

The perceptual consequences and neural basis of monocular occlusions

Inna Tsirlin

A dissertation submitted to the Faculty of Graduate Studies
in partial fulfilment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Graduate Program in Psychology
York University
Toronto, Ontario
June, 2013

©Inna Tsirlin

Abstract

Occluded areas are abundant in natural scenes and play an important role in stereopsis. However, due to the treatment of occlusions as noise by early researchers of stereopsis, this field of study has not seen much development until the last two decades. Consequently, many aspects of depth perception from occlusions are not well understood. The goal of this thesis was to study several such aspects in order to advance the current understanding of monocular occlusions and their neural underpinnings. The psychophysical and computational studies described in this thesis have demonstrated that: 1) occlusions play an important role in defining the shape and depth of occluding surfaces, 2) depth signals from monocular occlusions and disparity interact in complex ways, 3) there is a single mechanism underlying depth perception from monocular occlusions and 4) this mechanism is likely to rely on monocular occlusion geometry. A unified theory of depth computation from monocular occlusions and disparity was proposed based on these findings. A biologically-plausible computational model based on this theory produced results close to observer percepts for a variety of monocular occlusion phenomena.

Dedication

I dedicate this work to my beautiful daughter Sophie Nina Zaharescu and my beloved grandmother Nina Vasil'evna Tentser.

Acknowledgements

I would like to thank my supervisors, Professors Laurie Wilcox and Robert Allison, for believing in me and providing me with the greatest support from all standpoints. They made my experience in CVR wonderful, and I would not have been able to complete this work without them. I am very grateful to the rest of my thesis committee: Professors James Elder, Julie Harris, Richard Wildes and Joseph DeSouza, for their encouragement, insightful comments, and hard questions. My family and especially my husband Andrei Zaharescu, have been of tremendous help, continuously supporting and encouraging me to pursue an academic career. Special thanks go to James Elder, Albert Rothenstein and Jenny Read for enlightening discussions and help finding relevant literature. Finally, I would like to acknowledge my friends and colleagues in the laboratory: Bob Hou, Debi Stransky, Aliya Solski, Kazuho Fukuda, Sandra Lasek, Leslie Deas, Arthur Lugtigheid and many others in the Center for Vision Research and outside of it. They made me feel at home and shared my passion for research.

Contents

| | |
|--|------------|
| Abstract | ii |
| Dedication | iii |
| Acknowledgements | iv |
| 1 Introduction | 1 |
| 1.1 Introduction to stereopsis and monocular occlusions | 1 |
| 1.2 Qualitative depth from monocular occlusions | 3 |
| 1.3 Quantitative depth from monocular occlusions | 8 |
| 1.3.1 Two-object arrangements | 8 |
| 1.3.2 Illusory occluders | 10 |
| 1.3.3 Temporal monocular occlusions | 11 |
| 1.3.4 Monocular gap stimuli | 12 |
| 1.4 Interaction of monocular and binocular features | 14 |
| 1.5 Temporal properties of depth acquisition from monocular occlusions . | 15 |
| 1.6 Seeing monocular zones from the cyclopean eye | 18 |

| | | |
|----------|--|-----------|
| 1.7 | Computational models | 20 |
| 1.8 | Objectives | 21 |
| 2 | The role of monocular occlusions in stereopsis | 26 |
| 2.1 | Objectives | 26 |
| 2.2 | Experiment 2.1 - Monocular occlusions define the shape of occluding surfaces | 28 |
| 2.2.1 | Methods | 28 |
| 2.2.2 | Results | 31 |
| 2.3 | Experiment 2.2 - Quantitative depth from monocular occlusions | 32 |
| 2.3.1 | Methods | 34 |
| 2.3.2 | Results and Discussion | 35 |
| 2.4 | Experiment 2.3 - Controlling for binocular matching | 36 |
| 2.4.1 | Methods | 38 |
| 2.4.2 | Results and Discussion | 38 |
| 2.5 | Experiment 2.4 - Disparity biasing | 42 |
| 2.5.1 | Methods | 43 |
| 2.5.2 | Results and Discussion | 44 |
| 2.6 | General Discussion | 45 |
| 3 | Interaction of monocular occlusions and disparity | 47 |
| 3.1 | Objectives | 47 |
| 3.2 | Experiment 3.1 - random-dot frame stimulus | 48 |

| | | |
|----------|--|-----------|
| 3.2.1 | Methods | 48 |
| 3.2.2 | Results and Discussion | 50 |
| 3.3 | Experiment 3.2 - Disparity bias in the stimulus of Gillam and Nakayama (1999) | 52 |
| 3.3.1 | Methods | 53 |
| 3.3.2 | Results and Discussion | 58 |
| 3.4 | Experiment 3.3 - Disparity bias in the stimulus of Liu et al. (1994) | 61 |
| 3.4.1 | Methods | 62 |
| 3.4.2 | Results and Discussion | 64 |
| 3.5 | Experiment 3.4 - Control | 67 |
| 3.5.1 | Methods | 68 |
| 3.6 | General Discussion | 70 |
| 4 | Quantitative depth from monocular occlusions - one or many mechanisms? | 74 |
| 4.1 | Objectives | 74 |
| 4.2 | Experiment 4.1 - Replicating Nakayama and Shimojo (1990) and Gillam et al. (2003) | 77 |
| 4.2.1 | Methods | 78 |
| 4.2.2 | Control Experiments | 82 |
| 4.2.3 | Results and Discussion | 83 |
| 4.3 | Experiment 2.4 | 89 |
| 4.3.1 | Methods | 91 |
| 4.3.2 | Results and Discussion | 91 |

| | | |
|----------|--|------------|
| 4.4 | Computational Simulations | 93 |
| 4.4.1 | Methods | 94 |
| 4.4.2 | Results and Discussion | 97 |
| 4.5 | General Discussion | 99 |
| 4.5.1 | The source of quantitative depth in two-object arrangements | 99 |
| 4.5.2 | Why does camouflage not work? | 103 |
| 5 | The role of disparity in depth from monocular occlusions | 108 |
| 5.1 | Objectives | 108 |
| 5.2 | Methods | 109 |
| 5.3 | Results | 111 |
| 6 | A biologically-plausible computational model of depth from monocular occlusions and binocular disparity | 119 |
| 6.1 | Objectives | 119 |
| 6.2 | Biological plausibility | 123 |
| 6.3 | Model principles | 125 |
| 6.4 | Implementation details | 126 |
| 6.5 | Model formulation | 127 |
| 6.5.1 | Initial disparity computation - Energy neurons and inter-neural connections | 127 |
| 6.5.2 | Monocular occlusions detection | 131 |
| 6.5.3 | Computing the width of monocular areas | 138 |
| 6.5.4 | Other constraints | 139 |
| 6.5.5 | Edge detection | 141 |

| | | |
|----------|---|------------|
| 6.5.6 | 3D surface cells - final computation of disparity | 142 |
| 6.5.7 | Final selection of disparity | 148 |
| 6.6 | Model evaluation | 148 |
| 6.7 | Results | 152 |
| 6.8 | Discussion | 161 |
| 6.8.1 | Model features | 161 |
| 6.8.2 | Limitations and future improvements | 162 |
| 6.8.3 | Model predictions | 164 |
| 7 | Conclusions | 167 |
| 7.1 | Overview and contributions | 167 |
| 7.2 | Outstanding issues and implications | 172 |
| 7.3 | Conclusions | 175 |
| A | Symbols and acronyms used in the thesis | 176 |
| B | Auxiliary data | 179 |
| | Bibliography | 179 |

List of Tables

| | | |
|-----|--|-----|
| 6.1 | Biologically-inspired models of stereopsis with specific treatment of monocular occlusions. | 122 |
| 6.2 | Stimuli used to test the model. | 149 |
| 6.3 | Model parameters used in testing. | 152 |
| A.1 | Acronyms used in the thesis. | 176 |
| A.2 | Symbols used in Chapter 6. | 176 |

List of Figures

| | | |
|------|--|----|
| 1.1 | Geometry of stereopsis | 2 |
| 1.2 | Monocular occlusion geometry | 4 |
| 1.3 | Computing depth from disparity and monocular occlusions | 5 |
| 1.4 | Qualitative depth from monocular occlusions | 7 |
| 1.5 | Interaction of monocular and binocular features | 13 |
| 1.6 | The cyclopean eye and the monocular zones paradox | 18 |
| 2.1 | Stimuli used in Experiments 2.1 and 2.2 | 27 |
| 2.2 | Results of Experiment 2.1 for all observers | 31 |
| 2.3 | The viewing geometry of the stimuli in Experiments 2.1 and 2.3 | 33 |
| 2.4 | Results of Experiment 2 for all observers except for SL | 34 |
| 2.5 | An example of the stimuli with 2.7% density | 38 |
| 2.6 | Stimuli used in Experiment 2.3 | 39 |
| 2.7 | Results of the qualitative task used in Experiment 2.3 | 40 |
| 2.8 | Results of the disparity-matching task of Experiment 2.3 | 41 |
| 2.9 | Stimuli used in Experiment 2.4 | 42 |
| 2.10 | Results of Experiment 2.4 | 44 |

| | | |
|------|---|-----|
| 3.1 | Stimuli used in Experiments 3.1 | 50 |
| 3.2 | Results of Experiment 3.1 | 52 |
| 3.3 | Stimuli used in Experiment 3.2 | 55 |
| 3.4 | Results of parts 1 and 2 of Experiment 3.2 | 58 |
| 3.5 | Individual results of Experiment 3.2 | 61 |
| 3.6 | Stimuli used in Experiment 3.3 | 65 |
| 3.7 | Results of parts 1 and 2 of Experiment 2.3 | 66 |
| 3.8 | Individual results of Experiment 3.3 | 67 |
| 3.9 | Stimuli used in Experiment 4 | 68 |
| 3.10 | Results of Experiment 3.4 | 70 |
| 4.1 | Stimuli used in Nakayama and Shimojo (1990) | 75 |
| 4.2 | Stimuli used in Experiments 4.1 | 78 |
| 4.3 | Timeline of Control Experiments 4.1-A and 4.1-B. | 82 |
| 4.4 | Results of Experiment 4.1 | 84 |
| 4.5 | Results of control Experiment 4.1-A | 87 |
| 4.6 | Stimuli used in Experiments 4.2 | 90 |
| 4.7 | Results of Experiment 2 | 93 |
| 4.8 | Results of the computational simulations | 96 |
| 4.9 | Perception of camouflage | 104 |
| 5.1 | Stimuli used in the computational simulations | 110 |
| 5.2 | Simulation results for the RDS | 112 |
| 5.3 | Simulation results for the random-dot frame stimulus | 113 |
| 5.4 | Simulation results for the Gillam & Nakayama stimulus | 113 |

| | | |
|------|--|-----|
| 5.5 | Simulation results for the Liu et al. stimulus | 114 |
| 5.6 | Simulation results for a two-object arrangement with a bar | 114 |
| 5.7 | Simulation results for a two-object arrangement with a disc | 115 |
| 5.8 | Simulation results for the monocular gap stimulus | 115 |
| 5.9 | Simulation results for the monocular intrusion stimulus | 116 |
| 5.10 | Simulation results for different occlusion widths | 117 |
| 6.1 | Model overview | 128 |
| 6.2 | Disparity profile | 131 |
| 6.3 | Left-to-right check example | 135 |
| 6.4 | Biologically-plausible implementation of the left-to-right check | 136 |
| 6.5 | Detecting the width of monocular occlusions | 139 |
| 6.6 | Computing reliability | 140 |
| 6.7 | Using occlusion geometry to assign disparity in occluded regions | 145 |
| 6.8 | Computing and propagating disparity in monocular regions | 147 |
| 6.9 | Model results for the map image | 154 |
| 6.10 | Model results for the RDS | 154 |
| 6.11 | Model results for the random-dot frame stimulus | 155 |
| 6.12 | Model results for the Gillam & Nakayama stimulus | 156 |
| 6.13 | Model results for the Liu et al. stimulus | 156 |
| 6.14 | Model results for a two-object arrangement with a bar | 157 |
| 6.15 | Model results for a two-object arrangement with a disc | 158 |
| 6.16 | Model results for the monocular gap stimulus | 159 |
| 6.17 | Model results for the monocular intrusion stimulus | 160 |

B.1 Results of the disparity-matching task of Experiment 2.3 including
observer SL 179

Chapter 1

Introduction

1.1 Introduction to stereopsis and monocular occlusions

Light reflected from objects in the world creates images of these objects on the retinae of the two eyes. These retinal images are not identical; the interocular separation yields a difference in position of the imaged objects. Consider the situation depicted in Figure 1.1. The eyes are fixated on the cube and its image falls on corresponding locations on the retinae. The image of the sphere, located closer to the observer, falls on different locations on the two retinae. This difference is referred to as positional disparity. If corresponding points on the two retinae can be found, positional disparity can be extracted and the relative depth order of objects can be determined with simple geometry (see Figure 1.3-B). Other types of binocular disparity exist, such as size disparity and vertical disparity. In principle, any difference in the images of the two eyes can be referred to as disparity. In this thesis, the terms ‘disparity’ or

‘binocular disparity’ always refer to horizontal positional disparity, unless specified otherwise.

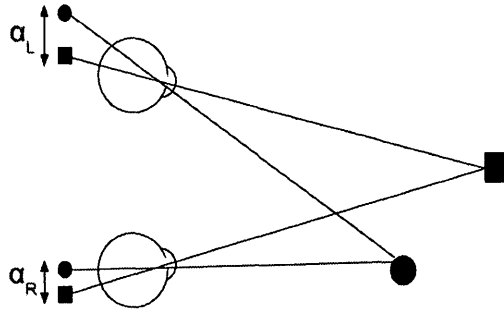


Figure 1.1: Geometry of stereopsis. The eyes are fixated on the cube such that there is no disparity between the positions of the images of the cube in the two eyes. The images of the sphere fall on different locations in the retinae. The distances of these images from the fovea, the central region of the retina, are equal to α_L and α_R in the left and the right eye respectively. The difference in the distances is the relative disparity of the circle.

The difference in the vantage points of the two eyes does not only give rise to positional disparity, but also creates monocular areas visible to one eye only as shown in Figure 1.2-A. These areas arise due to physical occlusion of objects by other objects in the scene and thus referred to as ‘monocular occlusions’ or ‘binocular half-occlusions’ in the computational literature (other monocular areas within the visual field exist in the location of the contralateral optic disc and outside of the region of binocular overlap). Importantly, monocularly occluded areas have no physical correspondence in the image of the other eye.

Leonardo da Vinci was one of the first scientists to discuss the phenomenon of monocular occlusions (Da Vinci, 1877). Closer to our times, Adolf von Szily demonstrated the role of monocular occlusions and disparity in defining cyclopean shapes in silhouette figures (Ehrenstein and Gillam, 1998, English translation). However, it

was not until about 20 years ago that research into monocular occlusions has started to develop. This delay was most likely related to the focus of stereoscopic research on solving the correspondence problem, or how the matching points in the images of the two eyes are found. This problem was highlighted by the popularization of random dot stereograms (RDS¹) by Julesz (1960). In RDS, depth is easily perceived although there are no visible objects and there is an abundance of potential false matches. Early computational models of stereopsis were directed at solving the issue of matching binocular images and thus ignored monocular areas in the input images or treated them as noise (Marr and Poggio, 1976, 1979).

1.2 Qualitative depth from monocular occlusions

In one of the earliest demonstrations of depth from monocular occlusions, Adolf von Szily showed that monocular features in silhouette figures could induce depth percepts in fused images and that the depth order of surfaces depended on the eye of origin of the monocular occlusions (Ehrenstein and Gillam, 1998, English translation)².

The first experimental evidence for the effect of monocular occlusions on perceived depth came from studies that did not examine monocular occlusions directly, but used stimuli containing monocular regions. Lawson and Gulick (1967) studied the mechanisms of subjective contour formation from stereopsis. They compared depth estimation in RDS with zero disparity and monocular occlusions (they did not identify them as such, but treated this phenomenon as a variant of size disparity) to depth estimation in RDS where depth was based only on binocular disparity. They found

¹All acronyms used in this thesis are listed in Table A.1 in Appendix A.

²An alternative explanation to the origin of depth percepts in these figures is discussed in Section 4.5.2.

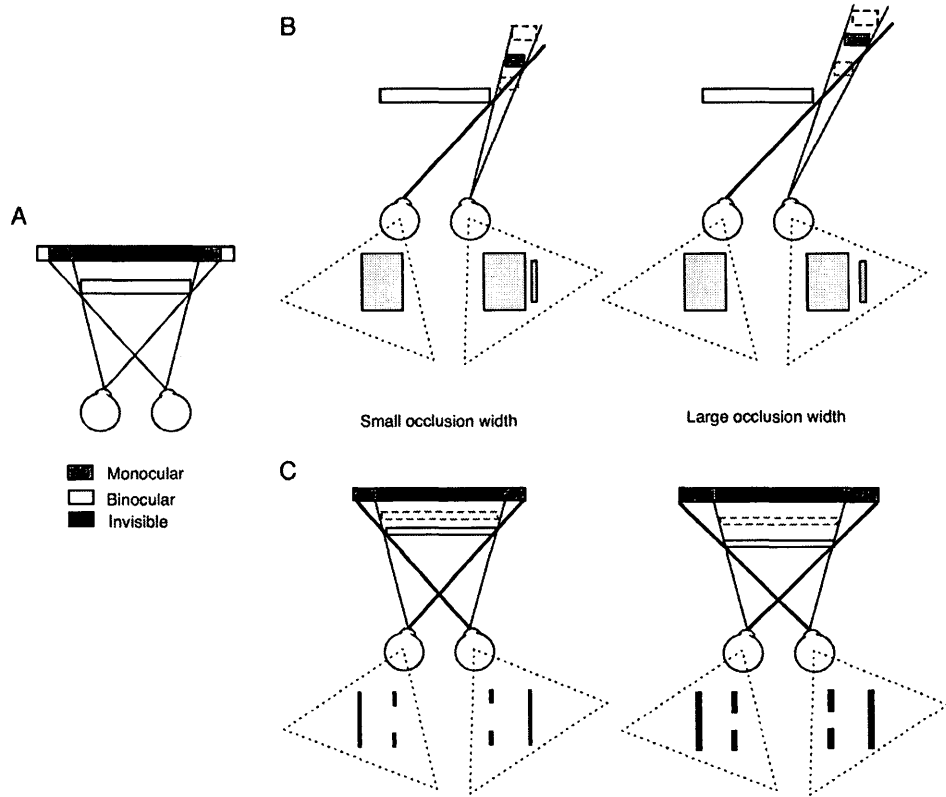


Figure 1.2: Monocular occlusion geometry. (A) A foreground surface occludes regions of the background in each eye on the contralateral side. (B) In this two-object arrangement (Nakayama and Shimojo, 1990), a larger surface (rectangle) occludes a stand alone smaller object (bar). The line of sight from the left eye (bold black line) that does not see the bar, constraints the minimum possible depth of the occluded object. It cannot be located closer to the occluder (red dashed outline) since it will be seen by the left eye. It could be positioned farther without violating viewing geometry (black dashed outline). The magnitude of the minimum possible depth is determined by the lateral separation between the object and the occluder in the eye that sees the occlusion. The larger the separation, the larger is the minimum possible depth between the two objects as shown in the right-hand schematic. (C) Similar geometric rules also apply to illusory occluder stimuli, for example that of Gillam and Nakayama (1999). The minimum possible depth of the illusory occluder on each side is constrained by the lines of sight from the contralateral eyes. Larger occluded regions yield larger minimum possible depth between the occluded region and the illusory occluder.

that the perceived configuration of the stimulus was different but the magnitude of perceived depth was comparable. Similarly, Kaufman (1965) conducted experiments

with random-letter stereograms composed of binocular and monocular regions. He observed that the presence of monocular areas created a perception of depth in the absence of disparity.

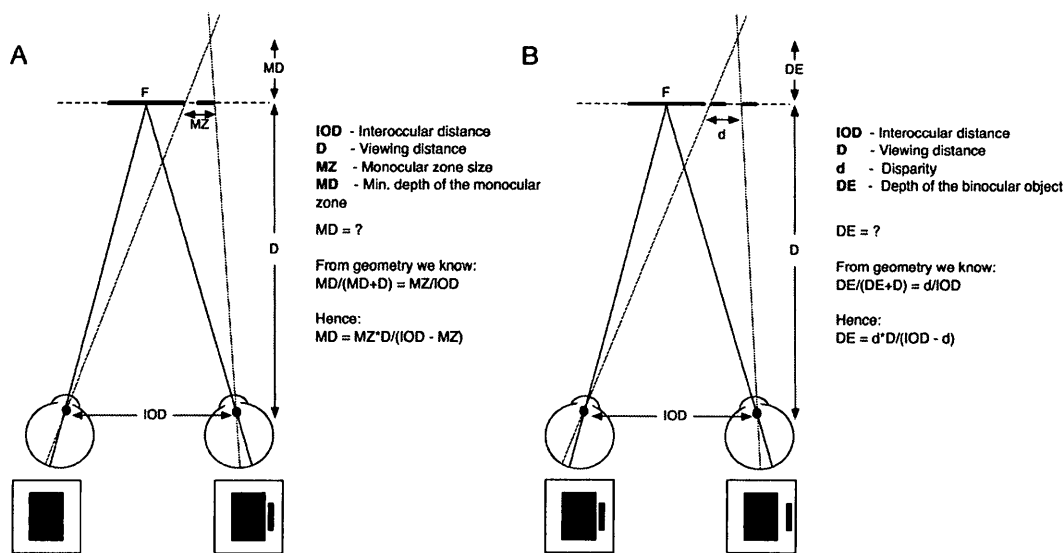


Figure 1.3: Computing depth from disparity and monocular occlusions. (A) shows how the magnitude of perceived depth can be calculated from the (on screen) size of a monocular region. (B) shows how the magnitude of perceived depth can be calculated from (on screen) disparity. The calculations are very similar and reflect the tight coupling between the two depth cues. The rectangular frames on the bottom of the image show the scene that is visible to the two eyes.

Nakayama and Shimojo (1990) showed that monocular objects consistent with the geometry of occlusion were perceived as farther in depth than those that did not comply with occlusion geometry (see Figure 1.4-A). They also demonstrated that subjective contours in depth can arise from monocular occlusions alone, even in the case where there are no binocular elements defining the occluder. In a nod to Leonardo da Vinci's original observations, and to distinguish it from conventional stereopsis,

Nakayama and Shimojo (1990) grouped these and other monocular occlusion phenomena under the name ‘da Vinci stereopsis’. Accordingly, in this thesis, the term ‘da Vinci stereopsis’ will be used to describe any phenomena where depth is perceived on the basis of monocular occlusions with no involvement of disparity.

Another early study showed that horizontal disparity was not necessary to perceive depth in the presence of monocular regions (Anderson, 1994). The stimulus consisted of vertical lines with small binocular dots positioned along them (see Figure 1.4-F). The dots and the lines had zero disparity³, while the lines differed in length in the two eyes. Observers consistently perceived the lines and the dots as positioned on a frontoparallel plane, behind a diamond-shaped aperture formed in the white background. This percept could not have been based on the vertical disparity of the line endings, since in this case the lines would have appeared slanted.

In the last 20 years, many other studies have demonstrated qualitative depth from monocular occlusions. It has been shown that monocular occlusions can induce percepts of illusory occluding contours (Cook and Gillam, 2004; Gillam and Grove, 2004; Gillam and Nakayama, 1999) (see Figure 1.4-B,C), perceived depth between two objects (Hakkinen and Nyman, 1997; Nakayama and Shimojo, 1990) (see Figure 1.4-A) and other depth phenomena (Forte et al., 2002; Pianta and Gillam, 2003b; Sachtler and Gillam, 2007) (see Figure 1.4-D-F). These studies will be discussed in detail in the following sections.

³In this thesis, the amount of disparity is always specified with respect to the intended plane of fixation in the respective experiments unless specified otherwise.

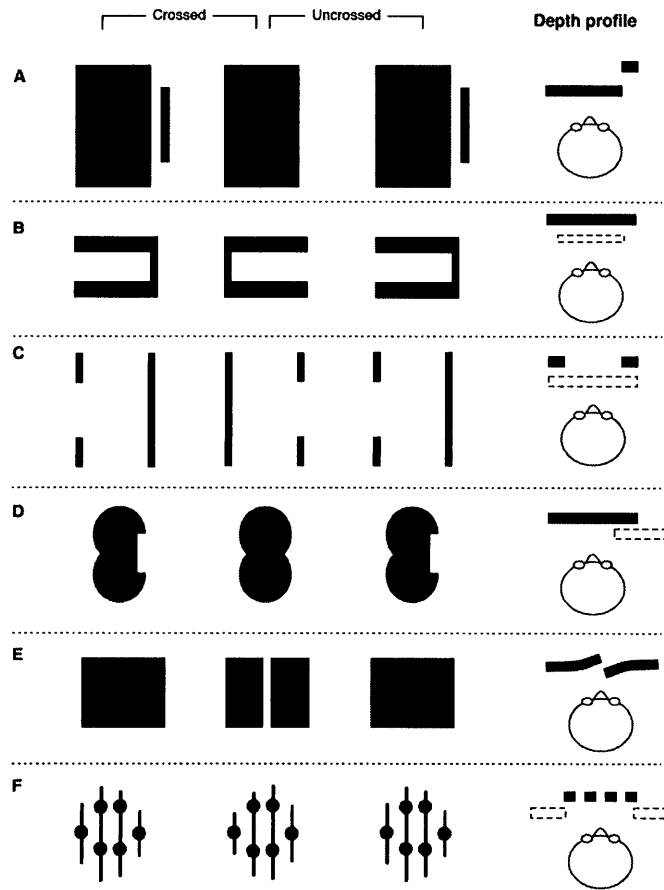


Figure 1.4: Qualitative depth from monocular occlusions. (A) The monocular bar is perceived to lie beyond the binocular rectangle (Nakayama and Shimojo, 1990). (B) An illusory white rectangle appears in front of a larger black rectangle (Liu et al., 1994). (C) An illusory rectangular surface appears in front of the black lines (Gillam and Nakayama, 1999). (D) An illusory intrusion appears in front of the black figure eight (Cook and Gillam, 2004). (E) The monocular gap creates a percept of two surfaces bending in depth away from each other (Gillam et al., 1999). (F) Stimuli with vertical monocular occlusions invoke the perception of an illusory diamond-shaped aperture (Anderson, 1994). Here and in all subsequent figures, the left and the central half-images are arranged for crossed-fusion and the central and the right half-images for divergent fusion.

1.3 Quantitative depth from monocular occlusions

As outlined above, there is strong evidence that monocular occlusions can induce a percept of depth. But can this percept be quantified? The magnitude of perceived depth in binocular features depends on disparity magnitude. With monocular regions, in most cases there is no viable match in the other eye, so binocular disparity cannot be computed. However, the depth of monocular regions and features could potentially be deduced from occlusion geometry. As shown in Figure 1.2, the minimum possible relative depth between an occluded object or region and the occluder is constrained by the line of sight from the eye that does not see the occluded region. This constraint yields a linear relationship between the minimum relative depth of an occluded region and its width. As the width of the occluded region increases, the minimum relative depth increases as well. The visual system could use this relationship, referred to as the ‘minimum depth constraint’, to establish the relative depth between the occluder and the occluded region or object (see Figure 1.3-A). Note, however, that in the case shown in Figure 1.2-B,C (and in the majority of other cases) the maximum possible depth is not constrained. Thus the estimates the visual system might make, based only on the minimum depth constraint, are likely not to be as reliable as those based on binocular disparity.

1.3.1 Two-object arrangements

Nakayama and Shimojo (1990) provided the first demonstration of quantitative depth from occlusion geometry. In their stimulus, shown in Figure 1.4-A, a binocular rectangle was presented with zero disparity. In one half-image a thin monocular bar

was positioned next to the rectangle. When the bar was positioned to the right of the rectangle in the right eye, the configuration was consistent with it being located behind the rectangle in depth and occluded by the rectangle in the left eye (Figure 1.4-A). Nakayama and Shimojo (1990) found that for a limited range of offsets, the perceived depth of the monocular bar increased as its distance from the binocular rectangle (i.e. occluded region width) increased. A similar result was obtained in two-object stimuli by Hakkinen and Nyman (1996). No quantitative depth was found by Nakayama and Shimojo (1990) in two-object arrangements that were consistent with camouflage geometry (this issue is discussed in detail in Chapter 4).

Gillam et al. (2003) proposed that quantitative depth in the stimuli of Nakayama and Shimojo (1990) is simply an instance of double-matching, in which the monocular bar in the right eye is matched to the edge of the occluding rectangle in the left eye (which is simultaneously matched to the rectangle edge in the right eye). Double-matching in these stimuli predicts the same perceived depth as the minimum depth constraint. Gillam et al. (2003) tested this hypothesis by replacing the bar in the Nakayama and Shimojo (1990) stimulus with a small disc. They argued that the disc was unlikely to be matched to the edge of the rectangular occluder because of the differences in size and shape. They found that the monocular disc stimulus did not produce quantitative depth percepts, a result that seemed to support the double-matching hypothesis. However, Gillam et al. (2003) used a different range of occlusion widths for the disc stimulus than that for the line stimulus, which could have affected their results as discussed in Chapter 4.

1.3.2 Illusory occluders

In one of the first demonstrations of the illusory occluder phenomenon, Liu et al. (1994) presented a stimulus consisting of a black rectangular bracket (\sqsubset) on a white background (see Figure 1.4-B). The gap of the bracket pointed to the left in the left half-image and to the right in the right half-image, such that the vertical contour of the bracket in one eye had no match in the other eye. During stereoscopic viewing, the stimulus was perceived as an illusory white rectangle floating in front of a black rectangle. The authors argued that this percept occurred since the vertical parts of the brackets were interpreted as occluded and an illusory occluder was constructed to account for the monocular occlusions. Liu et al. (1994) found that perceived depth magnitude of the illusory occluder was proportional to the width of the vertical part of the bracket in accordance with the minimum depth constraint. However, it was subsequently suggested that quantitative depth in this stimulus could have arisen due to binocular matching (Gillam, 1995; Liu et al., 1995, 1997).

Gillam and Nakayama (1999) modified the Liu et al. (1994) stimulus by removing the horizontal lines and corners and leaving only vertical lines, which precluded any possibility for binocular matching. When fused, their stimulus was perceived as an illusory rectangular surface hovering in front of the black lines (see Figure 1.4-C). The amount of depth perceived in this stimulus increased with the increase in the thickness of the lines (i.e. the width of the monocular zone), although it was somewhat greater than the depth predicted by the minimum depth constraint (Gillam and Nakayama, 1999; Mitsudo et al., 2006).

Cook and Gillam (2004) showed that quantitative depth could also be perceived from monocular occlusions when the monocular feature was in the form of a back-

ground intrusion (see Figure 1.4-D). When viewed stereoscopically, the intrusion was seen as an illusory surface occluding part of the background surface in depth. The amount of depth perceived in the illusory occluder depended on the width of the intrusion. However, in this case depth could also have been based on simple binocular matching as discussed in Chapter 5.

Horizontal size disparity is normally interpreted by the visual system as surface slant. However, difference in size could also arise from occlusion of a part of the object in one eye by an illusory occluder camouflaged against the background. In several publications, Grove and Gillam (Gillam and Grove, 2004; Grove et al., 2005; Grove and Gillam, 2007) showed that when a series of horizontal lines of different length aligned on one side were shortened by the same amount in one eye, an illusory occluder could be perceived on the aligned side of the lines. Moreover, when the degree of shortening of the lines varied gradually, the illusory occluder was perceived as slanted in depth indicating the presence of quantitative depth based on occlusion geometry.

1.3.3 Temporal monocular occlusions

A related phenomenon of quantitative depth from monocular features based on occlusion geometry has also been found in temporal monocular occlusions. Temporal occlusions arise when objects move behind other objects or apertures.

Shimojo et al. (1988) demonstrated the perception of quantitative depth in a monocular bar moving behind a binocular aperture. The bar was present only in one eye at a time with different interocular delays. According to occlusion geometry, a bar positioned further in depth from the aperture will have a larger interocular

delay than a bar positioned closer. Shimojo et al. (1988) found that in their stimuli perceived depth increased as the interocular delay increased from 0 to around 80 ms.

Brooks and Gillam (2006) suggested that in the stimuli of Shimojo et al. (1988) depth could have been perceived by matching the bar in one eye to the bar in other eye (although if this were true, the disparity would have been the same for all interocular delays - equal to the size of the aperture). They conducted their own study where they controlled for the possibility of such matching and confirmed that bars moving behind a camouflaged occluder can elicit the percept of quantitative depth. In other words, as the quasi-disparity (the disparity of the occluder resulting from the combination of the bar's speed and interoccluar delay) increased so did the perceived depth of the occluder. In control experiments, Brooks and Gillam (2006) showed that this phenomenon cannot be accounted for by integration of disparity over time or using temporal disparity. The authors concluded that quantitative depth in their stimuli was perceived solely on the base of monocular occlusions.

1.3.4 Monocular gap stimuli

Gillam et al. (1999) presented another type of stimulus where quantitative depth was perceived without horizontal disparity. In their stimulus, one half-image contained a black rectangle, while the other a monocular gap in the center of the rectangle (see Figure 1.4-E). When viewed stereoscopically the stimulus appeared as two surfaces with their inner edges bending away and towards the observer. The magnitude of depth between the inner edges depended on the width of the monocular gap. This phenomenon was partially preserved even when the outer edges of the rectangles had zero disparity (Pianta and Gillam, 2003a). In a later study using the same

stimuli, Pianta and Gillam (2003b) showed that the monocular gap stimuli produced quantitative depth similar to that produced by matched stimuli with conventional disparity.

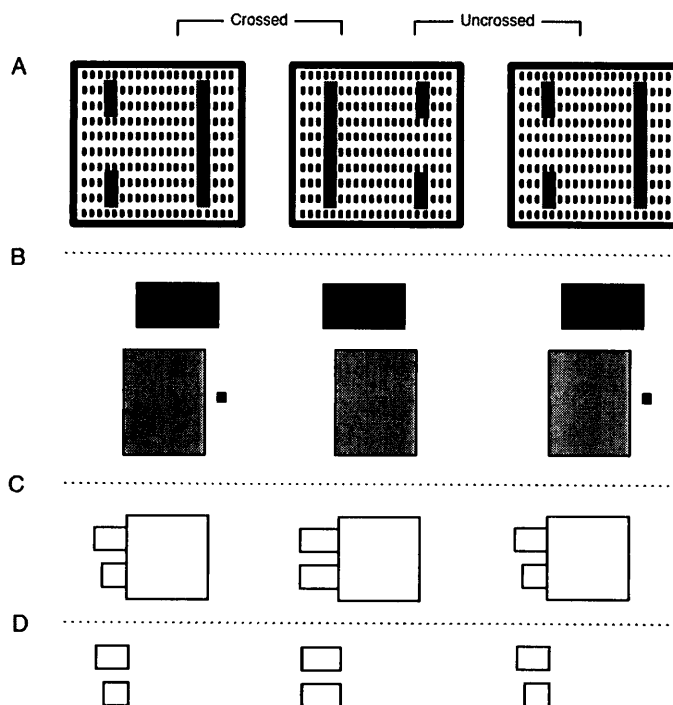


Figure 1.5: Interaction of monocular and binocular features. (A) The depth of an illusory occluder captures a wallpaper pattern (Hakkinen and Nyman, 2001) (B) The perceived depth of a monocular dot can be biased by a proximate binocular surface in a two-object arrangement (Hakkinen and Nyman, 1996). (C)+(D) Size disparity does not result in perceived surface slant when an occlusion interpretation is available (Hakkinen and Nyman, 1997).

1.4 Interaction of monocular and binocular features

Depth within monocularly occluded regions seems to interact in complex ways with depth in binocular regions. Anderson and Nakayama (1994) showed that contrast information available at the boundaries of monocular and binocular areas of a stereogram determined the depth in ambiguous stimuli. They proposed that such behaviour could be produced by hypothetical detectors, which respond to edges between binocular and monocular areas.

Hakkinen and Nyman (2001) also demonstrated that monocular occlusions could affect the perception of depth in ambiguous binocular patterns. In their stimuli, they combined modified Kanizsa figures with repetitive texture patterns, which could be matched in several ways. The Kanizsa figures were modified to match the stimuli used by Liu et al. (1994), such that parts of the figures were occluded and no conventional binocular disparity was present (see Section 1.3.2-A). They showed that in these configurations the central part of the texture was still captured due to the presence of monocular occlusions. To completely exclude the possibility of binocular matching, they used the vertical line stereograms introduced by Gillam and Nakayama (1999) instead of the Kanizsa figures (see Figure 1.5) and showed that capture was still observed.

In an earlier study, Hakkinen and Nyman (1996) examined the effect of an inducing surface on depth localization of monocular features. They presented the observers with a two-object arrangement: an occluding rectangular surface, a monocular dot and an inducing surface above the occluder (see Figure 1.5-B). They showed that

when the inducing surface had large uncrossed disparities with respect to the occluder, the perceived depth of the occluded dot was larger than when the inducing surface had a crossed disparity with respect to the occluder. This is consistent with the unidirectional restrictions imposed by occlusion geometry (see Section 1.3).

Subsequently, Hakkinen and Nyman (1997) demonstrated that the perception of stereoscopic slant from images with different widths in the two eyes, was diminished when an occlusion interpretation was possible (see Figure 1.5-C). They also showed that the borders of the occluder and the occluding surface did not need to be adjacent for the phenomenon to occur as long the occlusion interpretation was possible.

1.5 Temporal properties of depth acquisition from monocular occlusions

Several studies have examined the temporal characteristics of depth from monocular occlusions, aiming to relate them to the stages of processing of binocular disparity. Gillam and Borsting (1988) compared the latency for stereoscopic depth perception of RDS with monocular regions filled with texture matching that of the binocular regions to RDS with blank monocular regions. They found that, for larger RDS disparities (which yield larger monocular areas), it took observers significantly longer to perceive depth from disparity in stimuli with blank monocular areas in comparison to those with textured monocular areas. Similarly, Grove and Ono (1999) found that significantly more time was required to perceive stereograms with occluded zones textured differently from the surrounding binocular regions than when they had similar texture.

Wilcox and Lakra (2007) showed that the complete removal of monocular occlusions in photographs of self-occluding objects increased latencies for depth perception. The size of the monocular zone that was removed correlated with the time it took to perceive depth in the stimuli. Taken together, these findings suggest that monocular occlusions are processed at least as early as binocular disparity and are used to compute disparity-based depth.

Mitsudo et al. (2005) examined whether the overestimation of depth in the illusory occluder stimuli of Gillam and Nakayama (1999) was coded at the early stages of depth processing from stereopsis or was assigned at later stages due to configurational considerations. They used a visual search task with disparity noise and showed that the overestimation was present already at the early stages of visual processing suggesting monocular occlusions were processed along with disparity.

Shimojo and Nakayama (1990) used stimuli with both geometrically plausible and implausible (monocular region to the left of the occluder in the right eye) occluded monocular regions. The binocular and monocular regions in their stereograms were colored differently and the observers had to indicate when the monocular areas were suppressed and blended into the background. They found that the monocular areas that complied with occlusion geometry were seen clearly at the depth of the background, while the inconsistent monocular areas were suppressed or seen at the depth of the front surface. The authors proposed that since detection of geometrically consistent monocular occlusions requires eye-of-origin information, depth from occlusions must be computed at a very early stage of the visual cortex (V1), where the input of the two eyes combine but are still distinguishable (see also Nakayama (1996)).

Forte et al. (2002) came to a different conclusion. They used displays in which spatially filtered noise strips were seen beyond binocularly defined bars. The noise was monocular and no binocular correspondence could be established between the two eyes. Forte et al. (2002) found that as the noise orientation changed from horizontal to vertical (becoming discontinuous from strip to strip) it became harder for the observers to perceive the background in depth. They proposed that a higher level process is involved in computing depth from monocular occlusions, which establishes continuity of the monocular pieces in the two eyes. However, they did not consider that initial depth estimates from monocular occlusions could have been made as early as those from disparity and altered later by the high-level process.

Kuroki and Nakamizo (2006) found that depth in phantom line stereograms (Gillam and Nakayama, 1999), gap stimuli (Gillam et al., 1999) and RDS was scaled with vergence angle and viewing distance. They proposed that the similarity of this effect to the scaling of depth from disparity is suggestive of a common mechanism.

Sachtler and Gillam (2007) examined depth discrimination as a function of stimulus duration using stimuli with disparity-based and occlusion-based depth percepts. They found that the degradation in performance as stimulus duration was reduced was very similar for both types of stimuli, suggesting that both types of depth are established at an early stage of visual processing. However, since Sachtler and Gillam (2007) did not use a post-stimulus mask, the visual processing time was not adequately limited.

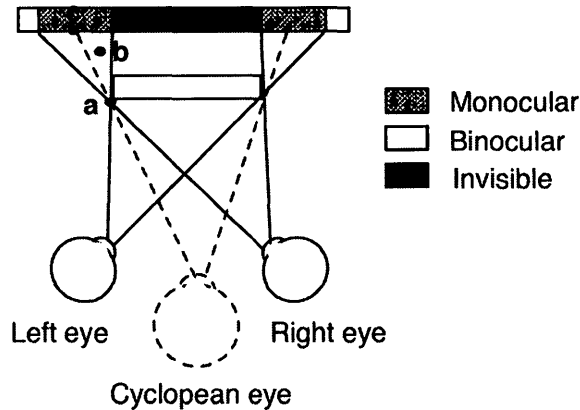


Figure 1.6: The cyclopean eye and the monocular zones paradox. The background surface is partially occluded by the foreground surface. Point 'b' is visible to the left eye, but should not be visible to a physical eye located at the position of the cyclopean eye due to the occluder. Point 'b' needs to be to the left of 'a' to be visible to the cyclopean eye. However, we see 'b' to the right of 'a', which means that 'a' and 'c' are not in the same visual direction. Adapted from Erkelens et al. (1996).

1.6 Seeing monocular zones from the cyclopean eye

It has been suggested that during binocular viewing we view the world as if seen from a central point located midway between the two eyes referred to as the 'cyclopean eye' (Erkelens et al., 1996; Ono et al., 2002). Since all areas and objects seen by each eye are included in the cyclopean view, monocular occlusions should also be visible to the cyclopean eye. However, this creates a problem because, as shown in Figure 1.6, from the location of the cyclopean eye, the edge of the occluding object in one eye and the monocular area in the other eye occupy the same visual direction. This means

that both cannot be seen by the cyclopean eye at the same time (Erkelens et al., 1996). Erkelens et al. (1996) examined the perceived visual direction of an occluder by aligning lines located at the depth of the occluded background to the perceived edge of the occluder. Observers consistently saw the left edge of the occluder to the right of its cyclopean direction and the right edge of the occluder to the left of its cyclopean direction. Based on these results and a few demonstrations Erkelens et al. (1996) concluded that near monocular areas the scene is perceived as seen from the appropriate physical eye (left or right) and not from the cyclopean eye.

Ono et al. (2003) provided a different interpretation to these findings. They suggested that the scene is still viewed from the cyclopean eye, however, there is an expansion and compression of nonfixated surfaces near depth discontinuities to allow for all seen objects and areas to be represented in the cyclopean view. Both proposals predict distortions in perceived shape of nonfixated occluding or occluded surfaces due to the shift in the visual direction of the edges of occluding surfaces and occluded regions. van Ee and Erkelens (2000) tested this hypothesis by asking observers to judge the aspect ratio of a rectangle either seen on its own or occluding another surface. The perceived aspect ratio did not depend on the presence of an occluded surface or on the depth of fixation and there was no distortion in the perceived shape of the occluded region. The authors concluded that the mechanism that determines perceived aspect ratio is dissociated from the mechanism that determines perceived direction. Ono et al. (2003) suggested that the shape distortion resulting from the shifts in visual direction is corrected by higher-level shape and direction processes.

1.7 Computational models

In the last two decades, the importance of the explicit detection of monocularly occluded areas has been stressed in the computer vision literature (for review see Egnal and Wildes (2002)). Many computational algorithms of stereopsis explicitly recover monocularly occluded regions (referred to as ‘binocular half-occlusions’ in the computer vision literature) to improve computed disparity maps (Lin and Tomasi, 2004; Min and Sohn, 2008; Sizintsev and Wildes, 2007; Sun et al., 2005; Zitnick and Kanade, 2000). Biologically-inspired models of stereopsis have recently started to address this issue as well (Assee and Qian, 2007; Cao and Grossberg, 2005; Hayashi et al., 2004; Watanabe and Fukushima, 1999).

Watanabe and Fukushima (1999) proposed a two-stage neural network where monocular occlusions are treated as a depth cue along with disparity. First, luminance edges are extracted and classified as binocular or monocular. The second stage in the network has two different layers that process information concurrently and interact over time. One layer reconstructs 3-D surfaces and the other discriminates monocular occlusions. Both processes are governed by several constraints including the smoothness and uniqueness constraints common to many stereopsis algorithms and several constraints arising from occlusion geometry.

Hayashi et al. (2004) modified the Watanabe and Fukushima (1999) model by replacing the edge detectors with energy neurons (a biologically-plausible model of disparity detectors) and adding several extra constraints. The smoothness constraint for the 3-D surfaces model was modified to include a mechanism to stop the spread of disparity at borders with monocularly occluded areas. The authors also added temporal dynamics and an interocular inhibition constraint to the monocular occlusion

detection layer. These modifications allowed the model to reproduce the binocular rivalry phenomenon.

Cao and Grossberg (2005) have developed a biologically-plausible model of stereopsis with a modular, hierarchical architecture. Monocular occlusions are not detected explicitly but said to be signalled by monocular cells (cells receiving inputs only from one eye) that detect luminance edges and compete with binocular edge cells at each image location. This model also uses energy-type binocular neurons and the uniqueness constraint in the initial estimation of disparity.

Assee and Qian (2007) designed a model based on a coarse-to-fine energy-model (Chen and Qian, 2004) with an additional layer of binocular edge selective V2 neurons (von der Heydt et al., 2000). Each V2 neuron receives feed-forward connections from a group of V1 neurons. Half of these V1 neurons are tuned to one disparity and the other to another disparity, such that a V2 cell fires maximally when its receptive fields is centered on a depth step of its preferred disparities. The authors assumed that a depth step signifies the presence of monocularly occluded region whose width is equal to the disparity difference between the two surfaces creating the depth step. The disparity assigned to the monocular region is that of the more distant surface.

1.8 Objectives

The literature described in this chapter shows that monocular occlusions are not treated as noise by the visual system, but instead play a significant role in binocular vision. Given the abundance of monocularly occluded areas in natural scenes, an understanding of how the brain detects and uses monocular occlusions is vital for es-

tablishing a complete picture of stereopsis and depth perception in general. Although existing research provides a good basis for the study of the mechanisms involved in perception of depth from monocular occlusions, many aspects of this phenomenon are not yet understood. It is the goal of this thesis to investigate several such aspects in an effort to advance the current understanding of monocular occlusions, their role in stereopsis and their neural underpinnings.

The primary role of monocular occlusions in stereopsis

The pioneering investigators of monocular occlusions, Gillam and Borsting (1988), Nakayama and Shimojo (1990) and Anderson (1994), have suggested that the primary function of monocular regions in stereoscopic depth perception is to define depth discontinuities and the boundaries of the occluding objects in depth. In support of this hypothesis, Nakayama and Shimojo (1990) showed that introducing a monocularly occluded region in a sparse RDS created a smooth illusory edge instead of the jagged edge perceived in the absence of occlusion (see also Lawson and Gulick (1967) and Kaufman (1965)). However, due to the properties of the stimulus, the difference between the percepts was not very salient since many of the subjects could not spontaneously see it. The first goal of this thesis is to introduce a novel stimulus that produces a robust depth percept, and use it to evaluate the role of monocular occlusions in localizing depth discontinuities and defining the shape of an occluding surface. These experiments are described in Chapter 2.

The interaction of monocular occlusions and disparity

Depth from disparity and monocular occlusions has been shown to interact in complex ways. For example, monocular occlusions can effect perceived depth from ambiguous disparity (Hakkinen and Nyman, 2001). It has also been demonstrated

that in two-object arrangements an additional binocular surface can bias the perceived depth of the monocular object (Hakkinen and Nyman, 1996). A similar effect was observed with the novel illusory-occluder stimulus in the experiments described in Chapter 2. Consequently, the second goal of this thesis is to investigate the biasing effect of binocular features on depth from monocular occlusions, in particular in illusory occluder stimuli. Studying this interaction will shed light on the conditions under which depth from the two sources interact and the nature of these interactions. This set of experiments is described in Chapter 3.

Quantitative depth from monocular occlusions - one or many mechanisms?

Nakayama and Shimojo (1990) demonstrated quantitative depth in two-object arrangements and proposed that it is based on the constraints imposed by occlusion geometry. Their findings were later questioned by Gillam et al. (2003), who suggested that quantitative depth percepts in their stimuli arose from conventional disparity. As described in Section 1.3, quantitative depth based on occlusion geometry has been demonstrated in other classes of monocular occlusion stimuli. Thus, the findings of Gillam et al. (2003) suggest that da Vinci stereopsis is a stimulus dependent phenomenon since it relies on different mechanisms to compute quantitative depth in different occlusion configurations. Parsimony favours the explanation that there is a single mechanism responsible for quantitative depth in all types of geometrically constrained occlusion stimuli. Moreover, there were methodological issues with the Gillam et al. (2003) study that might have affected their results. Chapter 4 examines the nature and source of depth in the class of stimuli used by Nakayama and Shimojo (1990) to understand whether da Vinci stereopsis relies on one type of mechanism or

its function is stimulus dependent.

Double-matching and depth from monocular occlusions - a computational analysis

There have been other cases, beside that of Nakayama and Shimojo (1990), in which depth in monocular occlusion stimuli has been attributed to binocular matching (Brooks and Gillam, 2006; Gillam, 1995; Pianta and Gillam, 2003b). This necessarily implies double-duty matching in which the monocular feature is matched to a binocular feature that has its own correct match. In other words, the binocular feature is matched twice, once to itself, and once to the monocular feature. This is problematic since double-duty matching has not been demonstrated unequivocally (Howard and Rogers, 2002). Moreover, many models of stereopsis have successfully used a uniqueness constraint on matching of the points between the two images, such that each image point can be matched to only one other point in the fellow eye (e.g. Marr and Poggio (1976, 1979)). Finally, other experiments described in previous chapters of this thesis, suggest that depth can be based solely on occlusion geometry. The extent to which depth in monocular occlusion stimuli relies on binocular (double-duty) matching can be easily evaluated by using model disparity detectors and examining their responses to a range of monocular occlusion stimuli. These computational experiments are described in Chapter 5.

A biologically plausible model of depth from monocular occlusions and disparity

Although many computational algorithms that use monocular occlusions to build disparity maps have been proposed, these algorithms are not intended to describe biological functioning and for the most part are not biologically-plausible (see defini-

tion in Chapter 6). Several biologically-inspired computational models of stereopsis, described in Section 1.7, do concentrate on describing the neural mechanisms underlying da Vinci stereopsis. However, the majority of these models have biologically implausible components and none of them have a demonstrated ability to obtain depth in a variety of monocular occlusion stimuli. The final goal of this thesis is to introduce a biologically-plausible model of depth from monocular occlusions and disparity that incorporates the findings of the experiments described in this thesis and is able to compute veridical depth maps for a variety of monocular occlusion stimuli. The model is described in Chapter 6.

Chapter 2

The role of monocular occlusions in stereopsis

2.1 Objectives

The first goal of this thesis was to design a stimulus with a robust depth percept that can be used to evaluate the role of monocular occlusions in localizing depth discontinuities and defining the shape of an occluding surface.

In this stimulus, shown in Figure 2.1, the presence of monocular occlusions drastically alters the perceived shape of the occluding surface. It is composed of a random-dot frame with zero disparity, and a bipartite central area. The left part of this area is a random-dot square with crossed disparity and the right part is blank and devoid of any disparity information. In the absence of a monocular region to the right of the blank region, the foreground surface, closer to the observer, is perceived as a textured square (Figure 2.1-A). Adding a monocular random-dot strip to the right of the

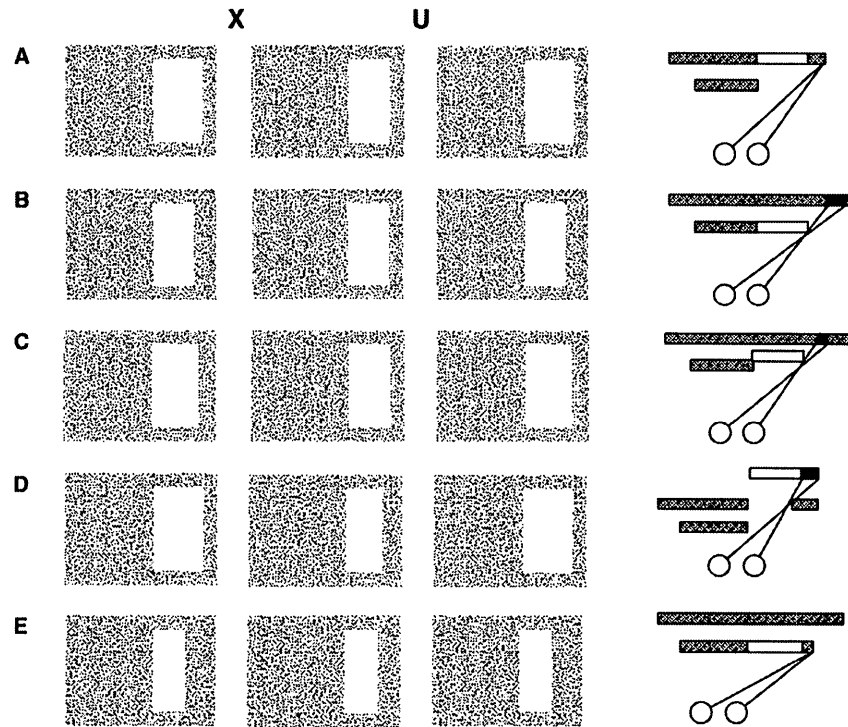


Figure 2.1: Stimuli used in Experiments 2.1 and 2.2. The rightmost column illustrates the typical corresponding cyclopean percepts where the random-dot texture is shown in gray, blank region in white and occluded regions in red. (A) No-Occlusion condition - there is no monocular region to the right of the blank area. The blank area is perceived as part of the background. (B & C) Occlusion condition - there is an occluded region in the right-eye's image to the right of the blank area. When the size of the monocular region is equal to the disparity of the central random-dot square (as in B), the blank area is perceived as part of the foreground. When the size of the monocular region is smaller than the disparity of the random-dot square (as in C), the blank area is perceived in between the background and the foreground. (D) Reverse-Occlusion condition - the monocular region is located in the left-eye's image and the blank area is perceived as lying behind the background. (E) Binocular-Strip condition - there is a binocular random-dot strip to the right of the blank area, which has a disparity equal to that of the central random-dot square. In this case, the blank area is perceived as part of the foreground.

blank region, changes the perceived configuration since an illusory occluder is created to account for the monocular occlusion. Now, the shape of the foreground surface is rectangular and it is composed of a textured square and a blank region (Figure

2.1-B,C). In the first experiment, this phenomenon is demonstrated empirically with naive observers.

This novel stimulus, which is referred to as the 'random-dot frame stimulus', also allows to evaluate the quantitative properties of da Vinci stereopsis. As was discussed in Section 1.3, quantitative depth in da Vinci stereopsis could be based on the minimum depth constraint, which depends on the width of the occluded region (see Figure 1.2). Although it is likely that the visual system relies on the minimum constraint, we do not have a complete understanding of the conditions for quantitative depth perception from monocular occlusions (see Section 1.3). The properties of the stimulus presented here provide a unique opportunity to manipulate occlusion geometry to generate uni- or bi-directional constraints on the minimum and maximum possible depth of the illusory occluder. By manipulating these properties, the conditions under which the visual system uses these constraints to assign a precise depth to the illusory occluder are investigated in detail in Experiment 2-4.

2.2 Experiment 2.1 - Monocular occlusions define the shape of occluding surfaces

2.2.1 Methods

Observers

Five observers naive to the purpose of the experiments participated in the study. All observers had normal or corrected-to-normal visual acuity and at least 20" stereoacuity as measured with the Randot[®] stereoacuity test.

Apparatus

Scripts for stimulus presentation were executed on a G5 Power Macintosh using Python 2.5. Stimuli were presented on a pair of CRT monitors (ViewSonic G225f) arranged in a mirror stereoscope with a viewing distance of 0.45 m. The resolution of the monitors was set to 1280×960 pixels and the refresh rate to 75Hz. At this resolution and viewing distance, each pixel subtended $2.24'$ of visual angle. The monitors were linearized and matched using a photometer to measure the gamma function. Observers used a chin rest to stabilize head position during testing.

Stimuli

The random-dot frame (background) of the stimulus shown in Figure 2.1-A and B, was $22.4'$ wide and was positioned at zero disparity. The central random-dot square subtended $2.24^\circ \times 2.24^\circ$ and had a crossed disparity of $4.48'$. The element density of the random-dot regions was 25% and each element was $2.24' \times 2.24'$. The whole stimulus subtended $4.2^\circ \times 3.2^\circ$. Dots were black on a grey background. Four variants of the stimulus were presented to the observers in Experiment 2.1:

1. **No-Occlusion** - In the first condition there was no monocular region on the border between the blank region and the random-dot frame (Figure 2-A).
2. **Occlusion** - In the second condition a monocular random-dot region was added in the right eye's image on the border between the blank region and the frame. The width of the region was $4.48'$, equal to the disparity of the textured foreground square. This stimulus corresponded to a depth arrangement where the blank region appears in front of the background (Figure 2-B).

3. **Reverse-Occlusion** - This stimulus was the same as the Occlusion stimulus, but the monocular random-dot region was added to the left eye image. This stimulus corresponded to a depth arrangement where the blank region appears behind the background (Figure 2-D).
4. **Binocular-Strip** - Instead of the monocular region used in occlusion and reverse-occlusion stimuli, a binocular random-dot strip was added on the border between the blank area and the background. The disparity of this strip was $4.48'$, equal to the disparity of the textured foreground square (Figure 2-E).

Note, that when the monocular region was introduced, normal binocular matching was not possible between the dots of the monocular region in one eye and the dots of the right-hand edge of the blank area in the other eye since these dots were uncorrelated. Consequently, the perceived depth of the illusory surface could not be based on conventional stereoscopic matching.

Procedure

At the beginning of each trial, the observers fixated on a white square ($31' \times 31'$) for one second after which the fixation mark was replaced by the stimulus. The stimulus presentation time was unlimited and observers could move their eyes freely. The observers were asked to judge the perceived shape ('square' or 'rectangle') of the foreground of the RDS in a forced-choice task and indicate their response using a gamepad. The instructions were presented in a written form for consistency. The presentation order was randomized. Observers completed 10 trials for each condition.

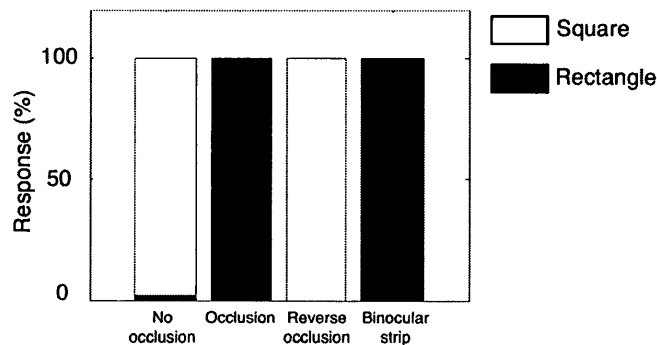


Figure 2.2: Results of Experiment 2.1 for all observers. The different types of stimuli are plotted on the abscissa. The ordinate shows the percentage of observer responses 'square' (in white) and 'rectangle' (in black) - for each stimulus.

2.2.2 Results

Figure 2.2 shows that when the occluded region was absent (Figure 2.1-A), the blank area was perceived at the depth of the background frame and the foreground region appeared square and limited to the random-dot texture. The presence of an occluded region along the vertical border of the blank area and the background (Figure 2.1-B), created a dramatic change in percept; all observers now perceived the blank area at the depth of the random-dot square, so the foreground was a continuous opaque rectangle which was monochromatic on one side and textured on the other. This suggests that the monocular region determined the location of a depth discontinuity and triggered the construction of a 3D surface. When the monocular region was added to the left half-image, instead of the right (Figure 2.1-D), all observers again reported that the foreground was square. In this case, the blank area appeared to be shifted behind the background frame elements. In the stimulus where the monocular region was replaced with a binocular strip of texture with disparity equal to that of the

foreground random-dot square (Figure 2.1-E) all observers perceived a rectangular surface, similar to that perceived in stimuli with monocular regions (Figure 2.1-B).

2.3 Experiment 2.2 - Quantitative depth from monocular occlusions

The stimulus introduced in Experiment 2.1 can be used to examine the nature of quantitative depth from monocular occlusions. As shown in Figure 2.3-A, in the Occlusion condition, where the blank area is seen in front of the random-dot frame, both the minimum and maximum possible depths are constrained. If the visual system uses geometric constraints specified by monocular occlusions then it should be able to localize the illusory region in depth with fairly high precision since the possible depth is constrained on both sides. The exact location of the occluder in depth will then depend on the width of the monocular region. In the Reverse-Occlusion condition, where the blank region is perceived beyond the random-dot frame, only the minimum depth is constrained (see Figure 2.3-A). Hence, in this condition, if the visual system adopts the minimum constraint, the perceived depth of the illusory occluder should also depend on the size of the monocular region. However, since the possible depth here is constrained only in one direction, the perceived location of the illusory occluder in depth might not be as precise as in the Occlusion condition. In both conditions, if quantitative depth can be perceived from monocular occlusions, then depth estimates should increase linearly with the width of the monocular regions in accordance with the geometric constraints.

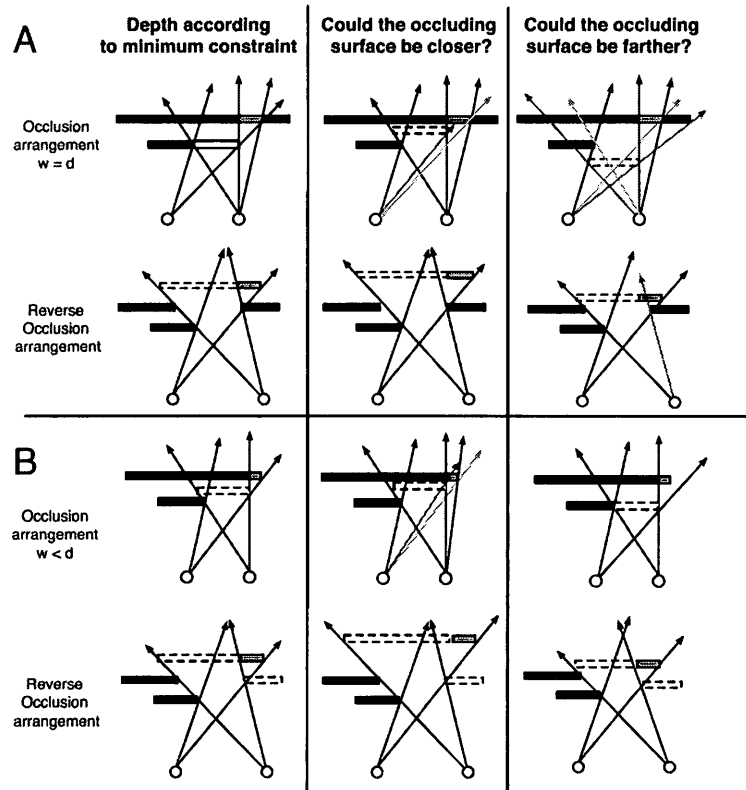


Figure 2.3: The viewing geometry of the stimuli in Experiments 2.1 and 2.3. The left column shows the depth arrangement according to the minimum depth constraint. The central and the right columns examine whether the viewing geometry allows for the illusory surface to be positioned closer to or farther from the background frame accordingly. The black rectangles represent the random-dot texture and the white rectangles with a dashed outline represent the blank region. Monocular occlusions are shown in bright green. Arrows show the lines of sight, with light gray arrows showing the lines of sight that comply with the minimum depth constraint and red arrows indicating the lines of sight that violate the geometric constraints. (A) In the Occlusion stimuli of Experiments 2.1 and 2.2 the depth of the blank region is constrained on both sides. In the Reverse-Occlusion stimulus, only the minimum possible depth of the blank region is constrained. (B) When the right portion of the random-dot surround is removed, geometric constraints are weakened and the depth becomes unconstrained in one direction in the Occlusion arrangement and in both directions in the Reverse-Occlusion arrangement.

2.3.1 Methods

The observers and apparatus were the same as in Experiment 2.1. The stimulus was the same as in Experiment 2.1 except that the central random-dot square had a crossed disparity of $13.44'$ and the width of the occluded region was one of $0, 4.48', 8.96'$ or $13.44'$. The monocular region was added either to the images of the left or the right eye resulting in 7 different conditions (2 eyes \times 4 widths). The observers were asked to adjust a disparity probe (antialiased black circle with radius $13.44'$), using a gamepad, to match the depth of the blank portion of the central area of the RDS. The probe was presented $\sim 1^\circ$ to the left of the RDS and its initial disparity was chosen at random. The disparity of the depth probe was adjusted in steps of 1.12 minutes (0.5 a pixel).

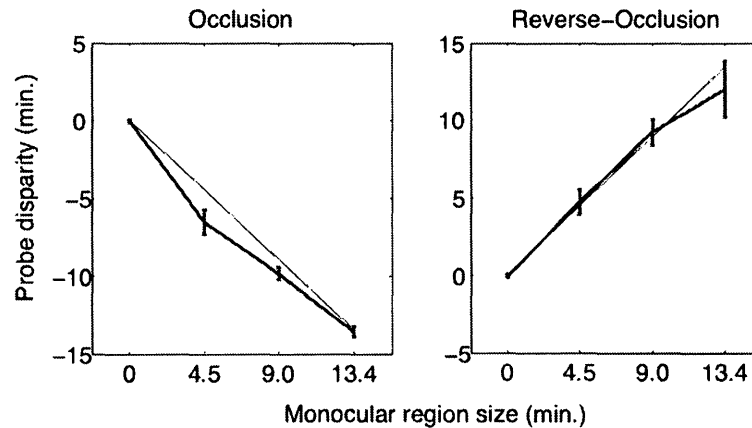


Figure 2.4: Results of Experiment 2 for all observers except for SL. The ordinate shows the disparity of the probe and the abscissa shows the size of the monocular region. Negative values are assigned to crossed depth. The error bars show ± 1 standard error of the mean. The thin black line indicates the predicted disparity.

Statistical analysis

The data for this and the subsequent experiments in this chapter were analyzed using one-way repeated measures ANOVA (linear regression analysis could also be used here, however rep. mes. ANOVA is preferable here since regression requires the data points to be independent of each other). Data for each condition (Occlusion and Reverse-Occlusion) were analyzed separately using an alpha level of 0.01.

2.3.2 Results and Discussion

The results of Experiment 2.2 are shown in Figure 2.4. In the Occlusion condition, when monocular occlusions of variable width were added to the right half-image, the blank area appeared positioned at intermediate depths, between the random-dot square and the background (e.g. compare B and C in Figure 2.1). Observers' depth estimates increased significantly with the increase in the size of the monocular region ($F(3, 15) = 421, p < 0.001$). In the Reverse-Occlusion condition, the blank region appeared to lie at different depths behind the surround depending on the width of the monocular region. Similar to the Occlusion condition, perceived depth increased significantly with the width of the monocular region ($F(3, 15) = 16.9, p < 0.001$). One of the observers (SL) did not see depth in this condition. Consequently she was removed from the sample when the data was combined to create Figure 2.4. Mean data for all observers including SL are shown in Appendix B. Note, that the standard errors are slightly larger in the Reverse-Occlusion condition (especially for the largest width) indicating that the depth estimates among observers were more variable in this condition than in the Occlusion condition. This could occur since the occlusion geometry in the Reverse-Occlusion condition restricts the possible depth of

the illusory occluder only in one direction which makes makes depth estimates less constrained.

2.4 Experiment 2.3 - Controlling for binocular matching

The results of Experiment 2.2 suggest that quantitative depth can be perceived from monocular occlusions in the random-dot frame stimuli. However, in light of the findings of Gillam et al. (2003) and Liu et al. (1997) it is important to rule out any possibility of binocular matching. One possible scenario is that size disparity between the blank regions in the two eyes, introduced due to the addition of the monocular region, created a percept of slant in the blank region. However, this explanation can be rejected since neither slant nor depth are perceived in the No-Occlusion condition although due to the disparity shift of the random-dot square, there is also size disparity between the blank regions in this condition (see Figure 2.1-A). Moreover, when asked, observers did not report slant percepts in the stimuli of Experiments 2.1 and 2.2.

Another possible explanation for the results of Experiment 2.2 involves binocular disparity. When the stimuli contain monocular occlusions, the random-dot textures defining the right-hand edges of the blank area in the right and the left half-images are uncorrelated. Consequently, these random dots cannot be coherently matched by the stereoscopic system. However, the texture-defined edges could potentially be matched rather than the individual texture elements. Importantly, the disparity between these edges would be equal to the width of the monocular region and would

predict the same depth percepts.

There are several reasons why this scenario is not very likely. First, note that in this scenario double matching would have to take place. Initially, the binocular dots of the right-hand part of the random-dot frame would be matched in the two eyes. Then the texture edge formed by these same dots in one eye would have to be matched to the edge formed by the monocular region in the other eye.

Second, observer SL (an experienced observer with good stereopsis) could not see quantitative depth in the Reverse-Occlusion stimuli although she could easily perceive depth in the Occlusion condition. If quantitative depth in the present stimuli was indeed based upon stereoscopic matching, then this observer should have seen depth in both conditions. On the other hand, individual differences in perception of quantitative depth from occlusions have been reported before in the literature (Cook and Gillam, 2004).

The third argument concerns the element density in the stimuli. Matching texture borders requires the borders to be well defined. In Experiments 2.1 and 2.2 the stimulus dot density was set to 25%, creating densely textured edges. Figure 2.5 shows an example of the stimuli with density of 2.7%. In (A) the width of the occluded region is equal to the disparity of the random-dot square and in (B) it is smaller. The texture borders in these stimuli are not as well defined due to low density, yet, the illusory occluder is still seen at different depths in (A) and (B). However, even when a sparse texture is used a border contour can still be discerned. Consequently, to evaluate the possible role that binocular disparity played in Experiments 2.1 and 2.2, the right portion of the random-dot surround was removed from the stimuli (see Figure 2.6) such that the right-hand texture border created by the monocular

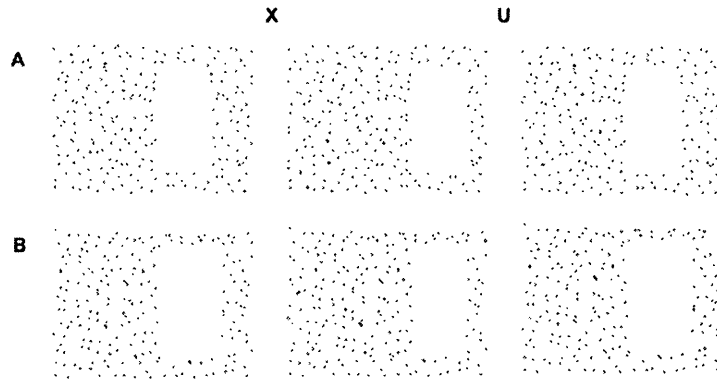


Figure 2.5: An example of the stimuli with 2.7% density (actual stimuli had 25% density), both panels showing the Occlusion condition. In (A) the size of the occluded region is equal to the disparity of the random-dot square and in (B) it is smaller than the disparity of the random-dot square.

occlusion could not be matched to anything in the other eye. Experiments 2.1 and 2.2 were then repeated using the modified stimuli.

2.4.1 Methods

The observers, apparatus and procedure used in the two parts of Experiment 2.3 were the same as in Experiments 2.1 and 2.2 respectively. The stimulus was modified by removing the rightmost portion of the random-dot background (see Figure 2.6) so it subtended $3.74^\circ \times 3.2^\circ$.

2.4.2 Results and Discussion

The results of Experiment 2.3 for the qualitative task (rectangle vs. square) shown in Figure 2.7, were virtually identical to those of Experiment 2.1 (Figure 2.2). The

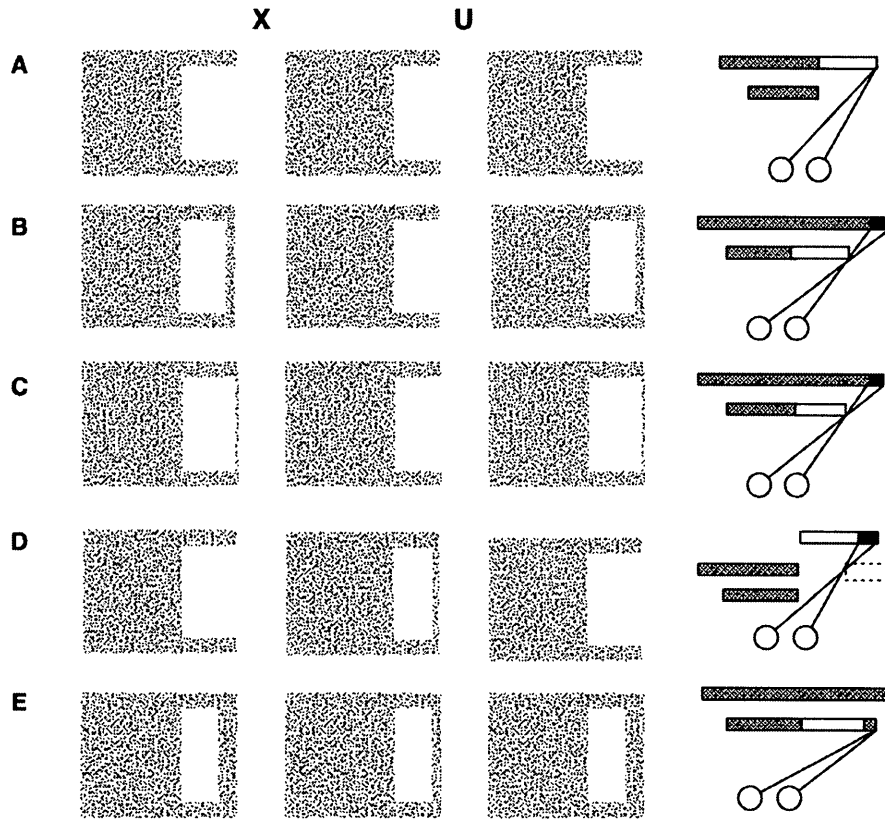


Figure 2.6: Stimuli used in Experiment 2.3. The stimuli are identical to those used in Experiments 1 and 2 except that the rightmost portion of the background has been removed (see legend for Figure 2.1).

presence of the occluded region created a depth edge, which triggered the percept of an illusory occluder, while in the absence of monocular occlusions the blank area was perceived as part of the background. Consequently, this effect does not depend on binocular matching and is based solely on monocular information present in the stimulus. However, the results of the disparity-matching task shown in green in Figure 2.8, were different from those of Experiment 2.2 (shown in blue). Regardless of the width of the monocular occlusion, the presence of monocular occlusions in

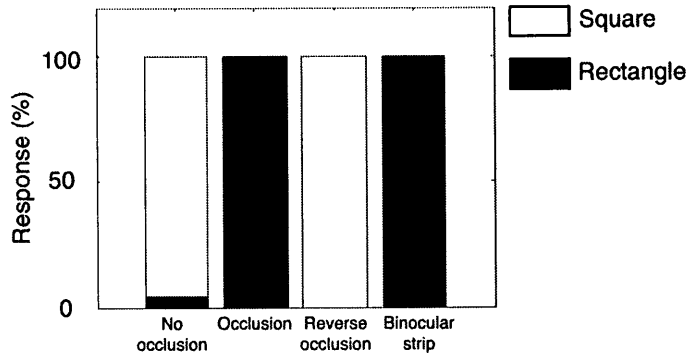


Figure 2.7: Results of the qualitative task of Experiment 2.3 for all observers. The different types of stimuli are plotted on the abscissa. The ordinate shows the percentage of observer responses square (in white) and rectangle (in black) - for each stimulus.

the right half-image (Occlusion condition) yielded a percept of the blank area at the same depth as the random-dot square (ANOVA with three non-zero widths gave $F(2, 10) = 3.05, p = 0.092$). When the monocular regions were introduced in the left half-image (Reverse-Occlusion condition), the blank area was perceived to lie at the same short distance behind the surround regardless of the width of the monocular occlusion ($F(2, 10) = 2.44, p = 0.136$).

Two important points are raised by these data. First, they offer additional evidence that monocular occlusions influence the perceived depth of the illusory occluding surface. In the Occlusion condition the presence of monocular occlusions, regardless of their size, creates a percept of an illusory occluder seen in depth. Second, these results suggest that binocular matching of the texture edges may have been responsible for the quantitative depth percepts in Experiment 2.2. However, note that in the modified stimuli the geometric constraints are weaker than in the original stimuli. For both occlusion configurations in our original stimuli (Figure 2.3 (A)), the minimum

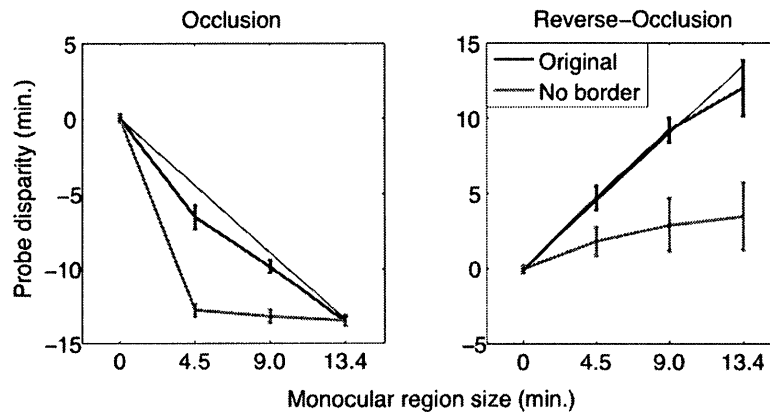


Figure 2.8: Results of the disparity-matching task of Experiment 2.3 for all observers except SL. The ordinate shows the disparity of the probe and the abscissa shows the size of the monocular region. The blue line shows the depth estimates with the stimuli used in Experiment 2.2 (Original) and the green line shows the depth estimates with the modified stimuli with the right-hand border removed (No border). Negative disparity values are assigned to crossed depth. The error bars show ± 1 standard error of the mean. The thin black line indicates the predicted depth.

depth or both minimum and maximum depths were constrained. Hence, the visual system could have relied on this information to localize the blank region. In the modified Reverse-Occlusion stimuli the magnitude of the depth from occlusions was not constrained at all after the removal of the right-hand border as shown in Figure 2.3-B. This could explain the absence of quantitative depth in these stimuli in Experiment 2.3. In the Occlusion stimuli, after the removal of the right-hand border, only the minimum depth remained constrained (see Figure 2.3-B). Although the visual system could have used this constraint to precisely localize the illusory occluder in depth, the occluder was always perceived at the depth of the binocular square. It is possible that the illusory occluder was captured by the strong disparity signal of the adjacent random-dot square. Since the depth of the illusory surface was restricted in only one

direction (minimum), it could have been pulled in the unrestricted direction, towards the binocular square that had a strong disparity signal. This possibility was tested in Experiment 2.4.

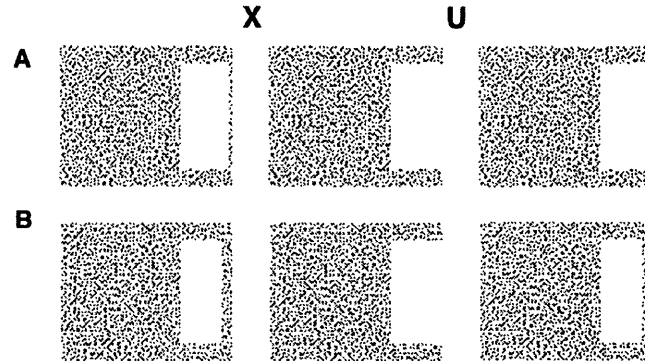


Figure 2.9: Stimuli used in Experiment 2.4. The stimuli from Experiment 2.3 were modified by placing the random-dot square at zero disparity. In (A) the width of the monocular region is smaller than in (B).

2.5 Experiment 2.4 - Disparity biasing

To examine the effect of the disparity of the random-dot square on the perceived depth of the illusory occluder, the disparity of the binocular square was set to zero in the stimulus with no right-hand border (see Figure 2.9). The disparity-matching experiment (Experiment 2.2) was then repeated. In these stimuli, binocular matching of texture-defined edges cannot take place since there is no right-hand border. Binocular capture should not affect the depth of the illusory occluder perceived in the presence of the monocular region since the binocular square has zero disparity and cannot pull the right edge of the illusory occluder in the unrestricted direction.

Consequently, any increase in the perceived depth of the illusory surface with the increase in the width of the monocular region would suggest that quantitative depth is based purely on occlusion information. The modified stimuli are shown in Figure 2.9. Note that now the illusory surface looks slanted. Its left edge is at zero disparity, alongside the random-dot texture, and its right edge is elevated due to the presence of monocular occlusions. This slant is not likely to be caused by size disparity between the images of the two eyes because there is no right-hand border and no clear indication where the blank region ends in one eye. Instead, the slant occurs due to the interpretation of the occluded region as part of the binocular frame located at the fixation plane. The blank area on the side of the occlusion is then interpreted as an occluding surface with crossed depth.

2.5.1 Methods

Three naive observers participated in this experiment; two from the original sample (MV and DS) and a new observer (MT). The new observer completed the disparity matching task with stimuli from Experiments 2.2 and 2.3 before participating in Experiment 2.4. The stimuli were the same as in Experiment 2.3, except that the random-dot square had zero disparity (see Figure 2.9). Observers were asked to set the disparity probe to the perceived depth of the right edge of the illusory occluder for stimuli with different monocular region widths. Only the Occlusion condition, with the monocular occlusion in the right eye, was tested. The procedure for this experiment was exactly the same as in Experiment 2.2.

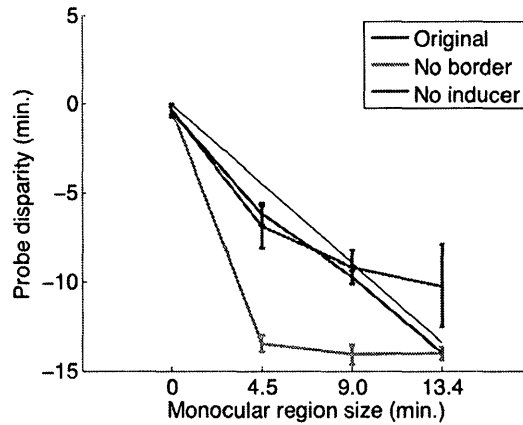


Figure 2.10: Results of Experiment 2.4. The ordinate shows the disparity of the probe and the abscissa the size of the monocular region. The blue, the green and the red lines show the depth estimates from Experiment 2.2 (Original), Experiment 2.3 (No border) and Experiment 2.4 (No inducer) respectively. The error bars represent ± 1 standard error of the mean. The thin black line shows the predicted disparity based on occlusion geometry.

2.5.2 Results and Discussion

All observers perceived the illusory occluder as slanted in depth and could make veridical depth judgments of its right-hand edge. As shown in Figure 2.10, the mean perceived depth of the illusory occluder for 3 observers increased with the increase in the width of the monocular region $F(3, 6) = 39.66, p < 0.001$. Perceived depth for the smallest three widths of the monocular region was very similar to that in the original experiment, while for the largest width the depth in the current stimuli was underestimated on average. The relatively larger error bars for the largest occlusion width reflect individual differences in the estimates. These data suggest that quantitative depth could be perceived in our stimulus on the basis of monocular occlusions alone.

2.6 General Discussion

The experiments described in this chapter provide direct evidence that the visual system uses monocular occlusions to localize depth discontinuities and to define the shape and depth of the occluding surfaces. In the stimuli used in these experiments, the blank region was perceived as part of the background and the foreground had a square shape in the absence of a monocularly occluded region. The presence of a monocular region at the border in the right eye (Occlusion condition) signalled a depth edge and triggered a percept of an occluding surface composed of blank and textured parts. When the monocular region was placed in the corresponding location in the other eye (Reversed-Occlusion condition), it signalled to the visual system that the blank region was positioned behind the textured surround creating an aperture through which the occluded region and the blank region were seen. The occluder in this case was the background frame. Interestingly, when the right portion of the background frame was removed, an illusory intruding edge was perceived in its place seemingly to account for the monocular occlusion (see Figure 7-D).

Experiment 2.2 showed that the magnitude of the perceived depth of the illusory occluder increased as the width of the monocular region increased in both Occlusion and Reverse-Occlusion configurations. This result is consistent with the restrictions imposed by the viewing geometry and it is likely that the visual system used these constraints to assign a precise location in depth to the illusory surface. When the experiment was repeated with the right-hand border removed, the illusory surface was always perceived at the depth of the random-dot square in the Occlusion condition, and at some distance behind the random-dot frame at the Reversed-occlusion condition. It is possible that binocular matching between the texture-defined edges

of the random-dot frame and the occluded region took place in Experiment 2.2 and the removal of the edge prevented this matching. On the other hand, the geometric constraints, which the visual system would have to rely on to extract depth from occlusions, were weakened with the removal of the right-hand border. In the Reversed-Occlusion condition, the position of the illusory surface was not restricted at all (except in sign) after the removal of the border, while in the Occlusion condition only the minimum depth remained restricted. I proposed that in the Occlusion condition, the illusory occluder was perceived at the depth of the random-dot square at all monocular region widths because the strong disparity signal of the square pulled the illusory surface in the direction unrestricted by the geometry. Indeed, when the square was given zero disparity, quantitative depth perception was restored, although depth was underestimated at the largest width of the monocular region. Taken together, these experiments reveal that the visual system is able to utilize the geometric constraints imposed by monocular occlusions to localize occluding surfaces in depth. However, monocular occlusions seem to provide a relatively weak cue to quantitative depth. The presence of an unambiguous disparity signal in the proximity of an illusory occluder can alter the perceived depth of the occluder when its depth is not completely restricted by the viewing geometry. These results demonstrate clearly that the visual system uses monocular occlusions to identify the location and direction of depth discontinuities and object boundaries in a scene. Further, for ambiguous surfaces the occluded regions help define object shape and estimate the object's position in depth. Thus, monocular occlusions are not simply a by-product of stereoscopic matching, but an important stage in the identification of depth discontinuities in a complex visual environment.

Chapter 3

Interaction of monocular occlusions and disparity

3.1 Objectives

The experiments of Chapter 2 suggested that a binocular surface could bias the perceived depth of the illusory occluder in random-dot frame stimuli, when occlusion geometry provided only partial constraints on the position of the occluder in depth. Hakkinen and Nyman (1996) demonstrated a related phenomenon in two-object arrangements. In these displays a monocular dot was placed a certain distance from a binocular rectangle, which was perceived as occluding the dot. Another binocular rectangle (inducer) was placed above the occluder (see Figure 1.5-B). The perceived depth of the monocular dot increased as the inducer moved farther behind the occluder. Interestingly, in the original article where two-object arrangements were used first, Nakayama and Shimojo (1990) anecdotally reported a similar phenomenon.

The next set of experiments presents a general test of the hypothesis that the perceived depth of an illusory occluder can be biased by a binocular feature in the direction unrestricted by occlusion geometry. Examining this phenomenon will allow to study the nature of the interactions between binocular disparity and monocular occlusions and its dependence on geometric constraints. Experiment 3.1 provides additional evidence for the biasing effect of binocular disparity in stimuli of Chapter 2. Experiment 3.2 demonstrates how this phenomenon generalizes to other illusory occluder stimuli. Experiment 3.3 shows that, in a stimulus where the illusory occluder is localized by both occlusion and some disparity information, most observers rely on the disparity cue and thus exhibit little bias. Finally, Experiment 3.4 shows that the biasing does not take place in stimuli where the depth of occluding surfaces is defined by a reliable disparity signal rather than monocular occlusions.

3.2 Experiment 3.1 - random-dot frame stimulus

This experiment examines the effect of the random-dot square disparity on the perceived depth of the illusory occluder in the random-dot frame stimuli of Chapter 2.

3.2.1 Methods

Observers

One experimenter and four naive observers participated in the study. All observers had normal or corrected-to-normal visual acuity and at least 20'' stereoacuity as measured with the Randot[®] stereoacuity test.

Apparatus

The apparatus was the same as in the experiments of Chapter 2 except that the viewing distance was set to 0.6 meters (since the stereoscope was re-calibrated). At this resolution and viewing distance, each pixel subtended $1.77'$.

Stimuli

The stimuli were the ones used in Experiments 2.3 and 2.4 (see Figure 3.1), but the dimensions in visual angles were slightly smaller due to the increased viewing distance of the apparatus (random-dot frame width - $17.7'$, random-dot square size - $1.77^\circ \times 1.77^\circ$, blank region size - $0.88^\circ \times 1.77^\circ$, dot size - $1.77' \times 1.77'$ and complete stimulus size - $2.9^\circ \times 2^\circ$). The width of the monocular strip was either $3.5'$ or $7.08'$. For stimuli with the $3.5'$ monocular strip the disparity of the binocular square was set to $3.5'$, $7.08'$ or $10.62'$. When the monocular region was $7.08'$ wide the disparity of the binocular square was set to $7.08'$, $10.62'$ or $14.8'$. In total there were 6 different stimuli types (3 square disparities x 2 monocular region widths).

Procedure

The procedure was the same as in Experiment 2.2 (the probe had a $10.6'$ radius, and was adjusted in steps of $0.88'$). Each stimulus condition was presented 20 times in random order with 120 trials in total (20 times x 6 conditions). The experiment was completed in one session. Before the beginning of the experiment the observers were given a short training session to familiarize them with the task.

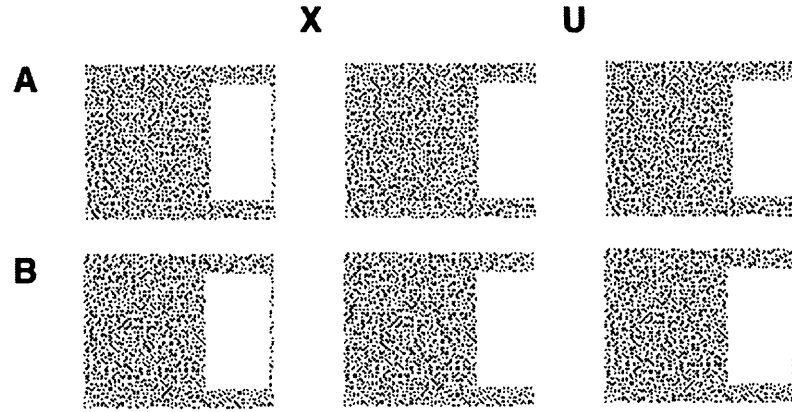


Figure 3.1: Example of stimuli used in Experiment 3.1. In (A) and (B) the monocular region has the same width. In (B) the binocular textured square has a larger disparity than in (A) and hence the perceived depth of the blank region in this case is larger.

Statistical analysis

The Friedman test, which is a non-parametric equivalent of one-way repeated measures ANOVA, was used to analyze the data (this test was chosen here to avoid trying to satisfy the requirements of the ANOVA test). Data for each monocular occlusion width was analyzed separately using an alpha level of 0.01.

3.2.2 Results and Discussion

Figure 3.2 shows the results of Experiment 3.1. The estimated disparity of the blank region is shown as a function of the disparity of the textured square and monocular region width. The dashed blue and the dotted red lines show the theoretical minimum disparity specified by the occlusion constraints for each monocular region width. If the textured square had no effect on the perceived depth of the blank region then observers estimates should follow these lines (perceived depth of the blank region

should remain constant). However, it is clear from the figure that the perceived depth of the blank region followed the disparity of the textured square regardless of the size of the occluded region. As the disparity of the textured square increased so did the estimated disparity of the blank region, as if the textured binocular square pulled the blank region towards itself. Statistical analysis showed a highly significant effect of the disparity of the random-dot square on the perceived depth of the occluder for both occlusion widths ($\chi^2 = 10, p < 0.01$ for both widths).

It is possible that the blank region is perceived as slanted since its depth is defined by monocular occlusions on the right side only. Since only one disparity probe was used in Experiment 3.1, slant was not assessed. To verify that the blank region was perceived as a frontoparallel surface, a subset of conditions in Experiment 3.1 was retested with a slightly different task. Only the monocular area width of 3.54' was used, with the same 3 inducer disparities as in the original experiment. Two disparity probes were used; one was centered above the right edge of the blank region and the other above its left edge. Only one probe appeared on each trial to avoid biasing observers. The probes and the stimuli were presented in random order. Two observers, naive as to the purpose of the experiment participated in the study. These were new observers, who did not participate in Experiment 3.1, therefore they did not have expectations regarding the probe settings. They were asked to align the probes to the perceived depth of the appropriate edges of the blank region. In all conditions observers set the two probes to virtually the same disparity, as would be expected if the blank region was perceived as a fronto-parallel surface.

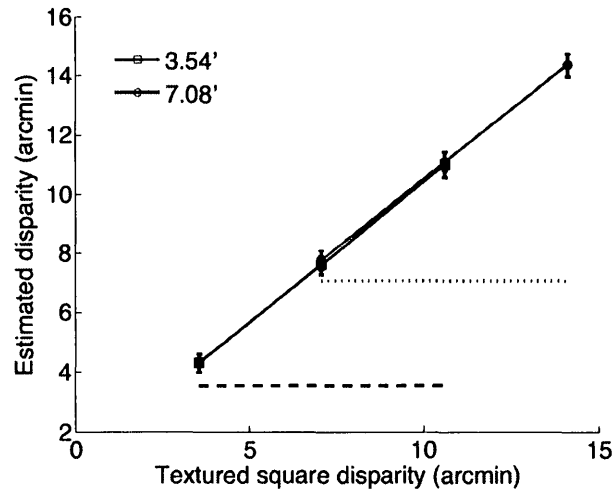


Figure 3.2: Results of Experiment 3.1 for five observers. The inducer disparity is plotted on the abscissa and the estimated disparity of the blank region on the ordinate. Blue and red lines show the data for stimuli with occluded region of width 3.54' and occluded region of width 7.08' respectively. Dashed blue and dotted red lines show the theoretical depth of the blank region if the depth of square had no effect on the depth of the blank region. Error bars show ± 1 standard error of the mean.

3.3 Experiment 3.2 - Disparity bias in

the stimulus of Gillam and Nakayama (1999)

This experiment examined whether disparity biasing generalizes to other illusory occluder stimuli, in particular the stimulus designed by Gillam and Nakayama (1999) shown in Figure 3.3. This stimulus was chosen since the depth of the illusory occluder in this stimulus relies solely on occlusion geometry. Using this stimulus, Gillam and Nakayama (1999) found that the perceived depth of the illusory occluder increased as the thickness of the bars increased, in accordance with occlusion geometry (see Figure 3.3 A and B). Experiment 3.2 tested whether a binocular inducer could bias the depth magnitude of the illusory occluder to the point that its perceived location

was determined by the inducer disparity instead of the minimum depth constraint.

3.3.1 Methods

Observers

Two experimenters and four naive observers participated in the study. All observers had normal or corrected-to-normal visual acuity and at least 20" stereoacuity as measured with the Randot stereoacuity test.

Apparatus

The apparatus was the same as the one used in Experiment 3.1 except that the scripts for stimulus presentation were executed using Psychtoolbox (v. 3.0.8) for MATLAB (v. 7.4) (Brainard, 1997).

Procedure

Experiment 3.2 consisted of three parts:

Occluder-Only - The first part replicated the original experiment of Gillam and Nakayama (1999). Their stimulus was presented with three different bar widths and observers were asked to estimate the depth of the illusory occluder using a disparity probe. In pilot experiments, the original protocol used by Gillam and Nakayama was adopted by displaying the probe and the stimulus simultaneously and allowing unlimited viewing time. However, the less experienced observers found it difficult to estimate the depth of the illusory occluder under unrestricted viewing conditions. Consequently, in Experiment 3.2 the presentation time was limited to 150 ms. The observers first fixated the zero-disparity nonius lines and, when they were well aligned,

pressed a button to initiate a trial. After the stimulus was presented, the disparity probe appeared and observers adjusted its depth to match the remembered depth of the illusory occluder (initially the probe had zero disparity). They were instructed to judge the depth at the vertical edges of the illusory occluder since the monocularly occluded areas define the shape and the depth of the vertical edges in this stimulus, and the depth signal is then interpolated between them. Thus depth at the edges of illusory occluders is most suitable to assess the effects of occlusion constraints and their interaction with disparity. Each of the three bar widths was presented 20 times in random order for a total of 60 trials in a single session.

Inducer-Only - In the second part of the experiment, the observers were asked to estimate the perceived depth of the binocular inducer only. This permitted a direct comparison between the perceived depth of the inducer and the perceived depth of the illusory occluder in the subsequent Inducer+Occluder condition. The procedure was the same as in part one. Each of the 6 inducer disparities was presented 20 times in random order for a total of 120 trials. Observers completed this part in two sessions of 60 trials each.

Inducer+Occluder The third part of this experiment was designed to reveal any biasing from the binocular inducer on the perceived depth of the illusory occluder. The stimulus used here was the original Gillam and Nakayama (1999) stimulus with a binocular inducer (disc) centered between the stimulus lines (see Figure 3.3 C and D). The procedure was the same as in the other two parts. There were 18 conditions (6 inducer disparities x 3 widths) repeated 20 times for 360 trials in total. Observers completed this part in four sessions of 60 trials each.

The three parts of this study were completed in the same order by all observers: 1)

Occluder-Only 2) Inducer-Only and 3) Inducer+Occluder conditions. This order was chosen to familiarize the observers with the two component stimuli before they were tested with the combined stimulus. Before the first session of each part, observers were shown the stimulus with an extended presentation time to familiarize them with the stimulus and then they were given a short practice session with the short exposure duration used in subsequent testing.

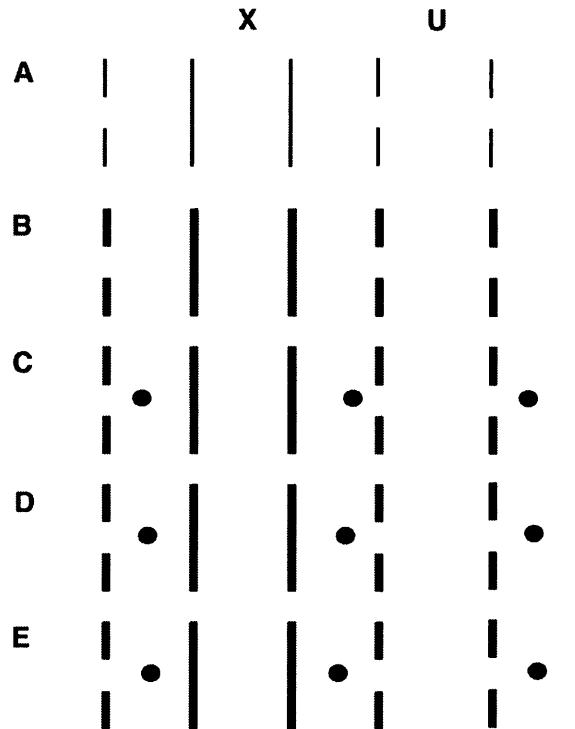


Figure 3.3: Stimuli used in Experiment 3.2. Typically observers interpret the gaps in the bars as monocular occlusions and perceive an illusory occluding surface floating in front of the bars. In (B) the defining bars are thicker than in (A) and hence the illusory occluder appears to have more depth. In (C) and (D) a binocular disc (the inducer) with crossed disparity is placed midway between the two bars. In (C) the disparity of the inducer is larger than in (D) and the illusory occluder is seen farther from the defining bars even though the bars have the same thickness. In (E) the inducer has a smaller crossed disparity than that specified by the geometric constraints. The edges of the occluder in this case are still seen in front of the bars, but its center appears concave.

Stimuli

Occluder-Only - The stimuli used in this part replicated those used by Gillam and Nakayama (1999). The bar length was 2.5° , the horizontal separation between the bars was 1.5° and the vertical gap in one of the bars was $44'$. The width of the bars was either $1.77'$, $3.54'$ or $5.31'$. The selected range of bar widths was similar to that used in the original study. The bars were black on a gray background and had zero disparity (see Figure 3.3 A and B)

The subsequently presented disparity probe consisted of two bars identical to the stimulus bars but without the monocular gap so that there was no percept of an illusory occluder. A black disc, $17.7'$ in diameter, was positioned at the center of the bar configuration. The black disc could be moved in depth in $1.77'$ steps. The two bars had zero disparity with respect to the plane of the screen and served as a reference for the depth of the disc. The disc was antialiased to allow for subpixel shifts. Both the stimulus and the probe were positioned in the center of the display.

Inducer-Only- This stimulus consisted of two bars with the same dimensions as the stimulus in the Occluder-Only condition (width $3.54'$), but without the monocular gap so that there was no percept of an illusory occluder. The inducer was a white disc, $10.6'$ in diameter, positioned at the center of the line configuration with disparity of $3.54'$, $0'$, $-3.54'$, $-10.62'$, $-14.16'$ or $-17.7'$. The choice of disparity values is discussed in detail below. The disparity probe in this part was the same as in the Occluder-Only condition.

Inducer+Occluder This stimulus was a combination of those used in the preceding two conditions. The inducer disc was placed in the center of the illusory occluder stimulus (see Figure 3.3 C - E). The three bar widths were the same as in

the Occluder-Only condition and the six inducer disparities were the same as in the Inducer-Only condition. Both crossed and uncrossed disparities were used for the inducer, although according to the geometric constraints the illusory surface must always be positioned some distance in front of the stimulus lines. Uncrossed (and small crossed) disparities were included to verify that the disparity biasing is indeed constrained by occlusion geometry as proposed in Chapter 2 and that the observers were estimating the depth of the illusory surface, and not the depth of the inducer. The disparity probe in this part was the same as in the Occluder-Only part.

Statistical analysis

Occluder-Only / Inducer-Only As in Experiment 3.1 the Friedman test was used to assess the main effects of bar width or inducer disparity with an alpha of 0.01.

Inducer+Occluder Due to inter-observer variability, data obtained in this condition were analyzed separately for each observer. A linear regression analysis was used with perceived depth as the dependent and inducer disparity as the independent variables. The slopes of the regression lines showed the relationship between inducer disparity and perceived depth of the occluder; the larger the slope, the stronger the relationship. A randomization procedure was used to generate an empirical distribution for the F statistic corresponding to each regression model. This is a non-parametric technique that avoids making explicit assumptions about the underlying distribution of the data. The data corresponding to each bar width were analyzed separately. To assess the role of geometric constraints on depth perception in our stimuli two regression models were computed for each bar width. The first model used data corresponding to inducer disparities that did not violate the geometric constraints

(different for each line width) and the second model used the data corresponding to the rest of the disparities. The alpha level was 0.01.

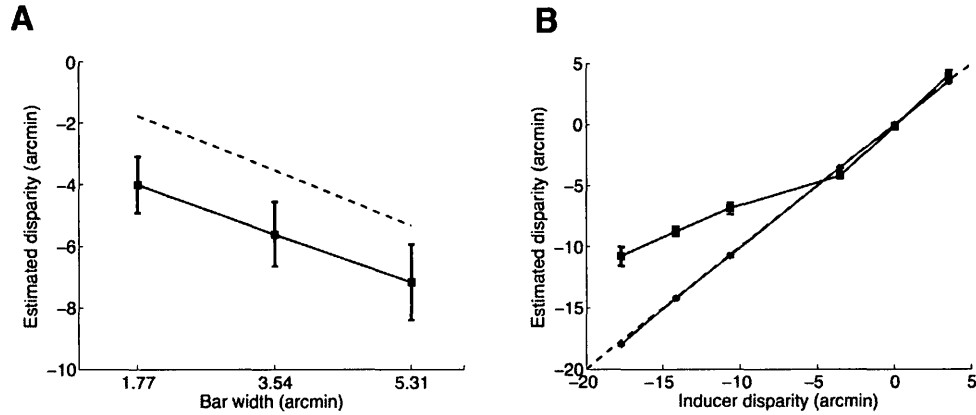


Figure 3.4: Results of Experiment 3.2. (A) Results of the Occluder-Only condition. The bar width is plotted on the abscissa and the estimated disparity on the ordinate. (B) Results of the Inducer-Only condition. Inducer disparity is plotted on the abscissa and the estimated disparity on the ordinate. Blue solid lines show the mean data for six observers and the black dashed lines show the theoretical disparity. Error bars show \pm standard error of the mean. The purple line in (B) shows the mean data for observers IT and DS for the control experiment (see text for details).

3.3.2 Results and Discussion

Occluder-Only Since all observers had the same pattern of results, mean data are shown in Figure 3.4-A. There was a significant increase in estimated disparity with increasing bar width ($\chi^2 = 12, p < 0.01$). These data confirmed the results reported by Gillam and Nakayama (1999) including their observation that depth estimates were larger than those predicted by the minimum constraint.

Inducer-Only As in the Occluder-Only condition, the data were consistent across all observers, thus mean data are shown in Figure 3.4-B (blue line with square

markers). As expected, observers' estimates of the inducer disc disparity increased significantly with increasing disparity ($\chi^2 = 25, p < 0.001$). At larger inducer disparities observers underestimated the disparity. Since the probe appeared after the presentation of the stimulus it is possible that the time delay reduced the accuracy of observers' judgments, more so for larger inducer disparities. To verify this, the experiment was repeated for two observers (IT and DS) with unlimited presentation time and the probe presented simultaneously with the stimulus (the probe was a small disc positioned to the right of the stimulus). The disparity estimates in this case, as shown in Figure 3.4-B (purple line with circular markers), were veridical.

Inducer+Occluder The black line with no markers in Figure 3.5 shows the results of the Inducer-Only task: the estimated disparity of the inducer in the absence of the illusory surface. The colored lines with different markers show the estimated disparity of the illusory surface in the presence of the inducer for different line widths. If the perceived disparity of the illusory occluder was not affected by the inducer, there should be little change in the estimated disparity of the illusory occluder as the disparity of the inducer changed. This clearly was not the case for five of the six observers. Three observers DS, IT and AS showed a strong bias; when the inducer disparity increased in the crossed direction, beyond the disparity specified by the minimum constraint, the disparity estimates became significantly biased in the crossed direction (mean slopes - DS - 0.36, IT - 0.48, AS - 0.23; all slopes differed from zero with $p < 0.01$). A similar trend can be seen for observers LT and LW, although the biasing was weaker (mean slopes - LT - 0.14, LW - 0.16; all slopes differed from zero with $p < 0.01$). Observer FZ showed no significant effect of the inducer on the perceived depth of the illusory occluder (mean slope - 0.015).

The configurations in which the occluder had an uncrossed or a smaller crossed disparity than the minimum specified by the geometric constraints created a very distinctive percept. That is, the illusory surface appeared to warp so that its edges sat toward the observer and its center was at the depth of the inducer (see Figure 3.3-E). The edges of the occluder were perceived to lie in front of the stimulus bars consistent with occlusion geometry. As indicated in the Section 3.2.1, observers were instructed to judge the depth of the illusory surface at its edges. As the data show, when the inducer had a small crossed or an uncrossed disparity observers still indicated that the illusory surface was in front of the stimulus lines (probe settings were always at a crossed disparity). The lack of a significant effect of inducer disparity on the perceived depth of the occluder edges in these conditions suggests that: 1) the observers were performing the task correctly, estimating the depth of the illusory occluder at its edge, and not the depth of the inducer and 2) the constraints imposed by the viewing geometry restrict the biasing effect of the binocular feature on the edges of the illusory occluder.

Note that the difference between the estimates made for stimuli with different line widths (indicated by differently coloured lines) seems to decrease as the disparity of the inducer increases beyond the minimum specified by the geometric constraints. This is an expected effect since there is no reason to assume a different rate of biasing for stimuli with lines of different thickness.

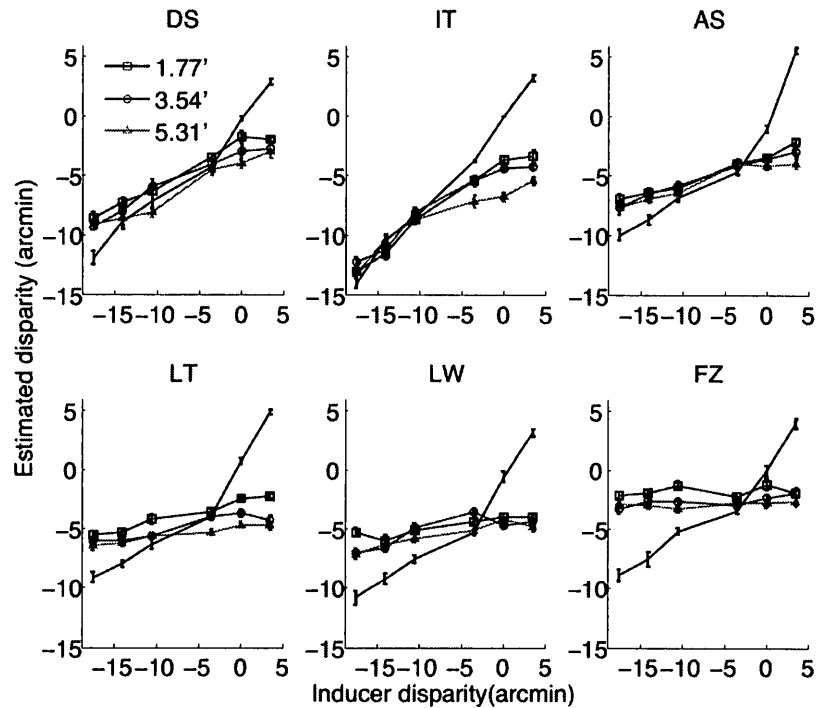


Figure 3.5: Results of Experiment 3.2 for six observers. Inducer disparity is plotted on the abscissa and the estimated disparity on the ordinate. The solid colored lines with different markers show the data for different line widths from the Occluder+Inducer part. The black line with no markers shows data from the Inducer-Only part. Error bars show +/- standard error of the mean.

3.4 Experiment 3.3 - Disparity bias in the stimulus of Liu et al. (1994)

Liu et al. (1994) also presented an example of illusory surface perception from monocular occlusions, using a different stimulus configuration. Their stimulus was described in detail in Section 1.3.2 of the Introduction and is shown in Figure 3.6. Liu et al. (1994) found that the perceived depth magnitude of the illusory rectangle increased as the width of the vertical part of the bracket increased, in accordance with the

minimum depth constraint. However, Gillam (1995) argued that while there were no matching vertical luminance edges, the visual system could match the horizontal luminance edges with disparities given by their endpoints. Liu et al. (1995, 1997) pointed out that the terminations of the horizontal edges in their stimulus have opposite contrast polarity and hence are not an optimal binocular matching primitive. They suggested instead that the letter corners could be matched and they demonstrated this phenomenon using computational modeling.

This stimulus represents a good test case for the limitations of the disparity bias demonstrated in Experiments 3.1 and 3.2. According to Gillam (1995) and Liu et al. (1994, 1997) depth in this stimulus can be based on both monocular occlusions and binocular disparity. Since occlusion geometry provides only partial constraints on the position of the illusory surface, it is likely that disparity represents a more reliable cue. If the visual system relies only on disparity in this stimulus, the perceived depth of the illusory occluder should not be strongly affected by a nearby binocular feature. On the other hand, if depth is perceived primarily on the basis of monocular occlusions there should be a depth biasing effect similar to that seen in Experiment 3.2.

3.4.1 Methods

Observers

One experimenter and five naive observers participated in the study. All observers had normal or corrected-to-normal visual acuity and at least 20'' stereoacuity as measured with the Randot stereoacuity test.

Procedure

The experiment consisted of the same three parts as Experiment 3.2. The procedure was also the same, except that in this case observers viewed the probe and test stimulus simultaneously (preliminary experiments showed that observers performed well under such conditions) as was the case in Liu et al. (1994) study. The presentation time was unlimited and observers pressed a button when they completed each trial. Observers completed the same three conditions outlined in Experiment 3.2: Occluder-Only, Inducer-Only and Inducer+Occluder.

Stimuli

Occluder-Only - The stimuli (3.6) used in this part of the experiment replicated the stimulus of Liu et al. (1994). The width of the the vertical portion of the bracket was one of 1.77', 7.08' or 12.39'. These widths covered the range used by Liu et al. (1994). The total size of the bracket was $2.06^\circ \times 1.47^\circ$ and the height of the gap inside the bracket was 1° . The bracket was black on gray background and was presented with zero disparity in the center of the screen. The disparity probe was an antialiased black disc with diameter of 17.7', positioned to the right of the stimulus. The disparity of the probe could be changed in 1.77' steps.

Inducer-Only - In this part of the experiment the bracket had the same parameters as in the Occluder-Only condition (side-bar width 7.08'), but without the monocular gap so that no illusory occluder was perceived. The inducer was a white disc, 10.6' in diameter, positioned vertically in the center of the bracket. The inducer disparity was one of 3.54', 0', -1.77', -7.08', -14.16', -17.7' or -21.24'. The choice of disparity values was governed by the same geometric considerations as in Experiment

3.2. The disparity probe was the same as in the Occluder-Only condition.

Inducer+Occluder This part of the experiment combined the stimuli from the other two conditions. The inducer disc was placed in the center of the illusory occluder stimulus (see Figure 3.6-C and -D). Three widths of the vertical positions of the bracket were used as in the Occluder-Only condition and seven inducer disparities as in the Inducer-Only condition. The disparity probe was the same as in the other two conditions.

Apparatus and statistical analysis

The apparatus and the statistical analysis for each of the conditions were the same as in Experiment 3.2.

3.4.2 Results and Discussion

Occluder-Only Since all observers showed the same pattern of results, the mean data were shown in Figure 3.7-A. As in the original experiment (Liu et al., 1994), there was a significant increase in estimated disparity as the width of the vertical bars was increased ($\chi^2 = 12, p < 0.01$). However, on average, the disparity was overestimated as it was in Experiment 3.2. In Liu et al.'s experiment disparity estimates were closer to the theoretical prediction (the minimum constraint), although one of their three observers seemed to underestimate and another to overestimate the perceived depth.

Inducer-Only As in the first part, the data were consistent across all observers and were combined in Figure 3.7-B. As expected, observers estimates increased significantly as the disparity of the inducer was increased ($\chi^2 = 12, p < 0.01$). Here the estimates were more accurate than in Experiment 3.2 which is expected since the

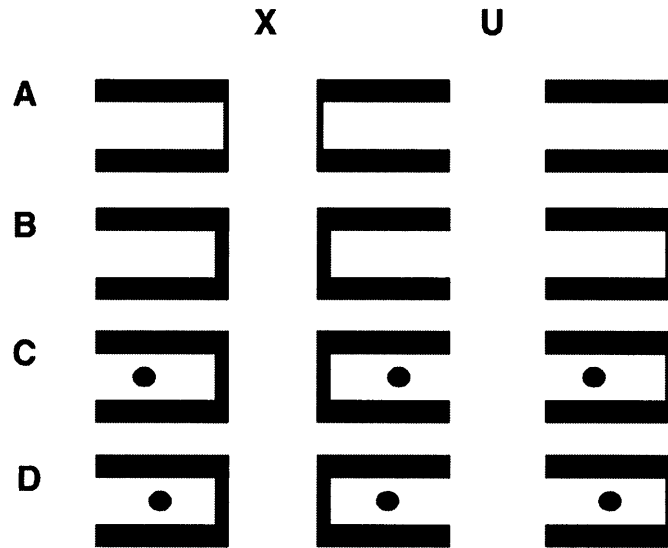


Figure 3.6: Examples of stimuli used in Experiment 3.3. The visual system interprets the absence of one of the side-bars as a monocular occlusion and creates a percept of an illusory occluding surface floating in front of a black rectangle. In (B) the side-bar is thicker than in (A) and hence the illusory occluder is perceived closer to the observer. In (C) a binocular disc (the inducer) with crossed disparity is placed in the center of the bracket. Most observers perceive the disc at a different depth than the illusory occluder. In (D) the inducer has a larger uncrossed disparity than that specified by the geometric constraints. The edges of the occluder in this case are still seen in front of the bars, but its center appears concave. In the actual stimulus, the background was gray and the inducer disc was white.

probe was presented concurrently with the stimulus.

Inducer+Occluder Data for this condition are shown in Figure 3.8. The black line without markers shows the results of the Inducer-Only task, that is, the estimated disparity of the inducer in the absence of the illusory surface. The solid colored lines with different markers represent the estimated disparity of the illusory surface in the presence of the inducer for different vertical bar widths. If the inducer had no effect on the perceived depth of the illusory occluder, there should be no change in the estimated disparity of the illusory occluder as a function of inducer disparity. Indeed, most of the observers showed little effect of the inducer on the perceived depth of

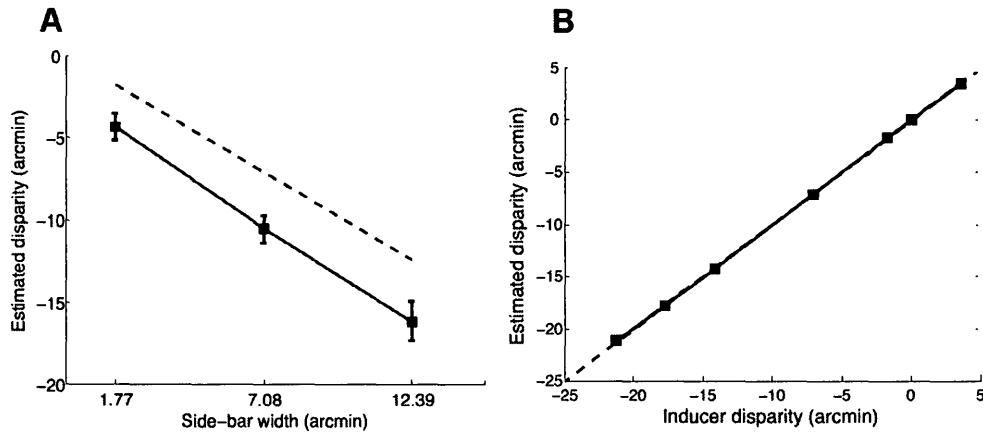


Figure 3.7: Results of Experiment 2.3. (A) Results of Occluder-Only part. The side-bar width is plotted on the abscissa and the estimated disparity on the ordinate. (B) Results of Inducer-Only part. Inducer disparity is plotted on the abscissa and the estimated disparity on the ordinate. The solid blue lines show the mean data for six observers and the dashed black lines show the theoretical disparity. Error bars show \pm standard error of the mean.

the occluder. However, observer DS showed a markedly different pattern of results. In her data there was a clear bias in the perceived depth of the illusory occluder in the presence of the inducer when the inducer had a larger crossed disparity than the minimum crossed disparity specified by the viewing geometry (mean slope - 0.6; all slopes differed from zero with $p < 0.1$).

The relatively small effect of the binocular inducer in the Liu et al. (1994) stimulus for 5 of the 6 observers suggests that these observers relied primarily on the binocular disparity signal in this stimulus. However, DS's data show that it is possible for the visual system to use monocular occlusion information to define the depth of the illusory surface under these conditions.

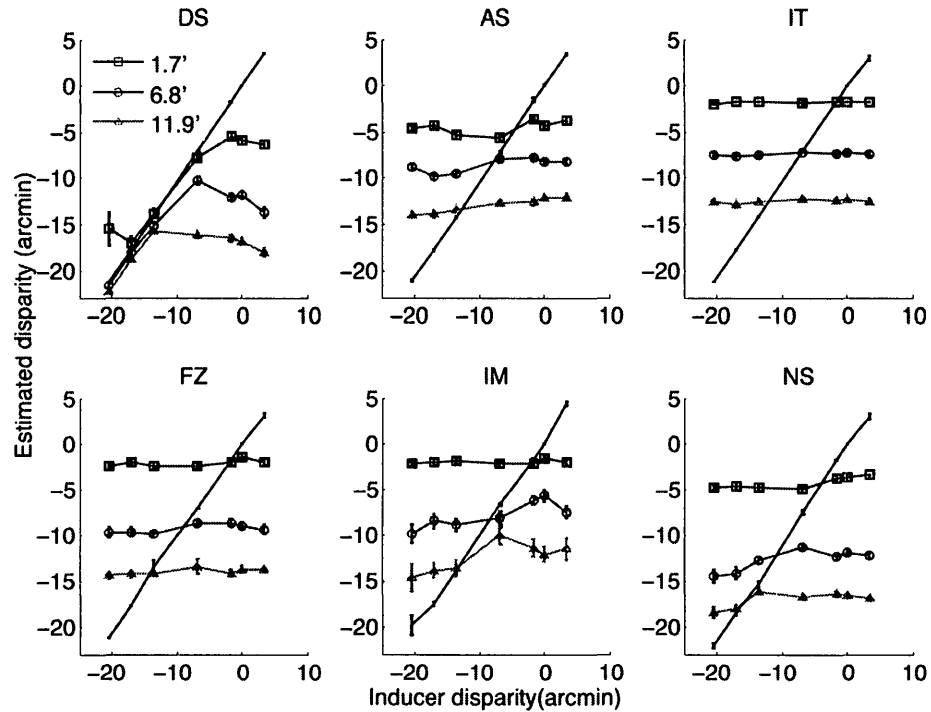


Figure 3.8: Results of Experiment 3.3 for six observers. Inducer disparity is plotted on the abscissa and the estimated disparity on the ordinate. The solid colored lines with different markers show the data from the Occluder+Inducer part for different side-bar widths. The black line without markers shows data from the Inducer-Only part. Error bars show +/- standard error of the mean.

3.5 Experiment 3.4 - Control

In Experiments 3.1-3.3 an implicit assumption was made that the perceived depth of surfaces defined by disparity could not be biased by a binocular feature placed in the center of the surface. Although this assumption makes intuitive sense, it must be verified empirically. Moreover, it is important to confirm that the disparity biasing exhibited by observer DS in Experiment 3.3 was indeed due to her reliance on depth from occlusions. It is possible that, for this observer, even surfaces defined

by unambiguous disparity could have had their depth biased by proximate features with a different disparity.

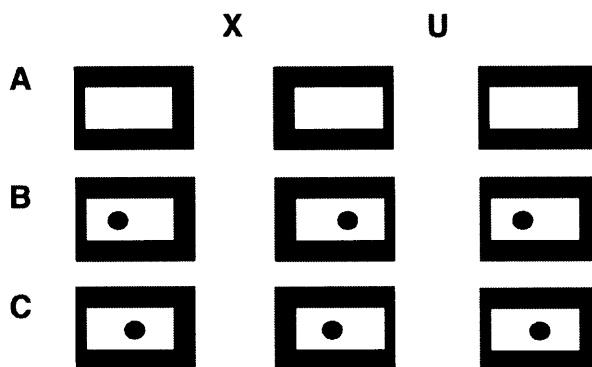


Figure 3.9: Stimuli used in Experiment 4. When fused, the stimulus is perceived as a small, white rectangle in front of a bigger black rectangle (in the actual experiment the background and the small rectangle were gray). In (B) a binocular disc (the inducer) with crossed disparity is placed in the center of the small rectangle. In (C) the inducer has a smaller crossed disparity than that of the small rectangle. In the actual stimulus, the black rectangle was placed on a gray background and the inducer was white.

3.5.1 Methods

Observers and apparatus were the same as in Experiment 3.3.

Stimulus

The stimulus, shown in Figure 3.9 was composed of a black rectangle subtending $2.06^\circ \times 1.47^\circ$ and a gray rectangle (same color as the background) subtending $1.6^\circ \times 1^\circ$. The disparity of the gray rectangle was one of $1.77'$, $7.08'$ or $12.39'$ and the black rectangle was presented with zero disparity at the center of the screen. The inducer was a white disc, $10.6'$ in diameter, positioned in the center of the gray rectangle. The

disparity range of the inducer and the disparity probe were the same as in Experiment 3.3.

Procedure

The observers were asked to adjust the disparity probe to match the disparity of the gray rectangle which appeared to lie in front of the black rectangle. As before, they were told to ignore the white disc. The disparity probe was presented simultaneously with the stimulus and the presentation time was not limited. Each stimulus condition was presented 20 times in random order. There were 21 different conditions (7 inducer disparities x 3 occluder disparities) and 420 trials in total. Observers completed this part in four sessions of 105 trials each.

Statistical analysis

The results of all observers were virtually identical, therefore their data were averaged for this analysis. Three regression models were fit, one for each surface disparity and the significance of the slopes was assessed using a randomization technique (see Experiment 3.2).

Results and Discussion

Figure 3.10 shows the results of Experiment 3.4. It is clear that the perceived depth of the small, gray rectangle was not affected by the inducer disparity for any of the observers, including DS. Statistical analysis showed no significant relationship between the inducer disparity and the estimated disparity of the rectangle. These results show that depth from disparity in the configuration used in Experiment 3.4

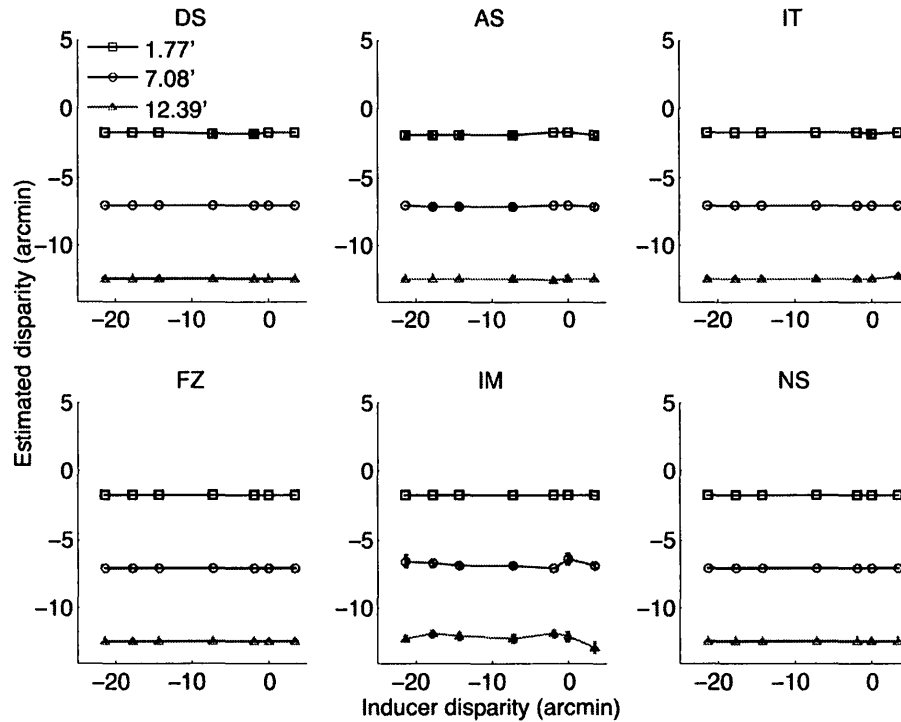


Figure 3.10: Results of Experiment 3.4 for six observers. Inducer disparity is plotted on the abscissa and the estimated disparity on the ordinate. Differently colored lines with different markers represent the disparity of the gray square. Error bars show +/- standard error of the mean

is not subject to the biasing demonstrated for depth from monocular occlusions in Experiments 3.1-3.3. The results also confirm that the pattern of results shown by observer DS in Experiment 3.4 was due to her reliance on monocular information.

3.6 General Discussion

These experiments show that the perceived depth of illusory occluders, induced by the presence of monocular regions, can be biased by a binocularly-defined feature in

several different illusory-occluder stimuli. This disparity-induced biasing was limited by the constraints specified by occlusion geometry. That is, a binocular inducer that had an uncrossed or a smaller crossed disparity than the minimum crossed disparity specified by the occlusion geometry, could not pull the perceived depth of the edges of the illusory occluder in the uncrossed direction. However, when the inducer had a larger crossed disparity than the minimum crossed disparity, then the perceived depth of the (whole) illusory occluder was biased toward the inducer. In Chapter 2 it was shown that when the depth of the illusory occluder was constrained in both directions, the binocular inducer had no effect on the perceived depth of the occluder. Taken together these findings show that the visual system is guided by geometric constraints in estimating qualitative and quantitative depth from monocular occlusions.

The disparity biasing demonstrated in the present experiments is also likely to be modulated by the size of the binocular inducer and the strength of its disparity signal. The bias created by the stimuli used in Experiment 3.1 was greater than that seen in Experiment 3.2 and the results of Experiment 3.1 were more homogenous. In Experiment 3.1, the inducer was a (relatively) large surface - larger than the adjacent illusory occluder - and its disparity was defined by multiple elements. In Experiment 3.2, the inducer was a single small disc - smaller than the illusory occluder and its disparity was defined by the vertical components of the circular contour. Consequently, this inducer's disparity signal was weaker than the signal of the inducer in Experiment 3.1. It is possible that intensifying the disparity signal of an inducer by increasing its extent, number and/or the length of its vertical edges might result in a stronger biasing effect.

Investigation of the biasing phenomenon in Experiment 3.3 revealed an interest-

ing tradeoff between monocular occlusion and disparity cues. The Liu et al. (1994) stimulus, used in Experiment 3.3, contains both monocular occlusion and binocular disparity cues to the depth of the illusory occluder (Gillam, 1995; Liu et al., 1997). Since monocular occlusions provide an ambiguous cue to depth in this stimulus (only the minimum possible depth of the illusory occluder is restricted), the observers were expected to rely on the binocular disparity signal to establish the depth of the illusory occluder. Grove et al. (2002) provided evidence to support this hypothesis when they examined the effect of the content of monocular zones on depth perception from occlusions in stimuli including those of Gillam and Nakayama (1999) and Liu et al. (1994). Although perceived depth in the former stimulus was reduced when the monocular regions were textured differently from the background, the depth in the Liu et al. stimulus was not. Five of the observers in Experiment 3.3 indeed relied on disparity with the Liu et al. stimulus and showed little biasing. However, one observer showed a strong bias, similar to her results in Experiment 3.2. Experiment 3.4 showed that the bias was not present for this or any other observer in stimuli where the occluding surface had luminance defined edges on both sides. However, the stimuli used in Experiment 3.4 had a strong disparity signal, while in Liu et al. (1994) stimulus the disparity signal was relatively weak. The weakness of the disparity cue might have caused observer DS to rely on the monocular occlusion information instead. Another example of the preference for monocular occlusions in stimuli with both cues present was provided by Hakkinen and Nyman (2001). They showed a disparity capture-like effect with Gillam and Nakayama (1999) stimuli overlaid on a wall-paper pattern. The binocular pattern was captured by the depth induced by monocular occlusions so that it was perceived at the depth of the illusory occluder. When considered along

with the results of Experiment 3.3, these findings suggest that some observers rely on monocular occlusions to determine the location in depth of a surface in stimuli where disparity information is present but is ambiguous or weak. Thus, monocular occlusions can serve as the primary cue to quantitative depth even in the presence of a disparity signal.

There have been several other cases in the literature in which depth percepts were initially assumed to be based on monocular information, but were later attributed at least in part to binocular disparity (Gillam et al., 2003; Pianta and Gillam, 2003b). Results presented in this chapter, based on a range of stimulus configurations, suggest that the presence or absence of the biasing phenomenon might be useful as a general 'litmus' test to determine whether depth magnitude in a given stimulus configuration is perceived primarily via monocular occlusions or binocular disparity. The degree to which a disparity-defined inducer can bias the perceived depth of a monocular feature or an illusory occluder in the unconstrained direction can reveal the extent to which the visual system relies on monocular geometry to specify depth magnitude.

Chapter 4

Quantitative depth from monocular occlusions - one or many mechanisms?

4.1 Objectives

Nakayama and Shimojo (1990) presented the first evidence that the visual system might use the minimum depth constraint to localize a monocularly occluded object in depth using two-object arrangements. Their stimulus is described in detail in Section 1.3. Figure 4.1 shows two possible arrangements of the stimulus - 1) the monocular bar is occluded from the rectangle and 2) the monocular bar is camouflaged against the rectangle. In both cases, as shown in Figure 4.1-B, the location in depth of the monocular bar is partially constrained by the line of sight to the eye that does not see the bar. According to the minimum depth constraint hypothesis, Nakayama and Shi-

mojo (1990) proposed that the perceived depth of the monocular bar should increase as occlusion width increases. This was found to be true in occlusion arrangements but not in camouflage arrangements.

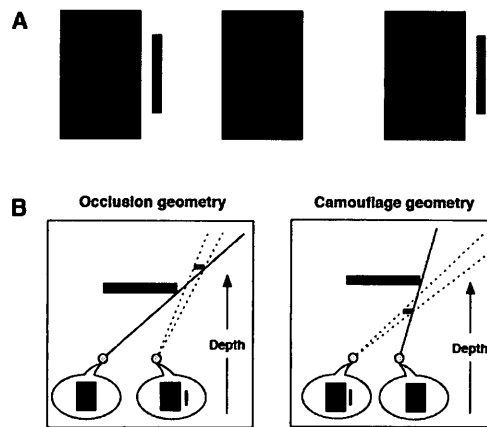


Figure 4.1: (A) Stimuli used in Nakayama and Shimojo (1990). For crossed fusion the left and the central images comply with occlusion geometry and the central and the right images with camouflage geometry. For divergent fusion the situation is reversed. (B) Occlusion and camouflage geometry. The position of the thin bar in depth is partially constrained by the line of sight from the eye that does not see the bar (solid black line). The minimum possible depth of the bar is determined by the distance between the outermost edges of the bar and the rectangle.

Gillam et al. (2003) proposed that quantitative depth in the stimuli used by Nakayama and Shimojo (1990) is simply an instance of double-matching, in which the monocular bar is matched to the edge of the occluding rectangle in the other eye (which is also simultaneously matched to itself). They tested this hypothesis by comparing perceived depth from two types of stimuli: 1) a stimulus with a rectangular textured binocular occluder and a thin monocular line and 2) the same stimulus but with the line replaced by a small monocular disc (see Figure 4.2). They argued that the disc was unlikely to be matched to the edge of the rectangular occluder

because of the differences in size and shape. They found that the monocular line stimulus produced quantitative depth percepts while the monocular disc stimulus did not. However, Gillam et al. (2003) used a different range of occlusion widths for the disc and the line stimuli, as discussed in Experiment 4.1, which could have affected their results.

As discussed in the Section 1.3 and shown in Chapter 2, quantitative depth based solely on occlusion geometry exists in other types of da Vinci stereopsis arrangements, such as illusory occluding surfaces and moving stimuli. Thus, the results of Gillam et al. (2003) suggest that da Vinci stereopsis is a stimulus dependent phenomenon since it relies on different mechanisms for the computation of quantitative depth in different occlusion configurations. However, parsimony favours a single mechanism for quantitative depth in all types of geometrically constrained occlusion stimuli.

This chapter describes three different experiments used to examine the source of depth percepts in two-object arrangements to understand whether da Vinci stereopsis relies on one type of mechanism or the mechanisms are stimulus dependent. Experiment 4.1 replicates the experiments of Nakayama and Shimojo (1990) and Gillam et al. (2003) to test the hypothesis of Gillam et al. (2003) using comparable conditions for all stimuli. In Experiment 4.2 the role of disparity in two-object arrangements is examined using the disparity bias 'litmus test' developed in Chapter 3 to differentiate depth from monocular occlusions and depth from disparity. Finally, the response of model disparity detectors to depth in two-object stereograms is investigated to understand whether these detectors can predict depth percepts in two-object stimuli.

4.2 Experiment 4.1 - Replicating Nakayama and Shimojo (1990) and Gillam et al. (2003)

In their experiments, Gillam et al. (2003) measured occlusion width as the distance from the edge of the occluder to the center of the monocular object instead of the outer edge of the monocular object (personal communication). Since in their experiments the monocular line was only 10" wide, the range of true occlusion widths when the line was the target was close to the widths specified in their article, 5' to 35'. However, for the disc that was 7' in diameter, the actual range of occlusion widths was 8.5' to 38.5'. This difference in test conditions is important since it has been shown that quantitative depth in two-object arrangements is perceived only over a narrow range of occlusion widths (Gillam et al., 2003; Hakkinen and Nyman, 1996; Nakayama and Shimojo, 1990). In particular, Gillam et al. (2003) found that even in their monocular line stimulus (Figure 4.2-B), quantitative depth was observed only for occlusion widths smaller than 12' – 15'. Therefore, it is possible that most of the occlusion widths they used with the disc stimulus were out of the range for quantitative depth in two-object arrangements. Experiment 4.1 replicates several conditions from Nakayama and Shimojo (1990) and Gillam et al. (2003), while carefully controlling occlusion width to equate the minimum predicted depth from occlusion for all monocular stimuli.

4.2.1 Methods

Observers

Two experimenters and four naive observers participated in this experiment. All observers had normal or corrected-to-normal visual acuity and at least 20'' stereoacuity as measured with the Randot stereoacuity test.

Apparatus

The apparatus was the same as in Experiments 3.2-3.4

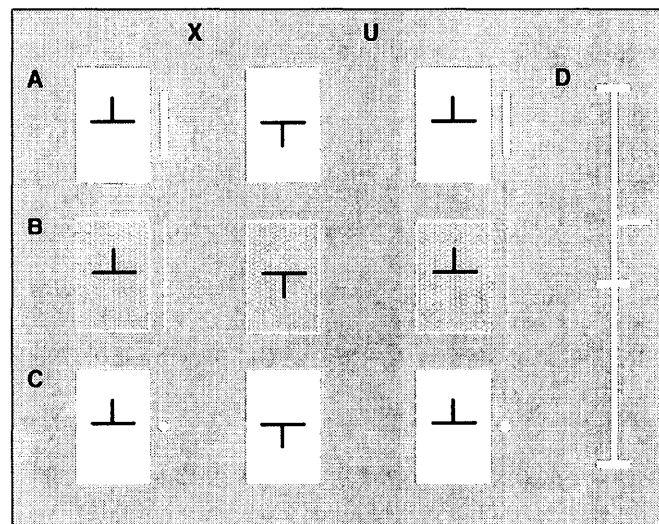


Figure 4.2: Examples of the stimuli and the ruler used in Experiment 4.1. (A) Stimulus which replicates that of Nakayama and Shimojo (1990) (B) Stimulus similar to that of Gillam et al. (2003). (C) The Nakayama and Shimojo stimulus with the monocular bar replaced with a monocular disc. For crossed fusion the left and the central images comply with occlusion geometry and the central and the right images with camouflage geometry. For divergent fusion the situation is reversed. (D) The virtual ruler used to measure perceived depth in Experiments 4.1 and 4.2.

Stimuli

Figure 4.2 shows the three types of stimuli used in Experiment 4.1. All three stimuli were composed of a rectangular binocular occluder and a smaller monocular object. All stimuli were drawn on a gray background.

Bar stimulus - this stimulus was identical, including angular size, to that used by Nakayama and Shimojo (1990) in their Experiment 2. It was composed of a solid, white binocular rectangle ($138' \times 418'$) and a solid, white monocular bar ($7.08' \times 331'$).

Line stimulus - this stimulus matched that used by Gillam et al. (2003) in terms of appearance. However, the dimensions of the occluder replicated those used by Nakayama and Shimojo (1990). That is, the occluder was the same size as in the bar stimulus but was comprised of a thin white outline $1.77'$ wide and filled with white random-dots each subtending $1.77' \times 1.77'$ (10% density). The monocular object in this case was a white line subtending $1.77' \times 418'$.

Disc stimulus - the occluder in this stimulus was the same as in the bar stimulus but the monocular object was a white disc $7.08'$ in diameter (equal to the width of the bar in the bar stimulus and similar in size to the disc used in Gillam et al. (2003)). The original occluder was chosen instead of the modified occluder used by Gillam et al. (2003) to permit a direct comparison with the stimulus of Nakayama and Shimojo (1990).

In all stimuli, the occluding rectangle was positioned at zero disparity in the center of the display and a black fixation cross was placed at its center. The vertical lines of the cross were Nonius lines ($3.5' \times 23'$) presented one to each eye, and the horizontal line ($42.5' \times 3.5'$) was presented to both eyes with zero disparity. In each case the monocular test object (bar, line or disc) was placed to the right of the

binocular occluder in the left or the right eye. When it was placed in the right eye, the configuration was consistent with occlusion geometry (see Figure 4.2). When the monocular object was placed in the left eye, the configuration was consistent with camouflage geometry. The monocular objects were only placed on one side of the rectangle since previous research has shown no effect of side (Gillam et al., 2003; Hakkinen and Nyman, 1996; Nakayama and Shimojo, 1990).

The distance from the right edge of the occluding rectangle to the right edge of the monocular object was varied in the same manner for all three stimuli. The occlusion width, corresponding to the minimum occlusion constraint, was 8.85', 12.39', 15.93', 19.47', 23.01' or 26.55'. This range encompassed the effective range for quantitative depth in two-object arrangements reported in the literature. In the binocular session (see below) the bar, line or disc was presented binocularly, 17.7' to the right of the rectangle. It was assigned either a crossed or an uncrossed disparity of 7.08' or 14.16'.

Procedure

Nakayama and Shimojo (1990) noted that the disparity probe they used for measurement, when presented simultaneously with the stimulus, affected the perceived depth of the monocular object in their experiments. On the other hand, sequential presentation of the stimulus and probe introduces a memory load to the task and increases response noise. Therefore, the observers were asked to use a virtual ruler presented simultaneously with the stimulus to indicate the perceived depth between the rectangle and the monocular object. The ruler was positioned 2.95° to the right of the center of the stimulus. It consisted of a vertical line subtending 3.5' × 531' bisected by a small horizontal line subtending 35.4' × 3.5' and another, adjustable, horizontal

cursor line ($35.4 \times 3.5'$). Observers moved the cursor along the ruler using a computer mouse (see Figure 4.2-D). They were instructed to match the distance between the central bisection mark and the cursor to the perceived depth of the monocular object relative to the binocular rectangle. Observers moved the cursor above or below the central bisection mark to indicate that the monocular object was seen behind or in front of the rectangle respectively. They also used this method to estimate depth in the binocular session. Observers were asked to fixate on the fixation cross while evaluating the depth in the stimulus and adjusting the ruler.

The three stimuli were shown in separate sessions whose order was randomized across observers. Each type of stimulus was shown 20 times for 240 trials per session (camouflage/occlusion \times 6 occlusion widths \times 20 repetitions). In each session, stimuli were presented in random order. Before each monocular stimulus session, observers completed a binocular session of 40 trials (4 disparities \times 10 repetitions) where the bar, line or disc was binocular and had either crossed or uncrossed disparity. The purpose of these binocular sessions was twofold, to train the observers in using the estimation method and to generate comparison data for the monocular conditions.

Statistical analysis

The data were first analyzed separately for the occlusion and the camouflage arrangements with a two-way repeated-measures ANOVA with occlusion width and stimulus type as factors. In a posthoc analysis, separate regression lines were fitted to the data corresponding to each stimulus type. The alpha level for all tests was 0.01.

4.2.2 Control Experiments

Three control experiments were performed in addition to the main experiment. Control Experiment 4.1-A was designed to rule out the influence of eye movements in the main experiment. In this experiment, two observers (one naive) performed a two-interval forced choice (2IFC) procedure to assess perceived depth with briefly presented stimuli (100 ms) (see Figure 4.3). Accurate binocular fixation was controlled and self-monitored with a fixation cross and Nonius lines. On each of the trials, one randomly-chosen interval contained a monocular bar or disc stimulus (Figure 4.2-A and Figure 4.2-C), with occlusion widths of 12.4', 15.9' or 19.5', and the other interval contained a corresponding binocular stimulus presented with disparity that varied according to the method of constant stimuli (25 repetitions of 7 disparities for each of the 3 occlusion widths). Observers were asked to indicate the interval in which the larger depth was perceived. A pilot experiment, in which stimuli were defined by binocular disparity, was conducted to confirm the feasibility of the task.

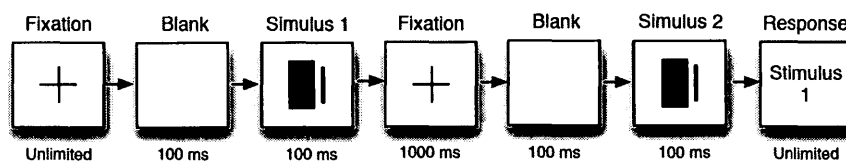


Figure 4.3: Timeline of Control Experiments 4.1-A and 4.1-B.

Control Experiments 4.1-B and 4.1-C assessed whether the estimation of the lateral position of the disc with respect to the rectangle was more difficult compared to the line and bar (a hypothesis that was proposed to account for some of the results of the main experiment). In Experiment 4.1-B the just noticeable difference (JND) was

measured in the lateral position of a binocular bar or disc with respect to the rectangle. The binocular disc or bar was presented with zero disparity (so that it appeared at the same depth as the rectangle) and the distance between the rectangle and the outer edge of the bar or disc was varied. Five observers (three naive) participated in a 2IFC discrimination procedure (same as in Experiment 4.1-A). On each trial, in one randomly-chosen interval, the distance was set at 15.9' and in the other interval the distance was one of 15.9', 16.8', 17.7', 18.6', 19.5', 20.4' or 21.2'. Observers were asked to indicate the interval in which the larger distance was presented. Experiment 4.1-C also used the bar and the disc presented dichoptically with zero disparity, but assessed the perceived magnitude of the lateral position judgments. The distance between the rectangle and the outer edge of the bar/disc was one of 12.4', 21.24' or 30.1'. Six observers (four naive) indicated the perceived distance from the edge of the rectangle to the outer edge of the bar or disc using the virtual ruler procedure used in the main experiment.

4.2.3 Results and Discussion

The left graph in Figure 4.4 shows the mean results for the sessions where the bar, line and disc were monocular. Depth estimates, expressed as equivalent disparity, are shown as a function of occlusion width, arrangement and monocular object type. Distances estimated using the virtual ruler were converted to equivalent disparities using a standard equation relating on screen disparity and depth (see Figure 1.3-B). The right side of the graph shows the data for the occlusion arrangement while the left side of the graph shows the data for the camouflage arrangement. The black solid lines show the predicted disparity the estimates should follow according to occlusion

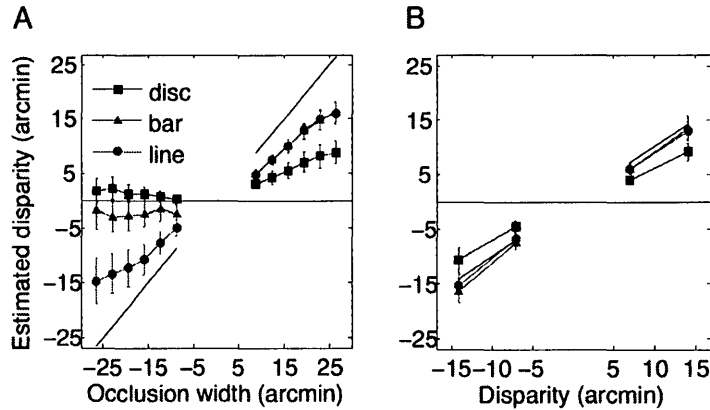


Figure 4.4: Results of Experiment 4.1 averaged across six observers. (A) Shows the results of the session with monocular objects and (B) with binocular objects. In (A) data for the occlusion arrangement are shown on the right and the data for the camouflage arrangement on the left. In (B) crossed disparities are shown on the left and uncrossed on the right. Estimated disparity for both graphs is plotted as a function of stimulus type and occlusion width/disparity. Red lines with triangular symbols represent the bar stimulus data, green lines with circular symbols represent the line stimulus data and blue lines with square symbols represent the disc stimulus data. The black solid lines show the predicted disparity. Error bars show +/- one standard error of the mean.

geometry.

For the bar and line stimuli in the occlusion arrangement, perceived depth increased with increasing occlusion width (main effect of occlusion width $F(5,25)=75.8$, $p < 0.001$; bar slope - $b=0.65\pm 0.054$ CI, line slope - $b=0.67\pm 0.048$ CI). On average, observers underestimated the depth relative to the predicted depth (slopes significantly different than 1). Two of the observers (LT and LW) followed the prediction closely, while the other four perceived less depth for both line and bar stimuli. These results were consistent with Nakayama and Shimojo (1990) where some observers' depth estimates were smaller than the minimum predicted from the occlusion geometry. In the camouflage condition quantitative depth on average was seen for the line stimulus ($b=0.65\pm 0.09$ CI) but not for the bar stimulus ($b=0.005\pm 0.1$ CI) (interac-

tion of stimulus and occlusion width $F(10,50)=6.46$, $p < 0.001$), in agreement with Gillam et al. (2003) and Nakayama and Shimojo (1990). On average, the bar was perceived slightly in front of the occluder and variability was noticeably higher in this condition. Some observers always perceived the bar to lie in front of the occluder, others always saw it behind or at the same depth; for some observers the percept changed from trial to trial.

Following the logic of Gillam et al. (2003), the disc stimulus is the crucial test of the double matching hypothesis. As shown in Figure 4.4, the mean perceived depth of the disc increased significantly as occlusion width increased ($b=0.35\pm 0.056$ CI). Consistent with the existing literature, no increase in depth was observed in the camouflage arrangement for the disc stimulus ($b=-0.1\pm 0.06$ CI).

It could be argued that some of the observers did not actually see quantitative depth in the disc stimulus and based their depth judgments purely on the occlusion width. This is highly unlikely since: 1) a similar strategy should have been used in the camouflage condition and that is clearly not the case. 2) The disparity estimates would be much smaller than observed since the estimated depth would be approximately equal to the actual separation of the monocular feature from the rectangle and not to the depth predicted from that separation.

Another potential explanation for these results is that the disc stimulus was more prone to double-matching than the disc stimulus of Gillam et al. (2003), since here the occluder and disc are both monochromatic. This possibility was evaluated by replacing the occluder in the disc stimulus with an occluder similar to that used by Gillam et al. (2003) (same occluder as in the line stimulus) and re-running the experiment with two observers (AC and IT). Both observers showed statistically significant

quantitative depth in the occlusion arrangement with this new stimulus, confirming the original findings.

Finally, quantitative depth in the monocular configurations could be based on alternating fixation. Observers could estimate depth from the difference in vergence angles created by alternating fixation between the occluder and the occluded object (possibly without awareness) or between the occluded object and the ruler (Enright, 1991). Since viewing time was unlimited, observers could potentially shift their gaze during a trial, despite instructions to maintain fixation. Further, the ruler was located 1.75° away from the edge of the occluder, which could potentially prompt the observers to break fixation. However, results of the control Experiment 4.1-A suggest that eye movements were not an important factor. In this experiment stimuli were presented for only 100 ms so observers should not have had enough time to initiate vergence eye movements while the stimuli were visible (e.g. Yang et al. (2002)). This procedure compared perceived depths from binocular disparity and monocular occlusions, to estimate the 'equivalent perceived disparity' of the monocular occlusion stimuli. For each monocular occlusion width the point of subjective equality (PSE) was estimated as the 50% point in the psychometric function (cumulative normal) fit to the data and plotted in Figure 4.5. It is clear that the perceived depth of the monocular features (bar and disc) increased significantly as a function of the occlusion width. Consequently, alternating fixation could be ruled out as an explanation of the results of Experiment 4.1. Interestingly, as in Experiment 4.1, perceived depth in Experiment 4.1-A was less than the depth specified by the minimum constraint for both stimuli ¹. This suggests that the underestimation observed in the monocular

¹It appears that, unlike Experiment 4.1, in Experiment 4.1-A the perceived depth of the disk stimulus was not much smaller than that of the line stimulus. However, in Experiment 4.1 depth

sessions of Experiment 4.1 (Figure 4.4-A) is not due to the estimation method used.

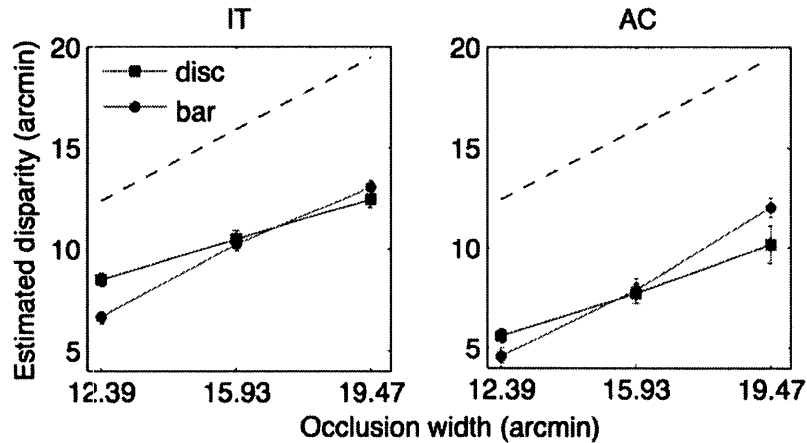


Figure 4.5: Results of control Experiment 4.1-A. PSEs are plotted as a function of occlusion width. Blue lines with square markers represent data for the disc stimuli and the green lines with round markers represent data obtained using the bar stimuli. Error bars show the ± 1 standard deviation computed by bootstrapping. The dashed black lines show the predicted depth according to the minimum constraint.

Although in Experiment 4.1 the disc stimuli in the occlusion arrangement produced quantitative depth percepts, the magnitude of perceived depth was significantly smaller than observed for the line and the bar stimuli (interaction of stimulus and occlusion width $F(10,50)=4.8$, $p < 0.001$; see slope statistics above). Interestingly, in the binocular session, when the bar, line and disc were binocular and had either crossed or uncrossed disparity, there was an underestimation of depth in the disc stimuli (see right graph of Figure 4.4) as well. The left side of the graph shows the

was estimated using the same virtual ruler for both stimuli, hence depth estimates were directly comparable. In Experiment 4.1-A the depth in the monocular stimulus was judged using a binocular version of the stimulus (bar or disk). Therefore, the properties causing the difficulty estimating the location of the disc in comparison to the bar, as reported in Experiments 4.1-B and 4.1-C, were present in both the target and the probe. This masked the differences in perceived depth in the two types of stimuli.

crossed disparity condition and the right side shows the uncrossed disparity condition. In both cases, observers see less depth when viewing the disc stimulus than the other two stimuli (interaction of stimulus and disparity - uncrossed $F(2,10)=7.2$, $p = 0.01$, crossed $F(2,10)=6.9$, $p = 0.012$ [at $\alpha = 0.05$]). Regression analysis showed that slopes for the line and bar stimuli for both uncrossed (bar $b=1.05\pm0.19$, line $b=1.0\pm0.17$, disc $b=0.72\pm0.15$) and crossed (bar $b=1.21\pm0.22$, line $b=1.21\pm0.21$, disc $b=0.84\pm0.22$) disparities were significantly different from zero and not significantly different from one (the slope of the predicted depth). For the disc stimulus the slope was significantly different from one for uncrossed disparities but not for the crossed.

It is possible that quantitative depth perceived in the monocular disc stimulus is smaller than in the other two conditions because this stimulus does not provide a reliable double-match. However, this does not explain why the binocular disc stimulus yields less perceived depth as well. A different account that encompasses both results appears more likely. Both depth from disparity and depth from monocular occlusion/camouflage in our stimuli require an accurate estimation of the relative position of the bar, line or disc with respect to the rectangle (or the bar, line or disc in the other eye). Due to its small spatial extent it might be harder to estimate the position of the disc compared with the bar or the line. Combined with the visual system's general bias towards smaller disparities (Goutcher and Mamassian, 2005) this might result in smaller perceived depth for the disc. Results from two control experiments support this hypothesis.

Control Experiment 4.1-B showed that the mean JND for the position of the disc stimulus ($M=1.84'$ $SD=0.28'$) was significantly larger ($t(4)=7.5$, $p < 0.01$) than the

mean JND for the position of the bar stimulus ($M=0.85'$ $SD = 0.17'$). The larger JND for the disc stimulus implies noisier estimates of its location relative to the rectangle. More noise in the disc condition could degrade depth estimates both from disparity and monocular occlusions compared to the bar stimulus.

In control Experiment 4.1-C, the perceived lateral distance between the disc or bar was directly compared with the edge of the rectangle using the virtual ruler (as in Experiment 4.1). Repeated-measures ANOVA confirmed that for all observers the perceived lateral distance of the disc or bar from the rectangle increased with physical distance ($F(2,10)=178.8$, $p < 0.001$) but was consistently larger for the bar than for the disc ($F(1,5)=29.38$, $p < 0.01$). The increase in estimated distance with true distance was greater for the bar than for the disc (interaction $F(2,10)=6.4$, $p = 0.01$). The finding that the perceived width of the 'occlusion' region is smaller for the disc stimulus than for the bar stimulus maps well onto the underestimation of perceived depth in the monocular disc stimulus in comparison to the bar stimulus in Experiment 4.1 (Figure 4.4-A). It also could explain the underestimation of depth observed with binocular disc stimuli since smaller perceived separations of the disc in the two eyes could translate into smaller disparities.

4.3 Experiment 2.4

This experiment was designed to probe the origin of depth in two-object arrangements using the disparity bias 'litmus test' developed in Chapter 3. There is some evidence in the literature that the depth of monocular objects in two-object arrangements can be biased by a nearby binocular feature. Nakayama and Shimojo (1990) noted

that the disparity probe they used affected the perceived depth of the monocular object. Similarly, Hakkinen and Nyman (1996) found that an additional binocular surface placed above a monocular dot in a two-object arrangement biased its perceived location. However, neither of these studies asked if this bias was specific to depth from monocular occlusions and so did not assess the bias in binocular versions of their stimuli. This experiment examines whether the depth of the monocular bar, line or disc used in Experiment 4.1 could be biased by an adjacent binocular surface. It was also assessed whether a similar bias was present when the bar, the line or the disc was presented binocularly. If the bias existed only for the monocular versions of the stimuli, this would suggest that depth percepts in the two cases were based on different mechanisms, namely monocular occlusion geometry in the first case and disparity in the second.

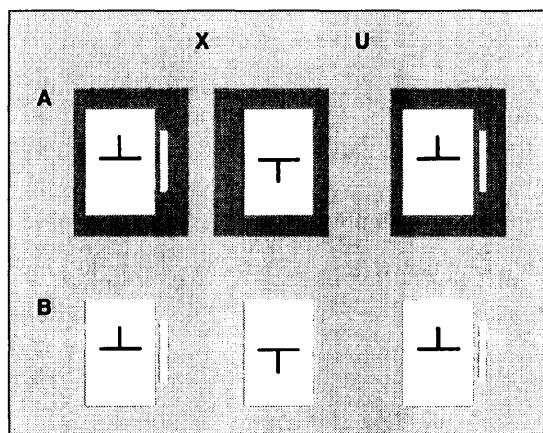


Figure 4.6: (A) Example of stimuli used in Experiment 4.2. The Nakayama and Shimojo (1990) stimulus is modified by adding an inducing surface with uncrossed disparity with respect to the occluding rectangle. (B) The same stimulus but without an inducing surface is presented for comparison.

4.3.1 Methods

The observers and apparatus were the same as in Experiment 4.1; however, only the occlusion arrangement was tested in this experiment. The occlusion stimuli (bar, line and disc) were modified by adding a uniform gray surface behind the occluding rectangle in depth. The surface subtended $4.07^\circ \times 7.25^\circ$ and had a disparity of $17.7'$. Both the occluder and the monocular features were located within the boundaries of the additional surface (see Figure 4.6). The observers were asked to ignore this additional surface and to perform the estimation task as in Experiment 4.1. Each session was preceded by a set of binocular trials where the bar, the line and the disc had uncrossed disparities of $7.08'$, $14.16'$, $21.24'$ or $28.32'$. The experiment was completed in three sessions, one for each type of feature (bar/line/disc), in random order.

4.3.2 Results and Discussion

Results from this experiment, along with these of Experiment 4.1, are presented in Figure 4.7. The mean depth estimates in Experiment 4.2 for monocular features (Figure 4.7-A dashed lines) are substantially larger than those obtained in Experiment 4.1 (Figure 4.7-A solid lines) for all feature types (this pattern was consistent across observers). However, there is little difference in the perceived depth of binocular features between the two experiments (Figure 4.7-B). To quantify the effect of the additional surface on the perceived depth in Experiment 4.2, the mean bias was computed for stimuli with monocular and binocular features. It was computed for each observer individually first, by subtracting the depth estimates for each occlusion width/disparity and stimulus type obtained in Experiment 4.1, from the depth

estimates obtained in Experiment 4.2 and then averaging these differences across occlusion width/disparity. The means for each stimulus type were then averaged across observers to obtain a final bias measurement for each condition. As shown in Figure 4.7-C, the addition of a binocularly-defined surface, behind the two-object configuration, biases the perceived depth of all monocularly defined features but not the binocular features. A two-way repeated-measures ANOVA on the mean bias data with stimulus type and experiment type (monocular/binocular) as factors, showed a significant main effect of experiment ($F(1,5)=60$, $p < 0.001$) and no significant effects of stimulus type ($F(2,10)=0.3$, $p=0.76$) or an interaction between the factors ($F(2,10)=0.3$, $p=0.76$). In post-hoc analyses, paired t-tests showed that for all three stimuli types the bias in the monocular and the binocular cases differed significantly (bar $t(5)=6.7$, $p<0.01$; line $t(5)=3.4$, $p=0.02$ [at $\alpha = 0.05$]; disc $t(5)=4.6$, $p<0.01$). In addition, one-sample t-tests showed that, for the binocular stimuli, the bias was not significantly different from zero for the bar and the line stimuli (bar $t(5)=0.6$, $p=0.55$; line $t(5)=1.1$, $p=0.3$) but was significantly different from zero for the disc ($t(5)=3.7$, $p=0.013$ [at $\alpha = 0.05$]). All t-tests were confirmed with a non-parametric Wilcoxon test.

These data show that, as in the experiments of Chapter 3, the depth of monocular but not binocular features can be easily biased in the direction of surrounding binocularly-defined features. Importantly, all three types of stimuli, (bar, line and disc) show a similar pattern of bias, suggesting that the same process, based on occlusion geometry, is used to assign depth.

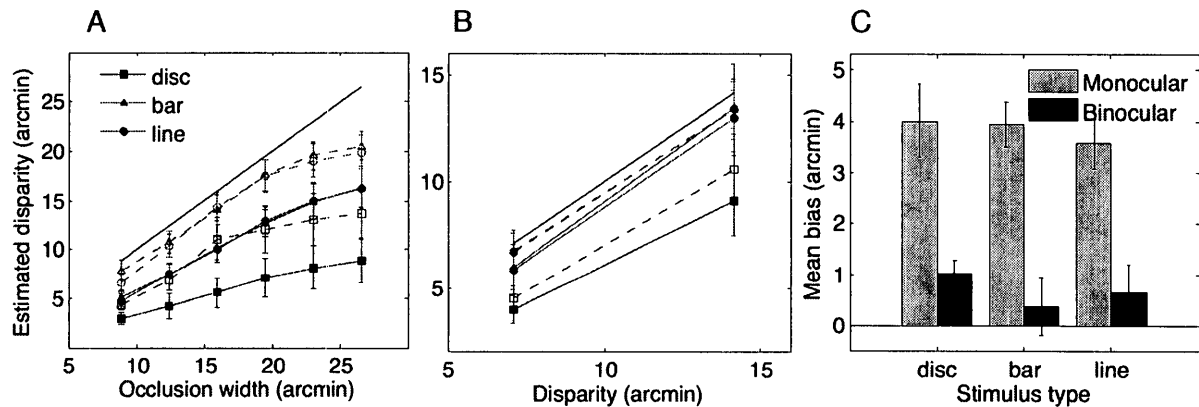


Figure 4.7: Results of Experiment 2. (A) Mean ($n=6$) disparity estimates obtained in Experiment 4.1 (solid lines, filled symbols) with monocular features are shown alongside those obtained in Experiment 4.2 (dashed lines, open symbols). (B) Same as in (A) but for binocular features. For both (A) and (B) estimated disparity is plotted as a function of occlusion width or disparity and differently colored lines with different markers correspond to different features (disc/bar/line). Black solid lines show predicted disparity. (C) Mean bias (see text)($n=6$). Light gray bars show the bias for monocular features and dark gray bars show the bias for binocular features. Error bars for all graphs show \pm one standard error of the mean.

4.4 Computational Simulations

In this experiment, a computational approach was used to assess the origin of quantitative depth percepts in da Vinci stereopsis. The double-matching hypothesis implies that a correspondence is made between the occluding and the occluded object. Any useful definition of a match implies a degree of similarity between matched images, which, in the simplest case, is a correlation. Moreover, it is widely believed that disparity detectors in V1 essentially perform a cross-correlation of the right and the left images (Banks et al., 2004; DeAngelis et al., 1995). Although V1 disparity detectors were found not to be sufficient for stereopsis, they are believed to be the first step in computing disparity for the perception of depth as well as for maintaining fusion

(Cumming and DeAngelis, 2001). To assess how well image correlation signals would predict depth percepts in two-object arrangements, a windowed cross-correlation algorithm was applied to the stimuli used in Experiment 4.1.

4.4.1 Methods

The sum of square differences (SSD) was used as a match metric (for review see Scharstein and Szeliski (2002)). The disparity that generated the maximum correlation value (minimum SSD) was chosen as the estimated disparity for each pixel. In cases where more than one disparity generated the maximum correlation value (i.e. multiple identical peaks), three methods of disparity assignment were compared:

1. The mean of the peak correlation disparities was computed.
2. The median of the peak correlation disparities was computed.
3. The disparity closest to zero was selected (zero bias).

Based on this algorithm a disparity estimate was obtained for each pixel and disparity maps were generated. A correlation goodness-of-match metric for each pixel was computed by dividing the peak correlation value of this pixel by the largest correlation response in the whole image. A correlation reliability metric for each pixel was computed as the difference between the normalized peak correlation responses for this pixel and the next largest response (i.e. two or more disparities with the maximum correlation value would generate a reliability of 0). The reliability metric is similar to the ‘peak ratio metric’ used to predict the locations where potential false matches could be made (Egnal et al., 2004; Little and Gillett, 1990). To accommodate disparity signals at different scales, three correlation window sizes were used 4, 10 and 20

px and two disparity ranges ± 20 px and ± 30 px². The algorithm was also applied separately from the left to the right image and from the right to the left image (the direction of the computation has been shown to impact the detection of monocular areas Egnal and Wildes (2002)). In total there were 36 different conditions for each stereogram (3 methods of depth estimation x 3 window sizes x 2 disparity ranges x 2 source images). Four types of stereograms were used:

1. Monocular bar stimulus from Experiment 4.1 (Figure 4.2-A)
2. Monocular disc stimulus from Experiment 4.1 (Figure 4.2-C)
3. Binocular bar stimulus from Experiment 4.1
4. Monocular bar stimulus with a black rectangle and white bar on a grey background (opposite contrast polarity).

The monocular stimuli were all consistent with the geometry of occlusion and the right edge of the monocular bar and disc was 7, 12 or 18 px away from the occluder. The binocular bar stimulus was identical in all respects to the monocular bar stimulus except that the bar was presented to both eyes and had a disparity of 4, 8 or 12 px. The binocular stimulus was used to compare the performance of the algorithm on monocular occlusion stimuli with its performance on similar binocular stimuli. The opposite polarity bar stimulus was used to demonstrate the dependence of the algorithm's response on stimulus properties.

²Some disparity selective neurons exhibit size-disparity correlation. That is smaller RFs are associated with tuning to smaller disparity. However this is a restriction that exists only for phase-shift neurons (see Chapter 6). The model neurons used here emulate position-shift neurons, whose disparity selectivity is not constrained by their size.

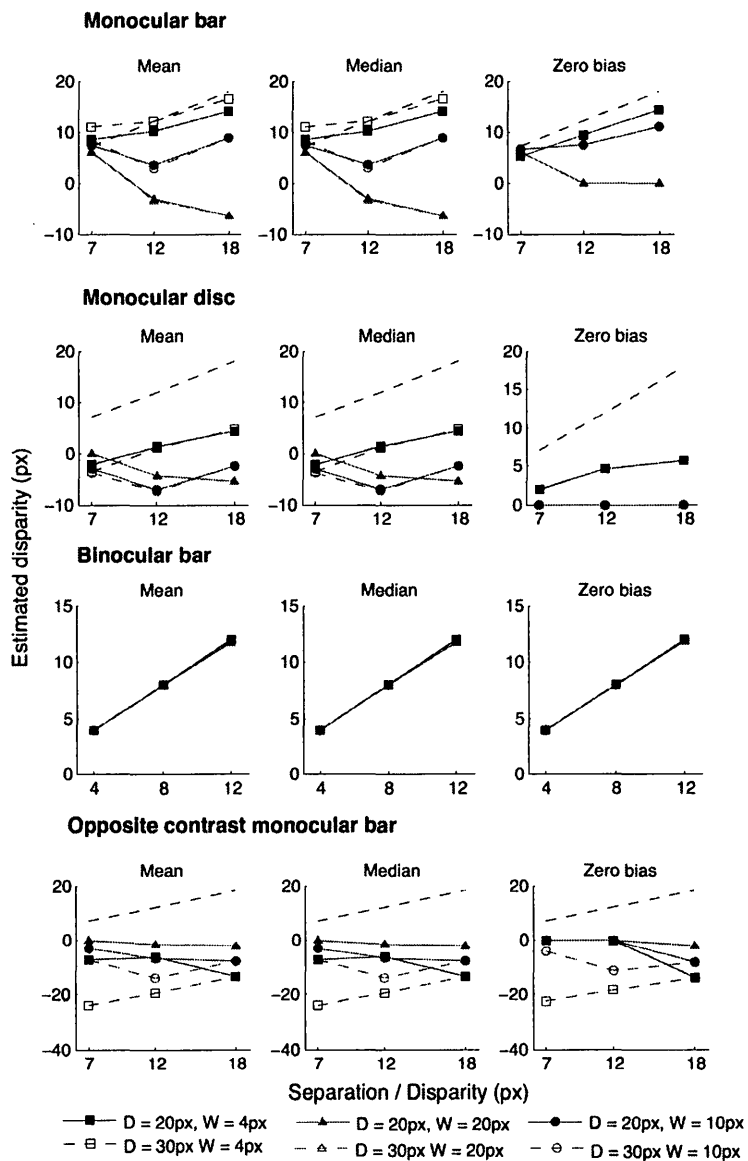


Figure 4.8: Results of the simulations for four types of stimuli (rows). The columns correspond to the three methods of disparity assignment and different correlation functions correspond to different window sizes and disparity ranges. For the monocular stimuli (rows 1,2 and 4) the mean estimated disparity of the pixels comprising the monocular object in the right eye is plotted as a function of occlusion width. For the binocular bar stimulus (row 3) the mean estimated disparity for the pixels comprising the bar in the right eye is plotted as a function of its disparity. Black dashed lines show the correct disparity (or theoretical disparity). Note the ordinate scale varies on different plots.

4.4.2 Results and Discussion

Figure 4.8 shows the results of the simulations for right-to-left cross-correlation. These results were chosen since the monocular features were in the right eye's image and thus right-to-left application gave better depth estimates. The left-to-right case is discussed below. Estimated depth for monocular objects is highly dependent on the parameters of the cross-correlation algorithm (Figure 4.8 first and second rows). The estimates vary widely as a function of window size, disparity range and disparity assignment method. No combination of parameters yields estimates close to those predicted by geometry (dashed black line). The combination that shows results most similar to the psychophysical data of Experiment 4.1 for both types of monocular objects (uncrossed depth that increases with occlusion width and is somewhat smaller than the predicted depth) combines the smallest correlation window (4 px) and the zero bias disparity assignment method (two top leftmost graphs, blue lines with square symbols). Other parameter combinations produce depth estimates that either do not increase with occlusion width or yield crossed or overly large disparities.

On the other hand, depth from disparity in the binocular bar stimulus (Figure 4.8 third row) is very robust to algorithm parameters. All methods, window sizes and disparity ranges yield the same estimated disparities, which are virtually identical to the predicted values. Importantly, disparity signal strength and reliability were found to be low for the monocularly occluded features and high for the binocular bar under all conditions. Moreover, the left-to-right simulation provided a set of depth estimates for the monocular regions that differ from those shown in Figure 4.8. However, in the case of the binocular stimulus, the estimates for the left-to-right simulation are the same as those for the right-to-left. In fact many stereo algorithms use what is called

a left-to-right check (see Chapter 6), comparing the output of the left-to-right to right-to-left applications, to identify monocularly occluded zones (Egnal and Wildes, 2002).

The ability of the correlation algorithm to match the monocular bar and disc to the occluder in the other eye is not surprising since cross-correlators are designed to respond to somewhat similar features in the two eyes (which is why they produce false matches in binocular areas). However, as these simulations show this signal is very noisy, unstable and highly dependent on the parameters of the detectors. The algorithm is also very sensitive to stimulus properties. As shown in the bottom row of Figure 4.8, no veridical quantitative depth can be predicted by any parameter combinations in the opposite contrast polarity bar stimulus. At the same time, observers perceive as much depth from this stimulus as they do from the same contrast polarity stimulus (Nakayama and Shimojo, 1990).

To conclude, the cross-correlation output for monocularly occluded regions in two-object configurations is noisy and parameter dependent. This contrasts with the vivid and robust quantitative depth experienced by human observers for both the monocular bar and dot stimuli and suggests that estimating depth in monocular regions requires more sophisticated mechanisms than simple correlation-based disparity detectors. These mechanisms might include separate monocular occlusion detectors that identify monocularly occluded areas (Egnal and Wildes, 2002), as well as higher-level neurons that would integrate information from disparity detectors and monocular occlusion detectors to establish a depth map.

4.5 General Discussion

4.5.1 The source of quantitative depth in two-object arrangements

The main focus of this set of experiments was to identify the source of quantitative depth in two-object arrangements in order to understand the mechanisms involved in da Vinci stereopsis. Several explanations for quantitative depth in two-object arrangements, like that used by Nakayama and Shimojo (1990), have been proposed. Below each explanation will be considered in turn and the most likely scenario will be discussed given the results of experiments reported in this chapter as well as data from the literature.

Double-matching

Gillam et al. (2003) proposed that depth in two-object arrangements can be accounted for by conventional disparity, assuming the edge of the occluder is matched to both itself and an edge of the monocular object in the other eye. Evidence for double-matching is a matter of debate for stereopsis in general (see Howard and Rogers (2002) pp. 127-137). More specifically, for two-object arrangements, a number of observations from the current data, and the existing literature, argue against this hypothesis. First, Experiment 4.1 suggests that the failure to find quantitative depth in the monocular disc stimulus in Gillam et al. (2003) could be attributed to the fact that the occlusion widths they used exceeded the range of da Vinci stereopsis. Following their logic, the fact that in Experiment 4.1 quantitative depth percepts were obtained using the disc stimulus argues against the double-matching hypothesis.

Second, it is well known that binocular disparities up to 1-2 degrees can yield

quantitative depth percepts (for review see Wilcox and Allison (2009)). Consequently, it is reasonable to assume that if depth perception in two-object arrangements relies on disparity it should be possible to obtain quantitative depth percepts for a wide range of occlusion widths. Thus, the very limited range of occlusion widths over which quantitative depth in two-object arrangement can be perceived (up to only 15' – 30' depending on the stimulus) is inconsistent with disparity based depth.

Third, Experiment 4.1 and Nakayama and Shimojo (1990) also found no evidence of quantitative depth in the camouflage condition for the bar and the disc stimuli. If double-matching takes place in two-object arrangements, quantitative depth should be seen in the camouflage condition since crossed and uncrossed disparities are treated (relatively) equally by the visual system. Quantitative depth was seen by five out of six observers in the camouflage condition with the line stimulus, which could be interpreted as a sign of double-matching. However, given the other arguments in this section, the results of Experiment 4.2, as well as the fact that occlusion geometry also predicts quantitative depth from camouflage, it does not seem likely that double-matching is the source of depth percepts in the line stimulus.

Fourth, in Experiment 4.2 I demonstrated that the perceived depth of all three monocular objects in two-object arrangements could be biased towards a far binocular surface, while the perceived depth of identical binocular objects was resilient to biasing. It could be argued that the disparity signal in two-object arrangements is weak due to double-matching and hence is more prone to influences from nearby binocular surfaces. However, Experiment 3.3 showed that even in stimuli with weak disparity signals most observers do not exhibit biasing.

Finally, computational simulations showed that simple model disparity detectors

produce a weak, noisy and parameter-dependent depth signal for monocular regions in two-object arrangements. This pattern is different from the stable, robust response to similar binocular stimuli. This discrepancy suggests that, although quantitative depth can sometimes arise in monocular regions due to fortuitous parameter choices in simple disparity detector models, the vivid depth percepts obtained in such stimuli are more likely to depend on a more sophisticated set of mechanisms. In sum, the weight of the evidence from the experiments in this chapter and the literature does not favour a double-matching account of depth percepts from two-object arrangements.

Sequential matching

Even in the absence of double-matching, depth in two-object arrangements could still be based on sequential stereopsis (Enright, 1991). When an observer fixates on the occluder, the edges of the occluder would be fused and the monocular object would be unmatched. When an observer fixates on the monocular object, it could be matched with the nearest edge of the occluder, leaving the occluder itself unmatched. The observer could estimate the relative disparity between the two objects based on the change in the vergence angle as they fixate at each location in turn. Control Experiment 4.1-A was designed to address this issue, and the results showed that quantitative depth is still seen in two-object arrangements when fixation is controlled and exposure durations are too brief to complete a vergence eye movement. The amount of depth perceived by both observers in Control Experiment 4.1-A was similar (and sometimes slightly exceeded) that estimated in Experiment 4.1. Hence this explanation can be ruled out.

Occlusion geometry

The explanation for da Vinci stereopsis offered by Nakayama and Shimojo (1990)

was based on occlusion geometry. They argued that the visual system could use the constraints on the minimum depth imposed by occlusion geometry to assign depth to the monocular objects in such stimuli. Data from Experiments 4.1 and 4.2 are consistent with this explanation as are the results of most other studies that used two-object arrangements (Hakkinen and Nyman, 1996; Nakayama and Shimojo, 1990). Experiment 4.1 shows that, for the majority of observers, there is quantitative depth in the disc stimulus in accordance with occlusion geometry. Experiment 4.2 provides new evidence that depth percepts in these stimuli are more likely to stem from occlusion constraints than from disparity since the depth of the monocular features in these arrangements can be biased by nearby disparity unlike the depth of similar binocular features.

There is a growing body of evidence that the visual system uses the constraints imposed by occlusion geometry to assign quantitative depth to illusory occluders (Anderson, 1994; Gillam and Nakayama, 1999; Grove and Gillam, 2007; Pianta and Gillam, 2003b). Quantitative depth based purely on occlusion geometry has also been found in moving stimuli (Brooks and Gillam, 2006; Shimojo et al., 1988). Thus if the constraints are used by the visual system in other types of stimuli, parsimony argues that the same mechanism is used with two-object arrangements.

Two findings seem to be at odds with the monocular geometry explanation. First, depth estimates in Experiment 4.1 were smaller than those predicted by the minimum depth constraint for four out of six of our observers. One explanation for this discrepancy is that the mechanisms estimating depth from monocular occlusion geometry are not as precise as disparity detectors. Note that this underestimation is also an issue for the double-matching explanation. Another issue is the absence of quanti-

tative depth in camouflage arrangements. However, this is a more serious problem for simultaneous or sequential double-matching accounts, as discussed in more detail below.

4.5.2 Why does camouflage not work?

Experiment 4.1 confirmed that quantitative depth is not perceived in camouflage arrangements when the monocular object is a bar or a disc. Observers perceived these monocular objects slightly in front, slightly behind or at the same depth as the occluder, but depth percepts did not vary consistently with occlusion width. Quantitative depth from camouflage configurations was perceived in the line condition in Experiment 4.1 (and in Gillam et al. (2003)), however, the variability was greater in this case than in the equivalent occlusion condition. If depth in da Vinci stereopsis is indeed based on double-matching (or another disparity-based process), the absence of quantitative depth in the camouflage condition with bar and disc stimuli is hard to explain. Gillam et al. (2003) proposed that in the original bar stimuli used by Nakayama and Shimojo (1990), the bar was shorter than the occluding rectangle and thus had strong relative size cue indicating it was positioned behind the occluder, which could have hindered depth perception in the camouflage condition. This explanation is unlikely since if the bar (or the disc) is made binocular, it is perceived in front of the rectangle and, more importantly, exhibits precise quantitative depth as was shown in Experiment 4.1, even though the same size cues are in place. Moreover, as shown in Figure 4.9-B when the monocular bar is extended to the full length of the occluder, the percept is very similar to the original stimulus shown in Figure 4.9-A. Crossed and uncrossed disparities are treated (relatively) equally by the visual

system, hence the absence of quantitative depth in the camouflage condition poses a serious problem for the double-matching hypothesis.

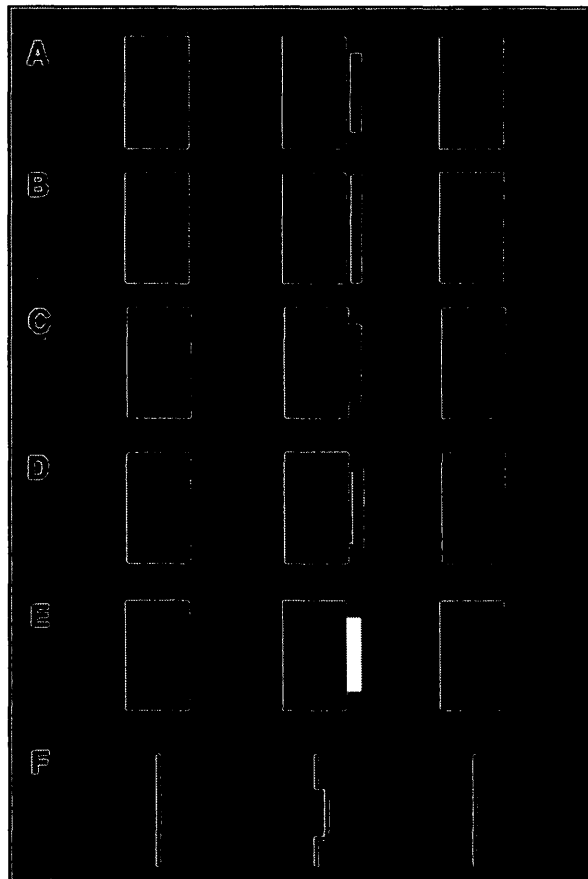


Figure 4.9: Perception of camouflage in different stimuli arrangements. (A) The original Nakayama and Shimojo (1990) two-object arrangement. Most observers perceive the bar as positioned slightly in front or slightly behind the rectangle. (B) The monocular bar has been extended vertically to the full length of the occluder. Most observers perceive it at the same depth as in (A). (C) The monocular bar has been extended laterally towards the rectangle to create an extrusion. The monocular area is more readily seen in front of the binocular rectangle. (D + E) These stimuli contain T and L-junctions as in (C); however, the depth of the bar is similar to that shown in (A). (F) Line stimulus replicating the contours of the edges in stimulus (C). Vivid depth is perceived in this stimulus demonstrating that in (C) depth could be perceived on the basis of binocular disparity.

If depth in two-object arrangements is based on occlusion geometry, camouflage configurations should also yield quantitative depth, unless these phenomena are not represented equally by the visual system (unlike crossed and uncrossed disparities). Nakayama and Shimojo (1990) proposed that this asymmetry makes sense ecologically because camouflage in two-object arrangements would occur only rarely in natural viewing as it requires an almost perfect match between the color and luminance of the occluder and the occluded object. Thus it is possible that the visual system simply cannot compute depth from monocular camouflage. Occlusion, on the other hand, frequently occurs at depth discontinuities and it is likely that cortical mechanisms have developed to compute depth in these areas.

Following this reasoning, it is possible that camouflage arrangements are reinterpreted as occlusion arrangements by creation of an illusory occluding edge on the other side (opposite to the binocular rectangle) of the monocular object (see also Assee and Qian (2007)). Observers in the experiments described in this chapter have reported seeing such edges in some camouflage stimuli and there are numerous examples of monocular occlusions promoting the perception of illusory occluding contours and surfaces (e.g. see Chapter 2). In this case, the occlusion width should have no effect on the perceived depth of the monocular object since its depth is estimated relative to the illusory edge (which presumably does not change its location relative to the monocular object). This explanation, however, does not account for the presence of quantitative depth in the line stimulus in Experiment 4.1 and in Gillam et al. (2003) experiments. It is possible that depth percepts in monocular occlusion stimuli depend on the size of the occluded object and the occluder. The disc and the bar had the same width in our experiments while the line was much thinner (as in Gillam

et al. (2003)); this hypothesis could account for the differences in percepts between the two sets of experiments but it requires further investigation.

Interestingly, quantitative depth in camouflage arrangements is readily perceived when the monocular object is extended towards the occluder such that no gap is left between the two objects and an 'extrusion' is formed as shown in Figure 4.9-C. Gillam et al. (2003) and Cook and Gillam (2004) suggested that cyclopean T or L-junctions might be the one of the factors affecting depth perception in these 'extrusion' stimuli compared with conventional two-object arrangements. Such junctions, they proposed, might serve as additional indicators of occlusion/camouflage relationships since T and L junctions in general play a role in figure-ground assignment in monocular and binocular images (Anderson, 1997; Anderson and Julesz, 1995). If this is the case, however, a similar, clear percept of crossed depth should be obtained from the stimuli shown in Figure 4.9-D and Figure 4.9-E, where cyclopean T and L-junctions are also present. However, in these cases the monocular regions are perceived by most observers to lie beyond the occluder.

It is more likely that the depth seen in extrusion stimuli is due to conventional disparity matching. That is, the right edge of the rectangle in the right image is matched to the right edge of the figure in the left image to produce the resulting depth profile. Importantly, this is an instance of one-to-one matching and not double-matching. To illustrate this point, a simplified, equivalent contour stereogram is shown in Figure 4.9-F, which produces the same depth profile as Figure 4.9-C. Consistent with the above explanation, depth is not readily seen in Figure 4.9-D since double-matching would be required here to obtain depth since the right edge of the binocular rectangle is visible. In Figure 4.9-E one-to-one matching is complicated because the

corresponding edges in the right and left eyes have opposite contrast polarity.

To conclude, it is not yet clear why some instances of the camouflage configuration do not produce quantitative depth perception. The ecological explanation proposed by Nakayama and Shimojo (1990) together with the reinterpretation of the camouflage arrangement as an instance of occlusion by an illusory surface seem to be the most viable alternatives.

Chapter 5

The role of disparity in depth from monocular occlusions

5.1 Objectives

The computational simulation that was used to probe the origin of depth percepts in two-object arrangements in Section 4.4 could be used to examine the mechanisms which underpin depth in other monocular occlusion stimuli. This type of analysis will reveal any binocular matching information available in these stimuli and will verify whether da Vinci stereopsis could rely solely on the output of disparity detectors. To this end, this chapter examines in detail the response of a set of simple disparity detectors to a wide range of monocular occlusions stimuli.

5.2 Methods

The model disparity detectors were implemented as a cross-correlation algorithm with SSD as a correlation metric, square correlation window, window sizes 4, 10 and 20 px and disparity ranges ± 20 px and ± 30 px. The disparity that generated the maximum correlation value (minimum SSD) was chosen as the estimated disparity for each pixel. In cases where more than one disparity generated the maximum correlation value the three methods of disparity assignment described in Section 4.4 were used - mean, median and zero bias. Match goodness and reliability metrics were also computed for each stimulus as in Section 4.4. Match goodness was computed as the ratio of the maximum correlation response of each pixel to the maximum correlation response in the whole image. Larger ratios indicated better matches for each pixel. The reliability metric for each pixel was computed as the difference between the normalized maximum correlation value and the next largest value. Thus if two or more disparities obtained the maximum correlation value the reliability would be equal to zero. These metrics reflected the quality of disparity estimates and allowed to gauge the efficiency of the correlation algorithm at each image location. The algorithm was applied to several different stimuli:

1. Standard wedding cake RDS with disparities of 4, 8 and 12 px (a smaller square in front of a bigger one) (Figure 5.1-A)
2. The random-dot frame stimulus introduced in Chapter 2 with disparity of 6 px and occlusion widths of 2, 4 and 6 px (Figure 5.1-B)
3. The illusory occluder stimulus of Gillam and Nakayama (1999) with occlusion widths of 3, 5 and 8 px (Figure 5.1-C)

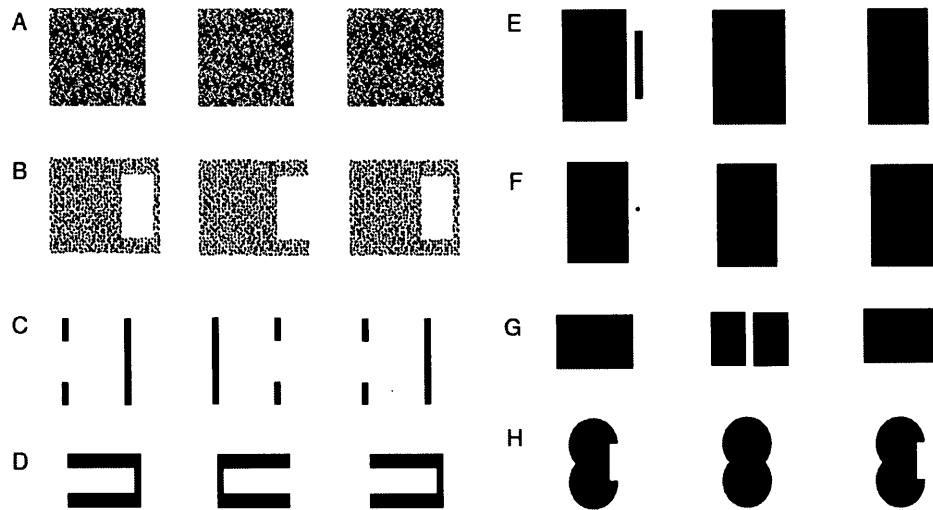


Figure 5.1: Stimuli used in the computational simulations. A) RDS B) Random-dot frame C) Illusory occluder from Gillam and Nakayama (1999) D) Illusory occluder from Liu et al. (1994) E) Two-object arrangement with a bar from Nakayama and Shimojo (1990) F) Two-object arrangement with a disc G) Monocular gap from Gillam et al. (1999) H) Monocular intrusion Cook and Gillam (2004)

4. The illusory occluder stimulus of Liu et al. (1994) with occlusion widths of 4, 8 and 12 px (Figure 5.1-D)
5. Two-object arrangement with a bar (Nakayama and Shimojo, 1990) with occlusion widths of 7, 12 and 18 px (Figure 5.1-E)
6. Two-object arrangement with a disc (Gillam and Nakayama, 1999) with occlusion widths of 7, 12 and 18 px (Figure 5.1-F)
7. Monocular gap stimulus of Gillam et al. (1999) with occlusion widths of 4, 8 and 12 px (Figure 5.1-H)
8. Illusory intrusion stimulus of Cook and Gillam (2004) with occlusion widths of

8, 10 and 12 px (Figure 5.1-I)

Image sizes ranged from 100×100 to 200×200 pixels. The algorithm was implemented in MATLAB 7.10 running on Mac OS X Version 10.7 on a 2.8 GHz Intel Core MacBook Pro.

5.3 Results

The responses of the algorithm to each stimulus type are shown in Figures 5.2-5.9. The figures show the results of simulation trials with intermediate occlusion widths, 10px window and ± 20 px disparity range. In all the diagrams, the monocularly occluded areas are outlined in red. Regions of interest, where depth should be perceived on the basis of monocular occlusion geometry are outlined in blue. In the case of the RDS the central square is the region of interest. In other stimuli the region of interest is either the monocular occlusion (and thus the blue outline coincides with the red) or the illusory occluding surface depending on the configuration. Each figure shows the following data:

1. The left and right images of the stimulus.
2. The disparity profile for the whole stimulus. A disparity profile is the correlation response plotted as a function of disparity. The overall disparity profile is computed by averaging the disparity profiles for all the image pixels.
3. The disparity profile for the region of interest. It is computed by averaging the disparity profiles for the image pixels falling within the outlined region.
4. The ground truth map where darker colours indicate closer disparities (depths).

5. The computed disparity map using the mean disparity selection method. Darker colours indicate more crossed disparities (closer depths). The map corresponding to the median disparity selection method is not shown since for all cases it is virtually identical to the mean method map.
6. The computed disparity map using the zero-bias disparity selection method. Darker colours indicate more crossed disparities (closer depths).
7. The match goodness map where darker colours indicate lower match goodness.
8. The reliability map where darker colours indicate lower reliability.

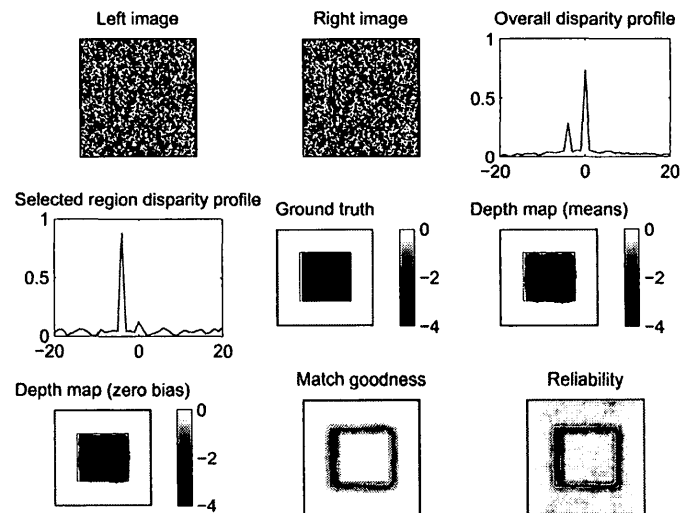


Figure 5.2: Simulation results for the RDS

Figure 5.10 summarizes the results of the simulations for all stimuli and parameters (occlusion widths, window sizes and disparity ranges). Each of the plots in Figure

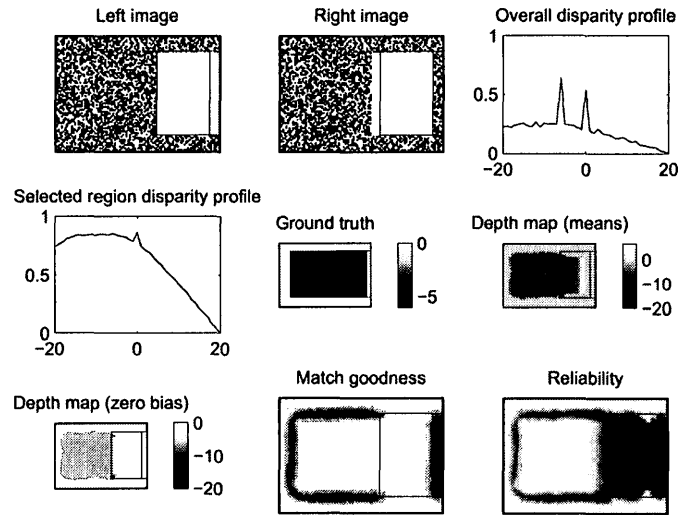


Figure 5.3: Simulation results for the random-dot frame stimulus.

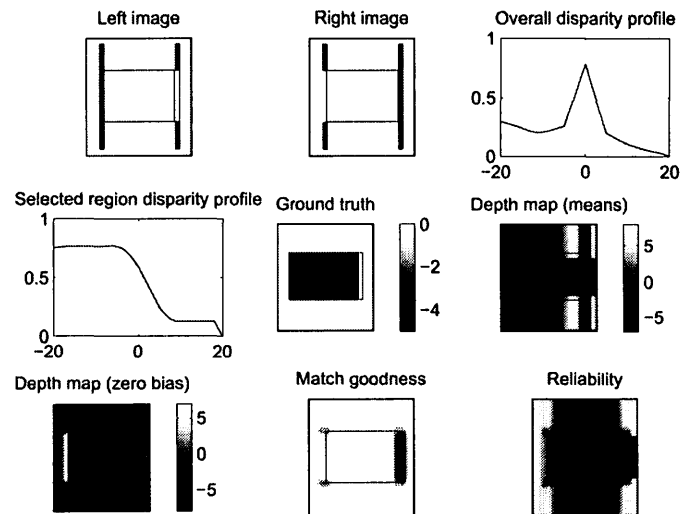


Figure 5.4: Simulation results for the Gillam & Nakayama stimulus.

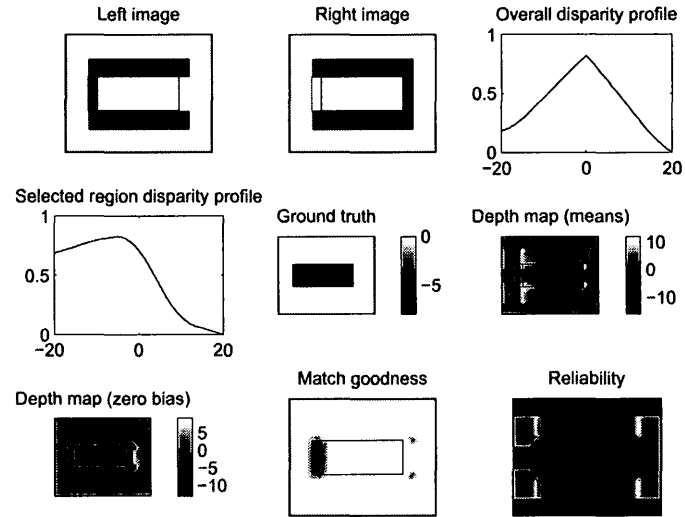


Figure 5.5: Simulation results for the Liu et al. stimulus.

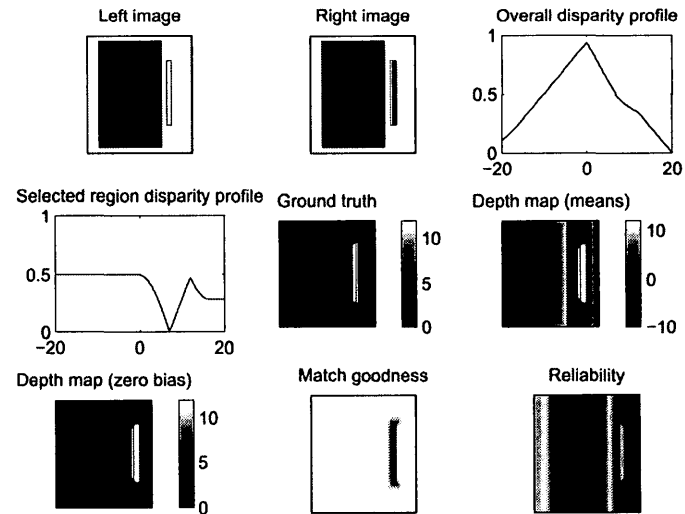


Figure 5.6: Simulation results for a two-object arrangement with a bar.

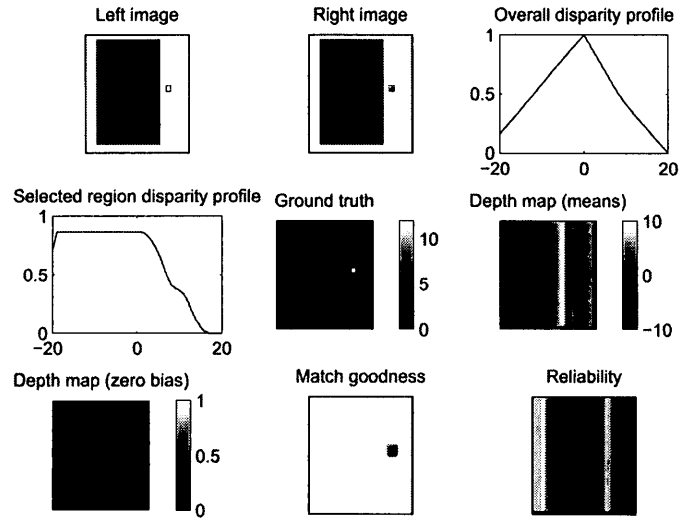


Figure 5.7: Simulation results for a two-object arrangement with a disc.

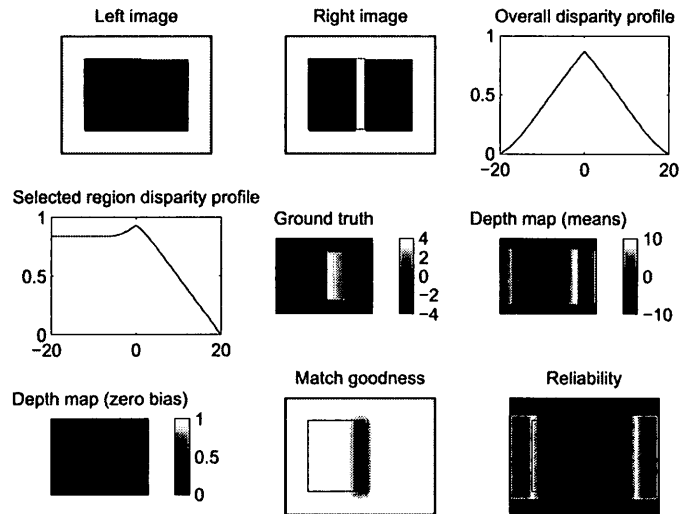


Figure 5.8: Simulation results for the monocular gap stimulus

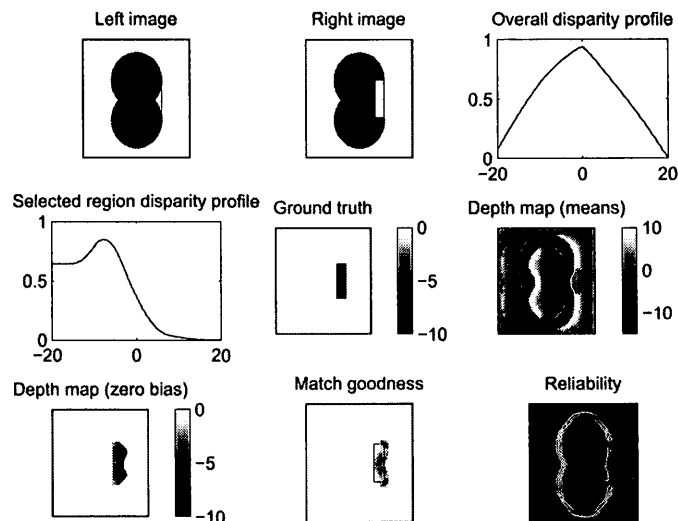


Figure 5.9: Simulation results for the monocular intrusion stimulus

5.10 shows the computed depth in the region of interest of one particular stimulus as a function of stimulus and disparity detectors parameters. Results for two different methods if disparity selection are shown for each stimulus type in side-by-side plots.

As can be seen in the figures, the simple stereo algorithm is not able to compute depth veridically for all occlusion stimuli. The zero-bias disparity assignment method gives estimates somewhat close to the ground truth only for the monocular intrusion stimulus and the two-object arrangement with the bar (however see Figure 4.8). The mean depth estimation method produces incorrect depth maps in all monocular occlusion stimuli. Moreover, when the depth estimates from different occlusion widths, window sizes, disparity ranges and depth selection methods are considered together (see Figure 5.10), it is obvious that the algorithm performance is highly sensitive to parameters. As expected the algorithm is able to compute the depth correctly in the

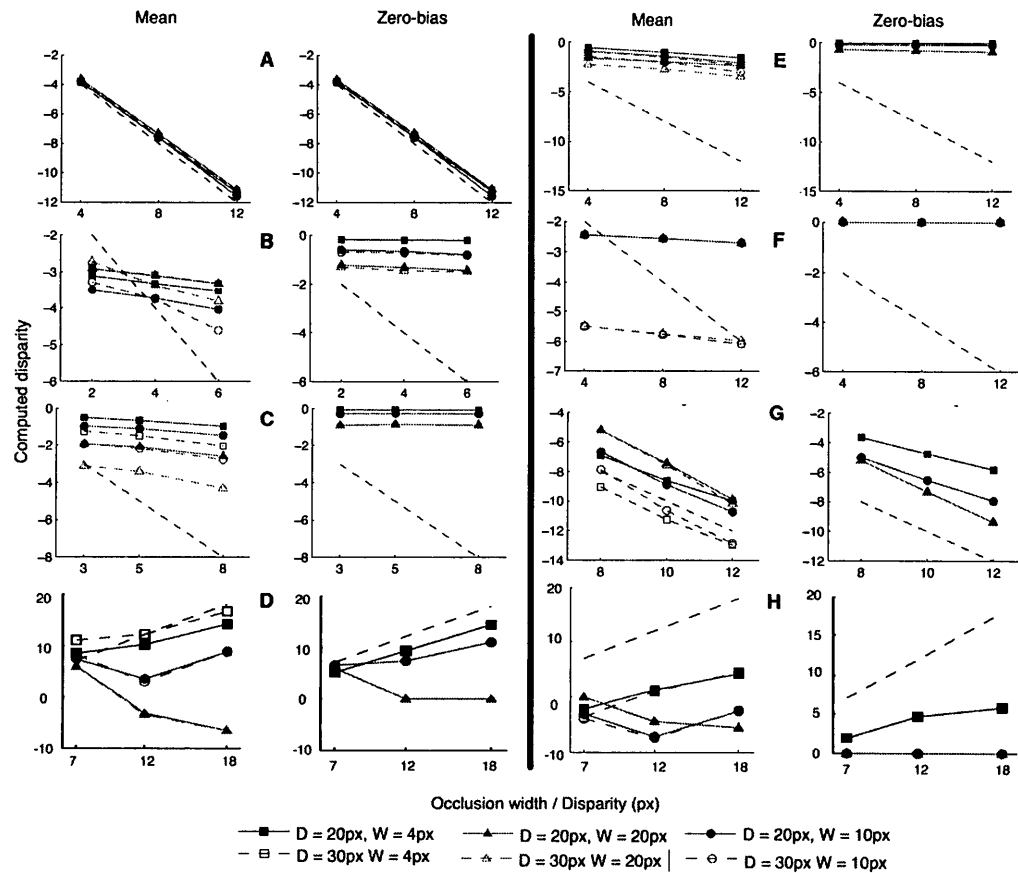


Figure 5.10: Simulation results for different occlusion widths. A) RDS B) Random-dot frame C) Illusory occluder (Gillam and Nakayama, 1999) D) Two-object arrangement with a bar (Nakayama and Shimojo, 1990). E) Illusory occluder (Liu et al., 1994). F) Monocular gap (Gillam et al., 1999). G) Monocular intrusion (Cook and Gillam, 2004). H) Two-object arrangement with a disc (Gillam et al., 2003). Differently coloured lines show the different algorithm parameters. Black dashed lines indicate the predicted disparity.

RDS stimuli (see Figures 5.2 and 5.10-A).

Except for the monocular intrusion stimulus, disparity estimates in all monocular regions have low match goodness and reliability. In comparison, textured binocular areas have high match goodness and reliability (see Figure 5.2). Textureless binocular

areas have high match goodness and low reliability (see for example Figures 5.5 and 5.4). In the case of the monocular intrusion stimulus, the monocular area shows a higher match goodness and reliability than the other monocular occlusion stimuli (see Figure 5.9). The results for the monocular intrusion stimulus support the argument made in Section 4.5.2 that depth in this stimulus is likely to be based on binocular one-to-one matching (similar results were obtained for this stimulus by Harris and Smith (2010)).

Taken together these results show that the matching information available in the majority of the monocular occlusion stimuli is not sufficient to compute correct depth maps in these stimuli. Thus the simple correlation-based disparity detectors cannot be the sole mechanism responsible for depth in monocularly occluded regions and illusory occluders. A more sophisticated mechanism is required, one that takes occlusion geometry into account in addition to binocular disparity. It is likely that da Vinci stereopsis entails explicit detection of monocular areas and subsequent processing of depth in these areas by a population of neurons dedicated specifically to this task. The difference in patterns of match goodness and reliability in monocularly occluded and binocular areas suggests that the visual system could use related metrics to distinguish monocular areas from binocular ones. Chapter 6 describes a computational model incorporating these and other principles that can successfully recreate depth maps in all the stimuli examined here.

Chapter 6

A biologically-plausible computational model of depth from monocular occlusions and binocular disparity

6.1 Objectives

Given the evidence that occlusions are important to resolving depth structure, it is important that they be considered in any general, biologically relevant model of stereopsis. The importance of the explicit detection and treatment of monocular occlusions has been recognized in the computer vision literature. Many stereo-algorithms employ different techniques to identify and assign depth to monocularly occluded areas (Belhumeur and Mumford, 1992; Jodoin et al., 2006; Lin and Tomasi, 2004; Min and Sohn,

2008; Sizintsev and Wildes, 2007, 2010; Sun et al., 2005; Zitnick and Kanade, 2000). Most concentrated on monocular occlusion detection methods, which are described and compared in Egnal and Wildes (2002). The methods for depth assignment to monocular areas have not been developed to the same extent. Many algorithms simply detect the occluded areas and present disparity maps with the occluded regions left blank (Jodoin et al., 2006; Lin and Tomasi, 2004; Sizintsev and Wildes, 2007; Zitnick and Kanade, 2000). Most other approaches include variations of disparity propagation from neighbouring binocular areas with various degrees of sophistication (Min and Sohn, 2008; Min et al., 2010; Wang et al., 2008; Yang et al., 2009).

These algorithms are primarily directed at improving the extraction of depth from disparity for machine vision and robotics and on the most part are not based on physiological or psychophysical findings. In the realm of computational models of biological vision (referred to hereon as biologically-inspired), monocular occlusions remain a relatively unexplored phenomenon. To the best of my knowledge there are only four models that can be classified as biologically-inspired and that provide some type of treatment of monocular occlusions. The mechanisms these models use for detection of occluded regions and for depth assignment in these regions are outlined in Table 6.1. This table also shows the stimuli that the models were tested on and the main shortcomings of each of the models. Several things can be concluded from the examination of Table 6.1. First, three of the models cannot be considered completely biologically-plausible since, among other things, all of them use binary representations of neuronal firing rates (Assee and Qian, 2007). Second, model testing is very limited as three of the models were tested with a maximum of two stimuli each. Third, none of the models explicitly incorporate the minimum depth constraint (i.e. width

of the monocular region) in order to assign depth in and around monocularly occluded regions. Finally, judging by their architecture, none of the models will be able to compute depth correctly for several important categories of monocular occlusion stimuli (e.g. illusory occluders or monocular gaps).

The goal of the modelling efforts described in this chapter is to develop a completely biologically-plausible model that will:

- ◆ Compute depth from disparity and monocular occlusions, separately and in combination.
- ◆ Be based on physiological findings and psychophysical data reported in this thesis and in previous studies.
- ◆ Adopt some of the successful techniques from computer vision, cast into a biologically-plausible form.
- ◆ Produce depth maps that closely correspond to observer percepts for a wide variety of stimuli including: photographs of real scenes, RDS, illusory occluder stimuli, two-object arrangements, monocular gap and monocular intrusion stimuli.

In the rest of the chapter, I first provide a working definition of biological-plausibility with which the model will comply and outline the main principles on which the model is based. I then provide the computational formulation of the model. Next, I outline the procedure and the results of testing of the model on a large battery of stimuli. Finally, I discuss the features, the limitations and the predictions of the model.

Table 6.1: Biologically-inspired models of stereopsis with specific treatment of monocular occlusions.

| Model | Occlusion detection | Assignment of depth in/from occlusions | Stimuli tested | Drawbacks |
|---|--|---|---|---|
| Cao and Grossberg (2005); Grossberg and Howe (2003) | No dedicated occlusion detectors. Monocular areas are identified implicitly through competition between monocularly identified luminance edges and binocularly identified edges. | Depth in occluded areas relies solely on disparity (double matching). | Two-object arrangement with a bar with same and opposite contrast polarities. Modified monocular gap stimulus. For the later two stimuli, the model generated results different from the psychophysical data. | <ul style="list-style-type: none"> • The method of occlusion detection would not be able to identify occluded areas in densely textured surfaces such as RDS due to the absence of clearly defined monocular luminance edges (contours). • This method will also not be able to reconstruct illusory occluders since occlusion geometry is not used and illusory occluders do not have luminance edges. • Neural output of the neurons is binary. |
| Watanabe and Fukushima (1999) | Explicit detection by left-to-right check (non biologically-plausible implementation). | Specific geometric constraints are used to assign depth to monocular occlusions. If a point is marked as occluded, there needs to be an occluding point. However, the width of the monocular region is not explicitly detected. | A random-dot stereogram with two monocular side-bands. The model made a qualitatively correct prediction of the monocular regions' depth but it is not clear whether the depth was correct quantitatively. | <ul style="list-style-type: none"> • Neural output is binary. • Initial disparity map is acquired using feature detection. • There is no mechanism in place to stop the spread of disparity beyond object edges. • Monocular occlusion detectors which fire but have no occluding points according to occlusion geometry are inhibited. Hence it is likely that this algorithm will not be able to detect depth in illusory occluder stimuli. |

Table 6.1 – continued from previous page

| Model | Occlusion detection | Assignment of depth in/from occlusions | Stimuli tested | Drawbacks |
|--|---|--|--|--|
| Hayashi et al. (2004) (based on Watanabe and Fukushima (1999)) | A variant of goodness of match metric, achieved by thresholding the output of disparity energy neurons. | Monocular areas are assigned the disparity of nearby binocular areas (the background surface). | None of the occlusion stimuli were tested. Instead a conventional random-dot stereogram showing one rectangle in front of another and a natural image were used. | <ul style="list-style-type: none"> • Neural output is binary. • Simply filling in disparity in monocular areas from surrounding binocular areas will not work for all cases of monocular occlusions in particular for two-object arrangements, illusory occluders, monocular gaps, monocular intrusions and possibly in other cases. |
| Assee and Qian (2007) | No explicit detection of monocular occlusions. Assumed to be located at the centre of the receptive fields of disparity-edge selective neurons. | Monocular areas are assigned the depth of the further surface that was detected by the disparity-edge selective neurons. | Used conventional RDS with a raised centre strip. Also tested a Panum's limiting case stimulus with different separations. | <ul style="list-style-type: none"> • This model will most likely not be able to predict depth in a range of stimuli where there is no texture on both sides of the monocular region such as illusory occluders, monocular gap, monocular intrusion and two-object arrangements. • No mechanism for disparity spreading was proposed. |

6.2 Biological plausibility

The term 'biologically-plausible' has been used to describe many different models and algorithms with various degrees of supporting evidence from the physiological and psychophysical literature. Thus in the absence of a reliable reference, a working definition of the term has to be established. In this thesis the term 'biologically-plausible' refers to computational units (neurons, neuronal connections or clusters of neurons) which:

1. Have been proposed in the physiological literature to describe real neurons as

measured by cell recordings and brain imaging (e.g. the energy model of disparity detectors, see below)

2. Have been proven theoretically to be a plausible neural mechanism based on physiological data (e.g. the normalization circuit, see below)
3. Are composed of a collection of mechanisms that comply to (1), (2) or both.

Based on this definition, the following computational units were used in the model:

- Neurons that have a distributed response (non-binary). Here, and in most models of this type, the neurons model the rate of firing of physical neurons not the magnitude of each discharge. It is well known that this rate is variable and cannot be represented by a binary function.
- Energy disparity neurons (Cumming and DeAngelis, 2001; Ohzawa et al., 1990)
- Summation, subtraction and multiplication of signals (described in numerous works)
- MAX and MIN operations (e.g. Yu et al., 2002)
- Normalization (e.g. Heeger, 1992)
- Inhibitory inter-neural connections (Bullier, 2001; Gilbert and Wiesel, 1983)
- Excitatory inter-neural connections (Bullier, 2001; Gilbert and Wiesel, 1983)
- End-stopped cells (hypercomplex cells) (Hubel and Wiesel, 1965) and on-off cells (Hubel and Wiesel, 1962)
- Rectification with respect to a threshold (this mechanism has been described in numerous works, e.g. Read and Cumming, 2003)
- Top-down modulation of inter-neural connections and receptive fields with respect to figure ground segregation (Angelucci et al., 2002; Hupe et al., 1998; Lamme, 1995; Schoenfeld et al., 2003)

6.3 Model principles

The model proposed here is based on several principles derived from the psychophysical and computational studies described in this thesis and in the literature:

Monocular occlusions are detected explicitly - The computational experiments in Chapter 5 have demonstrated that simple disparity detectors cannot be solely responsible for depth percepts from monocular occlusions. At the same time, computational algorithms that use explicit detection of monocular occlusions are able to successfully recreate depth maps in complex scenes (e.g Lin and Tomasi (2004); Min and Sohn (2008); Sizintsev and Wildes (2007)). Moreover, the experiments described in Chapters 2-4 show that the depth of monocular objects/areas and illusory occluders is likely to be computed on the basis of occlusion geometry rather than binocular disparity. This suggests that a set of mechanisms distinct from that involved in disparity computation is required.

The width of monocular regions is used to compute the depth in these areas and the surrounding binocular areas with a non reliable disparity signal - The experiments of Chapters 2-4 as well as previous studies (Gillam and Nakayama, 1999; Grove and Gillam, 2007; Pianta and Gillam, 2003a) have shown that the minimum depth constraint arising from occlusion geometry is used by the visual system to assign depth to monocular objects/areas and to illusory occluders. Thus the width of the monocular region, which can be used to compute the minimum depth constraint, should be the primary metric used by mechanisms designed to compute depth from monocular occlusions.

Depth from monocular occlusions and disparity is computed concurrently - It has been suggested in several studies that depth from monocular occlusions and from binocular disparity are likely computed simultaneously in the early stages of visual processing (Gillam and Borsting, 1988; Kuroki and Nakamizo, 2006; Mitsudo et al., 2005).

Depth signals from disparity detectors and monocular occlusion detec-

tors interact - Previous work (Hakkinen and Nyman, 1996, 2001) and the experiments of Chapters 2 and 3, have demonstrated that depth signals from monocular occlusions and binocular disparity interact in complex ways. Specifically, when the disparity signal is weak, the monocular occlusions signal can affect disparity-based depth and disparity signals were shown to affect depth from monocular occlusions in many instances.

Monocular occlusions determine the depth of illusory occluders - The experiments of Chapter 2 demonstrated how stimulus configurations can change drastically when monocular occlusions are introduced. To account for their presence in the absence of an explicit occluding edge, the visual system creates an illusory surface with depth determined by the occlusion geometry. Thus, depth signals based on the information in monocularly occluded areas can be used in order to compute the depth of illusory occluders in regions devoid of texture.

Camouflage is interpreted as occlusion - it is assumed in the model that monocular regions arise due to an occlusion of one object by another and not due to camouflage. Evidence from experiments with two-object arrangements suggests that the visual system might not be equipped to process camouflage and instead interprets it as occlusion (see extended discussion in Chapter 4). Moreover, this assumption greatly simplifies the computations required for determining depth in monocular regions and around them.

6.4 Implementation details

The model and the stimuli generating scripts were implemented in MATLAB 7.10 running on Mac OS X Version 10.7 on a 2.8 GHz Intel Core MacBook Pro. All the code is original, unless specified otherwise.

6.5 Model formulation

Before the biologically-plausible model was implemented, the principles outlined in the previous section were tested with a proof-of-concept algorithm. This algorithm, which was not biologically-plausible, had the same main components (see Figure 6.1) and was tested with on the same large battery of stimuli as the model. It could reproduce depth veridically in a variety of occlusion stimuli, proving that the proposed architecture is suitable for modelling da Vinci stereopsis. For brevity the detailed description of the proof-of-concept algorithm is omitted.

An overview of the model components is shown in Figure 6.1. Below I describe each of the components in detail and provide the mathematical formulations for all types of neurons engaged in the model. For convenience, Table A.2 in Appendix A summarizes all the symbols and functions used in the mathematical formulation of the model.

6.5.1 Initial disparity computation - Energy neurons and inter-neural connections

The initial computation of disparity is performed using a complex network of disparity detectors modelled as energy neurons (DeAngelis et al., 1995; Ohzawa et al., 1990).

The disparity energy model postulates that simple neurons compute the sum of the left and the right images filtered with respective receptive fields (RFs) (DeAngelis et al., 1995; Ohzawa et al., 1990). Accordingly, the response of a linear simple binocular neuron can be expressed as:

$$S = f_L : I_L + f_R : I_R = FI_L + FI_R, \quad (6.1)$$

where f_L and f_R are corresponding RFs in the left and right eyes, I_L and I_R are the left and the right image patches that fall on these RFs and ‘:’ is the Frobenius inner

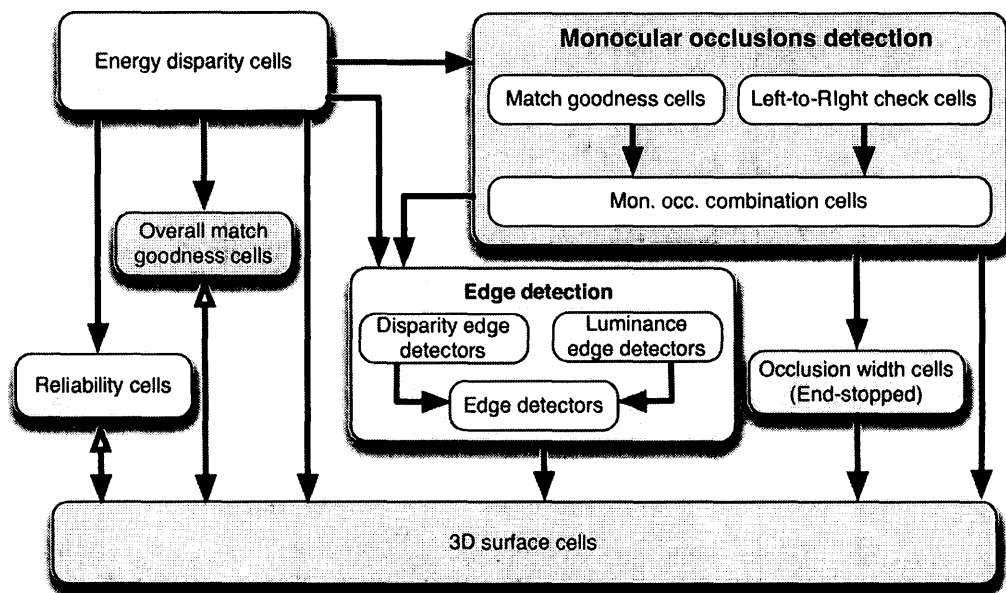


Figure 6.1: Model overview. Filled arrows show feed-forward connections and hollow arrow-heads show feed-back connections.

product, defined as $A : B = \sum_i \sum_j A_{i,j} B_{i,j}$. These neurons are postulated to be selective for position, disparity, spatial frequency and orientation (DeAngelis et al., 1995; Ohzawa et al., 1990).

Physiological studies (Anzai et al., 1997; DeAngelis et al., 1995; Prince et al., 2002) showed that the RFs of most simple binocular neurons can be described by Gabor functions. Disparity selectivity in these cells can be achieved through two mechanisms: 1) position shift - a shift between the positions of the RFs in the two eyes, and 2) phase shift - a shift in the phase of the Gabor in the two eyes. Since both phase and position shift mechanisms are used in the visual system (Anzai et al., 1999), both mechanisms are included in the model. The following Gabor filter was

used to describe the left RF of a simple binocular neuron:

$$f_L(x, y, d_L, \phi_L) = Gauss(x, y, d_L) \cdot Sin(x, y, d_L, \phi_L) \quad (6.2)$$

$$Gauss(x, y, d_L) = \frac{1}{2\pi\sigma_x\sigma_y} \cdot e^{-\frac{1}{2} \left[\left(\frac{(x+d_L)\sin\theta + y\cos\theta}{\sigma_x} \right)^2 + \left(\frac{(x+d_L)\cos\theta - y\sin\theta}{\sigma_y} \right)^2 \right]} \quad (6.3)$$

$$Sin(x, y, d_L, \phi_L) = \cos \left[\omega_0 \left((x + d_L) \sin\theta \right) + \phi_L \right] \quad (6.4)$$

where d_L is the position shift of the left receptive field, σ_x and σ_y are the horizontal and vertical Gaussian widths, θ is the preferred orientation, ω_0 is the peak preferred frequency, ϕ_L is the left phase parameter. The right RF has the same definition but with a positional shift - d_R and a phase shift - ϕ_R .

From this the preferred disparity of a simple cell S can be expressed as follows:

$$D_S = (d_L - d_R) + \frac{\phi_L - \phi_R}{\omega_0} \quad (6.5)$$

The phase-shift mechanisms are used for disparities smaller than π for each preferred frequency. For disparities larger than π , position-shift neurons are used. Three aspects need to be noted with respect to disparity computation. First, oriented RFs with phase-shift disparity tuning, actually do not code strictly horizontal disparity. They code disparity orthogonal to their orientation since the phase shifts orthogonally to the cell's orientation. Position-shift cells of all orientations always encode horizontal disparity. Thus, as a simple solution to this problem, only position shift neurons were used for orientations other than vertical. Second, the matches are made along the same epipolar lines in the two images (epipolar constraint). Third, the shift in the phase or the position of RFs which achieves disparity tuning is performed in

one eye only, as was done in previous works (e.g Chen and Qian (2004)).

The complex neurons sum the squared responses of two linear neurons in quadrature phase (DeAngelis et al., 1995; Ohzawa et al., 1990):

$$C_0 = S_1^2 + S_2^2 \tag{6.6}$$

The preferred disparity of the complex (model) cell is equal to that of the simple subunits but its output is not sensitive to the phase of the stimulus.

Mathematically, the classic energy model computes a quantity that comes close to a cross-correlation between images filtered with the neuron RFs. However, this quantity contains not only correlation information but also the monocular energy of the two images which makes it very prone to false matches (Fleet et al., 1996; Read, 2010; Read and Cumming, 2006). Given this, disparity estimates from the output of classic energy neurons are extremely noisy. This problem can be fixed easily by normalizing the output of the complex cells C_0 by the sum of squared monocular energy responses of the monocular RFs of the two simple cells S_1 and S_2 :

$$C = \frac{C_0}{FI_{1L}^2 + FI_{1R}^2 + FI_{2L}^2 + FI_{2R}^2}, \tag{6.7}$$

where $S_1 = FI_{1L} + FI_{1R}$ and $S_2 = FI_{2L} + FI_{2R}$.

Normalization with respect to monocular energy was proposed as one of the potential modifications which could bring the responses of the energy model closer to observed human performance (Allenmark and Read, 2011; Banks et al., 2004; Read, 2010; Read and Cumming, 2006) and it has also been demonstrated to be biologically plausible (Heeger, 1992; Read and Cumming, 2007).

Pooling across orientations and spatial frequencies, combined with local spatial pooling was proposed as another way to reduce the potential for false matches from energy neurons (Fleet et al., 1996). Thus in this model, orientation pooling is performed at each scale, and responses at all scales are pooled to produce the final

response. Spatial pooling (using a convolution with a 2D Gaussian) is performed for each orientation, each scale and on the final response. All pooling is performed by averaging the disparity profiles (see Figure 6.2) of the different cells.

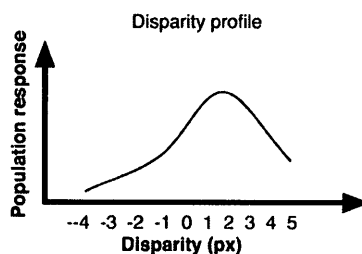


Figure 6.2: Disparity profile. The pooled population response $C_{x,y}^*$ for the location x, y is plotted as a function of disparity. In this case the population responds maximally to disparity of 2 px.

For each location x, y of the image (or retina) there is a population of complex cells C centred on x, y and tuned to the full ranges of disparity, orientation and spatial frequency as specified above. The final, pooled (as specified above) response of this population for the range of disparities $[-d_m, d_m]$ is labeled $C_{x,y}^*$ and represents the disparity profile of location x, y . It can be represented graphically as shown in Figure 6.2. The pooled response of the population to one particular disparity d is denoted $C_{x,y,d}$ (one point on the curve of Figure 6.2).

6.5.2 Monocular occlusions detection

Before describing the cells used to detect monocularly occluded areas in the model, I discuss the advantages and disadvantages of the main occlusion detection methods that have been proposed in the computer vision literature for occlusion detection (Egnal and Wildes, 2002):

- ◆ **Left-to-Right check** - In classic left-to-right check, disparity maps are computed first using the right image as the origin and then using the left image

as the origin. For binocular pixels (normally) the match made in one direction should correspond to the match made in other direction. Pixels for which the matches differ substantially are considered to be monocular pixels. This procedure is illustrated in Figure 6.3. This technique was found to be effective at detecting occlusions (Egnal and Wildes, 2002), however, it showed poor performance in areas with low spatial frequency.

- ◆ **Ordering constraint** - This constraint postulates that if point A is positioned to the left of point B in the left image, their matches will maintain the same order in the right image. In occluded regions, due to spurious matches, this constraint might not hold. Thus regions violating the ordering constraint are labeled as occluded. However, this constraint might not always hold even for binocular regions. For example, when A and B are very close in the left image, and A has a larger disparity than B such that in the right image their order is reversed. Moreover, even in occluded areas matching order could be preserved by chance. It was found to be the most conservative method out of the ones examined by Egnal and Wildes (2002).
- ◆ **Occlusion constraint** - this approach builds on the fact that occlusions occur at depth discontinuities. A disparity jump at neighbouring locations in the left image implies an occluded region in the right image. Although this method was found to work well in general (Egnal and Wildes, 2002), it has two disadvantages. First, it is prone to fail in the occluded regions of the left image when building an occlusion map of the right image. Since the matches in occluded regions tend to be spurious, this method might erroneously detect a depth discontinuity and signal an occluded region in the other image. More importantly, this approach will fail in areas devoid of texture due to inaccurate matches produced for binocular regions around monocular areas. Thus it is unlikely to detect occluded regions veridically in stereograms such as that of Gillam and Nakayama (1999) and Liu et al. (1994).

- ◆ **Bimodalities in disparity** similarly to the occlusion constraint this approach utilizes the fact that occlusions occur at depth discontinuities. Thus, a small region around an occlusion border would contain two disparities, that of the occluder and that of the occluded surface, yielding a bimodal histogram of disparity. Hence such regions could be identified as bordering occluded areas. Besides having the same drawbacks as the occlusion constraint, this approach only detects occlusion boundaries, not whole occluded areas. This approach was also found to be very sensitive to its parameters (Egnal and Wildes, 2002).
- ◆ **Match goodness jumps** The magnitude of the response of disparity detectors in binocular areas is normally quite high. In contrast, the magnitude of the response in occluded areas is low since no good match exists. The match goodness jumps method states that regions with high goodness of matches neighbouring regions with low goodness of matches indicate occlusion boundaries. Like the previous approach, this method only detects occlusion boundaries, not whole areas. However, this method was found to perform better than the other approaches in areas of low spatial contrast (Egnal and Wildes, 2002). A similar approach that can detect the whole occluded regions was used by Zitnick and Kanade (2000). They applied a minimum-response threshold to the highest correlation value for each pixel to identify occluded pixels.
- ◆ **Uniqueness constraint** This constraint was first proposed by Marr and Poggio (1976) as a general constraint on stereo matching. In its original form it states that each pixel in one image can have only one match in the other image. This constraint was later used in stereo algorithms to detect occluded areas. For example, Sun et al. (2005) modified the uniqueness constraint such that an occluded pixel could have no matches, while a visible pixel could have one or more matches. Accordingly, occluded pixels in the left image were defined as pixels that were not matched to any pixel in the right image given a right image disparity map. Disparity and occlusion maps were computed for both

images. Note that this approach is very similar to the left-right-check method. In fact left-right-check on its own implicitly imposes the uniqueness constraint on matching.

Based on the above findings and considerations I have chosen to detect monocularly occluded regions using a combination of two complimentary methods - left-to-right check and a (modified) match goodness heuristic, which were implemented in a biologically-plausible form. Egnal and Wildes (2002) found both techniques to be effective at detecting occlusions. However, left-to-right check was inferior in areas with low spatial frequency, while match goodness performed well in these areas. Moreover, both methods do not depend on reliable disparity estimates around the occluded regions and thus would be able to detect occluded regions in images with large textureless areas (such as the stimuli of Gillam and Nakayama (1999) and Liu et al. (1994)).

Left-to-right check

Before describing the neurons that compute the left-to-right check (see Figure 6.3), new notation needs to be introduced. Let the response of a population of complex cells $C_{x,y}^*$ with all the left-eye receptive fields fixed at location x, y be denoted as $C_{x,y}^{L*}$. Accordingly a population of complex cells $C_{x,y}^*$ with all the right-eye receptive fields fixed at location x, y is denoted $C_{x,y}^{R*}$. The response of the $C_{x,y}^{L*}$ cells tuned to a specific disparity is denoted $C_{x,y,d}^L$ for the left eye and $C_{x,y,d}^R$ for the right eye.

At each location x, y there are two types of neurons that together produce a left-to-right check response. The first type computes the summed difference between the disparity profiles of neurons $C_{x,y}^{L*}$ and $C_{x+d,y}^{R*}$ for a given disparity d . When location x, y has a well defined disparity d , then the $C_{x,y}^{L*}$ disparity profile will have a peak at d and the matching $C_{x+d,y}^{R*}$ profile will have a peak at $-d$. Thus when the profiles are shifted with respect to each other by d and their point-by-point difference is summed, this sum will be very small. On the other hand, disparity profiles for non-matching pixels

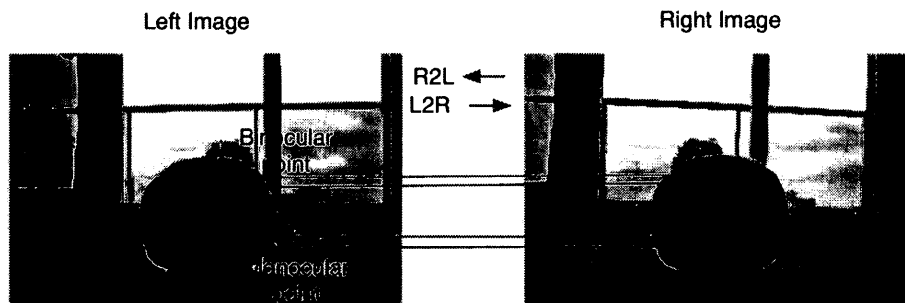


Figure 6.3: Left-to-right check example. Blue arrows indicate a right-to-left disparity computation and red arrows indicate a left-to-right disparity computation. In the example with the binocular point, both directions match the same points in the two images. In the case of the monocular tennis ball in the left image, in the left-to-right computation it is matched to some binocular point p (a black patch in this example) in the right image. However, in the right-to-left computation p is matched to itself in the right eye. Thus the match is different in the two computations and the tennis ball in the left image is marked as monocularly occluded.

will be different (after the appropriate shift) and will yield a relatively large difference. The shift is done simply by comparing each $C_{x,y,d}^L$ response with a $C_{x+d,y,d'-d}^R$ response. Before the subtraction, the two profiles are normalized and cubed. The cubing is done to amplify the true peak of the profile relative to the false peaks:

$$R_{x,y,d} = \sum_{d'=d_1}^{d_2} |n(C_{x,y,d'}^L)^3 - n(C_{x+d,y,d'-d}^R)^3| \quad (6.8)$$

where $d_1 = \max(-d_m, d - d_m)$ and $d_2 = \min(d + m, d + d_m)$ and the function $n(x)$ is a signal normalization function:

$$n(x) = \left[\frac{x}{\text{MAX}(x)} \right] \quad (6.9)$$

This computation is shown graphically in Figure 6.4. The difference between the two disparity profiles will be large when the peak of the left profile $C_{x,y}^{L*}$ is different from the peak of the appropriately shifted right profile $C_{x+d,y}^{R*}$. It will be small when

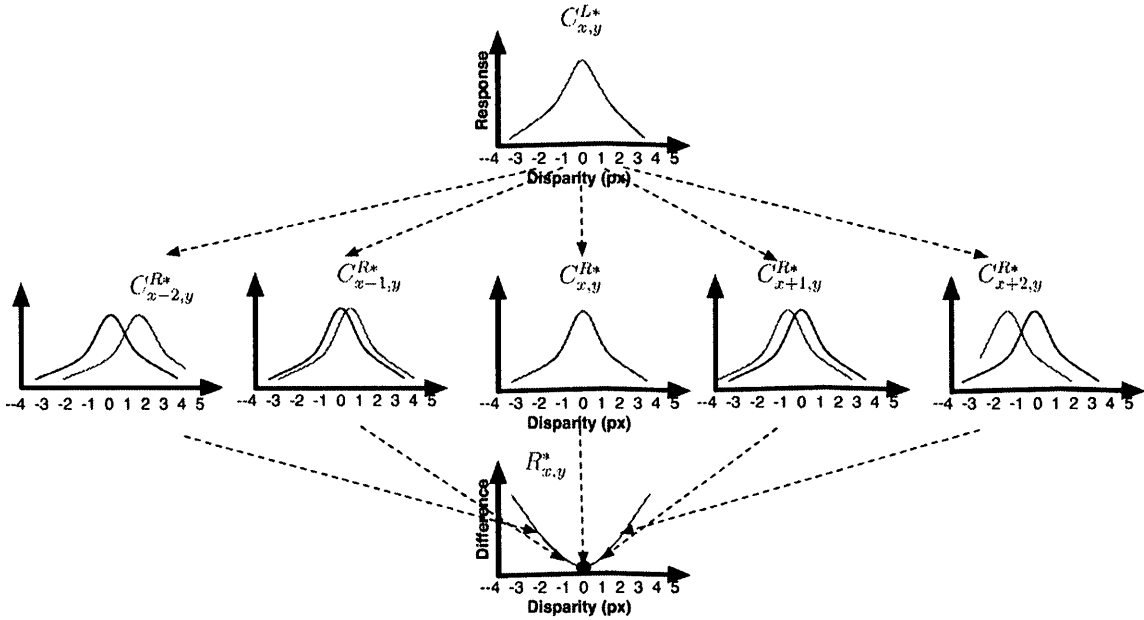


Figure 6.4: Biologically-plausible implementation of the left-to-right check. For each pixel x, y , the difference between the left-to-right disparity profile $C_{x,y}^{L*}$, shown in red, and all the potential matches $C_{x+d,y}^{R*}$, shown in black, is computed. This computation is performed by $R_{x,y,d}$ neurons. The minimum of the $P_{x,y}^*$ responses, shown as the red dot on the blue curve in the bottom row, is chosen as the left-to-right check response for location x, y .

the two profiles (appropriately shifted) have peaks at the same location. Thus for locations x, y with well defined disparity d' , the response of the neuron $R_{x,y,d'}$ should be very close to zero, indicating that both ‘left-to-right’ and ‘right-to-left’ computations yielded the same disparity estimate. In contrast, in monocularly occluded regions, which lack well defined disparity, all $R_{x,y,d}$ are likely to yield results much higher than zero. Thus at each location x, y out of all possible responses of $R_{x,y,d}$ neurons, the magnitude of the minimum response represents the likelihood that this location is monocularly occluded. Accordingly, $LRC_{x,y}$ neurons output the minimum $R_{x,y,d}$ response:

$$LRC_{x,y} = MIN(R_{x,y}^*) \quad (6.10)$$

where for convenience the collection of $R_{x,y,d}$ responses for all d in $[-d_m, d_m]$ is denoted $R_{x,y}^*$. Note that textureless areas would be identified as binocular by these neurons since the shape of their disparity profiles would be flat and identical (or similar) in both left-to-right and right-to-left computations thus producing a small $LRC_{x,y}$ response.

Match-goodness

The match-goodness metric is defined as the ratio of the strength of the maximum response for a given location to the maximum response within the whole population of disparity detectors. For monocularly occluded pixels the ratio will be lower than for binocular areas since no true match exists. Consequently, this ratio is then subtracted from 1, such that a higher response indicates a higher likelihood of the pixel to be occluded, and rectified:

$$MG_{x,y} = \left[1 - \frac{MAX(C_{x,y}^*)}{M^*} \right]_{\theta_1} \quad (6.11)$$

where $[x]_{\theta_1}$ indicates rectification with respect to the threshold θ_1 (notation adopted from Reynolds and Heeger (2009)) or with respect to 0 where $[x]_0$ is used (this notation is used throughout this chapter). Rectification is defined here as the nulling of a signal if it falls below the specified threshold. M^* is defined as follows:

$$M^* = \max(C_{x,y,d}) \text{ for all possible } x, y \text{ and } d. \quad (6.12)$$

Combining monocular detection cells

The two types of monocular occlusion detection cells complement each other as discussed above and hence the output of the two mechanisms is combined in a third

type of cell that signals the presence of monocularly occluded regions. These cells only fire when the combined input from the two different mechanisms is higher than a certain threshold θ_2 :

$$OCC_{x,y} = \left[n(MG_{x,y}) + n(LRC_{x,y}) \right]_{\theta_2} \quad (6.13)$$

6.5.3 Computing the width of monocular areas

After the monocular occlusions are detected, their width must be determined to compute the depth of monocular areas. The width computation is performed by a population of cells that have an end-stopped architecture with a wide excitatory centre (area EC) positioned between two narrow inhibitory bands (areas IB) and which receive input from the OCC cells described in Equation 6.13. The cells have different widths and respond optimally when a monocular region has a width and location matching those of the excitatory centre. These cells are illustrated in Figure 6.5. The response of each cell is normalized by the width of the excitatory region (this is done to make sure that large cells do not have a larger optimal response than smaller cells). For each possible occlusion width and for each location x, y there is a population of end-stopped cells with different parts of their excitatory centres positioned at x, y :

$$ES_{x,y,w,s} = \left[\frac{\sum_{x',y' \in EC-s} OCC_{x',y'} - \sum_{x',y' \in IB-s} OCC_{x',y'}}{w} \right]_{\theta_3} \quad (6.14)$$

ES cells fire when their response is larger than a threshold θ_3 . The response of the population of ES neurons tuned to different excitatory center locations s about the position x, y is denoted $ES_{x,y,w}^*$. The final response $W_{x,y,w}$ to each potential width w at each location x, y is equal to the response of the maximally excited neuron in $ES_{x,y,w}^*$. The response of W neurons is modulated by weighted input from the OCC neurons such that the width cells fire only if the location x, y was identified as occluded in the previous step:

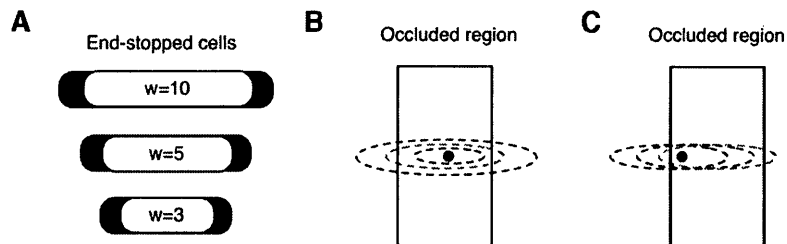


Figure 6.5: (A) The width of monocular occlusions is signalled by end-stopped cells of different widths. The excitatory centre changes size while the inhibitory side lobes have the same size for all cells. (B) At each location x, y there is a population of end-stopped cells tuned to different widths. (C) For each width there is a population of cells for which different parts of their excitatory centres fall on x, y . The largest response will be elicited from the cell with the excitatory centre which matches the size and the location of the monocularly occluded area (shown in dashed red line in (B) and (C)).

$$W_{x,y,w} = \left[\text{MAX}(ES_{x,y,w}^*) - OCC_{x,y} \times \gamma_1 \right]_0 \quad (6.15)$$

where γ_1 is the weight of the inhibitory connections from OCC cells. The response of the population of W neurons tuned to different widths at position x, y is denoted $W_{x,y}^*$ and referred to as the occlusion width profile (analogous to the disparity profile).

6.5.4 Other constraints

Reliability

The reliability metric is a measure of how reliable or robust disparity estimates are for a particular location. It is somewhat similar to the ‘peak ratio metric’ used to predict the locations where potential false matches could be made (Egnal et al., 2004; Little and Gillett, 1990), however, here it is used in the final computation of the 3D surfaces as a weight on excitatory connections between neurons. In locations devoid of texture, a maximal response might be obtained for several disparities. Although response magnitude at these disparities might be large (high match-goodness), these

disparity estimates are not reliable. Accordingly, reliability is estimated as the difference between the magnitude of the largest response of the population at each location x, y and the magnitude of the second largest response at this location (see Figure 6.6).

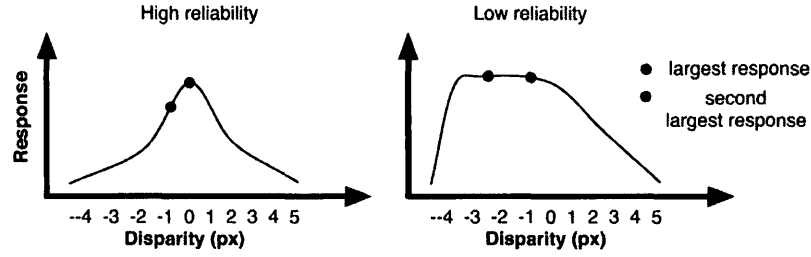


Figure 6.6: Computing reliability. Reliability is computed as the difference between the largest response and the second largest response of the population of neurons at each location. This difference is high when the response curve is steep as it is shown in the left-hand graph. In this case the disparity estimate can be considered reliable. On the other hand when the curve is flat since the population is responding similarly to many different disparities, the difference between the two points is small. In this case the disparity estimate is not reliable.

First, a population of neurons computes the difference between the maximum response in the population at x, y and the maximum response of the population with a response to one particular disparity d zeroed:

$$RE_{x,y,d} = MAX(C_{x,y}^{*-d}) - MAX(C_{x,y}^*) \quad (6.16)$$

where $C_{x,y}^{*-d}$ is the disparity profile at location x, y with the response to disparity d zeroed. If the true disparity at location x, y is d' such that $MAX(C_{x,y}^*) = C_{x,y,d'}$ then all the $RE_{x,y,d}$ for $d \neq d'$ will give a zero output. Only the neuron $RE_{x,y,d'}$ could have a response different from zero. This neuron will output the difference between the maximum and second maximum response. Consequently, reliability is computed as the maximum of all the $RE_{x,y,d}$ responses:

$$REL_{x,y} = MAX(RE_{x,y}^*) \quad (6.17)$$

Overall match goodness

Overall match goodness indicates the relative strength of the response of disparity detectors at each location. It is used in the final processing stage to build a 3D surface as a weight on excitatory connections between neurons. Match goodness for each location x, y is computed as the ratio of the maximum response for a given pixel to the maximum response within the whole population

$$OMG_{x,y} = \frac{MAX(C_{x,y}^*)}{M^*} \quad (6.18)$$

6.5.5 Edge detection

Disparity and luminance edges are detected and used to control the spreading of disparity signal from one object to another. Disparity and luminance edge maps are combined to create one edge map.

Disparity edges

Disparity edges are computed using simple on-off cells. These bipartite cells have an excitatory half and an inhibitory half and a vertically elongated shape. They operate on the output of the occlusion detection cells OCC . They signal optimally when their excitatory region (area ER) is positioned on a disparity edge (OCC cells are 'on' at this locations) and their inhibitory region (area IR) is located just of this edge.

$$DE_{x,y} = \sum_{x',y' \in ER} OCC_{x',y'} - \sum_{x',y' \in IR} OCC_{x',y'} \quad (6.19)$$

Luminance edges

Luminance edge detection is performed with the Sobel edge detector. It is a simple operator that consists of convolving two different linear filters (for vertical and horizontal edges) with the image to obtain a map of luminance gradients. This edge

detector has built-in version in MATLAB and was chosen for the model since it is biologically-plausible according to the working definition.

Luminance and disparity edges are added to produce a combined response:

$$ED_{x,y} = DE_{x,y} + LE_{x,y} \quad (6.20)$$

Instead of computing disparity and luminance edges, the model also includes an option of using a predefined edge maps loaded from an additional file.

The edge map is then used to perform rough object segmentation. This process is not implemented in a biologically-plausible way since it is a very complex process (with a whole literature dedicated to its understanding) and the implementation of a biologically-plausible version of this process is beyond the scope of this dissertation. It is assumed here that object segmentation is performed by a higher level process that sends feedback connections to the neurons in the lower visual areas. Feedback modulation based on figure ground relationships and object segmentation is a well documented physiological phenomenon (Angelucci et al., 2002; Hupe et al., 1998; Schoenfeld et al., 2003). Hence it is assumed that it is possible to provide a biologically-plausible implementation for this module. For the purposes of the model formulation the output of the higher-level neurons is denoted as $OBJ_{x,y,x',y'}$. It's output is positive when two locations x, y and x', y' belong to the same object and zero otherwise.

6.5.6 3D surface cells - final computation of disparity

In the final stage of the model, the 3D surface is constructed by aggregating information from all the previous stages in an iterative manner and assigning final disparities to both monocular and binocular locations. Disparities are computed differently for locations that were identified as monocularly occluded and those that were identified as binocular. The 3D surface neurons - $3D_{x,y,d}$ - sum weighted responses of two types of neurons: $MON_{x,y,d}$ that compute the depth for monocularly occluded regions and

$BIN_{x,y,d}$ that compute the depth for binocular regions. The input from these two types of neurons is modulated by inhibitory influences such that $BIN_{x,y,d}$ is suppressed if x, y is identified as a monocular location and $MON_{x,y,d}$ is suppressed when x, y is a binocular location:

$$3D_{x,y,d} = n \left(\left[BIN_{x,y,d} - OCC_{x,y} \times \gamma_1 \right]_0 + \left[MON_{x,y,d} - (\gamma_2 - OCC_{x,y} \times \gamma_1) \right]_0 \right) \quad (6.21)$$

where γ_2 is a constant. In all cases below where the output of the $3D_{x,y,d}$ neurons is used at the first iteration, when this output is just background noise, the disparity profiles are taken instead from the corresponding $C_{x,y,d}$ complex energy neurons. In other words, for the first iteration:

$$3D_{x,y,d} = C_{x,y,d} \quad (6.22)$$

Reliability and overall-match goodness are updated using the output of the $3D_{x,y,d}$ neurons after each iteration.

Depth in monocular areas

For monocularly occluded locations disparity is derived based on occlusion geometry, that is the width of the monocular region and the eye of origin, and the disparity information in neighbouring binocular areas. An assumption is made that the monocular region results from an occlusion arrangement (not from camouflage). The disparity estimates for occluded regions are obtained by collecting support from the neighbouring binocular regions via $BS_{x,y,d}$ cells and the neighbouring monocular regions via $MS_{x,y,d}$ cells:

$$MON_{x,y,d} = BS_{x,y,d} + MS_{x,y,d} \quad (6.23)$$

The contributing binocular regions are areas of size $H \times W$ to the left, right, above

and below the monocular region, labeled as NL, NR, NA and NB accordingly. Before being added, the disparity profiles $3D_{NX}$ in each region are averaged and weighted by the mean reliability RE_{NX} in the region. For an occlusion in the left eye the response is computed in the following way:

$$BS_{x,y,d} = \overline{3D}_{NL} \times \overline{RE}_{NL} + \overline{3D}_{NA,d} \times \overline{RE}_{NA} + \overline{3D}_{NB,d} \times \overline{RE}_{NB} + \max(\overline{3D}_{NR,d'} \times W_{x,y,w}) \times \overline{RE}_{NR} \quad (6.24)$$

where $d', w \in [d' + w = d]$

$$\overline{3D}_{NX,d} = \frac{\sum_{x',y' \in NX} 3D_{x',y',d}}{H \times W} \quad (6.25)$$

$$\overline{RE}_{NX} = \frac{\sum_{x',y' \in NX} RE_{x',y'}}{H \times W}$$

Since an occlusion arrangement is assumed, the disparity signal coming from the left, top and the bottom neighbouring areas are taken without adjustment. The occluded area is assumed to be coplanar with these regions. On the other hand, the area to the right of the occlusion is the occluding edge and it is suppose to have a more crossed disparity than that of the occluded region. According to occlusion geometry (i.e. the minimum depth constraint), the disparity of the occluded region in the left eye should be equal to the disparity of the occluding right edge plus the width of the occluded region. Thus the NR disparity signal is shifted by the width of the monocular region before the support from it is collected. Since the responses of all neurons in the model are distributed and not binary, the width profile $W_{x,y}^*$ and the mean disparity profile of the NR region $\overline{3D}_{NR}^*$, have to be carefully combined in order to be added to the new disparity profile for location x, y . Each possible disparity d can be achieved through several combinations of disparities d' and widths w . For example, disparity $d = 5$ can result from $(w = 3, d' = 2)$, $(w = 4, d' = 1)$,

($w = 5, d' = 0$) and other combinations. Consequently, for each such disparity d support is collected from all possible combinations of w and d' by multiplying $W_{x,y,w}$ with $\overline{3D}_{NR,d'}$. The maximum response of these possible w and d' combinations is taken as the response to each disparity d . This part of the computation in Equation 6.24 is illustrated in detail in Figure 6.7.

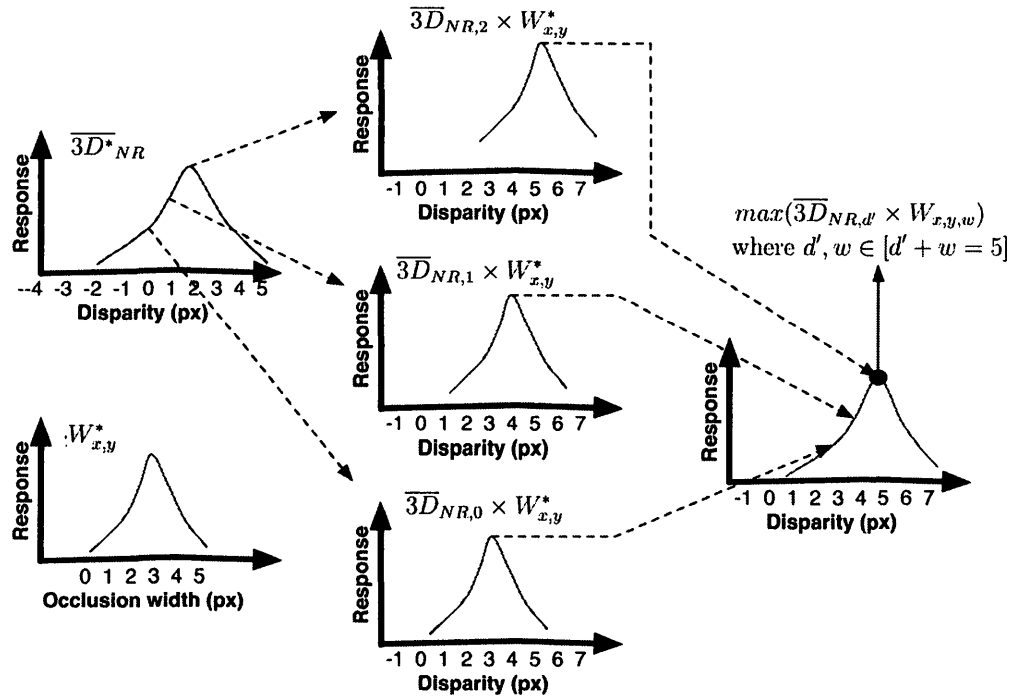


Figure 6.7: Using occlusion geometry to assign disparity in occluded regions. The red curve represents the mean disparity profile in the NR region, which is suppose to be occluding position x, y . The blue curve in the left column shows the width profile of location x, y . According to occlusion geometry the width of the occluded region needs to be added to the disparity of the occluding edge to produce the disparity of the occluded region. Since disparity and width representations are distributed, all possible combinations of widths and disparities need to be considered. The central column shows the width profile shifted by disparities $d' = 2, d' = 1$ and $d' = 0$ and multiplied by the appropriate response in the disparity profile. The right column shows the final disparity profile. Each point on this profile satisfies $d' + w = d$ and represents the maximum response out of all such combinations for disparity d .

Support from the neighbouring monocular locations is collected by summing the

disparity profiles of the locations within the support region SR that are identified as occluded. The contributions from each location within the support areas are weighted by a 2D Gaussian with standard deviation σ_1 centred on x, y , by the reliability and match goodness of each particular neighbouring location. Moreover, locations falling within the support neighbourhood but belonging to a different object are inhibited by feedback from a higher-level process (see Section 6.5.5) and do not contribute to the computation:

$$MS_{x,y,d} = \sum_{x',y' \in SR} \left[3D_{x',y',d} \times Gauss_{x',y'} \times RE_{x',y'} \times MG_{x',y'} \times OBJ_{x,y,x',y'} - (\gamma_2 - OCC_{x,y} \times \gamma_1) \right]_0 \quad (6.26)$$

Once the final disparity profile of an occluded location is established, the disparity signal can propagate into binocular areas with low reliability by the means of specially dedicated neurons. This propagation takes place if the reliability of the binocular area that is suppose to occlude the monocular area according to the geometry is low. This mechanism is illustrated in Figure 6.8. Disparity from occluded areas is propagated only horizontally and this propagation is stopped when either the next edge (luminance or disparity) is reached or reliability increases beyond a certain threshold. The propagation is mediated through long range excitatory connections.

$$PROP_{x+s,y,d} = \left[\max(3D_{x,y,d'} \times W_{x,y,w}) - (\gamma_2 - OCC_{x,y} \times \gamma_1) - ED_{x+s,y} \times \gamma_3 - RE_{x+s,y} \times \gamma_4 \right]_0 \quad (6.27)$$

where $d', w \in [d' + w = d]$ and γ_3 and γ_4 are the weights on the connections with edge detectors and reliability computing cells accordingly.

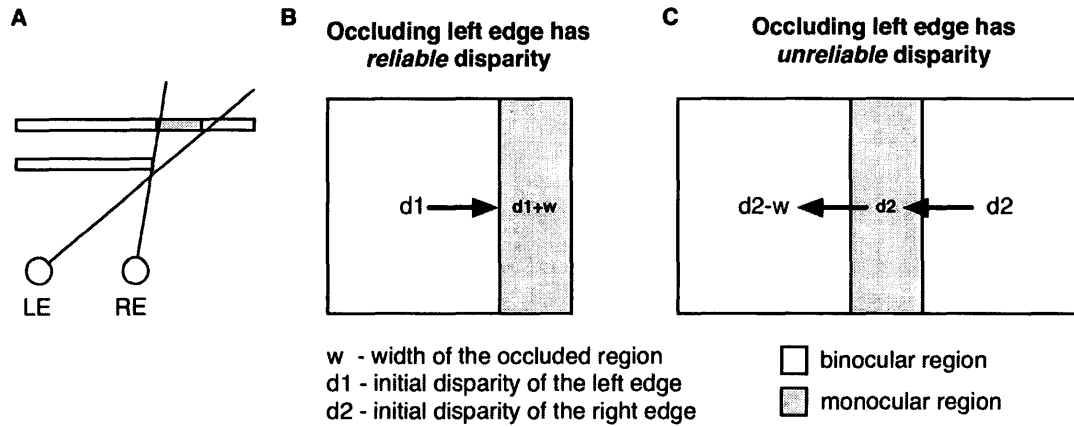


Figure 6.8: Computing and propagating disparity in monocular regions. (A) A bird's eye view of the observed scene. A foreground surface occludes part of the background surface from the left eye creating a monocular occlusion of width w . In all panels, binocular areas are shown in white and monocular areas in grey. (B) When the occluding foreground surface has a reliable disparity (it is textured or has well defined luminance edges), its disparity $d1$ is used to compute the disparity of the monocular region - $d1 + w$. In (C) the occluding surface does not have a reliable disparity (e.g. when the area to the right of the monocular occlusion has no texture or luminance defined edges). In this case the disparity of the monocular region - $d2$ is estimated using the disparity of the binocular area to the right of it. Then the disparity of the unreliable area to the left of the monocular occlusion is determined by using the occlusion width - $d2 - w$ - and propagated over the left edge.

Depth in binocular areas

The depth of binocular locations x, y is computed by accumulating support from the binocular locations in the surrounding region SR and the disparities propagated by the monocular occlusion regions (if any). In both cases the support is weighted by a 2D Gaussian with standard deviation σ_2 centred on x, y and by the reliability and match goodness of each particular neighbouring location. Moreover, locations falling within the support neighbourhood but belonging to a different object are inhibited by feedback from a higher-level process (see Section 6.5.5) and do not contribute to the computation:

$$\begin{aligned}
BN_{x,y,d} = & \sum_{x',y' \in SR} [3D_{x',y',d} \times Gauss_{x',y'} \times RE_{x',y'} \times MG_{x',y'} \times OBJ_{x,y,x',y'}] + \\
& \sum_{x',y' \in SR} [PROP_{x',y',d} \times Gauss_{x',y'} \times (\frac{1}{RE_{x',y'}}) \times OBJ_{x,y,x',y'}] \quad (6.28)
\end{aligned}$$

6.5.7 Final selection of disparity

Final depth maps were computed from the response of the $3D_{x,y}^*$ neurons after all the model iterations were completed using the zero-bias method. In this method the disparity corresponding to the maximum response of the population is taken as the true disparity at each point. If more than one disparity generates the maximum response, the disparity closer to zero is taken as the true disparity. This method is biologically-plausible since psychophysical studies showed that in ambiguous cases the visual system tends to prefer smaller disparities over larger ones (Goutcher and Mamassian, 2005). Moreover, this method gave estimates that were closest to the estimates made by the observers in the experiments of Chapter 4 (see Section 4.4).

6.6 Model evaluation

The model was tested on the battery of 25 images shown in Table 6.2. This battery includes the vast majority of the monocular occlusion stimuli types used in the psychophysical literature and in this thesis. This allows for a thorough validation of the model by comparing model performance against available psychophysical data. For each of the monocular occlusion stimuli, several occlusion widths were tested as specified in the third column of Table 6.2. For the RDS stimulus several disparities were tested. All synthetic images were generated using custom made MATLAB scripts together with depth ground truth and occlusion ground truth maps. The ground truth maps for the intermediate occlusion width / disparity values are shown

in Table 6.2 with the occluded regions outlined in red. For each of the synthetic images the coordinates of a region of interest were specified where depth is postulated to be perceived on the basis of monocular occlusions (or disparity in RDS). These regions are shown in blue on the ground truth maps (in some cases the blue outlines overlap with the red). The image and the ground truth of the real scene (captured under laboratory conditions) was taken from the Middlebury database (Scharstein and Szeliski, 2002). All the stereograms were rectified such that the same pixels fell along the same epipolar lines in the left and right half-images.

Table 6.2: Stimuli used to test the model.

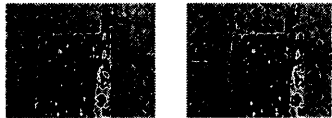
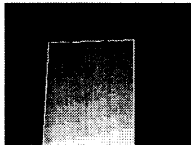





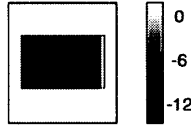
| Stimulus | Ground truth | Occlusion widths / Disparities (px) | Size (px) |
|---|--|-------------------------------------|-----------|
| <p>Map</p>  |  | NA | 280 × 216 |
| <p>RDS</p>  |  | -4,-8,-12 | 200 × 200 |
| <p>Random-dot frame</p>  |  | disparity -10, widths 4.7,10 | 260 × 194 |
| <p>Illusory occluder of Liu et al. (1994)</p>  |  | 8,12,16 | 226 × 324 |

Table 6.2 – continued from previous page










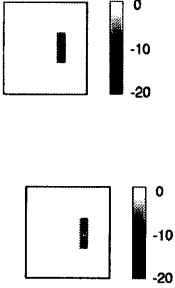
| Stimulus | Ground truth | Occlusion widths / Disparities (px) | Size (px) |
|--|---|-------------------------------------|-----------|
| <p data-bbox="391 491 673 544">Illusory occluder of Gillam and Nakayama (1999)</p>  |  | 8,12,16 | 324 × 246 |
| <p data-bbox="375 710 691 732">Two-object arrangement with a bar</p>  |  | 20,24,28 | 280 × 350 |
| <p data-bbox="375 902 691 923">Two-object arrangement with a disc</p>  |  | 20,24,28 | 280 × 350 |
| <p data-bbox="464 1285 599 1306">Monocular gap</p>  |  | 8,12,16 | 182 × 298 |

Table 6.2 – continued from previous page

| Stimulus | Ground truth | Occlusion widths / Disparities (px) | Size (px) |
|--|---|-------------------------------------|------------------|
| <p style="text-align: center;">Monocular intrusion</p>  |  | <p>16,20,24</p> | <p>324 × 260</p> |

Note that for the many of the cases presented here, there are no objective ground truth maps. The ground truth maps are either based on theory, on empirical data or both. For example, it is not clear what disparity should be assigned to textureless background surfaces in most of the synthetic images shown in Table 6.2 since many disparities would elicit a maximum response from the population of disparity detectors. I chose to assign a zero disparity to these regions since the visual system has a small disparity bias, and this preference is also reflected in the model in the form of the zero-bias disparity selection method. In some cases, Table 6.2 shows several ground truth maps for one stimulus since there are either several interpretations consistent with the stimulus configuration and occlusion geometry or observers reported several different percepts. In the analysis of the model results for these stimuli I use the ground truth map that is the closest to the model output.

The same model parameters were used for all stimuli during the testing phase. The parameters are listed in Table 6.3 in order of their appearance in the text. The edge maps used in the computations were drawn by hand (or pre-computed) and loaded from files. This was done in order to eliminate the effect of poor edge detection on the performance of the model and obtain the best results possible. This is a

Table 6.3: Model parameters used in testing.

| Parameter | Value | Parameter | Value |
|---|--|--------------|---------------|
| Energy neurons (EN) preferred spatial frequencies | 0.0625, 0.0877, 0.125, 0.1786 and 0.25 pixels per degree | θ_3 | 0.2 |
| EN preferred orientations | 0, 30, 60, 120, 150 degrees | γ_1 | 10 |
| EN Preferred disparities $[-dm, dm]$ | -30 to +30 pixels | γ_2 | 8 |
| EN RF aspect ratio | 2 | $H \times W$ | 2×15 |
| θ_1 | 0.3 | γ_3 | 10 |
| θ_2 | 1 | γ_4 | 10 |
| σ_1 | 15 | σ_2 | 5 |

reasonable alteration since edge detection is not the focus of the model. In pilot experiments, where the simulations were ran with the edges computed on the fly using edge detectors, the results were close to those reported here but somewhat noisier in a few cases.

The reported rates of true positives for occluded pixels were computed as the ratio of occluded pixels detected by the model to the total number of occluded pixels. The reported rates of false positives were computed as the ratio of binocular pixels signalled as occluded to the total number of occluded pixels. The two ratios represent an average of the ratios for the three occlusion widths /disparities unless specified otherwise.

6.7 Results

The maps computed by the model for each stimulus type are shown in Figures 6.9-6.17. All figures except for Figure 6.9, show (A) the stimuli, (B) detailed results of simulation trials with one of the occlusion widths (or disparities) and (C) plots showing the estimated mean disparity in the regions of interest for all occlusion widths (or disparities). Each detailed results plot (B) shows the following data:

1. The ground truth map where darker colours indicate closer depths.
2. The initial depth map output by the energy neurons where darker colours indicate closer depths.
3. The final depth map output by the model where darker colours indicate closer depths.
4. The occlusion ground truth map where white indicates occluded pixels.
5. The computed occlusion map where brighter colours indicate greater probability the pixel is occluded.
6. The computed map of occlusion widths where brighter colors indicate larger widths.

Figure 6.9 showing the results for the map image, does not have the occluded region maps since occlusion ground truth was unavailable for this image. Instead the figure shows the error maps that were computed by subtracting the ground truth from the initial and the final depth maps. The results for each type of stimulus are discussed separately below.

On average, the complete model reduces the percent of pixels with incorrect disparities by half compared to initial estimates made by the energy neurons. The overall rate of true positives in occlusion detection, averaged across all images, is 80% and the false positives is 30%. This performance was obtained with the same set of parameters for all images. Performance of the model can be greatly improved for each individual stimulus by selecting an optimal set of parameters for that stimulus.

Map photograph Figure 6.9 shows the model output for the a photograph showing a map leaning against a background. The initial output of the energy neurons provides very noisy estimates in the monocularly occluded region near the right edge of the map. The model improves on this result by detecting the monocular region and assigning it a proper disparity. The improvement is evident when the two error maps are compared.

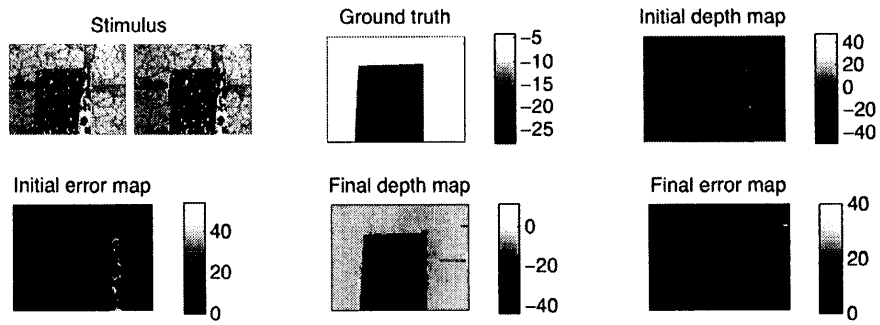


Figure 6.9: Model results for the map image.

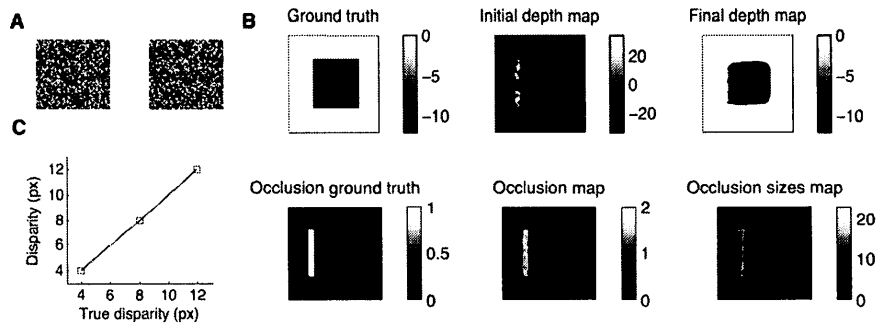


Figure 6.10: Model results for the RDS. (A) Stimulus. (B) Results for disparity -12. (C) Results for three different central square disparities.

RDS

Model results for the RDS are shown in Figure 6.10. As expected, the initial disparity estimates provided by the energy model are accurate in all areas except for the monocularly occluded region, to the left of the central square, which is characterized by spurious matches. The model corrects this problem by detecting the occluded pixels (80% true positives, 12% false positives) and assigning correct disparities in these locations. As shown in plot (C), model estimates of the disparity of the central square correspond to the predicted ones.

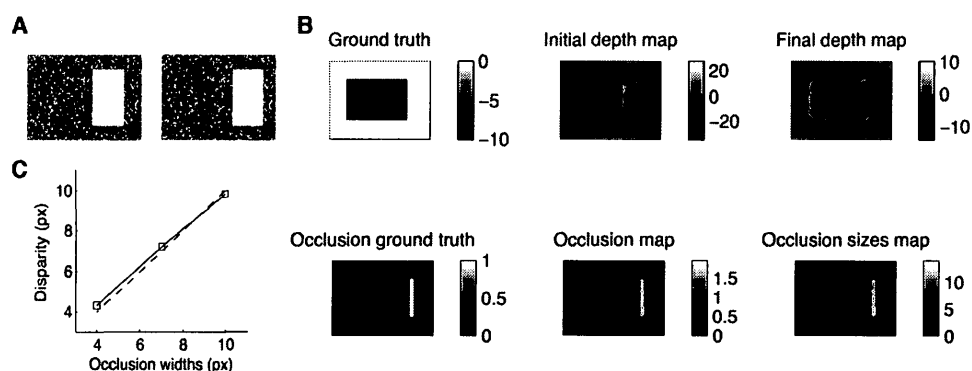


Figure 6.11: Model results for the random-dot frame stimulus. (A) Stimulus. (B) Results for disparity -10 and occlusion width 10. (C) Results for three different occlusion widths.

Random-dot frame stimuli

The results for the stimulus introduced in Chapter 2 are shown in Figure 6.11. The estimates of the energy model for the textured areas are accurate as expected, but the estimates for the central blank region are spurious and do not correspond to the percept of an illusory occluder at the same depth as the textured square. The model improves on these results by detecting the monocularly occluded region (71% true positives, 30% false positives) and propagating crossed disparity across the blank region, reconstructing the illusory occluder. Since the occluded region is textured, the width of the occluded region varies somewhat along its height, as can be seen in the occlusion and occlusion width maps. A similar, but a smaller effect, can be seen

in Figure 6.10 for the RDS. For the random-dot frame stimulus this variability results in a somewhat ‘striped’ disparity map for the blank region. This issue is expanded on in Section 6.8. Plot (C) shows that the disparity estimates for the illusory occluder increase as the occlusion width increases and the estimates lie on the predicted line in agreement with the psychophysical results reported in Chapter 2.

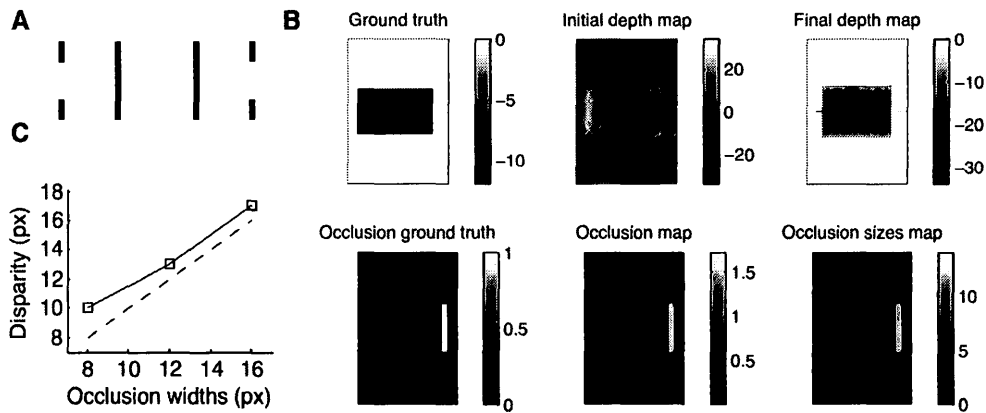


Figure 6.12: Model results for the Gillam & Nakayama stimulus. (A) Stimulus. (B) Results for occlusion width 8. (C) Results for three different occlusion widths.

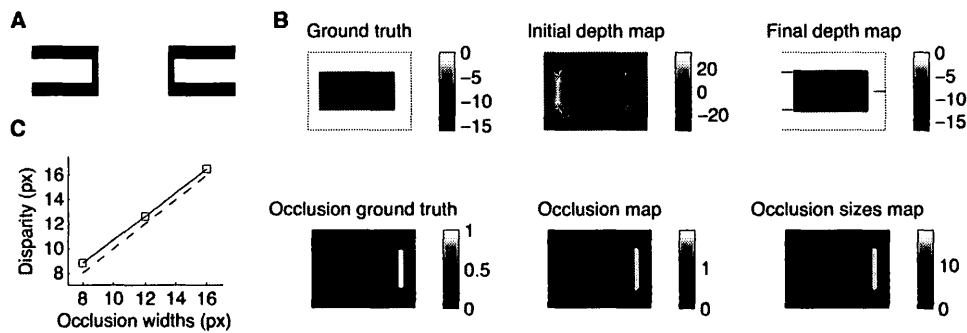


Figure 6.13: Model results for the Liu et al. stimulus. (A) Stimulus. (B) Results for occlusion width 16. (C) Results for three different occlusion widths.

Illusory occluder stimuli

The results for the two types of illusory occluder stimuli - Gillam and Nakayama (1999) and Liu et al. (1994) - are shown in Figures 6.12 and 6.13 respectively. Evidently for this type of stimuli the model is performing quite well. The monocularly occluded areas are detected accurately (95% true positives, 8% false positives) and the illusory surfaces are fully reconstructed although the initial estimates of disparity from the energy model fall short of the ground truth. As in the psychophysical data (see Chapter 3), the model predicts quantitative depth in both cases as shown in the (C) plots of Figures 6.12 and 6.13. Moreover, the model estimations replicate both the qualitative and the quantitative aspects of observer data. In the case of the Liu et al. (1994) stimulus, the estimates are very close to the predicted disparity values, while in the case of Gillam and Nakayama (1999) stimulus there is an overestimation of disparity similarly to the psychophysical data (see Figure 3.4 and 6.12). There are minor artifacts where disparity is propagating beyond the illusory surface, but these are quite small and most likely result from stray pixels being identified as occluded.

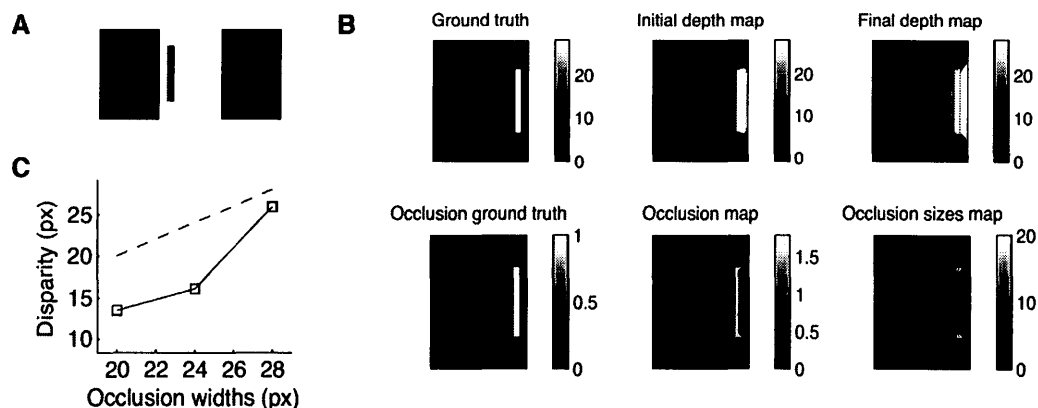


Figure 6.14: Model results for a two-object arrangement with a bar. (A) Stimulus. (B) Results for occlusion width 20. (C) Results for three different occlusion widths.

Two object arrangements

The results for the two types of two-object arrangements with a bar (Nakayama and Shimojo, 1990) and a disc (Gillam et al., 1999) - are shown in Figures 6.14 and

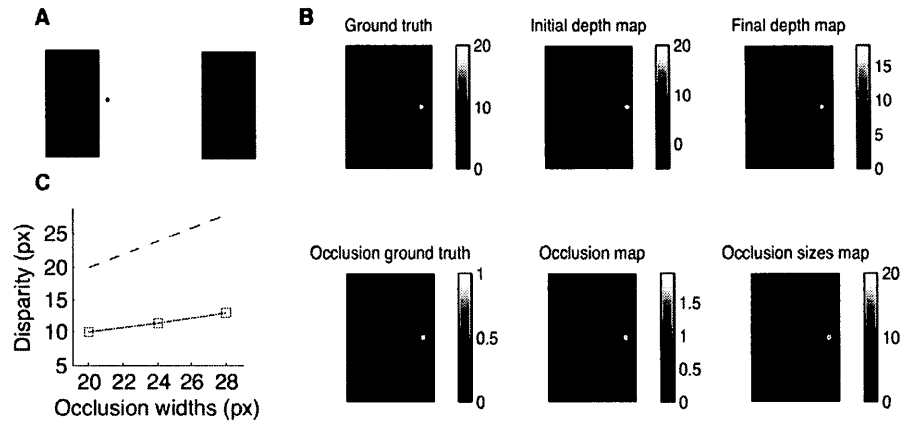


Figure 6.15: Model results for a two-object arrangement with a disc. (A) Stimulus. (B) Results for occlusion width 20. (C) Results for three different occlusion widths.

6.15 respectively. The outline of the disc is shown in red on all the maps to highlight that although on the initial depth map the pocket of uncrossed disparity is located beyond the contours of the disc, on the final depth map the disc is localized correctly. The occluded area is detected fairly accurately in this case (100% true positives, 30% false positives) although it is overestimated somewhat, most likely due to the relatively low spatial resolution of the energy neurons in comparison to the small size of the disc. Plot (C) shows that as occlusion width increases, so do the model disparity estimates. Importantly, disparity is underestimated quite substantially just like as it is in the psychophysical data provided in Chapter 4 (see Figure 4.4).

In the case of the bar, the initial disparity map provides a fairly accurate estimate of disparity, albeit for an area larger than that of the bar. These estimates are based on double-matching of the bar to the binocular rectangle since the uniqueness constraint is not imposed in the model. The monocularly occluded area detected by the model is narrower than the bar (45% true positives, 1% false positives). This occurs since the goodness of match of the right edge of the bar is quite high (due to double-matching) so the goodness-of-match occlusion detection metric fails to signal these parts as occluded. The model correctly predicts an increase in quantitative depth as the

occlusion width increases as shown in (C). Importantly, these disparity estimates are somewhat smaller than the predicted ones, a result that mirrors psychophysical data shown in Chapter 4 (see Figure 4.4). This occurs because the disparity estimates provided by the occlusion neurons and the binocular neurons are averaged in the computation of the disparity of the region of interest.

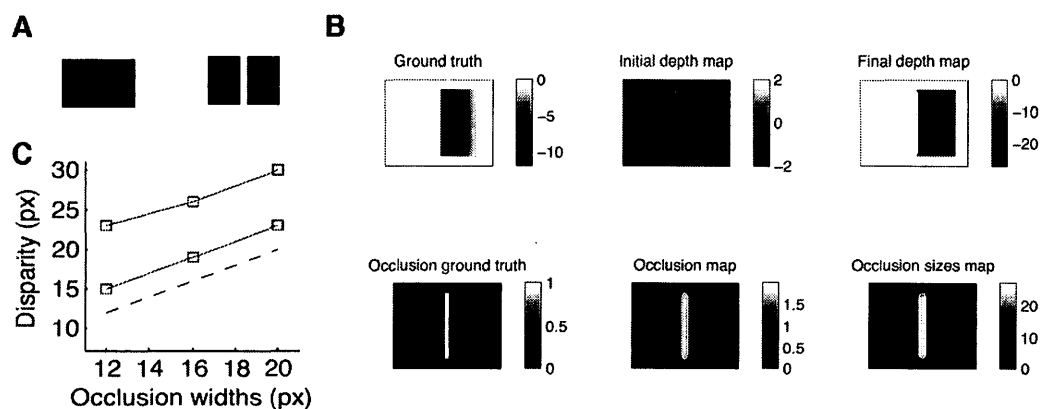


Figure 6.16: Model results for the monocular gap stimulus. (A) Stimulus. (B) Results for occlusion width 12. (C) Results for three different occlusion widths.

Monocular gap

The results for the monocular gap stimulus (Gillam et al., 1999) are shown in Figure 6.16. The model provides a disparity map that most closely resembles one out of three possible interpretations of this stimulus shown in Table 6.2. The model correctly detects the occluded area, however, its width is overestimated (93% true positives, 66% false positives) which results in greater disparity estimates for the side to the right of the gap. The estimated disparity also does not decrease with eccentricity towards the zero disparity right edge. This might be an artifact of the method of final disparity selection employed in the model. Only the disparity generating the maximum response is chosen as the true disparity of each pixel. When I examined the disparity profile of the pixels within the area closely, they showed two clear peaks similar in magnitude to each other. Hence a different disparity selection method could

yield a disparity which would be a weighted mix of the two peak disparities resulting in a gradual decrease of the disparity in the final map.

The model correctly predicts quantitative depth based on the width of the monocular region as shown in Figure 6.16-C. As discussed above, the disparities are overestimated due to the overestimation of the width of the occluded region. This aspect strictly depends on the threshold used in the goodness-of-match occlusion detection metric. The red line in (C) shows that when the threshold θ_1 (see Equation 6.11) is increased from 0.3 to 0.5, with all other parameters held constant, the model disparity estimates lie very close to the predicted disparities.

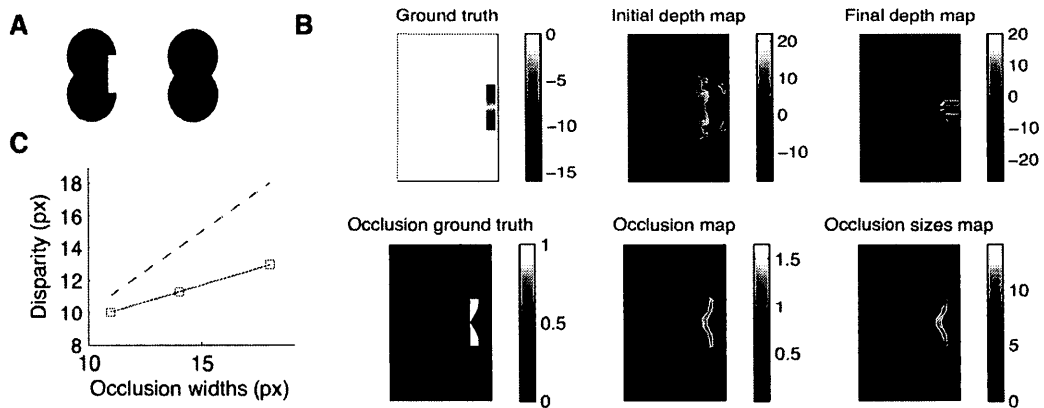


Figure 6.17: Model results for the monocular intrusion stimulus. (A) Stimulus. (B) Results for occlusion width 24. (C) Results for three different occlusion widths.

Monocular intrusion

The results for the monocular intrusion stimulus (Cook and Gillam, 2004) are shown in Figure 6.17. The initial estimate of the energy model shows a small curved surface on the edge of the figure eight, with crossed disparity of correct magnitude along the curviest points of the figure eight and a disparity close to zero at the midline of the figure eight. This is consistent with a one-to-one matching of the curved contour on one eye to the straight contour in the other eye, as was proposed in Chapter 4. However the disparity map is noisy and does not extend to the right of the figure eight

(creating an occluding surface) as it does when the figure is viewed stereoscopically. The model detects the narrow contour of the side of the figure eight as occluded but not the whole intrusion (true positives 54%, false positives 20%). This happens since the match goodness of the inside area of the occlusion is quite good and hence it is not signalled as occluded. The disparity is correctly propagated towards the right of the image, reconstructing an illusory intrusion. The model also predicts the increase in perceived depth with increasing occlusion width. Plot (C) shows the average disparity in the region of interest for both the ground truth and the computed disparity map.

6.8 Discussion

This chapter described a biologically-plausible model of depth from disparity and monocular occlusions. This model is based on principles established in psychophysical experiments described in this thesis and in the previous literature. Monocular occlusions are detected explicitly and take an active role in the construction of the final 3D map of surfaces in the image. Depth in occluded areas is established on the basis of monocular geometry and illusory occluding surfaces are constructed in cases when the disparity estimates for the occluding edge have low reliability. The features and predictions of the model and improvements that can be made to it are discussed below.

6.8.1 Model features

The proposed model offers several improvements over the existing biologically-inspired models of depth from disparity and occlusions along with several innovations which can advance the area of biologically-plausible modelling of stereopsis in general.

First, the model has been implemented in a biologically-plausible way with a distributed representation of neuronal firing rates all throughout the computation until the final selection of disparities is performed. This is an improvement over three

of the preceding models which used binary representations at all (or at several) stages.

Second, this is the first biologically-inspired model to use the occlusion width explicitly to compute disparity in occluded areas and to proactively propagate the disparity from the occluded areas into areas of low reliability. This feature allows the model to reconstruct illusory occluders and predict depth percepts in monocular gap and monocular intrusion stimuli. Other models of da Vinci stereopsis would not be able to compute depth in these stimuli correctly, judging by their architecture.

Third, this is the first biologically-inspired model of da Vinci stereopsis (and possibly the first biologically-inspired model of stereopsis) tested on a large set of different image types. It performed adequately on most of them, making improvements over the initial maps provided by the disparity energy model and producing disparity maps close to observer's percepts. It also predicted quantitative depth percepts from monocular occlusions for all stimuli where quantitative depth has been previously found in psychophysical experiments.

Finally, to the best of my knowledge, this model provided the first instance of a biologically-plausible implementation of the right-to-left check. It also proposed a new way of estimating the reliability of the disparity signal in a biologically-plausible way and demonstrated the importance of this metric in the construction of disparity maps. Both these metrics can be utilized by future biologically-inspired models of stereopsis.

6.8.2 Limitations and future improvements

Testing of the model with an extensive battery of images revealed some of the limitations of the model. First, occlusion detection in textured areas results in occluded areas with jagged edges. This indicates the need for a smoothing operation at either one or several of the following steps: 1) the two occlusion detection methods, 2) combination of the occlusion estimates from the two methods, 3) occlusion width estimation and 4) after the selection of optimal disparities for each location. Occlu-

sion maps can be further improved if they are refined using the luminance edge maps, similar to the way the edge maps are applied to the computations of the 3D surface neurons.

Second, in the case of two-object arrangements, the occluded area is not restricted only to the occluded object (bar or disc) but in fact includes the space between the monocular object and the occluder. The current model does not identify the textureless space between the occluded object and the occluder as occluded since its goodness of match is quite high. One way to address this issue is to introduce a third method of occlusion detection that is based on the uniqueness constraint (Egnal and Wildes, 2002). In essence this method flags locations for which a double match is made. When the estimates from the three methods of occlusion detection are combined, the reliability measure (which is low for uniform areas) could be used as a weighting factor, favouring the uniqueness method response over the other two. Note, however, that it is not yet clear whether in two-object arrangements only the monocular object or the whole region between the occluder and the monocular object is signalled as occluded. Thus this drawback of the model might actually reflect a property of the visual system (see related discussion in Section 6.8.3). The uniqueness constraint could also be used during the final computation of 3D surfaces in a form of inhibitory interactions between neurons occupying the same retinal locations but tuned to different disparities. In the current model this constraint is not implemented and double-matching is not controlled for as illustrated by the case of the two-object arrangements with a bar.

Third, the method of final disparity selection described in Section 6.5.7 is not optimal since it masks the presence of several peaks with similar magnitude in the disparity profiles. Due to that, the disparity maps of the monocular gap stimulus and the two-object arrangement with the bar did not show gradual disparity blending but rather a sharp change in disparity. A more sophisticated method of disparity selection is required that will combine the peak disparities in a meaningful way taking into account the differences between the disparities in question and between the

corresponding response magnitudes.

Fourth, in the current version of the model, disparity from monocular regions propagates over binocular areas with low reliability until an edge or an area of high reliability is reached. However, this can result in long streaks of propagated disparity across fields of textureless background (such as in the image of the map). It will be useful to introduce a Gaussian type weighting system on the propagation of disparity from monocular areas such that its influence tapers with distance.

Finally, currently the depth maps are computed as if the scene is observed from the point of view of one of the physical eyes. This is the convention for most algorithms and models of stereopsis, however, it contrasts with the notion of the cyclopean eye (Erkelens et al., 1996; Ono et al., 2002). In the future this issue needs to be considered taking into account the literature exploring the perceived visual direction near depth discontinuities (Erkelens et al., 1996; Ono et al., 2003, 2002).

6.8.3 Model predictions

Evaluation with the two-object arrangement stimulus showed that the depth in the monocular bar is a mix between depth computed by double-matching and that computed based on occlusion geometry. In the current model, the occlusion-based depth depends on the width of the monocular object: the greater the width is the greater the perceived depth will be. This suggests that larger bar widths will yield larger perceived depth even when the overall size of the monocular region remains the same. Testing this prediction will reveal whether the visual system computes the overall size of the monocular region, or just the size of the monocular object (also see Section 6.8.2).

The disparity in monocular areas is computed based on the monocular area width and surrounding binocular area disparities. The contribution of those binocular disparities is weighted by the reliability and the match goodness of the signals. Thus the signal that contributes most to the final disparity of the monocular region is the

one that has high reliability and high match goodness. Densely textured areas with non-repeating patterns tend to have such properties. Thus the model predicts that such areas could 'capture' monocular regions by propagating their disparity signal over these regions. For example, in the case of the bias experiment in Chapter 4, using a textured biasing surface should result in a larger biasing of the monocular object disparity.

In the model, disparity from monocular areas propagates into binocular areas when those have low reliability. Reliability is low when more than one match produces a high response from the population of energy neurons. This can occur in textureless areas such as the ones in several stimuli tested with the model. However, low reliability can also occur in areas with repeating texture (wallpaper patterns) since each element can be matched successfully to several others in the other eye and more than one global matching solution is available. Thus the model predicts that when a monocular region neighbours an area with a repeating pattern, the disparity computed for the monocular region can propagate into the binocular region. In fact, some evidence of such disparity propagation from monocular regions over wallpaper patterns was provided with an occlusion version of the Kaniza figure and the Gillam and Nakayama (1999) stimulus (Hakkinen and Nyman, 2001).

Another prediction made by the model is that a single object (e.g. a dot) presented to one eye only, while the other eye views a uniform field, would elicit a percept of depth through the creation of an illusory occluding edge. In fact, there is some evidence that stimuli of this type create qualitative depth percepts (Kaye, 1977; Wilcox et al., 2007), named monoptic depth, that depend on the stimulus position with respect to fixation. Although the model does not incorporate a fixation constraint, it could be introduced in a form of adjustment of relative disparities with respect to an assumed fixation position. The model also predicts that the perceived depth in such stimuli would increase with the increase in the size of the monocular objects. This aspect of monoptic depth remains to be explored.

Finally, the model makes a clear prediction about the perceived depth of the

occluded areas in self-occluding objects. Imagine a cylinder defined by random dots. Portions of this cylinder on the right side will be occluded from the left eye and vice versa. The model predicts that these self-occlusions will be perceived as fronto-parallel at the depth dictated by the width of the monocular areas. Although this seems like an unlikely percept, there is very little data on the perception of depth in self-occluding objects. This issue could be explored by using disparity probes to estimate perceived disparity in the monocularly occluded regions of volumetric objects. With sufficiently wide occluded regions, the perceived disparity could be estimated at both edges of the region to determine whether it is perceived as fronto-parallel as the model predicts.

Chapter 7

Conclusions

7.1 Overview and contributions

This thesis described psychophysical and computational studies of depth from monocular occlusions and binocular disparity. Occluded areas are abundant in natural scenes and play an important role in stereoscopic depth perception. However, due to the treatment of occlusions as noise by early researchers of stereopsis, this field of study has not seen much development until the last two decades. Consequently, many aspects of depth perception from occlusions are not well understood. The goal of this thesis was to study several such aspects in order to advance the current understanding of monocular occlusions and their neural underpinnings. In the thesis, I examined issues including the main role of monocular occlusions in stereopsis, their interaction with depth from disparity and the nature of the neural mechanisms underlying depth perception from monocular occlusions. I used a combination of psychophysical experiments and computational modelling in order to address these questions. The detailed findings for each research question were discussed within their respective chapters. This section highlights and summarizes the most important findings and contributions.

The primary role of monocular occlusions in stereopsis

Using the random-dot frame stimulus, which was specifically designed for this task, I showed that monocular occlusions play an important role in defining the shape and depth of occluding surfaces. The addition of a monocular texture strip to the stimulus caused a change in the perceived figure-ground relationships. The presence of the monocular region triggered a percept of an occluding surface, that was not perceived in the absence of the monocular texture. This was the first time the effect of monocular occlusions on the shape and depth of occluding surfaces was systematically examined, analyzed and unequivocally demonstrated. The particular configuration of the stimulus I designed also allowed me to study the role geometric constraints play in quantitative depth perception from occlusions. It has been proposed that geometric constraints govern the perception of depth in da Vinci stereopsis (Gillam and Nakayama, 1999; Nakayama and Shimojo, 1990). By manipulating the strength of these constraints in the random-frame stimuli, I confirmed that the visual system relies on occlusion geometry in da Vinci stereopsis. However, the experiments also revealed that when the constraints are weak, a binocular disparity signal can bias the perceived depth from occlusions in the direction unrestricted by geometry.

The interaction of monocular occlusions and disparity

In the next series of experiments, I studied the interactions between depth from disparity and monocular occlusions using a biasing effect exerted by binocular objects on depth from monocular occlusions. This effect was found in the random-dot frame stimulus and the purpose of the experiments in Chapter 3 was to determine whether this effect generalizes to other stimuli and to examine its limitations. These experiments demonstrated that the disparity bias is present in other illusory occluder stimuli (Gillam and Nakayama, 1999; Liu et al., 1994) and that it is limited by the constraints imposed by occlusion geometry. That is, the perceived location of illusory occluders in depth could be only biased in the direction unrestricted by the geometry. The bias was also minimized in a stimulus where weak disparity information was available and was completely eliminated in a stimulus with well defined disparity. Importantly, one

observer showed a very strong bias in the stimulus with weak disparity information (Liu et al., 1997), demonstrating that when the reliability of the disparity cue is low, the visual system could rely completely on da Vinci stereopsis. Taken together these findings reveal a complex system of interactions between depth from disparity and monocular occlusions, which is modulated by occlusion geometry and the relative strength of the two signals. Moreover, when considered collectively, the experiments of Chapter 2 and Chapter 3 emphasize the critical role occlusion geometry plays in da Vinci stereopsis. It is clear from these experiments that the visual system has specific knowledge of the geometric constraints on depth in occluded regions and this information is used to compute the depth maps of the visual environment.

Based on the results of Chapter 3, I proposed that the disparity bias can be used as a 'litmus' test for distinguishing the source of depth percepts in stimuli with monocular occlusions. This is an important contribution since until now there had been no unified and straightforward psychophysical method to ensure that the depth in stimuli designed to study occlusions, originates from monocular geometry, while virtually every new stimulus requires such verification.

Quantitative depth from monocular occlusions - one or many mechanisms?

In a set of psychophysical and computational experiments described in Chapter 4, I have examined the source of the depth percepts in two-object arrangements. It has been suggested that in these configurations quantitative depth relies on double-matching and not on occlusion geometry (Gillam et al., 2003). Since quantitative depth based on occlusion geometry was demonstrated in other types of configurations, this would imply that da Vinci stereopsis does not have a unified underlying mechanism, but instead is stimulus dependent. Using three different methods, including the disparity bias 'litmus test' developed in Chapter 3, I have shown that depth in two-object arrangements relies mostly on occlusion geometry. Experiment 4.1 showed that the absence of quantitative depth in the two-objects arrangement with a disc in Gillam et al. (2003) was likely the result of the larger range of occlusion widths

they used for this stimulus. The disparity bias 'litmus test', applied to the two-object arrangement stimulus, demonstrated that the depth of monocular features but not the binocular features could be biased in the direction unrestricted by the viewing geometry. Finally, a computational analysis of the response of model disparity detectors to images of two-object arrangements revealed that these detectors cannot signal depth reliably in these stimuli. All three experiments suggested that the depth percepts in two-object arrangements are unlikely to be disparity based. Considered together with previous literature, these findings indicate the da Vinci stereopsis relies on a single mechanism, which is likely to be based on occlusion geometry.

Double-matching and depth from monocular occlusions - a computational analysis

Using the computational technique developed in my study of two-object arrangements, I have conducted a computational analysis of the binocular matching information available in a large battery of other monocular occlusion stimuli. The goal was to completely rule out the hypothesis that simple disparity detectors, performing double-matching, could be the sole underlying mechanism of da Vinci stereopsis. I applied model disparity detectors to images of different stimuli where depth is believed to be based on monocular occlusions. The depth maps produced by the disparity detectors were highly dependant on parameters and did not correspond to observers' percepts in the vast majority of cases. This study provided the first extensive computational evidence that double-matching cannot account for da Vinci stereopsis.

A biologically plausible model of depth from monocular occlusions and disparity

The findings of Chapters 2 - 5 allowed me to consolidate the theory of da Vinci stereopsis by proposing a single, occlusion geometry based set of mechanisms for the computation of qualitative and quantitative depth from monocular occlusions. In Chapter 6 I described these mechanisms in a form of several key principles. First, monocular occlusions are explicitly identified by a set of dedicated neurons. Second,

depth in occluded regions and in surrounding binocular regions with low disparity reliability, is computed using occlusion geometry. That is, this depth is inferred using the width of the occluded region, eye-of-origin information and the disparity of the surrounding regions. Third, depth in binocular and monocular areas is computed concurrently and the two signals interact. For simplicity I will refer to this set of principles as the monocular occlusion geometry theory (MOG theory). Based on the MOG theory, I developed a biologically-plausible computational model and tested it on a large battery of images including natural scenes, RDS and many types of da Vinci stereopsis phenomena. The model produced results close to observer percepts for different types of monocular occlusion stimuli including both qualitative and quantitative depth percepts. These results demonstrate that the neural architecture proposed in MOG is a likely substrate of da Vinci stereopsis.

The MOG theory is one of the most important contributions of this thesis to the field of research of monocular occlusions. The previous literature on monocular occlusions tends to be stimulus-driven rather than theory-driven and is dispersed in terms of the direction and the extent of the theoretical discussions. The field lacks a clearly defined and well-founded theory of depth from monocular occlusions that could be easily tested using both psychophysical and physiological methods. MOG is one such theory. It is based on psychophysical, physiological and computational data and it makes clear predictions for the mechanisms involved in da Vinci stereopsis and their behavioural manifestations. Future studies can adopt this theory as a framework and test new hypotheses implied by the theory, or use it as basis for development of alternative theories.

Finally, it is also important to highlight the novelty of the general methodological approach of this thesis. The vast majority of previous work has concentrated on either the behavioural or the computational aspects of depth from monocular occlusions. Computational algorithms have considered only few of the psychophysical findings and most psychophysical studies have not proposed any modelling based on their data. The combined approach, adopted in this thesis, leads to more comprehensive

and clear results since the two methods inform and complete each other. Moreover, when developing a theory or a model of da Vinci stereopsis, it is important to consider many different types of occlusion phenomena in order to achieve generalizable results as has been demonstrated here.

7.2 Outstanding issues and implications

The physiology of da Vinci stereopsis

The body of work described in this thesis opens many new avenues for further research. Perhaps the most intriguing direction lies in the domain of physiology. Virtually no work has been done to explore the neural underpinnings of depth from monocular occlusions using physiological techniques such as single-cell recording. The theoretical principles and the computational model described in Chapter 6 make very specific predictions regarding the type of neurons that could support depth percepts from monocular occlusions. Several simple tasks could be devised to probe for occlusion sensitive cells in the striate and the extrastriate cortices (the latter are a more likely location for the neural architecture proposed in Chapter 6). For example, monocular areas in wedding-cake RDS can be used to locate neurons sensitive to occlusions. However, care has to be taken to distinguish between cells sensitive to uncorrelated stimuli (Poggio et al., 1988) (which are most likely involved in the detection of false matches) and those responding specifically to monocular occlusions. Brain imaging techniques such as fMRI could be used as a first step to identify the brain structures most likely to contain occlusion detectors.

The role of scale in depth from occlusions and disparity

Several questions raised by the studies described here can be addressed using psychophysical methods. Data from several experiments (see Chapters 3 and 4) have suggested that the scale of the stimulus components might play a role in depth from

monocular occlusions. For instance, there was a difference in quantitative depth percepts between the bar and the line stimuli in experiments of Chapter 4. Moreover, the size of the binocular inducer seemed to correlate with the strength of the biasing effect as was shown in Chapter 3. The role of relative size can be studied systematically by varying the sizes of the occluders and the occluded features.

The experiments described in Chapter 3, also demonstrated the effect of size on depth from binocular disparity. In the vast majority of psychophysical studies, quantitative depth perception from stereopsis was measured using a disparity probe. In Chapter 3, I used a translational method with a virtual ruler and showed that depth perceived from disparity in very small single objects might be underestimated in comparison to larger objects. This is a very intriguing finding and could be studied in more detail by manipulating the size, shape and configuration of objects and employing several methods of depth estimation.

The effects of the depth signal strength

The biasing of depth from monocular occlusions by disparity signals suggests that other weak depth signals could be subject to biasing by nearby stronger signals. For example, in the case of many false matches, a depth percept of a surface might be susceptible to biasing from surrounding surfaces. This can be studied by gradually decreasing the correlation in the central patch of a RDS and examining the effect of neighbouring correlated regions on its perceived disparity. Experiments of this type could reveal more details regarding the nature of connections between neighbouring disparity detectors.

Finally, the relative weakness of the depth signal generated by monocular occlusions implies that the propagation (or interpolation) of this signal might not be as efficient as that of a disparity signal in terms of its speed and extent. This aspect can be studied by comparing the speed and the extent of depth propagation from monocular occlusions and binocular disparity using two different variants of the same stimulus (e.g. the stimulus of Gillam and Nakayama (1999)). In one variant of the stimulus, the source of depth will be binocular disparity and in the other, depth will

be perceived on the basis of monocular occlusions. Exposure duration can then be varied in a task that would require to estimate depth in the central region of the stimuli. Exploring the difference between the propagation of occlusion and disparity based depth signals, will reveal the relative strength/extent of the inter-neural connections of the population of occlusion detectors and disparity detectors.

Temporal occlusions and viewing direction

As it stands now, the theory of da Vinci stereopsis proposed in this thesis does not encompass two aspects of depth from occlusions described in detail in Chapter 1. First, the MOG theory does not propose a solution to the conflict in the visual direction of the occluded and the occluding surfaces as viewed from the cyclopean eye (Erkelens et al., 1996; Ono et al., 2003; van Ee and Erkelens, 2000). This is an unresolved issue in the field of monocular occlusions, and more work is required to understand how the visual field is partitioned or distorted to accommodate occluded and binocular areas. Second, the MOG theory does not attempt to account for the phenomenon of depth from temporal monocular occlusions (Brooks and Gillam, 2006; Shimojo et al., 1988). Including this phenomenon would most likely require the addition of temporal dynamics to the system.

Self-occluding objects

Most objects in the world are volumetric and so parts of these objects will occlude other parts under certain viewing conditions. These self-occlusions received virtually no attention in the literature despite their obvious importance in natural scenes (Wilcox and Lakra, 2007). The computational model proposed in Chapter 6 predicts that self-occluded areas would be assigned a single depth based on their width making them fronto-parallel (see Section 6.8.3). Due to the complete lack of data on the perception of quantitative depth in self-occluded areas, it is difficult to evaluate the model's prediction. To study this issue, RDS of vertically oriented cylinders could be used with monocularly occluded areas on the left and the right sides of the cylinders. Perceived magnitude of depth in the occluded regions could be studied

by aligning disparity probes with the left and right edges of each occluded area to assess its curvature. In addition, an increase in the degree of the perceived convexity of the cylinder with the increase in occlusion width, could also indicate that the occluded regions form a curved surface in depth.

7.3 Conclusions

This thesis offers a new, integrative theory of depth from monocular occlusions and disparity backed by careful psychophysical experiments and extensive computational simulations. This theory makes clear predictions about the neural underpinnings of depth from monocular occlusions and the behavioural manifestations of this neural architecture.

Appendix A

Symbols and acronyms used in the thesis

Table A.1: Acronyms used in the thesis.

| Acronym | Description |
|---------|------------------------------|
| RDS | Random Dot Stereogram |
| RF | Receptive Field |
| JND | Just Noticeable Difference |
| PSE | Point of Subjective Equality |
| 2IFC | Two Interval Forced Choice |
| SSD | Sum of Squared Differences |
| MOG | Monocular Occlusion Geometry |

Table A.2: Symbols used in Chapter 6.

| Symbol | Description |
|-----------------|--|
| f_L and f_R | The left and the right RFs of a simple energy neuron. |
| I_L and I_R | The left and the right image patches that fall on these RFs. |

Table A.2 – continued from previous page

| Symbol | Description |
|--|---|
| $Gaus_L$ and $Gaus_R$ | The left and the right Gaussian functions of the simple cell RFs. |
| Sin_L and Sin_R | The left and the right sinusoids of the simple cell RFs. |
| d_L and d_R | The left and the right positional shifts of the simple cell RFs. |
| ϕ_L and ϕ_R | The left and the right positional shifts of the simple cell RFs. |
| σ_x and σ_y | The horizontal and vertical widths of $Gauss_{L/R}$. |
| θ | The preferred orientation of the simple cell RF. |
| ω_0 | The preferred spatial frequency of the simple cell RF. |
| C_0 | Classical complex energy cell |
| C | Complex energy cell with a normalized response |
| $C_{x,y,d}$ | Response of a population of C cells with RFs at x, y and tuned to disparity d pooled over scales and orientations. |
| $[-d_m, d_m]$ | The range of disparities the complex energy cells C are tuned to. |
| $C_{x,y}^*$ | The response $C_{x,y,d}$ for all disparities $[-d_m, d_m]$. Referred to throughout as the disparity profile |
| $C_{x,y}^{L*}$ and $C_{x,y}^{R*}$ | The response of a population of complex cells $C_{x,y}^*$ with all the left-eye / right-eye receptive fields fixed at location x, y . |
| $n(x)$ | A normalization function. |
| $R_{x,y,d}$ | The cells computing the difference between disparity profiles $C_{x,y}^{L*}$ and $C_{x+d,y}^{R*}$. |
| $LRC_{x,y}$ | The cells computing the left-to-right check. |
| M^* | A cell computing the maximal response of the whole population of disparity detectors. |
| $\langle x \rangle_{\theta_n} / \langle x \rangle_0$ | Rectification with respect to the threshold θ_n or 0 respectively. |
| $MG_{x,y}$ | The cells computing the match goodness metric for occlusion detection. |
| $OCC_{x,y}$ | Monocular occlusion detectors. |
| $ES_{x,y,w,s}$ | End-stopped cells, receiving input from $OCC_{x,y}$ cells, with an excitatory centre of width w which is shifted by s with respect to location x, y |
| $W_{x,y,w}$ | Cells computing the likelihood that x, y is located within a monocularly occluded region of size w |
| $W_{x,y}^*$ | The response of a population of $W_{x,y,w}$ cells tuned to different widths. Referred to as the occlusion width profile . |
| γ_n | The weight of inhibitory inter-neural connections. |

Table A.2 – continued from previous page

| Symbol | Description |
|------------------------|--|
| $C_{x,y}^{*-d}$ | The disparity profile at location x, y with the response to disparity d zeroed. |
| $RE_{x,y,d}$ | Cells computing the difference between the maximum response in $C_{x,y}^*$ and the maximum response in $C_{x,y}^{*-d}$. |
| $RE_{x,y}^*$ | The response of a population of $RE_{x,y,d}$ cells with different disparities zeroed. |
| $REL_{x,y}$ | Cells computing the reliability of disparity estimates at location x, y |
| $OMG_{x,y}$ | Cells computing the overall match goodness. |
| $DE_{x,y}$ | Disparity edges detectors. |
| $LE_{x,y}$ | Luminance edges detectors. |
| $ED_{x,y}$ | Combined edges detectors. |
| $OBJ_{x,y,x',y'}$ | Cells signalling whether x, y and x', y' belong to the same object. |
| $BIN_{x,y,d}$ | Cells computing final disparities for binocular locations. |
| $MON_{x,y,d}$ | Cells computing final disparities for monocularly occluded locations. |
| $3D_{x,y,d}$ | Cells computing final disparities by combining $BIN_{x,y,d}$ and $MON_{x,y,d}$ responses. |
| $BS_{x,y,d}$ | Cells aggregating support from binocular regions around monocularly occluded locations. |
| $MS_{x,y,d}$ | Cells aggregating support from monocular regions around monocularly occluded locations. |
| NL, NA, NB, NR | Regions to the left, above, below and to the right of a monocularly occluded region from where support is aggregated. |
| $H \times W$ | The height and the width of the support regions NL, NA, NB, NR . |
| $\overline{3D}_{NX,d}$ | Mean disparity profile of the support region NX . |
| \overline{RE}_{NX} | Mean reliability of the support region NX . |
| $Gauss_{x,y}$ | A Gaussian function centered on x, y . |
| $PROP_{x+s,y,d}$ | Cells propagating disparity signals from monocularly occluded locations x, y to binocular locations $x + s, y$. |

Appendix B

Auxiliary data

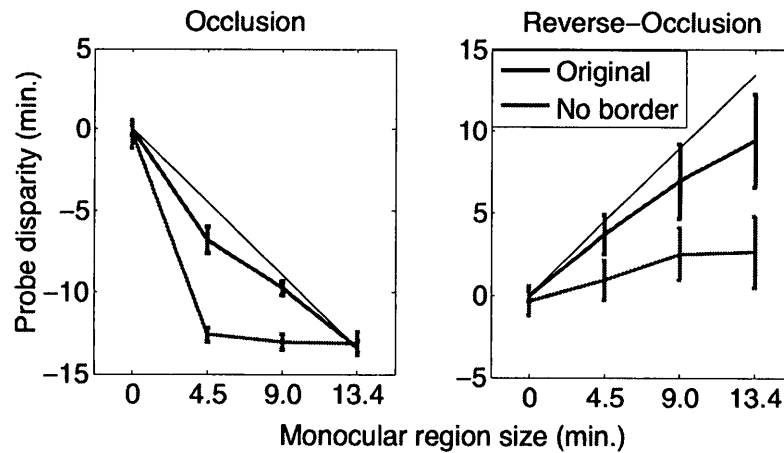


Figure B.1: Results of the disparity-matching task of Experiment 2.3 for all observers including SL. The ordinate shows the disparity of the probe and the abscissa shows the size of the monocular region. The blue line shows the depth estimates with the stimuli used in Experiment 2.2 (Original) and the green line shows the depth estimates with the modified stimuli with the right-hand border removed (No border). Negative disparity values are assigned to crossed depth. The error bars show ± 1 standard error of the mean. The thin black line indicates the predicted depth.

Bibliography

- Allenmark, F. and Read, J. (2011). Spatial stereoresolution for depth corrugations may be set in primary visual cortex. *BMC Neuroscience*, 12(Suppl 1):P263. 6.5.1
- Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature*, 367:365–368. 1.2, 1.4, 1.8, 4.5.1
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, 26(4):419–453. 4.5.2
- Anderson, B. L. and Julesz, B. (1995). A theoretical-analysis of illusory contour formation in stereopsis. *Psychological Review*, 102(4):705–743. 4.5.2
- Anderson, B. L. and Nakayama, K. (1994). Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychological Review*, 101(3):414–445. 1.4
- Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J.-M., Bullier, J., and Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, 22(19):8633–8646. 6.2, 6.5.5
- Anzai, A., Ohzawa, I., and Freeman, R. D. (1997). Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proceedings of the National Academy of Sciences of the United States of America*, 94(10):5438–5443. 6.5.1
- Anzai, A., Ohzawa, I., and Freeman, R. D. (1999). Neural mechanisms for encoding binocular disparity: receptive field position vs. phase. *Journal of Neurophysiology*, 82(2):874–890. 6.5.1
- Assee, A. and Qian, N. (2007). Solving da Vinci stereopsis with depth-edge-selective V2 cells. *Vision Research*, 47:2585–2602. 1.7, 4.5.2, 6.1, 6.1
- Banks, M. S., Gepshtein, S., and Landy, M. S. (2004). Why is spatial stereoresolution so low? *The Journal of Neuroscience*, 24(9):2077–2089. 4.4, 6.5.1

- Belhumeur, P. N. and Mumford, D. (1992). A bayesian treatment of the stereo correspondence problem using half-occluded regions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 506–512. IEEE. 6.1
- Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, 10:433–436. 3.3.1
- Brooks, K. R. and Gillam, B. (2006). Quantitative perceived depth from sequential monocular decamouflage. *Vision Research*, 46:605–613. 1.3.3, 1.8, 4.5.1, 7.2
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36:96–107. 6.2
- Cao, Y. and Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18:515–578. 1.7, 6.1
- Chen, Y. and Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural computation*, 16(8):1545–1577. 1.7, 6.5.1
- Cook, M. and Gillam, B. (2004). Depth of monocular elements in a binocular scene: The conditions for da Vinci stereopsis. *Journal of Experimental Psychology*, 30:92–103. 1.2, 1.4, 1.3.2, 2.4, 4.5.2, 5.1, 8, 5.10, 6.7
- Cumming, B. G. and DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience*, 24:203–238. 4.4, 6.2
- Da Vinci, L. (1877). *A treatise on painting (English translation)*. George Bell and Sons. 1.1
- DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1995). Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity? *Perception*, 24(1):3–31. 4.4, 6.5.1, 6.5.1, 6.5.1
- Egnal, G., Mintz, M., and Wildes, R. P. (2004). A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image and Vision Computing*, 22(12):943–957. 4.4.1, 6.5.4
- Egnal, G. and Wildes, R. (2002). Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133. 1.7, 4.4.1, 4.4.2, 6.1, 6.5.2, 6.8.2
- Ehrenstein, W. and Gillam, B. (1998). Early demonstrations of subjective contours, amodal completion, and depth from half-occlusions: "Stereoscopic experiments with silhouettes" by Adolf von Szily (1921). *Perception*, 27:1407–1416. 1.1, 1.2

- Enright, J. T. (1991). Exploring the third dimension with eye movements: Better than stereopsis. *Vision Research*, 31(9):1549-1562. 4.2.3, 4.5.1
- Erkelens, C., Muijs, A., and van Ee, R. (1996). Binocular alignment in different depth planes. *Vision Research*, 36(14):2141-2147. 1.6, 1.6, 6.8.2, 7.2
- Fleet, D. J., Wagner, H., and Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research*, 36(12):1839-1857. 6.5.1, 6.5.1
- Forte, J., Peirce, J. W., and Lennie, P. (2002). Binocular integration of partially occluded surfaces. *Vision Research*, 42:1225-1235. 1.2, 1.5
- Gilbert, C. D. and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience*, 3(5):1116-33. 6.2
- Gillam, B. (1995). Matching needed for stereopsis. *Nature*, 373:202. 1.3.2, 1.8, 3.4, 3.6
- Gillam, B., Blackburn, S., and Nakayama, K. (1999). Stereopsis based on monocular gaps: metrical encoding of depth and slant without matching contours. *Vision Research*, 39:493-502. 1.4, 1.3.4, 1.5, 5.1, 7, 5.10, 6.7, 6.7
- Gillam, B. and Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception*, 17:603-608. 1.5, 1.8, 6.3
- Gillam, B., Cook, M., and Blackburn, S. (2003). Monocular discs in the occlusion zones of binocular surfaces do not have quantitative depth: a comparison with Panum's limiting case. *Perception*, 32:1009-1019. (document), 1.3.1, 1.8, 2.4, 3.6, 4.1, 4.2, 4.2, 4.2.1, 4.2.3, 4.5.1, 4.5.2, 4.5.2, 5.10, 7.1
- Gillam, B. and Grove, P. (2004). Slant or occlusion: global factors resolve stereoscopic ambiguity in sets of horizontal lines. *Vision Research*, 44:2359-2366. 1.2, 1.3.2
- Gillam, B. and Nakayama, K. (1999). Quantitative depth for a phantom surface can be based on cyclopean occlusion cues alone. *Vision Research*, 39:109-112. (document), 1.2, 1.2, 1.4, 1.3.2, 1.4, 1.5, 3.3, 3.3.1, 3.3.1, 3.3.2, 3.6, 4.5.1, 3, 5.1, 6, 5.10, 6.3, 6.5.2, 6.2, 6.7, 6.8.3, 7.1, 7.2
- Goutcher, R. and Mamassian, P. (2005). Selective biasing of stereo correspondence in an ambiguous stereogram. *Vision Research*, 45(4):469-483. 4.2.3, 6.5.7

- Grossberg, S. and Howe, P. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research*, 43:801–829. 6.1
- Grove, P., Byrne, J., and Gillam, B. (2005). How configurations of binocular disparity determine whether stereoscopic slant or stereoscopic occlusion is seen. *Perception*, 34:1083–1094. 1.3.2
- Grove, P. and Gillam, B. (2007). Global patterns of binocular image differences resolve the ambiguity between stereoscopic slant and stereoscopic occlusion. *Vision*, 19(1):1–13. 1.3.2, 4.5.1, 6.3
- Grove, P., Gillam, B., and Ono, H. (2002). Content and context of monocular regions determine perceived depth in random dot, unpaired background and phantom stereograms. *Vision Research*, 42:1859–1870. 3.6
- Grove, P. and Ono, H. (1999). Ecologically invalid monocular texture leads to longer perceptual latencies in random-dot stereograms. *Perception*, 28:627–639. 1.5
- Hakkinen, J. and Nyman, G. (1996). Depth asymmetry in daVinci stereopsis. *Vision Research*, 36(23):3815–3819. 1.3.1, 1.5, 1.4, 1.8, 3.1, 4.2, 4.2.1, 4.3, 4.5.1, 6.3
- Hakkinen, J. and Nyman, G. (1997). Occlusion constraints and stereoscopic slant. *Perception*, 26:29–38. 1.2, 1.5, 1.4
- Hakkinen, J. and Nyman, G. (2001). Phantom surface captures stereopsis. *Vision Research*, 41:187–199. 1.5, 1.4, 1.8, 3.6, 6.3, 6.8.3
- Harris, J. M. and Smith, D. (2010). Biases and thresholds for depth perception from monocular regions of binocular scenes. In *Proceedings of the annual meeting of the Vision Sciences Society*. 5.3
- Hayashi, R., Maeda, T., Shimojo, S., and Tachi, S. (2004). An integrative model of binocular vision: a stereo model utilizing interocularly unpaired points produces both depth and binocular rivalry. *Vision Research*, 44:2367–2380. 1.7, 6.1
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(181-197):181–197. 6.2, 6.5.1
- Howard, I. P. and Rogers, B. (2002). Linking images. *Seeing in depth*. Oxford Press, pages 47–48. 1.8, 4.5.1

- Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160:106–154. 6.2
- Hubel, D. H. and Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28(2):229–289. 6.2
- Hupe, J., James, A., Payne, B., Lomber, S., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695):784–787. 6.2, 6.5.5
- Jodoin, P., Rosenberger, C., and Mignotte, M. (2006). Detecting half-occlusion with a fast region-based fusion procedure. In *British Machine Vision Conference*, volume 1, pages 417–426. 6.1
- Julesz, B. (1960). Binocular depth perception of computer generated patterns. *Bell Systems Technical Journal*, 39:1125–1162. 1.1
- Kaufman, L. (1965). Some new stereoscopic phenomena and their implication for the theory of stereopsis. *The American Journal of Psychology*, 78(1):1–20. 1.2, 1.8
- Kaye, M. (1977). Stereopsis without binocular correlation. *Vision Research*, 18:1013–1022. 6.8.3
- Kuroki, D. and Nakamizo, S. (2006). Depth scaling in phantom and monocular gap stereograms using absolute distance information. *Vision Research*, 46:4206–4216. 1.5, 6.3
- Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of Neuroscience*, 15(2):1605–1615. 6.2
- Lawson, R. and Gulick, W. (1967). Stereopsis and anomalous contour. *Vision Research*, 7:271–297. 1.2, 1.8
- Lin, M. H. and Tomasi, C. (2004). Surfaces with occlusions from layered stereo. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 26(8):1073–1078. 1.7, 6.1, 6.3
- Little, J. J. and Gillett, W. E. (1990). Direct evidence for occlusion in stereo and motion. *Image and Vision Computing*, 8(4):328–340. 4.4.1, 6.5.4
- Liu, L., Stevenson, S. B., and Schor, C. M. (1994). Quantitative stereoscopic depth without binocular correspondance. *Nature*, 367:66–68. (document), 1.4, 1.3.2, 1.4, 3.4, 3.4.1, 3.4.1, 3.4.2, 3.4.2, 3.6, 5.1, 4, 5.10, 6.5.2, 6.2, 6.7, 7.1

- Liu, L., Stevenson, S. B., and Schor, C. M. (1995). Matching needed for stereopsis. *Nature*, 373:203. 1.3.2, 3.4
- Liu, L., Stevenson, S. B., and Schor, C. M. (1997). Binocular matching of dissimilar features in phantom stereopsis. *Vision Research*, 37(5):633–644. 1.3.2, 2.4, 3.4, 3.6, 7.1
- Marr, D. and Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194(4262):283–287. 1.1, 1.8, 6.5.2
- Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological sciences*, 204(1156):301–328. 1.1, 1.8
- Min, D. and Sohn, K. (2008). Cost aggregation and occlusion handling with WLS in stereo matching. *IEEE Transactions on Image Processing*, 17(8):1431–1442. 1.7, 6.1, 6.3
- Min, D., Yea, S., and Vetro, A. (2010). Occlusion handling based on support and decision. In *IEEE International Conference on Image Processing*, pages 1777–1780. 6.1
- Mitsudo, H., Nakamizo, S., and Ono, H. (2005). Greater depth seen with phantom stereopsis is coded at the early stages of visual processing. *Vision Research*, 45:1365–1374. 1.5, 6.3
- Mitsudo, H., Nakamizo, S., and Ono, H. (2006). A long-distance stereoscopic detector for partially occluding surfaces. *Vision Research*, 46:1180–1186. 1.3.2
- Nakayama, K. (1996). Binocular visual surface perception. *Proceedings of the National Academy of Sciences of the United States of America*, 93:634–639. 1.5
- Nakayama, K. and Shimojo, S. (1990). Da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research*, 30(11):1811–1825. (document), 1.2, 1.2, 1.4, 1.3.1, 1.8, 3.1, 4.1, 4.1, 4.2, 4.2, 4.2.1, 4.2.1, 4.2.3, 4.3, 4.6, 4.4.2, 4.5.1, 4.5.2, 4.9, 4.5.2, 5.1, 5, 5.10, 6.7, 7.1
- Ohzawa, I., DeAngelis, G. C., and Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, 249:1037–1041. 6.2, 6.5.1, 6.5.1, 6.5.1
- Ono, H., Lillakas, L., Grove, P., and Suzuki, M. (2003). Leonardo’s constraint: Two opaque objects cannot be seen in the same direction. *Journal of Experimental Psychology*, 132(2):253–265. 1.6, 6.8.2, 7.2

- Ono, H., Wade, N., and Lillakas, L. (2002). The pursuit of Leonardo's constraint. *Perception*, 31:83–102. 1.6, 6.8.2
- Pianta, M. and Gillam, B. (2003a). Monocular gap stereopsis: manipulation of the outer edge disparity and the shape of the gap. *Vision Research*, 43:1937–1950. 1.3.4, 6.3
- Pianta, M. and Gillam, B. (2003b). Paired and unpaired features can be equally effective in human depth perception. *Vision Research*, 43:1–6. 1.2, 1.3.4, 1.8, 3.6, 4.5.1
- Poggio, T., Gonzalez, F., and Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *Journal of Neuroscience*, 8(12):4531–4550. 7.2
- Prince, S. J., Cumming, B. G., and Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque v1. *Journal of Neurophysiology*, 87(1):209–221. 6.5.1
- Read, J. and Cumming, B. G. (2003). Testing quantitative models of binocular disparity selectivity in primary visual cortex. *Journal of Neurophysiology*, 90:2795–2817. 6.2
- Read, J. C. (2010). Vertical binocular disparity is encoded implicitly within a model neuronal population tuned to horizontal disparity and orientation. *PLoS Computational Biology*, 6(4):e1000754. 6.5.1, 6.5.1
- Read, J. C. and Cumming, B. G. (2006). Does depth perception require vertical-disparity detectors? *Journal of Vision*, 6(12). 6.5.1, 6.5.1
- Read, J. C. and Cumming, B. G. (2007). Sensors for impossible stimuli may solve the stereo correspondence problem. *Nature Neuroscience*, 10(10):1322–1328. 6.5.1
- Reynolds, J. H. and Heeger, D. J. (2009). The normalization model of attention. *Neuron*, 61(2):168. 6.5.2
- Sachtler, W. L. and Gillam, B. (2007). The stereoscopic sliver: a comparison of duration thresholds for fully stereoscopic and unmatched versions. *Perception*, 36:135–144. 1.2, 1.5
- Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42. 4.4.1, 6.6
- Schoenfeld, M., Woldorff, M., Dzel, E., Scheich, H., Heinze, H.-J., and Mangun, G. (2003). Form-from-motion: MEG evidence for time course and processing sequence. *Journal of Cognitive Neuroscience*, 15(2):157–172. 6.2, 6.5.5

- Shimojo, S. and Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Research*, 30(1):69–80. 1.5
- Shimojo, S., Silverman, G., and Nakayama, K. (1988). An occlusion-related mechanism of depth perception based on motion and interocular sequence. *Nature*, 333:265–268. 1.3.3, 4.5.1, 7.2
- Sizintsev, M. and Wildes, R. (2007). Computational analysis of binocular half-occlusions. *Computational Vision in Neural and Machine Systems*. Cambridge University Press. 1.7, 6.1, 6.3
- Sizintsev, M. and Wildes, R. P. (2010). Coarse-to-fine stereo vision with accurate 3D boundaries. *Image and Vision Computing*, 28(3):352–366. 6.1
- Sun, J., Li, Y., Kang, S., and Shum, H. (2005). Symmetric stereo matching for occlusion handling. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 399–406. 1.7, 6.1, 6.5.2
- van Ee, R. and Erkelens, C. (2000). Is there an interaction between perceived direction and perceived aspect ratio in stereoscopic vision? *Perception and Psychophysics*, 62(5):910–926. 1.6, 7.2
- von der Heydt, R., Zhou, H., and Friedman, H. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, 40:1955–1967. 1.7
- Wang, L., Jin, H., Yang, R., and Gong, M. (2008). Stereoscopic inpainting: Joint color and depth completion from stereo images. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 6.1
- Watanabe, O. and Fukushima, K. (1999). Stereo algorithm that extracts a depth cue from interocularly unpaired points. *Neural Networks*, 12:569–578. 1.7, 6.1
- Wilcox, L. M. and Allison, R. S. (2009). Coarse-fine dichotomies in human stereopsis. *Vision Research*, 49(22):2653–2665. 4.5.1
- Wilcox, L. M., Harris, J. M., and McKee, S. P. (2007). The role of binocular stereopsis in monoptic depth perception. *Vision Research*, 47(18):2367–2377. 6.8.3
- Wilcox, L. M. and Lakra, D. C. (2007). Depth from binocular half-occlusions in stereoscopic images of natural scenes. *Perception*, 36:830–839. 1.5, 7.2
- Yang, Q., Bucci, M., and Kapoula, Z. (2002). The latency of saccades, vergence, and combined eye movements in children and in adults. *Investigative Ophthalmology and Visual Science*, 43(9):2939–2949. 4.2.3

- Yang, Q., Wang, L., Yang, R., Stewenius, H., and Nister, D. (2009). Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):492–504. 6.1
- Yu, A. J., Giese, M. A., and Poggio, T. A. (2002). Biophysiologicaly plausible implementations of the maximum operation. *Neural Computation*, 14:2857–2881. 6.2
- Zitnick, C. and Kanade, T. (2000). A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):675–684. 1.7, 6.1, 6.5.2