same platform as the lenses, and misalignment will cause the prism rotation axis to be non-vertical, which creates an image rotation. Rotated images will induce cyclo-disparity into the stimuli.

If the observer is correctly positioned with respect to the optics assemblies (and the observers' eyes are positioned above the rotation pins for the amblyoscope), the arms of the amblyoscope can be rotated to change the vergence stimulus. This rotation does not change the monitor position, so the prisms must be used to keep the CRT screens in the center of the field of view.

A.1.3 Focal power calibration

The lenses are precision ground to exact radii of curvature. As such their focal power is precise. However, once mounted, their separation and FLC spacers cause the thin lens formula to be a poor approximation for the system's focal power. The effective focal power is difficult to measure because it creates virtual images. I found that the best way to measure the focal power for the lenses was to use a Keplerian telescope. A telescope magnifies an image by its labeled power but it collapses the depth of field by the square of its power: $V_r = (M_A)^2 U_{obj}$, where V_r is the vergence of the image at the Ramsden (where an observer would view the telescope), M_A is the angular magnification of the telescope, and U_{obj} is the incoming light vergence of the object. For a discussion of Keplerian telescopes, see the Keating textbook, *Geometric and Physical Optics*.

I used a 10X Keplerian telescope designed for low vision patients (10x20 SpecWell 6deg with extra short focus) because it features an exceptionally large focal range. To prevent my eyes from accommodating, I used Tropicamide eye drops. I focussed the telescope through the lens system to the display's pixel grid and could then measure the distance for which the telescope was focussed for a real object. With the telescope, I was able to measure each focal state of the lens system to within several hundredth's of a diopter. These measurements were confirmed by others using the same technique.

The system has two options for tuning the focal distance. Coarse adjustments can be made by adding ophthalmic lenses to the front of the lens assemblies. Fine tuning of the focal distance is possible by making small adjustments to the monitor distance. These distance changes will create small size differences but we correct for these with a software calibration (Section A.3.2).

A.2 CRT calibration

A.2.1 Geometric undistortion

One of the limitations of CRT technology is that the image becomes distorted at high frame rates. The corrective options offered by the built-in monitor controls are too qualitative and too coarse to ensure rectilinear pixels. At 180 Hz, the display usually shows significant pincushion distortion. I have adapted the Bankslab calibration routine to undistort the display. To accomplish this we map the display distortions at 1 cm resolution. Then, during runtime of the software, we capture each completed frame in the back-buffer, and distort the pixel loci with the inverse of the display distortions. This ensures that the monitor distortion does not produce distorted images.

A.2.2 Gamma calibration

In order for depth filtering to work, intensity of the light between adjacent image planes must be distributed according to the depth-filtering algorithm. By default, the computer converts software intensity values to perceptual units via the gamma curve. ($L_{out} + C = L_{in}^{\gamma}$, where L_{out} is the screen luminance, L_{in} is the programmed luminance, γ is the gamma constant, and C is the black level) The purpose of this correction is to adjust the display so that the perceptual difference between a L_{in} value 0 and 0.1 is approximately the same as 0.9 and 1.0. In a physical metric like cd/m^2, 0 to 0.1 might be 1 to 4 cd/m^2, and 0.9 to 1 might be 80 to 100 cd/m^2. Typical γ settings are typically between 1.8 and 2.2.

For the display application, we do not want linear pixel values to yield linear brightness steps, but instead linear luminance steps. The gamma calibration linearizes the gamma response so that specifying an intensity of 0.5 produces half the luminance as an intensity of 1 and twice the luminance as 0.25. This calibration is done by changing the color-look-up-table (CLUT) in the graphics card.

A.3 Software calibration

The optics hardware has a number of undesirable properties that I addressed with software corrections. One of these issues is that lens system was in front of the eye. This means that power changes did occur in the pupil plane and therefore introduced magnification errors. Furthermore, if the eye is not exactly lined up with the lens system, the lenses can introduce a prismatic shift. These problems can all be corrected with software.

A.3.1 Vergence calibration

I used the prism and haploscope arm adjustments to get the monitors into roughly the center of the display system field of view. Then I used a bite-bar mounted far from the display to align my eyes directly behind the lens apertures. I then presented objects on the screen that were directly in front of each eye, and adjusted the software frusta such that the objects appeared in the center of the lens apertures. This ensured that when an object was programmed to appear at optical infinity that the eyes would be in parallel gaze. This step corrected for any miscalibration of the hardware vergence state and also corrected for any vertical misalignment. This procedure for eye alignment was taken from Akeley et al. (2004)[1].

A.3.2 Size calibration

Changing the focal power of the lenses changes the image size magnification by several percent. To calibrate for absolute object size, we do a utrocular calibration procedure. The lens from one side of the display is removed and a real object is set up at a comfortable viewing distance. If this object's distance, and size are measured, we can compute its visual angle. Then we adjust the size of a computer-generated image in the other eye's image so that it has the same (known) size. Then we reverse the process to calibrate the other eye's size. We only need to do this for one focal state as in the next step we will calibrate the other focal states with respect to the one calibrated focal state.

In addition to adjusting the angular size of objects in this stage, this also calibrates the disparity magnitudes. By ensuring that angular size of an object is calibrated, we have ensured that angular disparity measurements will also be correct.

A.3.3 Depth plane alignment

Once the size and position for one of the image planes is properly adjusted, the other planes can be calibrated in reference to the adjusted one. I directly adapted this procedure from Kurt Akeley's technique for the 3-mirrors display[1]. The procedure involves vernier alignment of lines; a reference line is on the calibrated plane, and a test line is adjusted on the other plane. The alignment is performed for central targets, to adjust for prismatic shift, and eccentric targets, to adjust for magnification.

An advantage of this technique, pointed out by Akeley, is that this procedure minimizes misalignment wherever the eye fixates. Because the nodal point of the eye is slightly anterior of the rotational center, the eye's lens translates by several millimeters during an eye rotation. Calibrating in multiple positions ensures that the image should be well-aligned at the fovea for any eye rotation, even if there are errors in the periphery.

Appendix B

Display choices and compatibility with time multiplexed volumetric display

Choosing an appropriate 2D monitor is critical when designing a volumetric display. The quality of the monitor will partially determine ghosting artifacts, refresh rates and resolution. I considered a number of different technologies and ultimately decided that the CRT was most appropriate for this first attempt at constructing the volumetric display. This chapter briefly describes the main display choices availalble and some of their strengths and weaknesses.

B.1 Cathode ray tubes (CRT)

Ultimately, I used an Iiyama CRT when constructing the volumetric display. It is a high end CRT with a maximum horizontal scan frequency of 136 kHz, which is nearly double that of a typical CRT. At 800x600, this display can support a vertical refresh rate of over 200 Hz. The maximum refresh rate will decline to under 180 Hz when the resolution is expanded to 1024x768 or higher. CRTs produce a high quality color image that is bright and unpolarized.

CRTs raster an electron beam through the vacuum of the display and excite the phosphor grid at the front of the display. One of the most attractive features of a CRT for this type of application is that the graphics card controls the CRT directly; the graphics card outputs the data as the CRT rasters the data to the screen. Following completely drawing a frame, the graphics card sends a vertical-sync-pulse which causes the electron guns to retrace from the lower right to upper left corner of the CRT. This TTL pulse is a useful synchronization signal for the lens controller boxes.

Another important property of the CRT is that the full image is drawn and fades very quickly. This type of rapid draw and decay is similar to a temporal impulse response. The

Feature	Ideal capability/ feature	Display attribute
Refresh rate	240 Hz	flicker and smooth motion
Image persistence	<< 5ms	temporal ghosting
Image brightness	400-800 cd/m^2	brightness
Light polarization	Unpolarized	chromatic effects
Screen blanking	>1ms	temporal ghosting
Image format	Direct view	compact system
Frame buffer	TTL synch pulse	hardware synchronization
Gamma response	$\gamma = 1$	optimize bit depth
Rectilinear pixels	Lithographic	geometric distortion
Color	Simultaneous	frame rate
Resolution	>1024x768	detail

Table B.1: The main characteristics that are relevant for the display. The desired values are for a system that has 4 or fewer image planes. With more image planes, the system would need to be faster and brighter.

frame is nearly fully extinguished before the next frame is drawn (frame blanking). However, not all of the phosophors fade at equal rates.

Some of the most serious problems with CRTs are that 1) they are becoming very hard to acquire, and high performance CRTs are no longer being manufactured. 2) The technology has reached the point where it is limited by fundamental principles. The phosphors are not fast enough or bright enough to support higher frame rates. 3) Operating CRTs at high frame rates can cause considerable distortion in the images.

B.2 Liquid crystal display (LCD)

This class of display is commonly used for desktop and laptop displays. It has gained much of its popularity because of its thin form factor compared with the CRT. The LCD panel is made of a series of stages. An illuminator provides a white light source to the panel, which is made up of three layers. The outer layers of the panel are linear polarizers, often oriented so that they are orthogonal to one another. They sandwich the pixel layer. They pixels are made of a mosaicing of colored filters and liquid crystal cells that can rotate the polarization of the light passing through them. Thus when a pixel is on, the liquid crystal cell rotates the incident polarized light by 90 degrees so that it passes through the second polarizing filter. An off-pixel does not rotate the light polarization and the two orthogonal polarizing filters extinguish the light.

This construction gives the LCD panel some advantages. Most importantly, they are emitting polarized light, which, if properly oriented with our system, could reduce the light attenuation of our optics assemblies by 50%. (In this case the light is being lost in the LCD

panel instead of in the lens assemblies.)

The key disadvantage of these displays is that they present images using "sample and hold" as opposed to an impulse response and thus normally require and use a lower refresh rate. The pixels of the LCD are addressed row by row, so that a new image sweeps across the display from top to bottom. The sweep of the image takes almost a complete frame interval. Making matters worse, the LCD pixels have a response time that can be as high as several milliseconds. This means that there is no instant in time in which the entire image is simultaneously in a steady state on the screen[144]. This also means that there will be no instant in time where the screen is fully blank, and the lens system could switch without causing ghosting. Some of these timing issues can be resolved through the use of clever backlighting schemes. One idea is to use a pulsed LED backlight to present the image at high intensity briefly.

There is a renewed engineering effort to redesign these LCD panels so they will be compatible with field sequential stereo and support higher frame rates for better motion reproduction (Although, some of the advertised 240 Hz displays do not have the input bandwidth to accept a 240Hz frame rate. Instead they read a 30 Hz or 60 Hz frame rate and interpolate the frames needed to present smooth motion at 240 Hz). Currently, there are no consumer LCD displays that could achieve a frame rate approaching 200Hz needed for a 4-plane flicker-free volumetric display. Furthermore, these displays use internal frame buffers with their own timing protocols. These protocols are generally inaccessible in the consumer product and it will be a challenge to ensure that the display and lens system are properly synchronized.

B.3 Digital light projection (DLP)

These displays rely on digital micromirror devices (DMD) and have found widespread use in projectors, both for home, business and cinema use. Their pixel grid is made of an array of individually addressable microscopic mirrors. These mirrors can be rapidly switched between reflecting a light source to the projection lens or to a light dump. The mirror is inherently a boolean device, either "on" or "off," but the displays use pulse-width modulation to achieve a full grayscale image. In this modulation, the mirrors are switched on for a fraction of the full frame time corresponding with their desired intensity. At the frame rates used, this technique is perceptually indistinguishable from an analog control over light intensity. There are two dominant features of this technology that would be valuable in the design of a volumetric display. 1) It would be easy to adapt this hardware to have a linearized gamma response which is important for depth filtering to be effective. (see Section A.2.2) 2) It would be a relatively minor change to rewrite the pulse width modulation protocols to include a short blank interval between frames where we could switch the lens.

There are two design implementations of this technology: the single-chip and 3-chip displays. In the single chip design, color is achieved through temporal multiplexing using a

color wheel. The wheel filters the light source to one of the color primaries before it reaches the DMD mirror chip which reflects the monochromatic component of the image through the lens to the display screen. The color wheel cycles to the next color position and the process is repeated to build the full image in color. Alternatively, the 3-chip DLP projectors use three colored light sources and three DMD chips. It superimposes the images for projection, so that the color image is transmitted simultaneously. The three chip projectors operate at 144 Hz and are normally found in cinema. The 1-chip projectors are an inexpensive consumer electronic device and typically operate at 100-200 Hz.

In principle, this technology could be well suited for volumetric displays. With the development kit, we could control the pulse-width modulation to insert a blank period between the frames and linearize the gamma in the projector. These changes would make the best use of the bit depth and would also minimize temporal ghosting. With some minor modifications, this system could even be used as a direct view display, where the observer was viewing the DMD chip directly.

We chose not to use these displays because neither the single-chip or 3-chip would out-of-the-box be ready to use in our display system. According to RealD, one of the main vendor for these 3-chip projector systems, the projectors only have an input bandwidth of approximately 66 Hz (which they would typically split between the stereo channels, and then would use multi-flash strategies to present at higher frame rates). These projectors are also enormous devices with tremendous light output and power demands. They would not be practical for use in a single viewer display.

The consumer class of DLP display could be modified to work for the volumetric display as a grayscale display. There would nevertheless remain technical challenges. This option would require purchasing the development kit for the DMD and building the housing necessary to use it as a display, but it should be possible to achieve grayscale refresh rates of 500 Hz.

B.4 Liquid crystal on silicon (LCOS)

These displays use liquid crystal technology in a similar way to LCD displays except that they are reflective instead of transmissive. Because the chips use silicon as their substrate, they can be much faster than LCD panels and produce higher quality images. The displays create color in the same ways as DLP, using either a single chip with a color wheel or a 3-chip design. They achieve grayscale modulation continuously by modulating the polarization rotation by the LC element.

These displays are advantageous because they support high frame rates and do not require pulse width modulation for brightness. The 3-chip projectors are also designed for consumers and are far more compact with smaller light sources. These displays would have the same problems with synchronization as the DLP and high end LCD displays.

Despite the similarity and LCOS and DLP, LCOS has two major disadvantages. The

most serious disadvantage of these displays is that they emit weakly polarized light. If the light were entirely polarized, as in LCD displays, we could use this initial polarization state as part of our optical design. If the light is entirely unpolarized, like in CRTs or DLPs, we can insert a linear polarizer to polarize the light at the beginning of our lens system. The weakly polarized light is incompatible with both approaches. Passing weakly polarized light through a linear polarizer causes chromatic and spatial irregularities in brightness throughout the image. The only way to effectively use the light in our system would be to project it onto a surface to break up the polarization. The second disadvantage is that it does not have a blank interval between frames. In order to eliminate ghosting during frame changes in the display, we would likeley need to modulate the light source, possibly via LEDs. A switchable illumination source could allow us to insert a blank interval between frames.

B.5 Organic light emitting diode (OLED)

This technology is very exciting because the pixels themselves are the light source. Whereas the other digital approaches modulate light that falls onto their pixel elements, OLED display pixels emit the light directly. This makes these displays well suited for a spatially multiplexed display, because they can transmit light, and emit light. LEDs also have very fast switch times, which would make them desirable for the temporal multiplexing of focus cues.

Unfortunately, at this time, these displays are not commercially available. It is a promising technology, but remains expensive and has a limited longevity. As these displays are developed for the consumer display market, they may also become a strong candidate for the volumetric display application.

Display Parameter	CRT	LCD	DLP	LCOS	OLED
Volumetric refresh rate	O	–	X	X	X
Image persistence	O	–	X	X	X
Image brightness	–	X	X	X	?
Light polarization	X	–	X	–	X
Screen blanking	O	–	Δ	?	X
Image format	O	O	Δ	Δ	Δ
Frame buffer	X	–	Δ	Δ	?
Gamma response	–	O	X	?	X
Rectilinear pixels	–	X	X	X	X
Resolution	–	X	X	X	?

Table B.2: Display effectiveness. This table shows how effectively the different display technologies are at achieving the requirements specified in Table B.1. The symbol 'X' denotes that it well satisfies the requirement, 'O' indicates that is somewhat adequate, Δ indicates that modification will be needed, '–' indicates it is inadequate, and '?' indicates that we do not yet know.

Appendix C

A breakdown in image plane geometry

Throughout the thesis, I have discussed many of the benefits of a time-multiplexed volumetric display. However, one disadvantage of this approach is that they have an undesirable geometric property: image "planes" can become warped. This error is mostly a theoretical problem, but nonetheless it is an important issue to address. When the plane of a display is optically shifted to another focal distance, the focal characteristics of that plane will cease to be consistent with a plane.

C.1 Focal warping of virtual-image-plans

Until now, I have primarily described the focal distances for point objects, but now I will consider the focal distance for fronto-normal planes. Viewed from the top, two such planes are visible to one eye in Figure C.1.

Consider a situation with two image planes, the near one is at 20cm, and the farther one is at 40cm. On each of these image planes, we would like to present a frontoparallel plane that subtends 30° of visual angle. The distance to a central position in that image plane (green dot, in Figure C.1A) will be equal to the nominal image plane distance, and thus also equal to the simulated plane distance. At eccentric positions on the plane (red dots) the distance from the eye's nodal point to the surface will be slightly further. However, because the simulated plane coincides with the real image plane, the red point still has a focal depth that is consistent with it lying on the plane. Panel A shows that for two simulated planes that fall on their appropriate image plane, all points in the plane will have the appropriate focal distance.

In Panel B, I show a display that uses a lens to simulate the far stimulus plane using the near image plane. Along the central axis, this technique works perfectly. The stimulus which is shown at a 5D distance is optically shifted to 2.5D by using a 2.5D lens. However, the eccentric point on the plane has a focal distance of 4.83D on the image plane. An ideal lens will shift this point to 2.33D. The 2.5D lens brings it approximately to the distance of

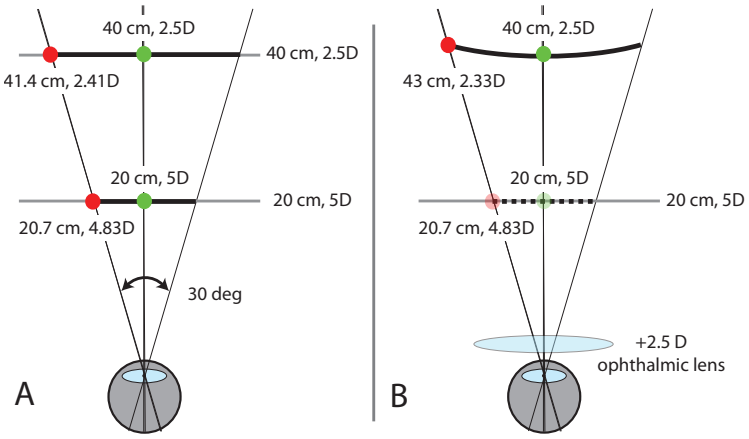


Figure C.1: Dioptric variation of fronto-parallel planes. Image planes are shown as horizontal gray bars, and frontoparallel stimulus planes are shown in heavy black. The green points are axial points lying on those planes, and the red dots are on the plane but at 15 deg eccentricity.

the plane at 40cm but not exactly. Its focal distance will be 0.08D farther than where it should be. The focal distance of the eccentric points of this plane no longer fall in a plane, but instead resemble a bowl.

This issue is a problem with the geometry of perspective projection to a plane in a system where the image planes are translated optically. In the "3-mirrors" display, described in Section 1.2.1, the image planes are spatially multiplexed, and have different optical distances to each point on the image plane, and thus the image planes are truly planes. This issue arises only in situations where a optical element shifts the position of the image plane so that its virtual distance no longer corresponds to its physical distance.

I carried out this analysis for several different image plane distances in Figure C.2. The abscissa represents eccentric location, and the ordinate represents distance in front of the eye. Panels A, B, and C represent physical display distances of 20 cm, 50 cm, and 200 cm respectively. With a near display, image planes at far distances become convex. When the display is far (as in Panel C), the near image planes are concave. The only positions that have "exactly" correct focal distances are the locations that are directly in front of the eye, and at an image plane that is at the same distance as the physical display.

C.2 Non-axial lens focal power

Radial astigmatism is another source of the error in focal distances in a display utilizing an optical element to shift image plane positions. The refractive power of a lens differs for off-axis viewing, and acquires a new focal power along both of its meridians relative to the new viewing angle. Light in the plane containing both the chief ray (principal ray), and

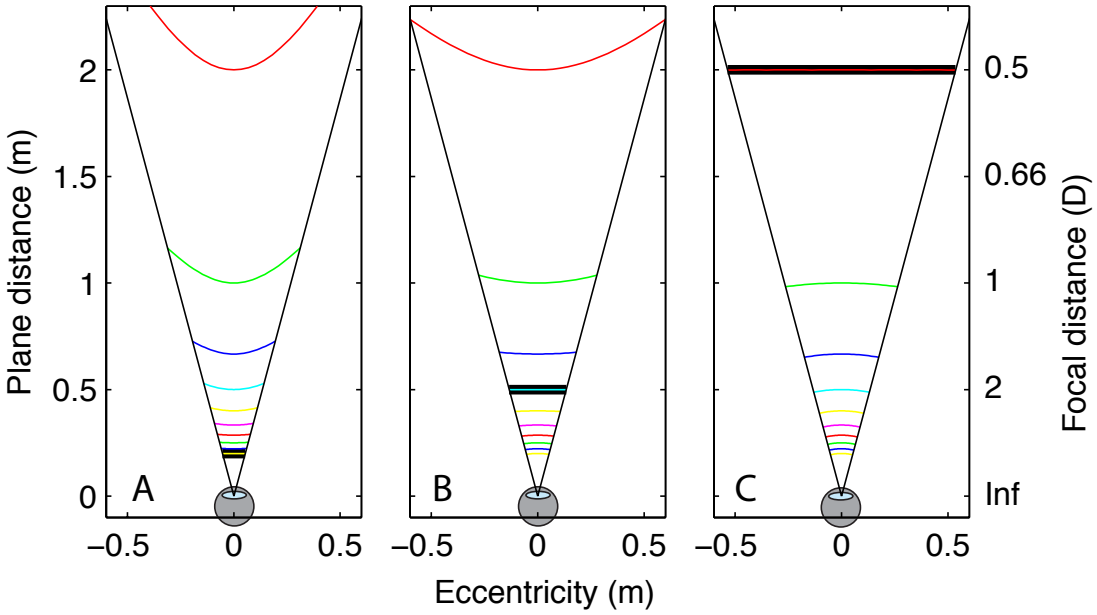


Figure C.2: Curvature of virtual image planes. The heavy black lines represent the position in space of the real image plane. The thin colored lines represent various virtual image planes created via lenses. Panel A: The real image plane is at 5D. Panel B: The real image plane is at 2D. Panel C: The real image plane is at 0.5D.

the optic axis of the lens is known as the tangential plane. The orthoganal plane that also contains the chief ray is the sagittal plane. The refractive powers of a lens of power P for an angle of ϕ, will be P_t in the tangential plane and P_s in the sagittal plane.

$$P_t = P\frac{2n + \sin^2\phi}{2n\cos^2\phi}; \qquad P_s = P(1 + \frac{\sin^2\phi}{2n}) \tag{C.1}$$

The new power of the lens is often described by its spherical equivalent power, which is the average of P_t and P_s.

The radial astigmatism from eccentric viewing will have a similar effect as the changes in the optical distance. This distortion of focal depth will stack with the path-length distortion to create a larger deviation from a plane's focal properties.

C.3 Conclusions

This problem may be academic in nature and not a serious source of concern. The image-plane design holds up nicely to other depth cues, including stereo and perspective information. Nevertheless, this issue demands serious consideration in the development of a head mountable display. In this type of display, the screen must be worn very close to the observer with lenses used to optically shift the image planes away from the viewer. Both the near distance of the display and the strong lenses will contribute to distorting the focal information at normal distances. If the display were attempting to simulate optical infinity, eccentric regions of the image would be myopic.

I extended the analysis of path-length differences to also include the influence of a thin lens. I have included this information in Table C.1. In this table, I computed the deviation of the focal distance for eccentric positions in a display. For up to a 15° azimuth, the deviations from normal focal distance remains minor when the simulated image plane is within 2D of the physical display distance. The deviations in this case remain below the typical threshold for blur discrimination (0.33 D).

In the situation of a head-mounted display, the compact system requires a powerful lens and a short distance to the display. In this case, (the lower panel of Table C.1), there are significant deviations from correct focal distances. This could lead to unacceptable focus cue artifacts. In addition to problems with the inappropriate blur gradient, the display would present inconsistent vergence–accommodation stimuli. If an object was moved closer to the observer, the vergence specified distance could be appropriately stimulated, however, the focal stimulus could be incorrect. The vergence stimulus would also change the monocular azimuths of the object, which could cause focus information to indicate that the object was moving further.

One way of dealing with these incorrect focus signals would be adapting the software to consider eccentricity in the depth-filtering computations. Another approach would be to

Angle	Display dist.	Image plane distance = 33 cm, 3D				
		$F_{desired}$	$F_{pathdistortion}$	P_{lens}	$F_{effective}$	deviation
0	5 D	3.00 D	3.00 D	2.00 D	3.00 D	0.00 D
5	5 D	2.99 D	2.98 D	2.01 D	2.97 D	0.02 D
10	5 D	2.95 D	2.92 D	2.05 D	2.87 D	0.08 D
15	5 D	2.90 D	2.82 D	2.12 D	2.71 D	0.19 D
20	5 D	2.82 D	2.70 D	2.22 D	2.48 D	0.34 D
0	1 D	3.00 D	3.00 D	–2.00 D	3.00 D	– 0.00 D
5	1 D	2.99 D	2.99 D	–2.01 D	3.01 D	–0.02 D
10	1 D	2.95 D	2.98 D	–2.05 D	3.04 D	–0.08 D
15	1 D	2.90 D	2.97 D	–2.12 D	3.08 D	–0.19 D
20	1 D	2.82 D	2.94 D	–2.22 D	3.16 D	–0.34 D
0	10 D	3.00 D	3.00 D	7.00 D	3.00 D	0.00 D
5	10 D	2.99 D	2.96 D	7.04 D	2.92 D	0.07 D
10	10 D	2.95 D	2.85 D	7.18 D	2.66 D	0.29 D
15	10 D	2.90 D	2.66 D	7.41 D	2.25 D	0.65 D
20	10 D	2.82 D	2.40 D	7.75 D	1.64 D	1.18 D

Table C.1: Desired vs. achieved focal distances for various display distances using a lens to achieve a 3.0 D (33cm) virtual-image plane. The columns from right to left represent: viewing angle eccentricity, the distance to the physical display, the correct focal distance for an object at that eccentricity at the image plane, the focal distance of that point with a lens system free of radial astigmatism, the spherical equivalent power of the appropriate thin lens, the focal distance of the object, and the deviation of the effective focal distance from the correct focal distance. Each panel represents a different display distance. The lower panel is a display distance consistent with use in a head–mounted display.

use a display system with a curved, or multifaceted, surface to reduce the dioptric range of distances present on the screen. Any approach would require careful engineering of the optics to minimize these undesirable effects while maximizing the display's field of view.