

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Perceptual Considerations for the Design and Display of Stereoscopic Imagery

**Permalink**

<https://escholarship.org/uc/item/9wm0v9qp>

**Author**

Held, Robert Thomas

**Publication Date**

2010

Peer reviewed|Thesis/dissertation

Perceptual Considerations for the  
Design and Display of Stereoscopic Imagery

by

Robert Thomas Held

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Copyright 2010  
by  
Robert Thomas Held

# ACKNOWLEDGMENTS

## **Family, Friends, and Mentors**

I owe a great deal to family, friends, and mentors. I want to thank Casey for her willingness to uproot and move, her constant, loving support, and her “real job,” which made it easier to live off a grad student salary and justified my references to her as my sugar mama. I am also very thankful to my parents, who have always been encouraging and provided an upbringing that made it easy to follow my interests and achieve my goals.

I’ve learned from some great people at Berkeley. Steve Conolly taught me how to teach and showed me that it’s important to take an active role in your department. Maneesh Agrawala, Rich Breiman, Pratik Mukherjee, James O’Brien, Austin Roorda, and Dan Vigneron put a lot of effort into my graduate education. They were always happy to meet up in their offices or in a cafe to discuss my research, give tips and suggestions, and perhaps most importantly, ask me questions to make sure I knew what I was doing and getting the most out of my education.

The Banks lab has been a wonderful place to be a grad student. The mixture of excellent research, intelligence, and great senses of humor made it so I didn’t want to be anywhere else. On projects, I worked closest with Johannes Burge, Chris Burns, Emily Cooper, Ahna Girshick, and David Hoffman and benefited greatly from their insights (and banter). Björn Vlaskamp, the undisputed favorite member of the lab, was always available to talk about research, idioms, or gadgets, and I cherished our regular trips to the CD store to buy and share music.

Finally, I want to thank Marty. As a mentor, he knows how to teach, challenge, compliment, and joke around. Most importantly, he knows how to balance all those things effectively. So whether I needed help with a research problem or wanted to talk about something else in my life, his door was always open.

## **Previously Published Material**

Most of the material in Chapter 1 already appeared in an article published in the Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization [Held and Banks, 2008]. The coauthor listed in this publication (Martin Banks) directed and supervised the research that forms the basis for Chapter 1. Similarly, most of the material in 2 was published in a manuscript that appeared in ACM Transactions on Graphics [Held et al., 2010]. The last author (Martin Banks) was the primary advisor on the project, with the tertiary author (James O'Brien) also providing direction and supervision. The following statement from the primary advisor summarizes my contribution to the project:

“Robert was the lead in designing the psychophysical task, ran most of the subjects, and processed the data. He also wrote all of the stimulus-generation and experiment code. The first draft of the manuscript was completed by him, then revised and augmented by the other three authors. The work is equal in scope to a dissertation chapter, as it began with a poorly understood psychological phenomenon, developed a perceptual model, and tested its predictions with psychophysical data.” - Martin S. Banks

# Perceptual Considerations for the Design and Display of Stereoscopic Imagery

Stereoscopic displays afford more accurate 3D percepts than conventional displays due to the added depth cue of disparity. However, 3D shape and scene layout are often misperceived when viewing stereoscopic displays. For example, viewing from the wrong distance alters an object's perceived size and shape. It is crucial to understand the causes of such misperceptions so one can determine the best approaches for minimizing them. We develop the mathematics of an existing geometric model for calculating misperceptions, and then describe common viewing situations in which the model fails to make a prediction. We show how the visual system's interpretation of vertical disparities can supplement the existing model and help predict the percepts associated with improper viewing of stereoscopic displays.

We also discuss blur as a previously under-appreciated depth cue present in both stereo and non-stereo images. We present a probabilistic model that explains how the pattern of blur in an image together with relative depth cues indicates the apparent scale of the image's contents. To examine the correspondence between the model/algorithm and actual viewer experience, we conducted an experiment with human viewers and compared their estimates of absolute distance to the model's predictions. We did this for images with geometrically correct blur due to defocus and for images with commonly used approximations to the correct blur. The agreement between the experimental data and model predictions was excellent. A semi-automated algorithm is included, which helps one apply the correct pattern of blur to change the apparent size of a scene. The algorithm and model allow one to manipulate blur precisely and to achieve the desired perceived scale efficiently.

Finally, we discuss the utility of stereoscopic displays for medical imaging. The technology's greatest benefits arise in applications in which monocular cues are uninformative. Its specific strengths and shortcomings are presented in terms of diagnostics, education, surgi-

cal planning, minimally invasive surgery, and telesurgery. General guidelines and common errors are also listed to help potential users avoid unwanted misperceptions and visual fatigue.

# CONTENTS

<b>Acknowledgments</b>	<b>iii</b>
Family, Friends, and Mentors . . . . .	iii
Previously Published Material . . . . .	iv
<b>Abstract</b>	<b>v</b>
<b>Introduction</b>	<b>1</b>
<b>1 Misperceptions in Stereoscopic Displays</b>	<b>3</b>
1.1 Related Work . . . . .	4
1.2 Predicted Distortions: Geometric Approach . . . . .	8
1.3 Failures of the Geometric Approach . . . . .	11
1.3.1 Epipolar Planes . . . . .	12
1.3.2 Previous Solutions to Skew Rays . . . . .	13
1.3.3 Vertical Disparity . . . . .	14
1.3.4 Perception with Skew Rays Present . . . . .	16
1.4 Future Work in Stereoscopic Misperceptions . . . . .	21
<b>2 Using Blur to Affect Perceived Distance and Size</b>	<b>22</b>
2.1 Background . . . . .	25
2.1.1 Defocus Blur in Computer Graphics and Photography . . . . .	25
2.1.2 Aperture and Blur . . . . .	27
2.2 Adjusting Blur to Modulate Perceived Distance and Size . . . . .	28
2.2.1 Tilt-and-shift Lenses and Linear Blur Gradients . . . . .	28
2.2.2 Comparing Blur Patterns . . . . .	30
2.3 Model: Blur as an Absolute Depth Cue . . . . .	33
2.3.1 Vision Science Literature . . . . .	33
2.3.2 Probabilistic Inference of Distance from Blur . . . . .	33
2.3.3 Impact on Previous Findings . . . . .	36
2.3.4 Perspective Cues . . . . .	37
2.3.5 Recovering Focal Distance in Practice . . . . .	38
2.4 Estimating Distance in Images with Manipulated Blur . . . . .	39
2.4.1 Applying Blur Patterns . . . . .	39
2.4.2 Calculating Best Fits to the Image Data . . . . .	41
2.4.3 Predictions of the Model . . . . .	42
2.4.4 Algorithm . . . . .	44
2.5 Psychophysical Experiment . . . . .	47
2.5.1 Methods . . . . .	48
2.5.2 Results . . . . .	48
2.6 Discussion . . . . .	50
2.6.1 Validity of Assumptions in the Model . . . . .	50
2.6.2 Algorithm Effectiveness . . . . .	52
2.6.3 Impact on Computer Graphics . . . . .	52



2.7	Future Work Investigating Blur as a Distance Cue . . . . .	58
<b>3</b>	<b>Stereoscopic Displays in Medical Imaging: A Review</b>	<b>59</b>
3.1	Depth and Displays . . . . .	59
3.1.1	Monocular Depth Cues . . . . .	60
3.1.2	Stereoscopic Depth . . . . .	61
3.2	Medical Applications of Stereoscopic Displays . . . . .	62
3.2.1	Diagnostics . . . . .	62
3.2.2	Medical Training . . . . .	65
3.2.3	Surgical Planning . . . . .	69
3.2.4	Laparoscopy . . . . .	70
3.2.5	Telesurgery . . . . .	73
3.2.6	Augmented-reality Surgery . . . . .	73
3.3	Summary of Benefits . . . . .	75
3.4	Summary of Drawbacks . . . . .	77
3.4.1	Hardware . . . . .	77
3.4.2	Resistance to New Technology . . . . .	78
3.4.3	Viewer Discomfort . . . . .	78
3.4.4	Stereoscopic Misperceptions . . . . .	79
3.5	Avoiding Misperceptions with Stereoscopic Displays . . . . .	81
3.6	Suggestions and Guidelines . . . . .	83
3.6.1	Suggestion 1: Parallel cameras . . . . .	84
3.6.2	Suggestion 2: Do not flip images . . . . .	84
3.6.3	Suggestion 3: Keep eyes centered and parallel relative to display . . . . .	85
3.6.4	Suggestion 4: Minimize vergence-accommodation conflict . . . . .	85
3.6.5	Suggestion 5: Appropriate use of pictorial blur . . . . .	86
<b>A</b>	<b>Skew-Ray Geometry</b>	<b>101</b>
<b>B</b>	<b>Data from Blur Experiment</b>	<b>107</b>

# LIST OF FIGURES

1.1	Geometry of stereoscopic image capture and display . . . . .	4
1.2	Example misperceptions for several stereoscopic acquisition and viewing situations . . . . .	9
1.3	Epipolar geometry of stereoscopic misperceptions . . . . .	13
1.4	The keystone effect produced by converging camera bodies . . . . .	14
1.5	Retinal and extra-ocular signals that can be useful for slant estimation . . .	15
1.6	Anaglyph stereograms captured using simulated cameras with parallel and converging optical axes . . . . .	17
1.7	Disparity fields for common viewing situations that produce skew rays . . .	19
2.1	Example urban images before and after application of tilt-shift blur . . . . .	23
2.2	Example urban and landscape images with and without added blur . . . . .	24
2.3	Schematic of blur in a simple imaging system . . . . .	28
2.4	Schematic of the Scheimpflug Principle . . . . .	29
2.5	Comparison of blur patterns produced by three rendering techniques: consistent blur, simulated tilt-and-shift lens, and linear blur gradient . . . . .	30
2.6	Possible combinations of focal distance and relative distance for example magnitudes of retinal-image blur . . . . .	34
2.7	Bayesian analysis of blur as cue to absolute distance . . . . .	35
2.8	Example GoogleEarth images with consistent, vertical-gradient (mostly consistent), and horizontal-gradient (inconsistent) blur patterns . . . . .	40
2.9	Model-based estimates of focal distances for GoogleEarth scene based on consistent, vertical-gradient (mostly consistent), and horizontal-gradient (inconsistent) blur pattern . . . . .	41
2.10	Schematic of variables pertinent to semi-automated blurring algorithm . . .	44
2.11	Input and output stages of semi-automated blurring algorithm . . . . .	47

2.12	Results of psychophysical blur experiment. . . . .	49
3.1	Random-dot stereogram with floating patches. . . . .	63
3.2	Endoscopic images during bowel surgery on an animal model . . . . .	71
3.3	Vergence-accommodation conflict in typical stereoscopic displays . . . . .	80
A.1	Diagram of vectors used to derive skew-ray geometry . . . . .	102
B.1	Individual subject data for blur experiment . . . . .	107

# INTRODUCTION

Stereoscopic displays have become commonplace as they have made their way into cinema [Lipton, 1982], medical imaging [Chan et al., 2005], and scientific visualization [Fröhlich et al., 1999]. Regardless of application, all stereo displays operate on the same basic mechanism: a unique image is presented to each eye. The differences between the images (known as a stereo pair) are interpreted by the visual system as depth information in a process known as stereopsis. The relative positions of an object’s projections onto the two retinas are used by the visual system to recover the distance to that object (see Chapter 1). Stereopsis is not the only source of depth information in typical images. In fact, the name “3D display” is a misnomer because displays have always contained three-dimensional information. As we discuss in Chapter 3, the key distinction in “3D displays” is that they provide stereoscopic depth information in addition to the monocular depth cues attainable with any display. Therefore, for the remainder of this dissertation we will use the name “stereoscopic display” or “stereo display” in place of the more common “3D display.”

As the use of stereoscopic displays has spread, the benefits and problems associated with them have become clearer. A well-documented problem is that perceived 3D shape and scene layout is often distorted. For instance, viewing a stereoscopic display from the wrong distance typically alters the perceived size and shape of an object [Masaoka et al., 2006; Woods et al., 1993]. In some applications, such as cinema, the distortions are not necessarily a serious problem [Lipton, 1982], but in applications like medical imaging or virtual reality [Deering, 1992], they can have grave consequences. In Chapter 1, we examine the causes of perceptual distortions in stereography and show that in some cases the standard model of such distortions is incorrect. We then describe how the model can be modified to include the visual system’s usage of vertical disparities to produce a more comprehensive understanding of the percepts that result from improper viewing conditions.

In addition to stereoscopic cues, we consider in Chapter 2 how the monocular cue of blur

can affect the perceived size and distance of imaged scenes. For example, cinematographers working with miniature models can make scenes appear life size by using a small camera aperture, which reduces the blur variation between objects at different distances [Fielding, 1985]. The opposite effect is created in a photographic manipulation known as the tilt-shift effect: A full-size scene is made to look smaller by adding blur with either a special lens or post-processing software tools [Laforet, 2007; Flickr, 2009; Vishwanath, 2008]. The strengths of these effects are in contrast to the findings of previous vision-science studies, which mostly concluded that blur was only a weak, ordinal cue to depth. We present a model for how the blur pattern in a scene can be combined with other depth cues to produce reliable estimates of size and distance. An algorithm is also included for semi-automatically applying blur to an image to change its apparent scale.

To provide a context for the models we develop in Chapters 1 and 2, Chapter 3 outlines how stereoscopic displays have been applied to medical imaging. Following a review of previous implementations and their outcomes, a summary is given of the strengths and weaknesses of stereoscopic medical displays. The model of stereoscopic misperceptions from Chapter 1 is used to explain how to achieve optimal viewing. Also, the blur-rendering techniques from Chapter 2 are revisited to discuss whether they are necessary for medical applications. Tips on avoiding visual fatigue are included to round out the set of guidelines.

# CHAPTER 1

## MISPERCEPTIONS IN STEREOSCOPIC DISPLAYS

It is useful to consider a stereoscopic percept as the product of three steps. (1) Images are acquired by stereo photography or generated by computer graphics. (2) The images are presented stereoscopically to a viewer. (3) The images are interpreted by the viewer's visual system. Geometric misperceptions are caused by inappropriate acquisition-viewing relationships (steps 1 & 2) such that the retinal images are not the same as those produced by the original scene. Perceptual misperceptions are produced by the viewer's visual system (step 3): the retinal images may each be geometrically correct, but visual cues such as vergence and accommodation cause them to be misinterpreted nonetheless [Watt et al., 2005]. The graphics, stereocinema, and virtual-reality literatures [Diner, 1991; Jones et al., 2001; Kusaka, 1992; Kutka, 1994; Leiser et al., 1995; Lipton, 1982; Masaoka et al., 2006; Son et al., 2002; Wartell et al., 2002; Woods et al., 1993; Yamanoue et al., 2006] have only used the geometric approach. We begin this chapter by developing the mathematics of the geometric approach and summarizing the predicted distortions. We then describe the limitations of the approach, especially in dealing with the vertical disparities produced by some viewing situations; these situations include rotation of the viewer's head relative to the display and using converging cameras to acquire images that are then viewed on a single-surface display. Finally, we describe an approach derived from vision science that provides a better characterization.

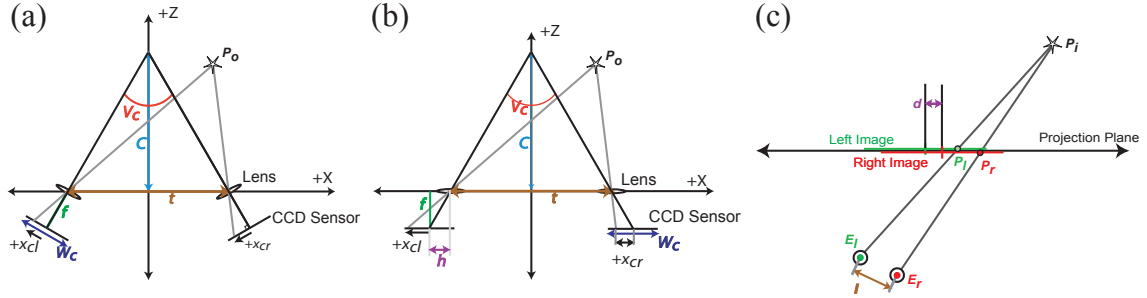


Fig. 1.1: (a) Image formation with converging cameras.  $\mathbf{P}_o$  is coordinates of point  $P$ ,  $f$  is camera focal length,  $t$  is separation between the cameras,  $C$  is distance to which the camera optical axes are converged,  $V_c$  is angle between cameras' optical axes,  $W_c$  is width of camera sensors,  $x_{cl}$  and  $x_{cr}$  are  $x$ -coordinates of  $P$ 's projection onto left and right camera sensors. (b) The cameras' optical axes can be made to converge by laterally offsetting the sensors relative to the lens axes.  $h$  is offset between sensor center and intersection of lens axis with the sensor. (c) Reconstruction of  $P$  from sensor images. Rays are projected from eye centers through corresponding points on picture. The ray intersection is estimated location of  $P$ .  $\mathbf{E}_l$  and  $\mathbf{E}_r$  are 3D coordinates of left and right eyes;  $\mathbf{P}_l$  and  $\mathbf{P}_r$  are locations of image points in the picture of  $P$  for left and right eyes;  $I$  is inter-ocular distance;  $d$  is distance between centers of pictures. The green and red horizontal lines represent the images presented to the left and right eyes, respectively.

## 1.1 Related Work

To describe the geometric approach, we use some derivations from Woods et al. [1993].

### Step 1: Acquisition (Object space to 2d camera sensors).

We first determine the 2D coordinates of a point in 3D space ( $P$ ) once projected onto the sensors of a pair of cameras. In 3D coordinates,  $X$  is the inter-camera axis,  $Y$  is the vertical axis perpendicular to the camera axis and running through the midpoint between the cameras, and  $Z$  is orthogonal to  $X$  and  $Y$ . The 3D coordinates of  $P$  are  $P_o$ . The cameras are specified by focal length  $f$ , sensor width  $W_c$ , and inter-camera separation  $t$ . Each camera has two axes: the lens axis, which bisects and is normal to the lens and is perpendicular to the image sensor plane, and the optical axis, which contains the lens center and sensor center. Camera alignment is specified by  $V_c$ , the angle between the cameras' optical axes, and by  $h$ , the displacement of the lens axis in the sensor plane. With the camera lenses parallel to one another, the optical axes can be parallel ( $h = 0, V_c = 0$ ) or

converging ( $h \neq 0, V_c \neq 0$ ) (Fig. 1.1A and 1.1B).  $P$ 's coordinates in the left and right cameras are  $(x_{cl}, y_{cl})$  and  $(x_{cr}, y_{cr})$ , where  $x$  and  $y$  are horizontal and vertical coordinates in the sensors:

$$\begin{aligned}
 x_{cl} &= f \tan \left[ \arctan \frac{t/2 + P_o(x)}{P_o(z)} - \frac{V_c}{2} \right] - h \\
 x_{cr} &= f \tan \left[ \frac{V_c}{2} - \arctan \frac{t/2 - P_o(x)}{P_o(z)} \right] + h \\
 y_{cl} &= \frac{P_o(y)f}{P_o(z)\cos\left(\frac{V_c}{2}\right) + (P_o(x) + \frac{t}{2})\sin\left(\frac{V_c}{2}\right)} \\
 y_{cr} &= \frac{P_o(y)f}{P_o(z)\cos\left(\frac{V_c}{2}\right) - (P_o(x) - \frac{t}{2})\sin\left(\frac{V_c}{2}\right)}
 \end{aligned} \tag{1.1}$$

**Step 2: Presentation (2d camera sensors to 2d projections).**

To present the stereo-camera images, the sensor coordinates  $(x_{cl}, y_{cl})$  and  $(x_{cr}, y_{cr})$  must be transformed to 2D picture coordinates  $(X_{sl}, Y_{sl})$  and  $(X_{sr}, Y_{sr})$ . In most applications, the pictures are presented on one display surface such as an LCD or projection screen. In vision science, they are often presented on two displays, one for each eye, in a device called a haploscope [Backus et al., 1999]. Single-surface displays are much more common, so we concentrate on them here. The pictures are characterized by their width  $W_p$ , and  $d$ , which is their horizontal displacement relative to one another. The ratio  $W_p/W_c$  is the magnification from the camera images to the picture. The 2D coordinates of corresponding points in the picture are:

$$\begin{aligned}
 X_{sl} &= x_{cl} \left( \frac{W_p}{W_c} \right) - \frac{d}{2} \\
 X_{sr} &= x_{cr} \left( \frac{W_p}{W_c} \right) + \frac{d}{2} \\
 Y_{sl} &= y_{cl} \left( \frac{W_p}{W_c} \right) \\
 Y_{sr} &= y_{cl} \left( \frac{W_p}{W_c} \right)
 \end{aligned}$$



### Step 3: Viewing (2d projections to percept).

The binocular viewer is positioned to view the pictures on the display surface. We use two new sets of 3D coordinates to describe this: one with its origin on the display surface and one with its origin at the viewer. For the first set,  $X$  and  $Y$  are the horizontal and vertical axes centered on the display surface and  $Z$  is orthogonal to them. In these coordinates, the eyes' positions are  $\mathbf{E}_l$  and  $\mathbf{E}_r$ . The positions of the points in picture are:

$$\mathbf{P}_l = (X_{sl}, Y_{sl}, 0)$$

$$\mathbf{P}_r = (X_{sr}, Y_{sr}, 0)$$

To determine the viewer's estimate of the location of  $P$  given  $\mathbf{E}_l$ ,  $\mathbf{E}_r$ ,  $\mathbf{P}_l$ , and  $\mathbf{P}_r$ , we project rays from the eye centers through the corresponding points in the picture. The estimated location of  $P$  is assigned to the point of intersection  $\mathbf{P}_i$  (Figure 1.1C). We want the location of  $\mathbf{P}_i$  specified in viewer coordinates, so we transform  $\mathbf{E}_l$ ,  $\mathbf{E}_r$ ,  $\mathbf{P}_l$ , and  $\mathbf{P}_r$  into  $\mathbf{E}'_l$ ,  $\mathbf{E}'_r$ ,  $\mathbf{P}'_l$ , and  $\mathbf{P}'_r$  in a viewer-centered system. There, the origin is midway between the eyes, which is  $\mathbf{E}_c$  in picture-centered coordinates;  $X$  is the inter-ocular axis,  $Y$  is the vertical axis, and  $Z$  is orthogonal to them.  $\mathbf{E}_c$  is subtracted from  $\mathbf{E}_l$ ,  $\mathbf{E}_r$ ,  $\mathbf{P}_l$ , and  $\mathbf{P}_r$ . The transformations are listed below:

$$\mathbf{E}'_l = R(\rho, \sigma)(\mathbf{E}_l - \mathbf{E}_c)$$

$$\mathbf{E}'_r = R(\rho, \sigma)(\mathbf{E}_r - \mathbf{E}_c)$$

$$\mathbf{P}'_l = R(\rho, \sigma)(\mathbf{P}_l - \mathbf{E}_c)$$

$$\mathbf{P}'_r = R(\rho, \sigma)(\mathbf{P}_r - \mathbf{E}_c)$$

where

$$R(\rho, \sigma) = \begin{bmatrix} \cos(\rho) \cos(\sigma) & \cos(\rho) \sin(\sigma) & -\sin(\rho) \\ -\sin(\sigma) & \cos(\sigma) & 0 \\ \sin(\rho) \cos(\sigma) & \sin(\rho) \sin(\sigma) & \cos(\rho) \end{bmatrix}$$

and

$$\rho = -\arctan\left(\frac{E_r(z) - E_l(z)}{E_r(x) - E_l(x)}\right)$$

$$\sigma = -\arcsin\left(\frac{E_r(y) - E_l(y)}{\|\mathbf{E}_r - \mathbf{E}_l\|}\right)$$

The intersection of rays originating at  $\mathbf{E}'_l$  and  $\mathbf{E}'_r$  and passing through  $\mathbf{P}'_l$  and  $\mathbf{P}'_r$  can then be found from:

$$\mathbf{E}'_l + (\mathbf{P}'_l - \mathbf{E}'_l)m = \mathbf{E}'_r + (\mathbf{P}'_r - \mathbf{E}'_r)n$$

$(\mathbf{P}'_l - \mathbf{E}'_l)m$  and  $(\mathbf{P}'_r - \mathbf{E}'_r)n$  represent the exiting rays;  $m$  and  $n$  are used to define points along those rays. When the two sides of the equation are set equal to each other, one can find the rays' intersection. The solutions for  $m$  and  $n$  are:

$$m = \frac{|(\mathbf{E}'_r - \mathbf{E}'_l) \times (\mathbf{P}'_r - \mathbf{E}'_r)|}{|(\mathbf{P}'_l - \mathbf{E}'_l) \times (\mathbf{P}'_r - \mathbf{E}'_r)|}$$

$$n = \frac{|(\mathbf{E}'_l - \mathbf{E}'_r) \times (\mathbf{P}'_l - \mathbf{E}'_l)|}{|(\mathbf{P}'_r - \mathbf{E}'_r) \times (\mathbf{P}'_l - \mathbf{E}'_l)|}$$

From this, we obtain:

$$\mathbf{P}'_i = \mathbf{E}'_l + (\mathbf{P}'_l - \mathbf{E}'_l)m$$

or

$$\mathbf{P}'_i = \mathbf{E}'_r + (\mathbf{P}'_r - \mathbf{E}'_r)n$$

These terms are identical if the intersection exists. We discuss non-intersecting rays in Section 1.3. We now have the estimated location of the point  $P$  in viewer coordinates. Misperceptions can be quantified by differences between  $\mathbf{P}_i$  and  $\mathbf{P}_o$ .

Before examining the consequences of incorrect acquisition and viewing parameters, it is useful to consider what it means to have those parameters correct. The picture presented to each eye has a center of projection (COP) whose position depends on image magnification ( $W_p/W_c$ ) and the orientation of the camera's optical axis relative to the sensor plane ( $h$  and  $V_c$ ). The separation between the COPs depends on inter-camera separation  $t$ , magnification  $W_p/W_c$ , and picture offset  $d$ . Two constraints must be satisfied for the viewing situation to match the viewing of the original scene. 1) Both eyes must be positioned at the appropriate COPs [Leiser et al., 1995; Wartell et al., 2002]. When the eyes are so positioned, the retinal images are the same while viewing the stereo picture as they would be while viewing the original scene. 2) The eyes' vergence (the angle between the eyes' optical axes) required to fixate a point in the virtual scene must be the same as the vergence required to fixate the corresponding point in the original scene [Leiser et al., 1995]. If the eyes are at the COPs and the correct vergence angles are not attainable, it is due to improper camera or display parameters. We will refer to viewing situations in which these constraints are satisfied as the proper viewing condition.

We are most interested in what happens when the viewing condition is not proper: specifically, when one or both eyes are not at the appropriate COPs and/or when the eye vergence is inappropriate. Incorrect positioning and vergence are common with single viewers and necessarily occur with multiple viewers.

## 1.2 Predicted Distortions: Geometric Approach

We implemented the geometric approach in software and investigated the consequences of modifying acquisition and viewing parameters. The investigation revealed viewing situations in which the geometric approach fails to produce a solution; we discuss those in Section 1.3.

Figure 1.2 shows the results of one investigation. We presented a 30cm cube at a distance

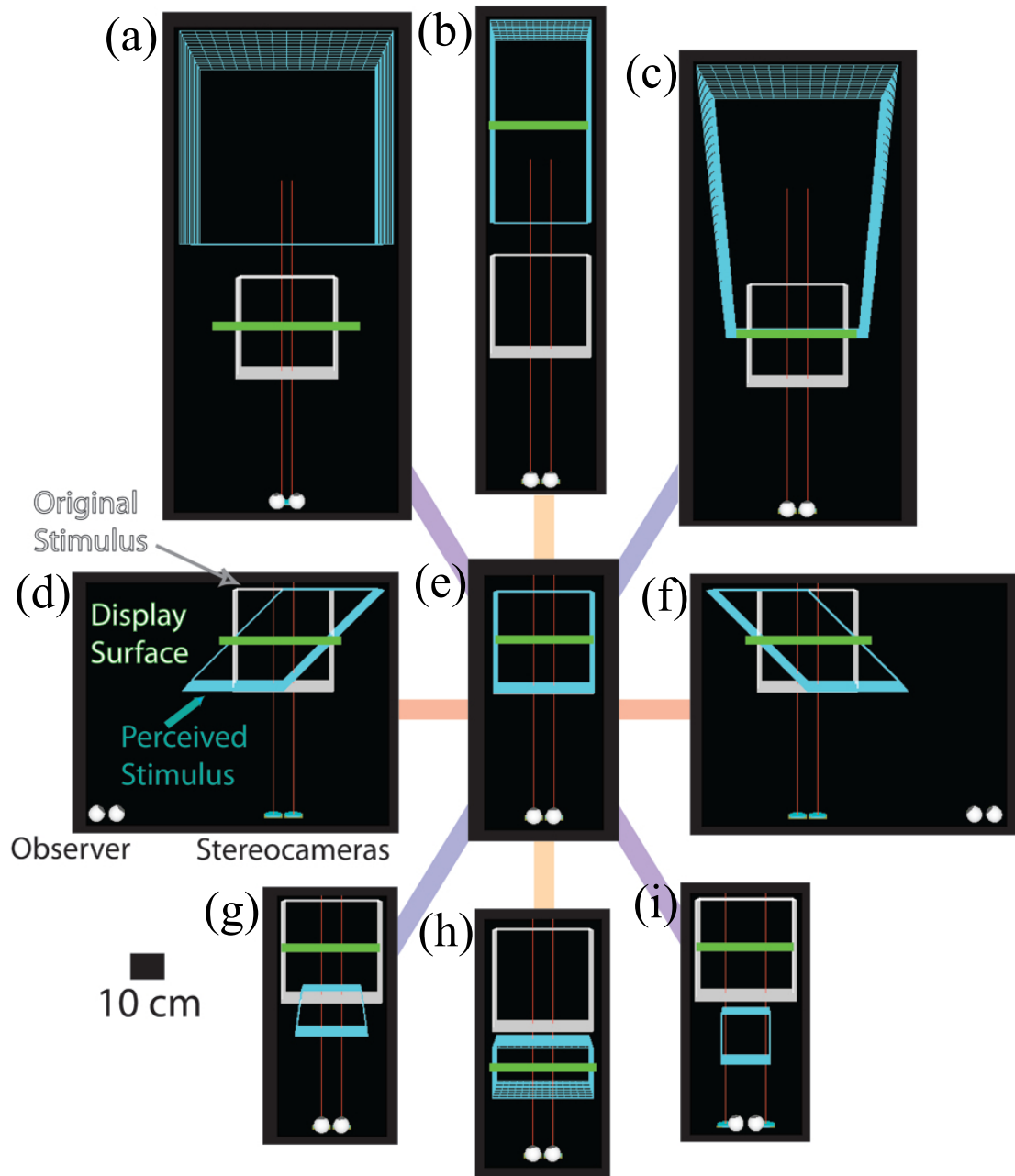


Fig. 1.2: Estimated 3D scenes for different acquisition and viewing situations. Each panel is a plan view of the viewer, stereo cameras, display surface, actual 3D stimulus, and estimated 3D stimulus. Red lines represent cameras' optical axes. (e) Proper viewing situation. Parameters are listed in Section 1.2. The actual and estimated stimuli are the same. (b) Viewer is too distant from picture. (h) Viewer is too close. (d) Viewer is too far to the left relative to the picture. (f) Viewer is too far to the right. (a) Cameras are too close together for viewer's inter-ocular distance. (i) Cameras are too far apart. (c) Distance between centers of the left and right stereo pictures (not depicted) is too great. (g) Distance between the centers of pictures is too small.

of 55cm from the cameras. In the proper viewing condition, the following parameters were used:

**Acquisition Parameters:**

Orientation of camera optical axes: Parallel

Inter-camera separation ( $t$ ): 6.2cm

Camera focal length ( $f$ ): 6.5mm

**Viewing Parameters:**

Magnification ( $W_p/W_e$ ): 84.6

Picture separation ( $d$ ): 6.2cm

Viewing distance: 55cm

Inter-ocular distance ( $I$ ): 6.2cm

Viewer position: Midpoint of inter-ocular axis on central surface normal of display.

Viewer orientation: Face parallel to display surface

We modified parameters independently to observe their effects on the estimated 3D percept. Each panel of Figure 1.2 presents the results for a set of parameters; (e) is the proper viewing condition.

Panels (b) and (h) show the consequences of moving the viewer respectively farther from (110cm) and closer to (27.5cm) the picture. When the viewer is too distant, the predicted perceived distance is greater and the predicted shape is stretched in depth. When the viewer is too close, the predicted perceived distance is less and the predicted shape is compressed in depth. These results are consistent with the analysis of Woods et al. [1993]. In (d) and (f), we translated the viewer left and right of the proper viewing position. The translation was parallel to the display surface and the viewer's head remained parallel to the surface. The predicted shape is skewed toward the viewer. A and I show the effects of inter-camera separation. The proper separation was equal to the inter-ocular distance of 6.2cm. In A, the cameras are 3.1cm apart, and the predicted stimulus is larger and farther away. In (i), the cameras are 12.4cm apart, and the predicted stimulus is smaller and closer to the

viewer than the original stimulus. (c) and (g) show the effects of picture displacement. In the proper viewing condition, the centers of the pictures were 6.2cm apart. Changing the picture displacement increases or decreases all of the disparities in the retinal images. When the pictures are separated by 7.5cm, the disparities are increased, and the result is a predicted stimulus that is farther away and stretched in depth. When the separation is 3.1cm, the predicted stimulus is closer and the shape is compressed in depth. Woods et al. [1993] did not investigate picture displacement, but this result could be derived from their analysis.

The results in Figure 1.2 are consistent with our empirical observations in these viewing situations and generally with the analysis of Woods et al. [1993]. There are, however, acquisition-viewing conditions for which the geometric approach does not yield a solution; we now turn to them.

### **1.3 Failures of the Geometric Approach**

In many viewing situations, rays from the eyes through corresponding points in the stereo pictures do not intersect, so the geometric approach cannot yield a solution for the predicted perceived stimulus. Interestingly, viewers in those situations perceive a coherent 3D scene, so the visual system finds a solution nevertheless. The presence of non-intersecting (skew) rays has been mostly unnoticed in the literature and the perceptual consequences have never been investigated. Our main contribution is an analysis of the causes of skew rays and a description of the manner in which the visual system finds a 3D solution when such rays exist.

### 1.3.1 Epipolar Planes

The causes of skew rays can be well understood in epipolar geometry. A point in real space and the two eye centers define a plane: the *epipolar plane* [Shapiro and Stockman, 2001]. It can be shown that two corresponding points in a stereo picture produce intersecting rays as long as they lie in the same epipolar plane (and are non-parallel). Consider a viewer of a stereoscopic picture with the eyes ( $E_l$  and  $E_r$ ) positioned at the COPs (Figure 1.3B). A ray from the left eye to point  $P_l$  and a ray from the right eye to  $P_r$  are identical to the rays that would have passed from the eyes to the original point  $P$ . Thus, they lie in the same epipolar plane as  $P$  and will intersect at  $P$  in virtual space. Now consider viewing the stereo picture with the eyes not at the COPs. The line segment between the COPs is the inter-COP axis. If the viewer is translated from the proper position, the inter-ocular axis will be parallel to, but not coincident with, the inter-COP axis. Rays from the two eyes to the corresponding points in the picture still lie in a common epipolar plane, so they will intersect in space. Therefore, a geometric solution exists for  $P_i$ . This is why the geometric approach could produce solutions to the viewing situations in the previous section. If the viewer's head is rotated about the inter-ocular axis (defined as "pitch"), the two axes remain coincident, so a solution still exists. But if the viewer's head is rotated about a vertical axis (yaw rotation) or a forward axis (roll rotation), the inter-ocular and inter-COP axes will be neither coincident nor parallel. In those cases, there are corresponding points in the picture that produce rays in different epipolar planes (Figure 1.3C). The rays are therefore not guaranteed to intersect, so there may be no solution for  $P_i$ . The Appendix provides a mathematical derivation of these results.

Mismatches between camera setup and display surface can also produce skew rays. For instance, the imaging sensors of converging cameras (lens axes converging) lie in different planes, but the resulting stereo pictures are usually displayed on one plane. The mismatch causes the left and right stereo pictures to exhibit "keystone" distortion (Figure 1.4), which

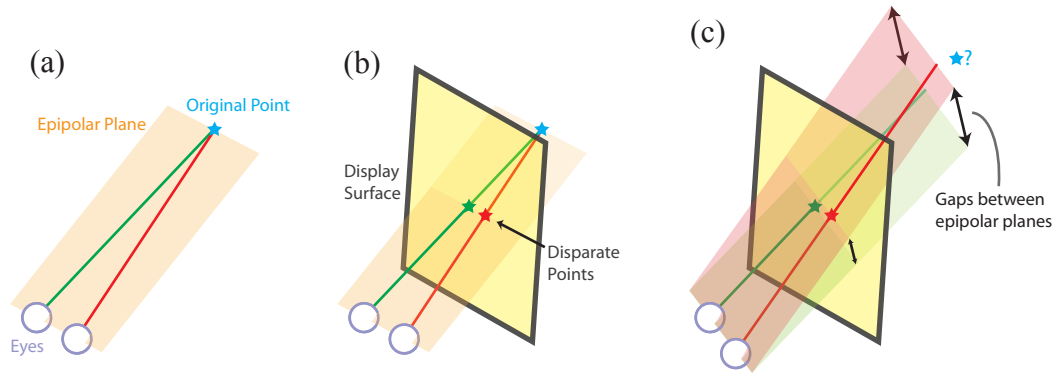


Fig. 1.3: Epipolar geometry. (a) In natural viewing, a point in space and the two eye centers define an epipolar plane. (b) If a viewer is correctly positioned relative to the picture, the rays emanating through the eyes and passing through a pair of corresponding points in the picture lie in the same epipolar plane and intersect in space. (c) With oblique viewing (head rotated about a vertical axis such that inter-ocular axis is not parallel to picture surface), rays will generally lie in different epipolar planes and never intersect.

creates non-zero on-screen vertical disparities between points that have non-zero  $Y$  coordinates in the picture. The vertical disparities produce rays that lie in different epipolar planes, so they do not provide a solution. A modification of the geometric approach provides a solution [Woods et al., 1993], but as we will show, the solution is very unlikely to match viewers' percepts.

### 1.3.2 Previous Solutions to Skew Rays

Most previous investigations of misperceptions in stereography have not discussed skew rays [Diner, 1991; Jones et al., 2001; Kusaka, 1992; Kutka, 1994; Leiser et al., 1995; Masaoka et al., 2006; Strunk and Iwamoto, 1990; Yamanoue et al., 2006], but a few have noted their existence in some situations [Agrawala et al., 1997; Wartell et al., 2002; Woods et al., 1993]. Only one of those studies considered the possible perceptual consequences: Woods et al. [1993] modified the geometric approach to accommodate skew rays created by improper acquisition and viewing settings. In particular, they observed that using converging cameras and a single display surface causes vertical disparities in the picture surface; they did not consider viewer rotations as we have done. In modifying the geomet-



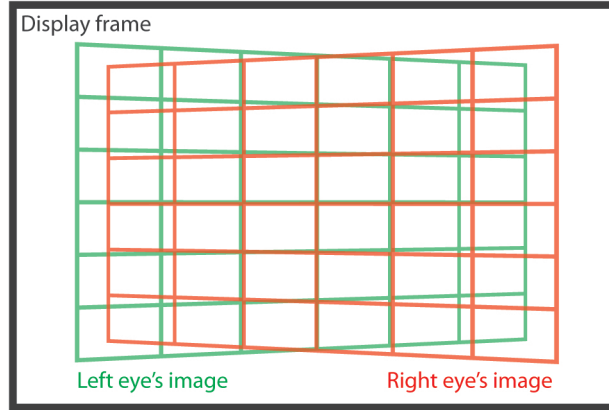


Fig. 1.4: The keystone effect. A rectangular grid was captured using converging cameras and displayed on a single flat display surface. Note the vertical disparities between the corresponding points in the corners.

ric approach, Woods and colleagues first determined which pairs of corresponding points had unequal on-screen  $Y$  coordinates ( $Y_{sl}$  and  $Y_{sr}$ ; Equations 1.1). They then reset the  $Y$  coordinates for each pair to the average  $Y$  value. This placed the on-screen points and the eye centers in a common epipolar plane, so ray intersections could be found. There are two important shortcomings with this approach. 1) It does not apply to yaw and roll rotations even though such rotations create skew rays (Figure 1.3). 2) Even in the converging-camera situation for which the approach does apply, vision-science findings strongly suggest that the 3D estimate will not match human percepts. We describe these findings next.

### 1.3.3 Vertical Disparity

A point in a real scene projects in the same epipolar plane for both eyes, but as we have said, epipolar geometry does not necessarily hold in the viewing of stereo pictures. Consequently, corresponding points in a stereo picture may project to different elevations in the two eyes, thereby creating non-zero vertical disparities in an epipolar coordinate system. Such non-zero vertical disparities are known to influence 3D percepts. An example is the *induced effect*. A lens is placed before one eye that magnifies the image vertically and creates non-zero vertical disparities. When this is done, a frontoparallel surface appears

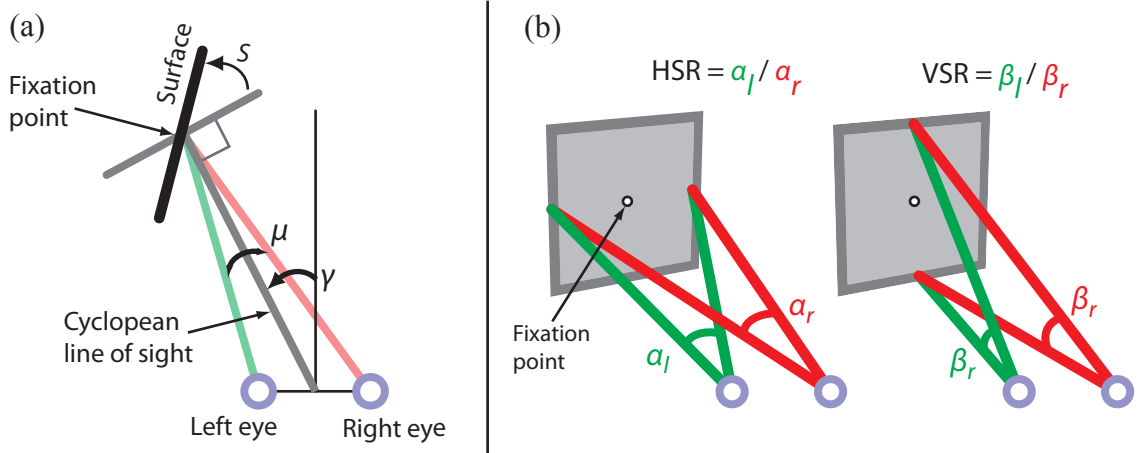


Fig. 1.5: (a) Plan view of viewer fixating a planar surface.  $S$  is the slant of the patch.  $\mu$  is eyes' horizontal vergence;  $\gamma$  is eyes' horizontal version. (b) Definitions of  $HSR$  (horizontal size ratio) and  $VSR$  (vertical size ratio).  $HSR$  is the ratio of the horizontal angles a surface patch subtends at left and right eyes.  $VSR$  is the ratio of vertical angles. Adapted from Backus et al. [1999].

slanted even though the horizontal disparities created by the surface are unaffected by the magnifier [Ogle, 1938]. There are many other perceptual consequences of altering vertical disparity, so it is well accepted in vision science that 3D percepts are a product of horizontal and vertical disparities [Backus et al., 1999; Banks et al., 2001; Rogers and Bradshaw, 1993, 1995]. Indeed, the visual system uses a number of depth cues to estimate the 3D structure of the environment. Many are monocular cues such as perspective and shading, which are beyond the scope of our discussion. But two estimation methods are based on stereopsis and should therefore be considered here. One stereoscopic estimation method is based on measuring horizontal disparities and eye position [Backus et al., 1999]:

$$S \approx -\arctan\left(\frac{1}{\mu} \ln(HSR) - \tan(\gamma)\right) \quad (1.2)$$

where  $S$  is the slant of a surface patch,  $HSR$  is the horizontal size ratio (a measure of horizontal disparity; defined in Figure 1.5),  $\mu$  is the eyes' vergence (defined in the figure), and  $\gamma$  is the eyes' version (defined in the figure). The geometric approach discussed here is identical to this means of estimating surface orientation. Another stereoscopic method

is based on measuring horizontal and vertical disparity and does not require an estimate of eye position [Backus et al., 1999]:

$$S \approx -\arctan\left(\frac{1}{\mu} \ln\left(\frac{HSR}{VSR}\right)\right) \quad (1.3)$$

where  $VSR$  is the vertical size ratio (a measure of vertical disparity; defined in Figure 1.5B), and  $\tilde{\mu}$  is a measure of vergence derived from the gradient of  $VSR$ .<sup>1</sup> The visual system uses both of these stereoscopic methods to estimate surface orientation from binocular disparity [Backus et al., 1999; Gårding et al., 1995; Rogers and Bradshaw, 1995]. When the two methods provide different estimates, the system's final estimate is a weighted average of the two with the weights determined by the relative reliabilities of the two methods [Backus et al., 1999; Rogers and Bradshaw, 1995].

### 1.3.4 Perception with Skew Rays Present

What is the visual system doing in situations that produce skew rays? We consider this problem by examining the three situations mentioned above.

#### **Condition 1: Observer Rotation in $X - Z$ Plane (Yaw)**

Figure 1.6 is a stereo picture of a cube. To see the perceptual consequences of a yaw rotation, rotate the picture about a vertical axis. The 3D percept changes in a few ways: the front and back surfaces appear to rotate relative to the viewer such that they remain roughly parallel to the picture surface; the front surface appears to rotate slightly less than the back surface, so the surfaces become non-parallel; the distance between the front and back surfaces appears to decrease. To understand the perceptual consequences, we need to consider the horizontal and vertical disparities created by the cube following a yaw rotation. Figure 1.7 plots those disparities as vectors; panels (b) and (e) show the disparities

---

<sup>1</sup>These equations apply for tilt 0. Extensions have been derived for all tilts [Banks et al., 2001].

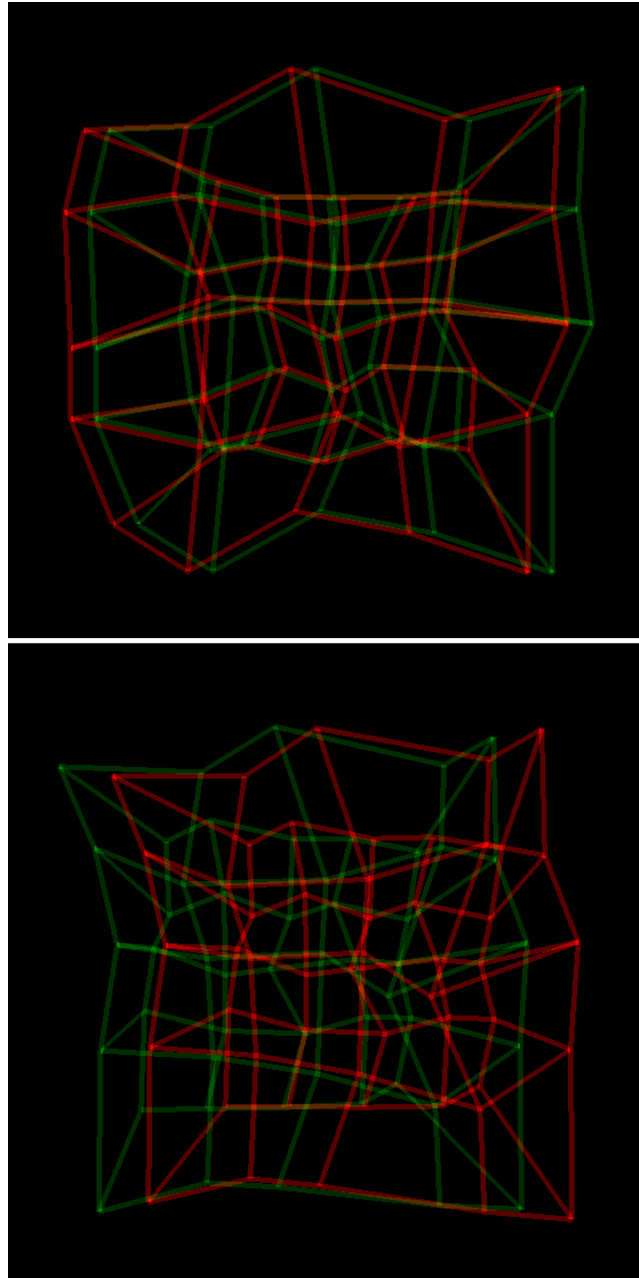


Fig. 1.6: Anaglyph stereograms captured with the acquisition settings listed in Section 1.2. Top: cameras with parallel optical axes. Bottom: cameras' optical axes were converged at 0.55m (center of cube). To view the stereograms, use red-green glasses with green filter over left eye. Try different viewing situations. 1) Move closer to and farther away from the page. 2) Move left and right while holding the head parallel to the page. 3) Position yourself directly in front of the page and rotate the head about a vertical axis (yaw) and then about a forward axis (roll). In each case, notice the changes in the cube's apparent shape. Points in the cube were randomly perturbed to lessen contributions of perspective cues to 3D percept.

associated with the cube's front and back surfaces, respectively. There are regions in the stimulus in which the vertical disparities reverse sign from the front to back surface; the upper left corner is an example. Such a sign reversal can never occur in natural viewing with aligned eyes.<sup>2</sup> For this reason, we cannot appeal to a natural situation to determine what the visual system perceives when a stereo picture undergoes a yaw rotation. The answer, however, is suggested by the vision science literature. Duke and Howard [2005] created stereograms of two transparent planes, one in front of the other. They applied one pattern of vertical disparity to one plane and the opposite pattern to the other plane; this creates reversals in the sign of vertical disparity (as we observed with yaw rotations). Viewers of these unnatural stimuli perceived different surface shapes for the front and back surface and those shapes are well predicted by a weighted combination of Equations 1.2 and 1.3, applied separately to the two surfaces. We found that the percept associated with yaw rotation while viewing stereo pictures is well predicted by a similar weighted combination of surface orientation estimates derived from the two means of estimation.

### **Condition 2: Observer Rotation in X-Y Plane (Roll)**

To see the consequences of a roll rotation, rotate the upper picture (Figure 1.6) about the forward axis. The 3D percept changes little with small rotations and then collapses with larger rotations as the visual system becomes unable to fuse the disparate images. As shown earlier, roll rotations cause non-intersecting rays, so once again the geometric approach cannot derive an estimate for the perceived 3D structure. To understand the perceptual effects, we again consider the horizontal and vertical disparities created by this viewing situation. Panels (a) and (d) in Figure 1.7 plots the disparities associated with the cube's front and back surfaces. The disparity pattern can be understood by considering how the eyes' positions change with head roll. If the roll is counterclockwise, the right eye moves up

---

<sup>2</sup>The vertical disparity associated with a point in space is non-zero if the point is to the left or right of straight ahead (i.e., not in the head's mid-sagittal plane) and above or below the plane of fixation (i.e., not in the visual plane). For any combination of such azimuth and elevation the vertical disparities of points at all distances have the same sign.

## Disparity Fields Produced By Improper Viewing

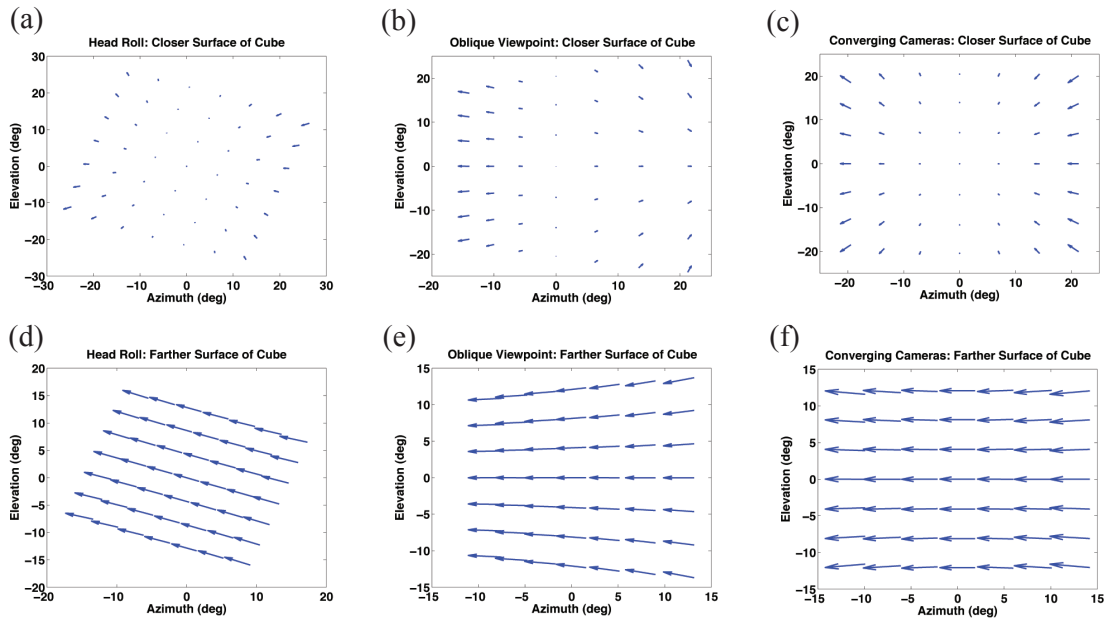


Fig. 1.7: Disparity as a function of azimuth and elevation. Fick coordinates (azimuth and elevation measured as longitudes and latitudes, respectively) were used. Vectors represent the direction and magnitude of disparities on the retinas produced by a stereoscopic image of a cube 0.3m on a side and placed 0.55m in front of the stereo cameras. Unless otherwise noted, the conditions listed in Section 1.2 were used to generate the figures. Arrow tails represent points on right eye's retina, and arrowheads represent corresponding points in left eye's retina. Panels (a), (b), and (c) contain points from the proximal face of the cube, where the eyes are fixating. (d), (e), and (f) represent the cube's distal face. In (a) and (d), the observer is viewing the display at a 45deg angle. In (b) and (e), the viewer's head has been rolled 20deg. In (c) and (f), the cameras converge at 0.55m.

and becomes closer to the upper right corner and farther from the bottom right corner. The opposite is true for the left eye. As a result, the upper right corner creates a larger retinal image in the right than in the left eye. The opposite is true for the bottom right corner. In both corners, the vertical disparities in epipolar coordinates have changed from zero with no roll to non-zero after roll. The horizontal disparities have been altered as well. The changes in vertical and horizontal disparity are proportional to the distance of the point from the rotation axis. Because of this, the perceived shape of the front or back surface should become curved, one corner bending toward the viewer and the opposite bending away. The amount of curvature depends on the magnitude of roll and whether the points on the picture have crossed or uncrossed horizontal disparity. The predicted deviation is only

significant with large rolls, so one expects little if any perceptual change for small rolls. We have been unable to observe the curvature effect because the ability to fuse the stimulus breaks down at the larger rolls where the effect is predicted.

### **Condition 3: Converging Cameras**

Figure 1.6 demonstrates the perceptual consequences of using converging cameras but a single display surface for viewing. The top panel is a stereo picture of a cube when the cameras were parallel and the bottom panel is a picture of the same cube when the cameras were converging. The 3D percepts for the two cases differ in two ways: the cube's front and back surfaces appear planar in the parallel-camera case and convex in the converging-camera case; the front and back surfaces appear closer to the viewer in the converging case. To understand the perceptual consequences of using converging cameras, we need to again consider the horizontal and vertical disparities. Using converging cameras causes keystoneing (Figure 1.4), thereby changing the pattern of vertical disparities. In particular, the horizontal gradient of vertical disparity is altered such that it specifies a nearer surface than is actually present; the alteration is different for near and far surfaces, but it always increases the vertical-disparity gradient. Converging cameras also alter horizontal disparities. Specifically, the horizontal gradient of horizontal disparity specifies a more convex surface than is actually present; again the change differs for near and far surfaces, but always increases the gradient. From Duke and Howard [2005], we know that the visual system is likely to estimate 3D structure by estimating the orientation and curvature of surfaces separately with a weighted combination of Equations 1.2 and 1.3. We found that this model predicts the percept associated with converging cameras quite well.

## 1.4 Future Work in Stereoscopic Misperceptions

To produce a truly useful model for stereoscopic misperceptions, the model's predictions must be tested against psychophysical data. This requirement certainly applies to the current case. In particular, it will be crucial to know how much weighting the visual system assigns to the perceptual estimates produced by the geometric and the vertical-disparity-based approaches outlined above. But for now, our analysis can provide rough guidelines for the design and evaluation of stereoscopic displays and viewing parameters. In Chapter 3, we provide more detail on how those guidelines may be applied to stereo medical displays.

It should also be noted that non-stereo pictorial cues can affect the 3D percept of a scene. In the next chapter, we discuss how the blur in an image can modulate the perceived distance and size of a scene, and how rendering or camera settings can be chosen to either avoid or enable such modulations.



## CHAPTER 2

# USING BLUR TO AFFECT PERCEIVED DISTANCE AND SIZE

The pattern of blur in an image can strongly influence the perceived scale of the captured scene. For example, cinematographers working with miniature models can make scenes appear life size by using a small camera aperture, which reduces the blur variation between objects at different distances [Fielding, 1985]. The opposite effect is created in a photographic manipulation known as the tilt-shift effect: A full-size scene is made to look smaller by adding blur with either a special lens or post-processing software tools [Laforet, 2007; Flickr, 2009; Vishwanath, 2008].

Figures 2.1 and 2.2 demonstrate the miniaturization effect. In Figure 2.2, the image in the upper left has been rendered sharply and to typical viewers looks like a life-size scene in San Francisco. The upper-right image has been rendered with a blur pattern consistent with a shorter focal distance, and it looks like a miniature-scale model. The two images in the lower row demonstrate how the application of a linear blur gradient can have a similar effect.

Clearly, blur plays a significant role for conveying a desired sense of size and distance. However, the way the visual system uses blur to estimate perceived scale is not well understood. Okatani and Deguchi [2007] have shown that additional information, such as perspective, is needed to recover scene scale. But a more detailed, perceptually based model will provide further insight into the effective application of blur. This chapter presents a general probabilistic model of distance estimation from blur. From the model, we develop an algorithm for manipulating blur in images to produce the desired apparent scale. We then validate the model and algorithm with a psychophysical study. Finally, we detail how



Fig. 2.1: (a) Rendering a cityscape with a pinhole aperture results in no perceptible blur. The scene looks large and far away. (b) Simulating a 60m-wide aperture produces blur consistent with a shallow depth of field, making the scene appear to be a miniature model. Original city images and data from GoogleEarth are copyright Terrametrics, SanBorn, and Google.

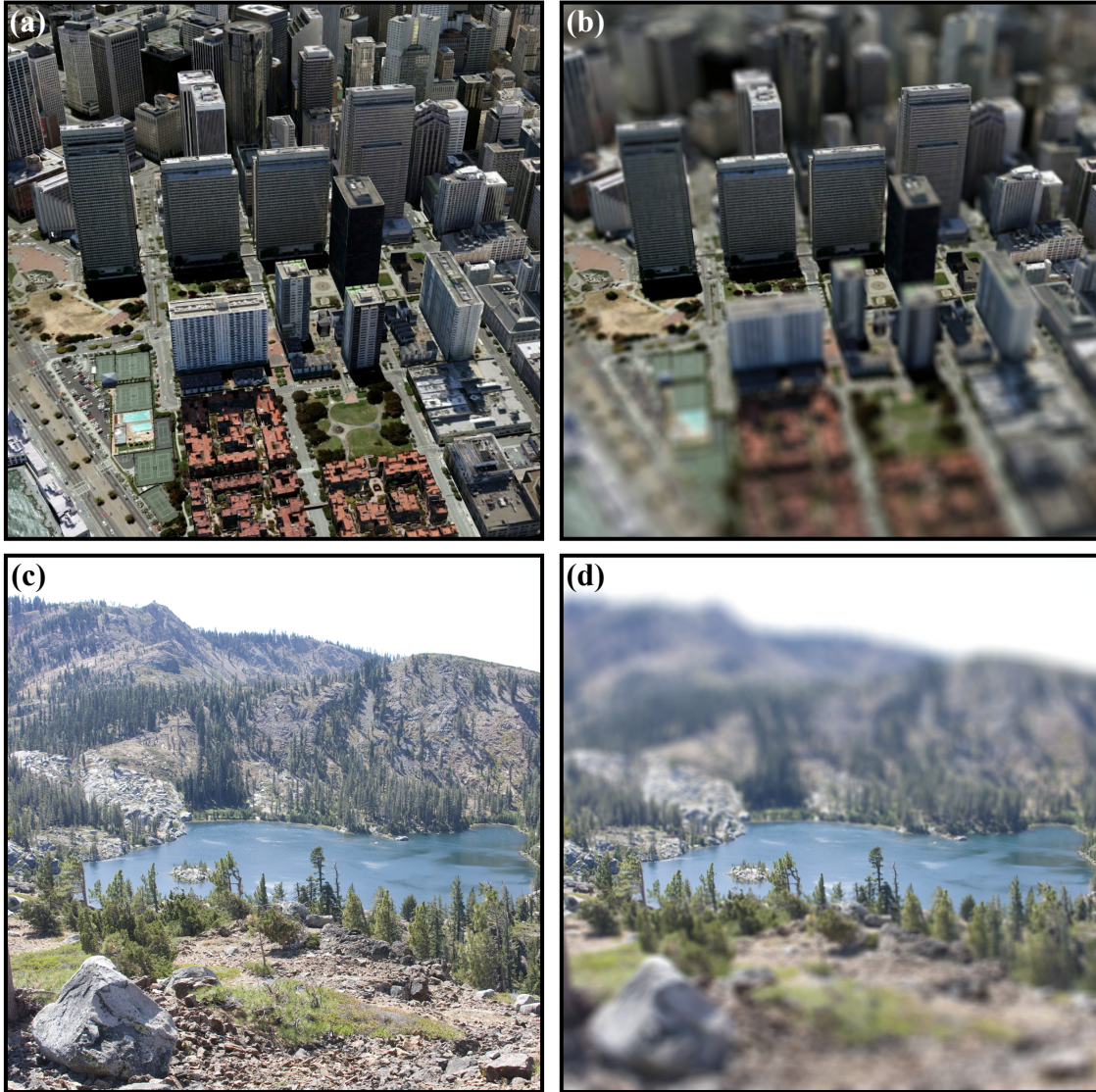


Fig. 2.2: Upper two images: Another example of how rendering an image with a shallow depth of field can make a downtown cityscape appear to be a miniature-scale model. The left image was rendered with a pinhole camera, the right with a 60m aperture. Lower two images: Applying a blur gradient that approximates a shallow depth of field can also induce the miniaturization effect. The effects are most convincing when the images are large and viewed from a short distance. Original city images and data from GoogleEarth are copyright Terrametrics, SanBorn, and Google. Original lake photograph is copyright Casey Held.

camera and rendering settings may be used to either accurately reproduce or modify the perceived scale of a scene.

## **2.1 Background**

### **2.1.1 Defocus Blur in Computer Graphics and Photography**

The importance of generating proper depth-of-field effects in synthetic or processed imagery is well established. Special-effects practitioners often manipulate the depth of field in images to convey a desired scale [Fielding, 1985]. For example, it is commonplace in cinematography to record images with small apertures to increase the depth of field. Small apertures reduce the amount of defocus blur and this sharpening causes small-scale scenes to look larger. Images created with proper defocus are also generally perceived as more realistic and more aesthetic [Hillaire et al., 2007, 2008].

Blur can also be used to direct viewers' attention to particular parts of an image. For example, photographers and cinematographers direct viewer gaze toward a particular object by rendering that object sharp and the rest of the scene blurred [Kingslake, 1992; Fielding, 1985]. Eye fixations and attention are in fact drawn to regions with greater contrast and detail [Kosara et al., 2002; Cole et al., 2006; DiPaola et al., 2010]. These applications of blur may help guide users toward certain parts of an image, but our analysis suggests that the blur manipulations could also have the undesired consequence of altering perceived scale. This suggestion is consistent with another common cinematic practice where shallow depth of field is used to draw a viewer into a scene and create a feeling of intimacy between the film subjects and viewer. A shallower depth of field implies a smaller distance between the viewer and subjects which creates the impression that the viewer must be standing near the subjects.

As depth-of-field effects are used more frequently, it is important to understand how to generate them accurately and efficiently. Some of the earliest work on computer-generated imagery addressed the problem of correctly rendering images with defocus blur. Potmesil and Chakravarty presented a detailed description of depth-of-field effects, the lens geometry responsible for their creation, and how these factors impact rendering algorithms [Potmesil and Chakravarty, 1981]. The seminal work on distribution ray tracing by Cook and colleagues discussed defocus and presented a practical method for rendering images with finite apertures [Cook et al., 1984]. Likewise, the original REYES rendering architecture was built to accommodate a finite aperture [Cook et al., 1987]. Development of the accumulation buffer was motivated in part by the need to use hardware rendering methods to generate depth-of-field effects efficiently [Haeberli and Akeley, 1990]. Kolb and colleagues described a method for rendering blur effects that are specific to a real lens assembly as opposed to an ideal thin lens [Kolb et al., 1995]. Similarly, Barsky investigated rendering blurred images using data measured from a specific human eye [Barsky, 2004].

Even with hardware acceleration, depth-of-field effects remain relatively expensive to render. Many methods for accelerating or approximating blur due to defocus have been developed [Fearing, 1995; Rokita, 1996; Barsky et al., 2003a,b; Mulder and van Liere, 2000], and the problem of rendering such effects remains an active area of research.

Researchers in computer vision and graphics have also made use of the relationship between depth and blur radius for estimating the relative distances of objects in photographs. For example, Pentland showed that blur from defocus can be used to recover an accurate depth map of an imaged scene when particular parameters of the imaging device are known [Pentland, 1987]. More recently, Green and colleagues used multiple photographs taken with different aperture settings to compute depth maps from differences in estimated blur [Green et al., 2007]. Others have created depth maps of scenes using specially constructed camera apertures [Levin et al., 2007; Green et al., 2007; Moreno-Noguer et al., 2007].

### 2.1.2 Aperture and Blur

When struck by parallel rays, an ideal thin lens focuses the rays to a point on the opposite side of the lens. The distance between the lens and this point is the focal length,  $f$ . Light rays emanating from a point at some other distance  $z_1$  in front of the lens will be focused to another point on the opposite side of the lens at distance  $s_1$ . The relationship between these distances is given by the thin-lens equation:

$$\frac{1}{s_1} + \frac{1}{z_1} = \frac{1}{f}. \quad (2.1)$$

In a typical imaging device, the lens is parallel to the image plane containing the film or CCD array. If the image plane is at distance  $s_0$  behind the lens, then light emanating from features at distance  $z_0 = 1/(1/f - 1/s_0)$  along the optical axis will be focused on that plane (Figure 2.3). The plane at distance  $z_0$  is the focal plane, so  $z_0$  is the focal distance of the device. Objects at other distances will be out of focus, and hence will generate blurred images on the image plane. We can express the amount of blur by the diameter  $c$  of the blur circle in the image plane. For an object at distance  $z_1$ ,  $c_1 = |A(s_0/z_0)(1 - z_0/z_1)|$ , where  $A$  is the diameter of the aperture. It is convenient to substitute  $d$  for the relative distance  $z_1/z_0$ , yielding:

$$c_1 = |A \frac{s_0}{z_0} (1 - \frac{1}{d})| \quad (2.2)$$

The depth of field is the width of the region centered around the focal plane where the blur circle radius is below the sharpness threshold, or the smallest amount of perceptible blur. Real imaging devices, like the human eye, have imperfect optics and more than one refracting element, so Eqs. 2.1 and 2.2 are not strictly correct. Later we describe those effects and show that they do not affect our analysis significantly. An important aspect of Eq. 2.2 is the inverse relationship between  $z_0$  and  $c_1$ . This relationship means that the blur at a given relative distance  $d$  increases as  $z_0$  decreases. In other words, the depth of field

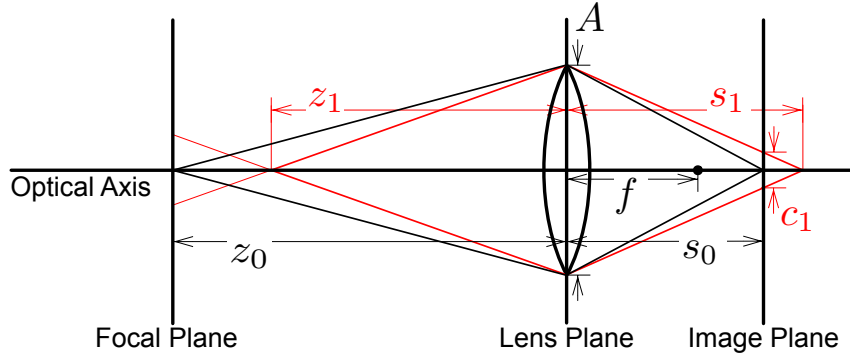


Fig. 2.3: Schematic of blur in a simple imaging system.  $z_0$  is the focal distance of the device given the lens focal length,  $f$ , and the distance from the lens to the image plane,  $s_0$ . An object at distance  $z_1$  creates a blur circle of diameter  $c_1$ , given the device aperture,  $A$ . Objects within the focal plane will be imaged in sharp focus. Objects off the focal plane will be blurred proportional to their dioptric ( $m^{-1}$ ) distance from the focal plane.

becomes narrower with closer focal distances. For this reason, a small scene imaged from close range generates greater blur than a scaled-up version of the same scene imaged from farther away. As explained in Section 2.3, it is this relationship that produces the perceived miniaturization in Figure 2.2.

## 2.2 Adjusting Blur to Modulate Perceived Distance and Size

### 2.2.1 Tilt-and-shift Lenses and Linear Blur Gradients

So far we have assumed that the imaging and lens planes are parallel, but useful photographic effects can be generated by slanting the two planes with respect to each other. Some cameras do so with a “tilt-and-shift” lens mount that allows the position and orientation of the lens to be changed relative to the rest of the camera [Kingslake, 1992]; other cameras achieve an equivalent effect by adjusting the orientation of the filmback. Rotation of the lens relative to the image plane affects the orientation of the focal plane relative to the optical axis (Figure 2.4). In such cases, the image and lens planes intersect along the so-

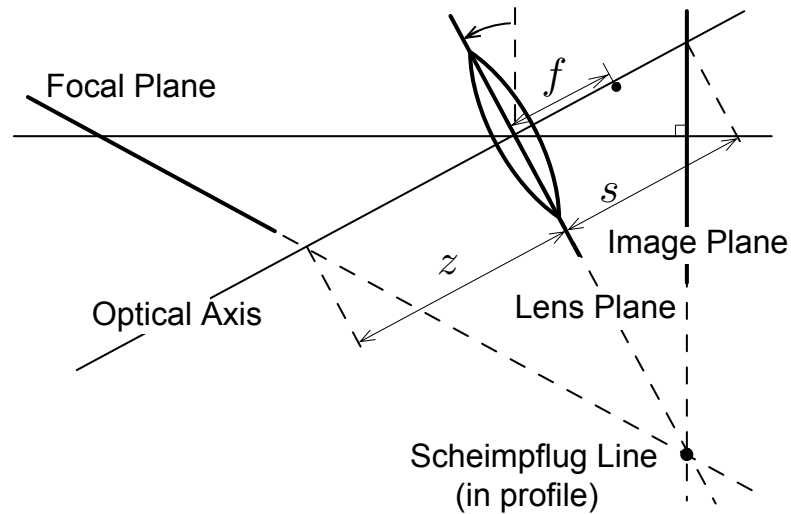


Fig. 2.4: The Scheimpflug Principle. Tilt-and-shift lenses cause the orientation of the focal plane to shift and rotate relative to the image plane. As a result, the apparent depth of field in an image can be drastically changed and the photographer has greater control over which objects are in focus and which are blurred.

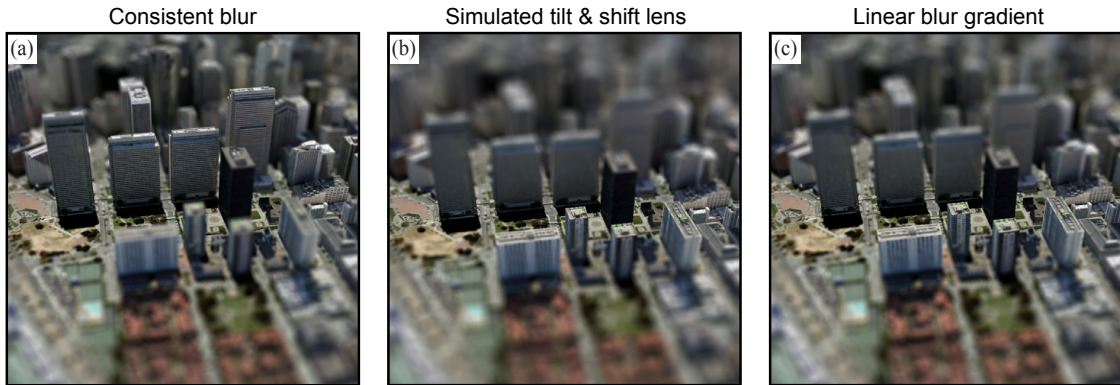
called Scheimpflug line. The focal plane also intersects those planes at the Scheimpflug line [Kingslake, 1992; Okatani and Deguchi, 2007]. Because the focal plane is not perpendicular to the optical axis, objects equidistant from the lens will not be blurred equally. One can take advantage of this phenomenon by tilting the lens so that an arbitrary plane in the scene is in clear focus. For example, by making the focal plane co-planar with a tabletop, one can create a photograph in which all of the items on the table are in clear focus. The opposite effect is created by tilting the lens in the opposite direction, so that only a narrow band of the tabletop is in focus. The latter technique approximates a smaller depth of field. The pattern of blur in the image is close to that produced by a slanted object plane photographed with a conventional camera at short range [Okatani and Deguchi, 2007]. McCloskey and colleagues [2009] showed that the pattern of blur produced by a slanted plane is a linear gradient, with the blur and distance gradients running in the same direction. Therefore, it stands to reason that a tilt-and-shift image could be similar to a sharply rendered image treated with a linear blur gradient. Indeed, most of the tilt-and-shift examples popular today, as well as Figure 2.2(d), were created this way [Flickr, 2009]. However, there can be large differences in the blur patterns produced by each method, and it would be useful to



know whether those differences have any impact on the perception of the image.

## 2.2.2 Comparing Blur Patterns

### Blur-rendering Techniques:



### Comparison of Blur-circle Diameters:

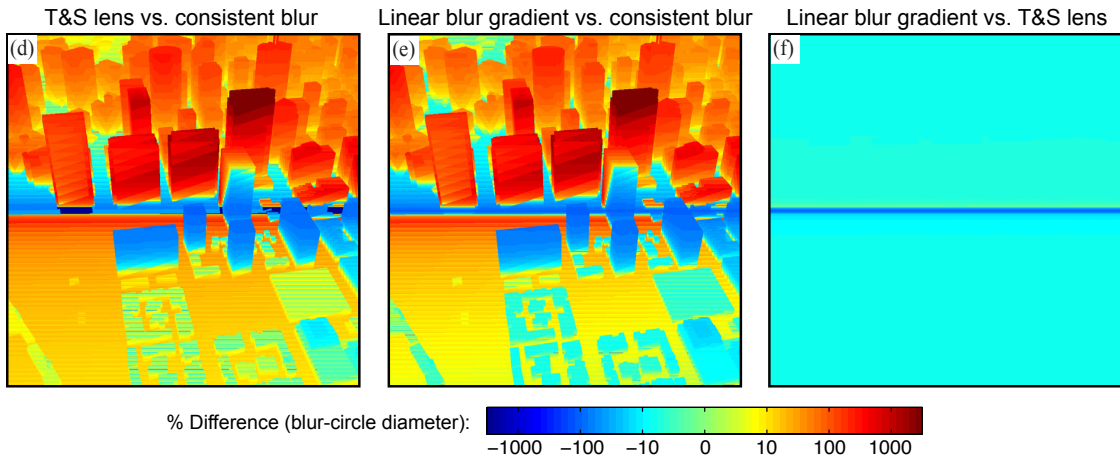


Fig. 2.5: Comparison of blur patterns produced by three rendering techniques: consistent blur (a), simulated tilt-and-shift lens (b), and linear blur gradient (c). The settings in (b) and (c) were chosen to equate the maximum blur-circle diameters with those in (a). The percent differences in blur-circle diameters between the images are plotted in (d), (e), and (f). Panels (d) and (e) show that the simulated tilt-and-shift lens and linear blur gradient do not closely approximate consistent blur rendering. The large differences are due to the buildings, which protrude from the ground plane. Panel (f) shows that the linear blur gradient provides essentially the same blur pattern as a simulated tilt-and-shift lens. Most of the differences in (f) are less than 7%; the only exceptions are in the band near the center, where the blur diameters are less than one pixel and not detectable in the final images.

As previously discussed, it is often useful to make the scene depicted in an image appear bigger or smaller than it actually is. Special-effects practitioners can make a scene look

bigger by recording with a small aperture or make the scene look smaller by recording with a large aperture [Fielding, 1985]. Consider recording a small scene located  $z_0$  meters away from the camera and trying to make it appear to be  $m$  times larger and located  $\hat{z}_0 = mz_0$  meters away from the camera. Assume that with a camera aperture diameter of  $A$ , the apparent size matches the actual size. Then, referring to the equations we developed in Section 2.1.2, the amount of blur we want to have associated with a given relative distance  $d$  is given by:

$$\hat{c}_1 = \left| A \frac{s_0}{\hat{z}_0} \left(1 - \frac{1}{d}\right) \right| = \left| A \frac{s_0}{mz_0} \left(1 - \frac{1}{d}\right) \right| = \left| \left(\frac{A}{m}\right) \left(\frac{s_0}{z_0}\right) \left(1 - \frac{1}{d}\right) \right| \quad (2.3)$$

Here we see that we can achieve the same amount of blur as encountered with a focal distance of  $\hat{z}_0 = mz_0$  by shooting the scene at distance  $z_0$  and setting the diameter of the camera aperture to  $\hat{A} = A/m$ . The aperture must therefore be quite small to make a scene look much larger than it is and this limits the amount of available light, causing problems with signal-to-noise ratio, motion, and so forth. Likewise the aperture must be quite large to make the scene look much smaller than it actually is, and such large apertures might be difficult to achieve with a physical camera. Because of these limitations, it is quite attractive to be able to use a conventional camera with an aperture of convenient size and then to manipulate blur in post-processing, possibly with blur gradients.

We quantified the differences in three types of blur—consistent, linear gradient, and tilt-shift—by applying them to 14 full-scale scenes of San Francisco taken from GoogleEarth (Figure 2.5(a) shows an example). In each image, we wanted to produce large variations in blur, as if viewed by the human eye (aperture  $\approx 4.6\text{mm}$ ) with a focal distance  $z_0$  of only 0.06m to the center of the scene. Because the actual focal distance was 785m, being consistent with a human eye at 0.06m meant that a virtual camera with a very large aperture of 60.0m was necessary. To produce each consistent-blur image, we captured many images of the same locale from positions on a jittered grid covering a circular aperture. We translated

each image to ensure that objects in the center of the scene, which were meant to be in focus, were aligned from one image to another. We then averaged those images to produce the final image. This approach is commonly used with hardware scan-conversion renderers to generate images with accurate depth-of-field effects [Haeberli and Akeley, 1990]. The tilt-and-shift images were generated in a similar fashion, but with the simulated image plane slanted relative to the camera aperture. The slant angles were chosen to produce the same maximum blur magnitudes as the consistent-blur images (slant =  $-16.6^\circ$  in Figure 2.5(b)). The direction of the slant (the tilt) was aligned with the distance gradient in the scenes. The distance gradient was always vertical, so the aligned blur gradient was also vertical. The maximum magnitudes of the gradients were set to the average blur magnitudes along the top and bottom of the consistent-blur images (Figure 2.5(c)). Thus, the histograms of blur magnitude were roughly equal across the three types of blur manipulation. For the linear blur gradients, blur was applied to the pixels by convolving them with cylindrical box kernels. A vertical blur gradient is such that all of the pixels in a given row are convolved with the same blur kernel.

We calculated the differences between the blur diameters produced by each rendering technique. The blur patterns in the tilt-and-shift-lens and linear-blur-gradient images were similar to each other (generally never differing by more than 7%; Figure 2.5(f)), but differed greatly from the pattern in the consistent-blur condition (Figure 2.5(c)(d)). The differences relative to consistent blur result from the buildings that protrude from the ground plane. In Section 2.5, we explore whether the differences affect perceived distance. This analysis was performed on all of our example images and we found that linear blur gradients do in fact yield close approximations of tilt-and-shift blur, provided that the scenes are roughly planar. This is why tilt-and-shift images and their linear-blur-gradient approximations have similarly compelling miniaturization effects.

Our next question is, why does blur affect the visual system's estimates of distance and

size? To answer this, we developed a probabilistic model of the distance information contained in image blur.

## **2.3 Model: Blur as an Absolute Depth Cue**

### **2.3.1 Vision Science Literature**

The human eye, like other imaging systems, has a limited depth of field, so it encounters blur regularly. Blur depends partly on the distance to an object relative to where the eye is focused, so it stands to reason that it might be a useful perceptual cue to depth. The vision science literature is decidedly mixed on this issue. Some investigators have reported clear contributions of blur to depth perception [Pentland, 1987; Watt et al., 2005], but others have found either no effect [Mather and Smith, 2000] or qualitative effects on perceived depth ordering, but no more [Marshall et al., 1996; Mather, 1996; Palmer and Brooks, 2008]. This conflicts with the clear perceptual effects associated with the blur manipulation in Figure 2.2. A better understanding of the distance information contained in blur should yield more insight into the conditions in which it is an effective depth cue.

### **2.3.2 Probabilistic Inference of Distance from Blur**

The physical relationship between camera optics and image blur can help us understand the visual system's use of retinal-image blur. For instance, if an object is blurred, is it possible to recover its distance from the viewer? To answer this, we return to Eq. 2.2. Regardless of whether a photograph or a real scene is being viewed, we assume that the visual system interprets the retinal image as being produced by the optics of the eye. Now the aperture  $A$  is the diameter of a human pupil,  $z_0$  is the distance to which the eye is focused, and  $d$  is the relative distance to the point in question. Figure 2.6 shows the probability of  $z_0$  and  $d$

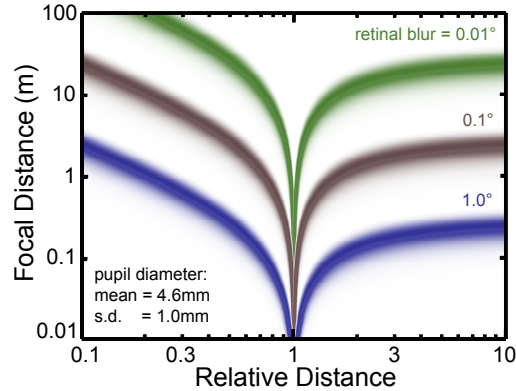


Fig. 2.6: Focal distance as a function of relative distance and retinal-image blur. Relative distance is defined as the ratio of the distance to an object and the distance to the focal plane. The three colored curves represent different amounts of image blur expressed as the diameter of the blur circle,  $c$ , in degrees. We use angular units because in those units, the image device’s focal length drops out Kingslake [1992]. The variance in the distribution was determined by assuming that pupil diameter is Gaussian distributed with a mean of 4.6mm and standard deviation of 1mm Spring and Stiles [1948]. For a given amount of blur, it is impossible to recover the original focal distance without knowing the relative distance. Note that as the relative distance approaches 1, the object moves closer to the focal plane. There is a singularity at a relative distance of 1 because the object is by definition completely in focus at that distance.

for a given amount of blur, assuming  $A$  is  $4.6\text{mm} \pm 1\text{mm}$  [Spring and Stiles, 1948]. For each blur magnitude, infinite combinations of  $z_0$  and  $d$  are possible. The distributions for large and small blur differ: large blur diameters are consistent with a range of short focal distances, and small diameters are consistent with a range of long distances. Nonetheless, one cannot estimate focal distance or relative distance from a given blur observation. How then does the change in perceived distance and size in Figure 2.2 occur?

The images in Figure 2.2 contain other depth cues—linear perspective, relative size, texture gradient, *etc.*—that specify the relative distances among objects in the scene. Such cues are scale ambiguous, with the possible exception of familiar size (see Section 3.3.2), so they cannot directly signal the absolute distances to objects. We can, however, determine absolute distance from the combination of blur and those other cues. To do this, we employ Bayes’ Law, which prescribes how to compute the statistically optimal (*i.e.*, lowest variance) estimate of depth from uncertain information. In the current case of estimating

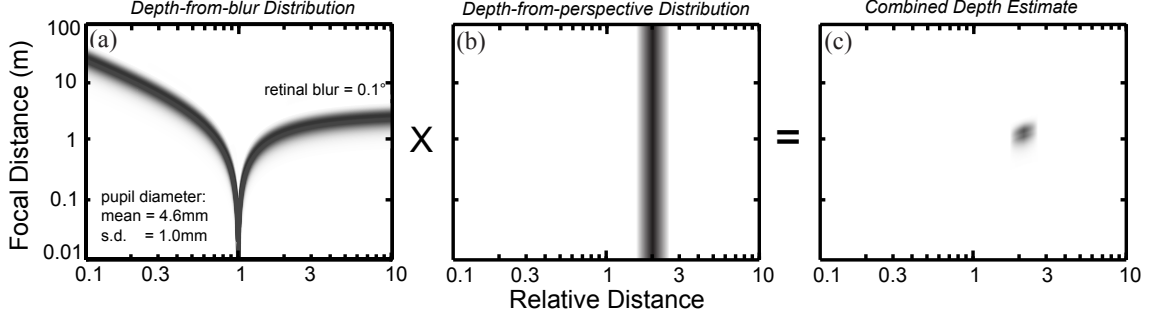


Fig. 2.7: Bayesian analysis of blur as cue to absolute distance. (a) The probability distribution  $P(z_o, d|c)$  where  $c$  is the observed blur diameter in the image (in this case,  $0.1^\circ$ ),  $z_o$  is the focal distance, and  $d$  is the relative distance of another point in the scene. Measuring the blur produced by an object cannot reveal the absolute or relative distance to points in the scene. (b) The probability distribution  $P(z_o, d|p)$  where  $p$  is the observed perspective. Perspective specifies the relative distance, but not the absolute distance: it is scale ambiguous. (c) The product of the distributions in (a) and (b). From this posterior distribution, the absolute and relative distances of points in the scene can be estimated.

distance from blur and other cues, estimates should be based on the posterior distribution:

$$P(z_o, d|c, p) = \frac{P(c|z_o, d)P(p|z_o, d)P(z_o, d)}{P(c, p)} \quad (2.4)$$

where  $c$  and  $p$  represent the observed blur and perspective, respectively. In this context, perspective refers to all pictorial cues that result from perspective projection, including the texture gradient, linear perspective, and relative size. Using a technique in Burge *et al.* [2010], we convert the likelihood distributions and prior on the right side of the equation into posterior distributions for the individual cues and then take the product for the optimal estimate.

Figure 2.7 shows the result. The left panel illustrates the relationship between focal distance and relative distance for a given amount of blur in the retinal image,  $P(z_o, d|c)$ . The middle panel shows the relationship between distance and perspective cues:  $P(z_o, d|p)$ . For two objects in the scene—one at the focal distance and one at another distance—one can estimate the ratio of distances to the objects from perspective. For instance, perspective cues may reveal the slant and tilt of the ground plane, and then the position of the objects along that plane would reveal their relative distances from the observer [Sedgwick, 1986]. The

variance of  $P(z_o, d|p)$  depends on the reliability of the available perspective cues: lower variance when the cues are highly reliable. The right panel shows the combined distribution derived from the products of the distributions in the left and middle panels. By combining information in blur and perspective, the model can now estimate absolute distance. We use the median of the product distribution as the depth estimate.

In summary, blur by itself provides little information about relative or absolute distance, and perspective cues by themselves provide little information about absolute distance. But the two cues in combination provide useful information about both distances. This constitutes our model of how blur is used in images like Figure 2.2 to provide an impression of absolute distance.

### 2.3.3 Impact on Previous Findings

As we said earlier, vision scientists have generally concluded that blur is a weak depth cue. Three reasons have been offered for its ineffectiveness. It is useful to evaluate them in the context of the model.

1. Blur does not indicate the sign of a distance change: that is, it does not by itself specify whether an out-of-focus object is nearer or farther than an in-focus object. It is evident in Figure 2.6 and Eq. 2.2 that a given amount of blur can be caused by an object at a distance shorter or longer than the distance of the focal plane. The model in Figure 2.7 makes clear how the sign ambiguity can be solved. The perspective distribution is consistent with only one wing of the blur distribution, so the ambiguity is resolved by combining information from the two cues.
2. The relationship between distance and blur is dependent on pupil size. When the viewer's pupil is small, a given amount of blur specifies a large change in distance; when the pupil is large, the same blur specifies a smaller change. There is no evi-

dence that humans can measure their own pupil diameter, so the relationship between measured blur and specified distance is uncertain. The model shows that distance can still in principle be estimated even with uncertainty about pupil size. The uncertainty only reduces the precision of depth estimation.

3. The visual system's ability to measure changes in retinal-image blur is limited, so small changes in blur may go undetected [Mather and Smith, 2002]. Blur discrimination is not well characterized, so we have not yet built corresponding uncertainty into the model. Doing so would yield higher variance in the blur distributions in Figure 2.6 and the left panel of Figure 2.7, much like the effect of uncertainty due to pupil diameter.

Thus, the model shows how one can in principle estimate distance from blur despite uncertainties due to sign ambiguity, pupil diameter, and blur discrimination. Furthermore, this estimation does not require that the focal distance be known beforehand, that more than one image recorded with different focal distances be available, or that the camera have a specially designed aperture.

### 2.3.4 Perspective Cues

The model depends on the reliability of the relative-distance information provided by perspective. In an image like the urban scene in Figure 2.2, linear perspective specifies relative distance quite reliably, so the variance of  $P(z_o, d|p)$  is small. As a consequence, the product distribution has low variance: *i.e.*, the estimates of absolute and relative distance are quite precise. In an image of an uncarpentered scene with objects of unknown size and shape, perspective and other pictorial cues would not specify relative distance reliably, and the variance of  $P(z_o, d|p)$  would be large. In this case, the product distribution would also have high variance and the distance estimates would be rather imprecise. Thus, the ability to estimate depth from blur is quite dependent on the ability to estimate relative distance



from perspective or other pictorial cues. We predict, therefore, that altering perceived size by manipulating blur will be more effective in scenes that contain rich perspective cues than it will be in scenes with weak perspective cues.

We have also assumed that perspective cues convey only relative-distance information. In fact, many images also contain the cue of familiar size, which conveys some absolute-distance information. We could incorporate this into the model by making the perspective distribution in Figure 2.7(b) two-dimensional with different variances horizontally and vertically. We chose not to add this feature to simplify the presentation and because we have little idea of what the relative horizontal and vertical variances would be. It is interesting to note, however, that familiar size may cause the pattern of blur to be less effective in driving perceived scale. Examples include photos with real people, although even those images can appear to be miniaturized if sufficient blur is applied appropriately.

### **2.3.5 Recovering Focal Distance in Practice**

The model can be implemented to estimate the focal distance  $z_0$  used to create a given image. First, the blur circle  $c_1$  and relative distance  $d$  are estimated at several locations in the image. Then, assuming some values for parameters  $A$  and  $s_0$ , Eq. 2.2 can be used to calculate  $z_0$ . Compiling the  $z_0$  estimates from all the example points provides a marginal distribution of estimates of the focal distance (Figure 2.9). The median of the marginal distribution may then be interpreted as the final estimate of  $z_0$ , with the variance of that distribution indicating the estimate's reliability. If the blur and depth information are either difficult to measure or not properly matched, the reliability will be low, and the blur in the image will have less impact on the visual system's estimate of the distance and size of the scene.

## 2.4 Estimating Distance in Images with Manipulated Blur

Our model predicts that the visual system estimates absolute distance by finding the focal distance that is most consistent with the blur and perspective in a given image. If the blur and perspective are consistent with one another, accurate and precise distance estimates can be obtained. We explored this notion by applying the procedure from Section 2.3.5 to images with three types of blur: (1) blur that is completely consistent with the relative distances in a scene (*consistent-blur condition*), (2) blur that is mostly correlated with the distances (*vertical-blur-gradient condition*), and (3) blur that is uncorrelated with the distances (*horizontal-blur-gradient condition*).

Fourteen scenes from GoogleEarth were used. Seven had a large amount of depth variation (skyscrapers) and seven had little depth variation (one- to three-story buildings). The camera was placed 500m above the ground and oriented down  $35^\circ$  from earth-horizontal. The average distance from the camera to the buildings in the centers of each scene was 785m.

### 2.4.1 Applying Blur Patterns

The consistent-blur rendering technique described in Section 2.2.2 was used. The diameters of the simulated camera apertures were 60.0, 38.3, 24.5, 15.6, and 10.0m. The unusually large apertures were needed to produce blur consistent with what a human eye with a 4.6mm pupil would receive when focused at 0.06, 0.09, 0.15, 0.23, and 0.36m, respectively. Figures 2.8(b) and (c) show example images with simulated 24.5m and 60m apertures. The vertical blur gradient was, as stated, aligned with the distance gradient. It was generated using the technique described in Section 2.2.2. Figures 2.8(d-e) are examples. The horizontal blur gradients employed the same blur magnitudes as the vertical gradients, but were orthogonal to the distance gradients. Figures 2.8(f-g) are examples.

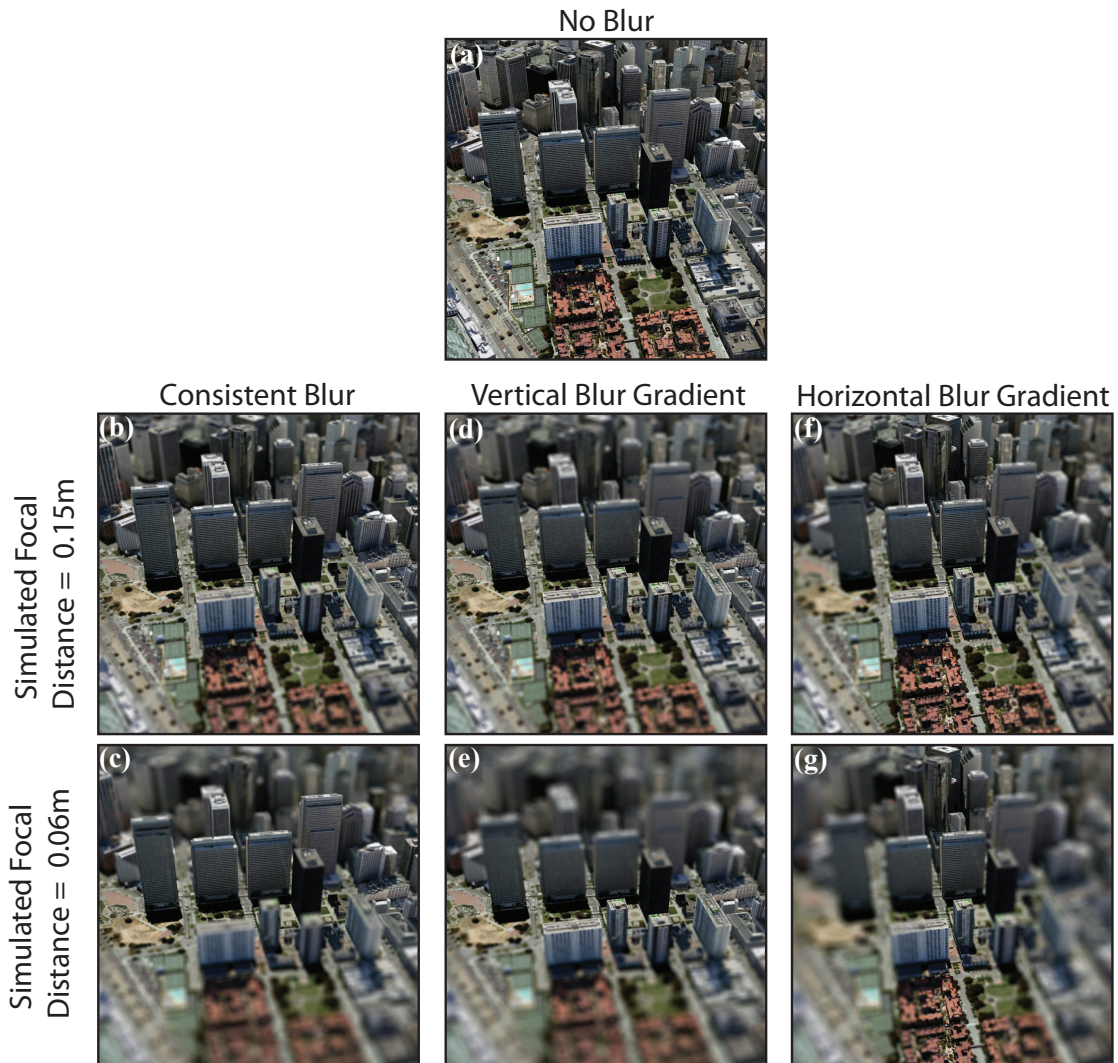


Fig. 2.8: The four types of blur used in the analysis and experiment: (a) no blur, (b-c) consistent blur, (d-e) linear vertical blur gradient, and (f-g) linear horizontal blur gradient. Simulated focal distances of 0.15m (b,d,f) and 0.06m (c,e,g) are shown. In approximating the blur produced by a short focal length, the consistent-blur condition produces the most accurate blur, followed by the vertical gradient, the horizontal gradient, and the no-blur condition. Original city images and data from GoogleEarth are copyright Terrametrics, SanBorn, and Google.

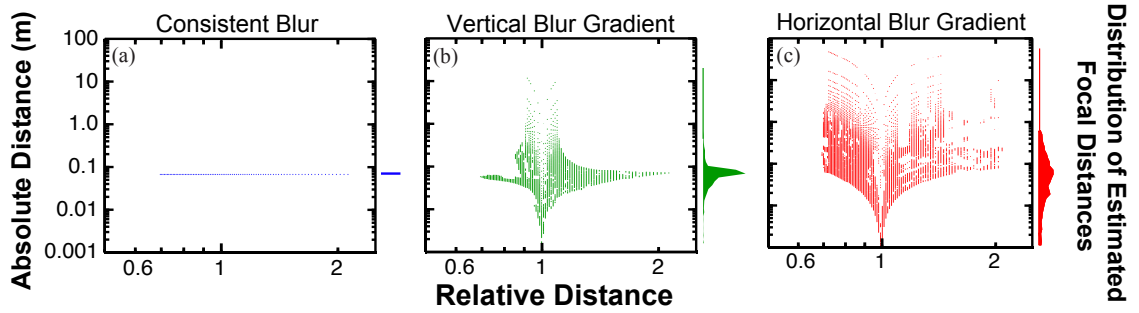


Fig. 2.9: Determining the most likely focal distance from blur and perspective. Intended focal distance was 0.06m. Each panel plots estimated focal distance as a function of relative distance. The left, middle, and right panels show the estimates for consistent blur, vertical blur gradient (mostly consistent), and horizontal blur gradient (inconsistent), respectively. The first step in the analysis is to extract the relative-distance and blur information from several points in the image. The values for each point are then used with Eq. 2.2 to estimate the focal distance. Each estimate is represented by a point. Then all of the focal distance estimates are accumulated to form a marginal distribution of estimates (shown on the right of each panel). The data from a consistent-blur rendering most closely matches the intended focal distance, resulting in extremely low variance. Though the vertical blur gradient incorrectly blurs several pixels, it is well correlated with the relative distances in the scene, so it too produces a marginal distribution with low variance. The blur applied by the horizontal gradient is mostly uncorrelated with relative distance, resulting in a marginal distribution with large variance and therefore the least reliable estimate.

## 2.4.2 Calculating Best Fits to the Image Data

To predict the viewers’ response to each type of blur, we applied the procedure in Section 2.3.5 to each image:

1. We selected pixels in image regions where blur would be most measurable, namely areas containing high contrast, by employing the Canny edge detector [Canny, 1986]. The detector’s parameters were set such that it found the subjectively most salient edges. We later verified that the choice of parameters did not affect the model’s predictions.
2. Recovering relative distance and blur information:
  - (a) Relative distances in the scene were recovered from the video card’s z-buffer while running GoogleEarth. These recovered distances constitute the depth map. In our perceptual model, these values would be estimated using perspec-

tive information. The z-buffer, of course, yields much more accurate values than a human viewer would obtain through a perspective analysis. However, our primary purpose was to compare the predictions for the three blur types. Because the visual system's ability to measure relative distance from perspective should affect each prediction similarly, we chose not to model it at this point. We can add such a process by employing established algorithms [Brillault-O'Mahony, 1991; Coughlan and Yuille, 2003].

- (b) For the consistent-blur condition, the depth map was used to calculate the blur applied to each pixel. For the incorrect blur conditions, the blur for each pixel was determined by the applied gradients.
3. To model human viewers, we assumed  $A=4.6\text{mm}$  and  $s_0=17\text{mm}$ . We assumed no uncertainty for  $s_0$  because an individual viewer's eye does not vary in length from one viewing situation to another. We then used Eq. 2.2 to calculate  $z_0$  for each pixel.
4. All of the estimates were combined to produce a marginal distribution of estimated focal distances. The median of the distribution was the final estimate.

The results for the example images in Figures 2.8(c), (e), and (g) are shown in Figure 2.9. The other images produced quantitatively similar results.

### 2.4.3 Predictions of the Model

First consider the images with consistent blur. Figure 2.9(a) shows the focal-distance estimates based on the blur and relative-distance data from the image in Figure 2.8(c). Because the blur was rendered correctly for the relative distances, all of the estimates indicate the intended focal distance of 0.06m. Therefore, the marginal distribution of estimates has very low variance and the final estimate is accurate and precise.

Next consider the images where the blur is mostly consistent with depth, due to the verti-

cal blur gradient. Figure 2.9(b) plots the blur/relative-distance data from the vertical-blur image in Figure 2.8(e). The focal-distance estimates now vary widely, though the majority lie close to the intended value of 0.06m. This is reflected in the marginal distribution to the right, where the median is close to the intended focal distance, but the variance is greater than in the consistent-blur case. We conclude that vertical blur gradients should influence estimates of focal distance, but in a less compelling and consistent fashion than consistent blur does. Although it is not shown here, scenes with larger depth variation produced marginal distributions with higher variance. This result makes sense because the vertical blur gradient is a poorer approximation to consistent blur as the scene becomes less planar.

Now consider the images with the imposed horizontal blur gradient. In these images, the blur is uncorrelated with the relative distances in the scene, so focal-distance estimates are scattered. However, the overall range of blur values and depth values are consistent with each other, which makes the median of the marginal distribution similar to those obtained with consistent blur and the vertical gradient. Critically, the variance of the distribution is much greater. The model predicts, therefore, that the horizontal gradient will have the least influence on perceived distance.

The analysis was also performed on images rendered by simulated tilt-shift lenses. The amount of tilt was chosen to reproduce the same maximum blur magnitude as the vertical blur gradients. The marginal distributions for the tilt-shift images were essentially identical to those for the vertical-blur-gradient images. This finding is consistent with the observation that linear blur gradients and tilt-shift lenses produce very similar blur magnitudes (Figure 2.5) and presumably similar perceptual effects.

These examples show that the model provides a useful framework for predicting the effectiveness of different types of image blur in influencing perceived distance and size. The horizontal-gradient results also highlight the importance of accounting for distance variations before applying blur.

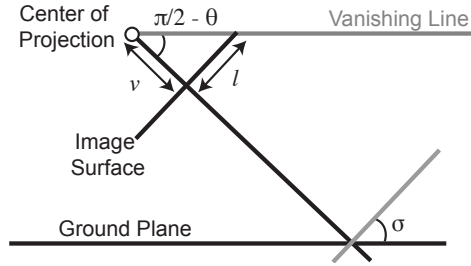


Fig. 2.10: Schematic of variables pertinent to the semi-automated blurring algorithm. Here, the image surface is equivalent to the monitor surface, and  $v$  and  $l$  are in units of pixels.  $\sigma$  indicates the angle between the ground plane's surface normal and the imaging system's optical axis. Refer to Algorithm 1 for details on how each value can be calculated from an input image. (Adapted from Okatani and Deguchi 2007)

#### 2.4.4 Algorithm

Our model predicts that linear blur gradients can have a strong effect on the perceived distance and size of a scene. But we used carefully chosen rendering parameters to produce the blur gradient images: the maximum blur magnitude was the same as that produced by a consistent-blur image taken at the intended focal length, and the tilt of the lens and the orientation of the gradient were perfectly aligned with the distance gradient in the scene. To simplify the application of the model to images, we developed a semi-automated algorithm that allows the user to load a sharply rendered image, indicate a desired perceived distance and size, and then apply the appropriate blur gradient to achieve that outcome.

We implemented the algorithm for scenes that are approximately planar globally, but it could be extended to scenes that are only locally planar. The user first sets the desired focal distance  $z_0$  and the viewer's pupil diameter  $A$ . To simulate the human eye,  $s_0$  is set to 0.017m [Larsen, 1971]. Next, the slant  $\sigma$  and tilt  $\tau$  of the planar approximation to the scene are estimated using one of two methods (slant and tilt are respectively the angle between the line of sight and surface normal, and the direction of the slant relative to horizontal). If the scene is carpentered, like the cityscape in Figure 2.11(a), we use the technique in Algorithm 1, originally described by Okatani and DeGuchi [2007]. If the scene is uncarpentered, like the landscape in Figure 2.11(b), then a grid is displayed over the image. The viewer uses

a mouse to rotate the grid until it appears to be parallel to the scene. The orientation of the grid yields  $\sigma$  and  $\tau$ . Parameters  $l$  and  $v$  (defined in Figure 2.10) are recovered from the settings (in our case, from OpenGL) that were used to render the grid on-screen. Finally, Algorithm 2 determines the amount of blur assigned to each pixel, then creates the final image.

```

prompt user to select two pairs of lines that are parallel in the original scene;
 $p_1$  = intersection of first line pair;
 $p_2$  = intersection of second line pair;
 $p_c$  = center of image;
 $vanLine$  = line connecting  $p_2$  and  $p_1$ ;
 $v = \sqrt{|p_{1_y}p_{2_y} + p_{1_x}p_{2_x}|}$ ;
 $l$  = distance between  $p_c$  and  $vanLine$ ;
 $\sigma = \pi/2 - \text{atan}(l/v)$ ;
 $\tau$  = angle between  $vanLine$  and the image's horizontal axis;

```

**Algorithm 1:** Determining  $\sigma$  and  $\tau$  from Parallel Lines

Figure 2.11 shows two example images produced by our algorithm. The scene is carpentered in panel (a), so the parallel-line-selection option was employed. Panel (b) shows the output of the algorithm, with an intended focal distance of 0.06m. The scene is not carpentered in Figure 2.11(c), so the user aligned a grid to be parallel to the predominant orientation of the scene. The resulting blurred image in panel (c) was designed to simulate a focal distance of 0.50m.

The algorithm provides an effective means for blurring images in post-processing, thereby changing the perceived distance and size of a scene. The semi-automated technique frees the user from the calculations needed to create nearly correct blur, and produces compelling results for images ranging from urban scenes to landscapes. Its effectiveness is supported by our model's predictions, which in turn were validated by the following psychophysical experiment.



```

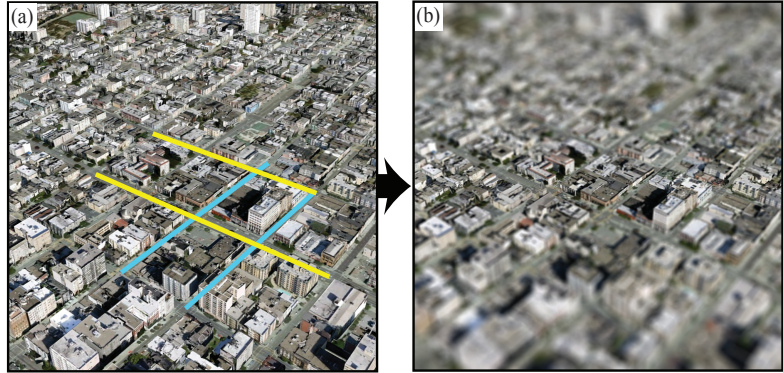
slantAxis = line that passes through  $p_c$  and is rotated  $\tau$  degrees from vertical;
for each pixel in the image do
    | distance = distance from pixel to  $p_c$ , projected onto slantAxis;
    |  $\epsilon = \text{atan}(\text{distance}/v)$ ;
    | if pixel is closer to observer than  $p_c$  then
    | |  $\epsilon$  is negative;
    | else
    | |  $\epsilon$  is positive;
    | end
    |  $\text{relativeDistance} = \cos(\sigma) * \cos(\epsilon) / \cos(\sigma + \epsilon)$ ;
    |  $\text{blurDiameterInRadians} = 2 * \text{atan}(|A * s_0 / z_0 * (1 - 1/\text{relativeDistance})| / 2 / s_0)$ ;
    |  $\text{blurDiameterInPixels} = \text{round}(2 * v * \tan(\text{blurDiameterInRadians} / 2))$ ;
end
finalImage = image composed of black pixels;
currentBlurDiameter = 0;
while currentBlurDiameter  $\leq \max(\text{blurDiameterInPixels})$  do
    | tempImage = image composed of every pixel with blurDiameterInPixels ==
    | currentBlurDiameter;
    | tempImageBlurred = convolve tempImage with cylindrical blur kernel of diameter
    | currentBlurDiameter;
    | finalImage = finalImage + tempImageBlurred;
    | currentBlurDiameter++;
end
return finalImage;

```

**Algorithm 2:** Calculating and Applying Blur to Image

Slant-estimation Technique:  
Parallel Lines

Intended Focal Distance:  
0.06m



Slant-estimation Technique:  
Manual Grid Alignment

Intended Focal Distance:  
0.50m

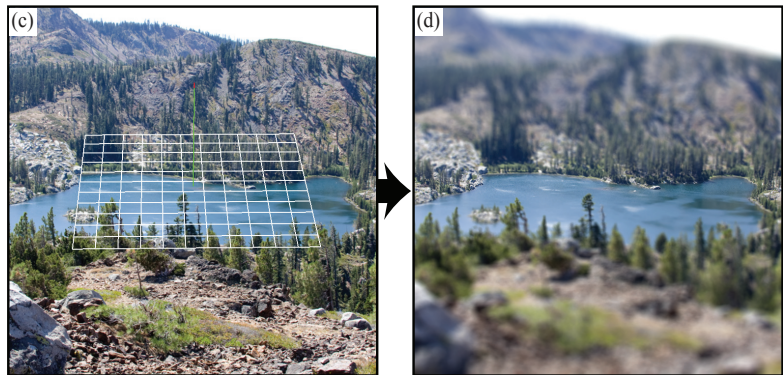


Fig. 2.11: Input and output of the semi-automated blurring algorithm. The algorithm can estimate the blur pattern required to simulate a desired focal length. It can either derive scene information from parallel lines in a scene or use manual feedback from the user on the overall orientation of the scene. (a) Two pairs of parallel lines were selected from a carpentered scene for use with the first approach. (b) The resulting image once blur was applied. Intended focal distance=0.06m. (c) A grid was manually aligned to lie parallel to the overall scene. (d) The blurred output designed to simulate a focal distance of 0.50m.

## 2.5 Psychophysical Experiment

We examined how well the model's predictions correspond with human distance percepts. We were interested in learning two things: (1) Do human impressions of distance accurately reflect the simulated distance when defocus blur is applied to an image, and (2) How accurately must defocus blur be rendered to effectively modulate perceived distance?

### **2.5.1 Methods**

We used the previously described blur-rendering techniques to generate stimuli for the perceptual experiment: consistent blur, vertical blur gradient (blur mostly consistent with depth), and horizontal blur gradient (blur inconsistent with depth). An additional stimulus was created by rendering each scene with no blur. The stimuli were generated from the same 14 GoogleEarth scenes on which we conducted the analysis in Section 4.4.

Each subject was unaware of the experimental hypotheses and was not an imaging specialist. They were positioned with a chin rest 45cm from a 53cm CRT and viewed the stimuli monocularly. Each stimulus was displayed for 3 seconds. Subjects were told to look around the scene in each image to get an impression of its distance and size. After each stimulus presentation, subjects entered an estimate of the distance from a marked building in the center of the scene to the camera that produced the image. Distances were entered in units of feet or meters using a numeric keypad. There were 224 unique stimuli, and each stimulus was presented seven times in random order for a total of 1568 trials. The experiment was conducted in four sessions of about one hour each. At the end, each subject was debriefed with a series of questions, including how they formulated their responses and whether the responses were based on any particular cues in the images. If the debriefing revealed that the subject did not fully understand the task or had based answers on strategies that circumvented the phenomenon being tested, his or her data were excluded.

### **2.5.2 Results**

Ten subjects participated, but the data from three were discarded. Two of the discarded subjects revealed in debriefing that they had done the task by estimating their height relative to the scene from the number of floors in the pictured buildings and had converted that height into a distance. They said that some scenes “looked really close,” but described a

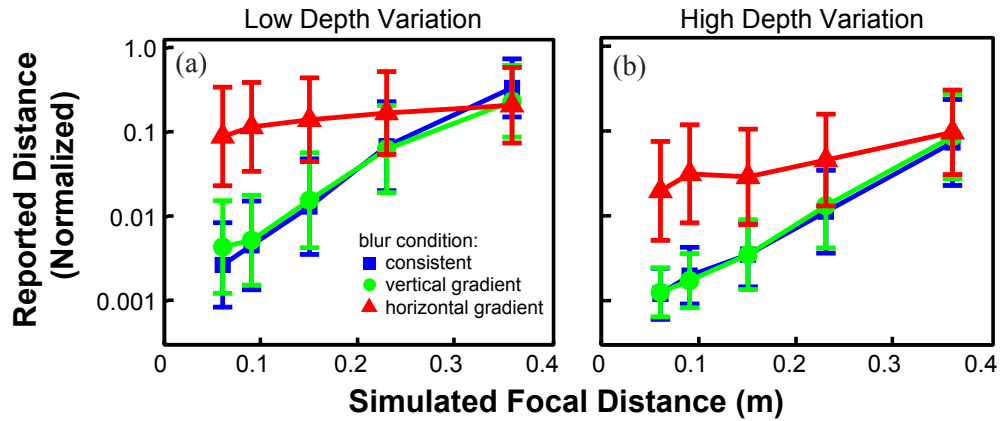


Fig. 2.12: Results of the psychophysical experiment averaged across the seven subjects. Panels (a) and (b) respectively show the data when the images had low and high depth variation. The type of blur manipulation is indicated by the colors and shapes of the data points. Blue squares for consistent blur, green circles for vertical blur gradient, and red triangles for horizontal blur gradient. Error bars represent standard errors. Individual subject data are included in the appendix.

conversion that scaled the perceived size up to the size of a real building. The third subject revealed that she had estimated distance from the amount of blur by assuming that the camera had a fixed focal distance and therefore anything that was blurrier had to be farther away.

Figure 2.12 shows the results averaged across the seven remaining subjects, with the left and right panels for the low- and high-depth-variation images, respectively. (Individual subject data are available in the appendix.) The abscissas represent simulated focal distance (the focal distance used to generate the blur in the consistent-blur condition); the values for the vertical and horizontal blur gradients are those that yielded the same maximum blur magnitudes as in the consistent-blur condition. The ordinates represent the average reported distance to the marked object in the center of the scene divided by the average reported distance for the no-blur control condition. Lower values mean that the scene was seen as closer and therefore presumably smaller.

All subjects exhibited a statistically significant effect of blur magnitude [3-way, repeated-measures ANOVA,  $F(5,30) = 13.8, p < 0.00001$ ], reporting that the marked object appeared smaller when the blur was large. The effect of magnitude was much larger in the consistent-

blur and vertical-blur-gradient conditions than in the horizontal-gradient condition, so there was a significant effect of blur type [ $F(3,18) = 14.7, p < 0.00001$ ]. There was a tendency for the high-depth-variation scenes to be seen as closer, but for blur magnitude to have a larger effect for the low-depth-variation scenes [ $F(1,6) = 2.27, p = 0.18$  (n.s.)].

All of the scenes in the experiment were oriented so the distances receded toward the top of the image. Therefore, we are using the data from the vertical-blur-gradient and horizontal-blur-gradient conditions to generally apply to blur that is mostly consistent or not consistent with the depths in a scene, respectively. The results show that perceived distance in human viewers is influenced by the pattern and magnitude of blur just as the model predicts. Consistent blur and aligned-linear-gradient blur (which is used in our semi-automated algorithm) yield systematic and predictable variations in perceived distance. Linear gradients that are not aligned with distance information yield a much less systematic variation in perceived distance.

## **2.6 Discussion**

### **2.6.1 Validity of Assumptions in the Model**

Our representation of the depth information conveyed by retinal-image blur was an approximation to information in the human visual system. Here we discuss four of our simplifying assumptions.

First, we represented the eye's optics with an ideal lens free of aberrations. Image formation by real human eyes is affected by diffraction due to the pupil, at least for pupil diameters smaller than 3mm, and is also affected by a host of higher-order aberrations including coma and spherical aberration at larger pupil diameters [Wilson et al., 2002]. Incorporating diffraction and higher-order aberrations in Figure 2.7(a) would have yielded

greater retinal-image blur than shown for distances at or very close to the focal distance: The trough in the blur distribution would have been deeper. The model estimates absolute distance from image regions with a wide range of relative distances, not just distances near the focal plane. Therefore, if the image contains a sufficient range of relative distances, the estimates are unaffected by the simplifying assumptions about the eye's optics. Additionally, monochromatic aberrations have been shown to cause different blur for defocused objects in front and behind the retina [Wilson et al., 2002]. This phenomenon could make it easier to determine whether a given observation of blur comes from an object closer or farther than the focal plane. The impact on our model would be asymmetric weighting of the distribution for relative distances less than or greater than one in Figure 2.7(a).

Second, we assumed that the visual system's capacity to estimate depth from blur is limited by the optics of retinal-image formation. In fact, changes in blur magnitude smaller than 10% are generally indiscriminable [Mather and Smith, 2002]. If we included this observation, the marginal distributions in Figure 2.9 would have larger variance than the ones shown, but the medians (and therefore the distance estimates) would be little affected.

Third, we assumed that the eye's optics are fixed. In fact, the optical power of the eye varies continually due to adjustments of the shape of the crystalline lens, a process called accommodation. Accommodation is effected by commands sent to the muscles that control the shape of the lens. Those commands are a cue to distance, albeit a variable and weak one [Wallach and Norris, 1963; Fisher and Ciuffreda, 1988; Mon-Williams and Tresilian, 2000]. In viewing real scenes, accommodation turns blur into a dynamic cue that may allow the visual system to glean more distance information than we have assumed. However, the inclusion of accommodation into our modeling would have had little effect because the stimuli were images presented on a flat screen at a fixed distance, so the changes in the retinal image as the eye accommodates did not mimic the changes that occur in real scenes. We intend to pursue the use of dynamic blur and accommodation using volumetric displays

that yield a reasonable approximation to the relationship in real scenes (*e.g.*, [Akeley et al., 2004]).

Fourth, our model assumes that the viewer was fixating at the center of each image, which was rendered sharply. In fact, each observer was instructed to look around the entire image, resulting in unnatural patterns of blur on the retina. In natural viewing, the object of fixation is usually in focus. Our study was primarily concerned with picture perception, and the subjects' task was to make judgments about the camera that produced the images, which may minimize the impact of the incorrect blur. But if the incorrect blur patterns were included in the model, the variance and reliability of the distance estimates would increase and decrease, respectively.

### **2.6.2 Algorithm Effectiveness**

The predictions of the model and the results of the psychophysical experiment confirmed the effectiveness of the linear blur gradients applied by our algorithm. Specifically, linear gradients and consistent blur were similarly effective at modulating perceived distance and size. Currently, the algorithm is most effective for planar scenes, and it is only useful for adding blur and making sharply focused scenes appear smaller. Further development could incorporate regional slant estimation to increase its accuracy for scenes with large distance variations, and include a sharpening algorithm to reduce blur and make small scenes appear larger.

### **2.6.3 Impact on Computer Graphics**

The model we developed explains a number of phenomena in which blur does or does not affect perceived distance and scale. Some of these phenomena occur in photography, cinematography, and graphics, so the model has several useful applications.

### Application: Natural Depth of Field

One of the main points of our analysis is that there is an appropriate relationship between the depth structure of a scene, the focal distance of the imaging device, and the observed blur in the image. From this relationship, we can determine what the depth of field would be in an image that looks natural to the human eye. Consider Eq. 2.2. By taking advantage of the small-angle approximation, we can express blur in angular units:

$$b_1 = 2 \tan^{-1}\left(\frac{c_1}{2s_0}\right) \approx \frac{c_1}{s_0} \quad (2.5)$$

where  $b_1$  is in radians. Substituting into the Eq. 2.2, we have:

$$b_1 = \left| \frac{A}{z_0} \left(1 - \frac{1}{d}\right) \right| \quad (2.6)$$

which means that the diameter of the blur circle in angular units depends on the depth structure of the scene and the camera aperture and not on the camera's focal length [Kingslake, 1992].

Suppose that we want to recreate the pattern of blur in the photograph that a human viewer would experience if they were looking at the original scene. We photograph the scene with a conventional camera and then have the viewer look at the photograph from its center of projection. The depth structure of the photographed scene is represented by  $z_0$  and  $d$ , different  $d$ 's for different parts of the scene. We can recreate in the photograph the blur pattern the viewer would experience when viewing the real scene. We can do this by adjusting the camera's aperture to the appropriate value. From Eq. 2.6, we simply need to set the camera's aperture to the same diameter as the viewer's pupil. If a viewer looks at the resulting photograph from the center of projection, the pattern of blur on the retina would be identical to the pattern created by viewing the scene itself. Additionally, the perspective information would be correct and consistent with the pattern of blur. This creates what



we call “natural depth of field.” For typical indoor and outdoor scenes, the average pupil diameter of the human eye is 4.6mm (standard deviation is 1mm). Thus to create natural depth of field, one should set the camera aperture to 4.6mm, and the viewer should look at the resulting photograph with the eye at the photograph’s center of projection. We speculate that the contents of photographs with natural depth of field will have the correct apparent scale.

### **Application: Simulating Extreme Scale**

We described how to manipulate blur to make small scenes look large and large scenes look small. These effects can be achieved by altering the blur pattern in post-processing, but they can also be achieved by using cameras with small or large apertures. Specifically, if the focal distance in the actual scene is  $z_0$ , and we want to make it look like  $\hat{z}_0$  where  $\hat{z}_0 = mz_0$ , Eq. 2.2 implies that the camera’s aperture should be set to  $A/m$ . In many cases, doing this is not practical because the required aperture is too restrictive. If the aperture must be quite small, the amount of light incident on the image plane per unit time is reduced, and this decreases the signal-to-noise ratio. If the aperture must be very large, it might not be feasible with a physically realizable camera. Consequently, it is very attractive to be able to adjust the blur pattern in post-processing in order to produce the desired apparent scale.

The demonstrations we showed here made large scenes look small. Figure 2.8(a) shows an image that was recorded with a focal length of  $\sim 800\text{m}$  and a pinhole ( $A \approx 0$ ) aperture. We made the image look small in panels (b) and (c) by simulating in post-processing focal lengths of 0.19 and 0.06m. We could have created the same images by recording the images with cameras whose aperture diameters were 24.5 and 60m, respectively, but this is clearly not feasible with a conventional camera. It is much more attractive to achieve the same effects in post-processing, and our algorithm shows how to do this.

Our analysis also applies to the problem of making small scenes look large. If we have an

image recorded with a particular aperture size, we want to reduce the blur in the image in the fashion implied by Figure 2.8. Our algorithm could potentially be used to determine the desired blur kernel diameter for each region of the image. However, implementation of this algorithm would require some form of deconvolution, which is prone to error [Levin et al., 2007].

### **Application: Using Other Depth Cues to Affect Perceived Scale**

Besides blur, several other cues are known to affect perceived distance and scale. It is likely that using them in conjunction with blur manipulation would strengthen the effect on perceived distance and scale.

Atmospheric attenuation causes reductions in image saturation and contrast across long distances [Fry et al., 1949], and serves as the motivation for the commonly used rendering method known as depth cueing. Not surprisingly, more saturated objects tend to be perceived as nearer than less saturated objects [Egusa, 1983]. In fake miniatures, the saturation of the entire image is often increased to strengthen the impression that the scene is close to the camera and small [Flickr, 2009]. Conversely, a reduction in saturation helps create the impression that the scene is far away and therefore large. It is also not surprising given the atmospheric effects, that high-contrast textures are perceived as nearer than low-contrast textures [Ichihara et al., 2007; Rohaly and Wilson, 1999]. We suspect, therefore, that adjusting image contrast would also be useful in creating the desired apparent size.

In principle, the acceleration of an object due to gravity is a cue to its absolute distance and size [Saxberg, 1987; Watson et al., 1992]. When an object rises and falls, the vertical acceleration in the world is constant. Thus, distant objects undergoing gravity-fed motion have slower angular acceleration across the retina than close objects. Human viewers are quite sensitive to this, but they use a heuristic in which objects generating greater retinal speed (as opposed to acceleration) are judged as nearer than objects generating slower reti-

nal speeds [Hecht et al., 1996]. This effect has been used in cinematography for decades: practitioners display video at slower speed than the recorded speed to create the impression of large size [Bell, 1924; Fielding, 1985].

The accommodative state of the viewer’s eye can affect perceived distance [Wallach and Norris, 1963; Fisher and Ciuffreda, 1988; Mon-Williams and Tresilian, 2000]. Thus, if an image is meant to depict a small scene very close to the eye, the impression of small size might be more convincing if the image is actually viewed from up close. Accommodation is, however, a weak cue to depth, so effects of actual viewing distance may be inconsistent.

### **Application: Blur and Stereo Displays**

Stereo image and video production has recently gained a great deal of attention. Several studios are producing films for stereo viewing and many movie houses have installed the infrastructure for presenting these 3D movies [Schiffman, 2008]. Additionally, many current-generation televisions are capable of stereo display [Chinnock, 2009]. It is therefore timely to consider blur rendering in stereo content.

Disparity, the cue being manipulated in stereo images, has the same fundamental geometry as blur [Schechner and Kiryati, 2000]. Disparity is created by the differing vantage points of two cameras or eyes, while blur is created by the differing vantage points of different positions in one camera’s or eye’s aperture. Consider two pinhole cameras with focal lengths  $f$ . The distance from the camera apertures to the film planes is  $s_0$ , and the distance between apertures is  $I$ . The cameras are converged on an object at distance  $z_0$  while another object is presented at  $z_1$ . The images of the object at  $z_0$  fall in the centers of the two film planes and therefore have zero disparity. The images of the object at  $z_1$  fall at different locations  $X_L$  and  $X_R$  creating a disparity of:

$$\delta = X_L - X_R = I \left( \frac{s_0}{z_0} \right) \left( 1 - \frac{1}{d} \right) \quad (2.7)$$

where  $d = z_1/z_0$  and  $1/s_0 = 1/f - 1/z_0$ . The connection to image blur is clear if we replace the aperture  $A$  in Eq. 2.2 with two pinholes at its edges. Then two images of  $z_1$  would be formed and they would be separated by  $c_1$ . From Eqs. 2.2 and 2.7, for cameras of focal lengths  $f$ ,

$$c_1 = (A/I)|\delta| \quad (2.8)$$

Thus, the magnitudes of blur and disparity caused by a point in a 3D scene should be proportional to one another. In human vision, the pupil diameter is roughly 1/12 the distance between the eyes [Spring and Stiles, 1948], so the diameters of blur circles are generally 1/12 the magnitudes of disparities. Because the geometries underlying disparity and blur are similar, this basic relationship holds for the viewing of all real scenes.

How should the designer of stereo images and video adjust blur and disparity? Because of the similarity in the underlying geometries, the designer should make the disparity and blur patterns compatible. To produce the impression of a particular size, the designer can use the rule of thumb in Eq. 2.8 to make the patterns of blur and disparity both consistent with that size. To do otherwise is to create conflicting information that may adversely affect the intended impression. Two well-known phenomena in stereo images and video—the cardboard cut-out effect [Yamanoue et al., 2000; Meesters et al., 2004; Masaoka et al., 2006] and puppet-theater effect [Yamanoue, 1997; Meesters et al., 2004]—may be caused by blur-disparity mismatches.

The blur-rendering strategy should depend, however, on how people are likely to view the stereo image. Consider two cases. 1) The viewer looks at an object at one simulated distance and maintains fixation there. 2) The viewer looks around the image, changing fixation from one simulated distance to another.

In the first case, the designer would render the fixated object sharply and objects nearer and farther with the blur specified by Eq. 2.8. By doing so, the blur and disparity at the viewer's

eyes are matched, yielding the desired impression of 3D structure. The blur rendering can guide the viewer's eye to the intended object [Kosara et al., 2001; DiPaola et al., 2010]. This would be common for entertainment-based content.

In the second case, the rule of thumb in Eq. 2.8 should probably not be applied. In real scenes, the viewer who looks at a nearer or farther object converges or diverges the eyes to perceive a single image and accommodates (*i.e.*, adjusts the eye's focal power) to sharpen the retinal image. If the rule of thumb were applied in creating a stereo image, objects at simulated distances nearer or farther than the sharply rendered distance would be blurred. The viewer who looks nearer or farther would again converge or diverge the eyes, but the newly fixated object would be blurred on the retina no matter how the viewer accommodated, and this would yield a noticeable and annoying conflict. On the other hand, if the image was rendered sharply everywhere, the viewer would experience a sharp retinal image with each new fixation, and this would probably be a more desirable outcome. As we discuss in Chapter 3, this notion could be important in applications like medical imaging, where the viewer may need to look at features throughout the scene.

Thus, blur rendering in stereo images should probably be done according to the rule of thumb in Eq. 2.8 when the designer intends the viewer to look at one simulated distance, but should not be done that way when the viewer is likely to look at a variety of distances.

## **2.7 Future Work Investigating Blur as a Distance Cue**

We described how the normal relationship between blur and accommodation is disrupted in pictures. Learning more about the consequences of this disruption will be valuable to the development of advanced displays in which the normal relationship can be approximated. It will also be crucial to learn how cues like disparity affect the visual system's use of blur as a depth cue.

# CHAPTER 3

## STEREOSCOPIC DISPLAYS IN MEDICAL IMAGING: A REVIEW

Because stereo displays can convey more accurate depth information than non-stereo displays, they can potentially benefit medicine in areas ranging from diagnostics to training. In particular, stereoscopic imaging could a) make complicated shapes and structures easier to identify, b) aid the user in assessing large data sets, and c) through its integration in virtual reality modules, decrease the cost of training and health care in general. However, the technology has drawbacks, including cost, equipment complexity, and the current necessity for eyewear. Stereo displays therefore need to demonstrate a clear advantage over existing techniques if they are to be widely adopted. In this chapter, we will discuss the advantages and disadvantages of existing medical applications of stereo displays. Tips and guidelines on the optimal use of stereo displays are then provided. We provide a concise guide for medical practitioners who want to assess the potential benefits of stereo displays before adopting them.

### **3.1 Depth and Displays**

To differentiate between stereoscopic and conventional displays, we begin with a discussion of the visual depth cues afforded by each technology.

### 3.1.1 Monocular Depth Cues

Any conventional display can present monocular depth cues. These are cues to depth that are useful to the visual system, even if they are acquired by only one eye. The depth cues most relevant to our discussion are perspective projection, occlusion, familiar size, shading, and the motion-based cues known as structure from motion and motion parallax.

Perspective projection, or how a 3D scene is projected onto a 2D image plane, offers several depth cues. Objects that are farther away from the imaging device are projected to smaller sizes than objects that are close. Parallel lines that recede into the scene (such as the lines on a road) usually project to converging lines in an image. Portions of the parallel lines that are spaced farther apart on the image are closer to the observer than portions that are spaced closer together. Perspective projection also produces texture gradients. Texture patches, like those that make up wallpaper, will project to have larger and more widely-spaced elements on the image when they are closer to the imaging device. The orientation of texture elements can also reveal 3D shape [Palmer, 1999].

Occlusion refers to objects blocking the view to each other. It is useful for ordering objects in depth, but it is generally less useful as a metric depth cue.

Familiar size refers to the estimation of scale and relative depth based on familiarity with the imaged objects in real life. For instance, seeing a human being in an image gives the viewer a reference point for the sizes and depths of other scene elements [Palmer, 1999].

The shading of a scene can be useful for recovering 3D shape information. Technically, shading can only be useful if one knows the position of the light source. However, it has been shown that the human visual system assumes most scenes are lit from above. This assumption allows one to constrain the problem and use shading to recover the 3D shapes of objects, even if the light source is not visible [Sun and Perona, 1998; O'Shea et al., 2008].

“Structure from motion” refers to the recovery of 3D shape information from the view of a rotating object. As the object rotates, the visual system is able to integrate information from the multiple viewpoints to assemble an impression of 3D shape. Motion parallax occurs when the viewer or imaging device is translated relative to a scene. The relative movement of the objects in the scene across the imaging plane provides information on their relative distances. Close objects will move more across the image than far objects. A familiar example is the view out a car window: bushes appear to rush by, but distant mountains seem stationary. The key distinction between structure from motion and motion parallax is whether the object or the viewer is the one moving.

### **3.1.2 Stereoscopic Depth**

Stereoscopic depth arises from the different images acquired by each eye. It is similar in concept to motion parallax, but instead of relying on a continually translating view of the image, the viewer is provided with two stationary, offset views. The source of depth information is the pattern of disparities between the two eyes’ images. The disparity for a given object is defined as the difference in the projected positions of that object in the two retinas. The human visual system is very sensitive to these disparities. Typical adult stereoacuity is around 15arcsec [Simons, 1981], which is equivalent to fixating on an object 1.0m away and distinguishing it in depth from an object only 1.2mm farther away. For objects close to fixation, disparity provides the most accurate depth information of any visual cue. Stereoscopic displays take advantage of this cue by presenting one unique image to each eye. If done correctly, the disparities provide the viewer with a much more accurate sense of depth and 3D shape than monocular cues alone. Stereo movies take advantage of the heightened impression of depth to produce more compelling visuals and to draw the viewer into the story. Meanwhile, stereo displays have also found use in a variety of scientific fields, including earthquake-epicenter visualization [Wells, 2002].



## **3.2 Medical Applications of Stereoscopic Displays**

Here we discuss several areas of medicine in which stereo displays have been implemented and outline their benefits and drawbacks.

### **3.2.1 Diagnostics**

The key benefits afforded by stereo displays to radiology are the abilities to a) perceptually separate features and tissue layers in an image and b) immediately recognize shapes without having to acquire extra images or rotate a 3D model. Both of these benefits would appear to be crucial for certain diagnostics, but new techniques must be proven to provide a significant improvement before practitioners will adopt them. With that in mind, we concentrate on the diagnostic applications of ophthalmic imaging, mammography, and vascular imaging.

#### **Ophthalmic Imaging**

Ophthalmic imaging deals with structures with multiple, thin tissue layers. It can be diagnostically relevant to differentiate between these layers, which may be difficult or impossible from only monocular cues. Two recent studies investigated the benefits of stereo imaging for the diagnosis of glaucoma [Abramoff et al., 2007; Bergua et al., 2009]. Abramoff et al. found that stereo images facilitated image segmentation and glaucoma diagnosis, while Bergua et al. noted that stereo photography provided improvements in the visualization of the neuroretinal rim shape, slope of the inner wall of the optic nerve, and cup depth. Another study applied stereo imaging to the diagnosis of diabetic retinopathy [Ahmed et al., 2006]. For that pathology, retinal thickening, elevated neovascularization and retinal detachment—all phenomena with subtle shape characteristics—are critical to the diagnosis. The addition of disparity information in stereo imagery improved the detectability of those

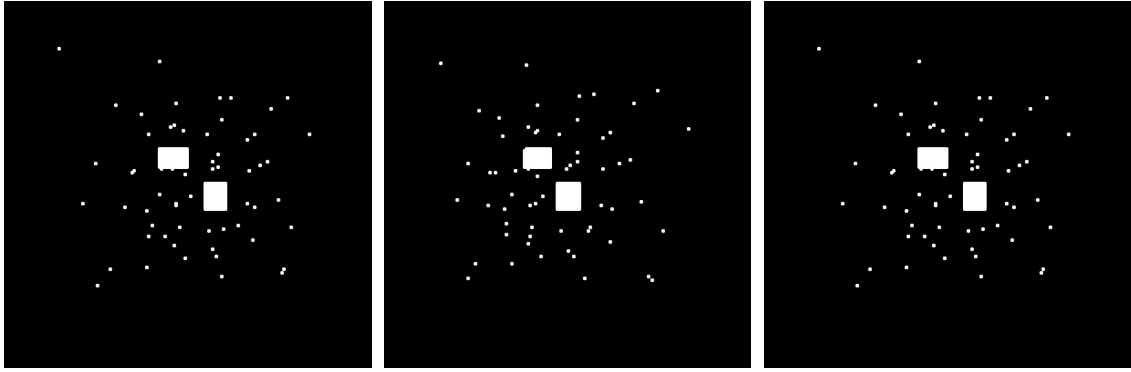


Fig. 3.1: Stereogram of a field of dots with floating patches. To view the stereogram, hold the page at arm's length and either a) cross your eyes so the left eye is directed at the middle image and the right eye is directed at the left image, or b) point the left eye at the middle image and the right eye at the right image. Without fusing the images, it is impossible to tell which of the patches is closer. However, once the images are fused, their depth ordering becomes apparent, as well as the 3D positions of the random dots.

processes, and the researchers noted improvements in specificity and sensitivity compared to non-stereo imaging.

## **Mammography**

Traditional monocular depth cues are of little to no use in mammography. Mammograms produce projections of all the tissue between the x-ray source and the collector, which precludes occlusion as a depth cue. Additionally, features in different layers can mask each other and become difficult or impossible to differentiate [Getty et al., 2008]. Familiar size is not present, nor any perspective-projection information. These problems are critical, as the specific 3D layout of microcalcifications are relevant to potential malignancy [Chan et al., 2005]. One solution is to acquire posterior-anterior and lateral projections to get orthogonal views of the tissue. However, it has been shown that it is more cost-effective and efficient to acquire a stereo pair of posterior-anterior images, as they allow the 3D shape of the tissue to become apparent, even if the two images are simply viewed side-by-side on a conventional display [Kelsey et al., 1982]. However, viewing the two images simultaneously on a stereo display provides the greatest benefit. This notion is apparent in

Figure 3.1. Each of the panels is an image of the same scene, composed of randomly placed dots and two rectangular patches placed at different depths. Viewed individually, one cannot recover which of the patches is closer, due to the lack of useful monocular depth cues. One could compare the positions of the patches within each image to determine their relative placement. However, if two of the images are fused as a stereo pair (see caption), the relative-depth information becomes immediately apparent. Microcalcification detection in stereo mammography is analagous to this example, and three recent studies have found significant benefit from the technology. In a large breast-cancer study involving roughly 1500 cases, Getty and colleagues [2008] found that stereoscopic imaging provided significant increases in true positives and decreases in false negatives. Anecdotally, the radiologists involved in that study also felt that the stereo images could be read more quickly than non-stereo images. In another study, receiver-operating-characteristic curves were created for radiologists observing mammograms with and without stereo displays [Chan et al., 2005]. Stereo was found to improve sensitivity for estimating the likelihood of malignancy and detecting microcalcifications and other masses. Finally, Tanaka et al. [2007] showed that stereo imaging helped observers separate nodes in lymphoscintigraphy that were originally considered to be one mass, based on non-stereo images. Thus, stereoscopic imaging clearly benefits mammography by providing useful depth information where monocular cues are weak and nearly useless.

### **Vascular Imaging**

Vascular imaging involves complicated, branching structures whose interconnections can be difficult to distinguish using conventional imaging techniques. For instance, if two branching structures superimpose in the image, it can be impossible to differentiate their connections. Stereo was applied to the problem in a study motivated by stereotactic surgery and renal-blood-vessel imaging [Sollenberger and Milgram, 1993]. A network of line seg-

ments was displayed to subjects, who performed path-tracing and line-segment-connectivity tasks. The display either showed a static, non-stereo image, a rotating image of the line network, or a stereo image. The rotating images, which provided structure from motion, and stereo images were the most useful for accomplishing the task, followed by the static, non-stereo image. Stereo is also useful with real medical data, as it has been shown to help radiologists isolate vessels at different depths [Serra et al., 1997] and diagnose the 3D vascular encasement of tumors [Sekiguchi et al., 1996]. New advances in photoacoustic imaging can produce 3D scans of larger tissue volumes, including the breast, at higher resolution than ultrasound or optical imaging without ionizing radiation [Xu and Wang, 2006]. Thus, the technology lends itself well to stereoscopic viewing. Finally, stereo appears to provide easy depth perception and a slight improvement in catheterization safety during angiography [Moll et al., 1998]. However, the radiologists involved in that study only gave the technology a lukewarm reception. The authors of the study suggest that the radiologists' familiarity with existing techniques may have biased them against the technology. While previous visualization methods for angiography are inherently more difficult to use, the radiologists in the study had become accustomed to the technology and knew how to work with it precisely and effectively. Introducing stereo imaging likely made it necessary for them to re-familiarize themselves with the equipment, which seemed inefficient. More studies will therefore be necessary to establish to whether the benefits of stereo imaging to angiography can overcome familiarity with existing technology. This outcome seems likely, as the ability of stereo to separate vessels in depth offers a significant advantage over current techniques.

### **3.2.2 Medical Training**

Stereo displays can aid medical training in anatomy and surgery by providing better spatial understanding of anatomical features. This is because the spatial abilities of students have

been correlated with their performance as surgeons [Luursema et al., 2006]. Put differently, students who can easily recognize and visualize 3D structure are more likely to perform well in surgery. Therefore, adding more shape information through stereo cues may facilitate students' learning of anatomy and improve their likelihood of success as practitioners. Meanwhile, the lack of familiar-size and perspective cues in surgeries without direct views of the tissue (e.g., laparoscopy) can make the procedures difficult for novices. As we discuss below, adding stereo images can ameliorate this problem and significantly improve the performance of novice surgeons on basic tasks.

Over the past decade, several researchers have studied the use of 3D anatomical models as pedagogical agents. Two early studies investigated the learning of wrist-bone anatomy. The authors concluded that canonical views are most important for learning anatomy, with little to no benefit afforded by multiple views (i.e., rotations) of a 3D model [Garg et al., 2001, 2002]. However, those studies did not provide continuous control of model rotation, which possibly disrupted structure from motion, and the carpals lie mostly in a 2D arrangement, which limits the added benefit of 3D visualization [Luursema et al., 2006]. Indeed, a later study found that students who are given control of the view of 3D ear models demonstrated better understanding of anatomical relationships than those presented only with static images [Nicholson et al., 2006]. This disparity is likely due to the added sense of shape provided by structure from motion. Another pair of studies examined whether stereo presentation and user interaction aided anatomical understanding. They found no clear benefit of stereo in an identification task. However, stereo improved performance on a task that required accurate perception of the relative positions of organs within an image. The results are logical because identification of different organs is likely possible from monocular cues like color and gross shape, while more subtle aspects like differences in depth and relative position would be easier to glean from stereo information.

In surgery, many errors are due to “misinterpretation of local anatomy compounded with

inadequate procedural knowledge” [Tendick et al., 1998]. Therefore, if students are provided with pedagogical agents that help them understand the anatomy and practice the procedure, fewer errors and generally better performance should ensue. Indeed, Tuggy and colleagues [1998] found that the use of a (non-stereo) virtual-reality sigmoidoscopy trainer resulted in faster insertion times, greater percentage of the colon visualized, and a general improvement in exam quality. Meanwhile, Ilgner et al. [2007] studied the use of stereo displays while teaching microscopic surgery. Microscopic surgery is difficult to teach in the operating room because only one surgeon can view the stereo microscope at a time. To address this issue, Ilgner et al. attached a stereo camera to the microscope. The junior surgeon performed the procedure, while the instructor and other medical students simultaneously viewed the surgical site on a stereo display. Instruction was facilitated, as the senior surgeon could easily direct the junior surgeon’s actions with the shared view. The stereo display also proved helpful for the junior surgeons to orient themselves in the unfamiliar setting of the patient’s anatomy.

Other studies have looked at the use of stereo displays for teaching laparoscopic [Taffinder et al., 1999; Patel et al., 2007] and pelvic procedures [Teveaarai et al., 2000]. Both laparoscopic studies found that stereo improved the ability of novices to perform basic laparoscopic tasks, with fewer errors and improved accuracy. Teveaarai and colleagues tested non-surgeons, non-celioscopic surgeons, and celioscopic surgeons with basic tasks on a pelvic trainer with and without stereo imaging. Stereo improved performance for each group. Most interestingly, stereo brought the performance of non-celioscopic surgeons up to the level of celioscopic surgeons using non-stereo imaging. This result supports the notion that the additional shape and depth information provided by stereo imaging may partially compensate for lack of familiarity with specific anatomy, or at least accelerate training.

Prystowsky et al. [1999] investigated whether a stereoscopic VR trainer helped students

learn intravenous (IV) catheter placement. The students demonstrated decreased insertion times and improved insertion quality over the course of the VR training, but their improvement did not translate to physical reality. The finding emphasizes that a full evaluation of stereo displays must include the benefits to practical knowledge and skills. Prystowsky and colleagues noted that the absence of translational improvement was likely due to the limited time allowed with the trainer and lack of feedback during training. The latter item will likely prove to be a crucial element in any successful VR training module, regardless of whether or not stereo imaging is included. Effective teaching involves a dialogue between educator and student, and that relationship should be supported, rather than replaced, by VR training. For instance, Johnson et al. [1998] have proposed a general VR system with built-in pedagogical agents with the intent to provide interactive training that monitors students, gives feedback, and answers questions.

Stereo imaging and virtual reality hold promise for improving medical training, but their usefulness may vary across students. Luursema et al [2006] investigated how stereo and interactive models benefited subjects with varying spatial abilities. Interestingly, they found that students with lower spatial abilities benefited more than students with high spatial abilities, which has important implications for teaching medical students. As stated earlier, spatial ability has been correlated with surgical skills. If stereo displays are implemented more widely (and are also available in the operating room), then the performance of professionals with low spatial abilities may improve more than for professionals with high spatial abilities. While this claim requires further work before it can be validated, stereo imaging and VR systems have the potential to offer more intuitive, accessible methods for teaching anatomy and surgery. Additionally, while the usefulness of cadavers and animals may never be fully matched, VR systems can provide an easily reusable, more cost-effective and humane alternative [Luursema et al., 2006; Pieper et al., 1991].

### 3.2.3 Surgical Planning

We have addressed the use of stereoscopic imaging for diagnostics, teaching, and the actual execution of surgical techniques, which are all integral to the planning phase of a medical procedure. To date, there is not much work on the benefits of stereo presentation specific to surgical planning. So in this section, we concentrate on 3D modeling and virtual reality, rather than stereoscopic imaging. Generally, the use of stereo visualization can be assumed to enhance the benefits of 3D modeling outlined below.

Regardless of a surgeon's expertise, proper preparation for each procedure is critical to success. Part of that preparation involves knowing the patient's specific anatomy and any complicating factors (e.g., presence of eloquent structures near a glioma). Medical images serve this purpose. Modern medical imaging data, such as computed tomography (CT) and magnetic resonance imaging (MRI), usually come in the form of a series of 2D slices. Interpreting such slices can be difficult, as one must mentally combine them to get a sense of the 3D anatomy. However, 3D models can be automatically generated from 2D-slice data, which offloads the 3D integration process from the viewer's brain to the computer. One can then focus on the pathology or surgical procedure at hand. For instance, one study on the planning of lung cancer resection found that 3D models decreased resection planning time by 30%, increased the accuracy of predicted resectability by 20%, and decreased the subjectively defined workload by 50%, compared to using 2D slices [Hu, 2005]. Other researchers have investigated the benefits of 3D data for surgeries on the liver [Fishman et al., 1996; Wigmore et al., 2001], heart [Hemminger et al., 2005], gastric cancer [Lee et al., 2003], mandible [Xia et al., 2000], and brain [Kikinis et al., 1996]. Gering et al. [2001] demonstrated how 3D models can help one route tumor resections to avoid eloquent structures in the brain. Meanwhile, Kikinis et al. [1996] presented neurological case studies, including venous malformations and intra- and extra-axial tumors, that used 3D modeling. The surveyed neurosurgeons reported that the 3D models contributed substantially



to planning several aspects of the procedures, including optimal craniotomy and cortisectomy sites, proximity of the targets to critical tracts and nuclei, the structural relationships between vascular structures, and the position of cranial nerves. The transition from 2D slice data to 3D models made anatomical evaluations much easier in each of these studies. Adding stereoscopic visualization would go another step further and make the process even more intuitive.

Along with the benefits of 3D models for procedure planning, virtual reality allows one to practice a surgery ahead of time, which can make the actual procedure feel much more familiar [Rosahl et al., 2006] and permit one to practice responses to rare complications [Burt, 1995]. Also, stereo visualization can make virtual reality more immersive. If the simulations can be made to appear more real-to-life, then it will be easier to translate the skills they teach to the operating room.

### **3.2.4 Laparoscopy**

Laparoscopy is a surgical technique in which thin instruments are inserted into a patient's body to remotely image the surgical site and perform the procedure. Commonly labelled "minimally invasive surgery," laparoscopy produces fewer complications and faster recovery times than traditional surgery [Tendick et al., 2000]. However, the indirect viewing can make it difficult to accurately manipulate the instruments [Taffinder et al., 1999]. The lack of familiar-size and perspective cues, particularly for novices, sometimes drives the surgeon to touch parts of the anatomy with the surgical instruments to gain points of reference, which is very inefficient (Figure 3.2). And for cholecystectomy, one cause of bile-duct injury can be misinterpretation of the anatomy [Tendick et al., 1998]. Additional depth information, namely through stereoscopic imaging, should improve the quality of laparoscopic procedures, at least for novice surgeons not yet well acquainted with the tools. Indeed, one study specifically credited stereo viewing for ameliorating the handicap of remote viewing



Fig. 3.2: Endoscopic images during bowel surgery on an animal model. Note the lack of familiar sizes, especially for novices unfamiliar with this view of the patient's anatomy. The lack of monocular visual cues makes it difficult to estimate distances between objects, especially in depth. Original figure by Hu, et al. [2009]

during laparoscopy [Taffinder et al., 1999].

The conclusions concerning the usefulness of stereo viewing have been mixed. Hofmeister et al. [2001] reported that only 50% of studies found an improvement in laparoscopic-task performance due to stereoscopic viewing. The metrics for those task-based studies included error rate and time to completion. Additionally, even in the cases in which stereo viewing provided an improvement, only 1/3 of those interviewed reported that they preferred it over non-stereo viewing [van Bergen et al., 1998]. Here we explore the specific benefits provided by stereo viewing, and the potential causes for the wide range of results.

Differences in surgical expertise appears to produce the varied outcomes. Laparoscopic experts have more refined skills, including the ability to glean meaningful depth information from monocular images, and therefore benefit less from stereo cues. Indeed, three of the early studies on stereo laparoscopy with negative results used subjects already acquainted with monocular viewing [Crosthwaite et al., 1995; Hanna et al., 1998; McDougall et al., 1996]. One study included a non-surgeon technician, but that person had experience teaching endoscopic techniques [Crosthwaite et al., 1995]. Another study separated low and high-experience surgeons based on whether they had completed fewer or more than 100 laparoscopic techniques [McDougall et al., 1996]. That number seems large and arbitrary, so any differences in ability between the groups are not evident. Finally, the last study used

surgeons who had completed a structured course on endoscopy and previously performed at least 20 laparoscopic cholecystectomies [Hanna et al., 1998]. Thus, the benefits for novice surgeons with little familiarity with laparoscopy cannot be determined from their results. If the subjects in these studies had already developed methods for performing non-stereo laparoscopy and become well-acquainted with that process, it seems logical that they would show little or no benefit from a new, unfamiliar viewing modality.

In fact, viewer experience appeared to be critical for the impact of stereo viewing in a number of laparoscopic studies in which novices demonstrated greater benefit than experts [Taffinder et al., 1999; Patel et al., 2007; Blavier and Nyssen, 2009]. As stated previously, it has also been shown that stereo viewing brings the performance of non-celioscopic surgeons closer to the level of celioscopic experts [Tevearai et al., 2000]. Additionally, the usefulness of stereo appears to depend on the difficulty of the task. Mueller-Richter et al. [2003] used subjects who were unfamiliar with endoscopic techniques and found that stereo laparoscopy did not result in improved performance in example laparoscopic tasks. However, the chosen tasks were relatively simple, including placing grains of corn into a bowl, placing pins in drilled holes, and moving a ring along a bent metal wire. It is conceivable that the depth cues available by a monocular display were sufficient to perform the tasks, so there was no additional benefit from stereo viewing. Jourdan et al. [2004] found that stereo viewing was most beneficial for more difficult tasks, like tying a knot and threading a needle. Meanwhile, the simple tasks they tested, including passing a rope between two graspers, cutting a ribbon at designated spots, and capping a needle, were not difficult and showed comparatively little advantage from stereo viewing. Other researchers have found similar results [Pietrabissa et al., 1994].

Thus, while it may seem that the benefits of stereoscopic laparoscopy are in doubt, a close look at previous studies reveals that significant improvements can be found with novice surgeons and complex tasks. Surgeons already well acquainted with non-stereo laparoscopy

are less likely to benefit from stereo viewing.

### **3.2.5 Telesurgery**

In laparoscopic surgery, the surgeon is removed from direct viewing of the surgical site. Telesurgery uses robotics and cameras to increase this separation, so a surgeon can perform operations from any location with a high-speed data connection. The implications are great for medicine in developing nations and the military; experts could perform surgery without traveling long distances or stepping into the battlefield. Patients could theoretically be treated sooner and have greater odds of survival. Compared to laparoscopy, telesurgery also aims to restore the degrees of freedom and force and tactile feedback typically lost by the instruments used in minimally invasive surgery. Previously, the benefits of stereoscopic imaging in telesurgery were limited by image quality and the ability to accurately produce stereoscopic images [Blavier et al., 2006]. However, the da Vinci Robot System has shown progress in this regard [Blavier et al., 2006; Byrn et al., 2007]. In the study by Byrn et al., subjects with varying levels of surgical experience used the da Vinci system to perform motor tasks that simulated basic laparoscopic surgical skills with and without stereo visualization. Subjects consistently completed tasks faster and with less error under the stereo conditions. Times to completion were reduced by 34-46% and errors were reduced by 44-66% [Byrn et al., 2007]. The inclusion of stereo visualization seems to reduce the apparent separation between the surgeon and the patient during telesurgery, and will likely be a critical feature in future systems.

### **3.2.6 Augmented-reality Surgery**

In augmented reality (AR), one's view of the world is supplemented with additional information, typically via a head-mounted display (HMD). In medical settings like neuro-

surgery, needle biopsies, and orthopedic surgery [Maurer et al., 2001], the technology can be used to superimpose previously acquired data over a patient's anatomy, essentially giving the surgeon "x-ray vision" [Nikou et al., 2000]. Another key advantage is the evaluation of imaging data in the context of the actual patient, from any point of view [Wendt et al., 2003]. The hope is that the additional information will lead to more accurate, less invasive procedures. For minimally invasive procedures, the surgeon could keep his or her eyes trained only on the patient, without needing to focus on a separate display to monitor instruments.

Several issues must be addressed to make AR useful without being cumbersome. The virtual objects must appear stationary in the real world, which demands accurate head tracking and high refresh rates. The deformation of soft anatomy, such as the brain, must be accurately monitored during surgery to correctly transform previously acquired imaging data [Edwards et al., 2001]. The virtual anatomy must also convincingly appear to be positioned within the patient. Stereoscopic visualization has been shown to help achieve this goal. Wendt et al. used a head phantom to test an AR system that used an HMD with stereoscopic overlays of MRI data. That study found that stereo cues and surface from motion allow subjects to accurately visualize targets within the phantom. Rosenthal et al. [2002] used a head-mounted display (HMD) to overlay a stereoscopic view of an ultrasound image onto the patient's body to facilitate placement of a biopsy needle. They used a breast phantom and found that the system produced smaller mean deviations between the needle placement and the desired target, compared to the standard process of using an ultrasound probe to find the target and attempting to insert the needle in the image plane. Wacker et al. [2006] verified that stereo-based AR systems using MR-imaging data allowed for accurate biopsy retrieval in a porcine model. Stereo AR has also been tested clinically with ear, nose and throat surgery and neurosurgery. Although some registration errors were encountered, the system improved the outcome in several cases [Edwards et al., 2001]. Generally, stereo-based AR systems have significant potential for enabling faster, more accurate med-

ical procedures.

Unfortunately, even with stereoscopic imaging, it can still be difficult to make the virtual anatomy appear to be within the patient. The issue is due to the strong monocular cue of occlusion—anything projected onto the skin will tend to appear to be outside the body [Maurer et al., 2001]. Several solutions are under investigation. One approach that appears to work well is to completely occlude a portion of the patient with a simulated “window” into the body [Bajura et al., 1992; Fuchs et al., 1998]. However, this solution prevents the surgeon from monitoring features on the skin surface. Maurer et al. found that wire-frame or dot-based models permit the visualization of both real and simulated anatomy with correct impressions of depth [Maurer et al., 2001]. Other solutions may involve displaying the virtual objects with partial transparency [Zhai et al., 1996]. More research will be required to determine the optimal balance between visualization of the virtual and real surfaces.

Despite some technical issues to be resolved, augmented reality’s ability to incorporate scan data and other medical information into a physician’s view of the patient should make diagnostics more intuitive and facilitate faster, more accurate surgeries.

### **3.3 Summary of Benefits**

Here we summarize the positive aspects of stereo imaging to provide hints about its effective implementation.

The disparities available in stereo images provide shape and depth information that make it easier to differentiate similar tissues and identify complex 3D structures. This observation is particularly important in settings like mammography and laparoscopy, where familiar-size and perspective cues can be weak or absent. In mammography, the use of x-ray projection confounds occlusion cues, which makes it impossible to order features in depth.

Similarly, in vascular imaging, vessels may become superimposed and difficult to differentiate [Sollenberger and Milgram, 1993]. Stereo cues help reduce this issue and make 3D structures more apparent. Since the 3D structure of tissue is relevant to pathology, the results are higher specificity and sensitivity in diagnostics. In laparoscopy and telesurgery, the view of the tissue is indirect and unfamiliar. Experienced surgeons can become accustomed to this setting and develop very efficient endoscopic techniques. However, novice surgeons can have difficulty gathering their bearings and recognizing features. This issue is partially avoided with stereo visualization, which provides strong spatial cues about the surgical site. Additionally, it becomes easier to correctly perceive the relative positions and orientations of the instruments. For simple tasks, the benefits of stereo are not significant. However, for tasks like knot-tying, stereo appears to cause a marked improvement in accuracy and time to completion [Byrn et al., 2007].

Novice surgeons can also benefit from the use of stereo visualization during training. Spatial abilities have been correlated with surgical abilities, and stereo presentation has been shown to improve subject performance on medically based spatial tasks [Luursema et al., 2006]. Therefore, it is logical that the use of stereo imaging may help students with lower spatial abilities better understand human anatomy and in turn perform well as surgeons.

Augmented reality combines live views of the patient with additional information, such as previously acquired medical images. When stereo visualization is included, the surgeon can have the illusion of looking into the patient at his or her anatomy. This can make procedures like incisions and biopsies much more intuitive and precise, since the surgical site can be known with greater accuracy [Wacker et al., 2006].

Even with non-stereo displays, the usefulness of 3D models has been well-established. When the models are compared to the 2D scan slices on which they are based, they have been shown to dramatically decrease the workload on physicians and improve diagnostics. Fatigue has also been reported while interpreting 2D-slice data [Byrn et al., 2007], which

may be avoidable with 3D models and stereo imaging, though the latter can also be a source of visual fatigue (discussed below). Virtual reality shows promise as a way to learn new procedures, practice for rare complications, and avoid the high cost of cadavers and practice animals.

## **3.4 Summary of Drawbacks**

Stereoscopic displays have potentially significant drawbacks. Here we divide the issues in a few basic categories.

### **3.4.1 Hardware**

A basic issue is the necessity for more hardware. In camera-based imaging modalities like laparoscopy, stereo cameras are required. Binocular imaging demands slightly more memory and CPU power, as well as special displays. As processing power continues to improve, the computing requirements will not be a large issue. But a bigger concern is that most implementations of stereo medical imaging must use either shutter glasses synchronized with the display [Chan et al., 2003; Goodsitt et al., 2000; Novotny et al., 2006; van Bergen et al., 1998; Vasilyev et al., 2008; Wenzl et al., 1994] or orthogonally polarized glasses [Crosthwaite et al., 1995; Hu et al., 2009; Jourdan et al., 2004; Mueller-Richter et al., 2003; Pietrzak et al., 2006; Song and Kang, 2009] to deliver a unique image to each eye. Autostereoscopic, or “glasses-free,” stereo displays are in development, but they do not yet offer the same image quality and resolution of glasses-based implementations [Matusik and Pfister, 2004]. Thus, though bulky eyewear can be a significant detractor for medical professionals [McDougall et al., 1996], they are currently a necessity. The equipment used to capture stereoscopic images can also complicate matters. Stereo imaging can require double exposures for applications like angiography [Moll et al., 1998]. The additional



exposure to ionizing radiation is clearly undesirable. However, for stereo mammography it has been claimed that the per-image radiation doses can be almost halved without any impact on the number of visible details [Maidment et al., 2003].

### **3.4.2 Resistance to New Technology**

Resistance to new technologies may be the greatest hurdle for stereo displays in medicine. As demonstrated in the previous laparoscopic review, expert surgeons benefited the least from stereo imaging, likely as a result of tricks and techniques they had already developed to compensate for the impoverished depth perception in monocular viewing. And even in an angiographic study in which stereo provided measurable benefits, the (experienced) radiologists surveyed gave it only a lukewarm reception. The authors of the study mentioned familiarity with existing techniques as a likely contributor to these opinions [Moll et al., 1998]. Therefore, it may be difficult to convince seasoned professionals of the advantages of stereo imaging. It would be unfortunate for those opinions to override its advantages for novices, which appear to be clear and consistent.

### **3.4.3 Viewer Discomfort**

Finally, viewer fatigue, usually in the form of eye strain or headache, is a recurring issue for stereo displays. The studies reviewed above reported varying rates of fatigue in their subjects, from very rare [McDougall et al., 1996; Ilgner et al., 2006] to relatively common [van Bergen et al., 1998; Pietrabissa et al., 1994]. General discomfort can come from wearing bulky eye wear. Objects that move quickly in depth, three-dimensional artifacts due to poor rendering, and unnatural blur can also be visually fatiguing [Lambooi et al., 2009]. Recently, it was shown that a phenomenon known as the vergence-accommodation conflict is a major source of discomfort and fatigue with stereo displays (Figure 3.3). Vergence

refers to the directions the eyes are pointed: they must be pointed at the same location in space for the viewer to perceive a single image rather than double images. Accommodation refers to where the eyes are focused: they must be focused to the appropriate distance to create sharp, rather than blurred, retinal images. In natural viewing, vergence and accommodation are tightly coupled, so the visual system can maintain single and sharp images [Martens and Ogle, 1959]. For instance, when one looks far away, the eyes' lines of sight are essentially parallel and they are focused at infinity. Meanwhile, when one looks at something close, the eyes point in to verge on the object and the lenses change focus to make it appear sharp on the retina. With stereo displays, however, vergence and accommodation are usually in conflict. The eyes can converge and diverge through the simulated 3D scene to fixate on various objects, but they must remain focused on the display screen to keep the images sharp (Figure 3.3(b)). It has been shown that vergence-accommodation conflict is a primary source of viewer fatigue and discomfort with stereo displays [Hoffman et al., 2008].

However, the amount of discomfort may be reduced through consideration of the image content. Consider objects near and far from the display surface. When the eyes verge on the nearer objects, there will be a smaller conflict between where the eyes are focused (the plane of the display) and where they would be focused for a real scene (the objects). Therefore the vergence-accommodation will be smaller. So if it is possible to translate the objects of greatest interest, whether they be in diagnostic images or in a surgical site, close to the plane of the display, eye strain and headaches may be reduced.

#### **3.4.4 Stereoscopic Misperceptions**

Adjusting the 3D content is directly related to the issue of stereoscopic misperceptions. As described in Chapter 1, unlike non-stereo images [Vishwanath et al., 2005], the interpretation of shape in stereoscopic images appears to depend significantly on viewer position.

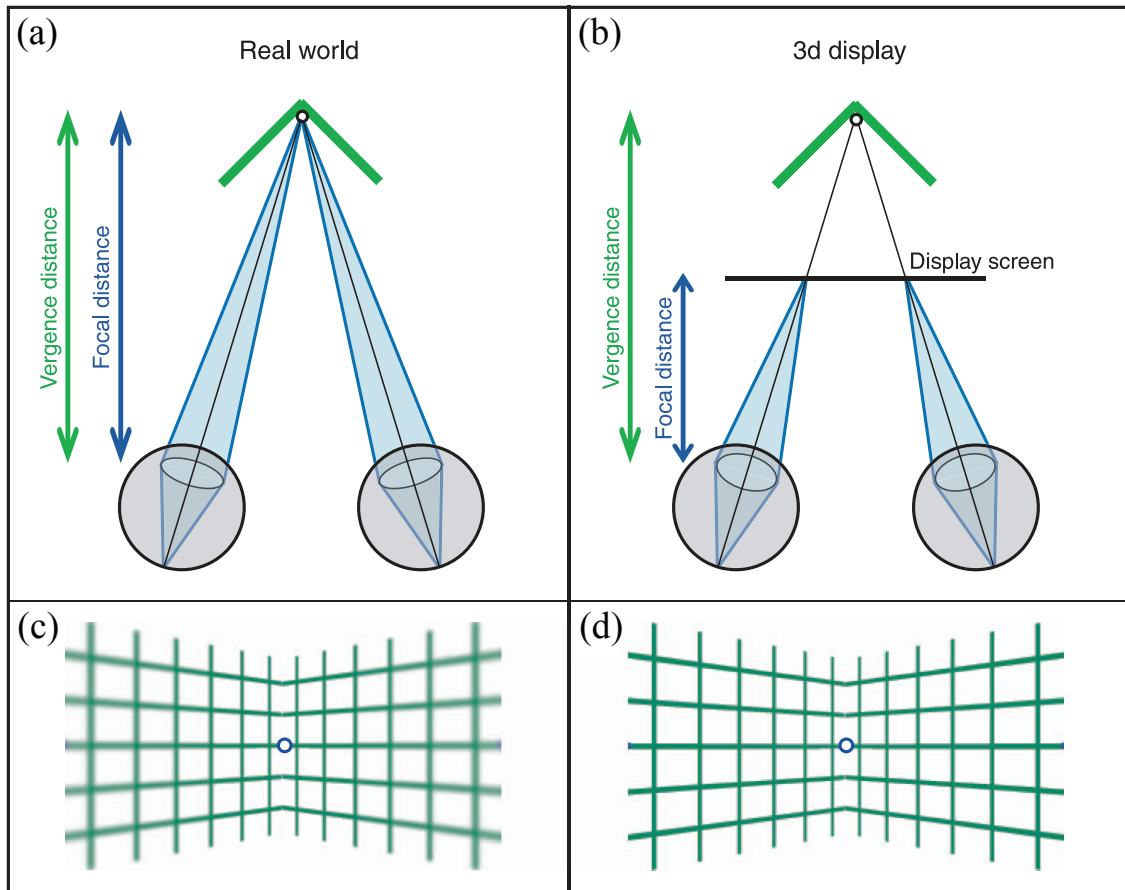


Fig. 3.3: Variations in vergence and accommodation with natural viewing and typical stereoscopic displays. (a) The eyes' vergence and accommodative states are coupled in natural viewing. Here, vergence and accommodation are both set to the far corner of of an open-hinge stimulus. The light from the edges of the hinge are physically closer to the eyes than the far corner, so they appear out of focus. (b) On a typical stereo display, vergence and accommodation are uncoupled. Vergence can vary through the 3D scene (here it is trained on the corner of the hinge), but accommodation must remain fixed on the surface of the display to keep the image sharp. Note that the entire hinge is imaged sharply, since all of the light is originating at the the surface of the display. Original figure by Hoffman, et al. [2008]

We discussed how there is technically only one “correct” viewing location for a conventional stereoscopic display, where the observer will veridically perceive the 3D contents. Additionally, the generation and display of stereoscopic content, whether it is based on computer models or stereo cameras, involve several interconnected variables that must be carefully set to avoid misperceptions. As a result, the proper capture, display, and viewing of stereo content requires significantly more effort than conventional images. That effort is probably only warranted for applications where disparity-based depth perception provides clear advantages over the already-present monocular cues.

### **3.5 Avoiding Misperceptions with Stereoscopic Displays**

The best way to ensure that a viewer correctly perceives an imaged scene is to deliver the exact same retinal images as those that would have been produced by the original scene viewed with the naked eyes. This amounts to placing the viewer’s eyes at the centers of projection (COP’s) of the images. For non-stereo images, placing the viewer’s eyes away from the COP’s does not necessarily impact their perception of the image, as long as they are viewing it binocularly. This does not apply to stereoscopic images. Refer to Chapter 1 for a detailed discussion of the variables that must be correctly set to ensure correct stereoscopic viewing. It should be noted that “correct viewing” may not always be desired, especially with medical and scientific visualization. The ability to modify an image and emphasize certain aspects can make it much easier to recover information. However, since the variables listed above interact nonlinearly and in sometimes unintuitive ways, it is important to understand how they affect the percept before adjusting them. We will limit our discussion primarily to camera orientation and stereo-image separation, as both are commonly modified in medical imaging.

Most stereoscopic displays, including those used in medicine and almost any display using

shutter glasses or polarized glasses, use a single display surface to present the left and right images. In Chapter 1, we showed that the stereo cameras' bodies should be parallel if their images will be displayed on those types of displays. This orientation is sometimes non-intuitive, and one may be tempted to use converging, or "toed-in," cameras. Several medical studies have used converging cameras [Chan et al., 2005; Getty et al., 2008; Maurer et al., 2001]. However, converging cameras produce images that are distorted if they are projected onto a typical stereo display. The images exhibit a keystone distortion (Figure 1.4) and the COP's for both the left and right images lie at the same location, centered relative to the face of the display. This setup makes it impossible for both eyes to be positioned at their correct locations, and produces misperceptions for every viewer. Parallel cameras, however, avoid the keystone distortion and permit placement of the eyes at the images' COP's. Some researchers have already begun using this setup [Ahmed et al., 2006]. The degree of misperceptions due to converging axes is not yet completely clear, so one cannot definitively claim that their use would negatively affect the outcome of medical procedures. However, there are no apparent advantages to using them, aside from ease of acquisition for some imaging hardware [Chan et al., 2005], so parallel cameras should be used whenever possible.

Any adjustments of the separation between the left and right stereo images will affect the pattern of disparities. As described earlier, disparities are the source of stereoscopic depth information, so if they are modified, the perceived shape will change as well. Increasing the stereo-image separation will compress the scene in depth and make it appear to move closer to the observer (Figure 1.2(c)). Decreasing the separation will make the scene appear to expand and move away from the observer (Figure 1.2(g)). The 3D shapes do not simply scale, but are distorted to varying degrees along the three principal axes. Therefore, applications that demand accurate spatial understanding of the entire scene should avoid adjustments to the stereo-image separation. Some applications may require one to investigate specific spatial relationships, like differentiating between layers of tissue. In those

cases, changes in the image separation may be useful.

Flipping the positions of the left and right images should always be avoided. Image flipping has been proposed as a convenient way to invert the depth in a stereoscopic image, with the rationale being that it is “easier to attend to objects seen in the foreground compared to those seen in the background” [Getty and Green, 2007]. The inversion occurs because the disparities flip sign, so features that once appeared in front of the display appear behind it and vice versa. This is the same principle as the pseudoscope, a device that flips the eyes’ images and makes convex surfaces appear concave and vice versa [Bernstein, 1910]. The technique would work for mammography, where the disparities are the primary source of depth information, though the benefits of viewing objects in the “background” versus “foreground” are questionable. However, it can have disastrous results for any images that contain perspective and occlusion cues, (e.g. laparoscopy or 3D models of the brain). While flipping the images’ positions inverts the stereo-defined depth, the monocular cues still indicate the original depth, so a conflict occurs. The resulting image is likely to appear confusing and visually fatiguing at best, and nonsensical at worst. Indeed, this is often reported when one uses a pseudoscope to look at a natural scene [Bernstein, 1910]. Therefore, stereo-image flipping should not be used.

### **3.6 Suggestions and Guidelines**

The correct interpretation of medical images is tantamount to properly treating a patient. For this reason, significant attention has been paid to the characteristics of the displays used by radiologists. Studies have investigated the various benefits and drawbacks of different technologies [Balassy et al., 2005; Sorantin, 2008], and the American College of Radiology publishes a “Practice Guideline for Digital Mammography,” which includes target values for the contrast, bit depth, and maximum luminance of medical displays [Pisano et al.,

2007]. As the previous section demonstrated, stereoscopic imaging introduces a host of new variables that must be carefully considered for optimal viewing. Here we present a list of general guidelines designed to help medical professionals use stereo displays with maximum benefit. The first three deal with misperceptions, while the last two address visual fatigue.

### **3.6.1 Suggestion 1: Parallel cameras**

As detailed in Chapter 1, converging cameras lead to image distortions and stereoscopic misperceptions when a single-surface display is used. Therefore, parallel camera bodies should be used. Occasionally, it can be beneficial to capture a scene so that a certain object will appear at the surface of the stereo display. This can be done by laterally offsetting the imaging sensor relative to the lens (Figure 1.2(b)). If the stereo images are generated using software, the process is equivalent to skewing each camera's frustum. The cameras' optical axes converge, with any object located at the point of convergence appearing at the surface of the display. Importantly, the image projections are still correct for viewing on a single-surface stereo display.

### **3.6.2 Suggestion 2: Do not flip images**

Flipping the position of stereoscopic images causes the disparity-defined depth to invert, but not the depth defined by monocular cues. It should never be used, unless the content does not contain monocular depth cues, as in mammography.

### **3.6.3 Suggestion 3: Keep eyes centered and parallel relative to display**

To avoid stereoscopic misperceptions, the viewer's eyes should be placed at the COP's of the stereo images. When a camera captures an image, its optical center is nearly always centered relative to the edges of the imaging plane. That is, the optical center lies on the central surface normal of the imaging surface. As a result, the COP for a projected image is nearly always on its central surface normal. The distance between the COP and the surface of the image can be difficult to determine on the fly, as it depends on several factors, including the display size. So as a general rule of thumb, one should try to center one's eyes relative to the display. The face should also be kept parallel to the display surface with the inter-ocular axis parallel to the top edge of the display. Rotating away from this position can cause unnatural disparity patterns and misperceptions (see Section 1.3.4). Note that these rules are most relevant for a single viewer. When multiple viewers are present, the best solution is to get everyone as close to the COPs as possible.

### **3.6.4 Suggestion 4: Minimize vergence-accommodation conflict**

Displays with near-correct focus cues are under development that minimize fatigue by maintaining the natural coupling between vergence and accommodation [Watt et al., 2005; Love et al., 2008]. However, currently available displays cannot avoid the vergence-accommodation conflict. The conflict will be worst when viewers fixate on objects that are presented away from the surface of the display, as their eyes must remain focused on the surface to keep the image sharp. One way to avoid fatigue is to present as much of the content as possible near the display surface. Aside from carefully capturing the content, one should avoid any manipulation of the images that will stretch the stereo-defined depth, such as modifications to the stereo-image separation. These suggestions are somewhat restrictive, so the individual user will need to weigh the risk of fatigue against the necessity for image manipulation.



### **3.6.5 Suggestion 5: Appropriate use of pictorial blur**

We generally suggest that medical-imaging data be rendered completely sharply with infinite depth of field. The term “depth of field” refers to the range of depths in an imaged scene that appear sharp. Anything outside the depth of field will appear blurry, with the amount of blur depending on the depth-wise distance between an object and the imaging system’s focal plane. Artistically, depth-of-field blur is useful to direct the viewer’s attention to certain objects [Cole et al., 2006; DiPaola et al., 2010; Kosara et al., 2002] and make a scene look more realistic [Hillaire et al., 2007, 2008]. We showed in Chapter 2 that blur can also act as a strong depth cue, as long as it is presented with another cue, like perspective projection. Therefore, if apparent realism is crucial, then rendering should include blur levels consistent with natural viewing. This lesson could be applied to medical-data visualization to make the content appear more real-to-life. However, blurring any part of the image makes that portion harder to evaluate. So the user would need to continually adjust the focal plane to lie on each object being evaluated to make it appear unblurred. Rendering the entire scene sharply removes this hassle. Therefore, while the inclusion of blur has artistic benefits, it typically should be omitted from medical images.

# CONCLUSIONS

It is important to understand the misperceptions that occur when viewing both conventional and stereoscopic pictures; without such an understanding, it will be very difficult to create displays and viewing situations that produce the desired 3D percept. With that in mind, we evaluated the standard model for predicting 3D percepts from stereoscopic displays. The standard model makes reasonable predictions for most, but not all image-capture and viewing settings. The settings that are problematic for the model involve rotation of the viewer's head relative to the display and the use of converging cameras for acquisition paired with a single display surface for viewing. We described findings in the vision-science literature that point to how the visual system determines 3D structure in these situations. In particular, the system uses vertical disparity as an additional signal for determining the structure.

We also investigated a monocular cue that can affect depth percepts in both stereo and non-stereo images. We showed how blur can be manipulated in images to make large things look small and small things look large. The strength of the effect stands in contrast with previous notions of blur as a weak depth cue. Our probabilistic model shows that absolute and relative distances cannot be estimated from image blur alone, but they can be estimated quite effectively in concert with other depth cues. We used this model to develop a semi-automatic algorithm for adjusting blur to produce the desired apparent scale. The results of a psychophysical experiment confirmed the validity of the model and usefulness of the algorithm.

We described how blur and stereo operate over similar domains, providing similar information about depth. A rule of thumb can be used to assure that blur and disparity specify the same 3D structure. As stereo images and video become increasingly commonplace, it will be important to learn what artistic variations from the natural relationship between blur and disparity can be tolerated.

The usefulness of stereo displays for medical tasks was also explored. It is apparent that

stereoscopic imaging will be most useful in scenarios where monocular cues, like perspective and familiar size, are not available to the viewer. The improvement in shape recognition is key. Examples include mammography, where stereo displays have been shown to improve the diagnostic abilities of radiologists. The benefits for novice surgeons working with indirect views of the surgical site are also evident, as performance can improve significantly with the use of stereo displays.

The technology also has drawbacks, such as the need for eyewear, possibility for eyestrain, and susceptibility to misperceptions when image-acquisition and viewing parameters are not properly set. Fortunately, these drawbacks can be addressed by improvements in display engineering and better understanding of the variables that affect human perception of stereoscopic content. Thus, the most important hurdle to the adoption of stereo medical displays will likely be clinical validation of the technology as superior to current techniques. While some clinical reports already exist, further work will better establish the advantages of stereo displays in practice.

# BIBLIOGRAPHY

- Abramoff, M., Alward, W., Greenlee, E., Shuba, L., Kim, C., Fingert, J., and Kwon, Y. (2007). Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Invest Ophthalmol Vis Sci*, 48(4):1665–1673.
- Agrawala, M., Beers, A. C., McDowall, I., Fröhlich, B., Bolas, M., and Hanrahan, P. (1997). The two-user responsive workbench: Support for collaboration through individual views of a shared space. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 327–332, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.
- Ahmed, J., Ward, T., Bursell, S., Aiello, L., Cavallerano, J., and Vigersky, R. (2006). The sensitivity and specificity of nonmydriatic digital stereoscopic retinal imaging in detecting diabetic retinopathy. *Diabetes Care*, 29(10):2205–2209.
- Akeley, K., Watt, S. J., Girshick, A. R., and Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Trans. Graph.*, 23(3):804–813.
- Backus, B. T., Banks, M. S., van Ee, R., and Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39(6):1143 – 1170.
- Bajura, M., Fuchs, H., and Ohbuchi, R. (1992). Merging virtual objects with the real world: seeing ultrasound imagery within the patient. *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 203–210.
- Balassy, C., Prokop, M., Weber, M., Sailer, J., Herold, C. J., and Schaefer-Prokop, C. (2005). Flat-panel display (LCD) versus high-resolution gray-scale display (CRT) for chest radiography: an observer preference study. *AJR American journal of roentgenology*, 184(3):752–6.
- Banks, M. S., Hooge, I. T. C., and Backus, B. T. (2001). Perceiving slant about a horizontal axis from stereopsis. *Journal of Vision*, 1(2):55–79.
- Barsky, B. A. (2004). Vision-realistic rendering: Simulation of the scanned foveal image from wavefront data of human subjects. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 73–81.
- Barsky, B. A., Horn, D. R., Klein, S. A., Pang, J. A., and Yu, M. (2003a). Camera models and optical systems used in computer graphics: Part I, Object-based techniques. In *Proceedings of the 2003 International Conference on Computational Science and its Applications (ICCSA'03), Montreal, Second International Workshop on Computer Graphics and Geometric Modeling (CGGM'2003)*, pages 246–255.
- Barsky, B. A., Horn, D. R., Klein, S. A., Pang, J. A., and Yu, M. (2003b). Camera models and optical systems used in computer graphics: Part II, Image-based techniques. In *Proceedings of the 2003 International Conference on Computational Science and its Applications (ICCSA'03), Montreal, Second International Workshop on Computer Graphics and Geometric Modeling (CGGM'2003)*, pages 256–265.

- Bell, J. A. (1924). Theory of mechanical miniatures in cinematography. *Transactions of the SMPTE*, 18:119.
- Bergua, A., Mardin, C. Y., and Horn, F. K. (2009). Tele-transmission of stereoscopic images of the optic nerve head in glaucoma via internet. *Telemedicine and e-Health*, 15(5):439–444.
- Bernstein, J. (1910). *The five senses of man : 91 woodcuts*. London, 8th ed. edition.
- Blavier, A., Gaudissart, Q., Cadriere, G., and Nyssen, A. (2006). Impact of 2D and 3D vision on performance of novice subjects using da vinci robotic system. *Acta Chirurgica Belgica*, 106(6):662—664.
- Blavier, A. and Nyssen, A. S. (2009). Influence of 2D and 3D view on performance and time estimation in minimal invasive surgery. *Ergonomics*, 52(11):1342–9.
- Brillault-O’Mahony, B. (1991). New method for vanishing-point detection. *CVGIP: Image Underst.*, 54(2):289–300.
- Burge, J. D., Fowlkes, C. C., , and Banks, M. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *Under Review*.
- Burt, D. (1995). Virtual reality in anaesthesia. *British Journal of Anaesthesia*, 75(4):472–480.
- Byrn, J., Schluender, S., Divino, C., Conrad, J., Gurland, B., Shlasko, E., and Szold, A. (2007). Three-dimensional imaging improves surgical performance for both novice and experienced operators using the da vinci robot system. *The American Journal of Surgery*, 193(4):519–522.
- Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679–698.
- Chan, H., Goodsitt, M., Hadjiiski, L., Bailey, J., Klein, K., Darner, K., and Sahiner, B. (2003). Effects of magnification and zooming on depth perception in digital stereomammography: An observer performance study. *Physics in Medicine and Biology*, 48(22):3721–3734.
- Chan, H. P., Goodsitt, M. M., Helvie, M. A., Hadjiiski, L. M., Lydick, J. T., Roubidoux, M. A., Bailey, J. E., Nees, A., Blane, C. E., and Sahiner, B. (2005). ROC study of the effect of stereoscopic imaging on assessment of breast lesions. *Medical Physics*, 32(4):1001–1009.
- Chinnock, C. (2009). Personal Communication.
- Cole, F., DeCarlo, D., Finkelstein, A., Kin, K., Morley, K., and Santella, A. (2006). Directing gaze in 3D models with stylized focus. *Eurographics Symposium on Rendering*, pages 377–387.
- Cook, R. L., Carpenter, L., and Catmull, E. (1987). The REYES image rendering architecture. *SIGGRAPH Comput. Graph.*, 21(4):95–102.

- Cook, R. L., Porter, T., and Carpenter, L. (1984). Distributed ray tracing. *SIGGRAPH Comput. Graph.*, 18(3):137–145.
- Coughlan, J. M. and Yuille, A. L. (2003). Manhattan world: Orientation and outlier detection by bayesian inference. *Neural Computation*, 15(5):1063–1088.
- Crosthwaite, G., Chung, T., Dunkley, P., and Shimi, S. (1995). Comparison of direct vision and electronic two-and three-dimensional display systems on surgical task efficiency in endoscopic surgery. *British Journal of Surgery*, 82(6):849–851.
- Deering, M. (1992). High resolution virtual reality. *SIGGRAPH Comput. Graph.*, 26(2):195–202.
- Diner, D. B. (1991). A new definition of orthostereopsis for 3D television. In *Proceedings of the 1991 IEEE International Conference on Systems, Man, and Cybernetics, Decision Aiding for Complex Systems*, (2):1053–1058.
- DiPaola, S., Riebe, C., and Enns, J. (2010). Rembrandt’s textural agency: A shared perspective in visual art and science. *Leonardo*, 43(2):145–151.
- Duke, P. A. and Howard, I. P. (2005). Vertical-disparity gradients are processed independently in different depth planes. *Vision Research*, 45(15):2025 – 2035.
- Edwards, P., Johnson, L., Hawkes, D., Fenlon, M., Strong, A., and Gleeson, M. (2001). Clinical experience and perception in stereo augmented reality surgical navigation. *Proceedings of the International Workshop on Medical Imaging and Augmented Reality*.
- Egusa, H. (1983). Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual field. *Perception*, 12:167–175.
- Fearing, P. (1995). Importance ordering for real-time depth of field. In *Proceedings of the Third International Computer Science Conference on Image Analysis Applications and Computer Graphics*, pages 372–380.
- Fielding, R. (1985). *Special effects cinematography*. Focal Press, Oxford, fourth edition.
- Fisher, S. K. and Ciuffreda, K. J. (1988). Accommodation and apparent distance. *Perception*, 17:609–621.
- Fishman, E., Kuszyk, B., Heath, D., Gao, L., and Cabral, B. (1996). Surgical planning for liver resection. *Computer*, 29(1):64–72.
- Flickr (2009). Flickr group: Tilt shift miniaturization fakes.
- Fröhlich, B., Barrass, S., Zehner, B., Plate, J., and Göbel, M. (1999). Exploring geoscientific data in virtual environments. In *VIS ’99: Proceedings of the conference on Visualization ’99*, pages 169–173, Los Alamitos, CA, USA. IEEE Computer Society Press.
- Fry, G. A., Bridgeman, C. S., and Ellerbrock, V. J. (1949). The effects of atmospheric scattering on binocular depth perception. *American Journal of Optometry and Archives of American Academy of Optometry*, 26:9–15.

- Fuchs, H., Livingston, M., Raskar, R., Colucci, D., Keller, K., Crawford, J., Rademacher, P., Drake, S., and Meyer, A. (1998). Augmented reality visualization for laparoscopic surgery. *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 934–943.
- Gårding, J., Porrill, J., Mayhew, J., and Frisby, J. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, 35(5):703 – 722.
- Garg, A., Norman, G., Eva, K., Spero, L., and Sharan, S. (2002). Is there any real virtue of virtual reality? The minor role of multiple orientations in learning anatomy from computers. *Academic Medicine*, 77(10):897–899.
- Garg, A., Norman, G., and Sperotable, L. (2001). How medical students learn spatial anatomy. *The Lancet*, 357(9253):363–364.
- Gering, D., Nabavi, A., Kikinis, R., Grimson, W., Hata, N., Everett, P., Jolesz, F., and III, W. W. (2001). An integrated visualization system for surgical planning and guidance using image fusion and interventional imaging. *Journal of Magnetic Resonance Imaging*, 13:967–975.
- Getty, D., D’Orsi, C., and Pickett, R. (2008). Stereoscopic digital mammography: Improved accuracy of lesion detection in breast cancer screening. *LECTURE NOTES IN COMPUTER SCIENCE*, 5116:74–79.
- Getty, D. and Green, P. (2007). Clinical applications for stereoscopic 3-D displays. *Journal of the SID*, 15(6):377–384.
- Goodsitt, M., Chan, H., and Hadjiiski, L. (2000). Stereomammography: Evaluation of depth perception using a virtual 3d cursor. *Medical Physics*, 27:1305.
- Green, P., Sun, W., Matusik, W., and Durand, F. (2007). Multi-aperture photography. *ACM Trans. Graph.*, 26(3):68.
- Haeberli, P. and Akeley, K. (1990). The accumulation buffer: Hardware support for high-quality rendering. *SIGGRAPH Comput. Graph.*, 24(4):309–318.
- Hanna, G., Shimi, S., and Cuschieri, A. (1998). Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy. *The Lancet*, 351:248–251.
- Hecht, H., Kaiser, M. K., and Banks, M. S. (1996). Gravitational acceleration as a cue for absolute size and distance? *Perception & Psychophysics*, 58:1066–1075.
- Held, R. T. and Banks, M. S. (2008). Misperceptions in stereoscopic displays: A vision-science perspective. In *APGV ’08: Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 23–32, New York, NY, USA. ACM.
- Held, R. T., Cooper, E. A., O’Brien, J. F., and Banks, M. S. (2010). Using blur to affect perceived distance and size. *ACM Transactions on Graphics*, 29(2):19:1–16.
- Hemminger, B. M., Molina, P. L., Egan, T. M., Detterbeck, F. C., Muller, K. E., Coffey, C. S., and Lee, J. K. T. (2005). Assessment of real-time 3D visualization for cardiotho-

- racic diagnostic evaluation and surgery planning. *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, 18(2):145–53.
- Hillaire, S., Lécuyer, A., Cozot, R., and Casiez, G. (2007). Depth-of-field blur effects for first-person navigation in virtual environments. In *VRST '07: Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, pages 203–206.
- Hillaire, S., Lecuyer, A., Cozot, R., and Casiez, G. (2008). Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments. *Virtual Reality Conference, 2008. VR '08. IEEE*, pages 47–50.
- Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue.
- Hofmeister, J., Frank, T., Cuschieri, A., and Wade, N. (2001). Perceptual aspects of two-dimensional and stereoscopic display techniques in endoscopic surgery: Review and current problems. *Surgical Innovation*, 8(1):12–24.
- Hu, T., Allen, P., Nadkarni, T., Hogle, N., and Fowler, D. (2009). Insertable stereoscopic 3D surgical imaging device with pan and tilt. *The International Journal of Robotics Research*, 28(10):1373–1386.
- Hu, Y. (2005). The role of three-dimensional visualization in surgical planning of treating lung cancer. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 646–649.
- Ichihara, S., Kitagawa, N., and Akutsu, H. (2007). Contrast and depth perception: Effects of texture contrast and area contrast. *Perception*, 36:686–695.
- Ilgner, J., Park, J., Labbé, D., and Westhofen, M. (2007). Using a high-definition stereoscopic video system to teach microscopic surgery. *Proceedings of SPIE-IS&T Electronic Imaging*, 6490:8–1–8–7.
- Ilgner, J. F. R., Kawai, T., Shibata, T., Yamazoe, T., and Westhofen, M. (2006). Evaluation of stereoscopic medical video content on an autostereoscopic display for undergraduate medical education. *Proceedings of SPIE*, 6055:605506.
- Johnson, W., Rickel, J., Stiles, R., and Munro, A. (1998). Integrating pedagogical agents into virtual environments. *Presence*, 7(6):523–546.
- Jones, G. R., Lee, D., Holliman, N. S., and Ezra, D. (2001). Controlling perceived depth in stereoscopic images. volume 4297, pages 42–53. SPIE.
- Jourdan, I., Dutson, E., Garcia, A., Vleugels, T., Leroy, J., Mutter, D., and Marescaux, J. (2004). Stereoscopic vision provides a significant advantage for precision robotic laparoscopy. *British Journal of Surgery*, 91(7):879–885.
- Kelsey, C., Jr, R. M., Jr, F. M., and Briscoe, D. (1982). Cost-effectiveness of stereoscopic radiographs in detection of lung nodules. *Radiology*, 142(3):611–613.
- Kikinis, R., Gleason, P., Moriarty, T., and . . . , M. M. (1996). Computer-assisted interactive three-dimensional planning for neurosurgical procedures. *Neurosurgery*, 38(4):640–651.



- Kingslake, R. (1992). *Optics in photography*. SPIE Optical Engineering Press, Bellingham, Wash.
- Kolb, C., Mitchell, D., and Hanrahan, P. (1995). A realistic camera model for computer graphics. In *Proceedings of ACM SIGGRAPH*, pages 317–324.
- Kosara, R., Miksch, S., Hauser, F., Schrammel, J., Giller, V., and Tscheligi, M. (2002). Useful properties of semantic depth of field for better f+c visualization. *Joint Eurographics - IEEE TCVG Symposium on Visualization*, pages 205–210.
- Kosara, R., Miksch, S., and Hauser, H. (2001). Semantic depth of field. pages 97–104.
- Kusaka, H. (1992). Apparent depth and size of stereoscopically viewed images. volume 1666, pages 476–482. SPIE.
- Kutka, R. (1994). Reconstruction of correct 3-D perception on screens viewed at different distances. *IEEE Transactions on Communications*, 42:29–33.
- Laforet, V. (2007). A really big show (May 31, 2007). *New York Times*.
- Lambooij, M., IJsselsteijn, W., Fortuin, M., and Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3):030201.
- Larsen, J. S. (1971). Sagittal growth of the eye. *Acta Ophthalmologica*, 49(6):873–886.
- Lee, S., Shinohara, H., Matsuki, M., Okuda, J., Nomura, E., Mabuchi, H., Nishiguchi, K., Takaori, K., Narabayashi, I., and Tanigawa, N. (2003). Preoperative simulation of vascular anatomy by three-dimensional computed tomography imaging in laparoscopic gastric cancer surgery. *Journal of the American College of Surgeons*, 197(6):927–936.
- Leiser, D., Bereby, Y., and Melkman, A. (1995). Minimizing distortions: Seating requirements for stereo projection rooms. 38:1231–1238.
- Levin, A., Fergus, R., Durand, F., and Freeman, W. T. (2007). Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26(3):70–1 – 70–8.
- Lipton, L. (1982). *Foundations of the Stereoscopic Cinema: A Study in Depth*. Van Nostrand Reinhold.
- Love, G., Hoffman, D., Hands, P., and Gao, J. (2008). High-speed switchable lens enables the development of a volumetric stereoscopic display. *IEEE Trans. Consum. Electron.*
- Luursema, J., Verwey, W., Kommers, P., Geelkerken, R. H., and Vos, H. J. (2006). Optimizing conditions for computer-assisted anatomical learning. *Interacting with Computers*, 18:1123–1138.
- Maidment, A. D. A., Bakic, P. R., and Albert, M. (2003). Effects of quantum noise and binocular summation on dose requirements in stereoradiography. *Medical Physics*, 30(12):3061–3071.

- Marshall, J., Burbeck, C., Ariely, D., Rolland, J., and Martin, K. (1996). Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America A*, 13:681–688.
- Martens, T. G. and Ogle, K. N. (1959). Observations on accommodative convergence; especially its nonlinear relationships. *American Journal of Ophthalmology*, 47:455-463.
- Masaoka, K., Hanazato, A., Emoto, M., Yamanoue, H., Nojiri, Y., and Okano, F. (2006). Spatial distortion prediction system for stereoscopic images. *Journal of Electronic Imaging*, 15(1):013002-1 – 013002-12.
- Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings of the Royal Society: Biological Sciences*, 263(1367):169–172.
- Mather, G. and Smith, D. R. R. (2000). Depth cue integration: Stereopsis and image blur. *Vision Research*, 40(25):3501–3506.
- Mather, G. and Smith, D. R. R. (2002). Blur discrimination and its relationship to blur-mediated depth perception. *Perception*, 31(10):1211–1219.
- Matusik, W. and Pfister, H. (2004). 3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *SIGGRAPH '04: SIGGRAPH 2004 Papers*.
- Maurer, C. J., Sauer, F., Hu, B., and Basile, B. (2001). Augmented reality visualization of brain structures with stereo and kinetic depth cues: System description and initial evaluation with head phantom. *Proceedings of SPIE: Medical Imaging*.
- McCloskey, M. and Langer, M. (2009). Planar orientation from blur gradients in a single image. pages 2318–2325.
- McDougall, E. M., Soble, J. J., Wolf, J. S., Nakada, S. Y., Elashry, O. M., and Clayman, R. V. (1996). Comparison of three-dimensional and two-dimensional laparoscopic video systems. *J Endourol*, 10(4):371–4.
- Meesters, L., IJsselstein, W., and Seuntjens, P. (2004). A survey of perceptual evaluations and requirements of three-dimensional tv. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(3):381–391.
- Moll, T., Douek, P., Finet, G., Turjman, F., and ... , C. P. (1998). Clinical assessment of a new stereoscopic digital angiography system. *Cardiovascular and Interventional Radiology*, 21:11–16.
- Mon-Williams, M. and Tresilian, J. R. (2000). Ordinal depth information from accommodation. *Ergonomics*, 43(3):391–404.
- Moreno-Noguer, F., Belhumeur, P. N., and Nayar, S. K. (2007). Active refocusing of images and videos. *ACM Trans. Graph.*, 26(3):67-1 – 67-9.
- Mueller-Richter, U., Limberger, A., and Weber, P. (2003). Comparison between three-dimensional presentation of endoscopic procedures with polarization glasses and an autostereoscopic display. *Surg Endosc*, 17(3):1432–2218.

- Mulder, J. D. and van Liere, R. (2000). Fast perception-based depth of field rendering. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 129–133.
- Nicholson, D., Chalk, C., Funnell, W., and Daniel, S. (2006). Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *MEDICAL EDUCATION-OXFORD-*, 40(11):1081.
- Nikou, C., Digioia, A., Blackwell, M., Jaramaz, B., and Kanade, T. (2000). Augmented reality imaging technology for orthopaedic surgery. *Operative Techniques in Orthopaedics*, 10(1):82–86.
- Novotny, P., Kettler, D., Jordan, P., Dupont, P., Pedro, J., and Howe, R. (2006). Stereo display of 3d ultrasound images for surgical robot guidance. *IEEE International Conference of the Engineering in Medicine and Biology Society, New York, NY*.
- Ogle, K. N. (1938). Induced size effect. I. A new phenomenon in binocular space-perception associated with the relative sizes of the images of the two eyes. *Archives of Ophthalmology*, 20:604–623.
- Okatani, T. and Deguchi, K. (2007). Estimating scale of a scene from a single image based on defocus blur and scene geometry. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8.
- O’Shea, J. P., Banks, M. S., and Agrawala, M. (2008). The assumed light direction for perceiving shape from shading. In *APGV '08: Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 135–142, New York, NY, USA. ACM.
- Palmer, S. (1999). *Vision Science: From Photos to Phenomenology*. MIT Press.
- Palmer, S. E. and Brooks, J. L. (2008). Edge-region grouping in figure-ground organization and depth perception. *Journal of Experimental Psychology: Human Perception and Performance.*, 24(6):1353–1371.
- Patel, H., Ribal, M., Arya, M., Nauth-Misir, R., and Joseph, J. (2007). Is it worth revisiting laparoscopic three-dimensional visualization? A validated assessment. *Urology*, 70(1):47–49.
- Pentland, A. P. (1987). A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine*, 9(4):523–531.
- Pieper, S., Delp, S., Rosen, J., and Fisher, S. (1991). Virtual environment system for simulation of leg surgery. *Proceedings of SPIE*, 1457:188.
- Pietrabissa, A., Scarcello, E., Carobbi, A., and Mosca, F. (1994). Three-dimensional versus two-dimensional video system for the trained endoscopic surgeon and the beginner. *Endoscopic surgery and allied technologies*, 2(6):315–317.
- Pietrzak, P., Arya, M., Joseph, J., and Patel, H. (2006). Three-dimensional visualization in laparoscopic surgery. *BJU International*, 98(2):253–256.

- Pisano, E. D., Seibert, J. A., Andriole, K. P., Breeden, III, W. K., Krupinski, E., Siegel, E. L., Williams, M. B., and Yaffe, M. J. (2007). *Practice Guideline for Determinants of Image Quality in Digital Mammography*. American College of Radiology.
- Potmesil, M. and Chakravarty, I. (1981). A lens and aperture camera model for synthetic image generation. *SIGGRAPH Comput. Graph.*, 15(3):297–305.
- Prystowsky, J., Regehr, G., Rogers, D., Loan, J., Hiemenz, L., and Smith, K. (1999). A virtual reality module for intravenous catheter placement. *Am J Surg*, 177(2):171–5.
- Rogers, B. and Bradshaw, M. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, 24(2):155–179.
- Rogers, B. J. and Bradshaw, M. F. (1993). Vertical disparities, differential perspective, and binocular stereopsis. *Nature*, 361:253–255.
- Rohaly, A. and Wilson, H. (1999). The effects of contrast on perceived depth and depth discriminations. *Vision Research*, 39:9–18.
- Rokita, P. (1996). Generating depth-of-field effects in virtual reality applications. *Computer Graphics and Applications, IEEE*, 16(2):18–21.
- Rosahl, S., Gharabaghi, A., Hubbe, U., Shahidi, R., and Samii, M. (2006). Virtual reality augmentation in skull base surgery. *Skull Base*, 16(2):59–66.
- Rosenthal, M., State, A., Lee, J., Hirota, G., Ackerman, J., Keller, K., Pisano, E., Jiroutek, M., Muller, K., and Fuchs, H. (2002). Augmented reality guidance for needle biopsies: an initial randomized, controlled trial in phantoms. *Medical Image Analysis*, 6(3):313–20.
- Saxberg, B. V. H. (1987). Projected free fall trajectories: 1. Theory and simulation. *Biological Cybernetics*, 56:159–175.
- Schechner, Y. Y. and Kiryati, N. (2000). Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 29(2):141–162.
- Schiffman, B. (2008). Movie industry doubles down on 3D. *Wired Magazine*.
- Sedgwick, H. A. (1986). *Space Perception*. Wiley.
- Sekiguchi, R., Satake, M., Oyama, H., and Wakao, F. (1996). Stereoscopic visualization system for clinical angiography. *Studies in health technology and informatics*, 29(690-3).
- Serra, L., Hern, N., Choon, C., and Poston, T. (1997). Interactive vessel tracing in volume data. *SI3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*.
- Shapiro, L. G. and Stockman, G. C. (2001). *Computer Vision*. Prentice Hall.
- Simons, K. (1981). Stereoacuity norms in young children. *Arch Ophthalmol-Chic*, 99(3):439–45.

- Sollenberger, R. and Milgram, P. (1993). Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(3):483–499.
- Son, J.-Y., Gruts, Y., Chun, J., Choi, Y. J., Bahn, J.-E., and Bobrinev, V. I. (2002). Distortion analysis in stereoscopic images. *Optical Engineering*, 41(3):680–685.
- Song, C.-G. and Kang, J. U. (2009). Design of the computerized 3D endoscopic imaging system for delicate endoscopic surgery. *J Med Syst*, pages 1–7.
- Sorantin, E. (2008). Soft-copy display and reading: What the radiologist should know in the digital era. *Pediatr Radiol*, 38(12):1276–84.
- Spring, K. and Stiles, W. S. (1948). Variation of pupil size with change in the angle at which the light stimulus strikes the retina. *Br. J. Ophthalmol.*, 32(6):340–346.
- Strunk, L. M. and Iwamoto, T. (1990). A linearly-mapping stereoscopic visual interface for teleoperation. In *Proceedings of the IEEE International Workshop of Intelligent Robots and Systems*, pages 429–436.
- Sun, J. and Perona, P. (1998). Where is the sun? *Nature Neuroscience*, 1(3):183–184.
- Taffinder, N., Smith, S. G., Huber, J., Russell, R. C., and Darzi, A. (1999). The effect of a second-generation 3D endoscope on the laparoscopic precision of novices and experienced surgeons. *Surg Endosc*, 13(11):1087–92.
- Tanaka, C., Fujii, H., Ikeda, T., Jinno, H., Nakahara, T., Suzuki, T., Kitagawa, Y., Kitajima, M., Ando, Y., and Kubo, A. (2007). Stereoscopic scintigraphic imaging of breast cancer sentinel lymph nodes. *Breast Cancer*, 14(1):92–9.
- Tendick, F., Downes, M., Cavusoglu, M., Gantert, W., and Way, L. (1998). Development of virtual environments for training skills and reducing errors in laparoscopic surgery. *Proceedings of the SPIE International Symposium on Biological Optics (BIOS'98)*, pages 36–44.
- Tendick, F., Downes, M., Goktekin, T., Cavusoglu, M., Feygin, D., Wu, X., Eyal, R., Hegarty, M., and Way, L. (2000). A virtual environment testbed for training laparoscopic surgical skills. *Presence: Teleoperators & Virtual Environments*, 9(3):236–255.
- Tevaearai, H. T., Mueller, X. M., and von Segesser, L. K. (2000). 3-d vision improves performance in a pelvic trainer. *Endoscopy*, 32(6):464–8.
- Tuggy, M. L. (1998). Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. *The Journal of the American Board of Family Practice / American Board of Family Practice*, 11(6):426–33.
- van Bergen, P., Kunert, W., Bessell, J., and Buess, G. F. (1998). Comparative study of two-dimensional and three-dimensional vision systems for minimally invasive surgery. *Surg Endosc*, 12(7):948–54.
- Vasilyev, N., Novotny, P., Martinez, J., Loyola, H., Salgo, I., Howe, R., and del Nido, P. (2008). Stereoscopic vision display technology in real-time three-dimensional

- echocardiography-guided intracardiac beating-heart surgery. *The Journal of Thoracic and Cardiovascular Surgery*, 135(6):1334.
- Vishwanath, D. (2008). The focal blur gradient affects perceived absolute distance [ECVP abstract supplement]. *Perception*, 27:130.
- Vishwanath, D., Girshick, A. R., and Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, 8(10):1401–1410.
- Wacker, F., Vogt, S., Khamene, A., Jesberger, J., Nour, S., Elgort, D., Sauer, F., Duerk, J., and Lewin, J. (2006). An augmented reality system for MR image-guided needle biopsy: Initial results in a swine model. *Radiology*, 238(2):497–504.
- Wallach, H. and Norris, C. M. (1963). Accommodation as a distance cue. *American Journal of Psychology*, 76:659–664.
- Wartell, Z., Hodges, L., and Ribarsky, W. (2002). A geometric comparison of algorithms for fusion control in stereoscopic HTDs. *IEEE Transactions on Visualization and Computer Graphics*, 8:129–143.
- Watson, J., Banks, M., Hofsten, C., and Royden, C. S. (1992). Gravity as a monocular cue for perception of absolute distance and/or absolute size. *Perception*, 21(1):69–76.
- Watt, S. J., Akeley, K., Ernst, M. O., and Banks, M. S. (2005). Focus cues affect perceived depth. *J. Vis.*, 5(10):834–862.
- Wells, N. (2002). Display of Munsell color values, earthquakes, and other three- and four-parameter datasets in stereo 3d. *Comput Geosci-Uk*, 28(5):701–709.
- Wendt, M., Sauer, F., Khamene, A., and Bascle, B. (2003). A head-mounted display system for augmented reality: Initial evaluation for interventional MRI. *RöFo: Fortschritte auf dem Gebiete der . . .*
- Wenzl, R., Lehner, R., Vry, U., Pateisky, N., Sevelde, P., and Husslein, P. (1994). Three-dimensional video-endoscopy: Clinical use in gynaecological laparoscopy. *The Lancet*, 344(8937):1621–2.
- Wigmore, S., Redhead, D., Yan, X., Casey, J., Madhavan, K., Dejong, C., Currie, E., and Garden, O. (2001). Virtual hepatic resection using three-dimensional reconstruction of helical computed tomography angioportograms. *Annals of Surgery*, 233(2):221–226.
- Wilson, B. J., Decker, K. E., and Roorda, A. (2002). Monochromatic aberrations provide an odd-error cue to focus direction. *Journal of the Optical Society of America A*, 19:833–839.
- Woods, A. J., Docherty, T., and Koch, R. (1993). Image distortions in stereoscopic video systems. *SPIE: Stereoscopic Displays and Applications IV*, 1915(1):36–48.
- Xia, J., Ip, H. H., Samman, N., Wang, D., Kot, C. S., Yeung, R. W., and Tideman, H. (2000). Computer-assisted three-dimensional surgical planning and simulation: 3D virtual osteotomy. *International journal of oral and maxillofacial surgery*, 29(1):11–17.

- Xu, M. and Wang, L. (2006). Photoacoustic imaging in biomedicine. *Review of Scientific Instruments*, 77:041101.
- Yamanoue, H. (1997). The relation between size distortion and shooting conditions for stereoscopic images. *Journal of the SMPTE*, 106(4):225–232.
- Yamanoue, H., Okui, M., and Okano, F. (2006). Geometrical analysis of puppet theater and cardboard effects in stereoscopic HDTV images. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(6):744–752.
- Yamanoue, H., Okui, M., and Yuyama, I. (2000). A study on the relationship between shooting conditions and cardboard effect of stereoscopic images. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(3):411–416.
- Zhai, S., Buxton, W., and Milgram, P. (1996). The partial-occlusion effect: Utilizing semitransparency in 3D human-computer interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(3):254–284.

# APPENDIX A

## SKEW-RAY GEOMETRY

In Section 1.3.1, we used epipolar geometry to illustrate the geometric approach's inability to provide a solution under certain viewing situations. Here we provide more detailed derivations that produce the same results. For a given 3D point in a stereo image, recall that the terms  $(\mathbf{P}'_l - \mathbf{E}'_l)m$  and  $(\mathbf{P}'_r - \mathbf{E}'_r)n$  represent the rays passing from the centers of the eyes ( $\mathbf{E}'_l$  and  $\mathbf{E}'_r$ ) to the corresponding points on the screen ( $\mathbf{P}'_l$  and  $\mathbf{P}'_r$ ) in viewer-centered coordinates. For two rays to intersect, they must be non-parallel and lie in a common plane. We determine whether two rays lie in a plane by first finding the plane defined by the two eye centers and a point in the left stereo picture, and then by finding the plane defined by the two eyes and the corresponding point in the right stereo picture. If the two planes are coincident, the rays must intersect, provided that they are non-parallel. We define the planes by their surface normals. The first normal is given by the cross product between two vectors that originate at the left eye and extend to the point in the left stereo picture ( $v_{l1}$ ) and to the right eye ( $v_{l2}$ ) (Figure A.1). The second plane is given by the cross product between the vectors originating at the left eye and extending to the point in the right stereo picture ( $v_{r1}$ ) and to the right eye ( $v_{r2}$ ).

$$\mathbf{v}_{l1} = (\mathbf{P}'_l - \mathbf{E}'_l)$$

$$\mathbf{v}_{r1} = (\mathbf{P}'_r - \mathbf{E}'_l)$$

$$\mathbf{v}_{l2} = (\mathbf{E}'_r - \mathbf{E}'_l)$$

$$\mathbf{v}_{r2} = (\mathbf{E}'_r - \mathbf{E}'_l)$$



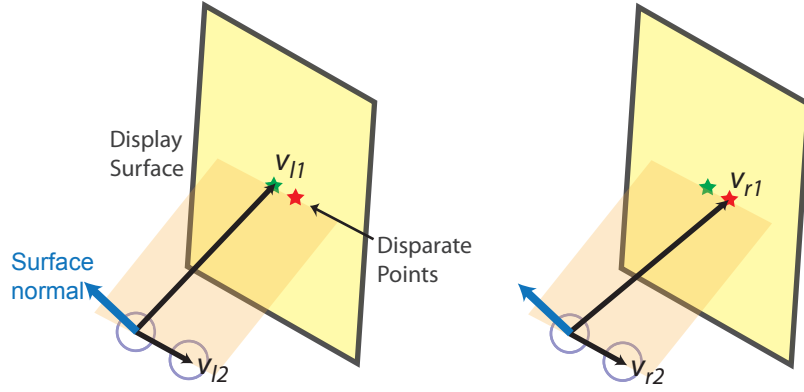


Fig. A.1: Planes defined using the centers of both eyes and either of the corresponding points in the pictures. The cross product of the illustrated vectors (from one eye to the other and from one eye to the image point) is a normal vector that defines the plane.

To include the viewer's position and orientation relative to the picture, we replace  $\mathbf{P}'_l$  and  $\mathbf{P}'_r$  with:

$$\mathbf{P}'_l = R(\rho, \sigma)(\mathbf{P}_l - \mathbf{E}_c)$$

$$\mathbf{P}'_r = R(\rho, \sigma)(\mathbf{P}_r - \mathbf{E}_c)$$

We can now investigate how viewer translation ( $E_c$ ) and rotation ( $R(\rho, \sigma)$ ) affect the above-defined planes.  $v_{l2}$  and  $v_{r2}$ , the same vectors originating at the left eye and extending to the right eye, can be expressed as  $(I, 0, 0)$ , where  $I$  is inter-ocular distance. Combining the equations, multiplying out the elements of  $R(\rho, \sigma)$ , and taking into account that  $P_l(z)$  and

$P_r(z)$  are both zero produces the following equations for the cross products:

$$\begin{aligned}
\mathbf{v}_{l1} \times \mathbf{v}_{l2} &= \mathbf{j}(I(\sin(\rho) \cos(\sigma)(P_l(x) - E_c(x)) \\
&\quad + \sin(\rho) \sin(\sigma)(P_l(y) - E_c(y)) + \cos(\rho)(-E_c(z)))) \\
&\quad + \mathbf{k}(I(\sin(\sigma)(P_l(x) - E_c(x)) + \cos(\sigma)(P_l(y) - E_c(y)))) \\
\mathbf{v}_{r1} \times \mathbf{v}_{r2} &= \mathbf{j}(I(\sin(\rho) \cos(\sigma)(P_r(x) - E_c(x)) \\
&\quad + \sin(\rho) \sin(\sigma)(P_r(y) - E_c(y)) + \cos(\rho)(-E_c(z)))) \\
&\quad + \mathbf{k}(I(\sin(\sigma)(P_r(x) - E_c(x)) + \cos(\sigma)(P_r(y) - E_c(y)))) \tag{A.1}
\end{aligned}$$

We now have surface-normal representations for the two planes that originate at  $\mathbf{E}_1$ . Before testing for equality, we normalize the vectors by dividing by their magnitudes  $|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|$  and  $|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|$ . Then, to determine if the planes are coincident, we check to see if their  $\mathbf{j}$  and  $\mathbf{k}$  terms are equal to one another (there are no  $\mathbf{i}$  terms in the equations above). The  $\mathbf{j}$  terms are:

$$\begin{aligned}
&\sin(\rho) \cos(\sigma)(P_l(x) - E_c(x)) \\
&\quad + \sin(\rho) \sin(\sigma)(P_l(y) - E_c(y)) + \cos(\rho)(-E_c(z)) \\
&= \frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} (\sin(\rho) \cos(\sigma)(P_r(x) - E_c(x)) \\
&\quad + \sin(\rho) \sin(\sigma)(P_r(y) - E_c(y)) + \cos(\rho)(-E_c(z))) \tag{A.2}
\end{aligned}$$

and the  $\mathbf{k}$  terms are:

$$\begin{aligned}
&\sin(\sigma)(P_l(x) - E_c(x)) + \cos(\sigma)(P_l(y) - E_c(y)) \\
&= \frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} (\sin(\sigma)(P_r(x) - E_c(x)) + \cos(\sigma)(P_r(y) - E_c(y))) \tag{A.3}
\end{aligned}$$

## Observer Translation

In this condition, the viewer is translated relative to the picture. The viewer's face is parallel to the picture surface and not rotated, so the  $X$ ,  $Y$ , and  $Z$  axes in the picture- and viewer-centered coordinate systems are parallel. Because the axes are parallel,  $\rho$  and  $\sigma$  in the rotation matrix  $R(\rho, \sigma)$  are 0. If these values are plugged in to Equations A.1, A.2, and A.3, the  $\mathbf{j}$  and  $\mathbf{k}$  terms of the surface normals become:

$$-E_c(z) = -\frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} E_c(z)$$

and

$$P_l(y) - E_c(y) = \frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} (P_r(y) - P_c(y))$$

where

$$\frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} = \frac{|\mathbf{j}(I(-E_c(z)) + \mathbf{k}(I(P_l(y) - E_c(y))))|}{|\mathbf{j}(I(-E_c(z)) + \mathbf{k}(I(P_r(y) - E_c(y))))|}$$

Examining the equations, it is clear that the two equalities are only valid if the  $Y$  coordinates of corresponding points on the picture are equal to one another. In other words, there must be no vertical disparities in the picture. On-screen vertical disparity is always zero when the images are captured with cameras whose lens axes are parallel, even if the optical axes are converged (i.e., the sensors are offset relative to the lens axes). Therefore, when the cameras' lens axes are parallel and the viewer's eye coordinates are parallel to the picture-centered coordinates, there are intersections for all ray pairs and the geometric approach provides a solution.

### **Observer Rotation in the $X - Z$ Plane**

Viewer rotation in the  $X - Z$  plane—"yaw" rotation—occurs when the viewer is positioned to the left or right of the proper position and turns the head toward the center of the stereo

picture.  $\rho$  is now nonzero, but  $\sigma$  is still zero. Equations A.1 and A.2 then become:

$$\begin{aligned} & \sin(\rho)(P_l(x) - E_c(x)) + \cos(\rho)(-E_c(z)) \\ &= \frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} (\sin(\rho)(P_r(x) - E_c(x)) + \cos(\rho)(-E_c(z))) \end{aligned}$$

and

$$P_l(y) - E_c(y) = \frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} (P_r(y) - E_c(y))$$

where

$$\frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} = \frac{|\mathbf{j}(I(\sin(\rho)(P_l(x) - E_c(x)) + \cos(\rho)(-E_c(z)))) + \mathbf{k}(I(P_l(y) - E_c(y)))|}{|\mathbf{j}(I(\sin(\rho)(P_r(x) - E_c(x)) + \cos(\rho)(-E_c(z)))) + \mathbf{k}(I(P_r(y) - E_c(y)))|}$$

In this case,  $P_l(y)$  and  $P_r(y)$  must be equal to  $E_c(y)$  for the dual equalities to be valid. The only corresponding points that will produce intersecting rays are points in the  $X - Z$  plane; all rays above and below that plane will be non-intersecting (skew rays), except when  $\mathbf{P}_l$  and  $\mathbf{P}_r$  are identical (zero disparity). Because rays for many points in the stereo picture do not intersect, the geometric approach cannot provide a prediction for what viewers should perceive. This is disappointing because yaw rotations are commonplace in the viewing of stereo media.

### Observer Rotation in the $X - Y$ Plane

With viewer rotation in the  $X - Y$  plane—“roll” rotation— $\sigma$  is nonzero and  $\rho$  is zero, which results in

$$-E_c(z) = -\frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} E_c(z)$$

and

$$\begin{aligned} & \sin(\sigma)(P_l(x) - E_c(x)) + \cos(\sigma)(P_l(y) - E_c(y)) \\ &= \frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} \sin(\sigma)(P_r(x) - E_c(x)) + \cos(\sigma)(P_r(y) - E_c(y)) \end{aligned}$$

where

$$\frac{|\mathbf{v}_{r1} \times \mathbf{v}_{r2}|}{|\mathbf{v}_{l1} \times \mathbf{v}_{l2}|} = \frac{|\mathbf{j}(-IE_c(z)) + \mathbf{k}(I(\sin(\sigma)(P_l(x) - E_c(x)) + \cos(\sigma)(P_l(y) - E_c(y))))|}{|\mathbf{j}(-IE_c(z)) + \mathbf{k}(I(\sin(\sigma)(P_r(x) - E_c(x)) + \cos(\sigma)(P_r(y) - E_c(y))))|}$$

Equations A.2 and A.3 are only valid if  $\mathbf{P}_l$  is identical to  $\mathbf{P}_r$ . Any non-zero disparity produces skew rays if the head is rolled relative to the picture. Thus, this is another situation in which the geometric approach cannot provide a prediction of viewers' percepts.

# APPENDIX B

## DATA FROM BLUR EXPERIMENT

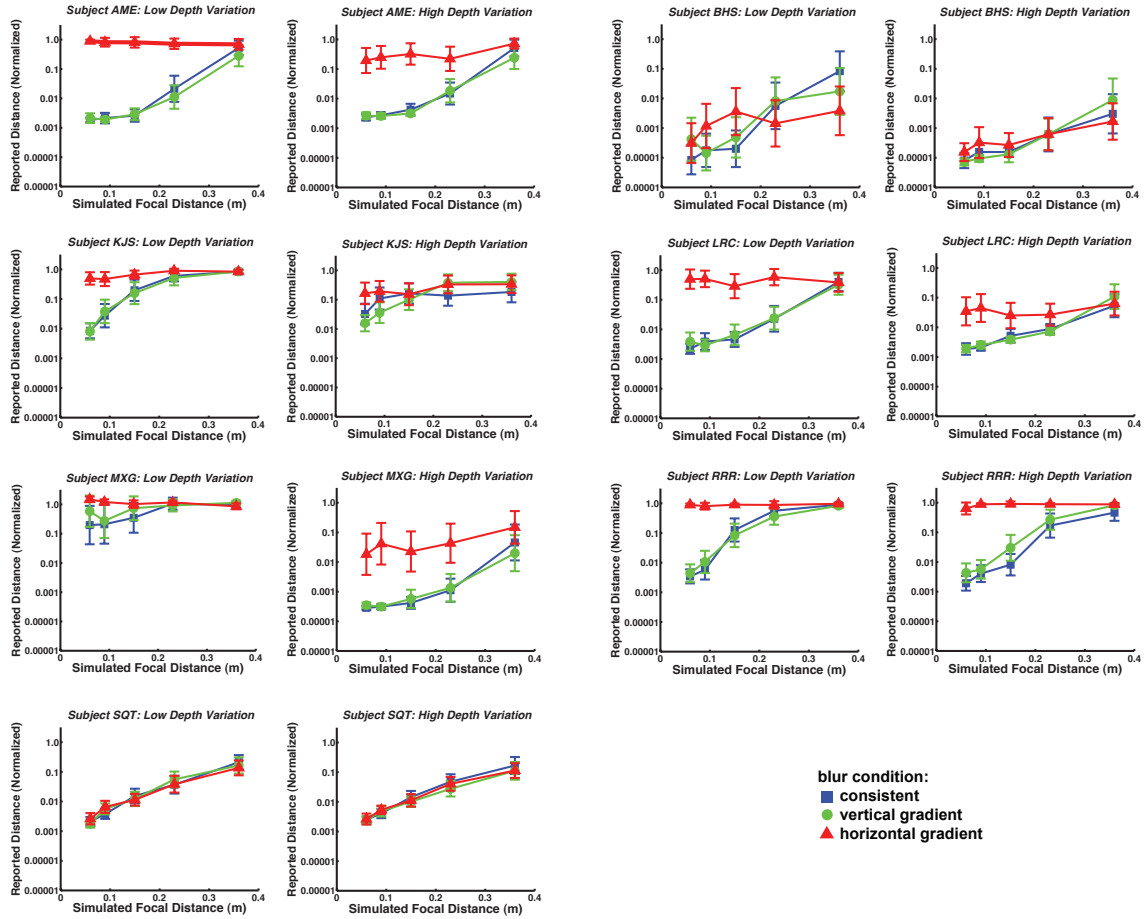


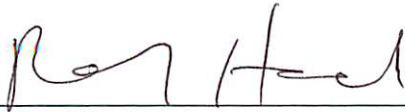
Fig. B.1: Individual subject data from the psychophysical blur experiment. For each scene, a subject's responses were normalized by settings for the no-blur condition. Vertical-blur gradients and consistent blur were equivalently effective at modulating perceived distance. In every subject but SQT and BHS, horizontal-blur gradients were the least effective. These results support the predictions of our probabilistic model.

**Publishing Agreement**

*It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.*

***Please sign the following statement:***

*I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*



\_\_\_\_\_  
Author Signature

5/1/2010

\_\_\_\_\_  
Date