Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yunfeng Li

Entitled Computational Models of 3D Shape Perception

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Zygmunt Pizlo
Chair

Richard Schweickert

Gregory Francis

Avinash C. Kak

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Zygmunt Pizlo

Approved by: James M. LeBreton 12/01/09
Head of the Graduate Program Date

Graduate School Form 20
(Revised 10/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation:

Computational Models of 3D Shape Perception

For the degree of Doctor of Philosophy

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Executive Memorandum No. C-22,* September 6, 1991, *Policy on Integrity in Research.**

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

Yunfeng Li
Signature of Candidate

11/30/09
Date

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/c_22.html

COMPUTATIONAL MODELS OF 3D SHAPE PERCEPTION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yunfeng Li

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2009

Purdue University

West Lafayette, Indiana

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor and committee chair, Dr. Zygmunt Pizlo, for his understanding, caring, guidance, and support during these years of my studying at Purdue. His mentorship was paramount in providing me the foundation for conducting mathematic psychology research and shaping my long-term career goals. My work origins from his idea about the role of simplicity constraints, especially maximum compactness, in human 3D shape perception. During the process of exploring, expanding and finally formulating this theory, Dr. Pizlo provided me brilliant intellectual insights as well as generous financial and equipment support. During the dissertation writing, he also spent a lot of time helping with my revision. Without his guidance and persistent help this dissertation would not have been possible.

I would like to thank Dr. Gregory S. Francis, Dr. Richard Schweickert, and Dr. Avinash C. Kak for their valuable input and comments on my work. I am particularly indebted to Dr. Gregory S. Francis, for his detailed edits on my dissertation.

I would also like to thank all of the members of the Pizlo research lab for their support and friendship. Especially I wanted to thank Dr. Tadamasa Sawada for his tech assistance for my defense presentation.

I am also very grateful to Ms. Julie Smith of the psychology department for her expertise and help on finalizing and formatting my dissertation.

Finally, I would like to thank my wife, Shen, and my daughter, Sylvia, for their love and support.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Appendix Figure

Appendix Figure Page

LIST OF SYMBOLS

Ξ: The set of all symmetric correspondences in an image.

': The symmetric corresponding point.

<a, a'>: a pair of symmetric points.

η: a 3D shape.

I: an 2D image

τ: The tilt of symmetry plane of a 3D shape.

π_{XY}: The image plane.

π_S: The symmetry plane of a 3D shape.

l_S: The intersection between an image plane and a symmetry plane.

α: The angle between an image plane and a symmetry plane.

Ψ: The set of all symmetric 3D interpretations of an image.

$V(\eta)$: The volume of a 3D shape η.

$S(\eta)$: The surface area of a 3D shape η.

$C(\eta)$: The compactness of a 3D shape η.

$H(\eta)$: The convex hull of a 3D shape η.

δ: Binocular disparity.

$O(A_i,A_j)$: The depth order between A_i and A_j.

M_η: The depth order matrix for η.

$\approx$: The equality between two depth order matrices.

Θ_η: The set of all symmetric 3D interpretations that have the same depth order matrix as that of η.

Q: The affine transformation from one 3D shape from the other.

e_m: The change coefficient from one 3D shape to the other along the direction m.

e_n: The change coefficient from one 3D shape to the other along the direction n.

k_m: The change magnitude from one 3D shape to the other along the direction m.

k_n: The change magnitude from one 3D shape to the other along the direction n.

$\varpi(\eta_1,\eta_2)$: The dissimilarity between two 3D shapes η_1 and η_2.

ε: Perceptual error.

ABSTRACT

Li, Yunfeng. Ph.D., Purdue University, December 2009. Computational Models of 3D Shape Perception. Major Professor: Zygmunt Pizlo.

In this study, two computational models were formulated to simulate human monocular and binocular 3D shape perception. In the monocular model, simplicity constraints (symmetry, planarity, maximum compactness and minimum surface area) were used to recover a 3D shape from its single image. In the binocular model, the ordinal depth of points in a 3D shape provided by stereoacuity was combined with the simplicity constraints to recover a 3D shape.

In two psychophysical experiments, human monocular and binocular 3D shape recovery was measured. The comparison between subjects' performance and the performance of the models showed that they were very similar. Specifically, monocular performance of both the subjects and the model was close to veridical for slants of the symmetry plane in the range between 30 and 60 deg. When slants were close to 0 deg or 90 deg (degenerate views), monocular performance deteriorated, but the type and the magnitude of errors were very similar in the subjects and the model. Binocular performance, on the other hand, was close to veridical for almost the entire range of slant of the symmetry plane. This is the first empirical study demonstrating veridical 3D shape perception and the first computational model that performs as well, or even better than the subjects do.

INTRODUCTION

We can easily perceive the objects around us as three dimensional and the percept is usually veridical. However, understanding the underlying processes is not easy. To produce the 3D shape percept, our visual system has to recover the 3D shape from the 2D retinal image(s). In the past 30 years scientists made significant progress towards understanding the process of 3D shape recovery and several theories have been proposed to explain this phenomenon. For example, Gibson (1979) claimed that our surrounding environment provides sufficient information that is directly recorded by the visual system. Poggio & Edelman's multiple view theory (1990) emphasized the role of learning in 3D shape perception: having registered many views of a 3D object in our memory through multiple encounters with the object, we can recognize the 3D object when a specific view appears again. However, there has been no satisfactory computational model of the underlying mechanisms.

Currently, many approaches to 3D shape perception are inspired by Marr's theory(Marr, 1982), especially his concept of 2.5D representation. In the context of Marr's theory, 3D shape perception includes two stages - viewer-centered representation of visible surfaces and an object-centered representation of the object. In the first stage, the visual system obtains information about orientations of visible surfaces from a variety of depth cues and in the second stage it derives the 3D shape representation from the surfaces and from the 3D models stored in memory. This theory inspired research on 3D shape perception during the last 20 years of the last century, specifically on the role of depth cues (Regan, 2000), the relationships among and combination of different depth cues (Landy et, al 1995, Hillis et, al, 2004),

the perception of 3D surfaces (Koenderink et al., 2004), and on matching 3D shapes (Ullman, 1996). Note that Marr's theory is essentially atomistic and that the philosophy behind it is that once we know the orientations of surfaces at many points, we can recover the 3D shape represented by these surfaces. Although the researchers who follow this approach and study the role of the depth cues provided some interesting results, there are several problems that remain unsolved. These problems are briefly discussed next.

The Percept of a 3D Shape may be More Accurate Than the Percept of 3D Surfaces

There is no empirical evidence that the percept of the orientation of surfaces is necessary for the 3D shape perception. There are two other possible scenarios. First, the percept of the orientation of surfaces may follow, rather than precede the percept of 3D shape. In other words, the percept of 3D surfaces may be the result of the 3D shape percept. Second, the orientation percept and the shape percept may be independent. Psychophysical experiments supports both of these possibilities. For example, there is a large body of experimental evidence showing that observers perceptually underestimate slants of surfaces (Gibson, 1950; Braunstein, 1968). At the same time, however, their 3D shape percept is often veridical. If one is committed to Marr's paradigm, one is faced with a question of how to derive veridical 3D shape from non-veridical (biased) orientations of surfaces?

Li (2009) conducted an experiment testing the consistency between a subject's 3D shape percept and the percept of 3D surfaces. The subject was presented with an image of a parallelepiped (see Figure 1) and was asked to perform two types of judgments based on the 3D percept of the parallelepiped: (1) the 3D orientation (slant and tilt) of each of the three visible faces using an elliptical probe; (2) the 2D shape of each of the three visible faces. Based on these judgments, two 3D shapes were computed: one from 3D orientation and the other from 2D shape judgments. These two 3D shapes were significantly different. In the follow-up experiment, the same 2D image of a parallelepiped

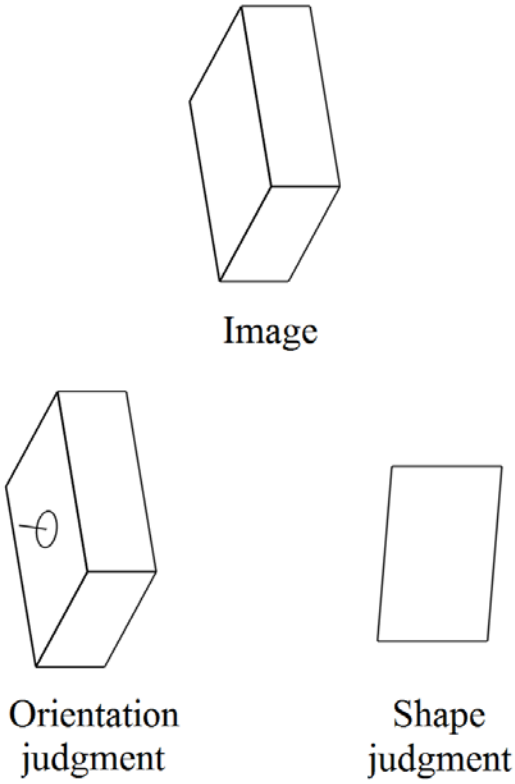


Figure 1. A stimulus in a single trial of the main experiment in Li (2009). The top is the image of a parallelepiped. The left bottom image illustrates the 3D orientation judgment and right bottom image illustrates the 2D shape judgment.

was presented at the top of a display, and the two computed 3D shapes were presented side by side at the bottom and they were rotating. The subject was asked to choose the 3D shape that was closer to their percept of the parallelepiped produced by the 2D image on top. All subjects strongly preferred the 3D shape computed from the 2D shape judgments, not the 3D shape computed from the 3D orientation judgments. These results strongly suggest that the 3D shape perception is not produced from the perception of orientations of 3D surfaces.

Depth Cues are not Necessary for 3D Shape Perception

In the case of images that do not have any surface depth cues, like line drawings, human observers can still perceive 3D shapes. Koenderink (1996) compared the subjects' 3D shape percept produced with two kinds of images – one was the image with depth cues (e.g. shading) and the other was just a line drawing represented by edges and contours. The subjects' performance with these two kinds of images was similar, which suggests that contours, not depth cues are important in 3D shape perception. Li (2005) extended these experiments using polyhedra with or without depth cues and found that once edges of the polyhedron were clearly visible, subjects could perceive the 3D shapes veridically. Depth cues (shading, texture) did not contribute to the 3D percept. Depth cues were only important when they served as edge cues, i.e., when they facilitated edge detection.

The Role of Simplicity Constraints

Li (2005) designed an experiment to test the role of symmetry and planarity constraints. In the experiment he tested subjects' shape discrimination using six kinds of 3D objects. Some objects satisfied both constraints: the objects were mirror-symmetric and the contours of their faces were planar (Figure 2a). Others, did not satisfy either (Figure 2 (c), (f)). Li found that when objects were symmetric and their faces were planar, the subject could easily tell whether two 3D shapes seen from different viewing orientations are same or different (see Figure 3). However, in the case of

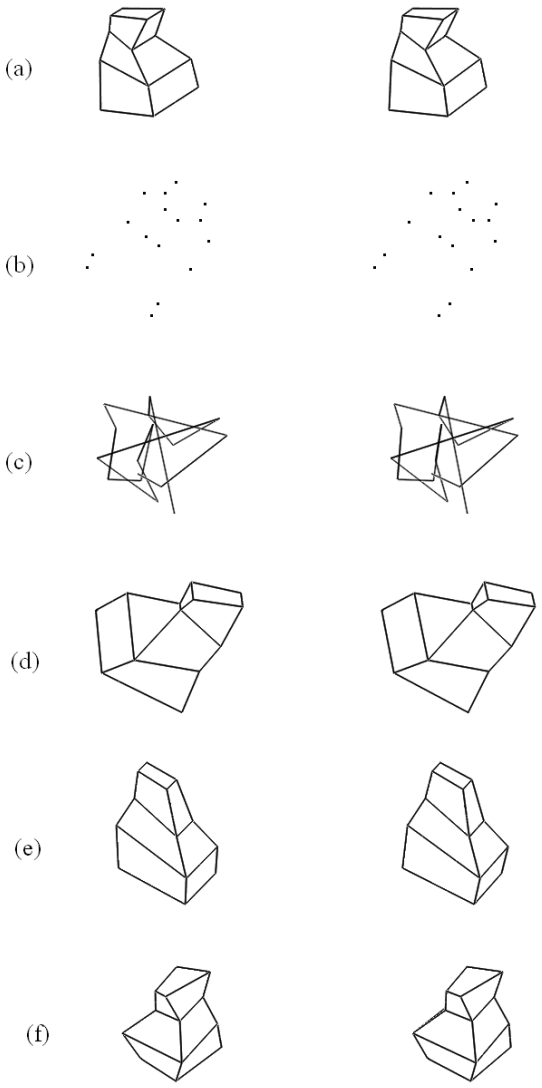


Figure 2. Stereoscopic pairs (for crossed fusion) of six types of stimuli used by Li (2005) (a) Polyhedron with one symmetry plane and planar surfaces. (b) 16 vertices which were obtained by removing the edges in stimuli (a). (c) Polygonal line. The 16 vertices were connected in a random order. (d) Partially non-planar, symmetric polyhedron. (e) Planar and asymmetric polyhedron. (f) Non-planar and asymmetric polyhedron.

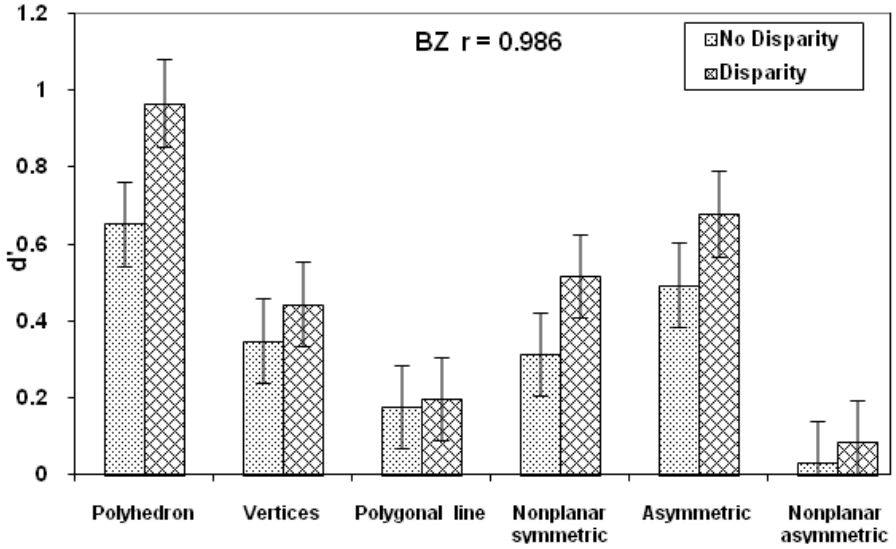


Figure 3. One naive subject's performance (d') for six types of objects. Higher d' represents higher performance (easier task).

non-planar and asymmetric objects, the performance was close to chance level. The effect of binocular disparity was small. Specifically, binocular performance in a given condition was not reliable, unless monocular performance was reliable. The fact that there was a high correlation between monocular and binocular performance suggests that the same perceptual mechanism is used.

This study showed that simplicity constraints are critical in determining the shape percept and the role of binocular disparity is secondary. Will the percept still be determined by constraints if depth cues provide conflicting information? This question was examined by Pizlo et al. (2005). Their subjects were presented with a pair of stereoscopic images of a cube. The image in the right eye was stationary and the image in the left eye was changing: it was a projection of a cube rotating (oscillating) around the vertical axis (refer to the demo at HTTP://VIPER.PSYCH.PURDUE.EDU/PIZLO_CUBES/). Consequently, the binocular disparity of the corner of the cube at the visible Y junction was changing. If binocular disparity were critical in determining the 3D shape percept, subjects would perceive a non-rigid cube and the corner would move back and forth along the visual line emanating from the right eye. However, all subjects reported that they perceived a rigid cube that oscillated around the vertical axis. This percept suggests that simplicity constraint could play more important role in determine human 3D shape perception than binocular disparity.

There has Been no Quantitative Model of 3D Shape Recovery Based on Depth Cues

Although some researchers formulated models that can reconstruct the orientation of surfaces from depth cues, there is no model that can recover the 3D shape from a real 2D image. The main problem is related to the unreliability of depth cues. When the surface orientation at one point is recovered with errors, combining these estimates across many surface points will lead to substantially larger errors in estimating the entire shape of a 3D object.

The four problems listed above suggest that surfaces and depth cues are not sufficient to explain the percept of 3D shapes. Information provided by edges and curves in a 2D image are probably more important for 3D shape perception. In particular, 2D symmetric edges and curves provide all the information that is needed to apply 3D simplicity constraints.

Why are simplicity constraints, like planarity and symmetry, so important in 3D shape perception? Simplicity constraints are important because they can dramatically decrease the number of possible 3D interpretations of a 2D image. It is known that three points in a 3D space define a plane uniquely and consequently the position of every point on the plane can be determined once its 2D image coordinates are given. Without the assumption of planarity, one would have to recover each surface point independently, which can lead to as many free parameters as there are image points. Mirror symmetry is even more powerful. As Vetter & Poggio (1994) showed, given a single 2D image of a mirror symmetric object allows one to compute a second image of the 3D object. This image is called a virtual image. Now, one is faced with a problem of recovering a 3D shape from two images (views). This is a much easier problem compared to 3D recovery from only one view (Ullman, 1996).

Based on the simplicity constraints, we developed computational models to explain human monocular and binocular 3D shape perception. Our models are the elaboration of earlier models by Marill (1991), Leclerc and Fischler (1992), Sinha (1995) and Chan et al. (2006).

MODEL

When we view a symmetric 3D object (η_0) monocularly or binocularly, the 3D object is projected on the retina(s) and forms retinal image(s) (I). To perceive the 3D shape, our visual system needs to recover the 3D shape from the 2D retinal image(s). Next we will introduce our models and explain how they recover a 3D shape from its image(s). This section includes four parts: (1) introducing a monocular recovery model; (2) introducing a binocular recovery model; (3) deriving a method to measure the dissimilarity between 3D shapes; (4) deriving predictions from our recovery models.

Monocular Recovery

Monocular recovery refers to computing a 3D shape (η_0) from its single 2D image (I). We assume that the image (I) is a 2D orthographic projection of η_0. We further assume that the 2D contours of the object have been identified in the image. Finally, we assume that η_0 is a mirror-symmetric 3D shape. The case of mirror-symmetric objects is especially relevant to human shape perception because most important objects in our environment, such as animal and human bodies, as well as man-made objects are mirror-symmetric. In the rest of this document, we will use “symmetric” to mean “mirror-symmetric.”

To recover a symmetric 3D shape, we need to first establish symmetric correspondences among all points in an image. Recall that a 2D image of a 3D symmetric shape is itself not symmetric (except for degenerate cases). So, when we talk about establishing symmetric correspondences of image points, we mean determining which pairs of points in the 2D image are projections of pairs of symmetric points in the 3D shape. Such pairs of image points will be called pairs of corresponding points. For each point, we need to know where its

corresponding point is. In this paper, we use the apostrophe (') to represent the corresponding point. To simplify the model, we only consider a one-to-one correspondence. For example, we assume that each point a, has only one corresponding point a', and vice versa. This can be expressed as follows (a')' = a. We use Ξ to represent the set of all symmetric correspondences in an image. Suppose there are n correspondences in an image. Then Ξ can be expressed as follows

$$\Xi = \{<a_i, a_i'> \quad i = 1,2,3,\ldots,n\} \tag{1}$$

In a perfectly symmetric 3D shape, the lines connecting the symmetric pairs are parallel, and the orthographic projections of those lines in the image plane are also parallel. Using this parallelism property, we can determine the set of correspondences (Ξ) as follows.

1. Specify a direction τ as the direction of the lines connecting the corresponding pairs of points in an image. For example, in Figure 4(a), the direction of the line a_1a_1' is assumed to be τ.
2. For all other points (a_i) in one curve of the image, find their corresponding points (a_i') in another curve by using the direction τ. Specifically, in Figure 4(a), for a point a_2 in one curve, we draw a line that passes through a_2 and is parallel to a_1a_1'. The intersection (a_2') with the other curve is the point corresponding to a_2.

Applying this method, we can find the correspondences for each point on a curve. However, the set of correspondence (Ξ) obtained this way is not unique because it depends on the choice of τ. Note that τ is the tilt of the symmetry plane of the recovered 3D shape. Therefore, different tilts of the symmetry

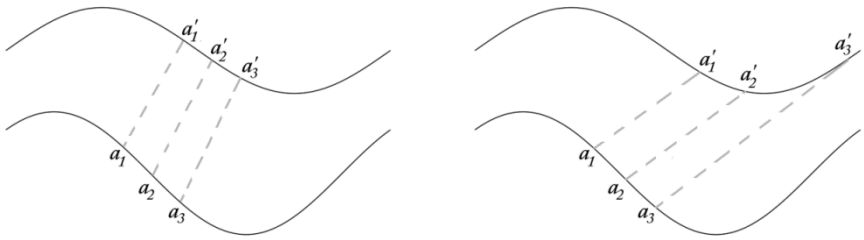


Figure 4. Illustration of two sets of symmetric correspondences for the same pairs of curves.

plane will lead to different correspondences (Ξ). Figure 4 illustrates how the set of correspondences (Ξ) is affected by τ. In Figure 4(a) and (b), there are three pairs of corresponding points (<a_1, a_1'>, <a_2, a_2'> and <a_3, a_3'>) and the lines aa_1',a_2a_2' and a_3a_3' are parallel. Although the images in Figure 4(a) and (b) are the same and points a_1, a_2 and a_3 are at the same positions, their corresponding points a_1', a_2' and a_3' are at different positions.

How our visual system determines Ξ is beyond the scope of this dissertation and it is left for future study. Here, we assume that the set of correspondences (Ξ) among all points is known and it is consistent with that of the original 3D shape η_0.

Once the set of corresponding points (Ξ) in an image is determined, the 3D shape can be recovered. First we set up the Cartesian coordinate system, which is defined as follows: The image plane is the XY plane. The X axis has the same direction as τ and the Y axis is orthogonal to the X axis in the image plane. The origin of the coordinate system can be set at an arbitrary point in the image plane. The Z axis is perpendicular to the image plane and indicates the direction in depth. Figure 5 illustrates the Cartesian coordinate system we use for the recovery of the 3D shape (η_0).

Using this coordinate system, recovering the 3D shape is equivalent to computing the Z value for each point in the image. We use the lower case letter to represent the coordinate of a point in the image plane and a capital letter to represent the coordinate of a recovered point in the 3D space. For any pair of corresponding points <a_i, a_i'>, their coordinates are (x_i, y_i) and (x_i', y_i'). The coordinates of the recovered symmetric pair <A_i, A_i'> are (X_i, Y_i, Z_i) and (X_i', Y_i', Z_i'). From the property of the orthographic projection, the following equations are satisfied:

$$x_i = X_i \quad (2)$$

$$y_i = Y_i \quad (3)$$

$$x_i' = X_i' \quad (4)$$

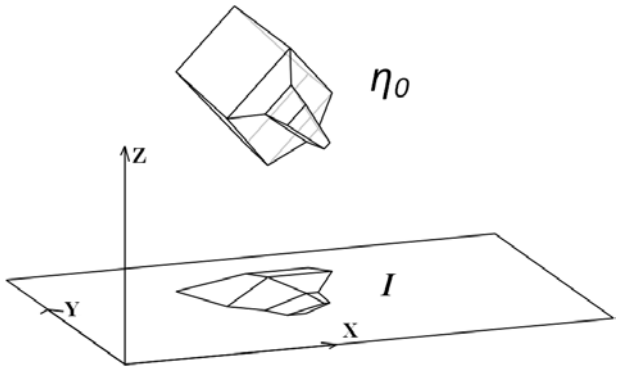


Figure 5. The coordinate system used for 3D shape recovery. XY plane is the image plane. X axis represents the direction of τ (the direction of lines connecting the corresponding points in the image). Y axis is orthogonal to X axis. Z axis is perpendicular to the image plane and it indicates the direction in depth.

$$y_i' = Y_i' \quad (5)$$

For the coordinate system defined above, the Y values for a and a' are the same. It follows

$$y_i = y_i' \quad (6)$$

From equations (3) and (5), we can obtain

$$Y_i = Y_i' \quad (7)$$

As a result, the points of any symmetric pair $<A_i, A_i'>$, have same Y value. Let π_S be the symmetry plane of the recovered 3D shape and l_S be the intersection of the image plane π_{XY} and π_S. Then we have the following properties:

1. A_i and A_i' are symmetric with respect to π_S. Hence the line A_iA_i' is perpendicular to π_S. Because l_S is on the plane π_S, l_S is perpendicular to A_iA_i'.
2. a_i and a_i' are the orthographic projections of A_i and A_i'. Hence the lines a_iA_i and $a_i'A_i'$ are perpendicular to the image plane π_{XY}. Because l_S is on the plane π_{XY}, l_S is perpendicular to a_iA_i and $a_i'A_i'$.
3. From (1) and (2) it follows that l_S is perpendicular to the plane defined by a_i, a_i', A_i, and A_i'. From this, it follows that l_S is perpendicular to the line a_ia_i'.
4. Because a_i and a_i' have the same y value, the line a_ia_i' is perpendicular to Y axis.
5. From (3) and (4) it follows that l_S is parallel to Y axis.

Without loss of generality, let l_S coincide with the Y axis (see Figure 6). Let α be the angle between the symmetry plane π_s and the image plane π_{XY} and the domain of α is (-90, 90). Then the normal of the symmetry plane (π_s) is [sin(α), 0, –cos(α)] and the plane is expressed by

$$\sin(\alpha)X-\cos(\alpha)\,Z = 0 \tag{8}$$

Since all symmetric pairs $<A_i, A_i'>$, are symmetric with respect to the symmetry plane π_s, A_i and A_i' must satisfy the following two conditions:

1. The midpoint of A_i and A_i' is on the symmetry plane π_s. It follows

$$0.5(X_i+X_i')\sin(\alpha)-0.5(Z_i+Z_i')\cos(\alpha)=0 \tag{9}$$

2. The line connecting A_i and A_i' is perpendicular to the symmetry plane π_s. It follows

$$(X_i-X_i')\cos(\alpha)+(Z_i-Z_i')\sin(\alpha)=0 \tag{10}$$

From equations (9) and (10), we obtain

$$Z_i = (-\cos(2\alpha)X_i+X_i')/\sin(2\alpha) \tag{11}$$
$$Z_i'=(-\cos(2\alpha)X_i'+X_i)/\sin(2\alpha) \tag{12}$$

Note, $X_i = (X_i')'$.Thus, equation (12) can be written as

$$Z_i'=(-\cos(2\alpha)X_i'+(X_i')')/\sin(2\alpha) \tag{13}$$

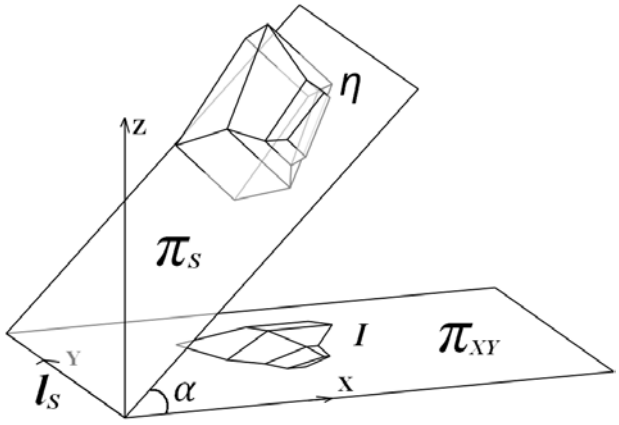


Figure 6. The illustration of 3D shape recovery. η is a recovered 3D shape from the image I. l_S is the intersection of the symmetry plane π_S of the recovered 3D shape and the image plane π_{XY} and it coincides with the Y axis. α is the angle between π_S and π_{XY}.

We see that equations (13) and (11) have the same form. Therefore, equation (11) can be used for all points, primed and unprimed. Based on equation (11) we can state the following:

1. To compute the Z value of a point (A_i), we need to know not only its X value, but also the X value of its symmetric counterpart (A_i'). In other words, both of their orthographic projections a_i and a_i' must be visible. For those pairs for which one of the points is hidden or invisible, equation (11) is not applicable. In this case, we will use a planarity constraint to recover these points - the details of this method are presented later in this dissertation.
2. In equation (11), X_i and X_i' are obtained from the image. The only unknown variable is α (the angle between the symmetry plane π_S and the image plane π_{XY}). Therefore, the recovery is uniquely determined by α. Let η(α) be the recovered 3D shape when the angle between π_S and π_{XY} is α. Figure 7 illustrates three possible recovered 3D shapes corresponding to three angles (α): -60, -45 and -30. Thus, once the set of symmetric correspondences in an image is set up, the 3D shape recovery is characterized by one free parameter.
3. Because Z(-α) = -Z(α) and the domain of α is symmetric, equation (11) is an odd function of α. For any two recovered points Z_i and Z_j, they have the following relation

$$Z_i(\alpha)-Z_j(\alpha) = -(Z_i(-\alpha)-Z_j(-\alpha)). \quad (14)$$

This means that in the two recoveries η(α) and η(-α), the distance between any two points is the same. It follows that these two 3D shapes are identical except that they are

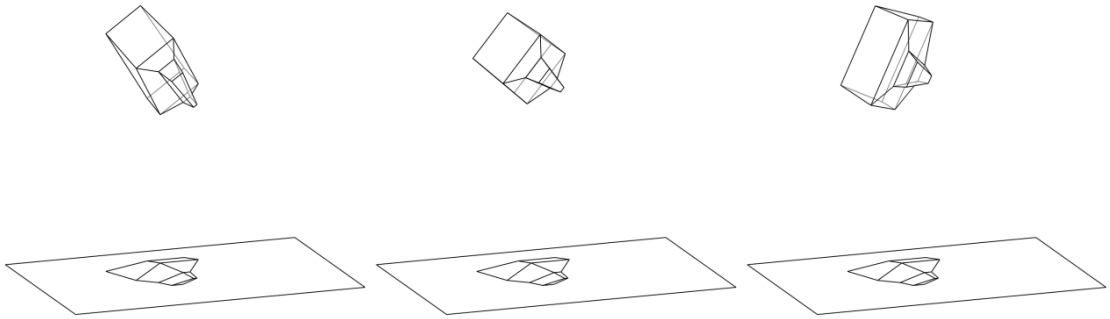


Figure 7. Three recoveries from the same image. The angles (α) between the symmetry plane of the recovered 3D shapes and the image plane are -60, -45, and -30, respectively.

inverted in depth. Note that |α| is the slant of the symmetry plane. Therefore, if the slant of the symmetry plane of the recovered 3D shape is known, two identical 3D shapes can be produced. These two shapes are related to each other by depth reversal.

To summarize, given a 2D orthographic projection (I) of a symmetric 3D shape (η_0), if we know the symmetric correspondence (Ξ) among all points, there are infinitely many symmetric 3D interpretations and they are determined by the angle (α) between the image plane (π_{XY}) and symmetry plane (π_S). Suppose all those symmetric 3D interpretations form a set Ψ, which can be expressed as follows:

$$\Psi_{(I,\Xi)}=\{\eta=\eta(I,\Xi,\alpha),\ \alpha\in(-90,90)\} \tag{15}$$

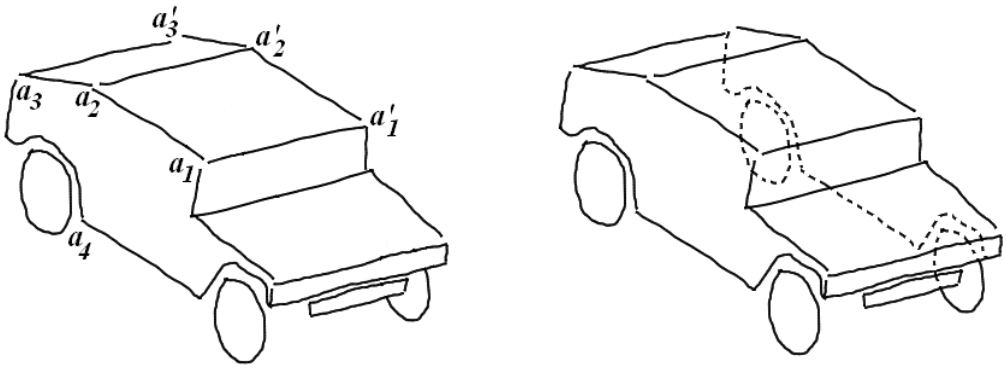
Expression (15) explicitly states that the key to the recovery of a unique symmetric 3D shape is to decide what the orientation of the symmetry plane is. We already know that the tilt of the symmetry plane determines the correspondences Ξ. It will be shown below that the slant of the symmetry plane determines the aspect ratio of the recovered 3D shape. Because we have assumed that the correspondences (Ξ) are known for a given image, the symbols I and Ξ will henceforth be ignored in equations and expressions. In particular, the expression (15) can be written as

$$\Psi=\{\eta=\eta(\alpha),\ \alpha\in(-90,90)\} \tag{16}$$

So far we have explained how to recover the depth of a pair of symmetric points when both points are visible. If one of the two points is occluded, these points cannot be recovered using the method described. For

example, in the case of point a_4 in Figure 8(a), its counterpart is invisible and we cannot use equation (11) to compute the Z value for this point. For this case, we will first use a planarity constraint to recover the visible point a_4. Specifically, we begin by recovering three pairs of visible points: $<a_1, a_1'>$, $<a_2, a_2'>$ and $<a_3, a_3'>$. If we know that a_4 is on the plane determined by a_1, a_2 and a_3, we compute the Z value for a_4. There are two ways to do it.

1. Using equation (11), we recover points (A_1, A_2 and A_3) as well as their symmetric counterparts from a_1, a_2, a_3 and a_1', a_2', a_3'.Then we compute the plane defined by A_1, A_2 and A_3. Finally, we compute the intersection between the plane and the line that passes through a_4 and is orthogonal to the image plane. The intersection is the recovered point (A_4) whose image is a_4. Because A_1, A_2 and A_3 , as well as their symmetric counterparts (A_1', A_2' and A_3') have already been computed, we know the symmetry plane. The symmetric counterpart of A_4 is obtained by reflecting A_4 with respect to the symmetry plane.
2. Suppose a_4' is the symmetric counterpart of a_4. Because a_4 is on the same plane as a_1, a_2 and a_3, a_4' must be on the same plane as a_1', a_2' and a_3'. Since these two planes are symmetric, their orthographic projections on the image plane are related by a 2D affine transformation (refer to Appendix A for the details of the proof). The parameters of a 2D affine transformation can be uniquely determined by 3 points. Because $<a_1, a_1'>$, $<a_2, a_2'>$ and $<a_3, a_3'>$ are the pairs of corresponding points in the 2D image, we can compute the parameters of the 2D affine transformation between these points. Next we apply this affine transformation to a_4 and obtain its invisible counterpart (a_4'). Then we can use



(a) (b)

Figure 8. The illustration of computing the hidden curves (or points) using the affine transformation method. (a) An image of car. <a_1, a_1'>, <a_2, a_2'> and <a_3, a_3'> are pairs of corresponding points. The symmetric counterpart of a_4 is hidden. (b) The hidden curves, computed by applying a 2D affine transformation, are shown as dashed lines.

equation (11) to compute the Z values of A_4 and A_4'. Figure 8(b) shows the hidden curves computed using this method[1].

The derivation described above shows that a symmetric 3D interpretation of an image is not unique. So, which 3D shape from the set (ψ) of recoveries corresponds to the subject's percept? We introduce two other constraints – maximum 3D compactness and minimum surface area. The compactness (C) of a 3D shape (η) is defined as $C(\eta)=V(\eta)^2/S(\eta)^3$, where V(η) represents the volume of η and S(η) represents the surface area of η. In the set ψ, we choose the 3D shape η_C that has the maximum compactness:

$$\eta_C=\arg\max(C(\eta)) \qquad \eta\in\psi \tag{17}$$

Similarly, the minimum surface area constraint chooses in the set ψ the 3D shape that has the minimum surface area:

$$\eta_S=\arg\min(S(\eta)) \qquad \eta\in\psi \tag{18}$$

Both of these constraints can lead to a unique 3D recovery which is close to the percept. We found, however, that the best results are produced by a combination of the two constraints. By best, we mean the most veridical and the closest to the subject's percept. How can these two constraints be combined? Note that minimizing surface area is equivalent to maximizing 1/S(η):

[1] The 2D affine transformation is equivalent, in this case to a simple rigid translation in the 2D image. This is because the two sides of the truck are parallel in 3D. However, in the general case, the 2D affine transformation will not be a rigid motion.

$$\arg\min(S(\eta)) = \arg\max(1/S(\eta)) \qquad \eta\in\psi \tag{19}$$

It follows that a combination of these two constraints is equivalent to maximizing $C(\eta) \oplus 1/S(\eta)$, which can be expressed as

$$\eta_M=\arg\max(C(\eta) \oplus 1/S(\eta)) \qquad \eta\in\psi \tag{20}$$

The binary operation $\oplus$ is defined as follows:

$$a\oplus b = a\times b^n \quad n\geq 0 \tag{21}$$

So, the combination of these two constraints is equivalent to

$$\eta_M=\arg\max(V(\eta)^2/S(\eta)^{3+n}) \qquad \eta\in\psi \tag{22}$$

The exponent n represents the relative weight assigned to the constraints. When n=0, the combination is equivalent to the maximum 3D compactness constraint; when $n=\infty$, the combined constraint is equivalent to the minimum surface area constraint. Our simulations showed that n=3 is optimal in the sense that the recoveries are close to veridcal. It follows that the combined constraint corresponds to maximizing $V(\eta)^2/S(\eta)^6$, which is equivalent to maximizing $V(\eta)/S(\eta)^3$. Therefore, the recovered 3D shape from the monocular model can be expressed as

$$\eta_M = \arg\max(V(\eta)/S(\eta)^3) \qquad \eta\in\psi \tag{23}$$

Suppose η_M is the recovered 3D shape when a 3D shape η_0 is viewed monocularly. Equation (23) defines the mapping between η_0 and η_M. We use the following function to represent this mapping,

$$\eta_M = f_M(\eta_0) \tag{24}$$

The above model can only be applied to those 3D shapes whose volumes and surfaces are well defined, like polyhedron. However, for many objects we meet in our daily life, like airplanes, trees, chairs, birds, the contours present in the image do not uniquely specify the volume and/or surface of the 3D object. This was the case with the truck in Figure 8. To make our model more general, we modified the above model: we compute the convex hull (H) for each recovered shape in ψ first. For a 3D shape η, its convex hull (H(η)) is the smallest convex 3D object that contains η (see Figure 9). Since the volume and surface area of a convex hull is well defined, we can compute its $V(H(\eta))/S(H(\eta))^3$. Then we find the 3D shape whose convex hull has the maximum V/S^3. This process can be expressed as

$$\eta_{m'} = \arg\max(V(H(\eta))/S(H(\eta))^3) \quad \eta \in \psi \tag{25}$$

This monocular model has been applied to recover many different kinds of 3D shapes, such as polyhedra (Li, Pizlo & Steinman, 2009) and parallelepipeds (Li, 2009). The recovery of real 3D objects, such as animals (Li & Pizlo, 2008), shows that maximizing V/S^3 of the convex hull works well. The model's recovery is very similar to human percepts. The following websites show examples: HTTP://WWW1.PSYCH.PURDUE.EDU/~SAWADA/MINIREVIEW/DEMO_POLY_A.HTML and HTTP://WWW1.PSYCH.PURDUE.EDU/~SAWADA/MINIREVIEW/DEMO_TRUCK.HTML.

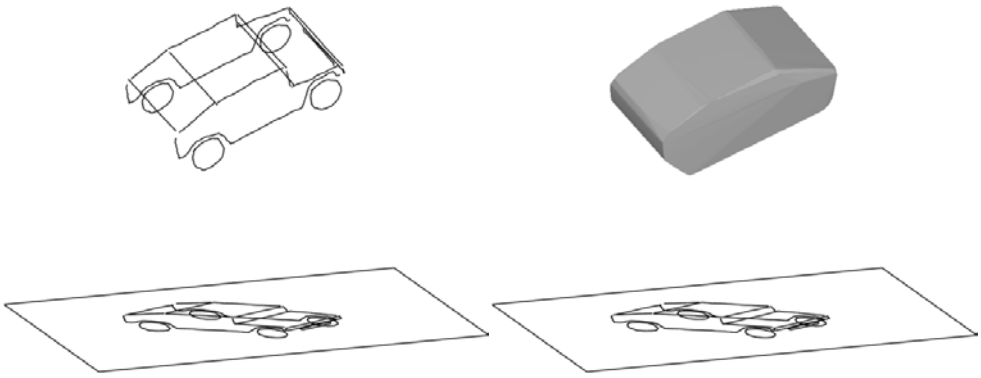


Figure 9. The left picture illustrates a recovered jeep from the image. The right one shows the convex hull of the recovered jeep.

Note, that the recovery from a real image is a little different from the recovery of a polyhedron from a synthetic image. In Appendix B, the process of recovery from a real image is described, including how noise in an image is handled.

Binocular Recovery

Because human eyes are about 6 cm apart, what the two eyes see is slightly different - the left eye sees more of the left side of an object and the right eye sees more of the right side. This difference is called binocular disparity. Although this phenomenon has been known for a long time (Euclid described it in about 300 B.C.), the relation between binocular disparity and depth perception hadn't been demonstrated until Wheatstone designed the first stereoscope in 1838. Wheatstone used two mirrors to reflect a pair of pictures (the pair of pictures is called a stereogram) to the two eyes separately. When the two images are fused, the object rendered in the pictures appears solid. This phenomenon can be explained from a geometrical view. Figure 10 illustrates the relation between depth and binocular disparity in a simple case of two points. Suppose the eyes fixate the point F at distance d from the eyes. F_L and F_R are the retinal images of F, and they fall on the fovea of the left and right eyes. Another point A is at a distance of Δd behind F. The visual angle between A_L and F_L in the left eye is α_L. If A_L is to the right of F_L, the sign of α_L is positive and if A_L is to the left of F_L, the sign of α_L is negative. The visual angle between A_R and F_R in the right eye is α_R (the rule for deciding about the sign is the same as in the left image). If two points on the left and right retina form the same visual angles with their corresponding foveas, these two points are called corresponding points. The binocular disparity (δ in radians) of the images of F is zero because F_L and F_R fall on the corresponding points. The binocular disparity of the images of A is α_L-α_R. Also note that binocular disparity of the images of A is equal to ω-θ. If the fixation distance (d) is much greater than depth (Δd) between A and F, and both points are close to the saggital plane, binocular disparity satisfies the following formula (Howard & Rogers 2002):

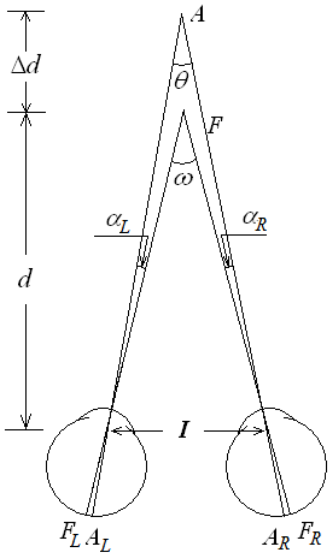


Figure 10. F is the fixated point at distance d from the eyes. F_L and F_R are the retinal images of point F and they are on the fovea of each retina. A is a 3D point which is Δd behind F. A_L and A_R are the retinal images of point A. The interocular distance is I.

$$\delta \approx \frac{I\Delta d}{d^2} \quad (26)$$

where I is the interocular distance. The relation between binocular disparity and relative depth is linear - the bigger the binocular disparity, the bigger the relative depth. If we assume that the interocular distance (I) and the distance (d) between fixation point F and the eyes are known, the relative depth of point A with respect to point F can be computed from binocular disparity.

Julesz's (1971) invention of a random-dot stereogram psychophysically proved that a 3D object can be perceived through binocular disparity only. The finding of disparity-tuned cells in monkey's primary visual cortex (Hubel & Wiesel 1963, Hubel, 1995), which indicates that binocular disparity is used by the visual system.

These geometrical, empirical and biological facts suggest that binocular disparity plays an important role in perceiving depth and 3D shapes. However, many experiments (e.g., Norman, Todd, Perotti, & Tittle, 1996; Todd & Norman, 2003; Johnston, 1991) showed that the 3D shape percept from binocular disparity is neither accurate nor reliable. Consider Johnston's (1991) experiment, which is representative for these groups of studies. She asked the subject to view a random-dot stereogram of an elliptical cylinder and adjust its depth so that it is perceived as circular. The viewing direction was orthogonal to the axis of the cylinder. She found that at an intermediate distance (about 1 meter), a subject's percept was close to veridical (that is a circular cylinder was perceived as circular). When the viewing distance was greater than 1 meter, subjects systematically underestimated the depth and, as a result, a cylinder that was stretched in depth, compared to a circular cylinder, was perceived as circular. The converse was true for distances less than 1 meter. The systematic error was up to a factor of 2.

There is, however, another type of binocular judgment, which is extremely reliable. Namely, the observers can make very accurate judgments

about order of points in depth (Blakemore 1970; Westheimer 1979). In fact, this judgment can be done with a precision that is one order of magnitude better than the distance between receptors on the retina (in technical jargon, this is called “sub-pixel resolution) (Westheimer & McKee, 1980). Because of this high precision, this judgment is called “hyperacuity” (it is also called stereoacuity).

To summarize, binocular vision is quite unreliable in judging the depths of points and 3D distances, but it is extremely reliable in judging depth order of points. How can the ordinal depth be incorporated into our shape recovery model? This is done in two steps:

1. Find the subset θ of ψ, in which the 3D shapes have the depth order as determined by stereoacuity;
2. In the subset θ, choose the one that has the maximum V/S^3.

The limits of stereoacuity are usually characterized by its threshold. Threshold refers to the smallest distance in depth that can be reliably detected. “Reliably” usually means 75% or 84% of the time. Therefore, for two points A_i and A_j, there are three possibilities for the decision about their depth order:

1. A_i is judged as farther than A_j. We write it as $A_i > A_j$;
2. The order is uncertain. We write it as $A_i \sim A_j$;
3. A_i is judged as closer than A_j. We write it as $A_i < A_j$;

Let $O(A_i,A_j)$ represent the judgment about depth order between two points:

$$O(A_i,A_j) = \begin{cases} 1 & A_i > A_j \\ 0 & A_i \sim A_j \\ -1 & A_i < A_j \end{cases} \quad (27)$$

The ability to detect the depth order is affected by several factors:

1. The depth difference (Δd) between A_i and A_j. If this difference is much larger than the threshold, depth order is easy to judge.
2. The angular distance between A_i and A_j. When the angular distance of two points increases, the stereoacuity threshold increases, as well.
3. The viewing distance (d). When the viewing distance increases, the separation in depth corresponding to threshold increases, according to the formula (26).
4. Eye movements. Wright (1951) showed that when the subject was allowed to change the fixation between two points, the stereoscopic acuity was better than that when he/she fixated on one point.
5. Context. The depth order judgment is affected by the surface on which the points reside. Norman & Todd (1998) showed that when two points are shown in total darkness, subjects can easily judge their depth order. However, when the points are perceived as lying on a surface, the stereoscopic threshold increases substantially.
6. Individual differences. Ogle (1950) tested two subjects' stereoscopic acuity and found that there were substantial differences between subjects: the difference in the stereoacuity threshold was up to a factor of 2. Large individual variability has been confirmed in a number of other studies.

From several papers that reported stereoacuity thresholds, we chose Rady & Ishak's (1955) study to simulate the human's ability to detect depth

order between points because their study was most comprehensive. Specifically, (1) several factors mentioned above were considered in their study (depth difference, angular distance, eye movements); (2) they used five angular separations of the two points, which is more than what could be found in other studies; (3) they measured the stereoscopic acuity for points without any context, which leads to the lowest thresholds. The simulation using their result will provide an upper limit for binocular shape recovery. Note that stereoacuity threshold changes with the change of angular separation of two points – large separation results in high threshold and small separation results in low threshold. This effect of angular separation on the threshold results from the non-uniform distribution of receptors in the human retina. This means that the simulation of stereoacuity threshold in human's depth order judgment represents the finite and non-uniform resolution on the retinas. The details of how we used these stereoscopic thresholds in our simulation model are given in Appendix C. The simulation model uses the following two properties: given two points A_i,A_j,

1. $O(A_i,A_j) = -O(A_j,A_i)$. This means that if a subject can tell that the point A_i is farther than A_j, he/she will also be able to tell that A_j is closer than A_i;
2. $O(A_i,A_i) = 0$. This means that if two points are at the same position, the depth order between them is uncertain.

For a 3D shape η_0, suppose there are n visible points when viewing it binocularly from some direction. Then we can build a n×n matrix ($M_{\eta 0}$) in which the value of each element (i,j) represents the depth order between A_i and A_j:

$$M(i,j) = O(A_i,A_j) \qquad i,j = 1,2,\ldots,n. \qquad (28)$$

We call this the “depth order matrix.” Corresponding to the properties of $O(A_i,A_j)$, the depth order matrix has the following two properties:

1. It is a skew symmetric matrix because M(i,j) = -M(j,i).
2. The elements on the diagonal are 0.

In a depth order matrix, the nonzero elements are called “valid” elements and the zero elements are called invalid elements. In order to discuss the properties of the depth order matrix, consider an example. Let M_1 and M_2 be two depth order matrices.

$$M_1 = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

The depth order matrix represents the state of depth order between points in a 3D shape. Having two (or more) depth order matrices, we can only compare them, but it does not make sense to perform ordinary matrix operations on them. For two depth order matrices M_k and M_l, if the element (i,j) in both M_k and M_l is valid, we call the element (i,j) comparable for M_k and M_l. Otherwise, the element is not comparable. For example, for the above two depth order matrices M_1 and M_2, the elements (1,2), (2,1), (2,3) and (3,2) are comparable. On the other hand, the elements (1,3) and (3,1) are not comparable because they are invalid in M_2. The equality between M_k and M_l is defined as follows: $M_k = M_l$ if and only if for all comparable elements (i,j), $M_k(i,j) = M_l(i,j)$. Note the equality between depth order matrices is not transitive, which means $M_k \approx M_l$ and $M_l \approx M_n$ does not imply $M_k \approx M_n$.

So, even though the values at (1, 3) and (3, 1) in M_1 and M_2 are different, M_1 are M_2 are equal because the values in all comparable elements

are the same. This definition of equality of depth order matrices will allow comparing two identical shapes presented at different viewing distances. When the viewing distance for one shape is very large, most of the depth orders will be uncertain. In the extreme case, this model will allow the comparison of two 3D shapes, one viewed monocularly and the other viewed binocularly (see below).

Figure 11 illustrates the comparison of the depth order matrices between two complex 3D shapes. Both shapes are possible 3D interpretations of the same 2D image, in which 13 points are visible. The shape on the left is labeled as η(-30) because the angle (α) between its symmetry plane and image plane is -30 degrees, and the one on the right is labeled as η(-45). Suppose both of them are viewed from the distance of 50cm (recall that the viewing distance is important in deciding whether the depth order for a pair of points can be discriminated). The depth order matrices for η(-30) and η(-45) are computed. For illustration purposes, the matrices are visualized and the cell with the value of 1, 0 or -1 is filled in red, gray or blue, respectively. Comparing the red and blue patches between these two color squares, we see that $M_{\eta(-30)} \neq M_{\eta(-45)}$. Since our visual system can tell that the depth orders are different for some pairs of points, these two 3D shapes will be considered different. Next we will explain how to use the depth order matrix to recover a 3D shape.

Suppose that: (1) a symmetric 3D shape η_0 is viewed binocularly from a distance d and its depth order matrix is $M_{\eta 0}$, (2) its orthographic projection on the cyclopean eye is I (cyclopean eye is an abstract concept referring to the information available if the camera were placed at the midpoint of the two eyes. This image can be estimated by combining the images from the left and right eyes); and (3) the set of all symmetric recoveries from the image I is ψ. We obtain a subset θ of ψ, in which for all recovered shapes, the depth order matrices are equal to $M_{\eta 0}$ when the shapes are viewed from the same distance d. This subset is expressed as follows

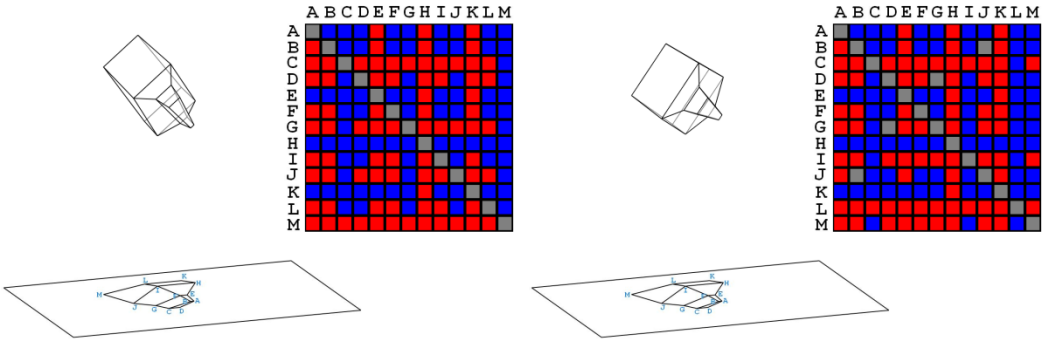


Figure 11. Two depth order matrices corresponding to two 3D shapes. The two 3D shapes are possible 3D interpretations of the same 2D image and they are viewed from a distance of 50cm. Both of them have 13 visible points. The color square represents the depth order between any two visible points. Red patch at (i,j) represents that point i is farther than point j, blue represents that point i is closer than point j, and gray represents that the depth order between point i and j is uncertain. By comparing the red and blue patches between the two squares, we see that these two depth order matrices are not equal.

$$\Theta_{\eta 0} = \{\eta \mid M_{\eta} \approx M_{\eta 0} \text{ and } \eta \in \psi\} \quad (29)$$

Note that stereoacuity alone cannot tell the difference between the 3D shapes in $\Theta_{\eta 0}$. Then, which one corresponds to our percept when we view η_0 binocularly? We conjecture that our visual system chooses the 3D shape that jointly maximizes 3D compactness and minimizes the surface area. In other words, the binocular recovery, similarly to the monocular recovery, will choose a 3D shape that has the maximum V/S^3 in the set θ. It is expressed as

$$\eta_B = \arg\max(V(\eta)/S(\eta)^3) \qquad \eta \in \theta \quad (30)$$

Now, the mapping between η_B and η_0 is defined as

$$\eta_B = f_B(\eta_0) \quad (31)$$

Like the symmetry constraint that reduces the degrees of freedom (DOF) of recovery from N to 1 (N is the number of visible points), stereoacuity limits the possible interpretations to a small set and this set usually leads to better performance. It is this reason why binocular performance is usually better than monocular performance. However, when the viewing distance increases, binocular performance becomes worse because the stereoacuity threshold, when expressed in cm, becomes higher. Our binocular model can account for this fact. It is known that stereoacuity is affected by the viewing distance. Large viewing distances lead to a less reliable judgment of depth order between points. Figure 12 illustrates the change of the depth order matrix when a 3D shape is viewed at two different distances. When the viewing distance is 100cm, the depth orders of only a few pairs of points are uncertain (see Figure 12(a)). However, when the viewing distance is increased to 1000cm, the depth order of most pairs of points is uncertain (in Figure 12(b),

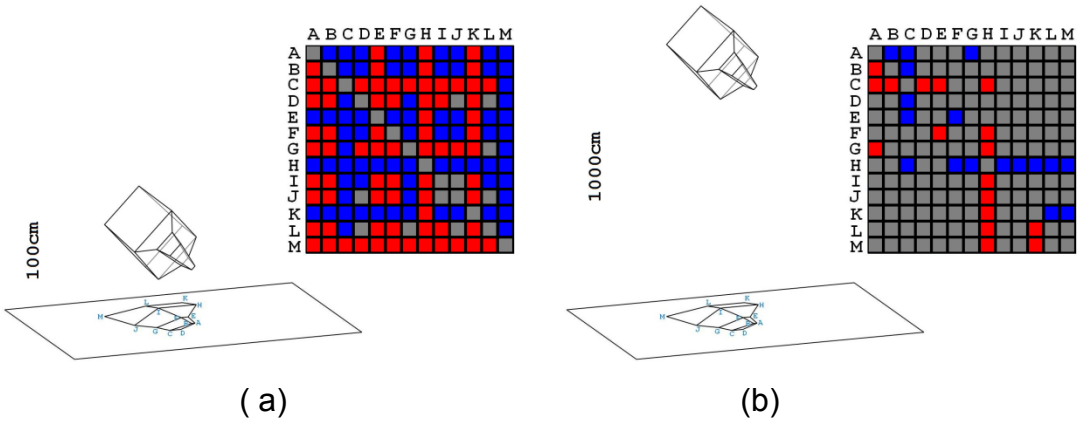


Figure 12. The comparison of the depth order matrices when a 3D shape is viewed at two different distances. (a) The viewing distance is 100cm and in its depth order matrix, the values of most elements are non-zero. (b) The viewing distance is 1000cm and in its depth order matrix, the values of most elements are zero.

most of cells are filled in gray). This means that the number of valid elements in this depth order matrix is small, which will lead to a larger set of possible 3D interpretations and consequently less reliable recovery of a 3D shape. Once the viewing distance is large enough, the depth order between all pairs of points is uncertain and the number of valid elements is decreased to 0. Then the set θ is the same as ψ. At this condition, binocular disparity has no effect and the recovery by the binocular model is same as that by the monocular model.

Note that expression (30) suggests that the mapping between the real 3D shape and the recovered 3D shapes is many-to-one, not one-to-one. In other words, we can find a subset (φ) of ψ in which each 3D shape has the same recovery (η_B). Or it is simply expressed as

$$\varphi = \{\eta \mid \eta_B = f_B(\eta), \eta \in \psi\} \tag{32}$$

For example, suppose η_{max} is the 3D shape that has the maximum V/S^3 in a set ψ. For every 3D shape η in the set ψ, we compute th corresponding subset θ_η in which all 3D shapes have the same depth order matrices as that of η. Then we form a set (φ), in which for each 3D shape η, η_{max} is an element of their corresponding subset θ_η. The set φ is expressed as follows:

$$\varphi = \{\eta | \eta_{max} \in \theta_\eta\} \tag{33}$$

According to our model, for all the 3D shapes in φ, the recovered 3D shape is η_{max}. This implies that the 3D shape percept will be the same for a number of different 3D original shapes.

So far, we introduced the monocular recovery model and the binocular recovery model. To test the psychological plausibility of these models, we need to run psychophysical experiments to measure subjects' performance. Before

that, we need to establish a way to measure the dissimilarity between two 3D shapes so that we can compare the empirical data and the simulation data.

The Measures of Dissimilarity Between Two 3D Shapes

Suppose η_1 and η_2 are two 3D shapes recovered from the same 2D image. They both are in the set ψ and the angles between their symmetry planes and image plane are α_1 and α_2 (see Figure 13). From equation (11), we know that for any point (a) in the image plane, the z-coordinate of its 3D recovery (A_1) in η_1 is

$$Z_1 = \frac{-\cos(2\alpha_1)X+X}{\sin(2\alpha_1)} \tag{34}$$

and the z-coordinate of its 3D recovery (A_2) in η_2 is

$$Z_2 = \frac{-\cos(2\alpha_2)X+X}{\sin(2\alpha_2)} \tag{35}$$

After subtracting the left-hand and right-hand sides, we obtain

$$Z_2 = \frac{\cos(2\alpha_1)-\cos(2\alpha_2)}{\sin(2\alpha_2)}X + \frac{\sin(2\alpha_1)}{\sin(2\alpha_2)}Z_1 \tag{36}$$

It can be seen that equation (36) represents a 3D affine transformation between η_1 and η_2:

$$\begin{pmatrix} X \\ Y \\ Z_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{\cos(2\alpha_1)-\cos(2\alpha_2)}{\sin(2\alpha_2)} & 0 & \frac{\sin(2\alpha_1)}{\sin(2\alpha_2)} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z_1 \end{pmatrix} \tag{37}$$

Let Q represent this 3D affine transformation matrix, so that the relation between η_1 and η_2 can be written simply as $\eta_2 = Q\eta_1$. Note that Q can be

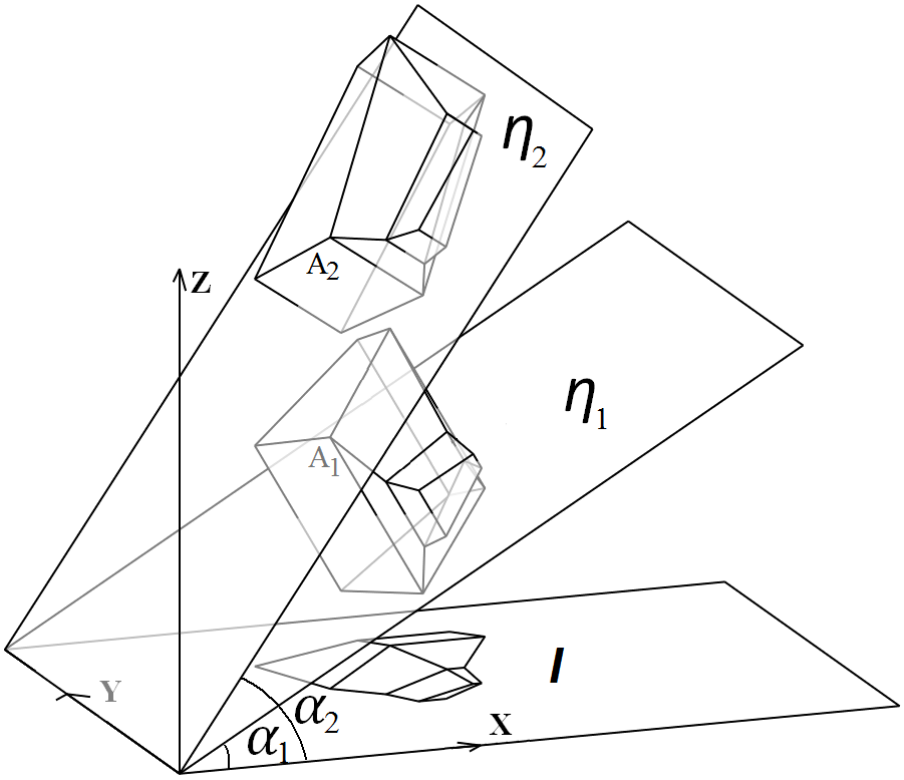


Figure 13. The comparison between two 3D shapes. η_1 and η_2 are two recovered 3D shapes from the image I. The angles between image plane and the symmetry plane of η_1 and η_2 are α_1 and α_2.A_1 is a point in η_1 whose corresponding point in η_2 is A_2.

decomposed and written as a product of three simpler matrices Q = USV', where

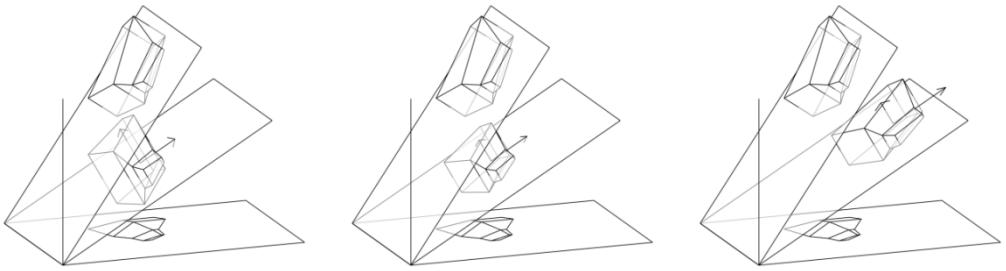
$$U = \begin{pmatrix} \cos(\alpha_2) & 0 & -\sin(\alpha_2) \\ 0 & 1 & 0 \\ \sin(\alpha_2) & 0 & \cos(\alpha_2) \end{pmatrix} \quad (38)$$

$$S = \begin{pmatrix} \frac{\cos(\alpha_1)}{\cos(\alpha_2)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{\sin(\alpha_1)}{\sin(\alpha_2)} \end{pmatrix} \quad (39)$$

$$V = \begin{pmatrix} \cos(\alpha_1) & 0 & -\sin(\alpha_1) \\ 0 & 1 & 0 \\ \sin(\alpha_1) & 0 & \cos(\alpha_1) \end{pmatrix} \quad (40)$$

U and V are orthonormal matrices, so they represent some 3D rotation operations. S is a diagonal matrix and it represents a similarity transformation. So, the affine transformation from η_1 to η_2 corresponds to a sequence of the following three simpler transformations. First, the object is rotated by $-\alpha_1$ around Y axis, then it is elongated or compressed along X and Z axes by factors of $\cos(\alpha_1)/\cos(\alpha_2)$ and $\sin(\alpha_1)/\sin(\alpha_2)$. Finally, the transformed 3D shape is rotated by α_2 around Y axis.

Conceptually, this product of three transformations can be simplified. Let's use V to define a new coordinate system. This eliminates the first rotation. The affine transformation corresponds now to the combination of a similarity transformation and a rotation. First, η_1 is stretched or compressed along new X and Y directions (in the original coordinate system the stretch and compression is performed along $[\cos(\alpha_1)\ 0\ \sin(\alpha_1)]$ and $[-\sin(\alpha_1)\ 0\ \cos(\alpha_1)]$. Then the transformed 3D shape is rotated by $(\alpha_2-\alpha_1)$ around Y axis. Figure 14 illustrates the process. In Figure 14(a), the angles (α) of the recovered 3D shapes η_1 and η_2 are 30 and 45 degrees. First, η_1 is compressed along the new Z direction (the direction orthogonal to the symmetry plane) by a factor of sin(30)/sin(45)



(a) (b) (c)

Figure 14. Illustration of the 3D affine transformation between two 3D recovered objects. (a) η_1 (the bottom) and η_2 (the top) are recovered from the same image and their corresponding angles α (the angle between the symmetry plane and the image plane) are 30 and 45 degrees. The two arrows indicate the directions along which η_1 will be compressed or stretched. (b) η_1 is compressed along the normal of its symmetry plane. (c) the resulting object is stretched along the direction indicated by the other arrow.

(see Figure 14(b)). Next, the resulting shape is stretched along the new X direction by a factor of cos(30)/cos(45) (see Figure 14(c)). After this similarity transformation, the two shapes (transformed η_1 and η_2) are identical except for their 3D orientation. Now, if we rotate the transformed η_1 by 15 degrees around Y axis, we obtain η_2.

The decomposition of Q shows that the shape changes (it is stretched and/or compressed) at the stage of the similarity transformation S. The directions of the shape changes are determined by V. Let m and n be the first and the third column vectors in V

$$m = [\cos(\alpha_1)\ 0\ \sin(\alpha_1)] \tag{41}$$

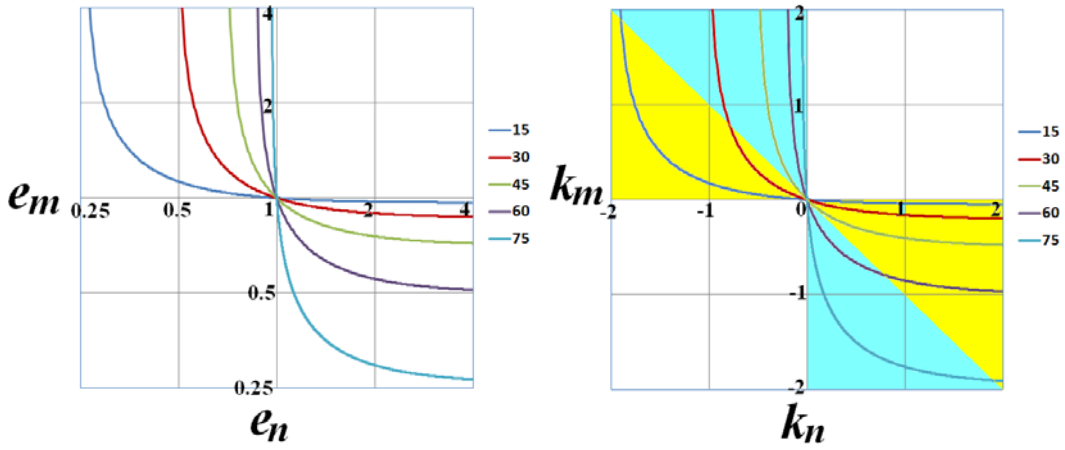
$$n = [-\sin(\alpha_1)\ 0\ \cos(\alpha_1)] \tag{42}$$

Then m and n represent the two directions of shape change. In the matrix S, the two elements ($\cos(\alpha_1)/\cos(\alpha_2)$, $\sin(\alpha_1)/\sin(\alpha_2)$) in the diagonal are the change coefficients along m and n. Note that because α_1 and $\alpha_2 \in (-90, 90)$, $\cos(\alpha_1)/\cos(\alpha_2)$ is always positive. However sin(α1)/sin(α2) can be either positive or negative. Negative means that the points on one side of the symmetry plane move to the other side after the affine transformation. Let

$$e_m = \left|\frac{\cos(\alpha_1)}{\cos(\alpha_2)}\right| \tag{43}$$

$$e_n = \left|\frac{\sin(\alpha_1)}{\sin(\alpha_2)}\right| \tag{44}$$

Then e_m and e_n represent the magnitude of change along m and n. When α_1 is fixed and α_2 is changing, the point (e_m, e_n) falls on a curve. Figure 15(a) shows the relation between e_m and e_n. Different curves correspond to different α_1. Note that when e_m=1, e_n is also equal to 1, and vice versa. This is because



(a) (b)

Figure 15. (a) When α_1 is fixed, (e_m, e_n) falls on a curve. The five curves correspond to the five slants of η_1, 15, 30, 45, 60 and 75 degrees. (b) When a point (k_m, k_n) falls in the yellow area, the dissimilarity $\varpi = k_n$. When it falls in the cyan area, $\varpi = k_m$.

e_m=1 implies that $\alpha_1 = |\alpha_2|$ and consequently e_n=1. This is illustrated in Figure 15(a) where all curves intersect at (1, 1).

Consider two values of e_m: 2 and 0.5. Both of these values represent a change by the same factor except that the former means that a 3D shape is stretched by a factor of 2 whereas the latter means that a 3D shape is compressed by a factor of 2. Since we are interested in measuring dissimilarity of two shapes, it is better to use a log transformation of e_m and e_n:

$$k_m = \log_2(e_m) \tag{45}$$

$$k_n = \log_2(e_n) \tag{46}$$

The sign(k_m) (or sign(k_n)) represents whether η_1 is elongated or compressed compared with η_2 along the direction m (or n), and $|k_m|$ (or $|k_n|$)represents the magnitude of change along the direction m (or n). For example, k_m=-1 means that η_1 is compressed along m by a factor of 2. Figure 15(b) shows the transformation from e_m-e_n space to k_m-k_n space. Similarly to e_m and e_n, if k_m=0, k_n is also equal to zero, and vice versa.

We will use max($|k_m|$,$|k_n|$) as a measure of dissimilarity (ϖ) between η_1 and η_2:

$$\varpi(\eta(\alpha_1),\eta(\alpha_2)) = \begin{cases} k_m & if\ |k_m| > |k_n| \\ k_n & if\ |k_m| \leq |k_n| \end{cases} \tag{47}$$

Figure 15(b) illustrates the relation between ϖ and (k_m, k_n). If a point (k_m, k_n) falls in the yellow area, $\varpi = k_n$. If it falls in the cyan area, $\varpi = k_n$. The dissimilarity measure $\varpi(\eta(\alpha_1),\eta(\alpha_2))$ has the following properties:

1. $\varpi(\eta(\alpha_1),\eta(\alpha_2))=0$ iff the two shapes $\eta(\alpha_1)$ and $\eta(\alpha_2)$ are identical except for a depth reversal. This happens when $|\alpha_1|=|\alpha_2|$ (the slants of the symmetry planes of $\eta(\alpha_1)$ and $\eta(\alpha_2)$ are the same).
2. $\varpi(\eta(\alpha_1),\eta(\alpha_2)) = -\varpi(\eta(\alpha_2),\eta(\alpha_1))$
3. $\varpi(\eta(\alpha_1),\eta(\alpha_2)) = \varpi(\eta(90-\alpha_1),\eta(90-\alpha_2))$
4. $\varpi(\eta_{(l,\Xi)}(\alpha_1),\eta_{(l,\Xi)}(\alpha_2)) = \varpi(\eta_{(l',\Xi')}(\alpha_1),\eta_{(l',\Xi')}(\alpha_2))$. This property states that given any image and/or any symmetric correspondences, if the α's of two recoveries are known, the dissimilarity is known. In other words, the dissimilarity is independent of shape.

It is customary in studies of shape to express shape differences using aspect ratios, rather than the log of aspect ratio. The conversion of ϖ to aspect ratio can be done as follows:

$$\varepsilon(\eta_1,\eta_2) = |2^{\varpi(\eta1,\eta2)}-1| \qquad (48)$$

Now that we have a formal way to measure the dissimilarity between two 3D shapes, we can evaluate the accuracy of 3D shape recovery by the model and by the subject. We begin with testing the model.

Simulation

We randomly generated 3D abstract polyhedrons like those shown in Figure 16. Every polyhedron had 16 vertices. Their positions were randomly-generated in 3D space with the following constraints:

1. The object had planar faces;
2. The "front" part of the object was a box that was smaller than the box in the "back" (refer to Figure 16(a) for the illustration) and these boxes had a pair of coplanar faces;

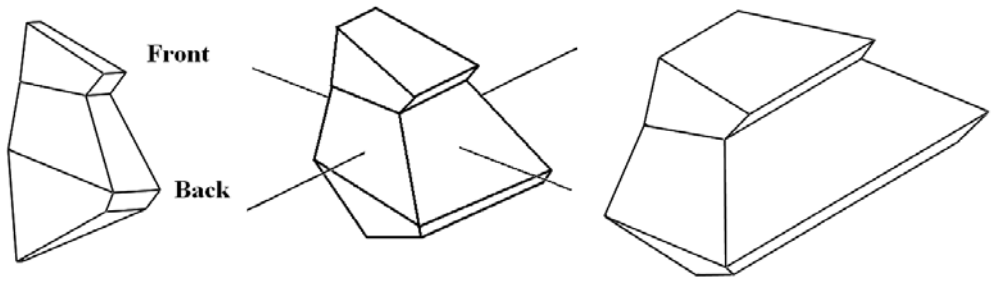


Figure 16. An illustration of polyhedra used for the simulation. The aspect ratio for each shape (from left to right) is: 1/3, 1 and 3.

3. The aspect ratio of the polyhedron varied between 1/5 and 5. The aspect ratio of a polyhedron was defined as the ratio of the thickness along two directions – one was the normal of its symmetry plane and the other was the normal of the coplanar face of the two boxes (see Figure 16(b)). The aspect ratios of the three polyhedrons in Figure 16 are 1/3, 1 and 3, respectively;
4. The overall size of the generated polyhedron was increased or decreased to fit a cube whose edge length was 10cm. In other words, the maximum length along X, Y or Z axis was 10 cm. The directions of X, Y and Z were defined relative to the observer: Z represents the direction in depth, the X axis is horizontal and the Y axis is vertical.

The slants of the symmetry plane ranged from 5 to 84 degrees with a step of one deg. Each slant was used 10 times. Totally, 800 polyhedra were generated. We recovered the 3D shapes using the monocular model. The averaged dissimilarity (ϖ) between the recovered and the original 3D shape for each slant was computed. The result is shown in Figure 17 (the blue curve). The abscissa represents the slant. The ordinate on the left represents the dissimilarity (ϖ) and that on the right represents the corresponding error in aspect ratio (ε). The simulation result shows that the model's performance is affected by the slant of the original shape. Specifically, the performance curve is an inverted U shape. When the slant is close to 45 degrees, the performance is better and the recoveries by the model are close to the original 3D shapes. When the slant is close to zero or close to 90 degrees (degenerate views), the performance is quite poor. We will call this relation the "slant effect."

Next, we applied our binocular model for three viewing distances: 50cm, 200cm or 800cm. The performance curves for different viewing distances are shown in Figure 17. Similarly to the monocular performance, the binocular

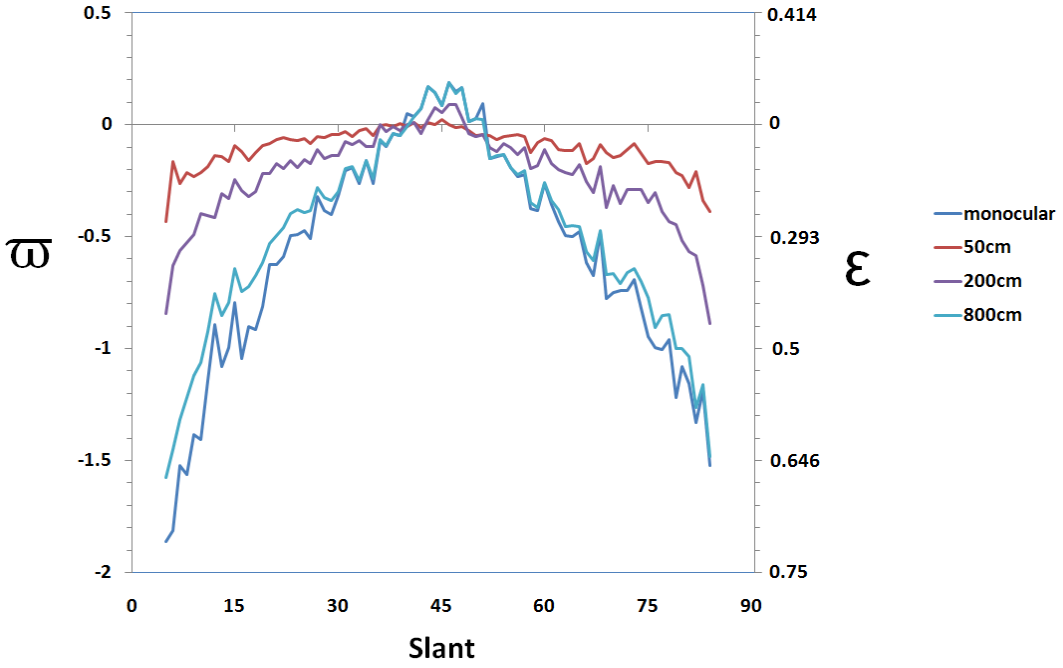


Figure 17. The monocular and binocular performance simulated by our models. The abscissa represents the slant of the real 3D shape. The ordinate on the left represents the dissimilarity (ϖ) between the recovered 3D shape and the original 3D shape and that on the right represents the corresponding error of aspect ratio (ε).

performance shows the slant effect, except that the magnitude of the effect is generally smaller. The magnitude of the slant effect increases when the viewing distance increases. Note that for the close viewing distance (50cm) the slant effect is extremely small, which means that the 3D shape recovery is always very close to veridical.

To summarize, in this section, we first presented a monocular model that is based on four simplicity constrains (symmetry, planarity, maximum compactness and minimum surface area). In the simulations, abstract polyhedrons, which include 16 vertices, were used. The result shows that when the slant of the symmetry plane of the polyhedron is close to 45 deg, the recovery is close to perfect. In the case of degenerate views (slant close to 0 or 90 deg), the monocular performance is poor. However, the model only made errors in one of the 15 parameters that characterize the 3D shape. This allows us to say that the 3D shape recovered by the model is always quite accurate. In the binocular model, which uses the information about depth order between points provided by stereoacuity, the 3D recovery is substantially better. Simulation results show that the binocular performance has a similar pattern as the monocular performance, which will lead to the continuity of percept when the viewing distance is changing. In a sense, the binocular model bridges monocular and binocular 3D shape perception. It is interesting to point out that a combination of the symmetry constraint (which is a non-metric constraint) and depth order information (which is also non-metric) produces a very good approximation to the metric shape.

Next, two psychophysical experiments are presented. These experiments measured human performance in 3D shape recovery. This performance was compared to the performance of our models.

PSYCHOPHYSICAL EXPERIMENTS

Experiment 1: Human's Performance in a 3D Shape Recovery Task – Fixed Depth, Varying Viewing Directions

In this experiment we tested the subject's binocular and monocular recovery of 3D shape. We also tested the subject's 3D shape recovery from motion parallax. In this condition, the images for the left and right eyes from the binocular test were presented successively to one eye.

Subject

Four subjects (MY, YL, ZP, ZS) participated this experiment. All subjects had normal or correct-to-normal vision. MY and ZS were naïve about the purpose of the experiment.

Stimuli

The polyhedral shapes were generated the same way as in the simulation experiment. Five slants of the symmetry plane of the polyhedra were used: 15, 30, 45, 60 or 75 degrees. Abstract shapes, rather than shapes of common objects, like chairs, couches or animal bodies, were used to make it possible to compare our model's performance with the performance of human observers. Human observers must be tested with abstract shapes to avoid familiarity confounds (Pizlo & Stevenson, 1999; Chan et al., 2006). Obviously, our model, which has no provision for "learning", is not subject to this problem. For the model all stimuli are novel, including those familiar to humans. Common objects could be used with the model, but this would make it impossible to compare human and the model performance.

Using OpenGL and shutter glasses, we set up a virtual reality system. Two slightly different images (stereoscopic images) for the subject's left and

right eyes were generated according to the subject's interocular distance. They were presented on a CRT display at the same rate as the refresh rate of the display. The subject viewed these images through shutter glasses that were synchronized with the display so that each eye received the image designed for this eye, only. For example, when the left glass was transparent and the right glass was opaque, the image for the left eye was shown. And when the right glass was transparent and the left glass was opaque, the image for the right eye was shown. When the refresh rate is high, the subject cannot perceive flicker and the two images are fused. In the experiment, the refresh rate was set to 100Hz. Thus, the image for each eye was updated at the rate of 50Hz.The size of the computer screen was 40 cm by 30 cm and the resolution was 1280 pixels by 1024 pixels.

The simulated viewing distance (i.e. the distance between the subject's eye and the center of the simulated polyhedron) and the distance between the subject and the monitor were the same and equal to 50cm.

Procedure

The subject viewed the images through the shutter glasses in a dark room. His/her head was supported by a chin-forehead rest. Two polyhedra were presented side by side and the separation was 13.3cm (see Figure 18). They were at the same level as the subject's eyes. On the left, the reference shape was presented under three viewing conditions:

1. Binocular viewing. The stereoscopic images of a stationary polyhedron were presented;
2. Monocular viewing. Subjects viewed one image of a stationary polyhedron monocularly. Specifically, when the left glass was transparent, the image for the left eye was presented. When the right glass was transparent, no image was presented and the screen was black.

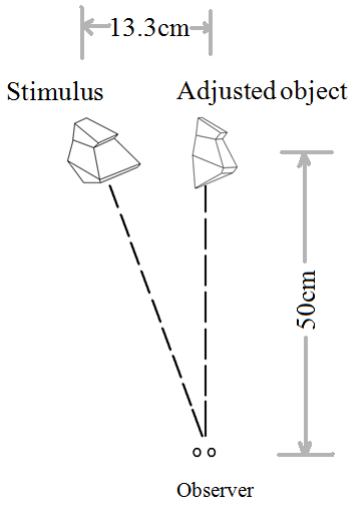


Figure 18. The illustration of the experimental setup. Two polyhedral shapes were presented side by side and the separation was 13.3cm. The simulated viewing distance was 50cm for both shapes.

3. Motion parallax. The images for the left and right eyes from the binocular viewing were presented alternately to the subjects' left eye at a rate of 2.5Hz.

On the right, a rotating 3D polyhedron was shown and subjects viewed it monocularly. This rotating polyhedron was selected from the family of symmetrical 3D shapes generated by our model (refer to equations (15) or (16)). Relative to the reference 3D shape, the orientation of the adjusted 3D polyhedron was first changed by 45 deg around the X axis first. The resulting polyhedron was then rotated around the Y axis at 80 degrees/second. This essentially guaranteed that none of the 2D images of the rotating polyhedron were identical to the images of the reference polyhedron (the viewing directions of the reference and the adjusted 3D shapes were different). This minimized the possibility that the subject used 2D features to produce a correct response. Viewing a rotating 3D polyhedron allowed many different views of the 3D shape to be seen in a short amount of time. The subject used a mouse to adjust the only parameter (α) to change the aspect ratio of the 3D shape until the adjusted 3D shape matched the percept of the 3D shape produced by the polyhedron shown on the left. To further minimize the effect of 2D artifactual cues, the average size of the 3D polyhedron on the right was 70% of the one shown on the left. At the start of each trial, α was set to a random value. There was no time limit for the adjustment.

There were two sessions for each condition and each session included 50 randomly generated polyhedra. Each slant (15, 30, 45, 60 or 75 degrees) was used 10 times in each session. On average, each session took about 20 minutes.

Results

For each trial, we computed the dissimilarity between the subjects' adjusted 3D shape and the original 3D shape. First, we computed the average dissimilarity for each condition (see Figure 19(a)). The binocular performance

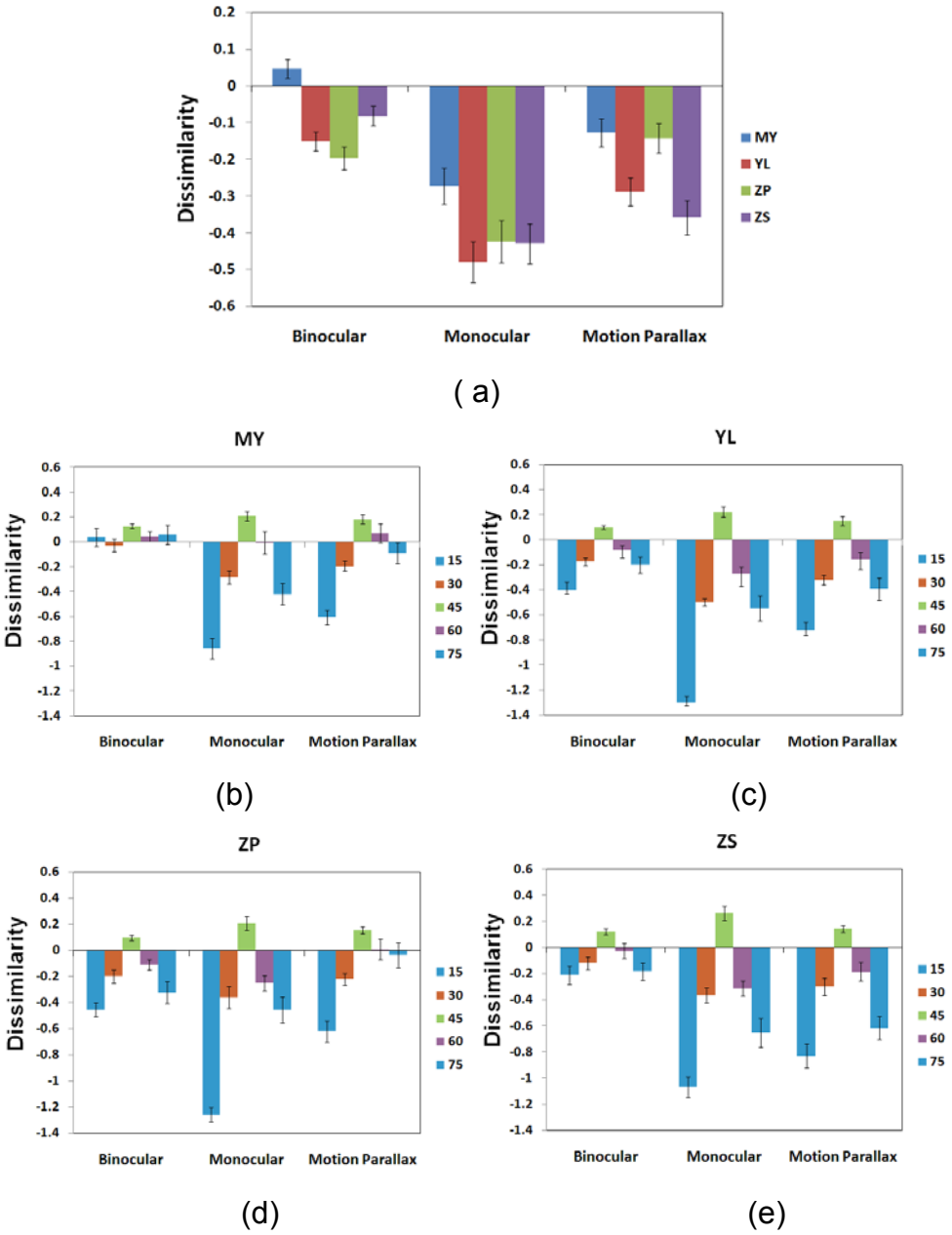


Figure 19. The dissimilarity between subjects' adjusted 3D shapes and the reference 3D shapes. (a) The subjects' average performance across the three viewing conditions. (b) (c) (d) and (e) individual subjects' performance for the three viewing conditions and for different slants.

was the best, the motion parallax performance was second best and the monocular performance was the worst. This pattern was consistent across all four subjects. The average dissimilarity between the adjusted 3D shape and the original 3D shape for the three viewing conditions were -0.102, -0.397 and -0.229. The corresponding perceptual errors (refer to equation (48)) were 7%, 26% and 15%.

Next, we plotted the subjects' performance for each slant. Figures 18(b)-(e) show the four subjects' performance. The pattern of results is quite similar across the subjects. Interestingly, performance of the two naïve subjects (MY, ZS) is not worse (and may be better) than that of the other two subjects. For each viewing condition, the performance tended to be best when the slant of the symmetry plane was 45 degree - the dissimilairty was close to 0 and the standard error was smallest. When the slant was far from 45 degrees, the performance tended to be worse. Note the inverted U pattern of results in all viewing conditions and all subjects (except binocular viewing of MY). This pattern is similar to that observed in the model simulations. In a control experiment, YL and ZP ran the binocular session for an additional three viewing distances: 100, 200 and 300 cm. On average, their perceptual error in recovering aspect ratio of a 3D shape increased from 11% at 50cm distance to 16% at 300cm distance. This moderate increase in recovery error is consistent with performance of the model shown in Figure 16.

Next, we evaluate the model's performance in binocular and monocular conditions using the same stimuli that the subjects viewed.

Stimulation

For each trial, a 3D shape was recovered by our model and it was compared with the reference 3D shape. Because we do not have a separate recovery model for the motion parallax condition, we only simulated the subjects' monocular and binocular performance (see Figure 20). Computationally, our binocular model could also be used as a model for motion parallax condition with two images. The model's binocular performance

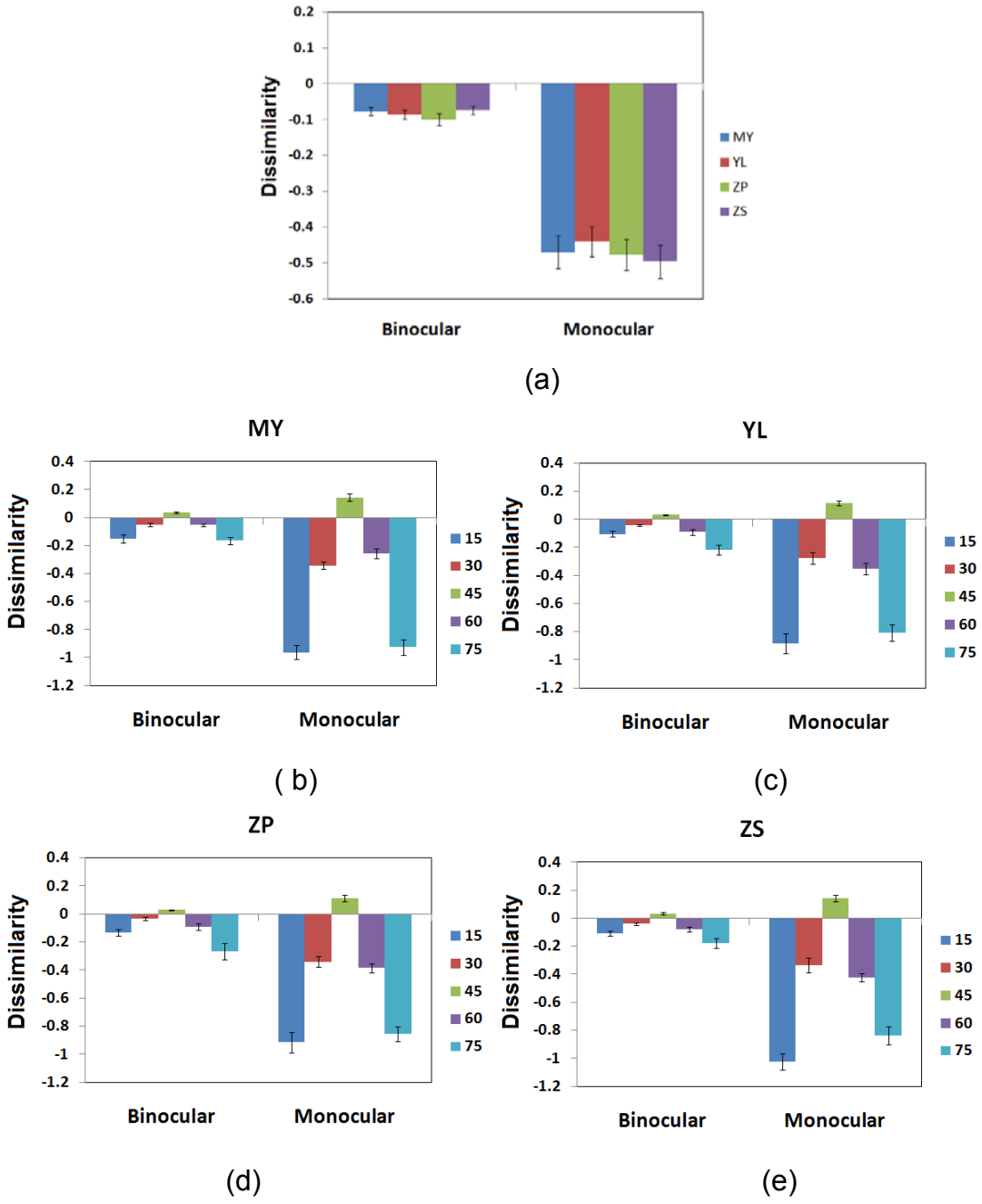


Figure 20. The performance simulated by our models. (a) The performance averaged across different slants. (b)-(e) The performance for each slant.

improves due to the depth order information provided by stereoacuity. Motion parallax provides equally good depth order information, so the benefit of adding the second image is expected to be similar in binocular and motion parallax conditions. This was roughly the case in our psychophysical experiment (see Figure 19).

The monocular recovery by the model was very similar to the psychophysical results (compare the graphs in Figures 17 and 18). The same was true for binocular recovery. However, the binocular performance by our model was slightly better than the subjects' performance. This is not surprising considering that our model simulated only one type of visual noise that is present in the human visual system, namely the noise represented by stereoacuity threshold. There are other sources of noise. For example, the human visual system has a limited discrimination ability in judging 3D aspect ratios. If we use a 3% Weber fraction for line length discrimination on the frontal plane (the Weber fraction is surely larger in 3D space) then comparing two aspect ratios will lead to threshold of at least $\sqrt{2}$ times 3%. This additional noise alone may be able to account for the difference between the subject's and the model's binocular performance. There are, however, other sources of variability. Our binocular model uses all pairs of points whose depth order can be judged. There are up to N^2 such pairs, for N visible vertices. It is very likely that due to the limitations of visual attention, the human observer uses only a subset of these pairs of points. Finally, note that when we computed the threshold of stereoacuity, we adopted Rady & Ishak's (1955) results. In their experiments, they measured the stereoacuity between points without any context. There have been reports that when an arbitrarily curved surface is used as context, the stereoacuity threshold increases (Norman & Todd, 1998). It is an open question as to whether stereoacuity is affected by the presence of a symmetrical shape.

The comparison between the subjects' perceived 3D shapes and the recovered 3D shapes by the models is shown in Figure 21. The average

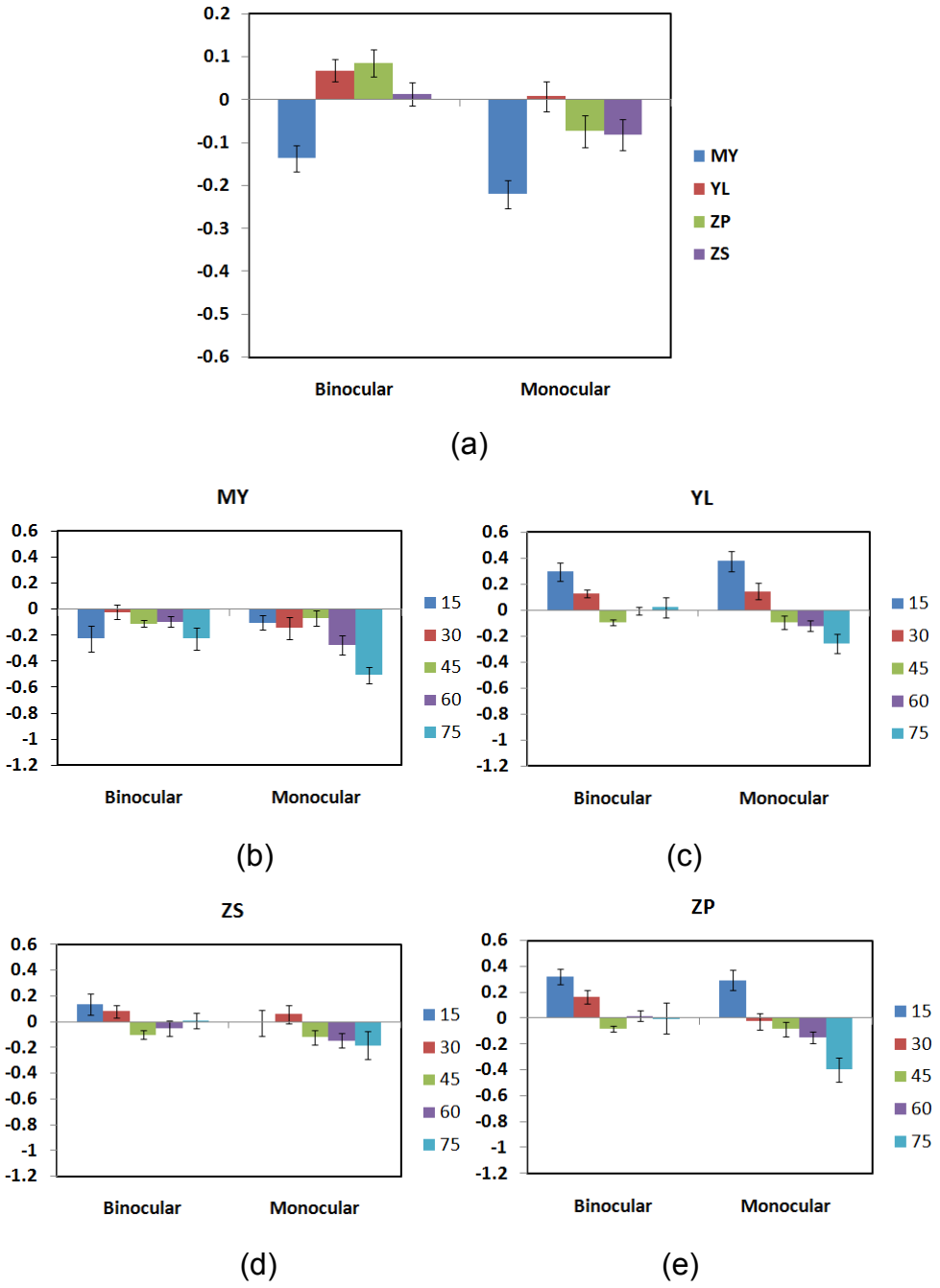


Figure 21. The dissimilarity between the subjects' perceived 3D shapes and the recovered 3D shapes by our models. (a) The dissimilarity averaged across different slants. (b)-(e) The comparison for different slants.

dissimilarity between subjects' percept and the recovery by the models was close to 0 for three of the four subjects (see Figure 21(a)). More importantly, the dissimilarities between the perceived shapes and recovered shapes are almost always smaller than the dissimilarities between the perceived shapes and the original shapes. This is especially clear in the case of monocular viewing (in binocular viewing both the subjects and the model produced close to perfect recovery). This means that our models can indeed explain to some degree the 3D shape percept. The same conclusion is generally true in the case of the comparison for individual slants (Figure 21 (b)-(e)). Only in a few cases the errors are not close to zero. For example, when the slant of the symmetry plane is 15 deg (the lateral surfaces of a symmetric 3D shape are facing the subject), the dissimilarity between the percept and the recoveries for YL and ZP (the non-naïve subjects) is greater than 0 for both monocular and binocular viewing conditions. This suggests that these subjects perceived more depth compression along the normal of the symmetry plane than the models. The fact that the other two subjects showed smaller dissimilarities, suggests that large differences between the model and the subject could be due to idiosyncratic factors, whose explanation is beyond the scope of this dissertation.

Discussion

Our psychophysical results show that the slant effect existed for all three viewing conditions, which implies that 3D shape perception in monocular vision, binocular vision and motion parallax probably involves the same mechanism. Performance of our models when they were applied to the same images was quite similar to the subjects' performance. This fact suggests that our visual system and our models use the same method to recover 3D shapes: (1) subjects and the models perceive the 3D shape with maximum V/S^3; (2) binocular vision helps determine the depth order between points and this extra information limits the size of the set of possible 3D interpretation and consequently improves performance.

Our psychophysical results show that the performance in the motion parallax condition was better than that in the monocular viewing condition. This suggests that the second image provides some extra information that limits the size of the set of possible 3D interpretations, like what binocular disparity does.

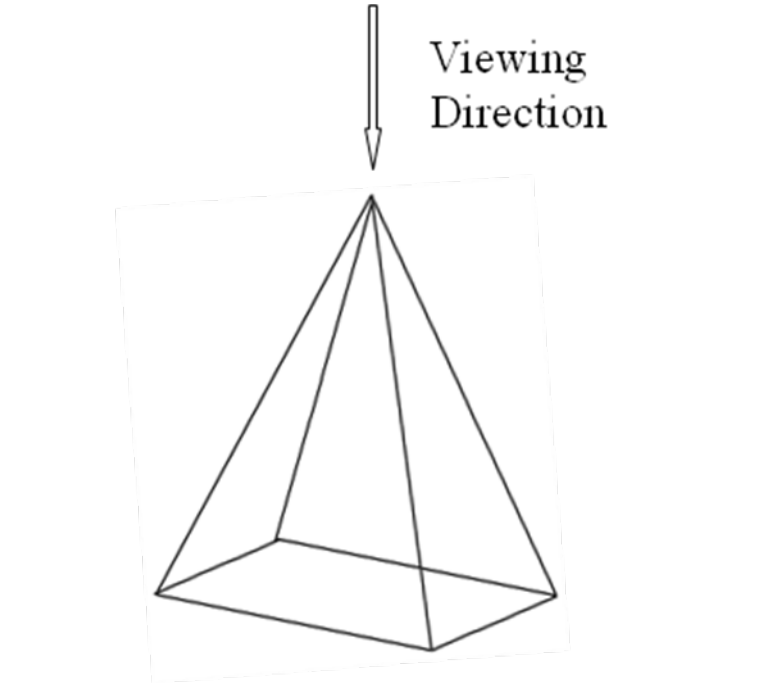
To summarize, the subjects' performance was similar to the prediction of our models. In addition, our psychophysical results suggest binocular shape perception is almost veridical. This result is different from the conclusions of many other studies. That used either degenerate viewing conditions or degenerate objects. In one of the most representative studies, Todd & Norman (2003) had the subject view two pyramids at different distances. Both pyramids were viewed from their top (Figure 22(a) illustrates the viewing direction and Figure 22(b) shows the stereoscopic images). The subject was asked to adjust the height of one pyramid to match the aspect ratios of the two pyramids. They found that the subject's percept was not accurate. Specifically, when the adjusted pyramid was farther than the reference, the adjusted height was systematically greater than veridical height. This result represents compression of the perceived depth. The perceptual error was up to 25%. Based on this result, the authors concluded that the binocular 3D shape perception is not veridical.

Clearly, Todd & Norman's conclusion about binocular shape perception is quite different from ours. We repeated Todd et al.'s experiment. We also repeated our Experiment 1 under slightly modified viewing conditions, to make the comparison as direct as possible.

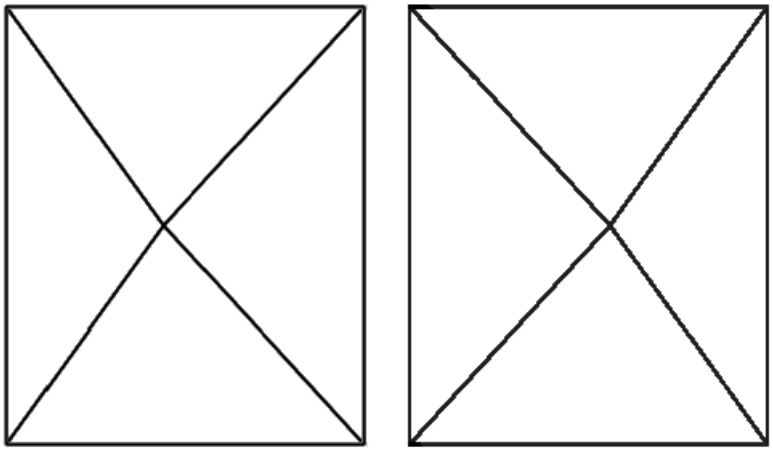
Experiment 2: Human's Performance in 3D Shape Recovery Task: Different Depths, Same Viewing Directions

Subjects

Four subjects (YL, ZP, ZS, ZX) participated in this experiment. All subjects had normal or corrected-to-normal vision. ZS and ZX were naïve about the purpose of the experiment.



(a)



(b)

Figure 22. The illustration of a pyramid used in experiment 2. (a) The pyramid was viewed from its top. (b) The stereogram (for crossed fusion) of the pyramid when it was viewed from the top.

Stimuli

The polyhedra were generated the same way as in Experiment 1. In addition, pyramids similar to those Todd & Norman (2003) used were generated (see Figure 22). The bottoms of the pyramids were rectangular and they were perpendicular to the visual axis. The size of each pyramid was scaled so that it fit inside a cube whose edge length was 10cm. 100 pyramids with random shapes were generated.

Procedure

Subjects viewed the image through the shutter glasses in a dark room and their heads were supported by a chin-forehead rest. The distance between the display and the observer was 100cm. Two 3D shapes –the adjusted and the reference shapes were displayed side by side. The adjusted 3D shape was right in front of the observer and the viewing distance was 125cm. The reference 3D shape was on the left. Relative to the adjusted 3D shape, it was moved 13.3cm to the left and 50cm closer to the observer. Finally, it was rotated 10.06 degrees around the Y axis, which was the angle formed by the centers of the two shapes and the observer's cyclopean eye (see Figure 23). This rotation guaranteed that the reference and the adjusted 3D shapes were viewed from the same direction. These viewing conditions (same 3D viewing orientation and different viewing distances) resembled closely the viewing conditions used in prior studies of binocular depth and shape constancy.

There were four viewing conditions for the reference 3D shape on the left. The first three were the same as the conditions in Experiment 1: (1) a polyhedron was viewed binocularly; (2) a polyhedron was viewed monocularly; and (3) the two images from the binocular condition were presented alternately to the subjects' left eye (motion parallax). The slant of the symmetry plane was 15, 30, 45, 60 or 75 deg. In the fourth condition, the pyramid was viewed from its top binocularly. That is, the viewing direction coincided with the height of the pyramid (see Figure 23).

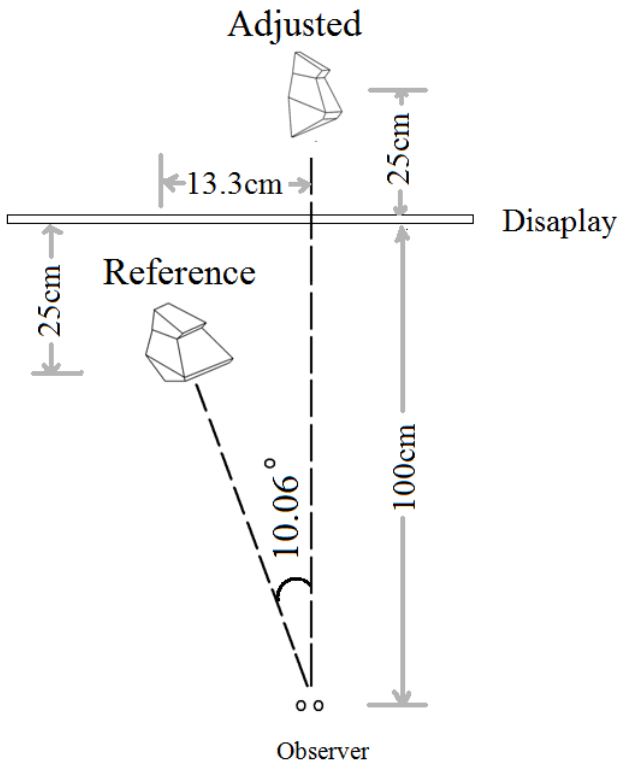


Figure 23. A schematic diagram of the viewing configuration used in Experiment 2. The distance between the display and an observer was 100cm. The adjusted object was right in front of the observer and the viewing distance was 125 cm. The reference 3D shape was on the left. Related to the adjusted 3D shape, it was moved 13.3cm to the left and 50cm closer to the observer. The angular separation between the reference and the adjusted 3D shapes was 10.06 degrees.

The adjusted 3D shape on the right was stationary and it was viewed binocularly. In the case of the first three viewing conditions (binocular, monocular and motion parallax), the task was similar to that in Experiment 1. Specifically, the subject used a mouse to adjust the parameter (α) to change the 3D shape until the adjusted 3D shape matched the reference 3D shape on the left. In the beginning of each trial α was set to a random value. In the fourth viewing condition, the subject's task was identical to that in Todd & Norman (2003) study. Namely, the subjects were asked to change the height of the adjusted pyramid until its aspect ratio matched the aspect ratio of the reference pyramid on the left. Note that because the pyramids were viewed from a degenerate viewing direction, the new binocular model described above cannot recover its shape. The reason is that the one-parameter family of symmetric pyramids contains the pyramids that are different from one another by a pure stretch along the depth direction. This stretch does change the depth order of any pairs of points. It follows that the subject in this task will be forced to use depth perception in matching 3D shapes. Depth matching is not needed in any of the other three conditions.

There were two sessions for each condition and each session included 50 randomly generated polyhedral or pyramids. For the binocular, monocular and motion parallax conditions, the slant of the reference 3D shape was 15, 30, 45, 60 or 75 degrees and each was used 10 times in each session.

Results

The average performance for each viewing condition is shown in Figure 24(a). The binocular performance with polyhedra was the best. The binocular performance with pyramids and monocular performance with polyhedra were the worst. The average dissimilarity between the adjusted 3D shapes and the reference 3D shapes were 0.085, -0.346, -0.043 and 0.365 for binocular, monocular, motion parallax and pyramid, respectively. When the dissimilarity is expressed as percentage error in adjusting aspect ratios, the average errors for the four conditions were 6%, 21%, 3% and 29%. Note that this is a systematic

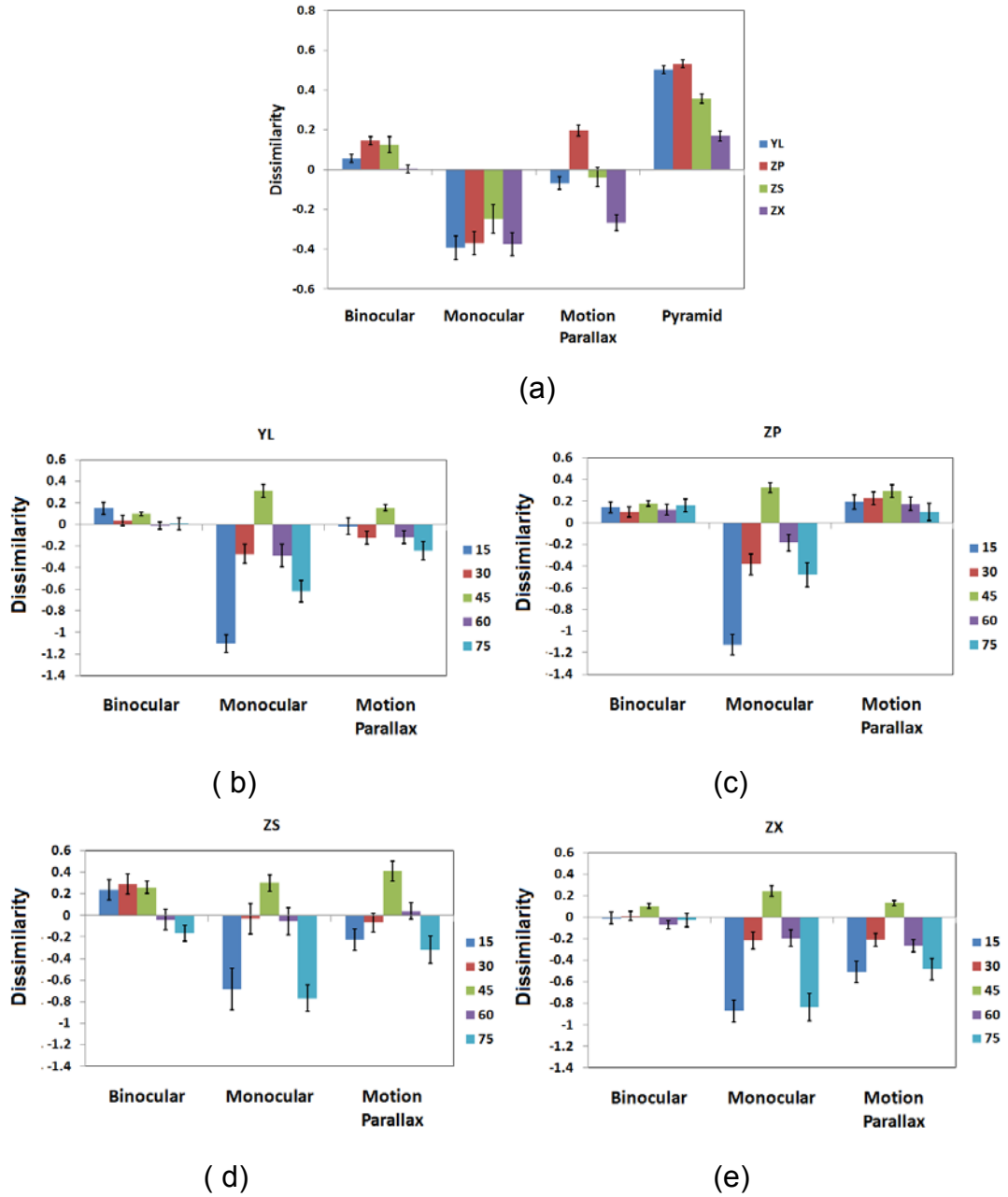


Figure 24. The subjects' performance for the four viewing conditions. (a) The average performance across the slants. (b)-(e) the subjects' performance for each slant in the three polyhedral conditions.

error. The large errors in monocular performance with polyhedron (21%) and binocular performance with pyramids (29%) mean that the 3D shape percept for these two conditions was substantially biased. Note that binocular viewing of a pyramid from a degenerate view led to worse performance than monocular viewing of a polyhedron from a non-degenerate view. This illustrates that binocular depth perception is quite unreliable. This contrasts with the nearly perfect performance in shape perception of a polyhedron under motion parallax and binocular viewing, where the errors were close to 0 (3% and 6%). The average standard deviation across the subjects for the motion parallax condition (0.189) was substantially higher than that for the binocular viewing condition (0.066), which suggests that binocular performance was overall better than motion parallax performance.

In the case of the monocular and binocular conditions, the magnitude of the systematic error was consistent across the four subjects. However, the individual variability was quite large for the motion parallax and pyramid conditions. Similar variability for the case of the pyramid condition was reported by Todd and Norman (2005).

Next, we plotted the subjects' performance for the three polyhedron conditions, as a function of the slant of the symmetry plane (see Figure 24(b)-(e)). Similarly to Experiment 1, monocular performance was quite good for slants 30-60 deg. For slants 15 and 75 deg performance was clearly worse. In the case of motion parallax, we could also observe the slant effect. However, the effect was much smaller than that in Experiment 1. In the case of the binocular performance of the polyhedra, the slant effect disappeared completely. The disappearance of the slant effect in Experiment 2 is not difficult to explain. Note that because the adjusted and reference 3D shapes were viewed binocularly from exactly the same viewing directions, the percept for the adjusted 3D shape should show the same slant effect as the percept of the reference 3D shape (the effect was not exactly the same because the viewing distances of the reference and adjusted 3D shapes were not the same; with

different viewing distances, the depth order matrices were slightly different). As a result, the two biases cancelled out. The slant effect for the monocular and motion parallax conditions are also expected to be smaller: the slant effect measured in Experiment 2 should be equal to the slant effect from Experiment 1 minus the binocular slant effect.

We conclude by pointing out that it is better to measure the shape percept by asking the subject to view the 3D shape from different viewing directions, as was done in Experiment 1. Using identical viewing directions is less than ideal. In fact, the only way to evaluate the degree of shape constancy is to use different viewing directions. Unfortunately, this is not how the 3D shape was studied in the past. Next we will simulate the processes for binocular and monocular performance.

Stimulation

Suppose that the reference 3D shape is η_R and the 3D shape recovered by the binocular model is η_R' when η_R is viewed from the distance of 75cm (the distance of the reference 3D shape in Experiment 2). Suppose next that the adjusted 3D shape is η_A and the 3D shape recovered by the binocular model is η_A' when η_A is viewed from the distance of 125cm (the distance of the adjusted 3D shape). If η_R' is the same as η_A', η_A and η_R will be perceived by the model as the same 3D shape.

Further, suppose that the set of all 3D shapes in the one-parameter family produced by the 2D cyclopean image of the 3D adjusted shape is ψ. For each 3D shape η_i in ψ, its recovered η_i' is computed and compared with η_R'. From equation (32), it is known that there is a range of η_is, whose recoveries are all identical to η_R'. In other words, small changes of the aspect ratio of the adjusted 3D shape will not change the 3D percept. Because these adjusted 3D shapes are perceptually equivalent, which means that they are perceived as the same 3D shape as that when η_R is viewed binocularly, we randomly chose one 3D shape among them to simulate the subjects' adjustment.

For the monocular performance, the simulation is done similarly. Suppose the reference 3D shape is η_M and the 3D shape recovered by the monocular model is η_M'. Then in the set of 3D shapes for the adjustment, we obtain the subset that are perceived as η_M'. In this subset, we chose one randomly to simulate the subjects' adjustment.

The dissimilarity between the 3D shapes adjusted by the model and the real 3D shapes was computed. Figure 25(a) shows the performance averaged across the five slants. Since the model was applied to the images that were actually used in the psychophysical experiments, the model's performance is slightly different across the four subjects simply because different subjects were tested with different, randomly generated stimuli. The pattern and magnitude of errors are similar to the subjects' performance shown in Figure 24. Figure 25(b)-(e) shows the simulation results for different slants. The slant effect for the binocular condition disappeared as was the case in the psychophysical results shown in Figure 25.

Discussion

In the psychophysical experiment, we repeated Todd & Norman's (2003) experiment with pyramids viewed from degenerate viewing directions and we replicated their results – the percept of the pyramid viewed from a degenerate viewing direction was not accurate and when the adjusted 3D pyramid was farther than the reference pyramid, the adjusted height was larger than the height of the reference pyramid, suggesting depth compression.

However, binocular performance with polyhedral objects viewed from non-degenerate viewing directions was much better. The average perceptual errors were less than 10%. What is responsible for the difference? We think that there are two factors:

1. In the previous studies of 3D shape perception, the subject was usually asked to view the 3D shape from a degenerate view. For example, the subjects in Todd & Norman's

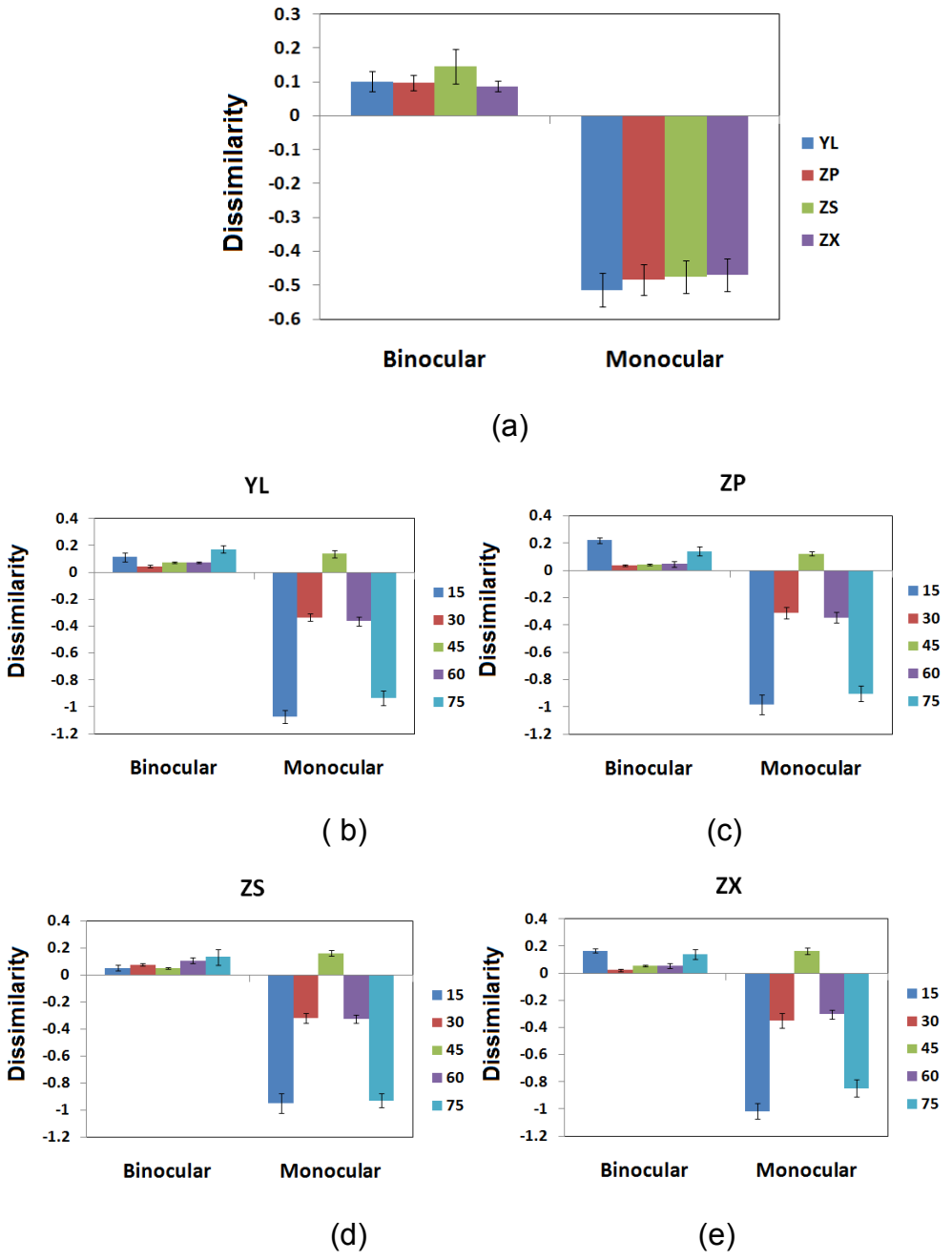


Figure 25. (a) Performance of the model averaged across the five slants. (b)-(e) the binocular and monocular performance for different slants.

experiment viewed the pyramid from its top. Our simulation results showed that binocular performance was worst when the slant of the symmetry plane was close to 0 or 90 degrees (see Figure 17). Note the slants of both symmetry planes of the pyramids used in Todd & Norman's (2003) experiment were 90 degree. For the degenerate view, the possible 3D interpretations were those produced by stretching the original 3D shape in depth. All pyramids in this family have the same depth order. Therefore, the contribution of stereoacuity is very limited, if any. The subjects have to make their judgment based on the distance between points in the depth direction. It is known that human's depth perception from binocular disparity is quite poor. Therefore, perceiving a 3D shape from its degenerate view cannot be veridical.

2. The objects used in previous studies of 3D shape perception were quite simple, like a cylinder, or a pyramid. According to our model, a simple 3D shape has a small depth order matrix, which suggests fewer limitations on the perceived 3D shape. Usually, small depth order matrix will lead to a large set of possible 3D interpretations and consequently will result in poor performance.

These two factors suggest that the following two conditions should be satisfied in order to perceive a 3D shape veridically – (1) the 3D shape should be complex and (2) it should be viewed from a non-degenerate view. These two conditions are almost always satisfied in everyday life.

SUMMARY AND CONCLUSIONS

In this dissertation, we presented two models to simulate human monocular and binocular 3D shape perception. The monocular model is based on simplicity constraints – symmetry, planarity, maximum compactness and minimum surface area. A recovery by the model usually includes three steps. Given a 2D image, we first apply the symmetry constraint to limit its possible 3D interpretation to a set that is determined by one free parameter. Then by applying the planarity constraint, we compute the hidden part of those 3D shapes. Finally, we combine the maximum compactness and minimum surface area constraints and choose a unique 3D shape in the set. This last step is one of the original contributions of our work: no one used these two constraints in 3D shape recovery before. Why should these two constraints be combined and why does the combination lead to a good recovery? We can answer this question from two perspectives.

1. Koffka (1935) pointed out that human 3D shape perception is a compromise between a simplicity principle and the 2D image. He interpreted the tendency toward the simplest possible shape as an external force and the retinal stimulus pattern an internal force. Recently, Griffiths & Zaidi (2000) used this idea to suggest that the perceived tilt of a 3D surface is a compromise between the 2D image of the surface and perceptual assumptions (a priori constraints). Our model provides a more direct application of Koffka's two-force mechanism to 3D shape. Namely, the maximum

compactness constraint represents the external force implementing the 3D simplicity constraint, and the minimum surface constraint represents the internal force. The minimum surface area can be viewed as the internal force because the lower bound on the surface area of a 3D shape is the surface area of the 2D image itself.

2. The combination of these two constraints is similar to the Bayesian process. Many scientists consider human vision as an information processing system, whose output is determined by the input (the 2D image in the case of 3D shape percept) and the built-in structure (the priors). A Bayesian model uses a combination of the prior and the likelihood to produce an optimal output (decision). The minimum surface area constraint, which leads to the most likely 3D interpretation, represents the likelihood function and the maximum compactness constraint represents the prior distribution function. Therefore, the combination of these two constraints is similar to the Bayesian process in which a maximum posteriori estimate is taken as an optimal estimation (Knill & Richards, 1996).

Consider now the binocular model. Previous studies of binocular depth perception showed that binocular vision is poor at judging the depth distances between points, but very good at determining their depth order(Blakemore 1970; Westheimer 1979; Westheimer & McKee, 1980). Combining the ordinal depth from binocular vision and the monocular shape recovery, we proposed a binocular model. When we view a 3D shape binocularly, the depth order information provided by stereoacuity limits the recovered 3D shape to a small set. Among the 3D shapes in the set, the model chooses the one that has the

maximum V/S^3 (the combination of maximum 3D compactness and minimum surface area constraints).

In Experiment 1, we measured the subject's performance and compared it with the performance of our model. We found that the subject's results were consistent with the prediction of our models: (1) both subjects' monocular and binocular performance show a slant effect; (2) the binocular 3D shape perception is close to veridical. The second of these two results is new; it is different from the results of most of the previous studies, in which the binocular 3D shape perception was reported to be highly inaccurate. Our further study (Experiment 2) revealed that the poor binocular performance reported in the past was due to the following two factors (1) in the prior studies the subjects were shown degenerate views of the 3D shape and/or (2) the objects were too simple. The difference between our results and the results of previous studies can be well accounted for by our model.

The consistency between the model performance and the subject's performance shows that our models of 3D shape perception are psychologically plausible. It suggests that the simplicity constraints play an important role in 3D shape perception and it also reveals the role of binocular vision in 3D shape perception.

Our models are not the first attempt to recover a 3D shape from 2D image(s). For example, a triangulation method is often used to recover a 3D shape from its several views (Hartley & Zisserman, 2003). The reliability of this method, however, strongly depends on the accuracy and precision with which the positions of points in images are measured. Chan et al.'s experiments (1999) showed that even a very small amount of noise in the images lead to large errors in the recovery. Such instability is absent in human 3D shape perception and in our binocular model. This means that the triangulation method is not as good as our new model, and furthermore that the triangulation method is not a good model of human shape perception.

Note that our current models have several limitations. Before they become general models of human shape perception, several of its aspects have to be further developed:

How to Recover a 3D Shape From a 2D Perspective Image?

In our daily life, the retinal image is a perspective projection of the 3D world. Our model, however, uses orthographic images. Orthographic projection is a good approximation to perspective projection when the size of a 3D shape is very small compared to the viewing distance. Can we recover a symmetric 3D shape from a perspective image? The answer is yes. If (1) the center of perspective projection is known (i.e. the camera is calibrated) and (2) the vanishing point of the lines connecting the symmetric points can be reliably estimated, the 3D shape can be recovered uniquely (refer to the Appendix D). Because of noise in an image, usually the estimate of the vanishing point is not accurate, especially when the vanishing point is very far from the center of the image. However, in such cases, the perspective distortions are small and can be ignored. In other words, the orthographic approximation is likely to be good enough.

How to Detect the 3D Symmetry?

In this study, we assumed that all symmetry correspondences in an image have been established before applying our model. Sawada et al. (2009) proved that any two curves in an image can be a projection of a pair of mirror symmetric 3D curves. Obviously, our visual system does not interpret every image as representing a 3D symmetric configuration. We speculate that the planarity constraint plays an important role in symmetry detection. Only when a 3D curve is on a plane, the projections of this curve and its symmetric counterpart will be considered as the projections of symmetric curves by our visual system. From Appendix A, we know that the projections of these two curves are subject to an affine transformation. Therefore, we can detect the symmetry by checking whether they are related by an affine transformation.

How to Find 3D Shapes in an Image?

So far, we only recover one 3D shape at one time. However, usually an image includes many 3D shapes. How to find them is essential for shape recovery. The method of detecting symmetry proposed above probably provides one way to solve this figure-ground organization problem. We can compare all curves in an image and see whether they can be interpreted as symmetric and at the same time, we can obtain the direction of the lines connecting the symmetric pairs (i.e., the tilt of the symmetry plane). Those curves with the same tilt will be grouped as an object.

The three aspects briefly discussed above are important for the application of our models to real images of real scenes. A lot remains to be done, but the models presented in the dissertation are a good starting point.

LIST OF REFERENCES

LIST OF REFERENCES

Attneave, F., & Frost, R. (1969). The determination of perceived tridimensional orientation by minimum criteria. Perception & Psychophysics, 6, 391-396.

Blakemore. C. (1970). The range and scope of binocular depth discrimination in man. Journal of Physiology, 211, 599-622

Brady, M., & Yuille, A. (1983). Inferring 3D orientation from 2D contour (an extremum principle). In W. Richards (Ed.), Natural computation (pp. 99-106). Cambridge, MA: MIT Press.

Braunstein, M. L. (1968). Motion and texture as sources of slant information. Journal of Experimental Psychology, 78, 247-253.

Chan, M. W., Stevenson, A. K., Li, Y., & Pizlo, Z. (2006). Binocular shape constancy from novel views: The role of a priori constraints. Perception & Psychophysics, 68, 1124-1139.

Chan, M.W., Pizlo, Z. & Chelberg, D.M. (1999) Binocular shape reconstruction: psychological plausibility of the 8 point algorithm. Computer Vision & Image Understanding 74, 121-137.

Edelman, S. (1999). Representation and recognition in vision. Cambridge, MA: MIT Press.

Gibson, J. J. (1950). The perception of visual surfaces. The American Journal of Psychology, 163, 367-384

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton Mufflin.

Griffiths, A. F., & Zaidi, Q. (2000). Perceptual assumptions and projective distortions in a three-dimensional shape illusion. Perception, 29, 171-200.

Hartley, R., & Zisserman, A. (2003). Multiple view geometry in computer vision. New York: Cambridge University Press

Hillis, J. M., Watt, S. J., Landy, M. S. & Banks, M. S. (2004). Slant from texture and disparity cues: optimal cue combination. Journal of Vision, 4, 1-24.

Howard, I. P., & Rogers, R. J. (2002). See in depth. Toronto, Ontario, Canada: I Poerteous.

Huang, T. S., & Lee, C. H. (1989). Motion and structure from orthographic projections. IEEE Transactions on Pattern Analysis & Machine Intelligence, 11, 536-540.

Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. Journal of Neurophysiology, 26, 994-1002.

Hubel, D. H. (1995). Eye, brain and vision. New York: Scientific American.

Johnston, E. P. (1991). Systematic distortions of shape from stereopsis. Vision Research, 31, 1351-1360

Julesz, B. (1971). Foundations of cyclopean perception. Chicago: The University of Chicago Press

King, M., Glenn, E., Tangeney, J., & Biederman, I. (1976). Shape constancy and a perceptual bias towards symmetry. Perception & Psychophysics, 9, 129-136.

Knill, C. D., & Richards, W. (1996). Perception as Bayesian inference. New York: Cambridge University Press.

Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L. (1992). Surface perception in pictures. Perception & Psychophysics, 52, 487-496.

Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L. (1996). Pictorial surface attitude and local depth comparisons. Perception & Psychophysics, 58, 163-173.

Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L., & Todd, J. T. (2004). Pointing out of the picture. Perception, 33, 513-169.

Koffka, K. (1935). Principles of Gestalt psychology. New York: Harcourt, Brace.

Landy, M. S., Maloney, L. T., Johnston, E. B. & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Research, 35, 389-412.

Leclerc, G. Y., & Fischler, A. M. (1992). An optimization-based approach to the interpretation of single line drawings as 3D wire frames. International Journal of Computer Vision, 9, 113-136.

Li, Y., & Pizlo, Z. (2007). Reconstruction of shapes of 3D symmetric objects by using planarity and compactness constraints. Proceedings of IS&T/SPIE Conference on Vision Geometry, Vol. 6499.

Li, Y., & Pizlo, Z. (May, 2008). Perception of 3D shapes from line drawings. Poster session presented at the annual meeting of the Society for Vision Science Society, Naples, FL.

Li, Y., Pizlo, Z., & Steinman, M. R. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal representation. Vision Research,49, 979-991.

Maniatis, L. (2008). The physical logic of visual perception (Doctoral dissertation, The American University, 2008). Dissertation Abstracts International, 69(07B). Abstract retrieved February 23, 2009 from First Search/Dissertation Abstracts International databse.

Marill, T. (1991). Emulating the human interpretation of line-drawings as three dimensional objects. International Journal of Computer Vision 6, 147-161.

Marr, D. (1982). Vision. New York: W.H. Freeman.

Norman, J. F., & Todd, J. T. (1998). Stereoscopic discrimination of interval and ordinal depth relations on smooth surfaces and in empty space. Perception, 27, 257-272.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. Journal of Experimental Psychology: Human Perception and Performance, 22, 173-186.

Ogle, K. N. (1950). Researches in binocular vision. Philadelphia: W B Saunders.

Perkins, D. N. (1972). Visual discrimination between rectangular and nonrectangular paralleopipeds. Perception & Psychophysics, 12, 396-400.

Pizlo, Z. (2001). Perception viewed as an inverse problem. Vision Research, 41, 3145-3161.

Pizlo, Z. (2008). 3D shape: Its unique place in visual perception. Cambridge, MA: MIT Press.

Pizlo, Z., Li, Y., & Steinman, R. M. (2006, August). A new paradigm for 3D shape perception. European Conference on Visual Perception. St. Petersburg, Russia.

Pizlo, Z., Li, Y., & Steinman, R. M. (2008). Binocular disparity only comes into play when everything else fails: A finding with broader implications than one might suppose. Spatial Vision, 21, 495-508

Pizlo, Z., & Stevenson, A. K. (1999) Shape constancy from novel views. Perception & Psychophysics, 61, 1299-1307.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. Nature, 343, 314-319.

Polya, G., & Szego, G. (1951). Isoperimetric inequalities in mathematical physics. Princeton, NJ: Princeton University Press.

Rady, A. A., & Ishak, I. G. H. (1955). Relative contributions of disparity and convergence to stereoscopic acuity. Journal of the Optical Society of America, 45, 530-534.

Regan, D. (2000). Human perception of objects. Sunderland, MA: Sinauer.

Rothwell, C. A. (1995). Object recognition through invariant indexing. Oxford, England: Oxford University Press.

Sawada, T., Li, Y., & Pizlo, Z. (2009). Any 2D image is consistent with a 3D symmetric interpretation. Pattern Recognition Letters. Manuscript submitted for publication.

Sinha, P. (1995). Perceiving and recognizing three-dimensional forms. Dissertation Abstract International, 56(12), 6873B. (UMI No. AAT 0576742) Retrieved September 16, 2008, from Dissertations and Theses database.

Todd, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. Perception & Psychophysics, 36, 97-103

Todd, J. T., & Norman, J. F. (2003). The visual perception of 3D shape from multiple cues: Are observers capable of perceiving metric structure? Perception & Psychophysics, 65, 31-47.

Ullman, S. (1979). The interpretation of visual motion. Cambridge, MA: MIT Press.

Ullman, S. (1996). High-level vision. Cambridge, MA: MIT Press.

Vetter, T., & Poggio, T. (1994). Symmetric 3D objects are an easy case for 2D object recognition. Spatial Vision, 8, 443-453.

Westheimer, G. (1979). Cooperative neural processes involved in stereoscopic acuity. Experimental Brain Research, 36, 585-597

Westheimer, G., & McKee, S. P. (1980). Stereogram design for testing local stereopsis. Investigative Ophthalmology & Visual Science, 19, 802-809.

Wright, W. D. (1951). The role of convergence in stereoscopic vision. Proceedings of the Physical Society B, 64, 289-297.

APPENDICES

Appendix A. Orthographic Projections of Two Planar Curves Related by Reflection are Related by a 2D Affine Transformation

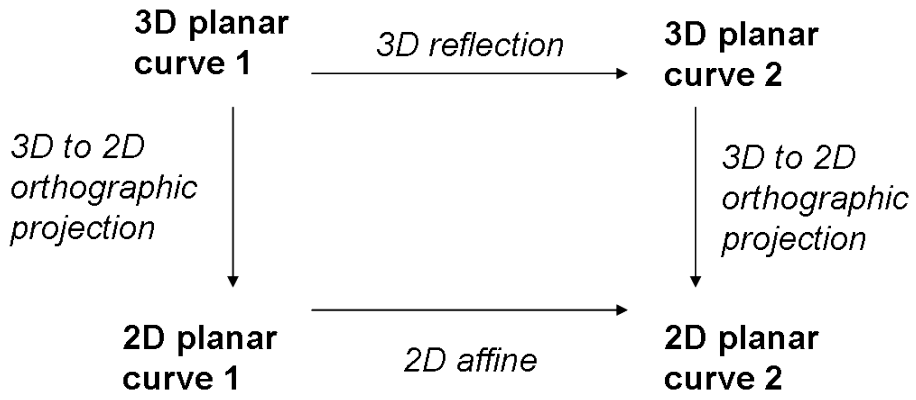


Figure 26. The relations analyzed in this appendix.

Suppose that the symmetry plane is

$$\pi_S: aX+bY+cZ+d =0 \quad \text{(A1)}$$

and A is a point on another plane

$$\pi: eX+fY+gZ+h=0 \quad \text{(A2)}$$

and its coordinates are $[X_A, Y_A, Z_A]$. Assume that π is not orthogonal to the image plane XY (this excludes degenerate views). It follows that $g\neq 0$. The symmetric point of A is A' whose coordinate is $[X_{A'}, Y_{A'}, Z_{A'}]$.

The points A and A' are symmetric with respect to the symmetry plane π_S if and only if they satisfy the following two conditions:

(1) The midpoint of A and A' is the on the plane π_S.

$$a(X_A+ X_{A'})+ b(Y_A+ Y_{A'})+ c(Z_A+ Z_{A'})+d = 0 \quad \text{(A3)}$$

(2) The line AA' is perpendicular to the symmetry plane π_S.

$$[X_A- X_{A'}, Y_A- Y_{A'}, Z_A- Z_{A'}]\times[a, b, c] = \mathbf{0} \quad \text{(A4)}$$

Note the 0 on the right-hand side of equation (A4) is written in bold font because 0 is a vector. The equation (A4) is equivalent to the following three equations:

$$a(Y_A-Y_{A'}) – b(X_A-X_{A'}) = 0 \quad \text{(A5)}$$

$$b(Z_A-Z_{A'}) - c(Y_A-Y_{A'}) = 0 \quad \text{(A6)}$$

$$c(X_A-X_{A'}) - a(Z_A-Z_{A'}) = 0 \quad \text{(A7)}$$

Note that equation (A7) is a linear combination of equations (A5) and (A6). Putting the three linear equations (A3), (A5) and (A6) together, we can write them as

$$\begin{pmatrix} a & b & c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}\begin{pmatrix} X_{A'} \\ Y_{A'} \\ Z_{A'} \end{pmatrix} = \begin{pmatrix} -a & -b & -c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}\begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} + \begin{pmatrix} -d \\ 0 \\ 0 \end{pmatrix} \quad \text{(A8)}$$

Let

$$W = \begin{pmatrix} a & b & c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix} \quad \text{(A9)}$$

Then the determinant of W is equal to

$$\det(W) = b(a^2+b^2+c^2) \quad \text{(A10)}$$

If b is not equal to 0, then W is a full rank matrix and we can compute the inverse of W. Pre-multiplying both sides of A8 by W^{-1}, we can express the coordinates of point A' as

$$\begin{pmatrix} X_{A'} \\ Y_{A'} \\ Z_{A'} \end{pmatrix} = \begin{pmatrix} a & b & c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}^{-1}\begin{pmatrix} -a & -b & -c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}\begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} + \begin{pmatrix} a & b & c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}^{-1}\begin{pmatrix} -d \\ 0 \\ 0 \end{pmatrix}$$

(A11)

Let

$$P = \begin{pmatrix} a & b & c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}^{-1}\begin{pmatrix} -a & -b & -c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}, \mathrm{T} = \begin{pmatrix} a & b & c \\ b & -a & 0 \\ 0 & c & -b \end{pmatrix}^{-1}\begin{pmatrix} -d \\ 0 \\ 0 \end{pmatrix}, V' = \begin{pmatrix} X_{A'} \\ Y_{A'} \\ Z_{A'} \end{pmatrix}$$

$$\text{and } \mathrm{V} = \begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix},$$

then equation (A11) can be simply written as

$$V'=PV+T \quad \text{(A12)}$$

Or

$$\begin{pmatrix} X_{A'} \\ Y_{A'} \\ Z_{A'} \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} \quad \text{(A13)}$$

Considering the first two rows, we obtain

$$\begin{pmatrix} X_{A'} \\ Y_{A'} \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} X_A \\ Y_A \end{pmatrix} + \begin{pmatrix} P_{13} \\ P_{23} \end{pmatrix} Z_A + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \quad \text{(A14)}$$

Because point A is on the plane π, from equation (A2), we obtain

$$Z_A = -\frac{h}{g} - \frac{e}{g} X_A - \frac{f}{g} Y_A \quad \text{(A15)}$$

Replacing Z_A in (A14) with (A15), we obtain

$$\begin{pmatrix} X_{A'} \\ Y_{A'} \end{pmatrix} = \frac{1}{g} \begin{pmatrix} gP_{11} - eP_{13} & gP_{12} - fP_{13} \\ gP_{21} - eP_{23} & gP_{22} - fP_{23} \end{pmatrix} \begin{pmatrix} X_A \\ Y_A \end{pmatrix} + \frac{1}{g} \begin{pmatrix} gT_1 - hP_{13} \\ gT_2 - hP_{23} \end{pmatrix} \quad \text{(A16)}$$

Let

$Q = \frac{1}{g} \begin{pmatrix} gP_{11} - eP_{13} & gP_{12} - fP_{13} \\ gP_{21} - eP_{23} & gP_{22} - fP_{23} \end{pmatrix}$ and $S = \frac{1}{g} \begin{pmatrix} gT_1 - hP_{13} \\ gT_2 - hP_{23} \end{pmatrix}$, then equation (A16)

can be simply written as

$$\begin{pmatrix} X_{A'} \\ Y_{A'} \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} X_A \\ Y_A \end{pmatrix} + \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} \quad \text{(A17)}$$

Equation (A17) indicates that [$X_{A'}$, $Y_{A'}$] and [X_A, Y_A] are related by a 2D affine transformation. Note, the [$X_{A'}$, $Y_{A'}$] and [X_A, Y_A] are 2D orthographic projections of A' and A, and A' and A are symmetric with respect to the plane π_S. Therefore, orthographic projections of two mirror symmetric planes are subject to affine transformation. Note there are six parameters in equations (A17) that determine the 2D affine transformation. Three pairs of symmetric correspondences in the image plane will be enough to compute these parameters.

If b is equal to 0 in equation (A10), from equation (A5) or (A6), we obtain

$$Y_A = Y_{A'} \quad \text{(A18)}$$

Combining equations (A2), (A3), (A7) and removing Z_A and $Z_{A'}$, we obtain

$$X_{A'} = kX_A + mY_A + n \quad \text{(A19)}$$

Where

$$k = \frac{(-a^2+c^2)g+2ace}{(a^2+c^2)g},\ m = \frac{2acf}{(a^2+c^2)g} \text{ and } n = \frac{2a(ch-dg)}{(a^2+c^2)g} \quad \text{(A20)}$$

Putting the two linear equations (A18) and (A19) together, we obtain

$$\begin{pmatrix} X_{A'} \\ Y_{A'} \end{pmatrix} = \begin{pmatrix} k & m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_A \\ Y_A \end{pmatrix} + \begin{pmatrix} n \\ 0 \end{pmatrix} \quad \text{(A21)}$$

Therefore, if b is equal to 0, $[X_{A'}, Y_{A'}]$ can be also written as a 2D affine transformation of $[X_A, Y_A]$.

Appendix B. Recovery From a Real 2D Image

There are several differences between the recovery from a synthetic image, like the image of a polyhedron, and the recovery from a real image, like the image of a bird.

1. An image of a polyhedron consists of line segments. For a line segment, we only need to recover its endpoints and the line segment can be recovered by connecting the recovered endpoints. In the case of a real image of a natural object, it is very unlikely that the curves on the image are straight. It follows that all points on each curve have to be recovered. Next, to apply our model to a real image, we need to extract curves from the image first. Detecting curves in a real image is itself a difficult problem. The human visual system solves it very well. However, there is still no algorithm that can find curves in real images, reliably. The problem of finding curves in a real image is outside the scope of this proposal. Our model has been be applied to curves in the image that have been extracted by hand.
2. When we recover a synthetic polyhedron from its 2D image, the information about which points in the 3D interpretation are symmetric is given. Again, in the case of polyhedra, the correspondence of the endpoints is sufficient because straight-line segments in 3D are produced (reconstructed) simply by connecting reconstructed endpoints. The situation is different when curves, rather than line segments, are recovered. In the case of curves, we need to know symmetry correspondences for all points on the curve. This is done as follows. We mark (by hand) which two curves in the image are symmetric in the 3D interpretation. Then, we indicate the direction in the image of the line segment connecting any symmetric pair of points. Recall, that the line segments connecting symmetric points in the 3D shape project to parallel line segments in an orthographic image of the shape. It follows

that all corresponding pairs of points in the image share the same direction.

3. In the case of polyhedra, the planarity constraint is always used because all faces of a polyhedron are planar. However, in the case of natural shapes, planarity is less common and is more difficult to establish. It follows that planarity will not be used as often. One implication of this fact is that the back, invisible part of the object will be more difficult to recover.
4. Not all curves found in the image will be used to recover a 3D shape. If a curve does not have a visible counterpart or the curve is not planar, there may not be enough constraints to recover it.
5. In the presence of noise, the line segments connecting images of symmetric points (symmetry line segments) are not exactly parallel. In such a case, there is no exact solution to the problem of 3D shape recovery. One way to remedy this problem is to correct the orientations of symmetry line segments before 3D recovery. This can be done by moving endpoints of each curve so that all symmetry line segments defined by the endpoints are parallel. Specifically, the endpoints are moved as little as possible (in the least squares sense) to make the symmetry line segments parallel. In the case of real images, noise is handled differently. Note that in the case of real images it is not known a priori which points are corresponding. The correspondence is established by using the direction of the symmetry line segments, which is indicated by hand. However, the noise on the curves in a real image could lead to spurious correspondences. This problem is illustrated in Figure 25. Consider the noisy curves in Figure 25(b). Note that both points, B and C are on a symmetry line segment emanating from point A. Point C leads to a spurious correspondence with point A. This problem can be eliminated by smoothing the curves (see Figure 25(d)).
6. There are two kinds of symmetric curves that are symmetric to themselves. These curves have to be identified, because they are

recovered separately. The first is any curve on the symmetry plane. The mirror image of each point on the curve is the point itself. The second is a curve that intersects the symmetry plane. For each point on the curve, its corresponding point is on the same curve, but on the opposite side of the symmetry plane. These two kinds of curves will be recovered after all other curves have been recovered.

7. For some shapes, like a chair, table, or spider, et al., it may be difficult to define their surface and volume. In fact, these objects do not have much volume. A small amount of volume may lead to unstable 3D recoveries when 3D compactness is maximized. Note, however, that these objects occupy a substantial amount of 3D space. Therefore, it is natural to compute their convex hulls and apply the maximum compactness and minimum surface area to the convex hull, instead of to the shapes themselves.

To summarize, the operations on how to apply our model to real images of real objects are as below:

1. Draw the contours in the image by hand. Indicate which curves are symmetric and which curves are on the same plane.
2. Smooth the curves in the image.
3. Indicate the direction of symmetry line segments and then compute all the correspondences.
4. The symmetry constraint is applied to recover all symmetric points (up to one unknown parameter).
5. The planarity constraint is applied to recover the hidden part (optional).
6. Obtain the orientation of the symmetry plane and recover those curves that are symmetric to themselves.
7. Compute the convex hull for each 3D interpretation within the one-parameter family of 3D shapes. Find the value of α that maximizes V/S^3.
8. Recover 3D curves using the α obtained in the step 7.

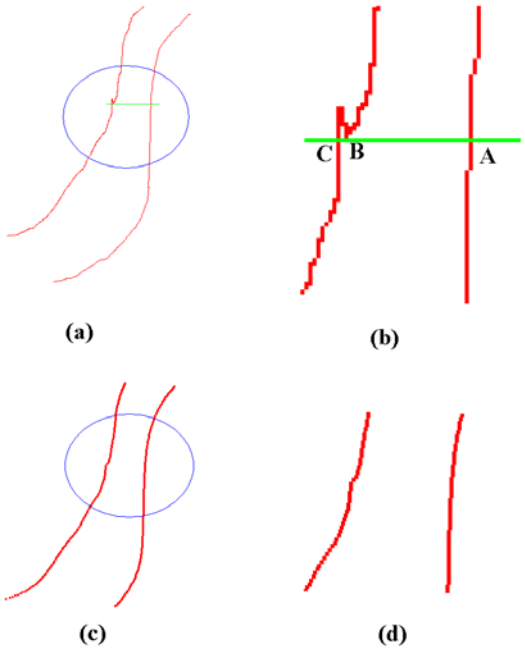


Figure 27. The illustration of how to eliminate spurious symmetry correspondences by smoothing curves. (a) A pair of original symmetric curves in an image. The green line indicates the direction of a symmetry line segment. (b) An enlarged image for the area in (a) encircled by the blue circle. Points B and C, which are on the curve on the left, are both possible symmetric counterparts of point A. (c) The pair of symmetric curves after smoothing. (d) An enlarged image for the area in (c) encircled by the blue circle.

Appendix C. Simulation of the Subjects' Ability of Detecting the Depth Order Between Points

Rady & Ishak (1995) measured 10 subjects' stereoscopic acuity. In their experiment, subjects viewed two illuminated circular apertures binocularly – one was the reference target and the other was the test target. The separation between the reference target and test target was controlled and it could be 7, 14, 25, 40 or 52 degrees. In each separation, the viewing distance of the reference target was fixed at 200cm. The viewing distance of the test target was changed and at each distance subjects were asked to judge which target was closer to the observer (see Figure 26). Then a psychometric curve between the response and the distance of the test target was obtained. The reciprocal of the standard deviation of the distribution of the response was used as the stereoscopic acuity at the corresponding separation.

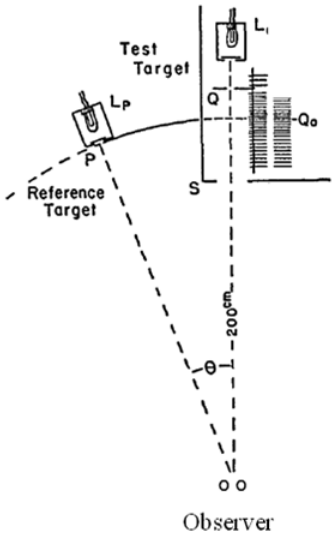


Figure 28. The apparatus Rady and Ishak (1955) used to measure human's stereoscopic acuity. The viewing distance of the reference target was 200cm.

Table 1 in Rady & Ishak's paper listed the averaged stereoscopic acuities across subjects for the different separation. Five stereoscopic acuities for different separations were measured in their experiment. To predict the stereoscopic acuity at the other separations, we drew a curve to fit the Rady & Ishak's results (see Figure 27). The expression for the fitting curve was:

$$k_{200} = 4.4/s + 0.22 \quad \text{(C1)}$$

Where k is the stereoscopic acuity and s is the separation (expressed in cm) between the reference target and the test target. According to the definition of stereoscopic acuity, the standard deviation for the response distribution at the separation s is

$$\sigma = 1/k_{200} \quad \text{(C2)}$$

or

$$\sigma(s) = \left(\frac{s}{4.4+0.22s}\right) \quad \text{(C3)}$$

Note that we put a subscript 200 after k. This is because Rady & Ishak's measure is for the viewing distance of 200cm. To derive the stereoscopic acuity at the other viewing distances, we need to know how the viewing distance affects the binocular disparity. From equation (26), we can derive

$$\Delta d = \delta d^2/I \quad \text{(C4)}$$

Equation (C4) indicates that for a given binocular disparity expressed in radians, the depth is proportional to the square of the viewing distance. Combining (C3) and (C4), the standard deviation, expressed in cm, at the distance d is

$$\sigma(d,s) = \left(\frac{d}{200}\right)^2 \left(\frac{s}{4.4+0.22s}\right) \quad \text{(C5)}$$

Therefore, when the viewing distance is d and the separation is s, the stereoscopic acuity is

$$k(d,s) = \left(\frac{200}{d}\right)^2 \left(\frac{4.4}{s} + 0.22\right) \quad \text{(C6)}$$

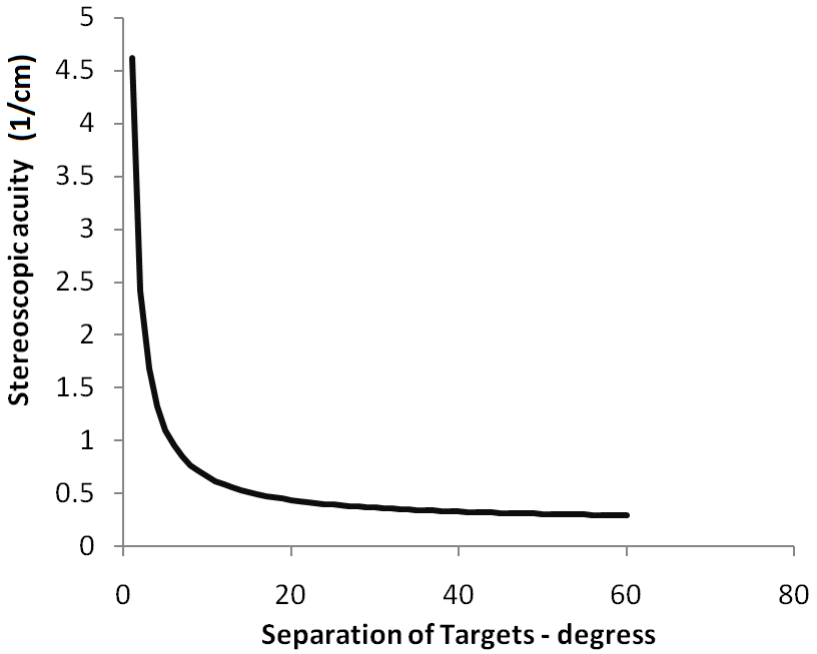


Figure 29. The stereoscopic acuity fitting curve for Rady and Ishak's results. The X axis is the separation between the reference target and the test target. The Y axis is the stereoscopic acuity.

Suppose A_i and A_j are two points, their viewing distances are d_i and d_j and the separation between them is s. In our experiment, subjects can change their fixation as they wanted. Therefore, the averaged viewing distance for A_i and A_j is considered as the viewing distance. According to equation (C5), we can compute the standard deviation of responses at the separation of s and the viewing distance of $(d_i+d_j)/2$.

To determine what subjects' judgment of depth order between A_i and A_j is, we should know their detection threshold. Threshold is usually defined as the percentile at which the signal can be successfully detected at 75%. We adopted this convention. If subjects can successfully tell the depth order between A_i and A_j at the rate of 75% or more than 75%, we assume that the subjects can tell the depth order. Otherwise, the depth order between A_i and A_j

is uncertain. Under the assumption of a normal distribution, the Z value that corresponds to 75% is 0.674. Therefore, the equation (27) can be written as

$$O(A_i, A_j) = \begin{cases} 1, & \frac{d_i - d_j}{\sigma((d_i + d_j)/2, s)} > 0.674 \\ 0, & \left|\frac{d_i - d_j}{\sigma((d_i + d_j)/2, s)}\right| \leq 0.674 \\ -1, & \frac{d_i - d_j}{\sigma((d_i + d_j)/2, s)} < -0.674 \end{cases} \quad \text{(C7)}$$

Appendix D. The Recovery of a Mirror Symmetric 3D Shape From its Perspective Image

Suppose c_1 and c_2 are the perspective projections of two 3D curves C_1 and C_2 that are mirror symmetric, and O is the projection center. Thus, the recovery from c_1 and c_2 is equivalent to computing c_1 and c_2's back projections that are mirror-symmetric with respect to some plane π_s.

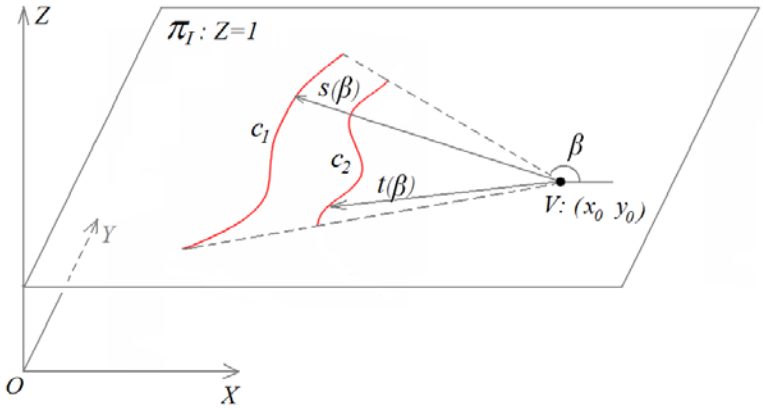


Figure 30. O is the projection center and Z=1 is the image plane. c_1 and c_2 are two arbitrary curves on the image plane. Point V is the vanishing point on the image plane.

Let the projection center O be the origin of the coordinate system and direction of the visual axis to be the z-axis. The image plane is perpendicular to the z axis. Without loss of generality, we assume that plane z=1 is the image plane π_I (see Figure 28). In a perspective projection, the lines connecting pairs of points (called corresponding points) in the image that are projections of pairs of symmetric points in 3D intersect at one point that is called the vanishing point. Let V(X_0, Y_0, 1) be the vanishing points on the image. Then, the points on c_1 and c_2, which are on a line emanating from the vanishing point are pairs of corresponding points. Let the angle between x-axis and the line emanating from vanishing point be β (see Figure 28). Let

$$v_1: [X_0+s(\beta)\cos(\beta), Y_0+s(\beta)\sin(\beta), 1] \quad (D1)$$

be a point on c_1, where $s(\beta)$ is the distance between v_1 and the vanishing point V. Similarly, let

$$v_2: [X_0+t(\beta)\cos(\beta), Y_0+t(\beta)\sin(\beta), 1] \quad (D2)$$

be a point on c_2, where $t(\beta)$ is the distance between v_2 and the vanishing point V. Because the angle β for v_1 and v_2 is the same, v_1 and v_2 are corresponding points. Suppose V_1 and V_2 are the perspective backprojections of v_1 and v_2. According to the property of perspective projection, we have

$$V_1: [Z_1(X_0+s(\beta)\cos(\beta)), Z_1(Y_0+s(\beta)\sin(\beta)), Z_1] \quad (D3)$$

$$V_2: [Z_2(X_0+t(\beta)\cos(\beta)), Z_2(Y_0+t(\beta)\sin(\beta)), Z_2] \quad (D4)$$

where Z_1 and Z_2 are the Z value for V_1 and V_2. Suppose that V_1 and V_2 are mirror symmetric with respect to a plane

$$\pi_s: aX+bY+cZ+d=0. \quad (D5)$$

Then, the following two equations are satisfied.

(1) The midpoint of V_1 and V_2 is on the plane π_s.

$$a(Z_1(X_0+s(\beta)\cos(\beta))+Z_2(X_0+t(\beta)\cos(\beta)))+b(Z_1(Y_0+s(\beta)\sin(\beta))+Z_2(Y_0+t(\beta)\sin(\beta)))+c(Z_1+Z_2)+2d =0 \quad (D6)$$

(2) The line connecting V_1 and V_2 is parallel to the normal of the plane π_s:

$$[Z_1(X_0+s(\beta)\cos(\beta))-Z_2(X_0+t(\beta)\cos(\beta)) \; Z_1(Y_0+s(\beta)\sin(\beta))-Z_2(Y_0+t(\beta)\sin(\beta)) \; Z_1+Z_2]\times[a \; b \; c] = [0 \; 0 \; 0] \quad (D7)$$

Equation (D7) is equivalent to the following three equations:

$$c(Z_1(Y_0+s(\beta)\sin(\beta))-Z_2(Y_0+t(\beta)\sin(\beta)))-b(Z_1-Z_2) \quad (D8)$$

$$c(Z_1(X_0+s(\beta)\cos(\beta))-Z_2(X_0+t(\beta)\cos(\beta)))-a(Z_1-Z_2) \quad (D9)$$

$$b(Z_1(X_0+S(\beta)\cos(\beta))-Z_2(X_0+t(\beta)\cos(\beta)))-a(Z_1(Y_0+s(\beta)\sin(\beta))-Z_2(Y_0+t(\beta)\sin(\beta)))=0 \quad (D10)$$

Note equation (D10) is linear combination of equations (D8) and (D9). Equations (D8) and (D9) can be rewritten as

$$(cY_0-b)(Z_1-Z_2)+c\sin(\beta)(Z_1s(\beta)-Z_2t(\beta))=0 \quad (D11)$$

$$(cX_0-a)(Z_1-Z_2)+c\cos(\beta)(Z_1s(\beta)-Z_2t(\beta))=0 \quad (D12)$$

We multiply both sides of (D11) and (D12) by $\cos(\beta)$ and $-\sin(\beta)$, respectively, and then add left-hand sides and right-hand sides.

$$(\cos(\beta)(cY_0-b)-\sin(\beta)(cX_0-a))(Z1-Z2) = 0 \quad (D13)$$

Note, equation (D13) should be satisfied for any symmetric pairs. It means for any value of β, equation (D13) should be satisfied, which means that:

$$cY_0-b = 0 \quad (D14)$$

$$cX_0-a = 0 \quad (D15)$$

From (C14) and (C15), we can obtain

$$b = cY_0 \quad (D16)$$

$$a = cX_0 \quad (D17)$$

Therefore, the symmetry plane can be rewritten as

$$cX_0X+cY_0Y+cZ+d = 0 \quad (D18)$$

or

$$X_0X+Y_0Y+Z+e = 0 \quad (D19)$$

where e = d/c. Note that the normal of the symmetry plane is $[X_0\ Y_0\ 1]$, which is the coordinate of the vanishing point. Note that the line connecting the origin (projection center) and the vanishing is parallel to the line connecting the recovered symmetric pairs. Because those lines are orthogonal to the symmetry plane, they indicate the direction of the normal of the symmetry plane. Therefore, the normal of the symmetry plane is equal to the vector from origin to the vanishing point. Equation (D18) shows that e is a free parameter, which indicates the position of the symmetry plane.

From (C16) and (C11), we obtain

$$Z_1s(\beta)-Z_2t(\beta) = 0 \quad (D20)$$

From (C20) and (C6), we obtain

$$Z_1 = -\frac{d}{c}\frac{2t(\beta)}{(t(\beta)+s(\beta))(X_0^2+Y_0^2+1)+2t(\beta)s(\beta)(X_0\cos(\beta)+Y_0\sin(\beta))} \quad (D21)$$

$$Z_2 = -\frac{d}{c}\frac{2s(\beta)}{(t(\beta)+s(\beta))(X_0^2+Y_0^2+1)+2t(\beta)s(\beta)(X_0\cos(\beta)+Y_0\sin(\beta))} \quad (D22)$$

From equations (D21) and (D22), the recovered shape is determined by the curves (t(β), s(β)) on the image and the vanishing point (X_0, Y_0). d/c that indicates the position of the symmetry plane is a scaling factor, which affects the size of the recovered object, but not its shape.

To summarize, for a perspective image of a mirror symmetric 3D shape, if the projection center and the vanishing point are known, the orientation of the symmetry plane and the 3D shape are uniquely determined. The size of the recovered is undetermined.

VITA

VITA

Yunfeng Li

PERSONAL INFORMATION

Address:
321 E Parkside Dr., Apt 213
Whitewater, WI, 53190

Email: Li135@purdue.edu

Telephone: (262)510-2981

EDUCATION

Ph. D. Mathematical and Computational Cognitive Science, Department of Psychological Sciences
Purdue University, Dec, 2009 (expected).
Dissertation: Computational models of 3D shape perception.
Advisor: Dr. Zygmunt Pizlo

M. A. Mathematical and Computational Cognitive Science, Department of Psychological Sciences
Purdue University, December, 2005.
Thesis: Binocular disparity vs. a priori constraints in 3D shape perception.
Advisor: Dr. Zygmunt Pizlo

M. A. Cognitive Psychology, Beijing University, July, 2001
Thesis: ERP research on perception without awareness
Advisor: Dr. Ying Zhu

B. S. Psychology, Beijing University, China, July, 1998
Honor Thesis: A research on memory binding
Advisor: Dr. Ying Zhu
Minor: Economics The Economic Research Center of China 1998
(Honor Science Program Certificate Beijing University, China 1997)

PUBLICATIONS

Pizlo, Z., Sawada, T., Li, Y., Kropatsch, W. & Steinman, R.M.(2009) A new approach to the perception of 3D shape based on verdicality, complexity, symmetry and volume. Vision Research. (In Press)

Li, Y. (2009). A new look on Perkins' law. Perception. (In Press)

Li, Y., Pizlo, Z., & Steinman, R. M. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal image. Vision Research, 49, 979-991.

Pizlo, Z., Li, Y., & Steinman, R. M. (2008). Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose. Spatial Vision, 21, 495–508.

Li, Y., & Pizlo, Z. (2007). Reconstruction of shapes of 3D symmetric object by using planarity and compactness constrains. Proceedings of the SPIE/IS&T electronic imaging. Vol. 6499.

Geng, H., Qi, Y., Li, Y., Fan, S., Wu, H., Zhu, Y. (2007). Neurophysiological correlates of memory illusion in both encoding and retrieval phases. Brain Research, 1136, 154-168.

Geng, H., Song, Q., Li, Y., Xu, S. & Zhu, Y. (2007) Attentional Modulation of Motion? Induced Blindness. Chinese Science Bulletin, 52 (8): 1063-1070.

Chan, M., Stevenson, A., Li, Y., & Pizlo Z. (2006). Binocular shape constancy from novel views: The role of a priori constraints. Perception and Psychophysics, 68, 1124-1139

Pizlo, Z., Li, Y., & Francis, G. (2005). A new look at binocular stereopsis. Vision Research, 45, 2244-2255.

Pizlo, Z., Li, Y., & Chan, M. (2005). Regularization model of human binocular vision. Proceedings of the SPIE/IS&T Computational Imaging Conference, 5674.

Geng, H., Song, Q., Li, Y., & Zhu, Y. (2005). The effect of attention to distractor on inhibitory processes in selective attention. Chinese Science Bulletin, 50, 1613-1619.

Geng, H., & Li, Y. (2002). Stroop effect under divided vs. focused attention: the influence of consciousness. Acta Scientiarum Naturalium Universitatis Pekinensis, 38, 421-426.

Geng, H., Fan, S., Li, Y. & Zhu, Y. (2002). The neuroscience study on consciousness. Acta Psychologica Sinica, 34, 91.

Geng, H., Zhu, Y., & Li, Y. (2001). False recognition: awareness, attention and stimulus quality. Acta Psychologica Sinica, 33, 104-110.