# A Perceptual Study of Occlusion and Luminance in 3-D Clutter

Milena Scaccia

Doctor of Philosophy

School of Computer Science

McGill University

Montreal, Quebec

September 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

© Milena Scaccia, 2019

#### ABSTRACT

Three-dimensional (3-D) cluttered scenes consist of many small occluding surfaces that are distributed over a volume. It is challenging to perceive such scenes due to limited visibility from occlusions. Nonetheless, occlusions can serve as a relative depth cue where points that are deeper in the clutter are less visible. In this thesis, we examine how well the human visual system uses information from occlusions paired with luminance cues for two types of perceptual tasks: (1) discriminating the depth of two target surfaces in the clutter and (2) discriminating the density of two halves of the clutter i.e. front versus back or left versus right. For the first task of discriminating depth, we tested the interaction between occlusions, color, and depth-luminance covariance (DLC). We tested both negative DLC i.e. dark-means-deep and positive DLC i.e. bright-mean-deep. For the second task of discriminating density of two halves of the clutter, we measured observers' bias and sensitivity for different parameters of clutter, namely the area and density of the occluders and the overall level of occlusion. We also ran model observers that compared the image occupancy of the two halves in order to gain more insight into the information available from the luminance, occlusion, and rotational motion cues for different scene parameters and for different tasks, namely front-back or left-right. In our research overall, we found that occlusions and luminance cues provide vital information for discriminating 3-D clutter. However, the information available is dependent on the task. For depth discrimination, we show that occlusions can help determine the sign of the DLC to resolve depth order, although there is a prior for negative DLC when the colors of the targets versus distractors differ. For more volume-dependant tasks such as density discrimination, we find that there is a subtle interplay between scene parameters such as occlusion, density, and area of the surfaces. For the front-back task, the level of occlusion affects the bias to see the front or back as denser, where low occlusion results in back bias and high occlusion results in front bias; also the front bias is reduced when using opposite luminance (white-black) on each half over using DLC. Weber fractions for the front-back task decrease as the density of surfaces increases, and for the left-right task they decrease as density increases and increase as the area of the surfaces increases. We find that varying density versus area has different effects on Weber fractions even though by design their variations produce the same changes in occlusion and image occupancy.

# ABRÉGÉ

Les scènes trois dimensionelles (3-D) composées de nombreuses petites surfaces sont difficile à percevoir en raison que les occlusions limitent la visibilité. Néanmoins, les occlusions peuvent servir comme indice de la profondeur relative où les points plus profonds sont moins visibles. Dans cette thèse, nous examinons dans quelle mesure le système visuel humain utilise les informations d'occlusions avec des indices de luminance pour deux types de tâches de perception: (1) la discrimination de la profondeur de deux surfaces cibles dans la scène et (2) la discrimination de la densité de deux moitiés de la scène, c'est-à-dire avant contre arrière ou gauche contre droite. Pour la premiére tâche de la discrimination de la profondeur, nous avons testé l'interaction entre les occlusions, la couleur, et la covariance profondeur-luminance (DLC). Nous avons testé les deux signes de la DLC, c'est-à-dire DLC négative où les surfaces plus profonds sont plus sombres, et DLC positive où les surfaces plus profonds sont plus brillantes. Pour la deuxième tâche de la discrimination de la densité des deux moitiés, nous avons mesuré le biais et la sensibilité des observateurs pour différents paramètres, à savoir la densité et la superficie des surfaces occluseurs, et le niveau d'occlusion. Nous avons aussi testé des observateurs modèles qui comparent l'occupation de l'image des deux moitiés afin de mieux comprendre les informations disponibles des indices de luminance, d'occlusion, et du mouvement rotatif pour différents paramètres de la scène et pour différentes tâches. Dans notre recherche, nous avons trouvé que les indices d'occlusions et de luminance fournissent des informations

essentielles pour la perception des scènes tridimensionnelles à plusieurs surfaces. Cependant, les informations disponibles dépendent de la tâche. Pour la discrimination de la profondeur, nous montrons que les occlusions peuvent aider à déterminer le signe de la DLC pour résoudre l'ordre de profondeur, bien qu'il y ait un prior pour la DLC négative quand les couleurs des cibles et distracteurs diffèrent. Pour la discrimination de la densité, nous constatons une interaction subtile entre les paramètres de la scène tels que l'occlusion, la densité et la superficie des surfaces. Pour la tâche avant-arrière, le niveau d'occlusion affecte le biais, où une faible occlusion engendre un biais de voir l'arrière plus dense, et une occlusion élevée engendre un biais de voir l'avant plus dense. Le biais frontale est réduite lorsqu'on utilise une luminance opposée (blanc-noir) sur chaque moitié au lieu de la DLC. Les fractions de Weber pour la tâche avant-arrière diminuent quand la densité des surfaces augmente, et pour la tâche gauche-droite, ils diminuent quand la densité augmente et ils augmentent quand la superficie des surfaces augmente. Nous constatons que la densité par rapport à la superficie des surfaces ont des effets différents sur les fractions Weber même si, par conception, leurs variations produisent les mêmes changements dans le niveau d'occlusion et de l'occupation de l'image.

#### ACKNOWLEDGEMENTS

First and foremost, I would like to thank my PhD supervisor Professor Michael Langer for his continuous guidance and support in the development of this thesis. Working with Professor Langer has enabled me to grow as a researcher and sharpen my communication skills. I am grateful for the invaluable help he has provided throughout the writing, editing and experiment design stages of the research. I would like to express my sincere gratitude to my PhD Committee, Professor James J. Clark and Professor Kaleem Siddiqi, for encouraging my progress and for their insightful feedback at all stages of my PhD. I wish to acknowledge professors and staff who also have advised me during courses, exams, or for general graduate study inquiries: Paul Kry, Prakash Panangaden, Jerome Waldispuhl, Brigitte Pientka, Doina Precup, Joelle Pineau, Laurie Hendren, Luc Devroye, Jörg Kienzle, Gregory Dudek, Jan Binder, Diti Anastasopoulos, among others. My thanks also extend to Professors Fred Kingdom and Nick Prins for their advice on psychometric curve fitting, and to Professors James Elder, Rob Allison, Curtis Baker, and Steven Zucker for the feedback they provided at conferences. I thank all those who have participated in my experiments. I thank my fellow graduate students for their friendship. I thank my family for their patience, love, encouragement, and support, and for the many hours spent in pilot experiments. This research was supported by a doctoral grant awarded to Milena Scaccia from the Fonds de Recherche du Quebec - Nature et Technologies (FRQNT), and an FRQNT team research grant and Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant to Michael Langer.

#### **CONTRIBUTION OF AUTHORS**

I am first author of the two manuscripts included in this thesis. The co-author of both manuscripts is my supervisor Michael Langer. The individual contributions of each author for each manuscript are as follows.

#### Signs of depth-luminance covariance in 3-D cluttered scenes

- Milena Scaccia: Developed the methods, created the figures (except Figure 3–1), calibrated the stimuli, designed and implemented the experiments, called participants, conducted the experiments to acquire the data, carried out the statistical analysis, interpreted the data, and wrote and edited the manuscript.
- Michael Langer: Supervised the development of the methods, the design of the experiments, created Figure 3–1, contributed to the interpretation of the data, revision of the manuscript, wrote/rewrote paragraphs in the literature review, and checked the references.

#### Density discrimination with occlusions in 3-D clutter

- Milena Scaccia: Developed the methods, created the figures, calibrated the stimuli, designed and implemented the experiments, called participants, conducted the experiments to acquire the data, carried out the statistical analysis, interpreted the data, and wrote and edited the manuscript.
- Michael Langer: Supervised the development of the methods, the design of the experiments, contributed to the interpretation of the data, co-wrote and edited the manuscript, and checked the references. In particular, held a significant contributing

role in the analysis and writing of the model observers Section 4.3.3, and rewrote and reorganized text in the Methods Section 4.2 and Discussion Section 4.4.1.

# TABLE OF CONTENTS

ABS	TRACI	Γ	ii
ABR	ÉGÉ .		iv
ACK	NOWL	EDGEMENTS	vi
CON	TRIBU	UTION OF AUTHORS	ii
LIST	OF TA	ABLES	ii
LIST	OF FI	GURES	ii
ABB	REVIA	TIONS	vi
1	Introdu	uction	1
	1.1 1.2 1.3	MotivationOrganization of ThesisContribution to Original Knowledge	1 3 4
2	Backg	round	6
	2.1	Ambient Occlusion	6 6 9 3 5 8
	2.2	1ransparency       2         2.2.1       Introduction       2         2.2.2       Stimuli       2         2.2.3       Discussion       2	20 20 22 22
	2.A	Derivation of Ambient Occlusion Models in 3-D Clutter22.A.1 Hedge Models2	.5 25

		2.A.2 Sphere Models	31
3	Signs	of depth-luminance covariance in 3-D cluttered scenes	34
	Prefac	e	34
	Abstra	act	35
	3.1	Introduction	37
	3.2	Method	43
		3.2.1 Apparatus	43
		3.2.2 Stimuli	44
		3.2.3 Design	47
		3.2.4 Procedure	51
		3.2.5 Participants	51
	3.3	Results	52
		3.3.1 Experiment 1	52
		3.3.2 Experiment 2	53
		3.3.3 Experiment 3	55
	3.4	Discussion	57
	3.5	Conclusion	59
	3.A	Color Calibration	60
	3.B	Model of Depth-Luminance Covariance	61
	3.C	Manipulating Visibility Cue in Experiment 2	64
	Ackno	owledgements	66
	Refere	ences	68
4	Densit	ty discrimination with occlusions in 3-D clutter	71
•	Denon		, 1
	Prefac	e	71
	Abstra	act	73
	4.1	Introduction	75
	4.2	Method	80
		4.2.1 Subjects	80
		4.2.2 Apparatus	80
		4.2.3 Stimuli	81
		4.2.4 Design	88
		4.2.5 Procedure	89
	4.3	Results	90
		4.3.1 Experiments 1 and 2 (front-back)	91
		4.3.2 Experiments 3 and 4 (left-right)	95

		4.3.3 Model observers	99
	4.4	Discussion	104
		4.4.1 Comparison between human and model observers	104
		4.4.2 Comparison with previous work	106
	4.5	Conclusion	108
	Ackn	owledgements	110
	Refer	ences	111
5	Discu	ssion and Conclusion	114
	5.1	Future work: Perception of Clumped 3-D Clutter	114
	5.2	Other Luminance Models	116
	5.3	Conclusion	119
А	Gamr	na Calibration	121
Refe	erences		125

# LIST OF TABLES

# Table

page

4–1	The 3 $\times$ 3 table shows values of mean density $\eta$ , area <i>A</i> , and number <i>N</i> of	
	surfaces for stimuli in our experiments. The occlusion factor $\lambda = \eta A$ is	
	constant in each column and increases from left to right. The area $A$ is	
	constant on the main diagonal and increases on the cross diagonal. The	
	mean density $\eta$ is constant on the cross-diagonal and increases on the	
	main diagonal.	81

# LIST OF FIGURES

Figure		page
2–1	Example ray-cast AO rendering of (a) Infinite Hedge and (b) Sphere. Example qualitative AO model rendering of (c) Infinite Hedge and (d) Sphere.	12
2–2	Mean luminance as a function of depth for a surface with normal parallel to the Z direction with frontoparallel (red) and randomly oriented (green) surrounding clutter. We plotted both the true ray-cast AO luminance values (dots), as well as the modeled luminance values (lines) of Equations 2.3 and 2.6, for frontoparallel and random orientated occlusions respectively. Error bars show standard deviation for 100 scenes.	16
2–3	Mean luminance as a function of depth for orientations $\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}$ , as per the (a) front-lit model in Equation 2.6, (b) back-lit model in Equation 2.7 and (c) front-and-back-lit models in Equations 2.6 + 2.7. Error bars show standard deviation for 100 scenes.	17
2–4	Standard deviation over 100 scenes as a function of occluder width while keeping $\lambda$ constant. When $\lambda$ is kept constant, the standard deviation increases as the density decreases and the area of the occluders increases.	18
2–5	Transparent occluders. Side blue opaque occluders were used to prevent depth reversals.	23
2–6	Opaque Occluders.	23
2.A.	1Luminance as a function of depth for orientations 0 to $\frac{\pi}{2}$ radians in a spherical volume. We show the ray-cast values and the model values from Equation 2.30. Error bars show standard deviation for 100 scenes.	33

3–1	Illustration of tunnel for the DLC- condition. The viewing position is from below. (a) The tunnel removes distractors that are between the depths of the near and far target and that could occlude the far target. We used the tunnel only in front of the far target because that was sufficient to equate the expected values of visibility of the two targets for any depth difference. This illustration does not show the random variations in the distractors positions and orientations. (b) In Experiment 2, we matched the distractors that fell in front of the targets (see highlighted black square). See Appendix 3.C for details.	46
3–2	Stimuli for dark-means-deep (DLC-) conditions with a black background. Rows represent distractor color and columns represent target color. On the diagonal, targets and distractors have the same color.	49
3–3	Stimuli for dark-means-near (DLC+) with a white background. On the diagonal, targets and distractors have the same color.	50
3–4	Mean and standard errors for Exp. 1 and 2 (depth) and Exp. 3 (luminance). The stimuli of Exp. 3 were the same as Exp. 1. The $3 \times 3$ layout of the plots corresponds to the conditions shown in Fig. 3–2 and 3–3	56
3.C.	1 (a) The visibility plot compares the mean target visibilities as a function of depth for Experiments 1 and 2. 1000 scenes were used for each data point. Target depths were chosen from $\{0, 2, 4,, 20\}$ and target visibilities were defined as the fraction of the target that was visible in the image (which we measured by counting pixels). The error bars show standard deviations (and not standard error). (b) The plot compares the mean absolute difference of target visibilities for Experiments 1 and 2 using 1000 scenes for each data point. Target depth differences $Z_{far} - Z_{near} = \Delta Z$ were chosen from $\{0, 4, 8,, 20\}$ . The standard deviations are almost identical for Experiments 1 and 2, so to avoid overlap they are shown only for Experiment 1 (with error bars)	67
4–1	Example stimuli for Experiments 1 and 3 (front-back and left-right, WB) at level $\Delta \eta = 0$ . The 3× 3 layout corresponds to Table 4–1. The arrows indicate the direction in which the variables $A, \eta, \lambda$ increase	84
4–2	Examples of stimuli for Experiments 1-4 at levels $\Delta \eta = \pm \frac{\eta}{2}, \pm \eta$ . In each case, the values of density $\eta$ and area <i>A</i> correspond to the middle condition in Table 4–1.	87

4–3	Biases $\alpha$ for human observers in Experiments 1 and 2. The nine plots correspond to the conditions in Table 4–1. Error bars indicate standard error of the mean. The gray lines indicate the $\Delta \eta$ levels in Equation 4.2 which were the levels used in the experiment. The black lines indicate the $\Delta \eta$ levels for which the mean number of pixels of front and back halves are equal. As we discuss later, these levels correspond to thresholds $\tau$ of a model observer that compares the image occupancies of the two halves.	92
4-4	Mean image occupancy of (a) front and back halves for Experiments 1 and 2, and (b) left and right halves for Experiments 3 and 4, for each of the nine conditions of Table 4–1 and for equal density in the two halves ( $\Delta \eta = 0$ ). The means are over 10,000 scenes. Occupancy is defined by the number of pixels of the corresponding front, back, left or right half sphere, divided by the number of pixels occupied by the image projection of the spherical bounding volume. Error bars show the standard deviation (not standard error) multiplied by a factor of 10 to better illustrate their relative magnitudes.	93
4–5	Mean Weber fractions for human observers for all four experiments. The nine plots correspond to the conditions in Table 4–1. Error bars show standard error of the mean.	98
4–6	Mean Weber fractions for model observers using 1-frame and 3-frame for front-back and left-right tasks. The nine plots correspond to the conditions in Table 4–1. Error bars shows standard error of the mean.	103
5–1	GLOW stimuli. (a) Depth discrimination of two targets with centered GLOW. (b) Depth discrimination of two virtual light sources	118
A-1	Visual Calibration (note the above image is scaled 60%). The left field is a rectangular wave grating with duty cycle 0.5. The voltage of the right rectangle is adjusted until the two fields appear equally bright. This is done to measure the relationship between voltage and displayed luminance.	123

### **ABBREVIATIONS**

- 2-D Two-Dimensional
- 3-D Three-Dimensional
- ANOVA ANalysis Of VAriance
- AO Ambient Occlusion
- DLC Depth-Luminance Covariance
- DLC- Negative Depth-Luminance Covariance
- DLC+ Positive Depth-Luminance Covariance
- DVR Direct Volume Rendering
- JND Just Noticeable Difference
- PSE Point of Subjective Equality

## CHAPTER 1 Introduction

#### 1.1 Motivation

Visualization of data is important for gaining insight to large amounts of information in problems arising in science, engineering and medicine. Typically, data sets are threedimensional (3-D) in nature and contain many overlapping surfaces that are distributed over a volume. We will refer to such scenes as *3-D cluttered scenes*. Examples from ecological perspectives may contain the foliage of bushes, shrubs, trees, as well as tall grass. Other examples can be found in the form of scientific visualizations, such as 3-D representations of biomolecules or of medical 3-D volume data such as a group of 2-D slice images obtained using magnetic resonance imaging (MRI). Such volume data may also contain semi-transparent surfaces to help reveal underlying structures in 3-D representations.

There have been few studies that examined how well human observers perceive 3-D cluttered scenes. When 3-D clutter contains opaque occluding surfaces, the visibility of underlying surfaces is reduced. Despite the challenge of occlusions, it has been shown that the visual system often is able to judge depth in 3-D clutter. For example, motion parallax and binocular disparity cues can be used to judge the depth order of two surfaces within the clutter [Langer, Zheng, and Rezvankhah, 2016], as well as the depth range of the clutter [van Ee and Anderson, 2001]. Langer et al. [2016] also showed that the visual system relies on probabilistic occlusion cues for judging depth order in 3-D clutter, where

the visibility of surfaces decreases with depth, and so the relative visibility of two surfaces within the clutter is a cue to their relative depth.

For volume data, such as 3-D medical data, a method known as Direct Volume Rendering (DVR) is commonly used to map data to color and opacity. In this case, occlusions may take the form of semi-transparent surfaces which increase the visibility of underlying structures, but may also lead to ambiguous spatial structure, such as ambiguous depth order [Boucheny, Bonneau, Droulez, Thibault, and Ploix, 2009] [Kersten, Stewart, Troje, and Ellis, 2006].

The motivation for the PhD research is to advance our perceptual understanding of 3-D cluttered scenes. Most cues from the human vision literature have not been extensively studied for viewing 3-D clutter. In this thesis, we explore cluttered stimuli that represent simplified, controllable forms of cluttered natural environments. In the manuscripts presented in Chapters 3 and 4 of the thesis, we explore how human observers use occlusion and luminance cues to perceive local and global structures in 3-D cluttered scenes. Specifically, we focus on two types of perceptual tasks: (i) a *depth* discrimination task where observers judge the relative depths of two target surfaces in the clutter, where the clutter obeys either a negative (dark-means-deep) or positive (bright-means-deep) depthluminance covariance (DLC) (Chapter 3) and (ii) a task where observers judge whether there is a spatial change in two halves of the cluttered volume. More specifically, we measure how well observers can detect differences in *density* between the two halves of the volume (Chapter 4). We address how the parameters of the clutter, namely the density and area of the surfaces, affect the perception of the clutter. We also compare DLC+/- to only having black or white delineate each half of the clutter. As part of the thesis work, we derived a DLC function that models ambient occlusion (AO) in 3-D cluttered scenes. The model is based on the ideas of Langer and Zucker [1994]'s cloudy day rendering model and Langer and Mannan [2012]'s probabilistic model of surface visibilities in 3-D clutter. The details of the derivation are found in Appendix 2.A. We use this DLC function for the stimuli in Chapter 3, where we examine both negative and positive signs of DLC in our experiments to determine when brighter surfaces appear closer. Note that the DLC functions used in the stimuli of Chapter 4 did not exactly follow this model. The Stimuli in Chapter 4 needed a DLC function that was more tailored to the task, which we describe in the Preface of Chapter 4.

#### **1.2 Organization of Thesis**

This thesis is presented in manuscript-based format, where Chapter 3 and Chapter 4 contain manuscripts that have been published or submitted to Journal of Vision. These manuscript chapters contain our work on depth and density discrimination, respectively. Before presenting the manuscripts, we include a preliminary Chapter 2 that includes back-ground information about ambient occlusion (AO), depth-luminance covariance (DLC), aerial perspective, transparency and volume rendering. These topics form the basis of our motivation to pursue perceptual studies on complex 3-D cluttered scenes. Chapter 2 has an Appendix 2.A where we derive AO models for 3-D clutter. Note that Appendix 2.A.1 has also been published in Appendix 3.B of the manuscript in Chapter 3. Chapter 2, Section 2.2 describes our early stimuli that explored opaque versus transparent, and DLC versus equiluminant occluder conditions. The insights we have gained from these stimuli have set the motivation for the work in our manuscripts of Chapters 3 and 4. In Chapter 5, we

present a comprehensive discussion and conclusion on all our findings. We also include Appendix A which outlines our procedure for gamma correction.

#### **1.3** Contribution to Original Knowledge

Our results contribute to our knowledge of the underlying performance of human visual perception of 3-D cluttered scenes. Our findings may also offer insights to the fields of computer graphics and visualization, in the form of guidelines for displaying 3-D cluttered scenes. The main contributions of our work are as follows:

- In the Appendix 3.B of Chapter 3, we present a depth-luminance covariance model for 3-D cluttered scenes that is based on Langer and Zucker [1994]'s cloudy day rendering model and Langer and Mannan [2012]'s probabilistic model of surface visibilities in 3-D clutter. This model can be used as an approximation to ray-cast ambient occlusion rendering. For more details, also consult Appendix 2.A of the thesis.
- In Chapter 3, we present new ideas for investigating and understanding human depth perception from luminance variations in 3-D clutter, where the observer's task was to discriminate the depth of two target surfaces embedded in clutter. We investigated interactions between occlusions, DLC-/+, and color for depth discrimination. We found that when the colors of the targets are different from distractors, human observers have a prior for DLC- (dark-means-deep). When the colors of targets and distractors were the same, human observers perform similarly well in DLC- (dark-means-deep) and DLC+ (bright-means-deep) conditions.
- In Chapter 4, we present a density discrimination study in 3-D clutter that addresses occlusion effects. The observer's tasks were to discriminate density in the front

versus the back half, or the left versus the right half of a cluttered volume. We measured bias and sensitivity for different parameters of the clutter, namely the density and area of the surfaces, the level of occlusion, and the type of luminance variation. For human observers, we show that the bias depends on the level of occlusion: when the level of occlusion is low, the bias to judge the back as denser is consistent with previous studies that used overlaid planes, and when the level of occlusion is higher, the bias crosses over to the front. Using a white-black luminance that clearly segmented the two halves reduces the front bias over using DLC. Weber fractions are also lower for white-black than for DLC. Weber fractions for human observers decrease as density increases for both front-back and left-right tasks which is consistent with previous work. The area of the elements did not affect Weber fractions for the front-back task, perhaps due to competing effects between the occlusion level and the likelihood of depth reversals. Weber fractions increase as area increases for the left-right task due to increased occlusion. We compared human observers that judge density to model observers that judge the image occupancies of the two halves against a known expected difference. For model observers, we show that the expected difference for the front-back follows a similar trend as the biases of the human observers, with a roughly constant offset between them. The trend in Weber fractions for model observers is similar human observers; however, for model observers the trend can be explained by the variation in the number of pixels in the two halves, i.e. when density increases and area decreases, there is less variation in the number of pixels which results in lower Weber fractions.

# CHAPTER 2 Background

In this chapter, we present the motivation for our research and describe previous work to provide context for the manuscripts presented in Chapters 3 and 4. We describe pertinent optical models, as well as their applications for rendering cluttered scenes.

#### 2.1 Ambient Occlusion

#### 2.1.1 Introduction

In order to render 3-D images, we rely on optical models that define the relationship between light and the particles in the volume. An optical model is useful in applications which map 3-D data to physical quantities that describe light interactions at the respective point in 3-D space. An example application is Direct Volume Rendering (DVR), which maps measurements from imaging devices or simulations to optical properties, such as color and opacity.

In the interaction between light and matter, light may be absorbed, scattered, or emitted by the medium. Because the solution to the complete light transport equation is computationally expensive, simplified models are commonly used. Models that only consider the emission-absorption components are generally used for DVR. There also exists a technique known as Aerial Perspective (or fog) that simulates the scattering of light through atmospheric particles by reducing the contrast of an object relative to its background as its distance to the viewer increases. Kersten et al. [2006] and Kersten-Oertel et al. [2014] have shown that fog is an effective depth cue for volumetric scenes with many occluding surfaces that are opaque or semi-transparent.

The models above only convey illumination cues within the volume particles without taking into account illumination effects from external light sources such as shadows. Such effects can provide additional depth cues, enhance the perception of small-scale spatial structures, and introduce greater realism. There exist global illumination techniques to compute these effects which are unfortunately expensive and therefore unsuitable for interactive rendering. However, there does exist a cheap approximation to global illumination known as *ambient occlusion* for adding shadows to objects lit with environment lighting [Christensen, 2002, Gritz et al., 2002, Landis, 2002]. Soft shadows are provided by darkening surfaces that are partially visible to the environment. The overall effect mimics how a scene would appear on a cloudy day and has been shown to enhance depth perception. Langer and Zucker [1994] showed that under diffuse lighting, such as on a cloudy day, surface concavities tend to be darker because the fraction of the diffuse source that is visible tends to be lower in a concavity. Langer and Bülthoff [2000] showed that, for smooth surfaces rendered under uniform diffuse illumination, the visual system takes account of luminance and, to some extent, surface normal variations when comparing depths of neighbouring surface points.

True ambient occlusion (AO) is traditionally obtained by estimating the percentage of rays cast about a hemisphere that reach the skylight. This is given by the ambient occlusion integral defined as follows. The luminance of a surface point  $\mathbf{X}_{\mathbf{p}} = (X_p, Y_p, Z_p)$  depends on the amount of hemispheric sky that is visible from it. The directions of the hemispheric sky are parametrized by polar angle  $\theta$  and azimuth  $\phi$ , and  $V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)$  is the visibility function which equals 1 if the hemispheric sky is visible from point  $\mathbf{X}_{\mathbf{p}}$  in direction  $(\theta, \phi)$ , and 0 otherwise. This yields the following expression for luminance at a surface point  $\mathbf{X}_{\mathbf{p}}$ ,

$$L(\mathbf{X}_{\mathbf{p}}) = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} V(\mathbf{X}_{\mathbf{p}}, \theta, \phi) \cos \theta \sin \theta d\theta d\phi.$$
(2.1)

The double integral above integrates over concentric circles of radius  $\sin \theta$  to obtain the surface area of the unit hemisphere. This area is weighted by Lambert's cosine law, which states that luminance from a Lambertian surface is proportional to the cosine of the angle  $\theta$  between the direction of the incident light ray and the surface normal. The total surface area of the cosine-weighted hemisphere is  $\pi$ , but we integrate only over the visible portions of the hemisphere. Dividing this result by  $\pi$  returns the luminance, a value between 0 and 1.

Equation 2.1 is difficult to solve analytically. A common technique is to sample rays using Monte Carlo sampling: this randomly samples cosine-weighted light directions over the hemisphere. The technique is by definition heavily dependent on scene geometry, because of ray-object intersections, and becomes increasingly expensive as the number of surfaces increases.

There exist several approximations of AO in the literature. Langer and Zucker [1994] derive constraints for surface aperture where ray directions are bounded by a cone centered on the surface normal. Another method known as directional occlusion shading (DOS) [Schott et al., 2009] also uses a cone instead of a hemisphere but the cone is centered along the viewing direction instead of the surface normal. Miller [1994] considers the geometric local and global accessibility of a point in order to darken deep areas of the scene that are not easily accessible (accessibility shading). Similarly, Stewart [2003]

shades a surface point according to uniform diffuse lighting that is blocked only by nearby occluders (vicinity shading). In screen-space ambient occlusion (SSAO) [Kajalin, 2009], shading happens in the pixel shader where pixel depth is used to form an AO map, there-fore applying AO as a post-processing step that is independent of scene complexity.

Some groups have derived mathematical models to approximate ray-cast AO for foliage-like cluttered scenes consisting of hundreds or thousands of small occluders. Researchers of botany have developed numerical models that describe how light penetrates different levels in a canopy and how the distribution of leaves or crops affects illumination within the canopy [Nilson, 1971] [Cescatti, 1997]. Reeves and Blau [1985] approximated the amount of light falling on a tree leaf through leaf self-shadowing. Luminance depended on the depth of the leaf and decreased exponentially with depth.

Hegeman, Premoze, Ashikhmin, and Drettakis [2006] derived a qualitative model of AO for trees based on the ideas in Reeves and Blau [1985]. They created a model to obtain a visually plausible estimation of AO that may not necessarily be physically accurate. They imposed certain restrictions, such as assuming that a tree is contained within a sphere and assuming that luminance varies as a function of tree density only and not occluder (leaf) size and shape.

#### 2.1.2 Theoretical Models

We present a high level description of our AO models, and we invite the reader to refer to Appendix 2.A for details of the derivation. The models we present borrow from the ideas of Nilson [1971], Cescatti [1997], Reeves and Blau [1985], and Hegeman et al. [2006]. Our models are also based on the ideas of Langer and Mannan [2012] that show how the probabilities of visible gaps, as viewed from the perspective of an observer, depend on the area *A* and density  $\eta$  of surfaces in the clutter, where the surfaces are assumed to be disks. The parameters *A* and  $\eta$  define a visibility factor  $\lambda = \eta A$ . For clutter that begins at depth  $Z = Z_0$ , the probability that a point **X**<sub>p</sub> at depth  $Z_p$  is visible is,

$$p(V(\mathbf{X}_p)) = exp\{-\lambda(Z_p - Z_0)\}.$$
(2.2)

This probabilistic model assumes that the distribution of the clutter elements follows a Poisson process and that the disks are frontoparallel, i.e. the surface normal face the Z direction.

If we consider the visibility function  $V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)$  in Equation 2.1 as a random variable, we can compute the probability that the hemispheric sky is visible along a ray using Equation 2.2. By plugging Equation 2.2 into Equation 2.1, we can obtain the expected luminance of a surface point  $\mathbf{X}_p$  as follows. When integrating over ray directions  $(\theta, \phi)$ , the term  $1/(\cos \theta)$  compensates for the length of ray passing through the clutter at incidence angle  $\theta$ . We assume the disks have surface normal facing the *Z* direction and so the area of these disks are foreshortened by  $\cos \theta$ . The cosine terms cancel out, and the solution to the integral in Equation 2.1 is,

$$E[L(\mathbf{X}_{\mathbf{p}})] = exp\{-\lambda(Z_p - Z_0)\}.$$
(2.3)

We use the model above as our DLC function in Chapter 3, where the luminance of a surface decreases exponentially with its depth  $Z_p$ . The model above is based on strong assumptions such that the clutter is uniform and infinite along XY as though it is an infinite hedge, and that the clutter elements are disks that are frontoparallel. While it is not physically accurate, it still captures a shadowing effect that is perceptually acceptable as we shall explore in the next subsection. See Figure 2–1 (a) for an example of ray-cast AO in a cluttered hedge (which was a time-consuming computation) and see Figure 2–1 (c) for AO rendered using the above model (Equation 2.3).

We also derived AO models for when occluders are randomly oriented in the scene. Our probabilistic visibility function is slightly different because when the disks are randomly oriented, the foreshortening of the disks will not depend on  $\cos \theta$ . Rather they will depend on a constant factor. A randomly oriented disk is expected to have half the projected area of a frontoparallel surface when a disk is projected on a plane perpendicular to ray direction ( $\theta$ , $\phi$ ). Therefore,  $\lambda = \frac{\eta A}{2}$  and,

$$p(V(\mathbf{X}_{\mathbf{p}}, \boldsymbol{\theta}, \boldsymbol{\phi})) = exp\{-\lambda \frac{(Z_p - Z_0)}{\cos \theta}\}.$$
(2.4)

So far, we have considered disks that were only lit from the front of the volume at depth  $Z = Z_0$ . When disks are randomly oriented, they can receive light from both the front and the back of the volume located at depth  $Z = Z_1$ . The probability that a ray cast from point **X**<sub>p</sub> to the back of the volume reaches the sky is:

$$p'(V(\mathbf{X}_{\mathbf{p}}, \boldsymbol{\theta}, \boldsymbol{\phi})) = exp\{-\lambda \frac{(Z_1 - Z_p)}{\cos \theta}\}.$$
(2.5)

To determine the expected luminance of a point, we consider rays cast both toward the front and the back of the volume using a separable equation, where the first term accounts for the front rays and the second term accounts for the back rays. Let  $\theta_n$  be the angle between the disk's surface normal and the Z direction. In the following equation, the cosine terms account for the horizon cut-off, e.g. any rays that are cast at an angle less than  $\frac{\pi}{2}$  are considered front rays and any rays that are cast at an angle greater than  $\frac{\pi}{2}$  are



(a)

(b)



Figure 2–1: Example ray-cast AO rendering of (a) Infinite Hedge and (b) Sphere. Example qualitative AO model rendering of (c) Infinite Hedge and (d) Sphere.

considered as back rays [Horn and Sjoberg, 1979],

$$E[L(\mathbf{X}_{\mathbf{p}})] = \cos^2(\frac{\theta_{\mathbf{n}}}{2}) * \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} p(V(Z_p, \theta, \phi) \cos \theta \sin \theta d\theta d\phi$$
(2.6)

$$+\cos^{2}\left(\frac{\pi-\theta_{\mathbf{n}}}{2}\right)*\frac{1}{\pi}\int_{0}^{2\pi}\int_{0}^{\pi/2}p'(V(Z_{p},\theta,\phi))\cos\theta\sin\theta d\theta d\phi.$$
 (2.7)

Note that the above models need to be numerically evaluated as they do not have closed-form solutions like the model in Equation 2.3. In the next subsection, we compare the above models to ray-cast AO. We used scenes consisting of either frontoparallel or randomly oriented square surfaces. We show that the models in Equations 2.6 and 2.7 capture ray-cast AO in clutter, but that AO can also be approximated using the closed-form Equation 2.3 because the surface orientation minimally affects luminance in 3-D clutter.

In Appendix 2.A, we also present the sphere model analogue to the infinite hedge model. The hedge model assumes infinite clutter on XY, whereas the clutter contained in the sphere is finite. We will similarly compare the theoretical model to ray-cast AO for cluttered spheres in the Appendix 2.A. Figure 2–1 (b) and (d) compare the ray-cast AO and the qualitative AO model rendering of the sphere, respectively.

#### 2.1.3 Simulation Results

We ran simulations over 100 scenes to compare our theoretical models to ray-cast ambient occlusion. Our model was evaluated using Matlab 2016a. To evaluate the model, we discretized the double integral by  $k = k_1k_2$  light directions  $\omega_{l=1,...,k} = (\theta_{i=1,...,k_1}, \phi_{j=1,...,k_2})$ ,

where 
$$\Delta \theta = \frac{\pi}{2k_1}$$
, and  $\Delta \phi = \frac{2\pi}{k_2}$ , and  $\theta_i = i\Delta \theta$  and  $\phi_j = j\Delta \phi$ ,  

$$\int_0^{2\pi} \int_0^{\pi/2} p(V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)) \cos \theta \sin \theta d\theta d\phi \approx \frac{1}{\pi} \sum_{j=1}^{k_2} \sum_{i=1}^{k_1} p(V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)) \cos \theta_i \sin \theta_i \Delta \theta \Delta \phi.$$
(2.8)

For the discretized integral, we used  $k = k_1k_2 = 1000 * 1000$ . The larger the k values, the finer the discretization, and the better the approximation of the integral. We evaluate the integral at a series of points over an interval, which we define below.

Our ray-cast scenes contained many occluding square surfaces. We used clutter density  $\eta = 0.125$  square centres per cm<sup>3</sup>. Our theoretical models are based on disks and so the average projected area of an unoccluded square at a given depth would need to be equal to the area of an unoccluded disk at that depth [Langer and Mannan, 2012]. In the scene, we used width w = 1 cm for each square, and in the model we defined  $A = \frac{w^2}{2}$  for randomly oriented clutter. *X* and *Y* ranges were both 80 cm in length, and the *Z* range was 20 cm. The *X* and *Y* ranges were made larger than the *Z* range because the model assumes infinite clutter about *X* and *Y*. Otherwise, the empirical data would deviate from the model because of an increasing number of rays hitting the *X* and *Y* walls. We examined squares along the center of the volume (X, Y) = (0, 0) and at eleven specific depths between Z = [0, 20] cm. Note that in viewer coordinates, the depths are between Z = [60, 80] cm because the viewer is located at Z = 0, the near plane is located at  $Z_0 = 60$  cm, and the far plane is located at  $Z_1 = 80$  cm.

We obtained values from ray-cast ambient occlusion using a cosine-weighted hemisphere with 128 rays. This was costly to render, but we did so in order to obtain true AO values to compare to our model. In Figure 2–2, we plotted luminance values for frontoparallel squares from Equation 2.3 for the purpose of showing that it can approximate the model in Equation 2.6. In Figure 2–3, we plotted luminance values for center squares at the eleven specific depths, with normals oriented at  $\frac{\pi}{6}$ ,  $\frac{\pi}{3}$ , and  $\frac{\pi}{2}$  radians to the *Z*-axis. To account for horizon cut-off in our simulation, we consider the following: for front-lit hedge (Equation 2.6), rays are only cast frontward (all rays that are greater than  $\frac{\pi}{2}$  radians from the *Z* direction are considered occluded), while for back-lit (Equation 2.7), rays are only cast backward. For front-and-back-lit (the summation of Equations 2.6 and 2.7), rays are cast in all directions. The mean luminance for each point was obtained over 100 scenes.

#### 2.1.4 Discussion

There is a non-zero standard deviation in the plots for the AO model because the luminance for the ray casting will not depend only on the orientation and depth of surfaces; it will also depend on the details of the particular clutter. As the density increases and the area of the occluders shrinks (holding  $\lambda$  constant) the standard deviation of the luminance will decrease. When density is larger and area is smaller, there is less variability in the number of pixels the square occupies [Langer and Mannan, 2012]. To put it another way, larger squares subtend larger visual angles, and thus variations in position and orientation of the squares contribute to a greater variance in the number of pixels occupied. This concept is related to some ideas we present in Chapter 4, where we show how higher standard deviation due to larger squares results in a decrease in model observers' sensitivity to differences in image occupancy.



Figure 2–2: Mean luminance as a function of depth for a surface with normal parallel to the *Z* direction with frontoparallel (red) and randomly oriented (green) surrounding clutter. We plotted both the true ray-cast AO luminance values (dots), as well as the modeled luminance values (lines) of Equations 2.3 and 2.6, for frontoparallel and random orientated occlusions respectively. Error bars show standard deviation for 100 scenes.



Figure 2–3: Mean luminance as a function of depth for orientations  $\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}$ , as per the (a) front-lit model in Equation 2.6, (b) back-lit model in Equation 2.7 and (c) front-and-back-lit models in Equations 2.6 + 2.7. Error bars show standard deviation for 100 scenes.

Figure 2–4 shows standard deviation increasing as occluder width increases. We plotted standard deviation for ray-cast AO in the hedge with randomly oriented squares for 100 scenes. We varied the width of the squares over 0.25, 0.5, 0.75, and 1 cm with corresponding densities 2, 0.5, 0.22, 0.125 occluder centers per cm<sup>3</sup>. Higher densities for smaller occluders became very expensive to compute; however we expect that as the occluder area reduces to zero (in which case occluders would just be a set of points), the standard deviation also reduces to zero.



Figure 2–4: Standard deviation over 100 scenes as a function of occluder width while keeping  $\lambda$  constant. When  $\lambda$  is kept constant, the standard deviation increases as the density decreases and the area of the occluders increases.

#### 2.1.5 Conclusion

In random cluttered scenes, we can use the models to render AO as an alternative to raycasting which is an expensive computation in scenes containing many small occluding surfaces. For the cases where there is no closed-form solution, one can numerically evaluate the AO integral over a set of inputs, and store these in a precomputed look-up table.

If the goal is not to be 100% physically accurate but rather to avoid precomputation and run-time costs while still getting a visually plausible result, one can compute a best fit to the solution of the integral. For an even faster solution, one can use the approximate closed-form solution as seen in Equation 2.3. This closed-form solution represents a depthluminance covariance (DLC) where luminance varies as a function of depth. In order to allow luminance to vary with both depth and orientation, we can multiply the equation by the cosine terms to obtain the luminance for any orientation. With such separable equations, we conclude that for random scenes containing many small occluding surfaces, AO is simply a Depth-Luminance Covariance (DLC) modulated by orientation.

In Chapter 3, we evaluate the use of DLC as a cue to depth. In pilot experiments, we tested whether the depth of two target surfaces embedded in clutter was perceived differently in clutter that was rendered with a simple DLC model in Equation 2.3 versus a model that more accurately captures AO such as the modulated DLC in Equations 2.6 + 2.7. We found that observers were not so sensitive to the subtle luminance variations caused by the orientations of the surfaces in our cluttered stimuli. The two models are almost the same, with orientation factored in for the latter. Thus, it was not surprising that observers performed the same when the models essentially produced the same luminance. Therefore, we chose to simply use DLC in Equation 2.3 to study how luminance variations affect human observers' perception of depth in 3-D clutter.

#### 2.2 Transparency

#### 2.2.1 Introduction

Renderings of clouds, smoke, medical data, etc. may contain elements that are semitransparent. There are different ideas on how the human visual system perceives transparency. Early theories are based on Metelli [1974]'s physical model of reflectance and transmittance. The physical model is based on a disc with an open sector that rotates over a bipartite background. When the rotation is very fast, it leads to the percept of a transparent surface. Based on this physical set-up, Metelli derived models for the reflectance and transmittance of the transparent surface, as well as the physical constraints that must be met to achieve perceptual transparency.

Singh and Anderson [2002] suggests that the human visual system makes errors in estimating the transmittance of a transparent surface, and so perceptual transparency should not be based on physical properties such as transmittance and reflectance. They argue that perceived transmittance is based on the Michelson contrast, i.e. it is proportional to the ratio of the contrast of the transparent region to the contrast of the underlying surface.

Anderson and Winawer [2005] present illusions that show identical texture patches appearing either white or black depending on their surroundings. For the dark surrounding the patches appear white and partially occluded by dark semi-transparent clouds; for the light surrounding the same patches appear black, visible through light semi-transparent clouds. This suggests that the human visual system segments an image based on contrast. That is, the regions that produce the highest contrast against their surrounds are seen in plain view, and the lower contrast areas are seen through a contrast-reducing medium.
Layered random dot stereograms can be thought of as partly transparent layers, i.e. the opacity within each layer is randomly chosen to be either 0 or 1. Stereo matching for such displays is generally regarded to be difficult [Akerstrom and Todd, 1988], although Tsirlin et al. [2008] showed that humans can segregate up to six simultaneous overlaid surfaces.

Typical studies on transparency considered only few layers. Volume-rendered scenes are generally more complex and may contain many occluding semi-transparent surfaces over a continuum of layers rather than a finite discrete set. Our experiments below will explore these type of stimuli. Although volume rendering is a classical technique in computer graphics and visualization, only a few studies have examined how well it provides quantitative depth information. Boucheny et al. [2009] conducted experiments involving depth discrimination in semi-transparent objects obtained with DVR. Subjects were to determine the depth order of two semi-transparent cylinders that had different widths, different luminances, and constant opacity. When the two cylinders were displayed statically, the observers had difficulty interpreting the transparent volumes correctly. However, in a dynamic context where the object rotated around a vertical axis, viewers were provided with strong dynamic cues to depth and were able to answer correctly. But this required careful tuning of luminance and constant opacity. Similarly, Kersten et al. [2006] compared monocular versus stereo viewing of a rotating semi-transparent cylinder, and asked subjects to judge the direction of rotation. Performance was at chance in a monocular viewing condition but well above-chance under stereo viewing.

In the next section, we describe stimuli from our initial attempts for experiments that explore depth-luminance covariance cues in semi-transparent versus opaque clutter.

#### 2.2.2 Stimuli

Prior to our manuscript work in Chapters 3 and 4, we provide a high-level description of our initial attempt for experiments. We tested how human observers perceive opaque versus semi-transparent surfaces which often are used in volume rendering. We also compared two luminance conditions: DLC- and uniform luminance. We viewed the stimuli using a stereo display coupled with head-tracking.

Our scenes consisted of two red target squares surrounded by  $11^3$  occluding surfaces. We refer to these as distractors. The task was to determine which of the target squares is closer to the human observer. Our stimuli had either opaque distractors, or partially transparent distractors with opacity  $\alpha = 0.4$ . See Figures 2–5 and 2–6 for examples of the transparent cluttered scene and opaque cluttered scene, respectively.

The red target squares always had luminance (1,0,0). However, we manipulated the luminances of the distractors. We either distributed the luminances according to negative depth-luminance covariance (DLC-) i.e. deeper surfaces are darker, or we used equiluminant surface elements (namely 0.8 in [0,1]). For the DLC- condition, a mapping from depth to surface luminance was chosen. These stimuli were designed before we developed models to capture AO in 3-D clutter as seen in Section 2.1, and so we chose to use a simple function where equal steps in depth gave equal steps in brightness (see the DLC function used in Chapter 4). Stimuli were gamma-corrected.

# 2.2.3 Discussion

The motivation for our stimuli was to determine whether DLC- would improve the depth perception of the red squares. In these particular stimuli, we were interested in determining whether DLC- could indirectly aid human observers in their task of discriminating



Figure 2–5: Transparent occluders. Side blue opaque occluders were used to prevent depth reversals.



(a) Uniform

(b) DLC-



the depths of the two red squares whose colors remained unchanged. We found that the thresholds for depth discrimination were lower (i.e. performance was better) for DLC-than uniform for opaque occluders. This suggests that DLC cues not only improve the 3-D depth perception of the distractors themselves as expected from previous DLC studies, DLC may also improve the depth perception of the targets that are embedded within the clutter. The idea here is that the distractors could provide a spatial frame of reference in which the depths of the targets can be compared. This pilot study provided limited evidence in support of this intriguing result and so we conducted further experiments in Chapter 3 to confirm and elaborate on it. In those experiments, we examine both negative and positive signs of DLC (DLC-/+), and we vary the luminance and colors of the targets, and background. We examine how the DLC sign of the occluders can influence when the brighter target appears closer or farther.

For the transparent conditions (i.e. opacity condition with  $\alpha = 0.4$ ), DLC- apparently offers no benefit over uniform. The reason DLC- did not improve performance in the transparent case may be that the visual system has difficulties disentangling transparency effects from surface color effects. Similar difficulties have been shown even in scenes with two layer transparency [Singh and Anderson, 2002], so it should hardly be a surprise that these difficulties exist for our cluttered scenes.

We used a display that combined motion parallax and stereo, as this has been shown to improve depth perception. Cho et al. [2014] carried out depth discrimination and ordering tasks in which the stimuli consisted of volumetric tubes that mimicked a 3-D medical scan of networks of blood vessels. Stereo was found to be a stronger cue than motion for volumetric data, and stereo and head-tracking together provided better performance than each of these cues on their own. However, a study by Kersten-Oertel et al. [2014] suggested that shading cues such as aerial perspective give better depth perception than these cues. We decided to use mono displays for the rest of the thesis, and focus on the luminance cues. In Chapter 3, we show that human observers are very capable of depth discrimination in monocularly-viewed scenes with DLC.

# Appendix

# 2.A Derivation of Ambient Occlusion Models in 3-D Clutter

## 2.A.1 Hedge Models

# **Front-Lit**

Our DLC- model combines the ideas of Langer and Zucker [1994]'s cloudy day rendering model and Langer and Mannan [2012]'s probabilistic model of surface visibilities in 3-D clutter. The former model assumes the scene is illuminated by a uniform hemispheric sky, centered in the Z axis direction. The latter model assumes a 3-D cluttered scene which begins at depth  $Z = Z_0$  and has infinite extent in the X and Y directions.

Under the cloudy day rendering model, illumination is assumed to be diffuse and non-directional. The luminance variations consider cast shadows only, but ignore interreflections. Thus, the luminance of a surface point  $\mathbf{X}_{\mathbf{p}} = (X_p, Y_p, Z_p)$  depends on the amount of hemispheric sky that is visible from it. The directions of the hemispheric sky are parametrized by polar angle  $\theta$  and azimuth  $\phi$ , and  $V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)$  is the visibility function which equals 1 if the hemispheric sky is visible from point  $\mathbf{X}_{\mathbf{p}}$  in direction ( $\theta, \phi$ ), and 0 otherwise. This yields the following expression for luminance at a surface point  $\mathbf{X}_{\mathbf{p}}$ ,

$$L(\mathbf{X}_{\mathbf{p}}) = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} V(\mathbf{X}_{\mathbf{p}}, \theta, \phi) \cos \theta \sin \theta d\theta d\phi.$$
(2.9)

The double integral above integrates over concentric circles of radius  $\sin \theta$  to obtain the surface area of the unit hemisphere. This area is weighted by Lambert's cosine law, which states that luminance from a Lambertian surface is proportional to the cosine of the angle  $\theta$  between the direction of the incident light ray and the surface normal. The total surface area of the cosine-weighted hemisphere is  $\pi$ . We integrate only over the visible portions of the hemisphere and so dividing the result by  $\pi$  returns the luminance, a value between 0 and 1.

For a 3-D cluttered scene, we consider visibility  $V(\mathbf{X_p}, \theta, \phi)$  as a random variable. We compute the probability that the hemispheric sky is visible along a ray,  $p(V(\mathbf{X_p}, \theta, \phi))$ , as done in Langer and Mannan [2012]. To get a closed form model of  $p(V(\mathbf{X_p}, \theta, \phi))$ , we assume the elements of the clutter are disks of area *A*, and we assume that the spatial distribution of the disks is a Poisson process with density  $\eta$  which is the average number of disk centers per unit volume. The parameters *A* and  $\eta$  can be lumped together as a single constant  $\lambda = \eta A$ .

The function  $p(V(\mathbf{X_p}, \mathbf{\theta}, \mathbf{\phi}))$  can be written in terms of  $Z_p$  only as follows. The length of any ray from  $\mathbf{X_p} = (X_p, Y_p, Z_p)$  to the edge of the clutter is  $(Z_p - Z_0)/\cos \theta$ , where the term  $1/(\cos \theta)$  accounts for the greater path length of a ray with greater polar angle  $\theta$ . To simplify the integral, we assume the elements of the clutter have surface normal facing the Z direction and so the area of these clutter elements are foreshortened by  $\cos \theta$  in the ray direction ( $\phi, \theta$ ). Then, the Poisson model yields:

$$p(V(Z_p, \theta, \phi)) = exp\{-\lambda \cos \theta (Z_p - Z_0) / \cos \theta\}.$$
(2.10)

The  $\cos \theta$  terms cancel out and so  $p(V(Z_p))$  depends only on depth,

$$p(V(Z_p)) = exp\{-\lambda(Z_p - Z_0)\}.$$
(2.11)

Then, the expected visibility of the sky in direction  $(\theta, \phi)$  is computed as,

$$E[V(Z_p)] = 1 \cdot p(V(Z_p)) + 0 \cdot (1 - p(V(Z_p))) = exp\{-\lambda(Z_p - Z_0)\}.$$
(2.12)

Using the definition of luminance of a surface point  $L(\mathbf{X}_{\mathbf{p}})$ , we use the above to compute the expected luminance of a surface point  $E(L(\mathbf{X}_{\mathbf{p}}))$ . The linearity property of expectation allows us to integrate over the expected visibilities of rays along individual directions,

$$E[L(\mathbf{X}_{\mathbf{p}})] = E[\frac{1}{\pi} \int_{0}^{2\pi} \int_{0}^{\pi/2} V(\mathbf{X}_{\mathbf{p}}, \theta, \phi) \cos \theta \sin \theta d\theta d\phi], \qquad (2.13)$$

$$= \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} E[V(\mathbf{X}_{\mathbf{p}}, \boldsymbol{\theta}, \boldsymbol{\phi})] \cos \boldsymbol{\theta} \sin \boldsymbol{\theta} d\boldsymbol{\theta} d\boldsymbol{\phi}, \qquad (2.14)$$

$$= \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} E[V(Z_p)] \cos \theta \sin \theta d\theta d\phi.$$
(2.15)

The resulting expected luminance of point  $\mathbf{X}_{\mathbf{p}}$ , written in terms of  $Z_p$  only, is given by,

$$E[L(Z_p)] = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} exp\{-\lambda(Z_p - Z_0)\} \cos\theta \sin\theta d\theta d\phi, \qquad (2.16)$$

Integrating the last equation yields,

$$E[L(Z_p)] = exp\{-\lambda(Z_p - Z_0)\}.$$
(2.17)

This is the rendering model that we use for DLC-, namely luminance is chosen to be proportional to  $E[L(Z_p)]$ . The rendering model is based on several assumptions. First, the illumination arrives from a hemisphere centered at the Z axis, and this illumination hemisphere has uniform luminance over all directions. Second, the clutter has infinite XY extent, which is not the case for cluttered scenes in our stimuli. Third, the luminance variations consider cast shadows only, but ignore interreflections. Fourth, the elements of the clutter have normals parallel to the Z direction. Because the model is based on rather strong assumptions, we cannot and do not claim that this model is photorealistic for the given scenes. Rather the model is meant to capture a qualitative shadowing effect that does occur in real scenes, namely when clutter such as foliage is illuminated under approximately diffuse light.

Thus far, we modelled the luminance of a frontoparallel surface surrounded by frontoparallel occluders. We can extend the above to model the luminance of a frontoparallel surface that is surrounded by *randomly oriented* occluders. When occluders are randomly oriented in the scene, the foreshortening of the disks will not depend on  $\cos \theta$  as as they previously did. Rather they will depend on a constant factor. A random oriented disk is expected to have half the projected area of a frontoparallel surface (when a disk is projected on a plane perpendicular to ray direction  $(\theta, \phi)$ ). Therefore we have  $\lambda = \frac{\eta A}{2}$ .

$$p(V(Z_p, \theta, \phi)) = exp\{-\lambda(Z_p - Z_0)/\cos\theta\},$$
(2.18)

and therefore,

$$E[L_0(\mathbf{X}_{\mathbf{p}})] = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} exp\{-\lambda(Z_p - Z_0)/\cos\theta\}\cos\theta\sin\theta d\theta d\phi.$$
(2.19)

Unlike Equation 2.16, the above integral does not have a closed-form solution. Therefore, we can numerically estimate it by evaluating the integral over a set of points in an interval. Note that the above integral still models AO for surfaces with normal parallel to the *Z* direction, but now the difference is that surrounding occluders are randomly oriented. To generalize the model to compute the luminance for a surface of *arbitrarily oriented normal* with randomly oriented occluders, we can approximate how luminance depends on both depth *d* and orientation  $\theta_n$  using a separable function:

$$E[L_{\theta_{\mathbf{n}}}(\mathbf{X}_{\mathbf{p}})] = \cos^2(\frac{\theta_{\mathbf{n}}}{2})E[L_0(\mathbf{X}_{\mathbf{p}})], \qquad (2.20)$$

where  $\theta_n$  is the angle between the surface normal and the *Z* direction. The cosine term accounts for the horizon cut-off, i.e. any rays that are cast at an angle greater than  $\frac{\pi}{2}$  are considered as occluded because they cannot penetrate the back of the volume [Horn and Sjoberg, 1979].

# **Back-lit**

The previous front-lit model only considered rays cast forward and cut off rays that penetrated the back of the volume. To consider the backward cast rays, we modify the probabilistic visibility function that a point at depth  $Z_p$  sees through the hedge by now considering distance from the back of the volume at depth  $Z_1$ :

$$p'(V(Z_p, \theta, \phi)) = exp\{-\lambda(Z_1 - Z_p)/\cos\theta\}.$$
(2.21)

Therefore,

$$E[L_{\pi-\theta_{\mathbf{n}}}(\mathbf{X}_{\mathbf{p}})] = \cos^{2}\left(\frac{\pi-\theta_{\mathbf{n}}}{2}\right)\frac{1}{\pi}\int_{0}^{2\pi}\int_{0}^{\pi/2}p'(V(Z_{p},\theta,\phi))\cos\theta\sin\theta d\theta d\phi.$$
(2.22)

#### **Front-and-Back Lit**

Recall in the last sections, we modelled AO for an infinitely deep hedge where hemispherical rays were cut-off at the horizon. If we instead consider a hedge of finite depth  $Z \in [Z_0, Z_1]$  where rays can penetrate both the front and back, we need to take into account all hemispherical rays.

We add Equations 2.20 and 2.22 for a total of front and back rays, again assuming that surrounding occluders are randomly oriented. We are essentially integrating over two hemispheres but applying different weights to the front and back hemispheres that depend on surface orientation, and the weights add up to 1. There is no contribution from  $E[L_{\pi-\theta_n}(\mathbf{X_p})]$  when  $\theta_n = 0$ :

$$E[L(\mathbf{X}_{\mathbf{p}})] = E[L_{\theta_{\mathbf{n}}}(\mathbf{X}_{\mathbf{p}})] + E[L_{\pi-\theta_{\mathbf{n}}}(\mathbf{X}_{\mathbf{p}})]$$
(2.23)

$$=\cos^{2}(\frac{\theta_{\mathbf{n}}}{2})*\frac{1}{\pi}\int_{0}^{2\pi}\int_{0}^{\pi/2}p(V(Z_{p},\theta,\phi)\cos\theta\sin\theta d\theta d\phi \qquad (2.24)$$

$$+\cos^{2}\left(\frac{\pi-\theta_{\mathbf{n}}}{2}\right)*\frac{1}{\pi}\int_{0}^{2\pi}\int_{0}^{\pi/2}p'(V(Z_{p},\theta,\phi))\cos\theta\sin\theta d\theta d\phi.$$
 (2.25)

If our goal is obtain a qualitative rendering which is simple but not necessarily physically accurate, we can approximate AO by replacing the integrals above with Equation 2.17 to give,

$$E[L(\mathbf{X}_{\mathbf{p}})] = E[L_{\theta_{\mathbf{n}}}(\mathbf{X}_{\mathbf{p}})] + E[L_{\pi-\theta_{\mathbf{n}}}(\mathbf{X}_{\mathbf{p}})]$$
(2.26)

$$\approx \cos^2(\frac{\theta_{\mathbf{n}}}{2}) * exp\{-\lambda(Z_p - Z_0)\}$$
(2.27)

$$+\cos^{2}\left(\frac{\pi-\theta_{\mathbf{n}}}{2}\right)*exp\{-\lambda(Z_{1}-Z_{p})\}.$$
(2.28)

where  $\lambda = \eta A$ .

### 2.A.2 Sphere Models

We consider a spherical volume with randomly distributed disks of radius *R*, such as leaves in a bush. Ambient light would penetrate the canopy from every direction. Let  $d_{\mathbf{n}}(\mathbf{X}_{\mathbf{p}}, \theta, \phi)$  be the distance from the disk's center to the sphere boundary in direction  $\omega = (\theta, \phi)$  centered on the disk's surface normal **n**. The probabilistic visibility function is

$$p(V(\mathbf{X}_{\mathbf{p}}, \boldsymbol{\theta}, \boldsymbol{\phi})) = exp\{-\lambda d_{\mathbf{n}}(\mathbf{X}_{\mathbf{p}}, \boldsymbol{\theta}, \boldsymbol{\phi})\}.$$
(2.29)

The directions  $\mathbf{u} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$  of the rays are cast about a hemisphere centered on the disk's surface normal  $\mathbf{n}$ . We can compute the distance d in our model by using the ray-sphere intersection point which would give the length of the ray. The parametric equation of the ray is  $\mathbf{x}(t) = \mathbf{p} + t\mathbf{u}$ . The expected luminance is given by,

$$E[L(\mathbf{X}_{\mathbf{p}})] = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} p(V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)) \cos \theta \sin \theta d\theta d\phi.$$
(2.30)

The above model reoriented the hemisphere so that it is centered about the disk's surface normal with arbitrary orientation relative to the +*Z*-axis. Alternatively, we can approximate the above using a separable equation as done in the hedge models, where we assume that the surface in question is frontoparallel (i.e.  $\mathbf{n} = (0,0,1)$ ) and account for orientation using the Horn and Sjoberg [1979] term. The luminance model is given by the following equation, where the first term accounts for the front-cast rays and the second

term accounts for the back-cast rays,

$$E[L(\mathbf{X}_{\mathbf{p}})] = \cos^{2}(\frac{\theta_{n}}{2}) * \frac{1}{\pi} \int_{0}^{2\pi} \int_{0}^{\pi/2} exp\{-\lambda d_{(0,0,1)}(\mathbf{X}_{\mathbf{p}}, \theta, \phi) \cos\theta \sin\theta d\theta d\phi \qquad (2.31)$$
$$+ \cos^{2}(\frac{\pi - \theta_{n}}{2}) * \frac{1}{\pi} \int_{0}^{2\pi} \int_{0}^{\pi/2} exp\{-\lambda d_{(0,0,-1)}(\mathbf{X}_{\mathbf{p}}, \theta, \phi) \cos\theta \sin\theta d\theta d\phi. \qquad (2.32)$$

Again, if we want a quick qualitative approximation, we can use the closed-form model in the place of the above integrals where we now replace the distance  $Z_p - Z_0$  in Equation 2.17 with the distance d from the point to the boundary of the sphere in the direction of the surface normal **n**,

$$E[L(\mathbf{X}_{\mathbf{p}})] \approx exp\{-\lambda d_{\mathbf{n}}(\mathbf{X}_{\mathbf{p}})\}.$$
(2.33)

Or if we want a separable equation,

$$E[L(\mathbf{X}_{\mathbf{p}})] \approx \cos^{2}(\frac{\theta_{n}}{2}) * exp\{-\lambda d_{(0,0,1)}(\mathbf{X}_{\mathbf{p}}, \theta, \phi)\}$$
(2.34)

$$+\cos^{2}\left(\frac{\pi-\theta_{n}}{2}\right)\exp\{-\lambda d_{(0,0,-1)}(\mathbf{X}_{\mathbf{p}},\theta,\phi)\}.$$
(2.35)

### **Sphere Simulation**

We compare ray-cast and model values for center squares at eleven specific depths for normals oriented at 0,  $\frac{\pi}{6}$ ,  $\frac{\pi}{3}$ ,  $\frac{\pi}{2}$  radians to the *Z*-direction (direction to the viewer).

For the raycasted data, 128 rays were cast from each square's center as usual. The radius of the sphere was 10 cm. We used 526 randomly oriented surfaces of width w = 1 cm, giving a density of  $\eta = 0.125$ . We chose the width w of each square so that the average projected area of an unoccluded square at a given depth would be equal to the area

of an unoccluded disk at that depth [Langer and Mannan, 2012], i.e.  $A = \frac{w^2}{2}$  for randomly oriented clutter.

See Figure 2.A.1 for plots of model simulation and ray-cast data. Equation 2.30 and Equation 2.31 + 2.32 produced similar results and so we display only results from the model in Equation 2.30.



Figure 2.A.1: Luminance as a function of depth for orientations 0 to  $\frac{\pi}{2}$  radians in a spherical volume. We show the ray-cast values and the model values from Equation 2.30. Error bars show standard deviation for 100 scenes.

# CHAPTER 3 Signs of depth-luminance covariance in 3-D cluttered scenes

# Preface

In this chapter, we examine the role of depth-luminance covariance (DLC) for the depth discrimination of two targets in 3-D clutter. We define the DLC function according to the model in Equation 2.3 of Chapter 2, which qualitatively captures a shadowing effect in real scenes. We use this model to examine both positive and negative DLC, where positive covariance means that deeper surfaces are brighter and negative covariance means deeper surfaces are darker. In Section 2.2, we previously addressed depth discrimination of two targets embedded in clutter, where occlusions obeyed a negative DLC. Our results suggested that human observers use the luminance information of the occluders to infer the relative depth of the targets. We elaborate on that result here. In these experiments, we let the luminance of the targets follow negative or positive signs of DLC. We aim to determine whether the occluders allow observers to infer the sign of the DLC in order to determine the relative depth of the two targets. We vary the color of the targets versus occluding surfaces, which we will refer to as *distractors* in the manuscript. We address a long-standing question in depth perception: when do brighter surfaces appear closer or farther? Our experiments are the first to address this question for 3-D cluttered scenes.

# Signs of depth-luminance covariance in 3-D cluttered scenes

Milena Scaccia<sup>1</sup> and Michael S. Langer<sup>2</sup> School of Computer Science, McGill University, Montreal, Quebec, Canada <sup>1</sup>milena.scaccia@mail.mcgill.ca

<sup>2</sup>langer@cim.mcgill.ca

Journal of Vision March 2018, Vol.18, 5. doi:10.1167/18.3.5

License: CC BY-NC-ND 4.0, https://creativecommons.org/licenses/by-nc-nd/4.0/ No modifications have been made to the original publication

# Abstract

In 3-D cluttered scenes such as foliage, deeper surfaces often are more shadowed and hence darker, and so depth and luminance often have negative covariance. We examined whether the sign of depth-luminance covariance plays a role in depth perception in 3-D clutter. We compared scenes rendered with negative and positive depth-luminance covariance where positive covariance means that deeper surfaces are brighter and negative covariance means deeper surfaces are darker. For each scene, the sign of the depth-luminance covariance was given by occlusion cues. We tested whether subjects could use this sign information to judge the depth order of two target surfaces embedded in 3-D clutter. The clutter consisted of distractor surfaces that were randomly distributed in a 3-D volume. We tested three independent variables: the sign of the depth-luminance covariance, the colors of the targets and distractors, and the background luminance. An ANOVA showed two main effects: subjects performed better when the deeper surfaces were darker and when the color of the target surfaces was the same as the color of the distractors. There was also a strong interaction: subjects performed better under a negative depth-luminance

covariance condition when targets and distractors had different colors than when they had the same color. Our results are consistent with a 'dark means deep' rule, but the use of this rule depends on the similarity between the color of the targets and color of the 3-D clutter.

**Keywords**: Depth Perception, Depth-Luminance Covariance, Occlusions, Clutter, Visibility.

## 3.1 Introduction

3-D cluttered scenes consist of many small surfaces that are distributed over a volume. Examples include the foliage of bushes, shrubs, trees, as well as tall grass. Depth perception in 3-D cluttered scenes is challenging since there are many occlusions present and the surfaces that make up the clutter often are only partly visible. Despite the challenge of occlusions, the visual system often is able to judge depth in 3-D clutter. For example, motion parallax and binocular disparity cues can be used to judge the depth order of two surfaces within the clutter [Langer, Zheng, and Rezvankhah, 2016], as well as the depth range of the clutter [van Ee and Anderson, 2001]. Langer et al. [2016] also showed that the visual system relies on probabilistic occlusion cues for judging depth order in 3-D clutter. The idea is that the fraction of a surface that is visible within 3-D clutter tends to decrease as the depth of the surface increases, and so the relative visibility of two surfaces within the clutter is a cue to their relative depth.

The experiments of Langer et al. [2016] considered geometric cues only, however, namely binocular disparity, motion parallax, and occlusions *i.e.* visibility. The surfaces in the clutter had random gray levels and there was no relationship between the luminance of the surfaces and their depth within the clutter. The experiments we report here take a different approach. We address scenarios in which luminance depends directly on depth. Either luminance increases with depth, or luminance decreases with depth. This manipulation explores a long-standing question in depth perception: when do brighter surfaces appear closer or farther? This question has been addressed using several different scene configurations in the past. The experiments that we report here are the first to address this question for 3-D cluttered scenes.

Here we review previous studies that have examined depth-luminance covariance. We begin with Schwartz and Sperling [1983] who were the first to consider how depthluminance covariance is combined with another depth cue, namely perspective. They used rotating Necker cubes consisting of lines rendered on a dark background. The luminance obeyed either a positive or negative depth-luminance covariance. When subjects judged the direction of rotation of the cube, they used a bright-means-near rule and ignored the perspective cues. In particular, when the brighter lines were deeper, subjects incorrectly perceived a deforming cube with the brighter lines in front rather than a rigid cube with brighter lines in the back. Dosher, Sperling, and Wurst [1986] went further by combining the depth-luminance covariance cue with binocular disparity, again using rotating Necker cubes under perspective. They modelled how subjects combined the cues both for static and dynamic displays. They found that darker lines tended to be perceived as deeper but that this depth-luminance covariance cue was weighted lower than the binocular disparity cue.

What might be the 'ecological optics' basis for a dark-means-deep rule for the Necker cube stimuli? Dosher et al. [1986] argued that if 3-D white scene lines were wires (cylinders) which projected to 2-D image lines in a black background, then the width of the image lines would decrease with distance and so the pixel intensity of the lines should decrease (from white to grey). The sign of this depth-luminance covariance would reverse for the case of black lines on a white background, namely the black lines would be brighter in the image as their scene depth increased, since a more distant line would project to smaller pixel sub-areas and the white background would fill the remaining pixel sub-areas. Schwartz and Sperling [1983] also noted informally that for Necker cubes rendered

as black lines on white background, subjects indeed used a bright-means-deep rule rather than dark-means-deep rule, and suggested that the sign of the depth-luminance covariance is determined more generally by the contrast of the line with the background rather than by luminance of the line per se.

Similar observations about contrast and perceived depth have been made for square patches viewed against a black or white background. When a light gray and a dark gray square are both viewed against a black background, the light gray square appears nearer. However, when the same squares are viewed against a white background, the dark gray square appears nearer [Farnè, 1977, Egusa, 1982]. It has been argued that these contrast effects are consistent with atmospheric scattering i.e. aerial perspective, namely a distant object will have lower contrast with the background since the luminance of both the object and background contain a common atmospheric component that increases with depth [O'Shea, Blackburn, and Ono, 1994].

There are other ecological optics theories of why a dark-means-deep or a brightmeans-deep rule might apply in a given situation. Under diffuse lighting such as on a cloudy day, surface concavities tend to be darker because the fraction of the diffuse source that is visible tends to be lower in concavities [Langer and Zucker, 1994]. This darkmeans-deep effect is modulated by factors, however, such as local variations in the surface normal, interreflections, glossiness, and non-uniformity of the diffuse source. Langer and Bülthoff [2000] showed that, for smooth surfaces rendered under uniform diffuse illumination, the visual system takes account of surface normal variations to some extent when comparing depths of neighbouring surface points and thus does not simply discriminate depths based on luminance. The question of when the visual system uses dark-meansdeep rule for interpreting shape from shading remains controversial [Tyler, 1998, Chen and Tyler, 2015, Todd, Egan, and Kallie, 2015].

Kim, Wilcox, and Murray [2016] investigated another aspect of the depth-luminance covariance by examining situations in which surfaces are perceived to emit or transmit light. They presented pairs of smooth terrain surfaces, one of which was rendered under diffuse lighting ('dark valley') and the other was depth inverted and assigned the same luminance at each corresponding surface position ('bright valley'). The only difference between the two cases was the depth information given by either stereo or rotational motion cues. Subjects were asked to judge which surface appeared to glow as if it had a light source inside or behind it. Subjects consistently chose the depth reversed surfaces ('bright valley') as glowing. Thus, subjects used the geometric cues from binocular disparity and dynamic occlusion to distinguish the hills and valleys, and subjects interpreted the luminance variations consistently with the shape defined by the geometric cues. The idea is similar to an observation by Langer [1999] that, in a photographic negative of a diffusely illuminated scene, concave regions tend to be brighter. So if the geometric cues that are given by occlusions constrain the 3-D geometry then the visual system perceives concave regions as glowing as if there were a light source and interreflections present. The key insight of Kim et al. [2016] was that this glow effect occurs even without occlusion cues being present, in particular, in their static binocular condition.

A negative depth-luminance covariance (dark-means-deep) also has been demonstrated using images of outdoor natural scenes using joint luminance and depth statistics [Potetz and Lee, 2003, Samonds, Potetz, and Lee, 2012]. These natural image statistics findings were consistent with neurophysiological findings that the disparity sensitivity of many neurons in monkey primary visual cortex covary with the neuron's local luminance contrast sensitivity. That is, cells with near or far disparities tend to be sensitive to local luminance maxima and minima respectively [Samonds et al., 2012]. These effects have been re-analysed and confirmed by Cooper and Norcia [2014] who also reported on a psy-chophysical study. Using natural images along with registered depth maps, they showed the perception of 3-D depth could be enhanced or diminished by biasing the image intensities towards a negative or positive depth-luminance covariance, respectively. For example, they found that for scenes that had a positive depth-luminance covariance, reducing the luminance of the far surfaces enhanced the perception of depth more than reducing the luminance of the near surfaces. This effect cannot be attributed simply to contrast enhancement, since darkening the background of a scene that has a positive depth-luminance covariance decreases the overall image contrast, yet such a manipulation was found to increase the perception of depth.

Finally, we consider some related work which addressed relationships between perceived depth, luminance, and color. Troscianko, Montagnon, Le Clerc, Malbert, and Chanteau [1991] showed that perceived surface slant of a tiled plane can increase when distant tiles are less saturated and both hue and luminance are held constant. They also showed that slant was not enhanced when luminance and saturation were held constant and only hue varied. The ecological basis for the saturation gradient effect is similar to what we discussed above, namely the atmospheric effect of aerial perspective. Here, colors from distant surfaces is less saturated because the color is mixed with the atmosphere which in their case had neutral hue. A different interaction between depth and color saturation arises from interreflections between surfaces. For example, in the case of diffuse lighting, surfaces that are deeper in the clutter tend to be more shadowed, and so a greater percentage of the illumination for deeper surfaces comes from light reflected off other surfaces. Such interreflection effects have been observed in more general scenes as well and can lead to hue and/or saturation gradients as a function of depth [Langer, 1999, 2001, Ruppertsberg, Bloj, and Hurlbert, 2008]. Moreover, there is some evidence that when depth information is given by stereo cues, the visual system can disentangle the interreflection effects and perceive the surface color [Bloj, Kersten, and Hurlbert, 1999]. This finding is similar to Kim et al. [2016]'s finding about stereo and glow discussed above.

The experiments that we report in this paper explored the effect of a depth-luminance covariance in 3-D cluttered scenes. The geometry of the scenes was similar to that of Langer et al. [2016] and the experimental procedure was also similar. In each trial, subjects were shown two target surfaces that were embedded in a cluttered 3-D volume and the task was to decide which of two target surfaces was closer. Our experiments varied the sign of the depth-luminance covariance, the background color of the scene (white or black), and the color saturations of the targets and the clutter surfaces. The sign of the depth-luminance covariance was given from the ordinal occlusion relationships. Our main goal was to examine when subjects used the depth-luminance covariance to perform the depth discrimination task.

The task of judging the depth order of the two targets was challenging for a few reasons. First, the targets did not overlap in the image and so there is no direct occlusion cue to determine their depth order. Second, since the targets were partly occluded by the random clutter, the target visibility was random. Moreover, we manipulated the clutter distribution to remove any systematic relation between the relative visibility of the targets and the depth order of the targets. The only information available for doing the task was the luminance differences between the targets. There is a limit in how accurately subjects can discriminate target luminances, however, especially in the presence of clutter since the clutter can produce local simultaneous contrast effects.

The third reason that the task is potentially difficult is that the correct answer on any trial depends on the sign of the depth-luminance covariance in the scene. This sign is readily available from occluders in the clutter. In particular, it is available from the luminance difference between the targets and the distractors that occlude the targets. However, it is unclear when subjects would take this depth-luminance covariance sign into account. In particular, in cases of a positive depth-luminance covariance, the positive sign conflicts with a default dark-mean-deep prior which has been shown in other studies to play a role in depth perception. A key goal of our experiments is to better understand when subjects use the sign of depth-luminance covariance.

# 3.2 Method

## 3.2.1 Apparatus

Images were rendered using OpenGL (Khronos Group, Beaverton, OR) and were displayed using a Dell Precision M6700 laptop (Dell, Round Rock, TX) with an NVIDIA Quadro K4000 graphics card (NVidia, Santa Clara, CA). The laptop's 17-in. LCD monitor was set to maximum brightness. A PR650 spectroradiometer (Photo Research, Syracuse, NY) was used to compute a luminance-to-RGB lookup table for each of the RGB channels. (For more details on color calibration and choice of colors, see Appendix 3.A.).

#### 3.2.2 Stimuli

Each scene consisted of two targets and of 1000 distractors. The two targets were  $4 \times 4$  cm squares that were separated in depth by a variable amount  $\Delta Z$ , and were separated horizontally by  $\Delta X = 10$  cm. Their XY positions were perturbed in a random XY direction by up to 1 cm. The targets were oriented so that their normal was parallel to the *Z* axis. Each distractor was a  $1 \times 1$  cm square, randomly oriented in 3-D and randomly positioned within a bounding box of size 20 cm  $\times 20$  cm.

The scene was rendered using perspective projection. The virtual subject's position was  $Z_0 = 60$  cm from the front and center of the clutter bounding box. The projection plane (display screen) was located at a depth of 70 cm from the subject, which corresponded to the center depth of the clutter.

Because the scenes were rendered under perspective projection, there was a size cue for depth. We removed this size cue for the targets only, similarly to Langer et al. [2016], by rescaling the height and width of each target to be proportional to its inverse depth. This ensured that the visual solid angle of a target at any depth (and in the absence of occlusions) would be equal to the visual angle of a target at the middle depth in the clutter. To help hide the fact that the two targets had the same visual angle in each trial, we jittered each target's aspect ratio so that the target was slightly rectangular, and we rotated each target by a random amount about the Z axis.

Subjects were seated so that their eye position was 70 cm directly in front of the screen corresponding to the virtual viewing position used in the rendering. We did not use a chin rest, so some variability in subject position was allowed. The task was performed monocularly with the non-dominant eye covered with an eye-patch. The height of each

subject's seat was changed so that the subject's eye would be roughly the same height as the center of the screen.

The luminance of the targets and distractors varied according to their depth Z as follows. For negative depth-luminance covariance (DLC-), the luminance was chosen to be proportional to  $exp(-\lambda(Z - Z_0))$ , where Z was the depth (cm) of the center of the distractor or target and  $Z_0$  was the depth where the clutter begins. We chose a decay factor of  $\lambda = .125$ , based on the density and size of the distractors [Langer and Mannan, 2012]. Thus, the intensities of the surfaces ranged from 1.0 down to  $exp(-.125 \times 20) \approx 0.082$ . For positive depth-luminance covariance (DLC+), luminance was proportional to  $exp(-\lambda(Z_0 + 20 - Z))$ . That is, it decayed exponentially with the Z distance to the back face of the volume. See Appendix 3.B for more details on how the DLC function is chosen.

In natural 3-D cluttered scenes such as foliage, it is common for targets embedded in the clutter to have a different color than the clutter itself. An extreme example is a red apple among green leaves. For this reason, we examined depth-luminance covariance not just for gray level surfaces but also for a few simple combinations of surface color. Three different colors were used for the targets and distractors: gray, unsaturated green, and high saturated green. The RGB values for the surfaces were scaled versions of the following: (1,1,1) for gray, (0.33,0.87,0.33) for unsaturated green, and (0,1,0) for saturated green. For any depth and for any DLC- or DLC+ condition, we scaled the RGB values so that the luminance (CIE Y) would depend on the depth, and be the same for any of the three colors (see Appendix 3.A for more details).

The visibility of a target is a strong cue for target depth [Langer et al., 2016]. Since the goal of our experiment was mainly to investigate luminance effects, we removed the depth information from the target visibilities. We did so by modifying the distribution of the distractors in the clutter, namely we removed distractors whose center point occluded the far target and so the tunnel's cross sectional area was the same as the far target's area. We only removed distractors whose depths were between that of the near and far targets. See Figure 3–1.



Figure 3–1: Illustration of tunnel for the DLC- condition. The viewing position is from below. (a) The tunnel removes distractors that are between the depths of the near and far target and that could occlude the far target. We used the tunnel only in front of the far target because that was sufficient to equate the expected values of visibility of the two targets for any depth difference. This illustration does not show the random variations in the distractors positions and orientations. (b) In Experiment 2, we matched the distractors that fell in front of the targets (see highlighted black square). See Appendix 3.C for details.

We did not use a tunnel for depth differences  $\Delta Z$  that exceeded 15 cm. The reason is that when the scenes are rendered with tunnels for high  $\Delta Z$  values, these tunnels become very apparent. In pilot studies we found that subjects became confused in these conditions which led to guessing behavior. By not using a tunnel when  $\Delta Z > 15$  cm, we gave these scenes a strong visibility cue which allowed subjects to perform the task easily and ensured that the staircases were well behaved.

Using a tunnel when  $\Delta Z \leq 15$  cm made the *expected* visibilities the same for the two targets for each depth difference. However, typically there were still differences in the visibilities of the two targets in each trial since the distractors were randomly positioned and oriented. If subjects were to mistakenly use these random differences in visibility in each trial as a depth cue, then this would compromise their performance and raise depth discrimination thresholds. To examine if subjects were indeed doing so, we carried out a second version of the experiment (Experiment 2) in which we manipulated the distractor distribution further to reduce the per trial variability in the visibility difference. See the highlighted area in Figure 3–1 (b) which indicated the distractors that we manipulated and see Appendix 3.C for the details of the manipulation.

We also ran a third version of the experiment in which subjects were asked to discriminate target luminance rather than depth. That is, we asked them which of the targets was brighter. This was done for two reasons. The first reason was to reassure us that judging relative luminance was not itself the limiting factor in our experiments. The second reason was to reassure us that participants were in fact discriminating depths, rather than just luminances in the main experiments.

# 3.2.3 Design

For each experiment, we used 36 different conditions which were defined by the following combinations  $(36 = 2 \times 2 \times 3 \times 3)$ :

• sign of depth-luminance covariance (DLC+, DLC-)

- background color (black, white)
- target color (gray, low or high saturation green)
- distractor color (gray, low or high saturation green)

Figure 3–2 shows examples of DLC- conditions. The rows and columns show the different distractor and target colors, respectively. Figure 3–3 shows examples of DLC+. For both figures, we chose backgrounds that are consistent with the depth-luminance covariance, namely black background for DLC- and white background for DLC+.

# Targets



Figure 3–2: Stimuli for dark-means-deep (DLC-) conditions with a black background. Rows represent distractor color and columns represent target color. On the diagonal, targets and distractors have the same color.

# Targets



Figure 3–3: Stimuli for dark-means-near (DLC+) with a white background. On the diagonal, targets and distractors have the same color.

#### 3.2.4 Procedure

In each trial, subjects were shown a scene consisting of distractors and two target rectangles. The task was to indicate which target was closer (to the subject) or which target was brighter, depending on the experiment. Subjects responded by pressing either the left or right arrow key on the keyboard.

For each experimental condition and for each subject, we estimated a depth discrimination threshold using a 1-up 1-down staircase. The values of  $\Delta Z$  in the staircase were chosen to target a proportion correct of approximately 78% [Garcıa-Pérez, 1998]. When the subject answered correctly or incorrectly, the distance  $\Delta Z$  between targets was reduced by a factor 0.8 or was increased by a factor 2.19, respectively. Staircase conditions were randomly interleaved. Each staircase began at level  $\Delta Z = 5$  cm and then terminated after ten reversals. To compute the thresholds, we averaged the log of the  $\Delta Z$  values of the last eight reversals.

Response time was limited to four seconds. If the subject did not respond in some trial, then a random choice was made and a red X mark would show on the screen. Additionally, there was a rest period after every 100 trials for as long as the subject wanted. The experiment typically lasted around forty-five minutes.

#### 3.2.5 Participants

For Experiment 1, twenty subjects participated with ages ranging from 19 to 44. For Experiment 2, eight new subjects participated with ages ranging from 19 to 60. For Experiment 3 (luminance discrimination), eight subjects participated with ages ranging from 28 to 60. Each subject was paid \$10. Subjects had little or no experience with psychophysics experiments. Each had normal or corrected-to-normal vision. We required subjects to pass the Ishihara's Test Chart for Color Deficiency. Subjects were unaware of the purpose of the experiments. Informed consent was obtained using the guidelines of the McGill Research Ethics Board which is consistent with the Declaration of Helsinki.

### 3.3 Results

Plots of the means and standard errors of the thresholds for all three experiments are shown in Figure 3–4. For each, we ran three-way repeated measures ANOVA's to test the effects of depth-luminance covariance (DLC-, DLC+), background color (black, white), and equality of target and distractor hue (diagonals versus off-diagonals in the figures). We report exact p values. A p value smaller than 0.05 was considered to be significant. These plots pooled the background (black versus white) conditions since this variable did not have a statistically significant effect (see below).

#### 3.3.1 Experiment 1

A dark-means-deep cue would predict thresholds to be lower for DLC- than DLC+ and this is indeed what we found with means 4.3 cm and 5.8 cm respectively. This difference was significant ( $F_{1,19} = 9.6, p = 0.006$ ). A second main effect was that thresholds were lower when targets and distractors had the same surface color (diagonals) than when they had different colors (off-diagonals) ( $F_{1,19} = 18.1, p = 0.0005$ ) with means 4.5 cm and 5.5 cm respectively. There was an interaction effect between DLC and color ( $F_{1,19} = 31.1, p = 0.00002$ ). When targets and distractors had the same color, thresholds were just slightly lower for DLC- than for DLC+ with means 4.4 and 4.7 cm respectively. When targets and distractors differed in color, thresholds were much lower for DLC- than DLC+ with means 4.2 cm and 6.8 cm, respectively. This suggests that subjects relied on a dark-means-deep prior more when the colors of targets differed from the colors of the distractors. This prior is correct for the DLC- stimuli but is incorrect for the DLC+ stimuli. We will return to this point in the Discussion section.

We had expected performance to be better when the background was consistent with the depth-luminance covariance, namely black background in a dark-means-deep condition and a white background in a dark-means-near condition, rather than vice-versa. However, the mean thresholds for consistent and inconsistent backgrounds were nearly the same, namely 5.0 cm and 5.1 cm ( $F_{1,19} = 0.01, p = 0.64$ ). We therefore pooled the results for the plots in Figure 3–4 for consistent and inconsistent backgrounds, showing 18 conditions instead of 36.

# 3.3.2 Experiment 2

The purpose of Experiment 2 was to reduce the per trial visibility difference between the two targets, in case subjects were mistakenly relying on the visibility difference to perform the task. (See Appendix 3.C.) Thresholds indeed were lower overall in Experiment 2 than in Experiment 1, with means 4.56 cm and 5.25 cm, respectively. A one tailed t-test on the signed differences in the mean thresholds for the 36 conditions of Experiments 1 versus 2 revealed a significant difference (t = 4.67, p < 0.0001).

Otherwise, the general trends were similar to Experiment 1. Some results did not reach significance, but this is unsurprising since we used only eight subjects for Experiment 2 compared to twenty for Experiment 1. Thresholds were lower for DLC- than DLC+ with means 3.3 cm and 5.6 cm respectively ( $F_{1,7} = 8.1, p = 0.02$ ). Thresholds were lower for diagonals than for off-diagonals with means 3.9 and 5.0 cm respectively ( $F_{1,7} = 2.76, p = 0.141$ ). The interaction between DLC and diagonal/off-diagonals also was close to significant ( $F_{1,7} = 0.35, p = 0.09$ ). Thresholds were slightly smaller for DLC-

than DLC+ for the diagonals with means 3.2 and 4.6 cm respectively. Thresholds were much lower for DLC- than DLC+ in the off-diagonal case, with means 3.3 cm and 6.6 cm respectively. In general, the DLC- thresholds were similar for all conditions, but the DLC+ thresholds were greater for the off diagonals. Finally, again there was no significant differences between consistent and inconsistent backgrounds ( $F_{1,7} = 0.6, p = 0.48$ ). The means for a consistent background versus a inconsistent one were 4.3 cm versus 4.6 cm.

#### 3.3.3 Experiment 3

We ran a control experiment in which subjects were asked to discriminate target luminance instead of depth. Thresholds were lower in Experiment 3 than in Experiment 1, with means 1.59 cm and 5.25 cm, respectively. A one tailed t-test on the signed differences in the mean thresholds for the 36 conditions of Experiments 1 versus 3 revealed a significant difference ( $t = 15.26, p < 10^{-16}$ ). This reassured us that judging relative luminance was not itself the limiting factor in our experiments. This also reassured us that participants were discriminating depths, rather than just luminances in the main experiments.

We did not run the luminance discrimination task again on the Experiment 2 stimuli, but we would expect performance there to be as good or better than what we found using the Experiment 1 stimuli, for the same reason that performance was better in Experiment 2 than Experiment 1.

No significant effects were found between conditions. DLC- and DLC+ had means 1.61 cm and 1.56 cm respectively ( $F_{1,7} = 0.58, p = 0.47$ ). Diagonal versus off-diagonal had means 1.64 cm and 1.53 cm respectively ( $F_{1,7} = 2.46, p = 0.16$ ). Consistent and inconsistent background colors had means 1.58 cm and 1.59 cm ( $F_{1,7} = 0.02, p = 0.89$ ).



Targets

Figure 3–4: Mean and standard errors for Exp. 1 and 2 (depth) and Exp. 3 (luminance). The stimuli of Exp. 3 were the same as Exp. 1. The  $3 \times 3$  layout of the plots corresponds to the conditions shown in Fig. 3–2 and 3–3.
#### 3.4 Discussion

A key finding in our experiments is that subjects behaved as if they used the sign of the depth-luminance covariance to perform the task. One natural strategy for using the sign information is as follows. Use the luminance and depth ordering of the distractors which is given by occlusions to determine the sign of the depth-luminance covariance. With this sign information, compare the two target luminances and choose the closer target according to the sign of the depth-luminance covariance of the distractors. When the target and distractor colors differ, rely less on the DLC sign information and instead rely more on a default dark-means-deep rule or 'prior'.

While the above strategy seems plausible, there is a second strategy that leads to essentially the same results. This second strategy is based on the contrast between targets and the distractors that occlude them. Recall that Schwartz and Sperling [1983] and others suggested that the perceived depth-luminance covariance is determined by the contrast of an object with the background rather than luminance per se. For our scenes, the direct relationship between depth and luminance also provides a contrast cue. However in our case, the cue is not the contrast between the luminance of the targets with respect to the background. Rather, it is the contrast between the targets and the occluders, namely the near target has a lower contrast with respect to its occluders than the far target has with respect to its occluders. To put it more simply, the near targets have a more similar luminance to their occluders than the far targets do to their occluders. This contrast cue is valid both in the DLC+ and DLC- conditions. If observers were to base their judgements of target depth on this contrast cue, then they would behave similarly in the task if they used the first strategy described above. Our experiment does not allow us to decide which of these two strategies better explains their behaviour. Further experiments would be needed to tease apart these two strategies.

Another key finding is that when the colors of target and distractors differ, subjects behaved as if they relied more on the dark-means-deep prior. This was somewhat surprising since the targets informally seem easier to segment from the distractors when the colors differ, and so one might expect that subjects could more easily compare them. However, Experiment 3 did not provide any evidence that subjects could discriminate the luminances of the targets better when the colors of distractors and targets differed. What seemed to happen in Experiments 1 and 2 when the targets and distractor colors differed is that the targets were seen as different types of objects than the distractors and so the target luminances were perceived as less related to the distractor DLC sign. Since there were no depth cues for the targets other than the DLC cue from the distractors, subjects may have just relied more on their prior for dark-means-deep.

Another finding worth discussing is that the background seemed to play little role. The background color is visible outside the X and Y limits of the clutter, but it is sometimes also visible within the clutter when one can see all the way through. In this case, it is difficult to perceive that the leaking background is indeed due to background as opposed to just another occluded surface. In the case that the background color is inconsistent with the depth-luminance covariance, a sliver of background that is visible within the clutter might just be perceived as an outlier. Therefore, we believe that it was the luminances of the clutter rather than the background that provided the main cue that subjects used. Perhaps in sparser scenes containing fewer occlusions, the background color would have more of an effect. One might ask if there were any other cues present in our stimuli that allowed subjects to perform the task, for example, perspective cues (size cues) from the distractors. To test if subjects were using information other than the sign information from DLC, we ran Experiment 1 again using three subjects (one author and two new naive subjects) and using equiluminant targets. The mean thresholds over all conditions were 11.5 cm which is well above the thresholds for the actual Experiment 1 but which is below the 15 cm limit of  $\Delta Z$ , where the visibility cue is present for our stimuli. Could it be that subjects achieved thresholds below 15 cm by using other cues? We believe the answer is no. We ran a model observer who guesses in conditions when the visibility cue has been removed ( $\Delta Z \leq 15$ ) and who answers correctly when the visibility cue is present ( $\Delta Z > 15$ ). We ran 300 staircases for such an observer. The mean threshold was approximately 12 cm which is similar to our three observers for the equiluminant targets. These results suggest that subjects relied entirely on the DLC cues, and the prior for DLC- to perform the task in our two main experiments.

#### 3.5 Conclusion

Our experiments contribute new ideas for investigating and understanding depth perception from luminance variations in complex scenes. In particular, we have shown that occlusions determine the sign of depth-luminance covariance in 3-D cluttered scenes and that subjects can make use of this sign information to discriminate the depths of targets embedded in the clutter. We showed that this sign information was combined with a prior for 'dark means deep' (negative depth-luminance covariance). Interestingly, subjects performed better when targets and distractors had the same color saturation, namely they relied less on a 'dark means deep' cue in that case. Further studies are needed to determine if there are other interactions of color, luminance, and perceived depth in 3-D cluttered scenes, for example, if there are interactions with other depth cues such as binocular disparity or motion parallax.

### Appendix

### **3.A Color Calibration**

When selecting the colors for the surfaces, we wanted to span as large a range of luminances as possible. Since the green channel has the largest luminance we used gray (1,1,1) and saturated green (0,1,0) to define two of our surface colors. As a third color, we used an unsaturated green which we defined by transforming the saturated green (0, 1, 0) from RGB to CIELUV, reducing the saturation correlate value by 50%, and then applying the inverse transformation back to RGB. This gave the RGB value (0.33, 0.87, 0.33). To transform from RGB to LUV, we first transformed from RGB to XYZ using the M matrix below, and then from XYZ to LUV.

To ensure that luminance varied with depth in the same manner for each of the three colors, we needed to know which RGB values of gray and unsaturated green had the same luminance value (*Y*) as the saturated green. We first determined the maximum luminances of gray  $Y_{Gray}$ , unsaturated green  $Y_{UGreen}$  and saturated green  $Y_{SGreen}$  using the transformation between RGB and CIE XYZ below. We rescaled the maximum RGB of gray (1,1,1) and unsaturated green (0.33,0.87,0.33) by multiplying by  $Y_{SGreen}/Y_{Gray}$  and  $Y_{SGreen}/Y_{UGreen}$ , respectively. This ensured we had the same maximum luminance for each color. Then, during the experiment, we scaled the RGB values of each surface color according to its depth depending on the condition.

The *Y* values above can be obtained via the transformation between RGB and CIE XYZ space as follows,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{M} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

We obtained the columns of **M** by using the spectroradiometer to measure the XYZ values for uniform RGB patches of  $(1,0,0)^T$ , green  $(0,1,0)^T$ , blue  $(0,0,1)^T$ . This gave

$$\mathbf{M} = \begin{bmatrix} 44.5 & 43.9 & 20.1 \\ 23.5 & 86.2 & 7.16 \\ 1.08 & 8.14 & 107 \end{bmatrix},$$

where the middle row is the luminance (*Y*) of each of the channels in  $cd/m^2$ . To obtain the luminances  $Y_{Gray}$ ,  $Y_{UGreen}$ , and  $Y_{SGreen}$ , we multiplied their corresponding  $(R, G, B)^T$ by the middle row of **M**.

### **3.B** Model of Depth-Luminance Covariance

Our DLC- model combines the ideas of Langer and Zucker [1994]'s cloudy day rendering model and Langer and Mannan [2012]'s probabilistic model of surface visibilities in 3-D clutter. The former model assumes the scene is illuminated by a uniform hemispheric sky, centered in the Z axis direction. The latter model assumes a 3-D cluttered scene which begins at depth  $Z = Z_0$  and has infinite extent in the X and Y directions.

Under the cloudy day rendering model, illumination is assumed to be diffuse and non-directional. The luminance variations consider cast shadows only, but ignore interreflections. Thus, the luminance of a surface point  $\mathbf{X}_{\mathbf{p}} = (X_p, Y_p, Z_p)$  depends on the amount of hemispheric sky that is visible from it. The directions of the hemispheric sky are parametrized by polar angle  $\theta$  and azimuth  $\phi$ , and  $V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)$  is the visibility function which equals 1 if the hemispheric sky is visible from point  $\mathbf{X}_{\mathbf{p}}$  in direction  $(\theta, \phi)$ , and 0 otherwise. This yields the following expression for luminance at a surface point  $\mathbf{X}_{\mathbf{p}}$ ,

$$L(\mathbf{X}_{\mathbf{p}}) = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} V(\mathbf{X}_{\mathbf{p}}, \theta, \phi) \cos \theta \sin \theta d\theta d\phi.$$

The double integral above integrates over concentric circles of radius  $\sin \theta$  to obtain the surface area of the unit hemisphere. This area is weighted by Lambert's cosine law, which states that luminance from a Lambertian surface is proportional to the cosine of the angle  $\theta$  between the direction of the incident light ray and the surface normal. The total surface area of the cosine-weighted hemisphere is  $\pi$ . We integrate only over the visible portions of the hemisphere and so dividing the result by  $\pi$  returns the luminance, a value between 0 and 1.

For a 3-D cluttered scene, we consider visibility  $V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)$  as a random variable. We compute the probability that the hemispheric sky is visible along a ray,  $p(V(\mathbf{X}_{\mathbf{p}}, \theta, \phi))$ , as done in Langer and Mannan [2012]. To get a closed form model of  $p(V(\mathbf{X}_{\mathbf{p}}, \theta, \phi))$ , we assume the elements of the clutter are disks of area *A*, and we assume that the spatial distribution of the disks is a Poisson process with density  $\eta$  which is the average number of disk centers per unit volume. The parameters *A* and  $\eta$  can be lumped together as a single constant  $\lambda = \eta A$ .

The function  $p(V(\mathbf{X}_{\mathbf{p}}, \theta, \phi))$  can be written in terms of  $Z_p$  only as follows. The length of any ray from  $\mathbf{X}_{\mathbf{p}} = (X_p, Y_p, Z_p)$  to the edge of the clutter is  $(Z_p - Z_0)/\cos\theta$ , where the term  $1/(\cos\theta)$  accounts for the greater path length of a ray with greater polar angle  $\theta$ . To simplify the integral, we assume the elements of the clutter have surface normal facing the *Z* direction and so these clutter elements are foreshortened in the ray direction  $(\phi, \theta)$ . Then, the Poisson model yields:

$$p(V(Z_p, \theta, \phi)) = exp\{-\lambda \cos \theta (Z_p - Z_0) / \cos \theta\}.$$

The  $\cos \theta$  terms cancel out and so  $p(V(Z_p))$  depends only on depth,

$$p(V(Z_p)) = exp\{-\lambda(Z_p - Z_0)\}.$$

Then, the expected visibility of the sky in direction  $(\theta, \phi)$  is computed as,

$$E[V(Z_p)] = 1 \cdot p(V(Z_p)) + 0 \cdot (1 - p(V(Z_p))) = exp\{-\lambda(Z_p - Z_0)\}.$$

Using the definition of luminance of a surface point  $L(\mathbf{X}_{\mathbf{p}})$ , we use the above to compute the expected luminance of a surface point  $E(L(\mathbf{X}_{\mathbf{p}}))$ . The linearity property of expectation allows us to integrate over the expected visibilities of rays along individual directions,

$$E[L(\mathbf{X}_{\mathbf{p}})] = E[\frac{1}{\pi} \int_{0}^{2\pi} \int_{0}^{\pi/2} V(\mathbf{X}_{\mathbf{p}}, \theta, \phi) \cos \theta \sin \theta d\theta d\phi],$$
  
$$= \frac{1}{\pi} \int_{0}^{2\pi} \int_{0}^{\pi/2} E[V(\mathbf{X}_{\mathbf{p}}, \theta, \phi)] \cos \theta \sin \theta d\theta d\phi,$$
  
$$= \frac{1}{\pi} \int_{0}^{2\pi} \int_{0}^{\pi/2} E[V(Z_{p})] \cos \theta \sin \theta d\theta d\phi.$$

The resulting expected luminance of point  $X_p$ , written in terms of  $Z_p$  only, is given by,

$$E[L(Z_p)] = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} exp\{-\lambda(Z_p - Z_0)\}\cos\theta\sin\theta d\theta d\phi.$$

Integrating the last equation yields,

$$E[L(Z_p)] = exp\{-\lambda(Z_p - Z_0)\}.$$

This is the rendering model that we use for DLC-, namely luminance is chosen to be proportional to  $E[L(Z_p)]$ . The rendering model is based on several assumptions. First, the illumination arrives from a hemisphere centered at the Z axis, and this illumination hemisphere has uniform luminance over all directions. Second, the clutter has infinite XY extent, which is not the case for cluttered scenes in our stimuli. Third, the luminance variations consider cast shadows only, but ignore interreflections. Fourth, the elements of the clutter have normals parallel to the Z direction. Because the model is based on rather strong assumptions, we cannot and do not claim that this model is photorealistic for the given scenes. Rather the model is meant to capture a qualitative shadowing effect that does occur in real scenes, namely when clutter such as foliage is illuminated under approximately diffuse light.

#### **3.C** Manipulating Visibility Cue in Experiment 2

In Experiment 1, the expected visibility was the same between the two targets, but there was variation in the visibilities of the random distribution of the clutter. This random variation often led to one target being more visible than the other. Visibility is a cue for depth discrimination 3-D cluttered scenes [Langer et al., 2016] and so it is possible that subjects were using the difference in visibility between targets to perform the task, even though in our stimuli this visibility difference contained no information for doing the task since the two targets had the same expected visibility for each depth condition.

In Experiment 2, we therefore attempted to reduce the subject's use of the visibility cue by reducing the random visibility differences between targets. We did so in each trial by matching the number of distractors falling in front of both targets. Specifically, we matched the number of distractors falling in front of the far target to that falling in front of the near target. We also matched the X, Y, and Z positions of the distractors relative to the center of each target, and we matched the slant angles, namely the angles between the Z axis and the occluder's surface normal. Note that we only matched the distractors that occluded the targets. The remaining visibility differences for the two targets were due to random interactions between the direction (tilt) of each distractor's orientation, the variations in aspect ratios and angles of the targets, and parallax between the targets are different.

In Figure 3.C.1 (a), we compare the mean target visibilities as a function of depth for Experiments 1 and 2. Target depths were chosen from  $Z = \{0, 2, 4, ..., 20\}$  cm and target visibilities were defined as the fraction of the target that was visible in the image (which we measured by counting pixels). In the plots, depths up to 8 cm show the visibilities for the near targets and depths beyond 12 cm show the visibilities for far targets. At depth 10 cm, there is no distinction between near and far targets. We used 1000 scenes for each data point in the plot.

The result in Figure 3.C.1 (a) shows that indeed the two targets had same expected visibility for each depth condition. The plots show an exponential fall in visibility up to depth Z = 10 cm [Langer et al., 2016] and then a rise in visibility beyond depth 10 cm. This rise is due to the tunnel in front of the far target, namely the size of the tunnel

increases as the far target depth increases (recall Fig. 3–1). Recall also that the tunnel is only used when the depth difference is less than 15 cm, and so the near and far targets are in the depth range  $10\pm7.5$  cm. This is why the visibility plot for the far target is so much lower for Z = 18 cm and 20 cm.

Figure 3.C.1 (b) shows the mean absolute difference in visibility between the near and far targets as a function of their depth difference  $\Delta Z$ . Recall that the depths in Figure 3.C.1 (a) were chosen from  $Z = \{0, 2, 4, ..., 20\}$  cm. Similarly, the depth differences in Figure 3.C.1 (b) were chosen from  $\Delta Z = \{0, 4, 8, ..., 20\}$  cm. For  $\Delta Z \leq 15$  cm, the mean absolute difference in visibility for Experiment 2 was approximately 30% less than Experiment 1. As expected, this reduction in visibility differences led to an improvement in performance from Experiment 1 to Experiment 2.

## Acknowledgements

This research was supported by a Doctoral Grant awarded to Milena Scaccia from the Fonds de Recherche du Quebec - Nature et Technologies (FRQNT), and an FRQNT Team Research grant and NSERC Discovery Grant to Michael Langer.



Figure 3.C.1: (a) The visibility plot compares the mean target visibilities as a function of depth for Experiments 1 and 2. 1000 scenes were used for each data point. Target depths were chosen from  $\{0, 2, 4, ..., 20\}$  and target visibilities were defined as the fraction of the target that was visible in the image (which we measured by counting pixels). The error bars show standard deviations (and not standard error). (b) The plot compares the mean absolute difference of target visibilities for Experiments 1 and 2 using 1000 scenes for each data point. Target depth differences  $Z_{far} - Z_{near} = \Delta Z$  were chosen from  $\{0, 4, 8, ..., 20\}$ . The standard deviations are almost identical for Experiments 1 and 2, so to avoid overlap they are shown only for Experiment 1 (with error bars).

### REFERENCES

- Marina G Bloj, Daniel Kersten, and Anya C Hurlbert. Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, 402(6764): 877, 1999.
- Chien-Chung Chen and Christopher W Tyler. Shading beats binocular disparity in depth from luminance gradients: Evidence against a maximum likelihood principle for cue combination. *PLOS one*, 10(8):e0132658, 2015.
- Emily A Cooper and Anthony M Norcia. Perceived depth in natural images reflects encoding of low-level luminance statistics. *Journal of Neuroscience*, 34(35):11761–11768, 2014.
- Barbara Anne Dosher, George Sperling, and Stephen A Wurst. Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, 26(6):973–990, 1986.
- Hiroyuki Egusa. Effect of brightness on perceived distance as a figureground phenomenon. *Perception*, 11(6):671–676, 1982.
- Mario Farnè. Brightness as an indicator to distance: relative brightness per se or contrast with the background? *Perception*, 6(3):287–293, 1977.
- Miguel A Garcia-Pérez. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12):1861–1881, 1998.
- Minjung Kim, Laurie M Wilcox, and Richard F Murray. Perceived three-dimensional shape toggles perceived glow. *Current Biology*, 26(9):R350–R351, 2016.
- Michael S Langer. When shadows become interreflections. *International Journal of Computer Vision*, 34(2-3):193–204, 1999.

- Michael S Langer. A model of how interreflections can affect color appearance. *Color Research Application*, 26(S1: Proceedings of the International Colour Vision Society): S218–S221, 2001.
- Michael S Langer and Heinrich H Bülthoff. Depth discrimination from shading under diffuse lighting. *Perception*, 29(6):649–660, 2000.
- Michael S Langer and Fahim Mannan. Visibility in three-dimensional cluttered scenes. Journal of the Optical Society of America A, 29(9):1794–1807, 2012.
- Michael S Langer and Steven W Zucker. Shape-from-shading on a cloudy day. *Journal of the Optical Society of America A*, 11(2):467–478, 1994.
- Michael S Langer, Haomin Zheng, and Shayan Rezvankhah. Depth discrimination from occlusions in 3D clutter. *Journal of Vision*, 16(11):11–11, 2016.
- Robert P O'Shea, Shane G Blackburn, and Hiroshi Ono. Contrast as a depth cue. *Vision Research*, 34(12):1595–1604, 1994.
- Brian Potetz and Tai Sing Lee. Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *Journal of the Optical Society of America A*, 20(7):1292–1303, 2003.
- Alexa I Ruppertsberg, Marina Bloj, and Anya Hurlbert. Sensitivity to luminance and chromaticity gradients in a complex scene. *Journal of Vision*, 8(9):3–3, 2008.
- Jason M Samonds, Brian R Potetz, and Tai Sing Lee. Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. *Proceedings of the National Academy of Sciences*, 109(16):6313–6318, 2012.

- Barry J Schwartz and George Sperling. Luminance controls the perceived 3-D structure of dynamic 2-D displays. *Bulletin of the Psychonomic Society*, 21(6):456–458, 1983.
- James T Todd, Eric JL Egan, and Christopher S Kallie. The darker-is-deeper heuristic for the perception of 3D shape from shading: Is it perceptually or ecologically valid? *Journal of Vision*, 15(15):2–2, 2015.
- Tom Troscianko, Rachel Montagnon, Jacques Le Clerc, Emmanuelle Malbert, and Pierre-Louis Chanteau. The role of colour as a monocular depth cue. *Vision Research*, 31(11): 1923–1929, 1991.
- Christopher W Tyler. Diffuse illumination as a default assumption for shape-from-shading in the absence of shadows. *Journal of Imaging Science and Technology*, 42(4):319–325, 1998.
- Raymond van Ee and Barton L Anderson. Motion direction, speed and orientation in binocular matching. *Nature*, 410(6829):690, 2001.

# CHAPTER 4 Density discrimination with occlusions in 3-D clutter

# Preface

In the two previous chapters, we explored how occlusions affect the relationship between luminance and depth, and examined how signs of depth-luminance covariance affect human observers' relative depth perception of two targets in 3-D clutter. In this chapter, we examine how luminance and occlusions affect human observer bias and sensitivity to a spatial change in a cluttered volume, where we vary the density of surfaces between the front and back halves and between the left and right halves of the volume. Our scenes are presented with kinetic depth.

In designing our stimuli, one challenge was to find a luminance cue that helped observers segment the two halves of the volume. The exponential DLC function used in Chapter 3 was found to be unsuitable for discriminating between two halves of the volume. Pilot experiments revealed that observers had difficulty perceiving which surfaces belonged to the front half versus the back half; for example it was assumed that the middle brightness 0.5 was located at the middle of the volume, and also the back of the volume tended to be hidden because of the exponential decay in visibility and luminance. Therefore, we decided to use a DLC function where equal steps in depth have equal steps in brightness. We compare DLC to a black-white condition where squares are only black or only white in each half. Most importantly, we examine how different parameters of clutter, namely the area and density of the squares and the overall level of occlusion affect performance in the density discrimination tasks. We will see below how these variations result in interesting effects in observers' performance. In order to gain further insight on the strategies used by humans, we also compare to model observers that perform the tasks by comparing the image occupancy of the two halves.

# Density discrimination with occlusions in 3-D clutter

Milena Scaccia<sup>1</sup> and Michael S. Langer<sup>2</sup> School of Computer Science, McGill University, Montreal, Quebec, Canada <sup>1</sup>milena.scaccia@mail.mcgill.ca <sup>2</sup>langer@cim.mcgill.ca To appear in Journal of Vision, 2019

### Abstract

We examined how well human observers can discriminate the density of surfaces in two halves of a rotating 3-D cluttered sphere. The observer's task was to compare the density of the front versus back half or the left versus right half. We measured how the bias and sensitivity in judging the denser half depended on the level of occlusion and on the area and density of the surfaces in the clutter. When occlusion level was low, observers in the front-back task were biased to judge the back as denser, and when occlusion level was high they were biased to judge the front as denser. Weber fractions decreased as density increased for both the front-back and left-right tasks, consistent with previous findings for 2-D density discrimination. Weber fractions did not vary significantly with area for the front-back task, but increased with area for the left-right tasks. We also ran model observers that compared the image occupancies of the two halves against a known expected difference. As the occlusion level increased, this expected difference followed a similar trend as the biases of the human observers, with a roughly constant offset between them. Weber fractions for human and model observers followed some similar trends, but there were discrepancies as well which can be partly explained by the information available to human versus model observers in carrying out their respective tasks.

Keywords: Density perception, numerosity perception, occlusions, clutter, visibility

### 4.1 Introduction

Three-dimensional (3-D) clutter consists of many small surfaces that are distributed randomly in a volume. Examples include foliage and branches of a tree, tall grass, etc. Little is known about how well the human visual system can judge the spatial distribution of 3-D clutter. Most studies of 3-D clutter have investigated clutter that is composed of many points or thin lines. Such studies have typically concentrated on depth cues such as motion and binocular disparity, and have ignored other cues such as occlusions and luminance. Occlusions are an important cue for perceiving the spatial distribution of 3-D clutter since points that are deeper in the volume are less likely to be visible, and so occlusions provide probabilistic information about depth. For example, Langer, Zheng, and Rezvankhah [2016] examined 3-D cluttered scenes consisting of squares that were randomly distributed in a volume, and showed that observers could use occlusion cues to discriminate the depth of target surfaces embedded in the clutter. Scaccia and Langer [2018] used similar scenes and showed there was an interaction between occlusion cues and color + luminance cues. Both of these studies investigated how well observers can discriminate the depths of two target surfaces within the clutter. In this paper, we address a different question, namely how well can observers discriminate the density in the two halves of a 3-D cluttered scene? We use the term 'density' rather than 'number' for our experiments, mainly because this is the term we used to explain the task to the subjects. Otherwise, there is no meaningful difference between the terms for our stimuli, since the two quantities are directly related in our stimuli.

Early studies measured how well human observers can directly estimate the number of dots presented. Taves [1941] and Kaufman, Lord, Reese, and Volkmann [1949] found that when the dot number was below seven, observers could count the dots, but for greater dot numbers the accuracy in estimating the number of dots decreased. A similar effect was found in studies that measured observers' ability to detect target patterns of dots against a noisy random-dot background [French, 1954, Barlow, 1978]. As the number of background dots increased relative to the number of target dots, the observer's ability to detect the target dots decreased.

Subsequent studies considered the relationship between perceived 2-D density and numerosity. There is some controversy about whether the mechanisms for perceiving the two are the same or not. Burr and Ross [2008] found that numerosity is a primal visual property because it is sensitive to adaptation and that observers estimate numerosity independently of density. This finding was challenged by Durgin [2008] who had shown that adaptation is determined by texture density and not numerosity [Durgin, 1995]. Other studies claimed that numerosity and density are not independent. In a task comparing two separate 2-D patches, Dakin, Tibber, Greenwood, Kingdom, and Morgan [2011] and Tibber, Greenwood, and Dakin [2012] found similar just noticeable differences (JNDs) in density and numerosity discrimination. They also showed perceived density and number were biased by increases in area, such that larger areas were perceived as denser and more numerous. Bell, Manson, Edwards, and Meso [2015] also found that perceived density was biased by increases in area, but did not find that perceived numerosity was biased by increases in area. They also showed that neither perceived density nor numerosity were biased by increases in volume. Several groups measured Weber fractions for numerosity discrimination. Burr and Ross [2008] and Ross and Burr [2010] showed that numerosity obeys Weber's law for low densities. For high densities, Anobile, Cicchini, and Burr [2014], Anobile, Turi, Cicchini, and Burr [2015] and Cicchini, Anobile, and Burr [2016] found that the thresholds increased only with the square root of density, corresponding to a decreasing Weber fraction. From this, they inferred that there exist different mechanisms for perceiving numerosity and density at low and high densities.

Other studies have examined density and number perception in depth layers. The results of these studies are pertinent to our current work as they explore density perception of the front versus back layers which is related to our task of judging the front versus back volume. Tsirlin, Allison, and Wilcox [2012] showed that for subjects to segregate two stereo-transparent planes, a greater inter-plane disparity is needed if the front plane is sparser than the back plane rather than vice-versa. They also found that for observers to perceive the front and back planes to be equally dense, the front plane needed to be more dense. They proposed that this bias was due to a figure-ground effect where the area between dots is assigned to the back plane. This back plane bias was also found in a moving dots study by Schütz [2012]. Aida, Kusano, Shimono, and Tam [2015] also showed that multiple depth layers viewed in stereo were perceived as having more dots than only one depth layer, even if the total number of dots was the same in the two cases. Again, this was thought to be due to the overestimation of the back surface.

Our current study considers volumes with many occluding surfaces. Occlusions have been neglected in previous studies of density estimation in volumes, namely these studies assumed the 3-D clutter consisted of lines and dots with little or no occlusions. Harris [2014] examined the perceived depth of cluttered scenes consisting of line elements. Both disparity gradients and number of elements were varied. Subjects judged a pair of stereotransparent planes to have a greater range of depth than a cluttered volume, even if in both cases the volume had the same depth and the same number of line elements. There was no significant effect from using a small versus large number of line elements. Goutcher and Wilcox [2016] also examined disparity volumes and tested how subjects discriminated the spread-in-depth and location-in-depth of the volume. They found that subjects used only the extreme disparity values to make their judgments. Sun, Baker, and Kingdom [2018] showed that binocular disparity affects perceived simultaneous density contrast where center and surround dots of a texture are presented at different disparities. They showed that simultaneous density contrast was reduced with larger plane separation or larger volumes.

Another variable that has been studied in density estimation is luminance. In nonoverlapping stimuli, Ross and Burr [2010] showed that perceived numerosity varied inversely with luminance, whereas perceived 2-D texture density did not. Tibber et al. [2012] found that varying the luminance contrast, had no effect on numerosity and density discrimination in 2-D. For overlapping surfaces presented with motion and with or without disparities, Schütz [2012] showed that the bias to see the back plane as more numerous was reduced when the front and back surfaces were assigned opposite contrast i.e. black or white dots on a gray background. This was presumably because it was easier to segment the front and back when they had different luminance than when they had the same luminance. However they found no evidence that a luminance difference facilitated density discrimination, as the JNDs were not affected.

The studies discussed above involved density or number discrimination of scenes consisting of dots or lines with little or no occlusion. Our study is fundamentally different in that we examine density discrimination in 3-D clutter in which occlusion effects are significant. We varied the amount of occlusion of surfaces in the 3-D clutter by manipulating both the size and density of the surfaces. The scenes were presented monocularly. However, a static monocular view of 3-D clutter yields only a weak percept of 3-D volume. To provide richer depth cues, we rotated each scene back and forth by a small angle. This gave a strong kinetic depth effect, and the dynamic occlusion cues typically provided enough information to resolve the two-fold rotation ambiguity.

Our goal was to examine the effects of occlusion when comparing the density in two halves of a cluttered volume. A brief overview of the experiments is as follows. Experiments 1 and 2 measure performance in discriminating density in the front and back halves of a cluttered volume. Experiment 1 uses black and white surfaces only, and Experiment 2 uses a luminance gradient namely a depth-luminance covariance. The black-white representation in Experiment 1 is similar to many 2-D texture studies and to layer studies such as Schütz [2012], except that they used two planes rather than two volume halves. The luminance gradient in Experiment 2 is similar to what we used in our previous work where we examined depth discrimination [Scaccia and Langer, 2018]. Experiment 2 allows us to compare performance in two types of luminance variation.

Experiments 3 and 4 examine how well observers can discriminate density in the left versus right halves of a volume. Both Experiments 3 and 4 use black and white surfaces. Experiments 3 and 4 were done to establish a baseline since the left-right task should be easier than the front-back task, as we argue below. They also allow us to probe whether observers use similar strategies in the front-back and left-right tasks, by comparing how the performance varies over scene parameters in the two different tasks.

We also present results for a model observer that counts pixels that correspond to visible surfaces in the two halves, and that compares the number of pixels or 'image occupancy' of the two halves. Our motivation for studying this model observer was two-fold. First, we wanted to understand the information that is available for doing that version of the task – namely a pixel level comparison of the two halves, rather than a surface density comparison. In particular, we wanted to understand how (if at all) this information varied between density and area conditions of our experiment. Second, we wanted to compare results of the human and model observers to see if human observers performance followed that of the model observers, which could indicate that human observers might be influenced by this image occupancy information.

### 4.2 Method

### 4.2.1 Subjects

Ten subjects participated in Experiments 1 to 3, with ages ranging from 19 to 71. The order of experiments 1 to 3 was randomized for each subject. Experiment 4 was run with six new subjects ages 19 to 61. Each subject was paid \$10. Subjects had little or no experience with psychophysics experiments and were unaware of the purpose of the experiments. Each had normal or corrected-to-normal vision. Informed consent was obtained using the guidelines of the McGill Research Ethics Board which is consistent with the Declaration of Helsinki.

### 4.2.2 Apparatus

Images were rendered using OpenGL (Khronos Group, Beaverton, OR) and were displayed using a Dell Precision T7610 workstation (Dell, Round Rock, TX) with an NVIDIA Quadro K4000 graphics card (NVidia, Santa Clara, CA). A 27-inch Apple monitor was used (Apple, Cupertino, CA). The display was gamma-corrected so that the displayed luminance values were proportional to the rendered gray level values.

# 4.2.3 Stimuli

The clutter in each scene consisted of a set of squares which were positioned and oriented randomly within two halves of a sphere of fixed diameter 24 cm. Each scene was defined by a mean density  $\eta$ , namely the number of surfaces per  $cm^3$ , and the area *A* of each surface. The total number of surfaces in each scene was  $N = \eta V$ , where *V* is the fixed volume of the sphere. The value *N* was rounded to the nearest integer. See Table 4–1 for the values of  $\eta$ , *A*, and *N* that were used in the different conditions of the experiments.

The values of  $\eta$  and A in Table 4–1 were chosen such that their product

$$\lambda \equiv \eta A \tag{4.1}$$

is constant within each column and increases from left to right. The parameter  $\lambda$  is called the *occlusion factor* [Langer and Mannan, 2012]. It is the expected total area of the surfaces in the clutter per  $cm^3$ . The greater the value of  $\lambda$ , the more occlusions tend to occur.

	$\lambda = 0.02$	$\lambda = 0.04$	$\lambda = 0.08$
r	$\eta = 0.06, A = 0.33, N = 434$	$\eta = 0.06, A = 0.67, N = 434$	$\eta = 0.12, A = 0.67, N = 869$
r	$\eta = 0.08, A = 0.25, N = 579$	$\eta = 0.12, A = 0.33, N = 869$	$\eta = 0.16, A = 0.5, N = 1158$
r	$\eta = 0.12, A = 0.17, N = 869$	$\eta = 0.24, A = 0.17, N = 1737$	$\eta = 0.24, A = 0.33, N = 1737$
_	4 1 701 0 0 (11 1	1 ( 1 '/	

Table 4–1: The 3 × 3 table shows values of mean density  $\eta$ , area *A*, and number *N* of surfaces for stimuli in our experiments. The occlusion factor  $\lambda = \eta A$  is constant in each column and increases from left to right. The area *A* is constant on the main diagonal and increases on the cross diagonal. The mean density  $\eta$  is constant on the cross-diagonal and increases on the main diagonal.

Note that the area *A* decreases and mean density  $\eta$  increases as one goes down each column of Table 4–1, and these variations exactly cancel to keep  $\lambda$  constant in each column. Also, the mean density  $\eta$  increases and area *A* is constant on the main diagonal (top left to bottom right), and area increases and mean density is constant on the cross-diagonal (bottom left to top right). These variations in  $\lambda$ ,  $\eta$  and *A* will be indicated by three arrows in subsequent figures. <sup>1</sup>

Our experiments measure how well observers can discriminate the density of surfaces in the front versus back halves and the left versus right halves. The levels of  $\Delta\eta$  in the two halves were chosen separately for each mean density value  $\eta$ , namely we defined nine density difference levels:

$$\Delta \eta = \{0, \pm \frac{\eta}{4}, \pm \frac{\eta}{2}, \pm \frac{3\eta}{4}, \pm \eta\}.$$
(4.2)

Thus the density of the two halves of each scene was  $\eta \pm \frac{\Delta \eta}{2}$ . Examples will be shown later in Figure 4–2.

We next consider the image formation model. Each cluttered scene sphere was rendered using perspective projection. The minimum and maximum depths of the clutter were defined to be at  $Z_0 = 58$  and  $Z_{max} = 82$  cm from the virtual subject's position, that is, the center of projection. This depth range corresponds to the diameter of the sphere.

<sup>&</sup>lt;sup>1</sup> Note the range of values of density and areas is larger in the central column, namely a factor of 4 instead of a factor of 2 range. A slightly cleaner design would have had  $(\eta = .08, A = 0.5, N = 579)$  and  $(\eta = .18, A = 0.25, N = 1158)$  in the top and bottom elements of the middle column. This would have yielded a gradual change in all the parameters within the top and bottom rows as well.

The projection plane or display screen was defined at  $Z_{mid} = 70$  cm which was the center depth of the sphere. Perspective effects were present, but were relatively weak since many surfaces were partly occluded and so there was large variation in image sizes of visible portions of the surfaces.

Scenes were rotated back and forth about the *X*-axis with an amplitude of 10 degrees and a rotational velocity of  $\pm 10$  degrees per second. Each stimulus was presented for four seconds, namely two periods of motion. The motion gave a strong kinetic depth effect. Moreover, dynamic occlusions between surfaces typically specified their ordinal depth, and so there was less of a tendency for depth reversals than what one typically has in 3-D cluttered scenes, namely if one uses small random dots. The motion may also help segment the front and back halves because the rotation yields opposite motion directions in the front versus back halves.

In Experiments 1 and 2, the observer's task was to indicate whether the front or back half was more dense. In Experiment 1, the squares in the front half had a different color than squares in the back half (black versus white). WB means white for the front half and black for the back half, and BW means black for the front and white for the back. We did not expect a difference between the WB and BW conditions, based on [Schütz, 2012] and [Tibber et al., 2012], but we included the two conditions just to be sure. Figure 4–1 shows examples of WB conditions. The different density and area combinations in the figure correspond to the entries in Table 4–1. Figure 4–2 (a) shows examples of the Experiment 1 stimuli for different densities in the two halves. Specifically, the four rows illustrate density differences  $\Delta \eta = \pm \frac{\eta}{2}, \pm \eta$  for each of the four Experiments.



Figure 4–1: Example stimuli for Experiments 1 and 3 (front-back and left-right, WB) at level  $\Delta \eta = 0$ . The 3×3 layout corresponds to Table 4–1. The arrows indicate the direction in which the variables  $A, \eta, \lambda$  increase.

In Experiment 2, we chose the luminance of the squares to vary continuously with depth Z. For negative depth-luminance covariance (DLC-), the luminance was chosen to be proportional to  $(\frac{Z_{max}-Z}{Z_{max}-Z_0})^3$ , where Z was the depth (cm) of the center of the square,  $Z_0$  and  $Z_{max}$  were the near and far limits of the clutter, as defined earlier. We chose this power law based on a simple  $Y^{\frac{1}{3}}$  approximation to the CIELUV's luminance factor, so that equal steps in normalized depth  $\frac{Z_{max}-Z}{Z_{max}-Z_0}$  in [0, 1] would have roughly equal steps in brightness (perceived luminance). Similarly, for positive depth-luminance covariance (DLC+), we chose luminance to be proportional to  $(\frac{Z-Z_0}{Z_{max}-Z_0})^3$ . We used a green background for Experiment 2 so that subjects could more easily distinguish between the surfaces and the background. Figure 4–2 (b) shows examples of the Experiment 2 stimuli.

Note that in Experiment 2, surfaces that are near the middle depth all have roughly the same luminance, regardless of whether they fall in the front half or back half. Thus, it is inherently difficult from the luminance information to decide if such surfaces belong to one half versus the other. Although the rotational motion cue provides some information for segmenting the front and back halves, this motion information is also weakest for surfaces near the middle depth of the volume, as these surfaces hardly move. For both of these reasons, for Experiment 2, observers must base their front-back judgment more on surfaces that are well short of or well beyond the middle depth. Since there is less information available in Experiment 2, we expect the Experiment 2 task to be more difficult and performance to be worse than in the Experiment 1 task where surfaces in the two halves are either white or black. As we will see later, performance was indeed worse for Experiment 2.

In Experiments 3 and 4, the observer's task was to indicate whether the left or right half of the clutter was more dense. The performance in this task gave us a baseline against which to compare the front-back performance, and also allowed us to compare human performance to that of a model observer who used the simple strategy of counting pixels that correspond to visible surfaces in the two halves. In Experiment 3, the surfaces in the front and back were WB and BW as in Experiment 1, but now the density was varied between the left and right halves. In Experiment 4, surfaces on the left and right were white and black (WB), respectively, or vice-versa (BW). Experiment 4 was added to test whether having black and white on the left and right halves as in Experiment 3. See Figure 4–2 (c) and (d) for examples of Experiments 3 and 4 stimuli.

The WB and BW scenes in Experiments 1, 3, and 4 were presented on a gray (0.5, 0.5) background. The luminance gradient scenes in Experiment 2 were presented on a green background, which had the same *Y* value as the gray background. See the Appendix in Scaccia and Langer [2018] for more details on display calibration and linearization.



Figure 4–2: Examples of stimuli for Experiments 1-4 at levels  $\Delta \eta = \pm \frac{\eta}{2}, \pm \eta$ . In each case, the values of density  $\eta$  and area *A* correspond to the middle condition in Table 4–1.

#### 4.2.4 Design

Our independent variables within each experiment were area *A* and density  $\eta$  and their product  $\lambda$ , and the sign of luminance variation across the volume (WB/BW or DLC-/DLC+). For each of the four experiments and for each combination of independent variables the stimulus levels  $\Delta \eta$  from trial to trial were chosen using a staircase procedure. We used a 1-up/1-down staircase with the nine stimulus levels  $\Delta \eta$  which were described above in Equation 4.2. The staircases targeted a proportion of choosing each of the two halves in 50% of the trials, *i.e.* the point of subjective equality (PSE). For each of the four experiments, the staircases for the different conditions were randomly interleaved. Each staircase began at a randomly chosen  $\Delta \eta$  level from the set of nine levels and then terminated after fourteen reversals.

Our dependent variables were bias  $\alpha$ , slope  $\beta$  and derived quantities JND and Weber fraction, which were defined as follows. Each staircase yielded the fraction of trials in which the subject chose the front half in Experiments 1 and 2 or right half in Experiments 3 and 4 as being more dense. We used the Palamedes toolbox [Prins and Kingdom, 2018] to fit a logistic function to these fractions

$$p(x; \alpha, \beta) = \frac{1}{1 + \exp(-\beta(x - \alpha))}$$

where x is one of the nine density difference levels  $\Delta \eta$ ,  $\alpha$  is the bias, and  $\beta$  is the slope.

We defined the just noticeable difference (JND) of density in the two halves to be value of  $x - \alpha$  such that x is the 75% threshold for choosing front or right, so JND  $\equiv \frac{\ln(3)}{\beta}$ . We then measured the density discrimination performance in terms of the following

quantity:

$$\frac{\text{JND}}{\eta + \frac{\alpha}{2}} \tag{4.3}$$

which we will refer to as a *Weber fraction*. For the front-back task, the denominator  $\eta + \frac{\alpha}{2}$  is the density of the front half at the point of subjective equality (PSE) since, when  $\Delta \eta = \alpha$ , the front and back half densities are  $\eta + \frac{\alpha}{2}$  and  $\eta - \frac{\alpha}{2}$  respectively. We used this density of the front half at the PSE as an estimate of the observer's perceived mean density, reasoning that the perceived density should be more reliable for the front half than the back half since the surfaces in the front are more visible. The Weber fraction is then the JND between the densities of the two halves, relative to the perceived mean density. Note that for the left-right task there is no bias and so the perceived mean is identical to the true mean in that case, and so the Weber fraction would simply be  $\frac{IND}{\eta}$ .

#### 4.2.5 Procedure

Subjects were seated so that their eye position was 70 cm directly in front of the screen, which corresponded to the virtual viewing position used in the rendering. The height of each subject's seat was adjusted so that the subject's eye would be roughly the same height as the center of the screen. We did not use a chin rest, so some slight variability in subject position was allowed. Subjects viewed the stimuli monocularly through the dominant eye. The non-dominant eye was covered with an eye-patch.

In Experiments 1 and 2, subjects responded by pressing the up or down arrow key to choose back or front half as denser, respectively. In Experiments 3 and 4, subjects responded by pressing either the left or right arrow keys to choose the left or right half as denser. Response time was limited to the four second duration of the stimulus. If the subject did not respond in some trial, then a random choice was made and a red X would

show on the screen. There was a rest period after every 100 trials for as long as the subject wanted. Prior to the experiments, subjects were given a practice session where they viewed ten trials from each experiment. This was done to ensure they were comfortable with the task and answering within four seconds. Each of the four experiments typically lasted around fifteen minutes. For the subjects that ran Experiment 1-3, these experiments were run in random order.

### 4.3 Results

For Experiments 1 and 2, we ran two-way repeated measures ANOVAs for bias and Weber fractions to test the effects of the occlusion factor  $\lambda$ , density  $\eta$ , area *A*, and the sign and type of luminance variation. For Experiments 3 and 4, we ran one-way ANOVAs for each of them rather than two-way ANOVAs that combined them, because these experiments had a different number and different set of subjects.

For all experiments, we found that the sign of the luminance variation had no effect. Therefore we do not report this factor. Instead we pool these conditions within each experiment. The lack of effect from sign of depth-luminance covariance is not entirely surprising, since this information specifies the ordinal depth of the two halves from the occlusions [Scaccia and Langer, 2018]. Although this information helps to segment the two halves [Schütz, 2012], this segmentation information is roughly the same for the two signs.

For all the statistical tests, we report exact p values except if p values are very small. A p value smaller than 0.05 is considered to be significant.

#### **4.3.1** Experiments 1 and 2 (front-back)

Figure 4–3 shows the mean of the biases  $\alpha$  in Experiments 1 and 2 (front-back). There was a strong main effect for the occlusion factor  $\lambda$  ( $F(2, 174) = 59.11, p < 10^{-19}$ ), namely a back bias ( $\alpha > 0$ ) for small  $\lambda$ , a near-zero bias for the middle  $\lambda$ , and a front bias ( $\alpha < 0$ ) for large  $\lambda$ . There was also a main effect for the type of luminance variation ( $F(1, 174) = 14.64, p < 10^{-3}$ ) with a more negative mean bias for Experiment 2 than Experiment 1, i.e. a greater bias to see the front as denser for Experiment 2 than for Experiment 1. There was also an interaction between the factors of  $\lambda$  and the type of luminance variation (F(2, 174) = 4.28, p = 0.015).

To gain some insight into how the biases depend on the occlusion factor  $\lambda$ , we consider *image occupancy* fractions, that is, the fraction of the pixels that correspond to surfaces in each half. Figure 4–4 (a) shows mean image occupancy fractions for the nine conditions of Table 4–1, specifically for the case of uniform density i.e.  $\Delta \eta = 0$ . For each experiment and within each condition, there is little variation in these values from scene to scene. The error bars show the standard deviations multiplied by 10 to better illustrate their relative magnitudes.

Within each row of Figure 4–4 (a), the means of front and back image occupancy fractions each increase with  $\lambda$ , and the difference between the means increases as well. Another way to view this trend in the image occupancy of the two halves is to let  $\Delta\eta$  vary instead of just taking  $\Delta\eta = 0$ , and to ask the question: what would  $\Delta\eta$  need to be for the two halves to have the same image occupancy? The answer depends on the occlusion factor: for a larger occlusion factor  $\lambda$ , the density difference  $\Delta\eta$  would need to be more negative. This effect is illustrated in Figure 4–3 by the black horizontal lines which show



Figure 4–3: Biases  $\alpha$  for human observers in Experiments 1 and 2. The nine plots correspond to the conditions in Table 4–1. Error bars indicate standard error of the mean. The gray lines indicate the  $\Delta\eta$  levels in Equation 4.2 which were the levels used in the experiment. The black lines indicate the  $\Delta\eta$  levels for which the mean number of pixels of front and back halves are equal. As we discuss later, these levels correspond to thresholds  $\tau$  of a model observer that compares the image occupancies of the two halves.


Figure 4–4: Mean image occupancy of (a) front and back halves for Experiments 1 and 2, and (b) left and right halves for Experiments 3 and 4, for each of the nine conditions of Table 4–1 and for equal density in the two halves ( $\Delta \eta = 0$ ). The means are over 10,000 scenes. Occupancy is defined by the number of pixels of the corresponding front, back, left or right half sphere, divided by the number of pixels occupied by the image projection of the spherical bounding volume. Error bars show the standard deviation (not standard error) multiplied by a factor of 10 to better illustrate their relative magnitudes.

the level of density difference  $\Delta \eta$  that would yield equal image occupancy for the front and back halves for each of the conditions. Note how these  $\Delta \eta$  values decrease as the occlusion factor  $\lambda$  increases.

Interestingly, the observer biases  $\alpha$  follow a similar trend. However, there is a crucial difference between these two trends, namely the observer biases are shifted relative to the density differences that give equal occupancy. In particular, the observer biases are closer to zero. This implies that observers are not merely judging the relative image occupancy of the two halves or the relative density of *visible* surfaces in the two halves, but rather they are judging the relative density of surfaces in the scene and taking account of the occlusion effects, albeit with a bias that depends on the amount of occlusion.

As the ANOVAs showed above, luminance type (Experiment 1 versus 2) affected bias as well. There was a main effect for this factor and there was also an interaction between luminance type and occlusion factor  $\lambda$ . Such effects are not surprising since the task in Experiment 2 is inherently more difficult than in Experiment 1, as discussed earlier in the Stimuli section, and the task is even more difficult when  $\lambda$  is larger. Observers may change their bias when the task is more difficult, and possibly rely more on perceived image occupancy than on perceived density in that case.

Figure 4–5 shows the means of the Weber fractions (recall Equation 4.3) for all four experiments. For Experiments 1 and 2, the occlusion factor  $\lambda$  effect was near significant (F(1,174) = 2.78, p = 0.057) with Weber fractions decreasing as  $\lambda$  increased. There was a strong main effect for the type of luminance variation  $(F(2,174) = 17.71, p < 10^{-4})$ , with greater Weber fractions for Experiment 2 (DLC) than Experiment 1 (WB). The greater

Weber fractions in Experiment 2 is not surprising because the task in Experiment 2 is inherently more difficult.

To explore the effect of the occlusion factor  $\lambda$ , we examined specific combinations of density  $\eta$  and area A. There was a main effect for changes in density along the main diagonal conditions (F(2,54) = 6.34, p = 0.003) and again a main effect for the type of luminance (Experiment 1 vs. 2) (F(1,54) = 7.18, p = 0.009). On the cross diagonal, there was no main effect from changes in area (F(2,54) = 0.87, p = 0.42), but there was again a main effect for the type of luminance (F(2,54) = 8.78, p = 0.004). There was also a main effect within columns (F(2,174) = 8.1,  $p < 10^{-3}$ ), with lower Weber fraction as we move down each column where density increases and area decreases, as well as type of luminance effect (F(1,174) = 18.8,  $p < 10^{-4}$ ). This result is consistent with the result on the main diagonal where density had an effect. We conclude that the near-significant Weber fraction effect of the occlusion factor  $\lambda$  in both Experiments 1 and 2 was primarily due to a strong density effect, not to occlusion *per se*. We will discuss these results further in the Discussion section.

### **4.3.2** Experiments 3 and 4 (left-right)

In Experiments 3 and 4, the task was to discriminate density for the left and right halves (see Figure 4–2 c, d). The left-right biases were all close to zero as expected, and so we do not show them in Figure 4–3.

For the Weber fractions, we ran one-way ANOVAs for Experiments 3 and 4, as they had a different number and different set of subjects. There was no significant effect for the occlusion factor  $\lambda$ , neither for Experiment 3 (F(2,87) = 2.55, p = 0.08) nor for Experiment 4, (F(2,51) = 1.8, p = 0.18). However, there was a strong main effect for density

(on the main diagonal where area was fixed and density varied) for both Experiment 3  $(F(2,27) = 24.3, p < 10^{-6})$  and for Experiment 4, (F(2,15) = 10.9, p = 0.001). There was a weaker but still significant effect of area (on the cross diagonal where density was fixed and area varied) both for Experiment 3 (F(2,27) = 5.24, p = 0.01) and Experiment 4 (F(2,15) = 7.6, p = 0.005). The density and area effects were in opposite direction, with Weber fractions decreasing as density increased, and Weber fractions increasing as area increased. These opposing effects might be the reason why there was no significant effect from the occlusion factor  $\lambda$  which is the product of density and area. There was also a main effect within columns, both for Experiment 3  $(F(2,87) = 14.62, p < 10^{-5})$  and Experiment 4  $(F(2,51) = 35.5, p < 10^{-9})$ , namely Weber fractions decrease moving down each column of Table 4–1 as density increases and area decreases. This is consistent with the density and area results found on the main and cross diagonals, respectively, namely that Weber fractions increased as density increased and as area decreased.

We compared Weber fractions for Experiment 3 versus Experiment 4 using a t-test. Recall that Experiment 3 had black and white separated in the front and back and Experiment 4 had black and white separated in the left and right halves. We suspected that Experiment 4 would be easier since subjects would not need to disentangle white versus black in each half. However, a one tailed t-test on the signed differences of the Weber fractions showed no significant difference (t = 0.81, p = 0.22).

Finally, we compared Weber fractions for Experiments 1 versus 3. This is an interesting and important comparison because the stimuli for the  $\Delta \eta = 0$  level in the two experiments were the same. The crucial difference between the two is how observers deal with the more challenging occlusion effects in the front-back task versus the left-right task. In particular, we expected performance to be worse in Experiment 1 since observers need to account for the different image occupancies of the front versus back which are due to occlusion, whereas in Experiment 3 (and Experiment 4) observers could perform the left versus right task in principle by just comparing the overall image occupancy in the two halves. A one-tailed t-test indeed showed that Weber fractions were much higher for Experiment 1 than Experiment 3 (t = 6.95,  $p < 10^{-6}$ ).



Figure 4–5: Mean Weber fractions for human observers for all four experiments. The nine plots correspond to the conditions in Table 4–1. Error bars show standard error of the mean.

#### 4.3.3 Model observers

To gain more insight into the effect of occlusions on the difficulty of the task, we present model observers that only compare the image occupancy of the two halves. We assume that the model observer in each condition is unbiased ( $\alpha = 0$ ), namely it knows the *expected* difference in pixel counts in the two halves. We refer to this expected difference in each condition as  $\tau$ . We define the front-back model observers to respond 'front' when the number of front pixels exceeds the number of back pixels by this threshold  $\tau$ . We computed  $\tau$  for each condition in advance, namely it was the mean difference in the number of pixels from visible front versus back surfaces over 10,000 scenes for that condition and for the  $\Delta \eta = 0$  level. These  $\tau$  values were computed using the data in Figure 4–4(a), and the density differences that correspond to the  $\tau$  values are plotted as black lines in Figure 4–3. Similarly, the left-right model observers compare the number of pixels corresponding to surfaces in the left and right halves. In this case the expected difference for the two halves is zero ( $\tau = 0$ ).

Before we discuss the performance of the model observers, we examine the data in Figure 4–4 (a). Within each of the three columns of Figure 4–4 (a), the mean image occupancy of the front half and back half is roughly constant. This is because the expected value of the number of the image occupancy of the front and back depends only on the occusion factor  $\lambda$ , which is constant within each column. However, the standard deviations of image occupancy are not constant within each column but rather they decrease from the top to bottom, typically by over 30%. The reason is that having a larger number of smaller surfaces drives the pixel counts of front and back surfaces to be closer to their expected values. (A similar effect was described in Figure 3 of Langer and Mannan [2012].) It

follows that the mean difference in the number of pixels of the front and back also will be constant within each column, and the standard deviations of the front-back difference will decrease from top to bottom within each column as well. Similarly, for Figure 4–4 (b), the means for left versus right are the same within each column, but again the standard deviation decreases moving down each column, and so the standard deviations of the difference will decrease moving down each column. Based on these observations about the standard deviations, we predicted that the model observer should be more sensitive to density differences moving down each column, both for the front-back task and the left-right task. We will see below that this prediction holds.

Another factor that affects the variation in the visible number of each half is the motion in the stimuli. Recall that in each trial, the stimulus is not just a static image, but rather it is a sequence of images *i.e.* the clutter rotates. Having a sequence of images increases the chances that any surface point will be visible at some frame during the sequence. To explore this factor, we compared performance of model observers that used a single frame to model observers that used multiple frames. The idea is that, by averaging the counts of the two halves over multiple frames, the model observer can reduce the variance (or standard deviation) in the pixel counts in the two halves and improve its performance. This is similar to how human observers can perform the task more easily with rotation present than with only a static image, as the rotation creates dynamic occlusions which help the human observer resolve the depth reversal ambiguity.

We compared the Weber fractions of a model observer that used just one frame corresponding to 0 degrees rotation with a model observer that used three frames, namely, 0 degrees and  $\pm 5$  degrees rotation about the *X*-axis. Using a similar method as in Figure 4–4, we found that the standard deviations for the image occupancies for each half of the sphere and for each condition were reduced by approximately 40% in each condition for three-frame observers relative to the one frame observers. We predicted that these reduced standard deviations in the image occupancy relative to the fixed differences in the mean image occupancies would decrease the model observer's Weber fraction.

Although the model observers only count pixels in the two halves in each frame, these counts are precise and they turn out to be sufficient for performing the task quite well. To measure exactly how well, we used a much more refined  $\Delta\eta$  range for the model observer experiments. We defined density difference levels for the model observer, by dividing each of the human observer's stimulus levels by 10, so the model observer's levels were

$$\Delta \eta = \eta_f - \eta_b = \{0, \pm \frac{\eta}{40}, \pm \frac{2\eta}{40}, \pm \frac{3\eta}{40}, \pm \frac{4\eta}{40}\}.$$

For each of the conditions of Experiments 1 and 4, we ran model observers on 100 staircases and fit psychometric functions to each staircase.

Figure 4–6 shows the mean of the model observer Weber fractions for the frontback task (white bars) and the left-right task (black bars). Two observations can be made, and both follow the predictions above. First, for each column of Figure 4–6, the Weber fractions indeed decrease as we move down the column. The reason is that when density  $\eta$  is larger and area *A* is smaller, there is less variability in the number of front pixels and back pixels for each condition and so the observer can detect more reliably whether the difference in front versus back pixel counts is greater than the expected value  $\tau$  of that difference. Second, the three frame model observer had lower Weber fractions than the one frame model observer. Thus, although the rotational motion itself (i.e. the image velocities of the surfaces) was not used by the model observer, the motion information did provide a significant benefit to the model observer, namely by providing more samples for comparing pixel counts in the two halves.

Finally, the model observer also shows a density effect on main diagonal and an area effect on cross diagonal. The model observer has no notion of the individual surfaces in the stimuli. The effects from density and area are due to the occlusion effects just discussed.



Figure 4–6: Mean Weber fractions for model observers using 1-frame and 3-frame for front-back and left-right tasks. The nine plots correspond to the conditions in Table 4–1. Error bars shows standard error of the mean.

### 4.4 Discussion

### 4.4.1 Comparison between human and model observers

We have compared the human observer biases to the model observer parameters in the Results section, and so we concentrate our discussion here only on the Weber fractions. For the front-back task, Weber fractions for the human and model observers decreased moving down each column of the  $3 \times 3$  plots. For model observers, this trend was explained by a reduced variability in the pixel counts of front and back. For human observers, the decrease in the Weber fractions was mainly due to density. This density effect has been shown previously by Anobile and colleagues, namely Weber fractions for density decrease as density increases. One might have expected an area effect as well, since increasing the area and holding density fixed creates more occlusions which should make the task more difficult because the deeper surfaces would be less visible. However, we did not find an area effect. The reason may be that varying area has a second occlusion effect that works in the opposite direction. While occlusions typically resolve the depth reversal ambiguity, they do not always do so: in particular, depth reversals are more likely to occur when the area of elements is small than when their area is large, that is, when there are fewer occlusions. When a depth reversal does occur, it tends to produce an incorrect response and so it follows that the depth reversals tend to produce more incorrect responses when the area of the elements is smaller. As the two occlusion effects work in opposite directions, we would expect these effects to partly cancel out - at least for the front-back task where depth reversals lead to incorrect responses.

For the left-right task, Weber fractions decreased moving down each column of the  $3 \times 3$  plots, both for model and human observers. For model observers, this trend was

explained by a reduced variability in the pixel counts of each half. For human observers, pixel count variability is unlikely to have played a role since it requires a precision in estimating image occupancy of the left and right halves that is presumably far greater than what humans are capable of. Rather, human observers seemed to show effects for both density and area, with a stronger main effect for density than for area. Weber fractions decreased with density which has been shown before by Anobile and colleagues. Weber fractions increased with area, with density held fixed, presumably because occlusions interfered with the visibility of deeper surfaces, as discussed above. Note that in the left-right task, the depth reversals can also occur when the area of the elements is smaller but these would not lead systematically to incorrect responses, unlike for the front-back task.

A key difference between the human and model observers is that the model observers have roughly the same Weber fraction for the front-back task as for the left-right task within each of the nine conditions, whereas the human observer Weber fractions are larger for the front-back task than for the left-right task. One reason is presumably that the model observer knows the expected difference in image occupancy for the two halves in each condition, whereas the human observer does not. (Recall that the staircases for all the different conditions were randomly interleaved.) In the front-back tasks, in particular, human observers not only have to assign the bright and dark luminances to the front and the back halves, but they also need to compare their densities to some internal standard that takes account of occlusions. Since there must be some uncertainty in what this internal standard should be on each trial in the front-back task, human observers naturally perform worse in this task than in the left-right task. Finally, we saw that the single-frame model observers had higher Weber fractions than the multiple frame model observers. We did not run our experiment on single frame stimuli for the human observers. The reason is that the one frame task is too difficult. As one can see by the single frame examples in Figure 4–1, even the front-back ordering can sometimes be difficult to discern, especially at the lowest values of the occlusion factor  $\lambda$ .

### 4.4.2 Comparison with previous work

Previous studies used a variety of parameters, including intrinsic parameters such as the density and area of the elements, and extrinsic parameters such as the overall area of the stimulus, the number of elements, and eccentricity.

With regard to the front versus back biases, our back bias finding with small occlusion factor  $\lambda$  was consistent with the back bias found in studies with depth layers [Tsirlin et al., 2012, Schütz, 2012, Aida et al., 2015]. These studies used a range of parameters and typically did not allow occlusions. For example, Schütz [2012] used a 2-D density range of 0.25-2 dots/deg<sup>2</sup>, an *N* range of about 20-150 dots, and a fixed 10 degrees diameter stimulus. Tsirlin et al. [2012] used a fixed density of about 20 dots/deg<sup>2</sup>, 3000 dots, and a fixed stimulus size of about 13 × 13 degrees. For our stimuli, the density range was approximately from 1.5 to 6 squares/deg<sup>2</sup>. Our smallest number of elements was 434 and our largest was 1737, and the circle bounding our volume had a diameter of 20 degrees. These values are thus in a similar range as studies above. This suggests that the back bias is robust to differences in densities, number of elements, and size of stimulus, as well as differences to the arrangement of the 3-D stimulus (layers or volume), at least when the amount of occlusion is low.

Regarding Weber fractions, we found them to decrease as density increased for both front-back and left-right human observers. We compare our results to Anobile et al. [2014] who were the first to show how Weber fractions varied with density. Using patches that were centered 13 degrees left and right of fixation and densities ranging from 0.02 to 4 dots/deg<sup>2</sup>, they found constant Weber fractions for lower densities and decreasing Weber fractions for higher densities, where the switch occurred at about 0.2 dots/deg<sup>2</sup>, depending on patch area. Anobile et al. [2015] subsequently showed that the switching point from constant to decreasing Weber fractions depends on eccentricity. For example, using centrally-presented patches of diameter 8 degrees and presented in sequence, they found that the switch from constant to decreasing Weber fractions occurred at much higher densities, namely about 2 dots/deg<sup>2</sup>, and that the switching point decreased when the patches were presented more peripherally. Our stimuli had diameter of 20 degrees and we did not control eye movements, and so they are a mix of central and peripheral presentation. Moreover, while our mean 2-D densities varied from 1.5-6 elements/deg<sup>2</sup>, our 2-D densities varied within each stimulus, namely greatest at the center of the projected sphere and diminishing to zero at the circular edge. Overall though, our 2-D densities and eccentricities were in the range that was similar to where Anobile and colleagues found decreasing Weber fractions as density increased, and so our results on decreasing Weber fractions are consistent with their findings.

As for comparing to previous work on luminance, our stimuli used a combination of bright and dark elements on a gray background, whereas Ross and Burr [2010] used equiluminant dots on a black background. They found that perceived numerosity increases with decreasing luminance, but perceived density does not. There was no obvious comparison to make because of the differences in stimuli between ours and theirs. Tibber et al. [2012] varied the luminance so that one patch had twice the contrast of the other, and they used a gray background. They did not find that such luminance variations affected density sensitivity. Our results were consistent with this finding.

### 4.5 Conclusion

Our results provide new insights into the perception of density of 3-D clutter. Most previous studies ignored occlusion effects, whereas our study addressed occlusion effects directly. The biases that we found depended on the level of occlusion. The bias to see the back as more dense in our low occlusion scenes was consistent with previous findings, which shows that the back bias previously found for overlaid planes with no occlusions extends to volumes with a low occlusion factor  $\lambda$ . The bias crossed over to the front when occlusion was higher. This dependence of bias on the level of occlusion has not been previously reported. We also found that these front and back biases did not depend on the density  $\eta$  and the area *A* of the elements per se, but rather they depended on the product of these two variables, namely the occlusion factor. Finally, we found that the biases roughly followed the difference in image occupancy: as the occlusion factor increased, there was a greater bias to see the front as more dense, and this bias roughly followed the difference in occupancy of the front and back surfaces.

We used several cues to help observers perceive the relative depth of surfaces in the clutter in the presence of occlusions. We used rotating 3-D clutter rather than static clutter so that the rotation would provide a strong kinetic depth effect, and the occlusions generally resolved the two-fold ambiguity in the rotation direction, especially with larger

occlusion levels. The motion also provided a cue for density discrimination, namely it provided multiple views of the volume which revealed more points that were deeper in the volume. The rotation may also have helped observers segment the surfaces into front and back halves since the rotation yields opposite motion directions in front versus back halves. We also used differences in luminance in the two halves to help observers to perceptually segment the front and back halves of the 3-D clutter. For the front-back task in particular, we compared a white-black luminance condition versus a depth-luminance covariance condition. Using black and white in the two halves reduced the front bias and also increased the sensitivity to density differences, in particular, it decreased the Weber fractions overall.

We found differences in Weber fractions for the front-back versus left-right tasks. For the front-back task, Weber fractions decreased when density increased which is consistent with previous studies. However, the area of the elements of the clutter did not seem to have an effect. We believe this is because increasing the area leads to two competing effects. On the one hand, it creates more occlusions which makes the task more difficult because the back half is less visible, but on the other hand it also reduces the likelihood of depth reversals and this reduced likelihood of errors makes the task easier. Future work could address the trade-off of these two competing effects. Note that, for the left-right task, a depth reversal would itself not lead to an error, and so the only area effect for the left-right task should be that an increase in area leads to more occlusions which should make the task more difficult. This would explain why human Weber fractions for the left-right task increased when area increased. In sum, we have seen that human observers use a combination of image cues to discriminate density in 3-D clutter. The effects we found for bias generalize previous results, and the effects we found for sensitivity (Weber fractions) seems to be consistent with what has been found for 2-D stimuli in previous studies, where occlusions were not studied. In 3-D clutter the combination of cues is more complicated because the density cues interact with other cues that are present, in particular, the area and the occlusion factor which is the product of density and area. One interesting topic for future work is how and why the Weber fraction varies with combinations of variables. One could also address these effects with different distributions of clutter and with different viewing conditions as well, for example, with motion parallax and stereo. There may be a subtle interplay between these variables, and it would be interesting to explore observer's strategies to deal with these subtleties in different tasks and over different ranges of scene and viewing parameters.

# Acknowledgements

This research was supported by a doctoral grant awarded to Milena Scaccia from the Fonds de Recherche du Quebec - Nature et Technologies (FRQNT), and an FRQNT team research grant and Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant to Michael Langer.

### REFERENCES

- Saori Aida, Tsutomu Kusano, Koichi Shimono, and Wa James Tam. Overestimation of the number of elements in a three-dimensional stimulus. *Journal of Vision*, 15(9):23, 2015.
- Giovanni Anobile, Guido Marco Cicchini, and David C. Burr. Separate mechanisms for perception of numerosity and density. *Psychological Science*, 25(1):265–270, 2014.
- Giovanni Anobile, Marco Turi, Guido Marco Cicchini, and David C. Burr. Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *Journal of Vision*, 15(5):4–4, 2015.
- Horace B. Barlow. The efficiency of detecting changes of density in random dot patterns. *Vision Research*, 18(6):637–650, 1978.
- Jason Bell, Aaron Manson, Mark Edwards, and Andrew Isaac Meso. Numerosity and density judgments: Biases for area but not for volume. *Journal of Vision*, 15(2):18, 2015.
- David Burr and John Ross. A visual sense of number. *Current biology*, 18(6):425–428, 2008.
- Guido Marco Cicchini, Giovanni Anobile, and David C Burr. Spontaneous perception of numerosity in humans. *Nature Communications*, 7:12536, 2016.
- Steven C. Dakin, Marc S. Tibber, John A. Greenwood, Frederick A. A. Kingdom, and Michael J. Morgan. A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences*, 108(49):19552–19557, 2011.
- Frank H. Durgin. Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human perception and performance*, 21(1):149, 1995.

- Frank H. Durgin. Texture density adaptation and visual number revisited. *Current Biology*, 18(18):R855–R856, 2008.
- Robert Stanton French. Pattern recognition in the presence of visual noise. *Journal of Experimental Psychology*, 47(1):27, 1954.
- Ross Goutcher and Laurie M. Wilcox. Representation and measurement of stereoscopic volumes. *Journal of Vision*, 16(11):16, 2016.
- Julie M. Harris. Volume perception: Disparity extraction and depth representation in complex three-dimensional environments. *Journal of Vision*, 14(12):11, 2014.
- Edna L. Kaufman, Miles W. Lord, Thomas Whelan Reese, and John Volkmann. The discrimination of visual number. *The American Journal of Psychology*, 62(4):498–525, 1949.
- Michael S. Langer and Fahim Mannan. Visibility in three-dimensional cluttered scenes. *Journal of the Optical Society of America A*, 29(9):1794–1807, 2012.
- Michael S. Langer, Haomin Zheng, and Shayan Rezvankhah. Depth discrimination from occlusions in 3D clutter. *Journal of Vision*, 16(11):11, 2016.
- Nicolaas Prins and Frederick AA Kingdom. Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the palamedes toolbox. *Frontiers in Psychology*, 9:1250, 2018.
- John Ross and David C. Burr. Vision senses number directly. *Journal of Vision*, 10(2):10, 2010.
- Milena Scaccia and Michael S. Langer. Signs of depth-luminance covariance in 3-D cluttered scenes. *Journal of Vision*, 18(3):5, 2018.

- Alexander C. Schütz. There's more behind it: Perceived depth order biases perceived numerosity/density. *Journal of Vision*, 12(12):9, 2012.
- Hua-Chun Sun, Curtis L. Baker, Jr., and Frederick A. A. Kingdom. Simultaneous density contrast and binocular integration. *Journal of Vision*, 18(6):3, 2018.
- Ernest Henry Taves. Two mechanisms for the perception of visual numerousness. *Archives* of Psychology (Columbia University), 265(47), 1941.
- Marc S. Tibber, John A. Greenwood, and Steven C. Dakin. Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision*, 12(6):8, 2012.
- Inna Tsirlin, Robert S. Allison, and Laurie M. Wilcox. Perceptual asymmetry reveals neural substrates underlying stereoscopic transparency. *Vision Research*, 54:1–11, 2012.

# CHAPTER 5 Discussion and Conclusion

The stimuli we explored in the thesis had simple distributions of clutter. In Chapter 3, the volume had clutter that was uniformly distributed. In Chapter 4, the clutter in each half of the volume was uniformly distributed but with different densities. In Section 5.1, we describe more complex distributions of clutter such as clumped clutter that we explored in our trajectory but remain topics of future work. We describe possible future directions for depth and density discrimination of clumped clutter. In Section 5.2, we also describe different depth-luminance models we attempted to study. In Section 5.3, we summarize the main results of the thesis work and conclude.

# 5.1 Future work: Perception of Clumped 3-D Clutter

There exist applications that present the challenging situation of non-uniform clutter. Examples include atoms of a molecule or leaves on a tree where leaves clump together. Some researchers have modelled clumping phenomena for the foliage case [Nilson, 1971].

One question to explore is: how can human observers use luminance information to infer the spatial relationships of objects in a clumped cluttered scene? If we consider a uniform diffuse illumination model i.e. ambient occlusion, we showed in Appendix 2.A, that for uniform clutter, luminance decreased exponentially with depth and so observers can use a rule such as "dark-means-deep". However, in clumped clutter, ambient occlusion would not necessarily exhibit an exponential decrease in luminance throughout the volume. Depending on the geometry of the scene, it is possible to have a surface that is highly occluded and located in a clump that is close to the observer. That surface would be darker than a surface that is less occluded and located in a further clump. Therefore, on a global scale, luminance specified by AO may not necessarily be correlated to depth. On a local scale, if clutter is uniformly distributed within each clump, then luminance correlates to depth locally. Of course, if we use a DLC function that ignores scene geometry, and simply shades all surfaces according to their depth in the scene, then luminance correlates to depth on a global scale. However, it is not clear whether using a luminance cue that ignores geometry best conveys the clumping properties of the scene.

It would be interesting to show that human observers use a rule that is more sophisticated than "dark-means-deep" for clumped cluttered scenes. This would be the clumped cluttered scene analogue to what was shown in [Langer and Bülthoff, 2000], that for smooth surfaces rendered under uniform diffuse illumination, the visual system does not simply discriminate depths based on luminance, but also takes account of surface normal variations when comparing depths of neighbouring surface points. In order to understand the spatial relationships of surface elements in clumped clutter, observers would need to first group surface elements according to the clump they belong to. Second, they would need to judge the relative depths of the clumps, and third, judge the relative depths of surfaces within each clump. If luminance cues do not specify the depth relationships, other cues such as the occlusion relationships between the clumps, binocular disparities, and motion parallax can help. Another cue worth considering is color or texture. If it is difficult to see the depths of the clumps relative to one another, or to see where one clump ends or the other begins, then assigning different colors or textural/material properties to each clump may help us perceive the distribution of clumps. Such a technique is applied in direct volume rendering applications where different materials are assigned different colors and opacities.

Aside from the depth perception of surfaces in clumped clutter, it would be interesting to examine how the visual system perceives the clumps themselves within the clutter. Important applications for detecting clumps in clutter lie in medical imaging, where the motivation is to find an anomaly within tissue, e.g. a tumor. In foliage-like scenes, a clump can indicate an object in a tree. A clump-detection experiment would be similar in flavor to the experiment in Chapter 4, where observers had to detect a spatial change in the clutter. Possible questions to ask are: how dense should the clump be to be detectable by human observers? What shading model best shows the clump?

# 5.2 Other Luminance Models

Another direction for future work is to explore how different luminance models would affect depth perception. We initially were interested in determining whether luminance variations caused by surface orientation would affect depth perception. In Appendix 2.A, we showed that for uniform clutter, the model that considers surface orientation produces almost the same luminances as the model that does not, and thus had no effect on performance. We then attempted to design other experiments that address how arbitrary depth-luminance models affect depth perception. We considered another type of luminance-gradient effect where luminance now decreases with distance from a specific point in the scene, for example, from the middle of the volume. We will call this type of luminance-gradient effect GLOW. Zavagno [1999] has presented this effect on a 2-D scene where a luminous mist spreads out from a central area and the background has the same luminance as the central area. Zavagno et al. reported that neither the central area nor the background

needed to be at a maximum luminance to be perceived as glowing. Gray centers were also reported as glowing as long as the background had a relatively high luminance. Also, the luminance gradient did not need to be linear or continuous for the central area to appear self-luminous.

Koenderink et al. [2007] explored the GLOW effect in the context of the 3-D visual light field. A spherical gauge object was introduced into the scene and observers adjusted its luminance so that it fit with the rest of the scene. This implicitly meant that they inferred the structure of the light field where three different physical light fields were used: open sun-light, overcast, and GLOW which had a light source in the center of the scene. Observers performed well in all scenes except when the gauge object was placed in a shadowed area of the volume. Observers had trouble detecting cast shadows.

In a 3-D cluttered scene, the occlusions provide ordinal depth information. When immersed in a light field, the luminances of the occluders provide additional ordinal information about the occluders. The presence of occluders also reveals the structure of the light field, as seen in [Koenderink et al., 2007].

We explored GLOW for discriminating the depth of two targets embedded in clutter. We used the same setup as the experiment in Chapter 3, except we used a sphere instead of a cube. We found that the way targets are set up for our scenes made it difficult to discriminate depth. The targets in our experiments are always of equal depth from the center of the sphere, with some jitter in the X and Y directions. Therefore distance from the center is almost equivalent for both targets, making their luminance almost equivalent, as seen in Figure 5–1 (a).



Figure 5–1: GLOW stimuli. (a) Depth discrimination of two targets with centered GLOW. (b) Depth discrimination of two virtual light sources.

We explored alternative GLOW stimuli for a depth perception task. For example, instead of asking observers to discriminate the depth of two shaded targets, we could ask them to discriminate the depth of two light sources within the clutter. The light sources could be virtual XYZ-lights such that luminance decreases away from their position, as shown in Figure 5–1 (b).

Performing the task would require that subjects use information about the global luminance distribution. Occlusions can provide insight on the relative depth of the two virtual light sources. For example, in Figure 5–1 (b), the further light source has a greater number of darker squares occluding it than does the closer light source. In Chapter 3, we showed that occlusions specified the sign of the DLC. We can probably similarly find that occlusions specify the directions of the GLOW luminance distribution, which would help infer depth in the scene. This could be another avenue for future work.

### 5.3 Conclusion

Our quest to understand the perception of 3-D clutter has led to interesting discoveries about the roles played by available cues, especially luminance and occlusions. Occlusions are inherent to 3-D clutter, and provide rich ordinal depth information. We have demonstrated that this information can be used to determine the sign of the DLC in order to infer the relative depth of two target surfaces (Chapter 3). This result has also helped observers resolve the sign of the luminance variation when discriminating the density of the front and the back halves of a cluttered volume (Chapter 4). We also showed that occlusions influence the direction of the bias for discriminating the density of the front versus back of a cluttered volume, where there is a bias to judge the back as denser when occlusion is low, and the bias crosses over to the front when occlusion is higher. There were also different effects on sensitivity depending on whether we varied the density or the area of the elements even if by design both these variations produced the same level of occlusion, namely Weber fractions decreased with density for both the front-back and left-right tasks and increased with area for the left-right task.

Evidently, occlusion and luminance cues play vital roles for depth and density discrimination in 3-D clutter. However, there exist scenarios where these cues do not necessarily guide observers to an accurate percept of the scene. For instance, in the depth discrimination study of Chapter 3, we found that there are subtle interactions between occlusions, color and luminance. When the colors of the targets differ from the distractors, occlusions do not necessarily help observers infer the sign of the DLC, as there is a prior for "dark-means-deep" (DLC-). Also, in Chapter 4, we found that for the front-back task, that there are competing effects between the likelihood of depth reversals and visibility, based on the occlusion level. When occlusion is low, a depth reversal is more likely to occur, but the back elements are more visible. When occlusion is high, depth reversals are less likely to occur but the back elements are less visible. Also, observers are unaware of the expected differences in visibility between the front and the back halves for different levels of occlusion, as this expected value varied with conditions randomly from trial to trial. This uncertainty could have contributed to the human observer biases.

In the density discrimination tasks of Chapter 4, we had expected human observers to take advantage of the rich 3-D information provided by the combination of DLC and motion. Using a luminance cue such as DLC can help resolve any depth order ambiguity. However, human observers perform better when squares are only black or only white in each half. Perhaps this is because it is easier to segment the two halves, and human observers may use motion to obtain more samples of this intensity information from multiple frames, as done by the model observers.

In conclusion, we have contributed new ideas for understanding the perception of 3-D clutter. We have shown how observers use the information available from occlusion and luminance cues for different tasks, namely depth discrimination and density discrimination tasks in 3-D cluttered scenes. As part of the thesis work, we have also derived an approximate model for ambient occlusion using a probabilistic model of surface visibilities from occlusions, which we used as a DLC function. We hope that the findings of this thesis can serve as guidelines for designing perceptually effective renderings of 3-D clutter, and inspire future research on the perception and visualization of complex 3-D cluttered scenes.

# APPENDIX A Gamma Calibration

When displaying stimuli for a psychophysics experiment, it is important that images are presented as intended. Images produced for a psychophysics experiment may not be displayed correctly to a monitor because monitors are non-linear with respect to input. We therefore must gamma-correct the stimuli before displaying them.

In the event that the monitor's gamma is unknown, or one wishes to verify gamma for a particular monitor, one may use a visual method or a tool such as a spectroradiometer to compute a gamma look-up table, as we describe below.

Also, because we use a calibrated monitor for our experiments, it is important that the brightness setting stay the same each time. This is because changing the brightness setting might not just scale the brightnesses but may cause a non-linear change, which one would need to measure in order to correct.

The relationship between the input value of a pixel and the displayed luminance can be modelled by the power function,

$$L = D^{\gamma}, \tag{A.1}$$

where *L* is the displayed luminance,  $D \in [0,1]$  is the input pixel value, and gamma is denoted as the positive constant  $\gamma > 0$ . The non-linearity can be corrected by applying the inverse relationship to the input so that

$$L = (D^{\frac{1}{\gamma}})^{\gamma}. \tag{A.2}$$

This is known as gamma correction.

Below, we describe how to determine the value of  $\gamma$  either by using visual calibration which does not require any specialized equipment, or by using a tool (such as a photometer or spectroradiometer).

# **Gamma Correction using Visual Calibration**

We use visual calibration where two fields are presented side by side [Colombo and Derrington, 2001]. The right is a patch of pixels for which the voltage can be uniformly adjusted and the left is a field in which a fraction Q of pixels is set to the maximum voltage (white, 1) and the rest are set to the minimum (black, 0). The subjects' task is to adjust the voltage so that both fields look equally bright. We measure the voltage D which produces a given fraction Q of maximum luminance  $L_{max} = 1$ ,

$$L = D^{\gamma} = QL_{max}, \tag{A.3}$$

e.g. to find the voltage D which produces the color that is halfway between white and black, we use Q = 0.5 to obtain 50% of the maximum luminance the monitor can produce.

To obtain an estimate of gamma, a smooth curve can be fit to the values of Q and D using least-squares. We linearize Equation A.3 by applying the natural log to both sides,

$$\gamma \ln D = \ln Q, \tag{A.4}$$

therefore we can use linear least squares. Alternatively, we can record the values of Q and D in a look-up table, and use interpolation to obtain D values for intermediate Q values.

Because we are using OpenGL for our experiments, we carry out the visual calibration using stimuli generated by OpenGL. The adjustable patch of pixels is generated using a



Figure A–1: Visual Calibration (note the above image is scaled 60%). The left field is a rectangular wave grating with duty cycle 0.5. The voltage of the right rectangle is adjusted until the two fields appear equally bright. This is done to measure the relationship between voltage and displayed luminance.

quad primitive. The observer can adjust the voltage of the quad's pixels using the up and down arrows of the keyboard. The other patch is generated as a rectangular wave grating with white and black bars. The spatial period of the grating is 10 pixels and the duty cycle (which is the proportion of pixels that are white) is varied from 0.1 to 0.9. We define the width of a line to be one pixel. E. g. if duty cycle is 0.1, each white line drawn is followed by nine black lines, for duty cycle 0.2 two white lines are followed by eight black lines, and so on and so forth. See Figure A–1 for an example with duty cycle = 0.5.

For each duty cycle, we adjust the voltage so that the grating and the patch appear equally bright, and record when a match is made. The recorded values can be used as input to a MATLAB routine for the linear least squares solution of gamma in Equation A.4, or stored in a look-up table.

# Gamma Correction using Spectroradiometer

We can alternatively use a spectroradiometer (we use the PR 650 model) to measure the output luminances of a series of input patches [0.1, 0.2, ..., 0.9, 1.0] in order to compute

a look-up table. This tool outputs the CIE XYZ values (see Appendix 3.A in Chapter 3) for a uniform surface patch, where the *Y* component specifies luminance.

Once we obtain XYZ values for the input patches, we can use interpolation to obtain estimates for intermediate luminance values. In an application, we can look-up the output luminance for a given input and apply the correction accordingly.

### REFERENCES

- Robin A Akerstrom and James T Todd. The perception of stereoscopic transparency. *Perception & Psychophysics*, 44(5):421–432, 1988.
- Barton L Anderson and Jonathan Winawer. Image segmentation and lightness perception. *Nature*, 434(7029):79, 2005.
- Christian Boucheny, Georges-Pierre Bonneau, Jacques Droulez, Guillaume Thibault, and Stéphane Ploix. A perceptive evaluation of volume rendering techniques. *ACM Transactions on Applied Perception (TAP)*, 5(4):23, 2009.
- Alessandro Cescatti. Modelling the radiative transfer in discontinuous canopies of asymmetric crowns. i. model structure and algorithms. *Ecological Modelling*, 101:263–274, 1997.
- Isaac Cho, Zachary Wartell, Wenwen Dou, Xiaoyu Wang, and William Ribarsky. Stereo and motion cues effect on depth perception of volumetric data. In *Stereoscopic Displays and Applications XXV*, volume 9011, page 901118. International Society for Optics and Photonics, 2014.
- Per H Christensen. Note 35: Ambient occlusion, image-based illumination, and global illumination. *PhotoRealistic RenderMan Application Notes*, 2002.
- Elisa Colombo and Andrew Derrington. Visual calibration of crt monitors. *Displays*, 22 (3):87–95, 2001.
- Larry Gritz, Tony Apodaca, Matt Pharr, Dan Goldman, Hayden Landis, Guido Quaroni, and Rob Bredow. Renderman in production. *ACM SIGGRAPH 2002 Course Notes*, 2002.

- Kyle Hegeman, Simon Premoze, Michael Ashikhmin, and George Drettakis. Approximate ambient occlusion for trees. *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, March 2006.
- Berthold KP Horn and Robert W Sjoberg. Calculating the reflectance map. *Applied optics*, 18(11):1770–1779, 1979.

Vladimir Kajalin. Screen space ambient occlusion. *Shader X*, 7(413):24, 2009.

- Marta A Kersten, A James Stewart, Niko Troje, and Randy Ellis. Enhancing depth perception in translucent volumes. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):1117–1124, 2006.
- Marta Kersten-Oertel, Sean Jy-Shyang Chen, and D Louis Collins. An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. *Visual-ization and Computer Graphics, IEEE Transactions on*, 20(3):391–403, 2014.
- Jan J Koenderink, Sylvia C Pont, Andrea J Van Doorn, Astrid ML Kappers, and James T Todd. The visual light field. *Perception*, 36(11):1595–1610, 2007.
- Hayden Landis. Production-ready global illumination. *Siggraph course notes*, 16(2002): 11, 2002.
- Michael S Langer and Heinrich H Bülthoff. Depth discrimination from shading under diffuse lighting. *Perception*, 29(6):649–660, 2000.
- Michael S Langer and Fahim Mannan. Visibility in three-dimensional cluttered scenes. *Journal of the Optical Society of America A*, 29(9):1794–1807, 2012.
- Michael S Langer and Steven W Zucker. Shape-from-shading on a cloudy day. *Journal of the Optical Society of America A*, 11(2):467–478, 1994.

Michael S Langer, Haomin Zheng, and Shayan Rezvankhah. Depth discrimination from occlusions in 3d clutter. *Journal of Vision*, 16(11):11–11, 2016.

Fabio Metelli. The perception of transparency. Scientific American, 230(4):90–99, 1974.

- Gavin Miller. Efficient algorithms for local and global accessibility shading. In *Proceed*ings of the 21st annual conference on Computer graphics and interactive techniques, pages 319–326. ACM, 1994.
- Tiit Nilson. A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meteorology*, 8:25–38, 1971.
- William T Reeves and Ricki Blau. Approximate and probabilistic algorithms for shading and rendering structured particle systems. In *ACM Siggraph Computer Graphics*, volume 19, pages 313–322. ACM, 1985.
- Mathias Schott, Vincent Pegoraro, Charles Hansen, Kévin Boulanger, and Kadi Bouatouch. A directional occlusion shading model for interactive direct volume rendering.
  In *Computer Graphics Forum*, volume 28, pages 855–862. Wiley Online Library, 2009.
- Manish Singh and Barton L Anderson. Toward a perceptual theory of transparency. *Psychological review*, 109(3):492, 2002.
- A James Stewart. Vicinity shading for enhanced perception of volumetric data. In *Visualization, 2003. VIS 2003. IEEE*, pages 355–362. IEEE, 2003.
- Inna Tsirlin, Robert S Allison, and Laurie M Wilcox. Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision*, 8(5):5, 2008.
- Raymond van Ee and Barton L Anderson. Motion direction, speed and orientation in binocular matching. *Nature*, 410(6829):690, 2001.

Daniele Zavagno. Some new luminance-gradient effects. *Perception*, 28(7):835–838, 1999.