A PSYCHOLOGICALLY PLAUSIBLE ALGORITHM FOR BINOCULAR SHAPE RECONSTRUCTION

A Thesis
Submitted to the Faculty

of

Purdue University

by

Moses W. Chan

In Partial Fulfillment of the
Requirements for the Degree

of

Doctor of Philosophy

May 1999

To my wife, Linh, and our son, Alexander.

ACKNOWLEDGMENTS

I would like to thank Prof. Delp and Prof. Pizlo for their guidance and encouragement. Prof. Pizlo has served as a good mentor. I would also like to thank Prof. Maciejewski and Prof. Chelberg for serving on my thesis committee. Finally, I would like to sincerely thank Prof. Mowle for being there for me when I needed him.

DISCARD THIS PAGE

TABLE OF CONTENTS

Page

DISCARD THIS PAGE

LIST OF TABLES

DISCARD THIS PAGE

LIST OF FIGURES

Figure Page

Appendix
Figure Page

Appendix
Figure

Appendix
Figure Page

DISCARD THIS PAGE

NOMENCLATURE

θ	= vergence angle
$\mathbf{E}$	= essential matrix
$\mathbf{R}$	= rotation matrix
$\mathbf{T}$	= translation vector
Π	= left image plane
Π'	= right image plane
I	= interocular distance
D	= viewing distance
$\mathbf{u}$	= left image point
$\mathbf{u}'$	= right image point
$\mathbf{U}$	= 3-D object point
H	= left shape hypothesis
H$'$	= right shape hypothesis

ABSTRACT

Moses W. Chan, Ph.D., Purdue University, May 1999. A Psychologically Plausible Algorithm for Binocular Shape Reconstruction. Major Professors: Edward J. Delp and Zygmunt Pizlo.

Binocular shape reconstruction is an inverse problem of inferring a 3-D similarity structure of an object from two perspective views. As most inverse problems, shape reconstruction is ill-posed and ill-conditioned, which implies that its solution is unstable, i.e., the reconstructed shape is extremely sensitive to noise present in the images. In order to accurately reconstruct a shape from noisy images, a priori knowledge (constraints) must be used. Two types of constraints are considered here: (a) system constraints; and (b) constraints of the geometrical properties of the objects.

We first modified the 8 point algorithm by adding a binocular fixation constraint (system constraint). The reconstruction performance of this algorithm is substantially better than that of the 8 point algorithm. We also compared the performance of these two algorithms to the performance of humans. The results show that the accuracy of the new algorithm is similar to that of humans, but the stability of the solutions is still very poor.

To obtain more stable solutions, geometrical constraints must be used to restrict a family of possible solutions. Recent psychophysical experiments showed that monocular cues related to 3-D topology, and constraints such as planarity of surface contours and symmetry of the object are crucial in shape perception. Based on these results, we developed a new algorithm for binocular shape reconstruction. In this algorithm, a 3-D shape is first obtained by means of monocular reconstruction, and then binocular disparity is used to correct the shape. The reconstruction performance of this new algorithm is substantially more stable and accurate than that of the first algorithm,

and it is similar to that of humans. The psychological plausibility of this method has been tested and confirmed in a psychophysical experiment.

1. INTRODUCTION

Binocular shape reconstruction is a process of inferring a 3-D similarity structure of an object from two perspective views. A similarity structure of an object refers to the geometrical properties of its contour or surface that are invariant under rigid motion and size scaling. It is assumed in this document that point correspondences between the two images are known. This document consists of the following parts: (a) we compare the performance of one of the best known computer vision algorithms (8 point algorithm) to the performance of human subjects. Such comparisons are not made very often, but they are useful in evaluating the quality of computer vision methods; (b) we determine to what extent the constraints that exist in human binocular vision can improve the computational efficiency of binocular reconstruction; (c) we evaluate the usefulness of the 8 point algorithm in modeling human vision; (d) we point out that geometrical constraints of the objects must be used to obtain stable reconstructions; (e) we propose a new approach to binocular shape reconstruction, and finally, we perform a psychophysical test that evaluates the psychological plausibility of this new approach. We begin with a review of computer vision and human vision research on binocular shape reconstruction.

1.1 Prior Computer Vision Research on Binocular Shape Reconstruction

We consider the problem of reconstructing the 3-D shape of an object from two perspective views. The geometry of the problem is sketched in Figure 1.1. Without restricting the generality we assume the coordinate system of the right camera, X′Y′Z′, to be the world coordinate system. The left camera with the coordinate system XYZ is arbitrarily translated and rotated relative to the right one. The optical axes of the

left and right cameras are aligned with the Z and Z′ axes respectively. The left image plane (Π) coincides with the XY plane, and the right image plane (Π′) coincides with the X′Y′ plane.

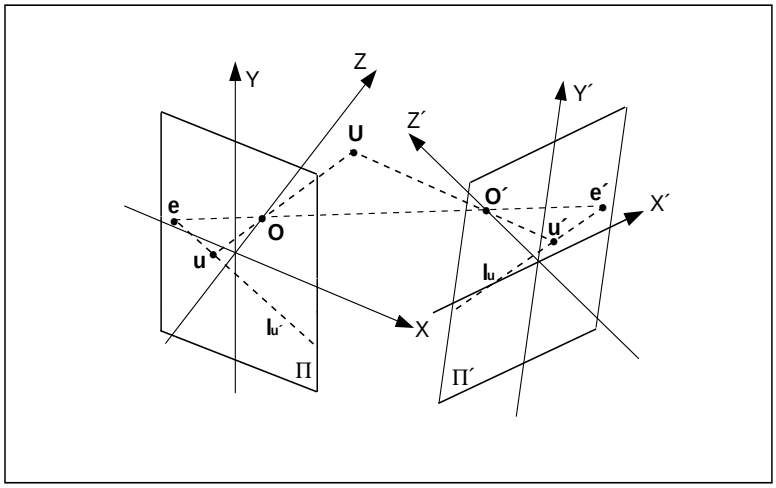


Fig. 1.1. Geometry of a binocular viewing system.

Before a shape is reconstructed by a binocular viewing system, the system must be calibrated. Calibration involves obtaining the intrinsic and extrinsic parameters of the cameras. The intrinsic parameters include: coordinates of the principal point, focal length, pixel's aspect ratio and the skew parameter (see [Fau93] for a comprehensive review of camera calibration). In most of this document we assume that the intrinsic parameters of the two cameras are known. The extrinsic parameters include the 3-D orientation and position of one camera relative to the other, which is represented by six independent parameters. Rotation about the X axis is known as the elevation, and rotation about the Z axis (i.e. optical axis of the left camera) is known as the torsion. The angle formed by Z and Z′ is called vergence.

The first method of obtaining the extrinsic parameters was provided by Kruppa [Kru13]. He showed that two perspective images of 5 points obtained by a camera with known intrinsic parameters determine a finite number of possible solutions. Kruppa's method was iterative; it involved obtaining eleven solutions from a non-linear equation (Faugeras and Maybank corrected Kruppa by pointing out that at most ten solutions

exist in this case [FM90]). Thompson proposed another iterative method, which involved solving a set of third order non-linear equations with five unknowns [Tho59]. Subsequently, a method which involved solving a set of linear equations was proposed by Longuet-Higgins [LH81], and independently by Tsai and Huang [TH84]. This method is known as the 8 point algorithm, and it has been widely used in the areas of binocular shape reconstruction and reconstruction of shape from motion. Below we present a brief description of the 8 point algorithm and its computational properties.

Let $\mathbf{U}$ be a 3-D object point, $\mathbf{O}$ and $\mathbf{O}'$ the optical centers of two cameras, and $\mathbf{u}$ and $\mathbf{u}'$ two projections of $\mathbf{U}$ on the image planes $\mathbf{\Pi}$ and $\mathbf{\Pi}'$ respectively (Figure 1.1). The intersections of the line that goes through $\mathbf{O}$ and $\mathbf{O}'$, and the image planes $\mathbf{\Pi}$ and $\mathbf{\Pi}'$ are the epipoles $\mathbf{e}$ and $\mathbf{e}'$ respectively. The semi infinite line starting from $\mathbf{e}$ and passing through $\mathbf{u}$ is called the epipolar line $\mathbf{l}_{\mathbf{u}'}$. Similarly, the epipolar line $\mathbf{l}_{\mathbf{u}}$ starts at $\mathbf{e}'$ and passes through $\mathbf{u}'$. The geometry of the camera configuration is governed by the well known epipolar constraint: for each point $\mathbf{u}$ on $\mathbf{\Pi}$, its corresponding point $\mathbf{u}'$ lies on the epipolar line $\mathbf{l}_{\mathbf{u}}$. Similarly, for each point $\mathbf{u}'$ on $\mathbf{\Pi}'$, its corresponding point $\mathbf{u}$ lies on its epipolar line $\mathbf{l}_{\mathbf{u}'}$.

It is assumed that the intrinsic parameters are known. Therefore, the image points can be represented by Cartesian coordinates: $\mathbf{u} = [u_x u_y 1]^t$ and $\mathbf{u}' = [u'_x u'_y 1]^t$. Consider two visual rays emanating from the image points $\mathbf{u}$ and $\mathbf{u}'$, and passing through their corresponding optical centers $\mathbf{O}$ and $\mathbf{O}'$. These two visual rays intersect in 3-D space if and only if the following equation is satisfied [LH81] [TH84]:

$$\begin{bmatrix} u'_x & u'_y & 1 \end{bmatrix} \mathbf{E} \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix} = 0 \tag{1.1}$$

where

$$\mathbf{E} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \mathbf{R}^t$$

$\mathbf{R}^t$ represents the transposition of $\mathbf{R}$ where $\mathbf{R}$ is a 3x3 matrix representing the rotation, and T_x, T_y and T_z represent the translation of the left camera relative to the right one. The 3x3 matrix $\mathbf{E}$ is known as the essential matrix. It encapsulates the epipolar geometry of the camera configuration. Since Equation 1.1 is a linear homogeneous equation with 9 unknowns (only 6 of them are independent), the matrix $\mathbf{E}$ has 8 parameters which can be computed up to a scale factor if 8 or more pairs of corresponding points are provided (one pair of points yields one equation). After $\mathbf{E}$ is computed, $\mathbf{R}$, T_x, T_y and T_z can be recovered up to a scale factor. Then, the 3-D object point $\mathbf{U}$ can be reconstructed (also up to a scale factor) by straightforward algebraic manipulations [LH81].

The 8 point algorithm usually determines a unique solution. However, multiple solutions can occur for some configuration of points. It was noted by Longuet-Higgins that the reconstruction is not unique if four or more points lie on a straight line, 6 or more points lie on a conic, or 7 or more points lie on a plane [LH81]. Shortly after the 8 point algorithm was published, Longuet-Higgins provided a necessary and sufficient condition for non-unique reconstruction: the reconstruction is not unique if and only if the points lie on a quadric surface that passes through both centers of projection [LH84]. A unique solution still can be computed even when fewer than 8 points are provided. Tsai and Huang showed that only 7 points are needed to obtain a unique reconstruction provided that the points do not lie on (a) a cone that passes through the origin of the world coordinate system, or (b) two planes with one passing through the origin [TH84]. Faugeras and Maybank [FM90] showed that up to ten solutions exist if only five pairs of corresponding points are provided. Other analyses of multiplicity of solutions can be found in [Neg90] [HNH90].

It is known that the 8 point algorithm is very sensitive to noise [SA89] [Har95]. In the presence of noise, the estimated extrinsic parameters are not accurate, which leads to substantial errors in the reconstruction of the 3-D points. One approach to reduce reconstruction errors is to improve the accuracy of the estimation of the extrinsic parameters. Spetsakis and Aloimonos pointed out that the 8 point algorithm

provides a solution which is not optimal in a maximum likelihood sense [SA89]. The 8 point algorithm would have been optimal if all 8 elements of the essential matrix had been independent. However, only 6 of them are independent. Spetsakis and Aloimonos showed that a non-linear least squares method, which estimates the 6 independent parameters, can be used to improve the estimation accuracy (a detailed error analysis of estimating extrinsic parameters can be found in [WHA89]).

Another approach to reduce reconstruction errors is to partially compensate for the error in the estimated extrinsic parameters by a triangulation method [HS97] [RCF95]. This method reconstructs a 3-D object point by computing the intersection of two visual rays. The intersection point exists in the absence of noise. However, if the image coordinates are corrupted with noise, then the visual rays are very unlikely to intersect in 3-D space. One way to estimate the location of a 3-D point in such cases is to obtain a point that minimizes the sum of the squared distances from the two visual rays. This point is known to be the mid-point of the common perpendicular to the two visual rays (Mid-Point method). The Mid-Point method has been used to reconstruct similarity and projective structures [HS97] [RCF95]. It was shown that in Euclidean space the Mid-Point method performs equally well as other computationally expensive minimization methods (see [HS97] for details). However, in projective space (the case of cameras with unknown intrinsic parameters) the Mid-Point method is unreliable, simply because of additional assumptions that are needed to impose a metric.

A similarity structure can also be reconstructed when only some (or none) of the intrinsic camera parameters are known. Hartley showed that the focal lengths, and the relative orientation and position of two cameras can be obtained if all other intrinsic parameters are known [Har92]. If none of the intrinsic parameters are known, then ground control points are needed to compute Euclidean structure of a scene [HGC92]. If ground control points are not available, then only projective or affine structures can be computed. Faugeras showed that in such a case Equation 1.1 can be written in a

more general form as $\mathbf{u}' = \mathbf{F}\mathbf{u}$, where $\mathbf{F}$ is a fundamental matrix [Fau92]. A detailed review of the fundamental matrix and its stability analysis can be found in [LF96].

Several algorithms for binocular reconstruction of similarity structure and motion estimation were formulated using the characteristics of human binocular vision. It is known that the human visual system is a fixating system. That is, the optical axes of the eyes intersect at a 3-D point. Thus, the fixating system has fewer degrees of freedom than the general binocular system used in the 8 point algorithm, and, as a result, it is computationally simpler. In the case of estimating camera motion relative to an object from two perspective views, Soatto and Perona [SP96] showed that the fixating system has only four degrees of freedom: one translational and three rotational. This is because the camera only translates along its fixation axis (the translations along X and Y axes are zero). In such a case, motion parameters can be estimated more accurately, as compared to the system with all six degrees of freedom. Other methods for motion estimation by a fixating system can be found in [RH94] [Taa92] [FA93].

In the case of binocular reconstruction using a fixating system, Gårding and Lindeberg [rL94] and Gårding et al. [rPMF95] showed that if torsion is zero, and the surface around the fixation point is locally planar, then it is possible to use the left and right image brightness gradients to compute a relief structure of a 3-D object. This structure preserves projective properties and depth ordering. If the vergence is known, then the local surface orientation of the object can be obtained. This algorithm operates directly on image brightness gradients; therefore, extraction of features such as points or lines is not needed. As a result, it is computationally simple.

Li et al. showed that if two cameras are mounted on a common elevation platform so that the system is fixating, and torsion is zero, then such a vision system has only two degrees of freedom (i.e. rotations about Y and Y′ axes). Note that in their system, the platform itself has two degrees of freedom, but since binocular reconstruction of shape requires only the information about the orientation and location of one camera relative to the other, these additional two degrees of freedom are not involved in the

computations [LBW96]. They showed that the fundamental matrix can be obtained with as few as three pairs of points if some intrinsic parameters of the cameras are known or the orientations of the cameras are known. The same camera configuration was used by Brooks et al. for self-calibration [BdAHB96]. They derived equations for computing the fundamental matrices for both static and moving stereo heads. These equations were based on the assumption that some of the intrinsic parameters are known. A similar approach was used by Chan el at. [CPC96] [CPC99]. They showed that 3 points are sufficient to recover the extrinsic parameters of a binocular fixating system and thus to reconstruct a 3-D scene if all intrinsic parameters are known.

1.2 Prior Human Vision Research on Binocular Shape Reconstruction

The fact that the images of an object seen by two eyes are slightly different was first stated by Euclid in about 300 B.C. However, the relationship between binocular disparity and depth perception was not demonstrated until Wheatstone designed the first stereoscope [Whe38]. He showed that fusing two images of a 3-D scene, obtained from slightly different viewing directions, leads to a single percept of the scene. Since then, extensive research has been conducted on perceptual mechanisms of human binocular perception. The main focus was on the human ability to reconstruct similarity and Euclidean structures. We begin with a treatment of reconstruction of Euclidean structure (i.e. distances). Prior theories involved a number of assumptions [HR95a]. Using the notation from the previous section, these assumptions are as follows:

1) All intrinsic parameters of the eyes are known.

2) The two eyes are located at the same height in the head coordinate system. This means that T_y is zero.

3) The two eyes are fixating at the same 3-D point. This means that the visual axes are coplanar. It follows that the elevations of the eyes are the same (i.e. the rotation about the horizontal axis of one eye relative to the other is zero).

4) Torsion in each eye is small (less than $\pm$ 8 degrees) [HR95b].

5) All non-zero extrinsic parameters (T_x, T_z, torsion and vergence θ) are known from extra-retinal cues.

To summarize these assumptions, all intrinsic and extrinsic parameters are assumed to be known from extra-retinal cues (some of them are assumed to be zero). In such a case, reconstruction can be performed even when only one point is in the field of view. The depth U_z of a point **U** is reconstructed first. Then, the other two coordinates of the point are reconstructed from its depth U_z and its image coordinates u_x and u_y using standard equations of perspective projection: $U_x = u_x U_z$, $U_y = u_y U_z$. The geometrical model of this reconstruction is shown in Figure 1.2.

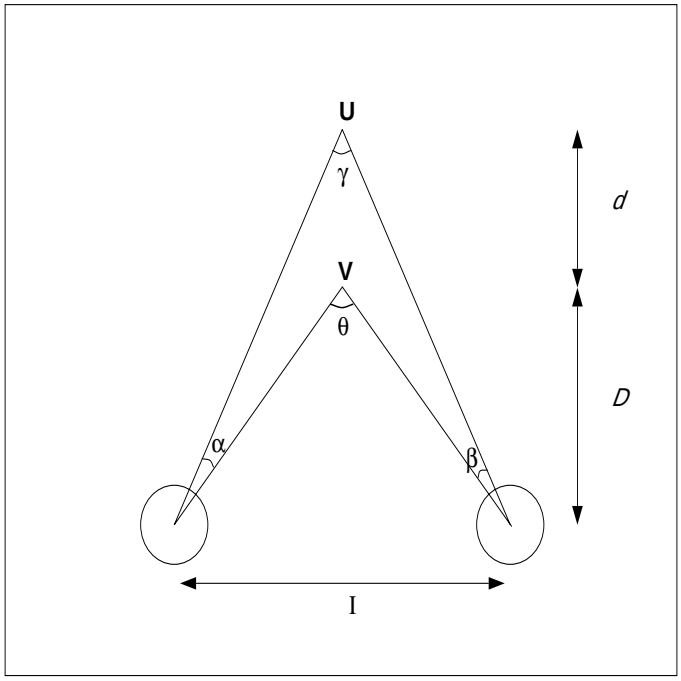


Fig. 1.2. Geometry of human binocular vision for reconstruction of 3-D points.

In this figure, d is the distance between the object point **U** and the fixation point **V**, I is the interocular distance (T_x), D is the distance between **V** and the observer. The angle of disparity (more exactly, horizontal disparity) $\delta = \theta - \gamma = \beta + \alpha$. Under the assumption that θ is small (so that $\theta \approx tan(\theta)$), and D is large relative to d (so that $D + d \approx D$), d can be approximated using the following equation [CF85]:

$$d \approx \frac{\delta D^2}{I} \tag{1.2}$$

where $D \approx I \ / \ \theta$.

According to this equation the relative depth d of a point $\mathbf{U}$ is computed based on the depth D of the fixation point, the interocular distance and the disparity angle. The depth of the fixation point itself is computed from vergence and interocular distance.

The fact that extra-retinal cues are not a reliable source of information about vergence has led researchers to derive a different method of depth reconstruction. Mayhew and Longuet-Higgins formulated a method based on vertical disparity (using the terminology from Section 1.1, vertical disparity is equal to $u_y - u'_y$) [MLH82]. In this method, vertical disparity of two image points is used to compute U_z directly without taking vergence into account. Recently, Rogers and Bradshaw showed that human subjects have the capability to use vertical disparity in depth judgements, but only if the simulated 3-D stimulus was very large [RB93]. One problem with using vertical disparity is that it is not applicable under symmetric viewing condition (i.e. when the object point is equally distant from both eyes) because $u_y - u'_y$ is zero for all distances.

To summarize the review of prior theories of perception of Euclidean structure, this structure was assumed to be obtained by first reconstructing the depths of points, and then the 3-D distances. Clearly, once distances are known, it is trivial to obtain the similarity structure (shape) because the similarity structure is represented by ratios of distances. Note that since depth is assumed to be a critical factor in the process of shape reconstruction, integrating depth cues, such as binocular disparity, texture, motion and shading is expected to improve shape reconstruction accuracy. Depth cues are usually assumed to be processed by different modules [ML89] [LMY90]. Three possibilities for depth cue interaction are generally considered. The first is vetoing; it is found when one strong cue completely overrides another weaker cue. The second and the third interactions are known as the weak and strong fusion [CY90]. In weak

fusion, depth estimates from different modules are linearly combined. Strong fusion on the other hand, involves interaction between modules prior to obtaining depth estimates.

Since prior theories assumed that shape reconstruction is based on reconstructing 3-D distances, precision and accuracy of the latter should limit those of the former. It is known that 3-D distances are not accurately nor reliably perceived by human observers [MLB90] [NTPT96]. This implies that either binocular shape reconstruction is not accurate, or it is not based on reconstructing 3-D distances. Recent psychophysical experiments showed that human subjects can judge shapes and relative depths accurately. Glennerster et al. measured the accuracy of matching the apparent depth of two surfaces at different distances [GRB96]. To perform this matching, the subjects did not need to estimate absolute depth of the two surfaces; only relative depth information (similarity structure) was needed. The subjects' judgments were accurate and reliable. Durgin et al. tested subjects' ability to match the shapes of two cones [DPOR95]. Again, this task required information only about similarity structure, rather than about Euclidean structure. Subjects' judgements were precise and accurate. All these results suggest that 3-D shape reconstruction is not based on reconstructing 3-D distances.

To summarize, prior theories of human binocular vision assumed that shape (similarity structure) is derived from 3-D distances (Euclidean structure). However, existing results suggest that shape perception does not have to be based on perception of distances. None of the existing human vision theories can account for these results. Interestingly, the 8 point algorithm does *not* perform reconstruction of distances, but this algorithm has never been tested as a psychological model. In this document we provide such a test.

2. PSYCHOPHYSICAL EXPERIMENT ON BINOCULAR SHAPE RECONSTRUCTION

The goal of this experiment was to measure the accuracy and precision of human binocular shape reconstruction. In order to evaluate the role of binocular, as opposed to monocular cues, we tested the subjects under both stereoscopic and monoscopic viewing conditions.

2.1 Method

2.1.1 Subjects

The author and two other subjects were tested. They all had normal, or corrected to normal vision. The subject's head was supported by a chin-forehead rest placed $50cm$ away from the monitor.

2.1.2 Stimuli

Twenty-five cylinders of revolution were used as stimuli. Each cylinder was generated from a distinct 2-D cubic B-Spline curve. The size of the simulated objects was about $8\times5\times5cm^3$ (visual angle of approximately 6 degrees). The geometry of computing the images of the stimuli is depicted in Figure 2.1. Let $X_oY_oZ_o$ be the object's coordinate system. The objects were displayed by means of a stereoscopic display system (The CrystalEyes System StereoGraphics Corporation) [Lip91]. This system consists of a pair of liquid-crystal display (LCD) glasses, a monitor, an infra-red emitter and a graphics display controller. Each LCD lens was electrically controlled to be opaque or transparent in synchronization with the display.

The images were alternately displayed in rapid succession in such a way that the right eye was occluded when the the left eye view (left image) was being presented

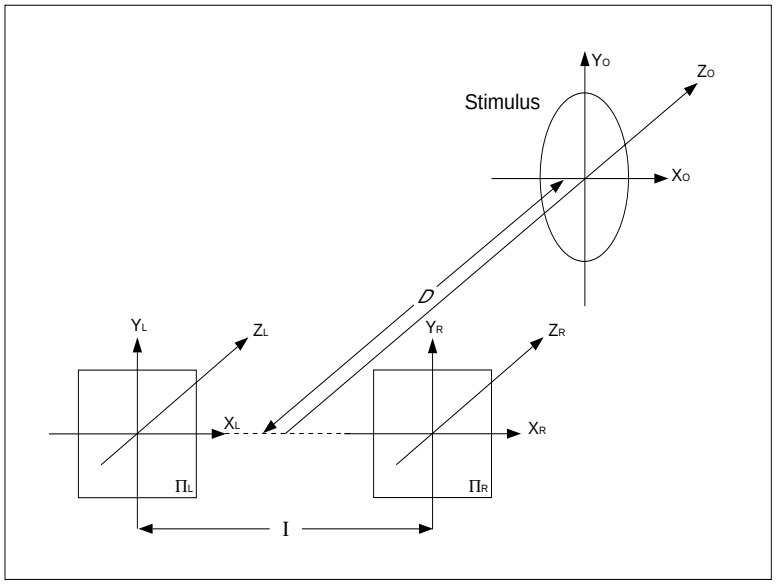


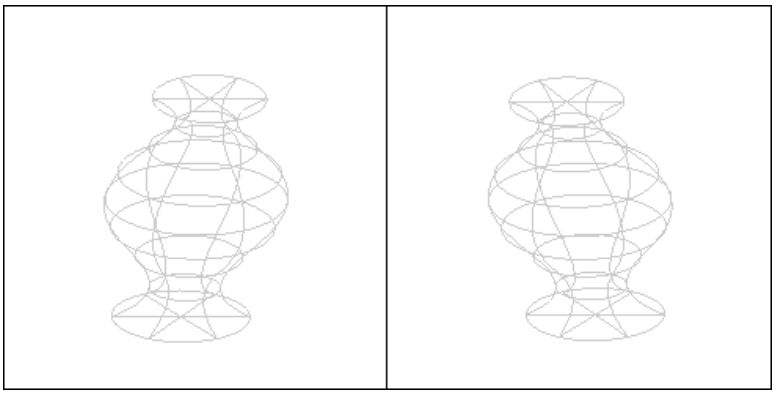
Fig. 2.1. Geometry of computing the images of the stimuli in the psychophysical experiment.

and vice versa. The left and right images were computed by assuming that the image planes Π_L and Π_R are coplanar, and the axes of their coordinate systems are parallel to the corresponding axes of the object's coordinate system. The simulated viewing distance D was $50cm$, and the simulated interocular distance I was $7cm$. In the monoscopic mode, the images were generated with I being zero. In this case, the left image was the same as the right image.

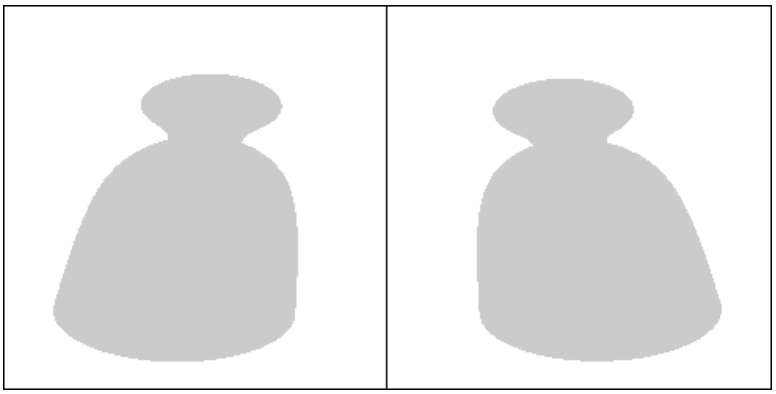
The switching rate of the display was $144/sec$. The luminance of the object through the active LCD glasses was $8.5cd/m^2$, and the luminance of the background was $32.0cd/m^2$.

Three types of monocular cues were used: occluding contour, wire frame, and shading. Examples of stereoscopic pairs of the stimuli are shown in Figure 2.2.

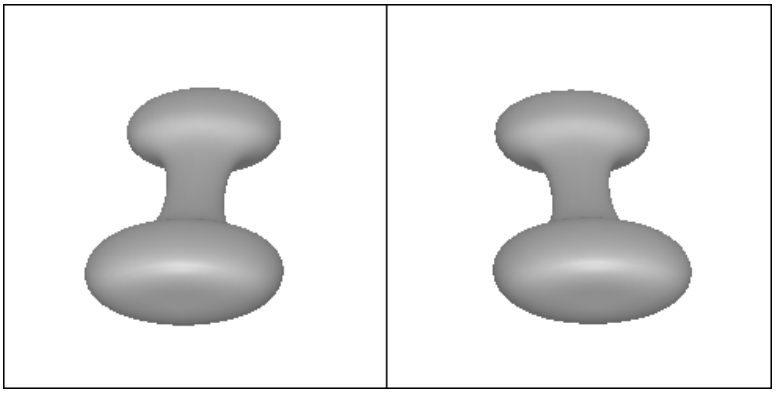
If integration of depth cues is involved in shape reconstruction as prior theories claim, the performance with wire frame and shaded stimuli should be much better than that with occluding contour stimuli. This is because both wire frame and shaded stimuli provide additional depth cues other than just occluding contour.



(a)



(b)



(c)

Fig. 2.2. Examples of stereoscopic pairs of the stimuli: (a) wire frame, (b) occluding contour and (c) shading.

2.1.3 Procedure

There were a total of six different sessions in this experiment representing all combinations of viewing condition (monoscopic vs stereoscopic) and monocular cue (occluding contour, wire frame and shading). The order of the sessions was random and different for every subject. In each trial, the subject was presented with two objects. The standard object was displayed on the left at an upright orientation. The comparison object was displayed on the right. It was obtained from the standard object by changing its aspect ratio. Specifically, this object was stretched or compressed along the direction of the axis of revolution (Y_o axis in Figure 2.1) and along the directions orthogonal to this axis. These transformations involved a random scaling factor from a uniform distribution within a range of 0.5 to 2.0. After the object had been transformed, it was rotated about the horizontal axis (X_o axis) so that its axis of revolution formed an angle with the frontal plane. The subject's task was to adjust the height of the standard object (using the mouse key) so that the aspect ratio of the standard object matched that of the comparison object. The procedure in which the aspect ratio is adjusted or discriminated has been commonly used in prior studies on shape perception (e.g. Thouless [Tho31], Stavrianos [Sta45], Durgin et al. [DPOR95], Glennerster et al. [GRB96], Johnston [Joh91], Pizlo and Salach-Golyska [PSG95]). A ratio of the two aspect ratios (called henceforth normalized aspect ratio) was computed in each trial. This ratio was used as a measure of the accuracy of the percept.

In each session containing 75 trials, 25 different objects were displayed with different rotation angles in a random order. Each object was displayed exactly three times and the rotation angles were uniformly distributed from 40 to 70 degrees. The results from the 75 trials were grouped into three sets corresponding to the magnitude of rotation: 40-50 deg, 50-60 deg, and 60-70 deg. The average normalized aspect ratio and the standard deviation of this ratio were computed from these sets.

2.2 Results and Discussion

Figures 2.3 - 2.5 show the results. The ordinate shows the normalized aspect ratio. Thus, "one" on the ordinate represents an accurate percept. The abscissa shows the rotation angle. In Figures 2.3 and 2.4, each data point is a mean computed from 75 adjustments (3 subjects, 25 trials per subject). In Figure 2.5, the monoscopic curve is the average of all curves in Figure 2.3, and the stereoscopic curve is the average of all curves in Figure 2.4. The height of each symbol is equal to two times the standard error of the mean. The standard error is obtained by dividing the standard deviation of the normalized aspect ratio from N trials by the square root of N. The dashed line indicates the cosine of the rotation angle and it represents (approximately) the projected aspect ratio. In the absence of 3-D information, the data points would be expected to lie on this line.

Figures 2.3 and 2.4 show the performance of human shape reconstruction for all three types of cues (shading, wire frame and occluding contour) under monoscopic and stereoscopic viewing conditions respectively. Since the performance of all subjects was similar, we present here their average performance. Several results are observed. First, it is seen that the accuracy deteriorated as the rotation angle increased in both viewing conditions for each depth cue. The magnitude of this decrease is equivalent to several standard errors. This implies that this decrease is statistically highly significant. Second, it is seen that the differences among the curves representing individual depth cues are comparable to the standard errors. Furthermore, shading (strong depth cue) does not lead to more accurate performance as compared to occluding contour (weak depth cue). Thus, adding depth cues leads to small and statistically insignificant changes of performance, and, in fact these changes are not even systematic. This suggests that shape reconstruction (monocular and binocular) does not involve integration of depth cues.

Figure 2.5 shows the performance for both monoscopic and stereoscopic viewing conditions averaged over all stimuli and subjects. It is seen that human performance

with stereoscopic viewing is better than that of monoscopic viewing. This difference is much greater than the standard errors, which implies that it is statistically highly significant. Overall, in stereoscopic viewing, the shape reconstruction was quite accurate: the average error of the reconstructed aspect ratio was smaller than 10% (the standard deviation of the distribution of individual reconstructions was 15%). This good performance is consistent with the results of Durgin et al. [DPOR95] and Glennerster et al. [GRB96]. It is also worth pointing out that monoscopic performance, although worse than stereoscopic performance, is quite good. This means that the subjects were able to perform the shape matching task by means of monocular shape recognition from a single perspective image, rather than reconstruction. The analysis of this result is beyond the scope of this paper (see [PR92] and [Low87] for theories of monocular shape recognition).

In conclusion, the high accuracy of binocular shape reconstruction along with the absence of depth cues integration suggest that binocular shape reconstruction is not based on reconstructing the Euclidean structure. In the next section we present the results from testing the 8 point algorithm. As pointed out in the previous section, the 8 point algorithm reconstructs shape without reconstructing depth and distances. Therefore, this algorithm is a possible psychological model.

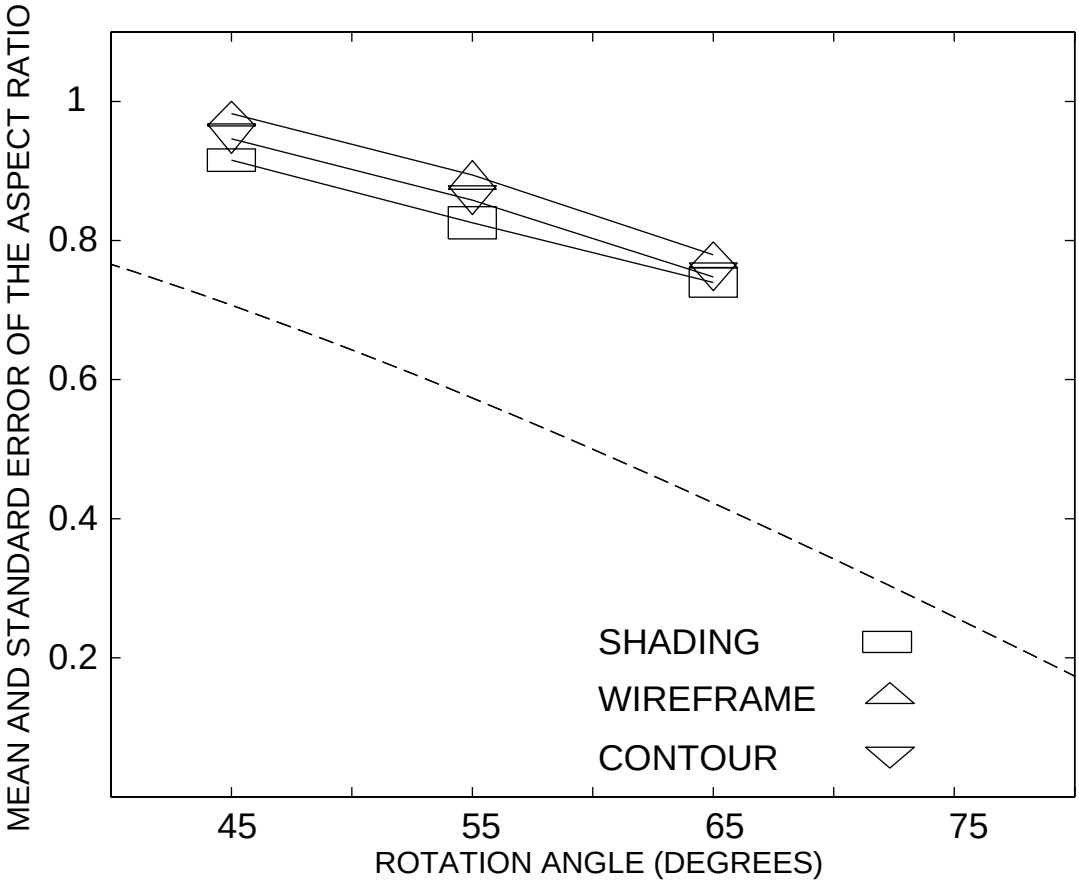


Fig. 2.3. Human shape reconstruction for different stimuli with monoscopic viewing (average from three subjects). The height of each symbol in each graph is equal to two times the standard error of the mean. The dashed line indicates the cosine of the rotation angle. Thus, this line represents (approximately) the projected aspect ratio. In the absence of 3-D information, the data points would lie on this line.

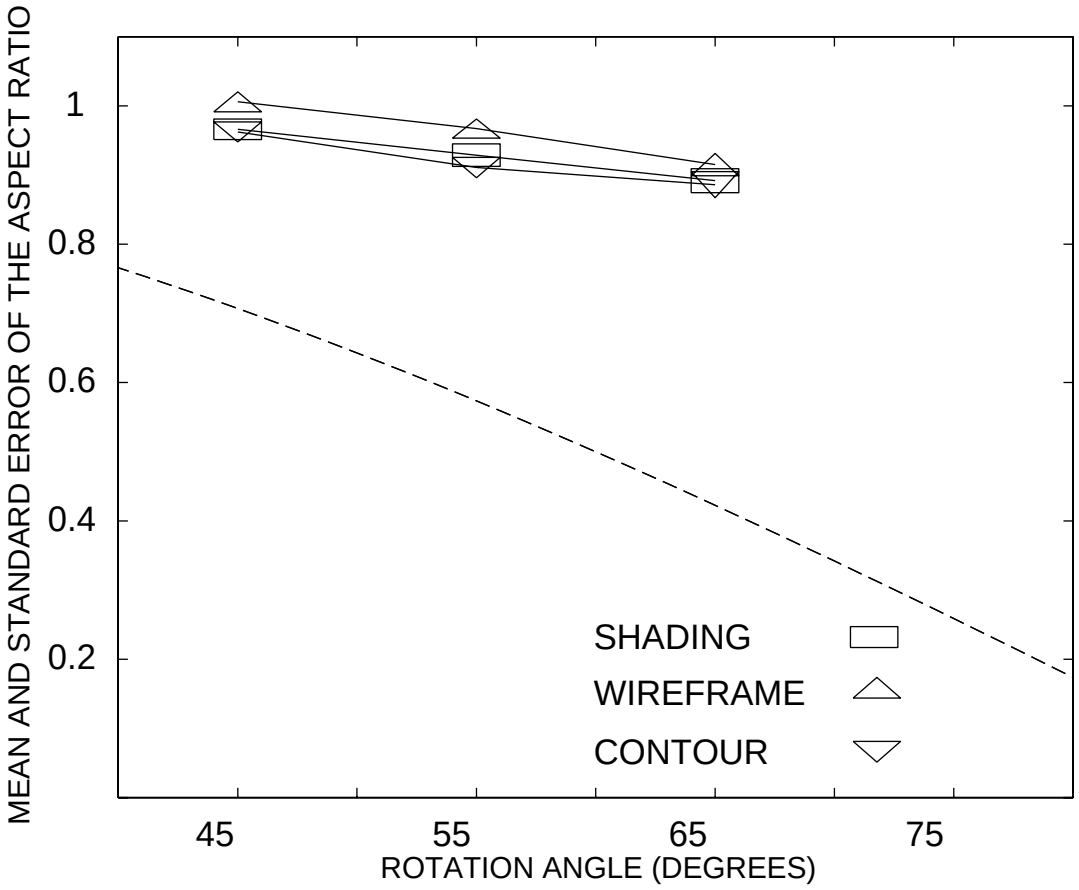


Fig. 2.4. Human shape reconstruction for different stimuli with stereoscopic viewing (average from three subjects). The height of each symbol in each graph is equal to two times the standard error of the mean. The dashed line indicates the cosine of the rotation angle. Thus, this line represents (approximately) the projected aspect ratio. In the absence of 3-D information, the data points would lie on this line.

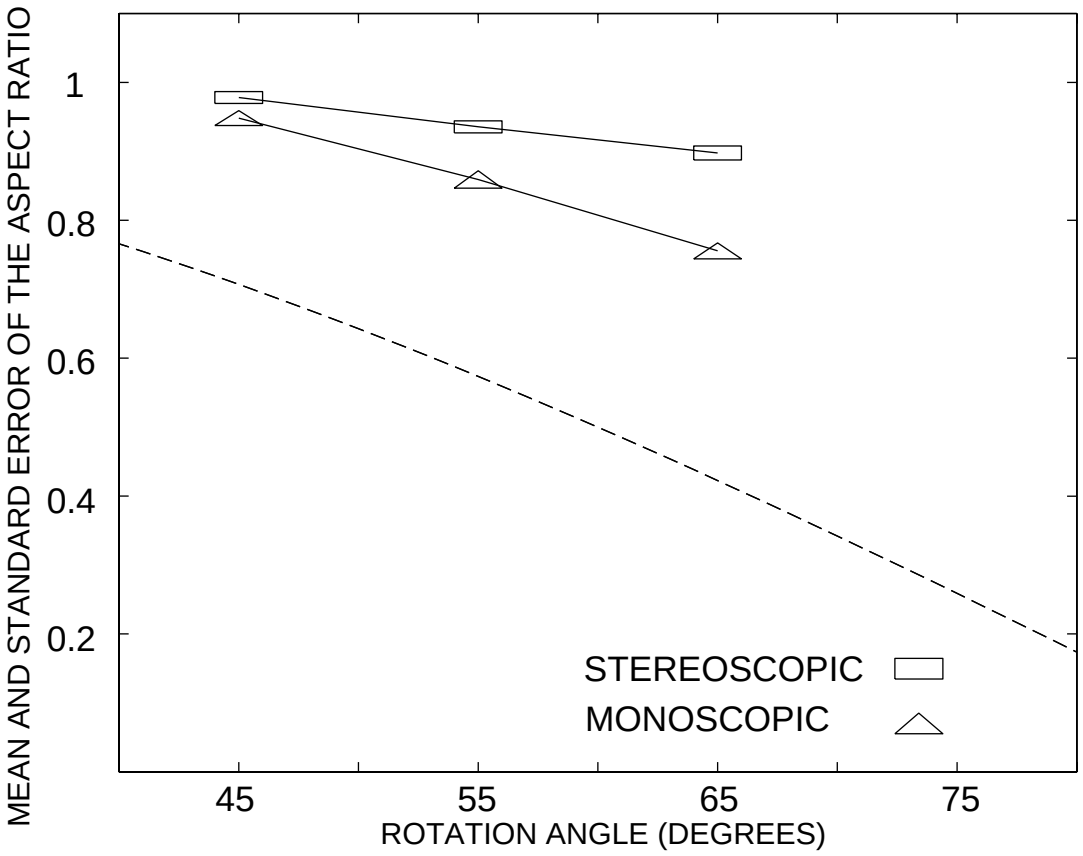


Fig. 2.5. Human shape reconstruction for different stimuli with average results of Figure 2.3 and 2.4. The height of each symbol in each graph is equal to two times the standard error of the mean. The dashed line indicates the cosine of the rotation angle. Thus, this line represents (approximately) the projected aspect ratio. In the absence of 3-D information, the data points would lie on this line.

3. COMPUTER ALGORITHMS FOR BINOCULAR SHAPE RECONSTRUCTION

3.1 Testing the 8-Point Algorithm as a Psychological Model

The purpose of this simulation experiment was to evaluate the shape reconstruction performance of the 8 point algorithm, and to examine the effect of a number of parameters on the reconstruction. The results of this simulation experiment will be compared to the results of our psychophysical experiment reported in Section 2. The viewing parameters and the procedure used in this simulation experiment were chosen to be similar to those in the psychophysical experiment. However, the stimuli were different. Recall that the 8 point algorithm is applied to images of distinctive points. Therefore, this simulation experiment involved a set of points in 3-D rather than cylinders (as in psychophysical experiment), and the aspect ratio was computed from three points (a triangle).

It is important to note that human performance in binocular judgments about aspect ratio of such simple stimuli as triangles is not worse than the performance we measured in this study for cylinders [PS98]. Therefore, this simulation experiment is likely to provide an adequate comparison of the performance of the 8 point algorithm to the performance of human subjects.

3.1.1 Stimuli

One hundred different objects were generated for each data point. Each object consisted of 9 or 36 random points generated within a unit cube with a constraint that 3 points formed a randomly shaped isosceles triangle whose height was vertical. The base of the triangle was horizontal and orthogonal to the line of sight. This triangle

was used to estimate the aspect ratio of the reconstructed object (see procedure below). The viewing parameters were as follows:

1) **Viewing distance:** the objects were placed at the distances of 4, 8 or 16 from the center of projection. These distances corresponded to vergence angles of 8, 4 and 2 degrees. Recall that in the psychophysical experiment the vergence was 8 degrees.

2) **Size:** three different sizes were used. They were obtained by scaling the original object by 1, 1/2 or 1/4. Thus, for viewing distance of 4, the object's size, expressed as a fraction of the viewing distance, was 1/4, 1/8 and 1/16. In the psychophysical experiment, the object's size was about 1/6 of the viewing distance.

3) **Noise:** uncorrelated Gaussian noise was imposed on each image coordinate with zero mean and standard deviation in the range from 0.1% to 10% of the diameter of the objects' image. Note that 1.4% of noise for an individual point leads to about 2% of uncorrelated noise for the distance between two points. This level of noise is known to exist in the human visual system [Wat87] [DLN+90].

3.1.2 Procedure

In each trial the object was randomly rotated about the horizontal axis by an angle of 45, 55 or 65 degrees. The aspect ratio of the object was randomized in the same way as the aspect ratio of the comparison object in the psychophysical experiment (Section 2). Each object was used 100 times at each rotation angle (total 10,000 trials per rotation angle). The base and the height of the isosceles triangle were randomly generated from within a range of 0.2 to 1.0.

Reconstruction was performed by first estimating the extrinsic parameters from the essential matrix (see Section 1.1), and then computing the 3-D object points using the Mid-Point method [HS97] [RCF95].

The accuracy of the reconstruction was evaluated by computing a normalized aspect ratio of the reconstructed object. The mean of the ratio was calculated for each experimental condition. Note that the aspect ratio was computed only for the

isosceles triangle (3 points); however, all points were used to estimate the extrinsic parameters.

3.1.3 Results and Discussion

The results are shown in Figures 3.1 - 3.5. The ordinate in Figures 3.1 - 3.4 represents the averaged normalized aspect ratio. Thus "one" on the ordinate represents an accurate reconstruction. The abscissa represents the rotation angle. Each data point is a mean computed from 10,000 reconstructions.

Figure 3.1 shows the reconstruction performance for different object sizes. Each object had 9 points. The viewing distance was 4 and the noise was 2%. The reconstruction performance is quite poor; the average ratio ranges from 1 to 3, which corresponds to errors of up to 200%. The performance was worse for smaller object sizes. This can be explained as follows. If the object's size (more exactly, the range of the object in depth) is small relative to the viewing distance, then perspective projection becomes approximately equivalent to parallel projection. In the absence of noise in the images, this fact has no effect on the accuracy of reconstruction. However, if noise is present, it may overshadow the perspective effects. Since two images obtained from parallel (or orthographic) projection are not sufficient for unique reconstruction of a 3-D object [AB89] [KvD91], the reconstruction is likely to be less accurate in the case of small objects when image noise is present.

Figure 3.2 shows the effect of the viewing distance on shape reconstruction of an object consisting of 9 points. The size of the object was 1 and the noise level was 2%. The reconstruction performance is poor. Again, up to 200% errors are obtained. The accuracy of the performance decreases as the viewing distance increases. Although the viewing distance is not an essential parameter in shape reconstruction when there is no noise in the image, the viewing distance affects the accuracy of shape reconstruction in the presence of noise. When the distance of the object from the camera is large, then the vergence angle becomes small so that the differences between the left and right

images produced by the difference in the viewing directions are small as compared to differences produced by noise. As a result, shape reconstruction deteriorates.

Figure 3.3 shows the performance of the algorithm for different amounts of noise. In this experiment, the object's size was 1, the viewing distance was 4 and each object consisted of 9 points. The results show that the algorithm underestimates aspect ratios when the noise level is 0.1%, and overestimates aspect ratios when the noise level is larger. The performance does not vary greatly for noise levels larger than or equal to 0.7%. However, the overall performance is poor.

Figure 3.4 shows the performance of the algorithm for different numbers of object points: 9 vs. 36. The object's size was 1, the viewing distance was 4, and the noise level was 2%. It is shown that the performance with 36 points is slightly better than that with 9 points. Specifically, the mean aspect ratio decreased by about 13%.

Finally, consider the variability of the reconstruction. The average standard deviation of the normalized aspect ratio for the experiments presented in this section was 3.25 (i.e. 325%), which represents extremely poor precision. Figure 3.5 shows the frequency distribution of the normalized aspect ratio for one condition (45 degrees rotation angle, 2% noise, viewing distance of 4, object size of 1, and 9 object points). It is seen that the histogram is highly skewed with some reconstructions representing 900% of error.

The comparison of the simulation results (for the conditions that were similar to those used in the psychophysical experiment) to the results of human subjects in stereoscopic viewing (Section 2) shows that the performance of the 8 point algorithm is different from the performance of subjects. Specifically:

1) The 8 point algorithm consistently overestimates aspect ratios for noise levels 0.4 - 10%, whereas human subjects tend to underestimate aspect ratios (recall, that the noise in the human visual system is equivalent to about 1.4%).

2) The accuracy of the reconstruction by the 8 point algorithm is substantially worse than that of human subjects. In the worst case, the mean error is about

200%. The mean error obtained from human subjects on the other hand, is smaller than 10%.

3) The precision of this algorithm is very poor. Specifically, the average standard deviation is 325%. This is much greater than the average standard deviation of the human reconstructions, which was 15%.

It can be concluded from these results that the 8 point algorithm is not an adequate model of human binocular shape reconstruction (a statistical analysis of the performance of the 8 point algorithm and human subjects is presented in Section 3.2.2). Specifically, the algorithm's performance is much worse than that of human subjects. We expect that an adequate model of human binocular shape perception will improve the performance of shape reconstruction. In the next section we will show how to modify the 8 point algorithm, by incorporating constraints that are present in human binocular vision, so that its performance becomes closer to the performance of human subjects.

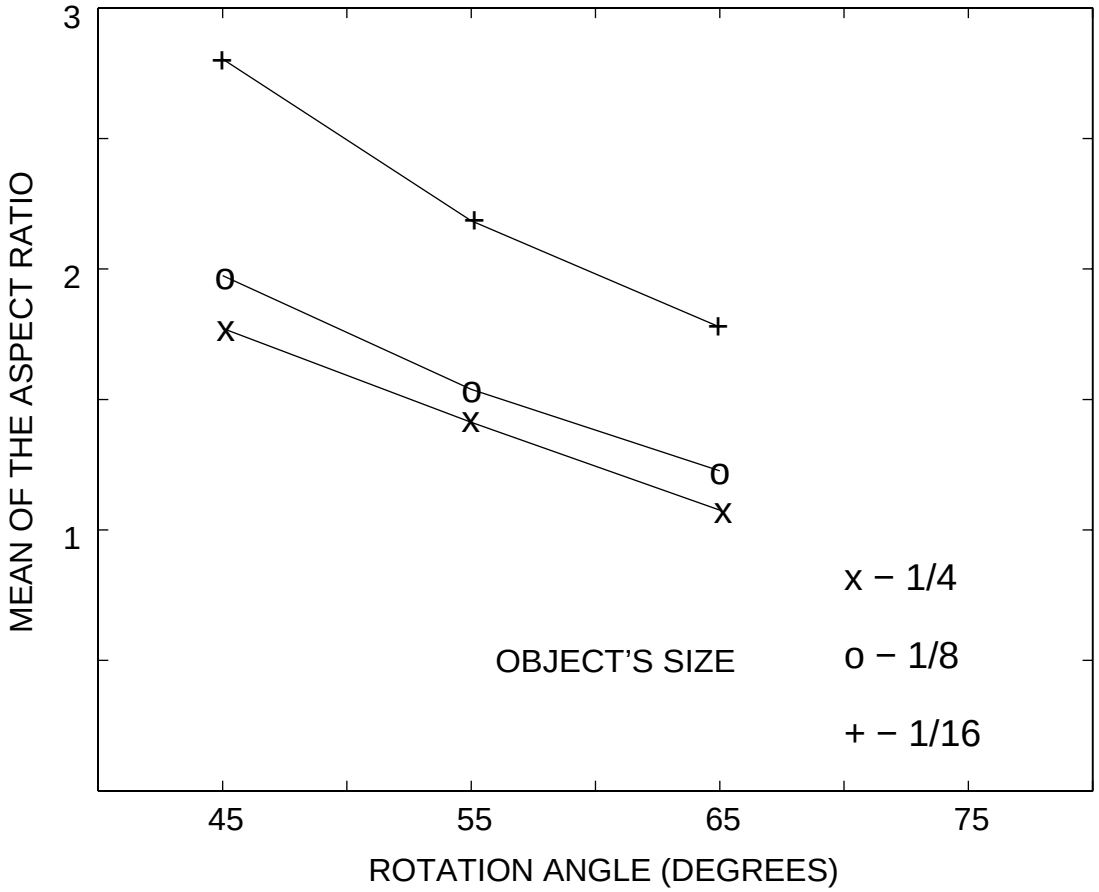


Fig. 3.1. The effect of object's size (expressed as a fraction of the viewing distance) shown in the performance obtained with the 8 point algorithm in reconstructing aspect ratio. The following viewing conditions were used: noise 2%, viewing distance 4, vergence 8 degrees, 9 object points, and object's size 1/4, 1/8, and 1/16 of the viewing distance.

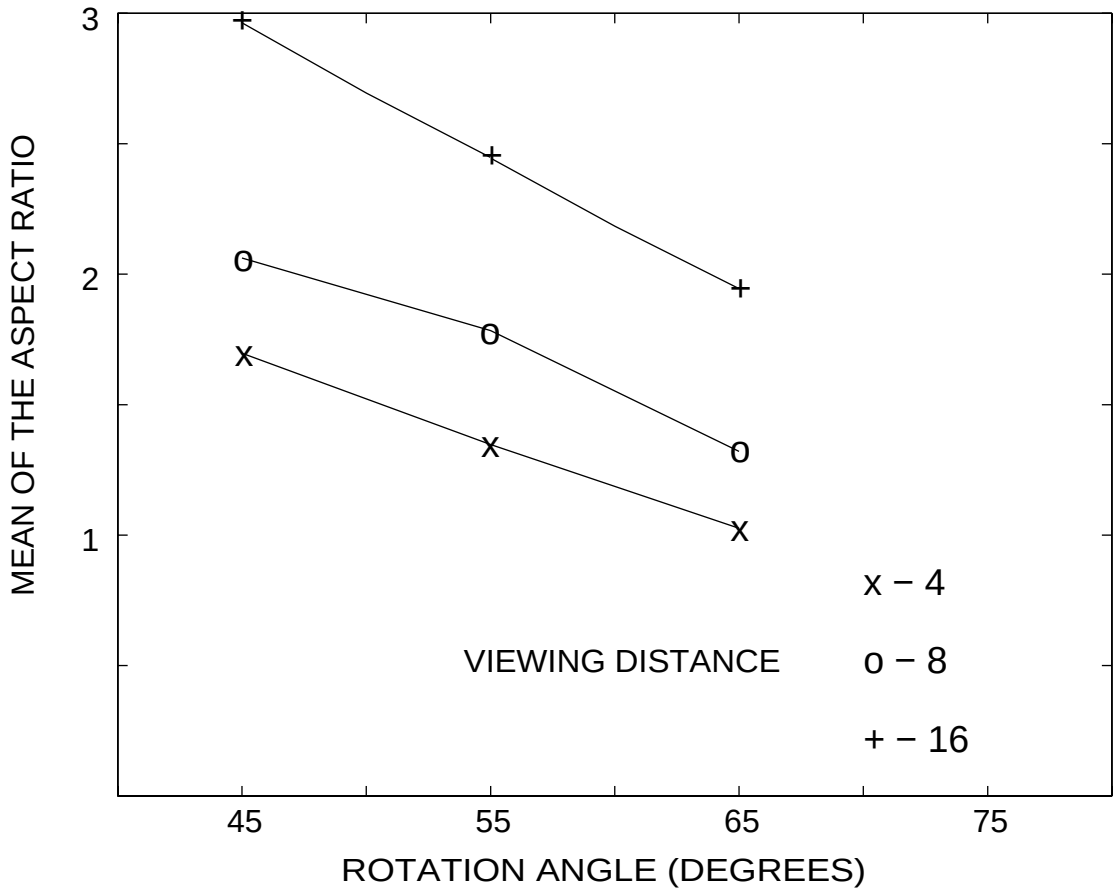


Fig. 3.2. The effect of viewing distance shown in the performance obtained with the 8 point algorithm in reconstructing aspect ratio. The following viewing conditions were used: noise 2%, object's size 1, 9 object points, and viewing distances of 4, 8, and 16 corresponded to vergence angles of 8, 4, and 2.

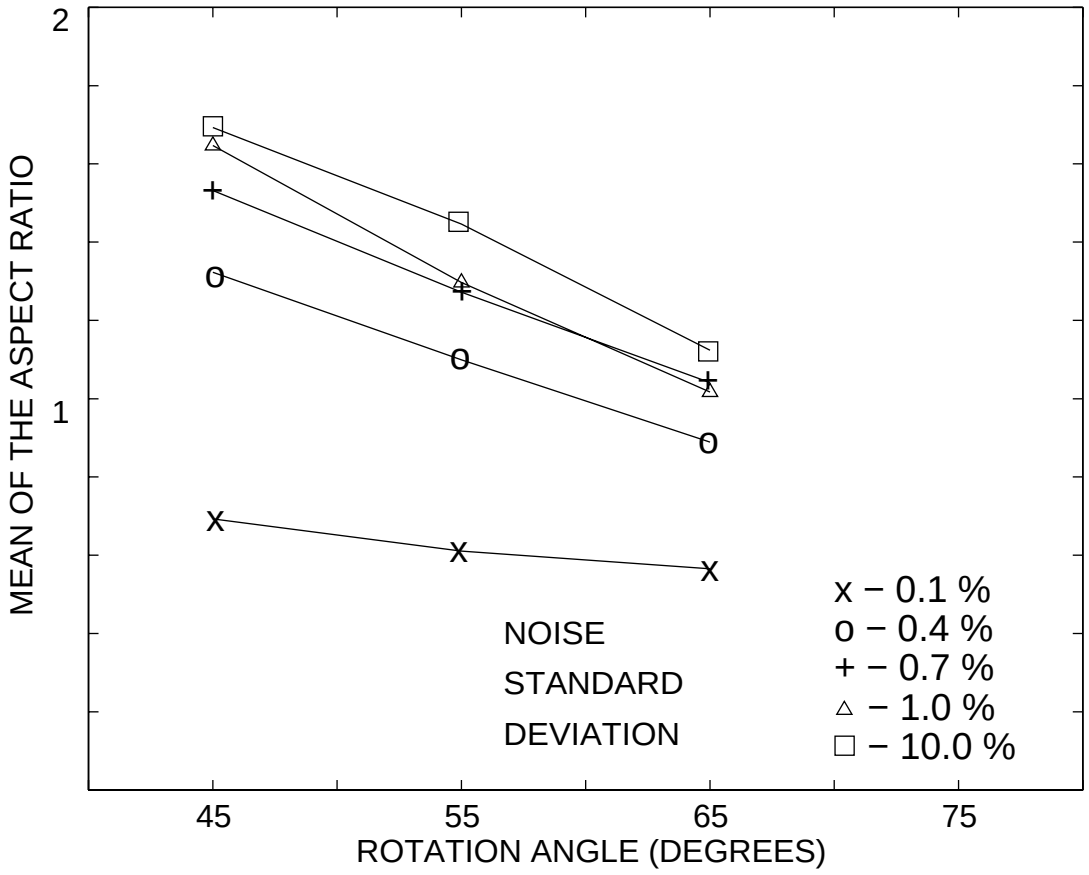


Fig. 3.3. The effect of noise shown in the performance obtained with the 8 point algorithm in reconstructing aspect ratio. The following viewing conditions were used: viewing distance 4, object's size 1, vergence 8 degrees, 9 object points, and noise 0.1%, 0.4%, 0.7%, 1.0%, and 10.0%.

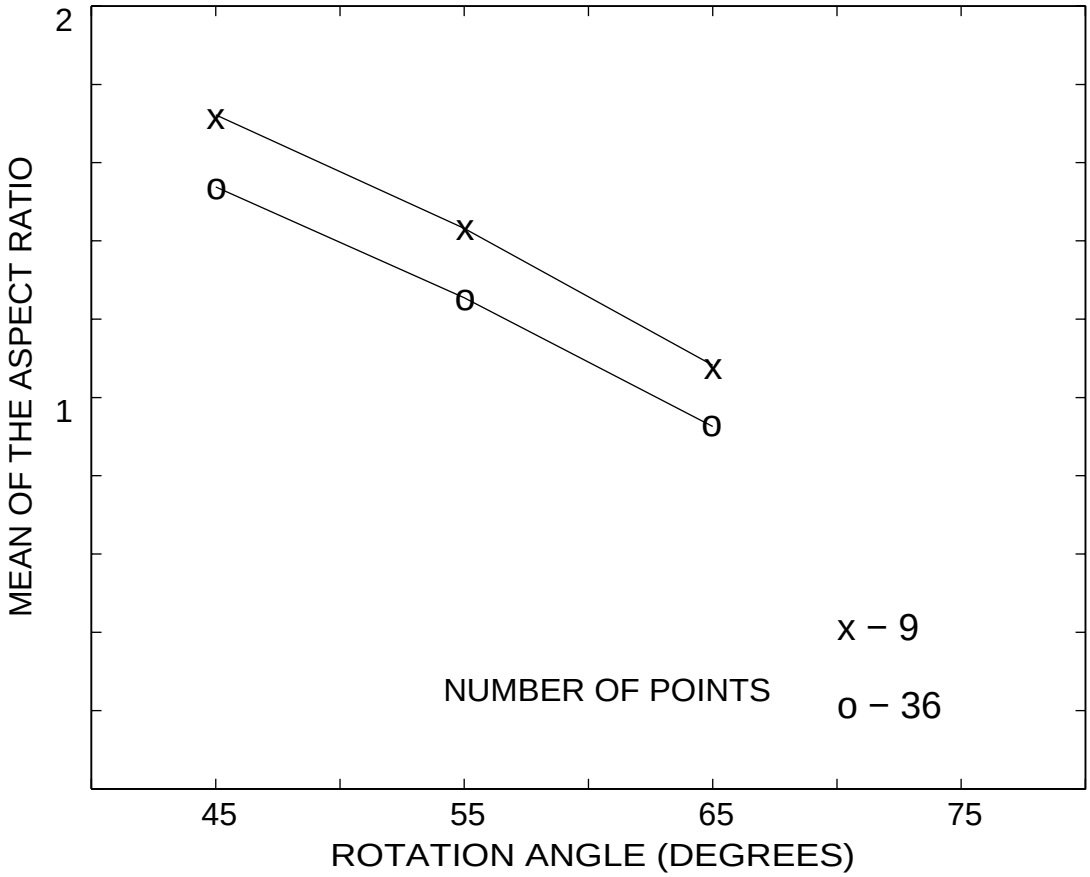


Fig. 3.4. The effect of number of points shown in the performance obtained with the 8 point algorithm in reconstructing aspect ratio. The following viewing conditions were used: noise 2%, viewing distance 4, object's size 1, vergence 8 degrees, and 9 and 36 object points.

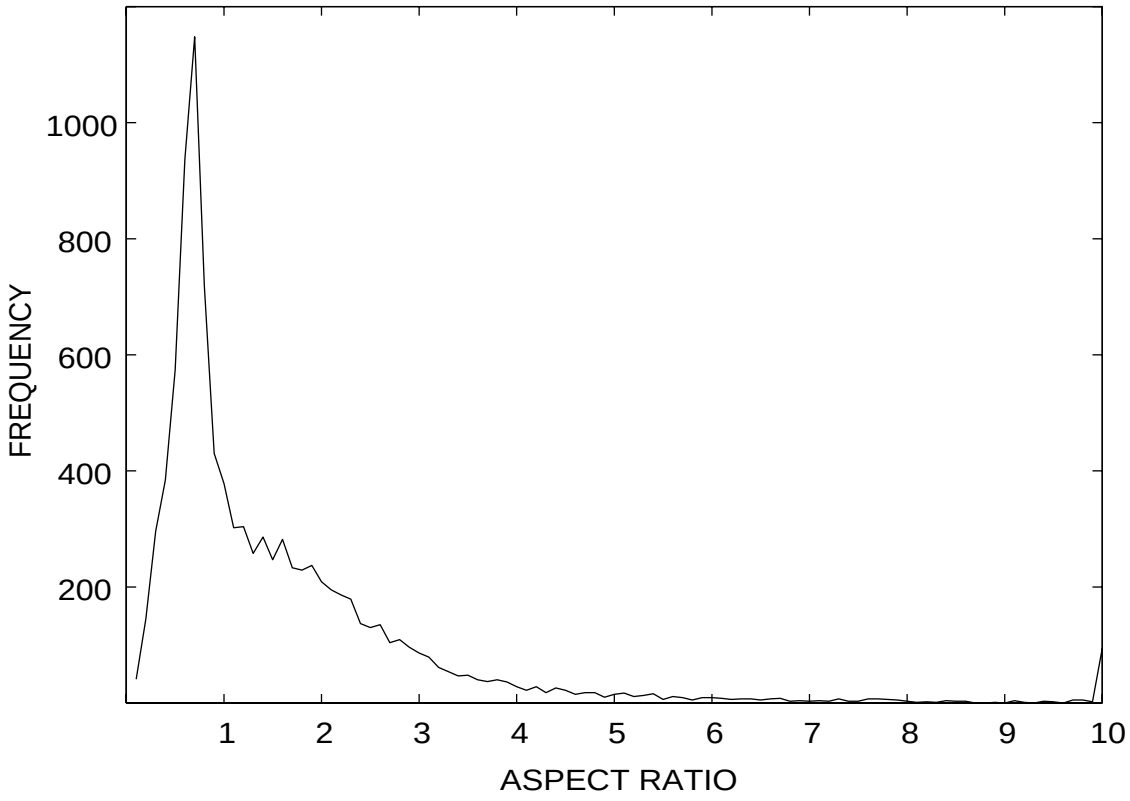


Fig. 3.5. Frequency distribution of the reconstructed aspect ratio obtained with the 8 point algorithm. The following viewing conditions were used: noise 2%, viewing distance 4, object's size 1, vergence 8 degrees, 9 object points, and rotation angle of 45 degrees.

3.2 Shape Reconstruction using Constraints of the Binocular Fixating System

In the 8 point algorithm, the rotation **R** and translation **T** represent any rigid motion in 3-D space. This means that in this algorithm there is no constraint on the relative orientation and position between the two cameras. However, human binocular viewing represents a fixating system whose geometry is more constrained. In this section we will show how to incorporate these constraints in the 8 point algorithm.

3.2.1 Constraints of the Human Binocular Vision

Theories of human binocular vision involve a number of constraints and assumptions about the intrinsic and extrinsic parameters (see Section 1.2). Here we assume that all intrinsic parameters are known and that some extrinsic parameters are zero. Specifically, we assume that:

1) the elevation of one eye relative to the other is zero;

2) the torsion of each eye is zero;

3) the two eyes are located at the same height in the head coordinate system (T_y is zero).

The cameras are free to rotate about the Y and Y′ axes as long as their optical axes intersect.

The remaining extrinsic parameters are assumed to be unknown. These constraints represent a binocular fixating system. We apply the above constraints to Equation 1.1 to yield the following equation:

$$\begin{bmatrix} u'_x & u'_y & 1 \end{bmatrix} \begin{bmatrix} 0 & -T_z & 0 \\ T_z & 0 & -T_x \\ 0 & T_x & 0 \end{bmatrix} \begin{bmatrix} a & 0 & b \\ 0 & 1 & 0 \\ -b & 0 & a \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix} = 0 \qquad (3.1)$$

where $a = cos\theta$ and $b = sin\theta$.

We scale the essential matrix by $1/T_x$, substitute $t = T_z/T_x$ and perform appropriate grouping to yield the following equation:

$$\begin{bmatrix} u'_y u_x & u'_y & -u'_x u_y \end{bmatrix} \begin{bmatrix} at+b \\ bt-a \\ t \end{bmatrix} = -u_y \tag{3.2}$$

The three unknown parameters $at + b$, $bt - a$, and t can be solved for by using a linear least squares technique for 3 (or more) pairs of corresponding image points. After these parameters are estimated, a, b and t are computed. Finally, the shape of a 3-D object is reconstructed using the Mid-Point method.

3.2.2 Simulation Experiments

The binocular fixating algorithm was tested using the same stimuli and procedure as described in Section 3.1. The results are shown in Figure 3.6 - 3.10.

First, we consider the reconstruction performance for different object sizes. Figure 3.6 shows that the performance of the binocular fixating algorithm is quite good; the mean reconstructed ratios range from 0.75 to 0.95. The performance slightly deteriorates as the object size decreases. Figure 3.7 shows that the accuracy of shape reconstruction decreases as the viewing distance increases. Figure 3.8 shows that the mean reconstruction is quite accurate even when the noise level is 10%. Figure 3.9 shows that the performance is slightly better when more points are used. Specifically, the mean aspect ratio increased by about 6%.

Finally, consider the variability of the reconstruction. The average standard deviation of the normalized aspect ratio from all the experiments was about 0.80 (i.e. 80%). Figure 3.10 shows the frequency distribution of the normalized aspect ratio for one condition (the same one that was used to produce Figure 3.5). It is seen that the histogram is skewed with some reconstructions representing 300% of error. The ratios, however, are substantially closer to one as compared to the ratios computed by the 8 point algorithm.

Clearly, the reconstruction accuracy and precision of the binocular fixating algorithm are substantially better than those of the 8 point algorithm. Specifically, the worst mean error of the binocular fixating algorithm is about 50% and the average standard deviation is about 80%, while the worst mean error of the 8 point algorithm is about 200% and the average standard deviation is about 325%.

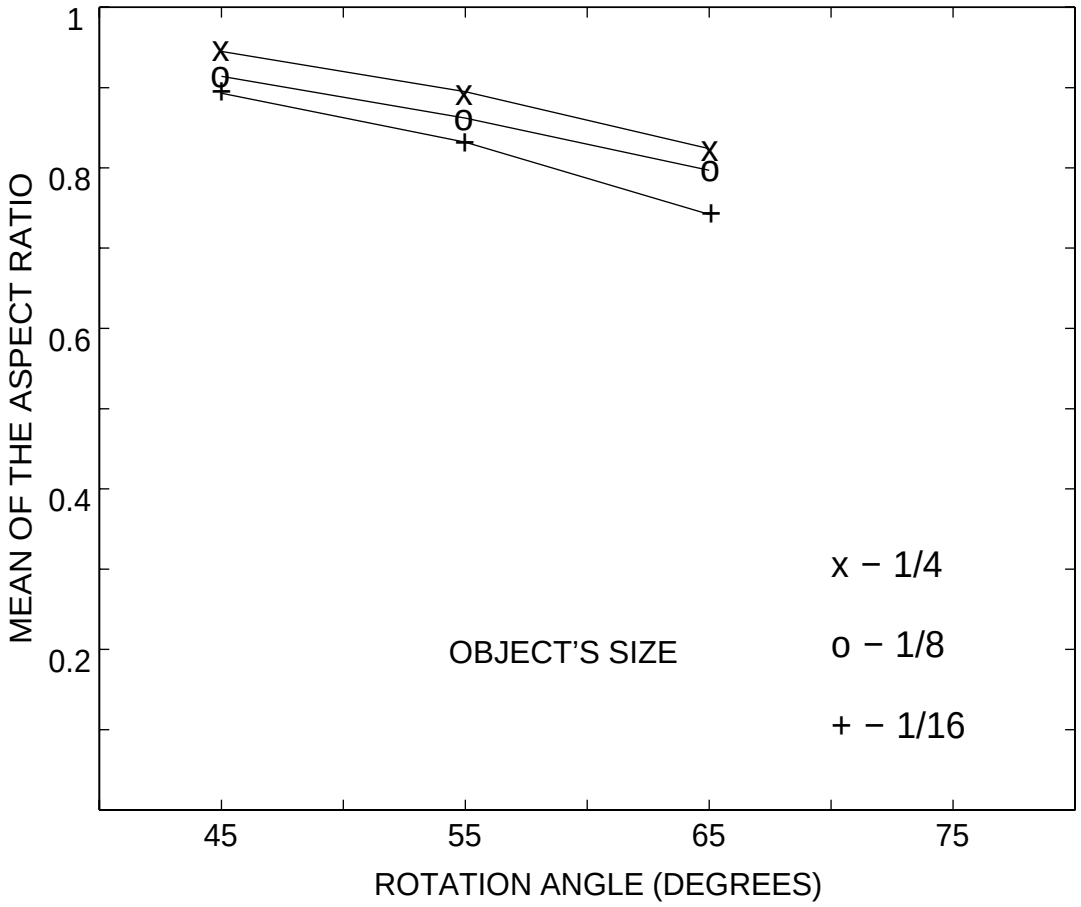


Fig. 3.6. The effect of object's size (expressed as a fraction of the viewing distance) shown in the performance obtained with the binocular fixating algorithm in reconstructing aspect ratio. The following viewing conditions were used: noise 2%, viewing distance 4, vergence 8 degrees, 9 object points, and object's size 1/4, 1/8, and 1/16 of the viewing distance.

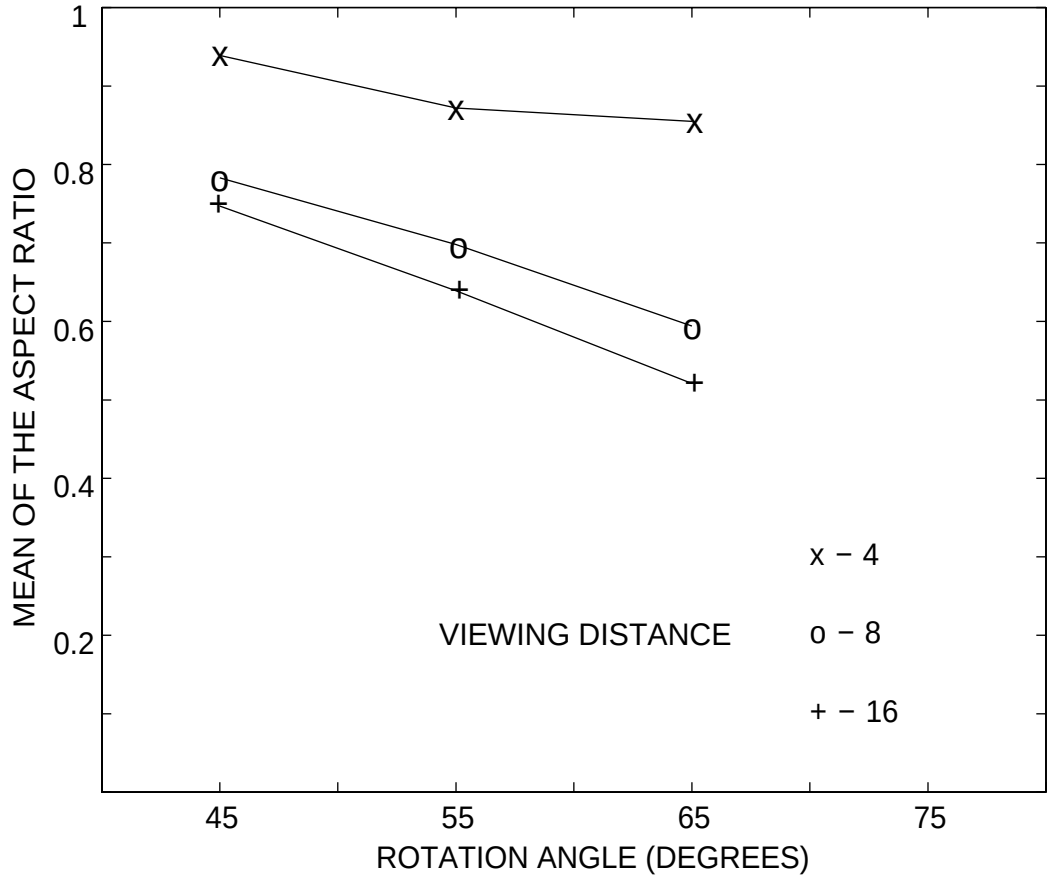


Fig. 3.7. The effect of viewing distance shown in the performance obtained with the binocular fixating algorithm in reconstructing aspect ratio. The following viewing conditions were used: noise 2%, object's size 1, 9 object points, and viewing distances of 4, 8, and 16 corresponded to vergence angles of 8, 4, and 2.

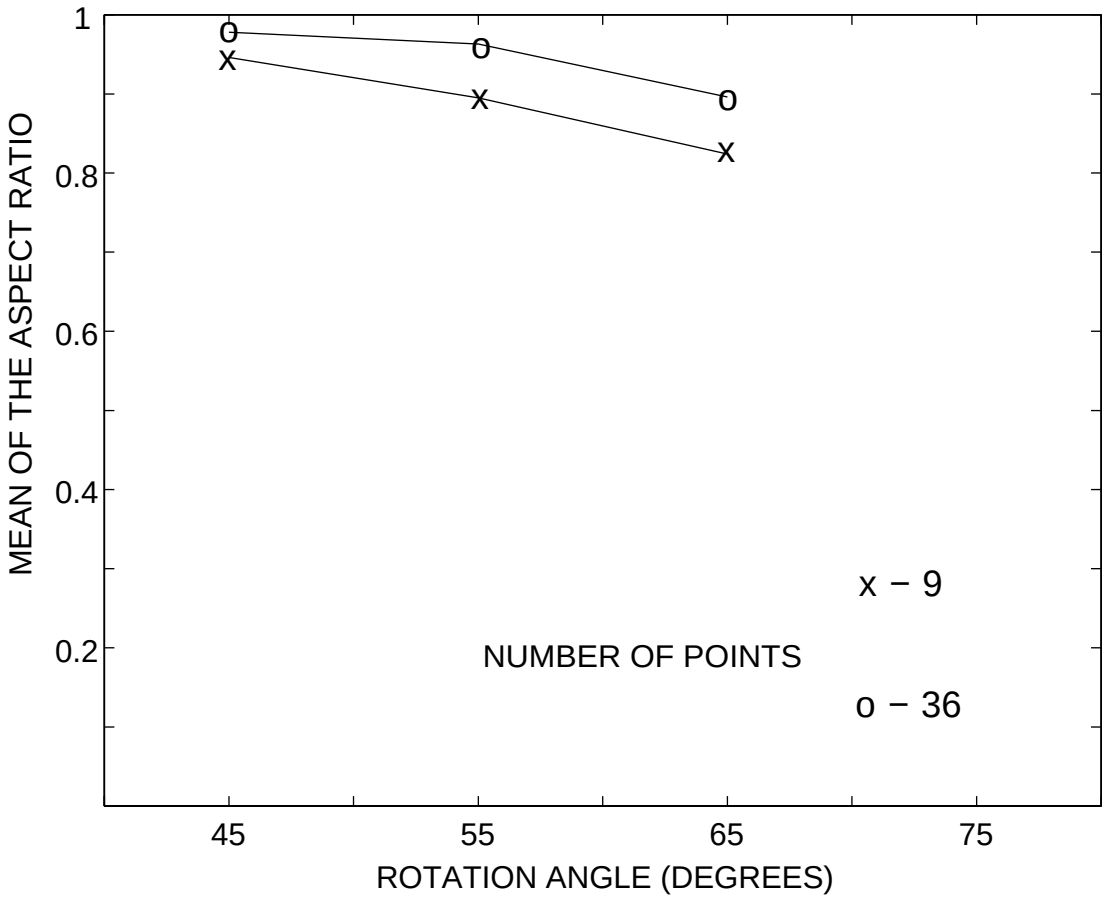


Fig. 3.8. The effect of noise shown in the performance obtained with the binocular fixating algorithm in reconstructing aspect ratio. The following viewing conditions were used: viewing distance 4, object's size 1, vergence 8 degrees, 9 object points, and noise 0.1%, 0.4%, 0.7%, 1.0%, and 10.0%.

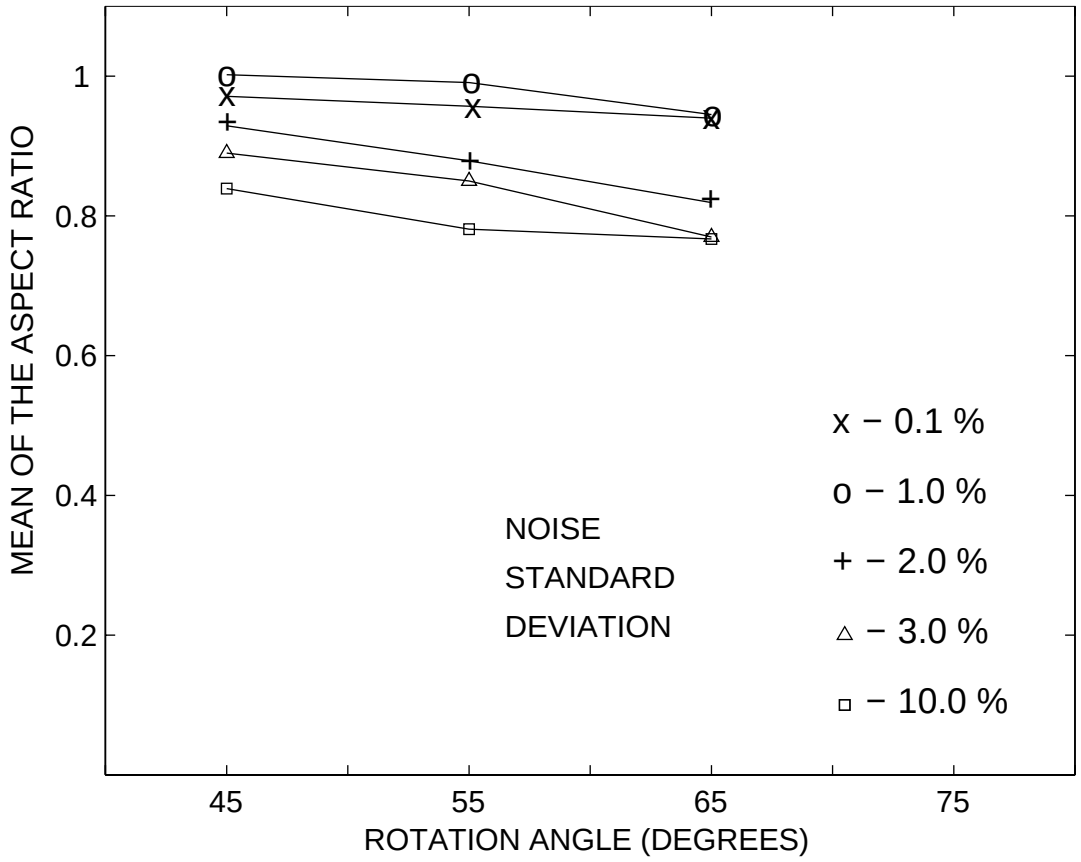


Fig. 3.9. The effect of number of points shown in the performance obtained with the binocular fixating algorithm in reconstructing aspect ratio. The following viewing conditions were used: noise 2%, viewing distance 4, object's size 1, vergence 8 degrees, and 9 and 36 object points.

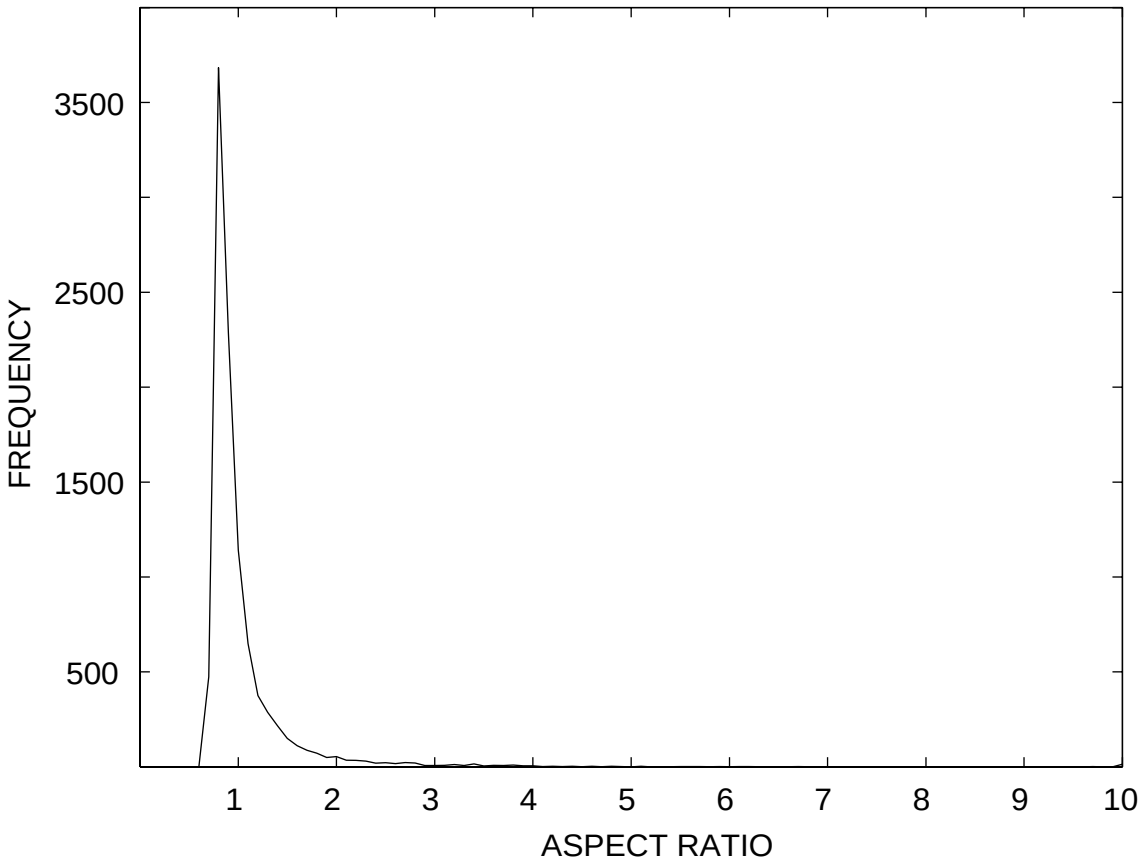


Fig. 3.10. Frequency distribution of the reconstructed aspect ratio obtained with the binocular fixating algorithm. The following viewing conditions were used: noise 2%, viewing distance 4, object's size 1, vergence 8 degrees, 9 object points, and rotation angle of 45 degrees.

3.3 Comparison of the performance obtained with the 8 point algorithm, binocular fixating algorithm, and human subjects

To compare the performance of both algorithms and human subjects, we plotted the results of the binocular fixating algorithm, the 8 point algorithm, and the human subjects on one graph (Figure 3.11). The simulation was done using viewing distance of 6, object size 1 and vergence 8 degrees. These values correspond to the viewing conditions used in the psychophysical experiment (Section 2). The noise level is 1.4% because this level is comparable to the noise in the human visual system. The curve representing human performance is identical to the stereoscopic curve in Figure 2.5.

Figure 3.11 clearly shows that the average performance of the binocular fixating algorithm is not very different from the performance of human subjects (although, this difference is statistically significant because the curves are separated by up to 10 standard errors). On the other hand, the average performance of the 8 point algorithm is substantially different from the performance of human subjects.

Note that the distributions of the aspect ratio estimates have long tails (Figure 3.5 and Figure 3.10). Thus, using the means as our statistics might have resulted in biased evaluations of the performance of the two algorithms. To remove this bias, we plotted the median of the normalized aspect ratio for both algorithms (Figure 3.12). It is seen that the performance of the 8 point algorithm as evaluated by a median, has improved, although it is still significantly different from the performance of human subjects (the curves are separated by 5-20 standard errors). Performance of the binocular fixating algorithm, on the other hand, deteriorated slightly.

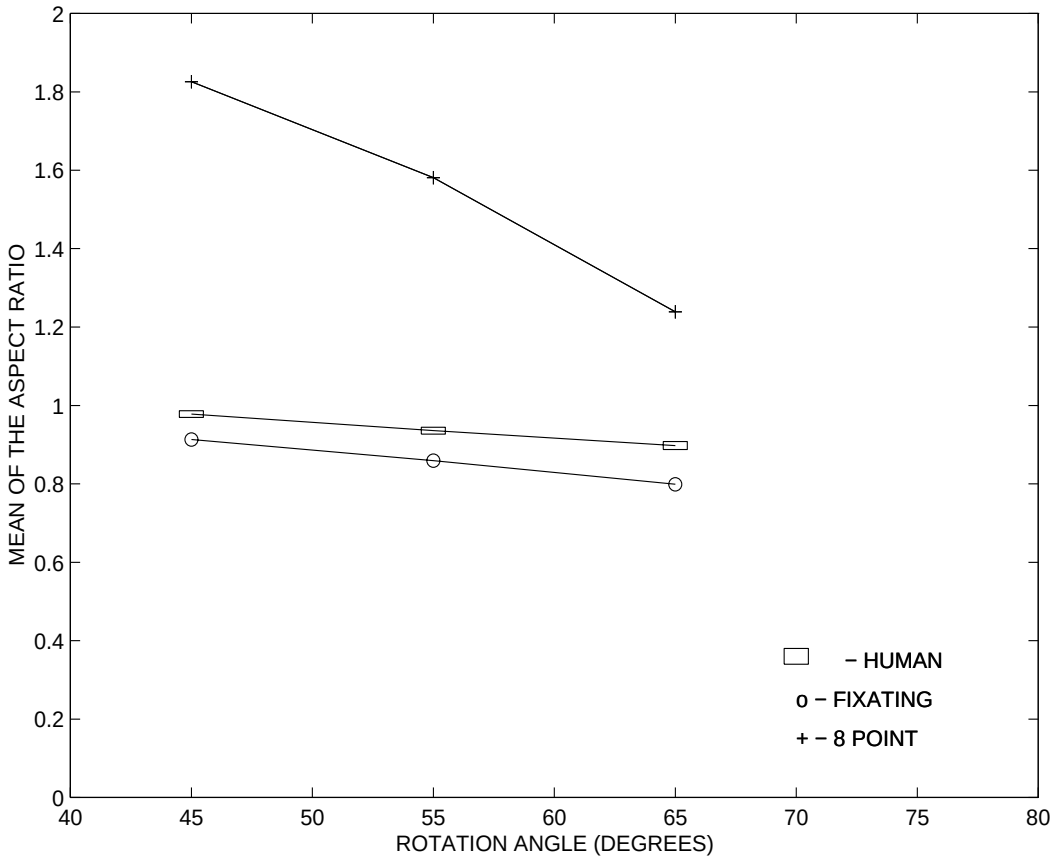


Fig. 3.11. Comparison of binocular reconstruction by human subjects, the 8 point algorithm, and the binocular fixating algorithm. The data points for human subjects represent average results of the three subjects in stereoscopic conditions. The height of each retangle is equal to two times the standard error of the mean. The data points for the two algorithms represent the mean of the reconstructions. The two were tested for objects consisting of 9 points, viewing distance 6, size 1, vergence 8 degrees and noise of 1.4%.

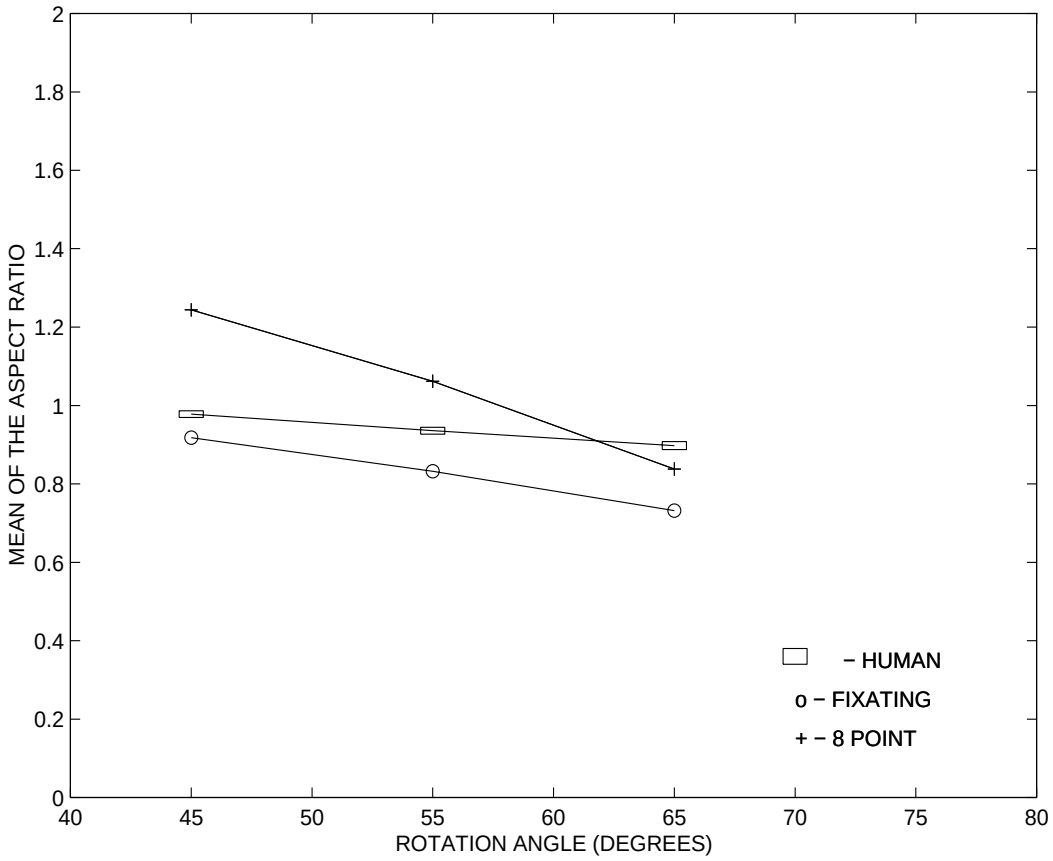


Fig. 3.12. Comparison of binocular reconstruction by human subjects, the 8 point algorithm, and the binocular fixating algorithm. The data points for human subjects represent average results of the three subjects in stereoscopic conditions. The height of each retangle is equal to two times the standard error of the mean. The data points for the two algorithms represent the median of the reconstructions. The two algorithms were tested for objects consisting of 9 points, viewing distance 6, size 1, vergence 8 degrees and noise of 1.4%.

To further evaluate the psychological plausibility of both algorithms, we analyze the variability of their reconstructions. Specifically, we use the F test [HWJ88] to compare the variance of the performance of both algorithms to that of human subjects. A null hypothesis is that the variances of two populations are the same. A criterion for rejecting this hypothesis is obtained based on a level of significance (probability of incorrectly rejecting the null hypothesis), and the number of trials or degrees of freedom (dof). A ratio of the two variances (F ratio) is computed, and then tested using the criterion for rejection. Here we use a conservative significance level of 0.01.

We first consider the 8 point algorithm. The F ratios (σ^2_{8point} / σ^2_{human}) are shown in Table 3.3. The dof of the numerator is 224 (3 subjects, 3 depth cues, 25 trials each yields a total of 225 - 1), and the dof of the denominator is 9999 (100 object, 100 trials each yields a total of 10000 - 1). The critical value for rejection is 1.28 [HWJ88]. The null hypothesis is rejected if the F ratio exceeds this value.

Table 3.1 F test results for σ^2_{human} and σ^2_{8point}.

Rotation Angle (degrees)	45	55	65
σ^2_{human}	0.0231	0.0213	0.0203
σ^2_{8point}	4.3956	5.7044	6.7752
F Ratio	188.6792	270.2703	344.8276

The hypothesis that σ^2_{human} is equal to σ^2_{8point} is rejected for all three rotation angles. It is clear that the performance of the 8 point algorithm and human subjects are statistically very different. Therefore, it can be concluded that the 8 point algorithm is not an adequate model of human binocular shape reconstruction.

We use the same test to compare the variance of the performance of the binocular fixating algorithm ($\sigma^2_{fixating}$) and human subjects (σ^2_{human}). The results are shown in Table 3.3.

The hypothesis that σ^2_{human} is equal to $\sigma^2_{fixating}$ is rejected for all three rotation angles. Thus, although the average performance of the binocular fixating algorithm is similar to that of human subjects, this algorithm is not a plausible psychological model because the variability of this algorithm is very different from that of human subjects. As seen in Table 3.3, $\sigma^2_{fixating}$ is greater than σ^2_{human} by a factor of 18 - 40. If the binocular fixating algorithm were used by the human visual system, then the human subjects would have to perform 18 - 40 reconstructions in order to obtain such high precision. It is quite possible that within a single trial, which lasted about 15 - 20 seconds, the subject did perform more than one reconstruction. However, it does not seem likely that the subject performed up to 40 reconstructions. This means that the binocular fixating algorithm presented here should be modified in order to improve its stability. This improvement is possible by incorporating prior constraints [CPS99].

Table 3.2 F test results for σ^2_{human} and $\sigma^2_{fixating}$.

Rotation Angle (degrees)	45	55	65
σ^2_{human}	0.0231	0.0213	0.0203
$\sigma^2_{fixating}$	0.4289	0.7329	0.8098
F Ratio	18.5529	34.3643	39.8406

Another issue which has to be pointed out is that the new algorithm operates on distinctive points, whereas human subjects perform binocular shape reconstruction equally well for wire objects (with distinctive points), as for shaded objects and objects represented by occluding contour (without distinctive points). This fact suggests that our algorithm should be generalized so that it can be applied to objects that do not have distinctive points.

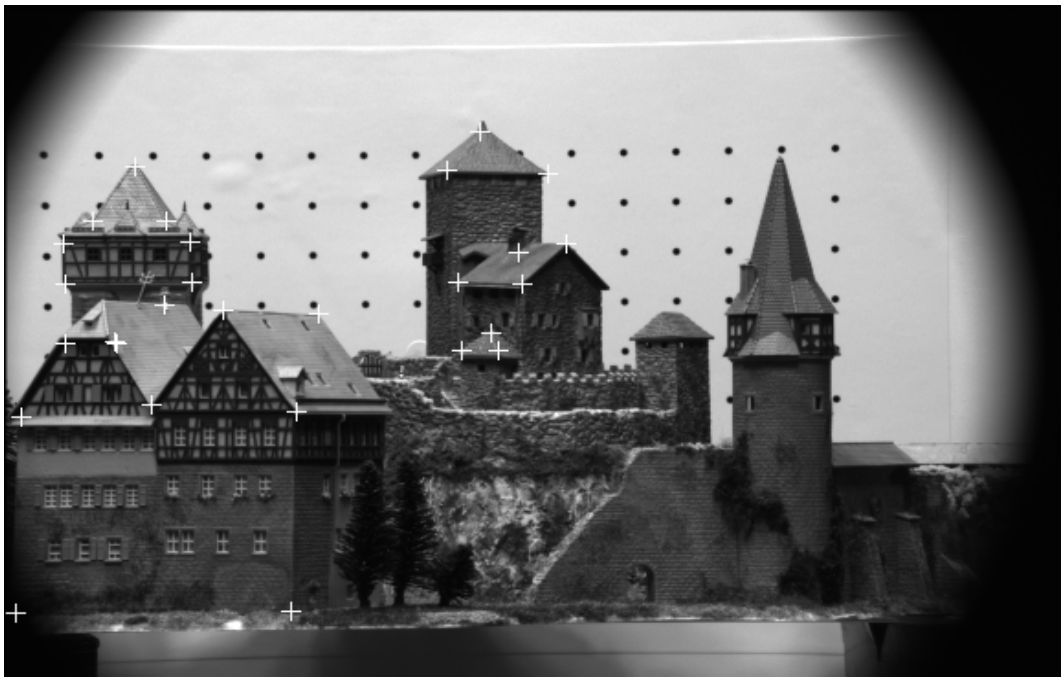
3.4 Experiments with Real Stereo Images

Finally, the two algorithms were tested using 10 pairs of real stereo images obtained from the Calibrated Imaging Laboratory at Carnegie-Mellon University. These 10 pairs of images were combinations of 5 images taken from different positions. Figure 3.13 shows a pair of these images. 28 points were marked on each image, and they were reconstructed using the binocular fixating algorithm and the 8 point algorithm. The accuracy of the reconstruction of the similarity structure was evaluated by:

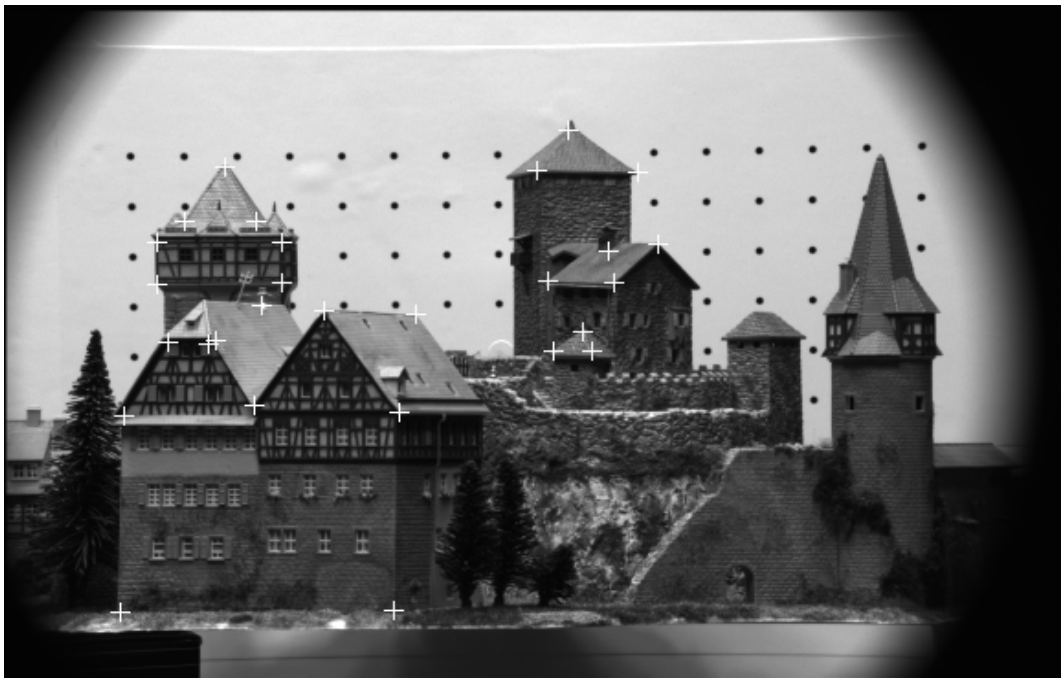
$$\mathbf{ratio}_i = \frac{\frac{\|\mathbf{U}'_i,\mathbf{U}'_k\|}{\|\mathbf{U}'_j,\mathbf{U}'_k\|}}{\frac{\|\mathbf{U}_i,\mathbf{U}_k\|}{\|\mathbf{U}_j,\mathbf{U}_k\|}} \tag{3.3}$$

where i $\neq$ j $\neq$ k $\neq$ i, $\|\mathbf{U}_i, \mathbf{U}_j\|$ denotes the Euclidean distance between the points $\mathbf{U}_i$ and $\mathbf{U}_j$, $\mathbf{U}_i$ represents the i^{th} original 3-D point and $\mathbf{U}'_i$ represents the i^{th} reconstructed point.

We plotted the frequency distribution of the reconstructed ratios of all marked points using the two algorithms (Figure 3.14). Again, a perfect reconstruction is represented by $\mathbf{ratio}_i$ of one. It is seen that most of the ratios produced by the binocular fixating algorithm are close to one. At the same time, the ratios produced by the 8 point algorithm are scattered in the range of 0.2 to 1.0. This shows that the performance of the binocular fixating algorithm is again better as compared to that of the 8 point algorithm.



(a)



(b)

Fig. 3.13. Left (a) and right (b) views of a castle. The points marked with "+" were used to test the 8 point algorithm and the binocular fixating algorithm.

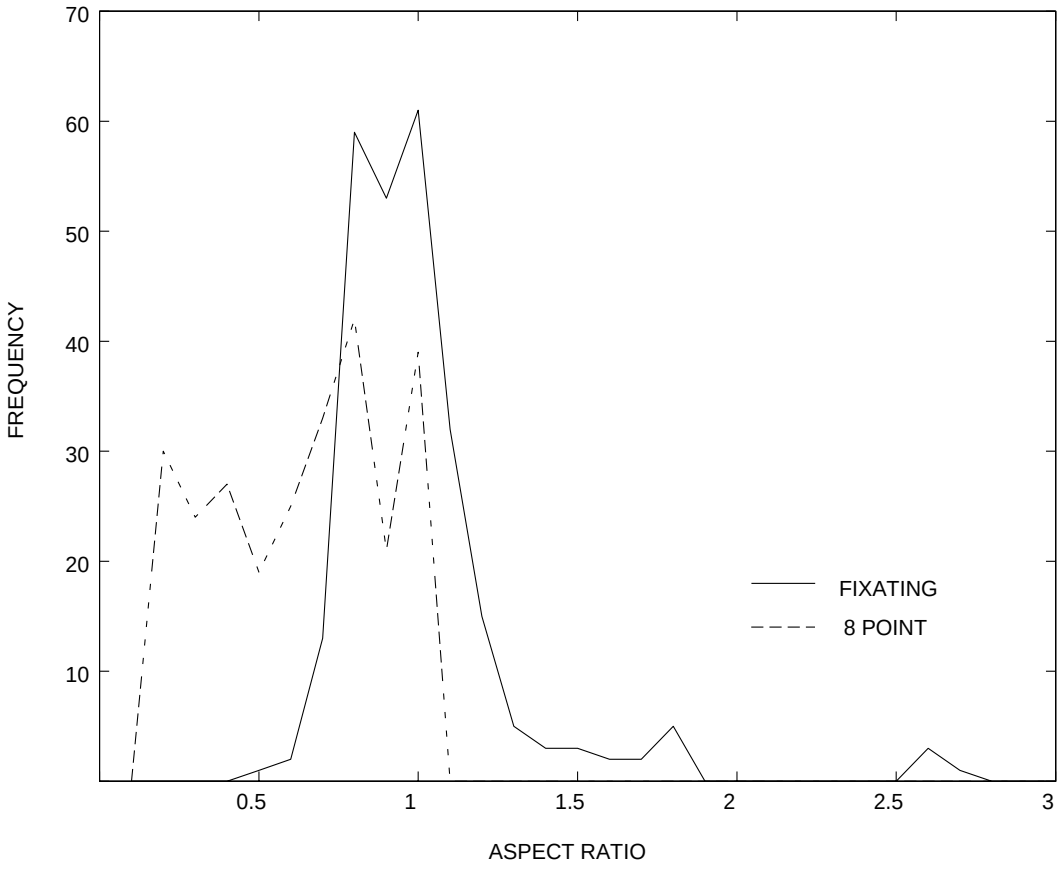


Fig. 3.14. Frequency distribution of the reconstructed ratios of distances for the images shown in Figure 3.13.

3.5 Summary

The 8 point algorithm has been widely used in computer vision research for shape reconstruction from two perspective views. Despite its popularity in computer vision applications, the 8 point algorithm has never been tested as a model of human binocular shape reconstruction. We compared the reconstruction performance of the 8 point algorithm to that of human subjects. We found that the performance of the 8 point algorithm is poor under a moderate amount of noise (the same amount of noise as that found in the human visual system), and it is quite different from that of human observers.

To improve the performance of the 8 point algorithm, we imposed the constraints of the human binocular system. The new algorithm can be used to compute the essential matrix with as few as three pairs of corresponding points. The results of the simulation experiments show that the binocular fixating algorithm is substantially more accurate and precise than the 8 point algorithm. Furthermore, the performance of the binocular fixating algorithm is similar to that of human subjects.

It is important to point out that this new algorithm can better account for existing psychophysical results on binocular shape reconstruction than prior theories. Specifically, this algorithm leads to accurate shape reconstruction for close viewing distances, and for scenes whose range in depth is not small as compared to viewing distance. This feature of the algorithm is consistent with recent results of Glennerster el at. [GRB96], Durgin el at. [DPOR95] and with our psychophysical results presented in this thesis. If the range in depth of the scene is small as compared to the viewing distance, shape reconstruction may lead to large errors. This feature is also consistent with prior psychophysical results [Joh91]. Finally, since our algorithm reconstructs similarity, but not Euclidean structure, it does not presuppose (as prior theories do) that the accuracy of distance perception must be better than that of shape perception. As pointed out in Section 1.2, existing psychophysical results show that shape is often perceived more accurately than distances.

4. THE ROLE OF GEOMETRICAL CONSTRAINTS IN SHAPE RECONSTRUCTION

It was shown in Chapter 3 that the performance of the binocular fixating algorithm is substantially better than that of the 8 point algorithm. However, it is still very unstable as compared to that of human subjects. This result is not surprising because binocular shape reconstruction is an inverse problem. As most inverse problems in natural sciences, binocular shape reconstruction is ill-posed and/or ill-conditioned [PT84]. A problem is well-posed if its solution (a) exists, (b) is unique, and (c) depends continuously on the initial data. Ill-posed problems fail to satisfy one or more of these criteria. A problem is ill-conditioned if its solution is unstable.

In order to illustrate how the binocular shape reconstruction problem can be solved, we first assume that:

1) the intrinsic parameters of the two cameras (i.e. coordinates of the principal point focal length, pixel's aspect ratio and the skew parameter) are known exactly;

2) the extrinsic parameters (orientation and position) of one camera relative to the other are known exactly;

3) correspondences of points in the two images are known;

4) there is no noise in the images.

Under such assumptions, an object point $\mathbf{U}$ can be reconstructed as an intersection of the visual rays emanating from the image points $\mathbf{u}$ and $\mathbf{u}'$, and going through their corresponding centers of projection $\mathbf{O}$ and $\mathbf{O}'$. This intersection always exists and is unique. However, if any of the assumptions listed above is not satisfied (and in real life none is), the two visual rays do not intersect, which means that binocular shape reconstruction does not have a solution (i.e. the problem is ill-posed). A solution can be obtained by any approximation method. For example, if the correspondences

of the points are known, then one can estimate the object point by a point that is closest to the two visual rays (mid-point method) [HS97]. However, such solutions are very unstable (i.e. the problem is ill-conditioned). To overcome the instability a number of methods have been proposed. Some representative methods are described next.

4.1 Prior Computer Algorithms for Shape Reconstruction

The fact that human subjects can accurately and precisely recognize and/or reconstruct shapes suggests that the human vision system uses constraints that stabilize the perceptual solution. Two types of constraints are considered here: (a) system constraints; and (b) constraints of the geometrical properties of the objects. In Section 3 we showed that system constraints are useful but not sufficient. This fact suggests that accurate and reliable reconstruction must also involve constraints on the family of possible solutions. In other words one has to use *a priori* knowledge about the objects. There are two classes of methods that take such constraints into account: model-based methods (also known as parameter fitting methods) and standard regularization methods.

Model-based methods presuppose that the object to be reconstructed belongs to a class of known objects (models), which can be represented by a small number of parameters. Let $\mathbf{x}$ be a vector of unknown parameters of the object, $\mathbf{y}$ the image of $\mathbf{x}$, and $\mathbf{A}$ a linear transformation which maps $\mathbf{x}$ to $\mathbf{y}$ ($\mathbf{Ax} = \mathbf{y}$). The parameters are obtained by minimizing $\|\mathbf{Ax} - \mathbf{y}\|^2$. Several algorithms for binocular reconstruction have been proposed based on this approach [EW87] [HA89] [ML92]. These algorithms are called single-stage because they perform stereo matching (i.e. they solve the correspondence problem) and object reconstruction in one step. One problem with this class of methods is that it assumes that the object to be reconstructed satisfies a predefined model exactly. This is too restrictive. To avoid this problem, one can use standard regularization methods [TA77], in which constraints are expected to be satisfied only approximately.

In standard regularization theory one finds $\mathbf{x}$ that minimizes:

$$E_R(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \lambda\|\mathbf{P}\mathbf{x}\|^2 \tag{4.1}$$

where $\|\mathbf{P}\mathbf{x}\|$ is called the stabilizing functional (constraint). $\mathbf{P}$ is a linear combination of derivatives of the solution and, thus, represents smoothness constraint. λ is called the regularization parameter. Under the assumptions that $\mathbf{A}$ and $\mathbf{P}$ are linear operators, $E_R(\mathbf{x})$ is convex, and a global minimum exists. This formulation is equivalent to (approximately) restricting the solution space to generalized splines, whose order depends on the order of the stabilizer $\mathbf{P}$.

Standard regularization theory has been applied to shape reconstruction from shading, texture, motion, and binocular disparity [Gri81] [Gri85] [PTK85] [BPT88]. Smoothness constraint, which is used in standard regularization theory, is useful in modeling continuous surfaces; however, it is not appropriate in modeling surface discontinuities. Terzopoulos [Ter86] proposed a class of controlled-continuity stabilizers, in which a weighting function is used to control the smoothness over the continuity and discontinuity regions. The controlled-continuity constraints were used by March in his binocular vision algorithms [Mar88] [Mar89].

If either $\mathbf{A}$ or $\mathbf{P}$ is a nonlinear operator, then $E_R(\mathbf{x})$ may not be convex. In such a case, a search is used to obtain $\mathbf{x}$. However, the presence of local minima poses a challenge; since an exhaustive search is impractical, one has to use a constrainted search, such as steepest descent. But then the solution depends strongly on the starting point. One way to overcome this problem is to use a continuation method [WTK87] [Lec89]. This method involves finding minima for a sequence of approximations of $E_R(\mathbf{x})$. The method first finds a global minimum of the coarsest approximation of $E_R(\mathbf{x})$, and then uses it as the starting point of a descent search at the less and less coarse approximations of $E_R(\mathbf{x})$, until it reaches $E_R(\mathbf{x})$. This method does not necessarily yield a global minimum, but it has been demonstrated to be quite robust.

Examples of the application of this method to surface reconstruction can be found in Fua and Leclerc [FL95] [FL96].

A stochastic version of the standard regularization theory is formulated using the Bayes's theorem [MMP87]. In this formulation, a posteriori distribution $P(\mathbf{x}|\mathbf{y})$, which represents the probability of an object $\mathbf{x}$ given the image $\mathbf{y}$, is obtained from:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}$$

The term $P(\mathbf{y}|\mathbf{x})$ is called the likelihood and it measures how well the object $\mathbf{x}$ matches the image $\mathbf{y}$, and $P(\mathbf{x})$ refers to the prior model. In this way, we can solve the reconstruction problem by obtaining an estimate of $\mathbf{x}$ that maximizes the *a posteriori* probability (MAP). For a given set of measurements, the term $P(\mathbf{y})$ is a constant normalizing term and it can be ignored. It is clear that maximizing $P(\mathbf{y}|\mathbf{x})$ is equivalent to minimizing

$$E_B(\mathbf{x}) = -logP(\mathbf{y}|\mathbf{x}) - logP(\mathbf{x}) \tag{4.2}$$

Note that Equation 4.2 is analogous to Equation 4.1. Bayesian approaches implement the smoothness constraint by Markov random field (Bouman and Sauer [BS93]), Gaussian function (Matthies [Mat92]) or Browian motion (Belhumeur [Bel96]).

To summarize, the performance of prior computer algorithms that rely solely on binocular disparity is extremely unstable. Constraints of the viewing system can be used to obtain more stable solution; however, the solution is still very unstable as compared to the performance of human observers. This suggests that geometrical constraints of the objects must be used to restrict the family of solutions. To incorporate geometrical constraints into the reconstruction process, standard regularization theory can be used. Prior algorithms based on standard regularization theory used surface smoothness as the regularizing constraint. The performance of these algorithms was quite good when the surfaces were richly textured. However, it is not

clear how the smoothness constraint can be applied to objects that do not have textured or shaded surfaces. Clearly, one should use other types of constraints as well. A recent psychophysical experiment demonstrated the operation of other constraints in perception of shape from motion [PS99]. This experiment will be briefly described next.

4.2 Perception of Shape from Motion

The experiment tested the accuracy of human 3-D shape perception using stimuli having various constraints. All stimuli involved 16 points in 3-D - vertices of a polyhedron. In the main condition, these points were generated with the following constraints: the faces of the object were planar polygons with 4 or more vertices, and the object had one plane of symmetry. The object was represented by showing the edges of the polyhedron. Examples of stimuli are shown in Figure 4.1 (for more examples see: http://bigbird.psych.purdue.edu/shapedemo).

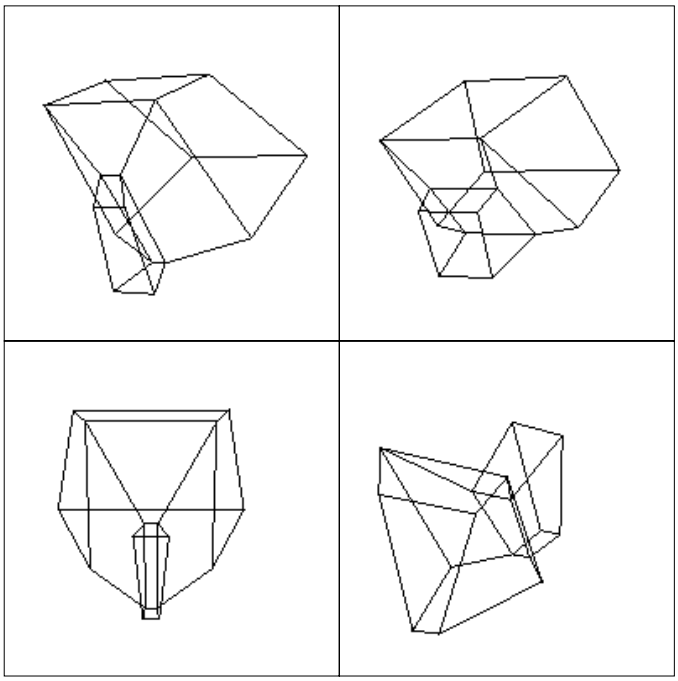


Fig. 4.1. Examples of stimuli used in the psychophysical experiments.

Nine conditions were tested. Each condition involved different number of constraints:

1) the polyhedron as shown in Figure 4.1;

2) vertices of the polyhedron without any edges;

3) polygonal line connecting the vertices in a random order;

4) partially non-planar, symmetric polyhedron. In this case, the planarity constraint of the faces was removed; however, the polyhedron had still a plane of symmetry. When a face is formed by 4 points, and it has a plane of symmetry, then the face is planar under the assumption that none of the points lie on the plane of symmetry. As a result, even though the planarity constraint was not explicitly used in generating the stimuli, some faces were planar;

5) planar, asymmetric polyhedron. This stimulus was obtained by using one half of the polyhedron from condition 1;

6) non-planar, asymmetric polyhedron. The contours of this stimulus were non-planar and the polyhedron was asymmetric;

7) three faces (quadrilaterals) of the polyhedron; each of two pairs of the faces shared one vertex;

8) three triangles. Each triangle was one part of a quadrilateral that was used in condition 7;

9) polygonal line connecting the seven vertices that were used in condition 8.

The stimuli were displayed by means of rotation around a vertical axis. Each experimental trial consisted of two short sequences of motion that represented either the same or different stimuli. The two sequences were 90 degrees apart. In a trial with different stimulus, the second sequence involved a new randomly generated stimulus. The subject's task was to discriminate whether the two sequences contained stimuli with the same or different shapes.

The performance was the best in condition 1. Recall that the polyhedra in this condition were generated with constraints: contours of faces were planar and the object was symmetrical. Furthermore, all edges of the polyhedron were visible, which

provided rich monocular cues related to 3-D topology. Conditions 2 and 3 led to much worse performance than condition 1. In fact, the performance in condition 3 was at chance level. Note that in condition 2 only symmetry constraint was present. In condition 3 no constraints were present. This shows that 3-D shapes can be reliably recognized and/or reconstructed if the objects contain a number of constraints, and monocular cues related to 3-D topology are present.

Conditions 4 and 5, where either planarity was partially removed or symmetry was absent, led to slightly worse performance than condition 1. Condition 6, where both planarity and symmetry were absent, led to performance that was much worse than condition 1. This shows that planarity (and also symmetry) constraint is important in 3-D shape perception.

Conditions 7-9 led to much worse performance than condition 1 and the performance in conditions 8 and 9 was close to chance level. The stimuli in these conditions were generated using only some faces and/or only some vertices of the original polyhedra; thus, the topological constraints were very weak. For example, changing the shape of one part of the object in conditions 7-9 does not have to be associated with any other changes of the object. In condition 1, on the other hand, it is not possible to change the shape of one face without changing the shapes of several other faces because the faces shared vertices and edges. The poor performance in conditions 7-9 shows that topological constraints are also involved in 3-D shape perception.

To summarize, constraints such as planar surface contours, 3-D topology, and symmetry are critical in perception of shape from motion. Interestingly, these constraints resemble closely the constraints that were previously identified in shape perception from a single image [AF69] [HB60] [Per72].

The fact that shape perception in human observers involves the same set of constraints in the case of perception from a single image and from multiple images (motion) suggests that these two cases involve similar mechanisms. Furthermore, it seems reasonable to conjecture that the same mechanism is involved in perception of shape from two images (binocular case). In the next section we present a new algorithm for

binocular shape reconstruction. This algorithm uses the constraints found in prior psychophysical experiments. Then we present results of the simulations as well as a psychophysical test of the algorithm's psychological plausibility.

5. A NEW APPROACH TO BINOCULAR SHAPE RECONSTRUCTION

As we discussed in Section 4.1, binocular disparity is not a reliable source of information that can be used to reconstruct shapes in the presence of image noise. Consequently, a 3-D shape reconstruction that is obtained directly from binocular disparity is very unstable even when 1% of noise is added to the image coordinates. The errors of reconstruction are so large that they often lead to topological reversals in 3-D. But 1% of noise on the image coordinates cannot be very harmful in monocular shape reconstruction (i.e. reconstruction from a single image) because it virtually does not affect the shape of the object's image. Indeed, prior computer algorithms [Mar91] [LF92] [Sug86] [SP97], which used constraints analogous to those described in the previous section, have demonstrated that shape reconstruction from a single image is stable, and almost never leads to topological errors. This means that in order to avoid the instability caused by binocular disparity, one should begin 3-D shape reconstruction monocularly, and only then use binocular disparity to improve the accuracy of the shape. This is the approach that we have adopted.

5.1 The New Method

An overview of our new method is shown in Figure 5.1. In this method, we first assign a relative depth ordering to the 3-D points. This relative depth ordering is a weak constraint which is not critical, but it improves computational efficiency of monocular shape reconstruction. One shape is obtained from each image. Note that this shape is just an initial "guess" of the 3-D shape of the object (called henceforth shape hypothesis). The final reconstruction uses information from the second image of the object to correct the shape hypothesis. Finally, one shape is selected from the

two corrected shape hypotheses as the solution. The implementation details of this method are presented next.

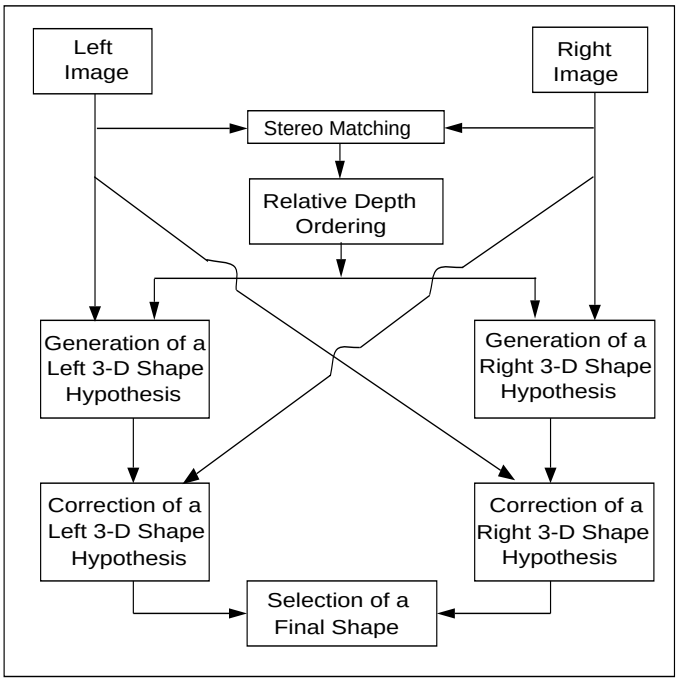


Fig. 5.1. Overview of the new binocular shape reconstruction algorithm.

It is assumed here that the intrinsic camera parameters are known (see [Fau93] [Har97] for a comprehensive review of camera calibration), and the feature (point) correspondences are known (see [BF82] [DA89] [VC95] [CN95] for matching algorithms).

The basic geometry of this problem is shown in Figure 5.2. We assume that a binocular fixating system is used, in which the left camera coordinate system XYZ is translated in X′ and Z′ directions, and rotated about the Y′ axis relative to the coordinate system X′Y′Z′. The axes Z and Z′ intersect at a fixation point. This camera configuration simplifies computation of extrinsic camera parameters (i.e. orientation and position of the left camera) [CPC99]. O and O′ are the centers of projection of the left and right cameras, I and I′ are the left and right images of the object on the image planes Π and Π′, and H and H′ are the left and right shape hypotheses respectively.

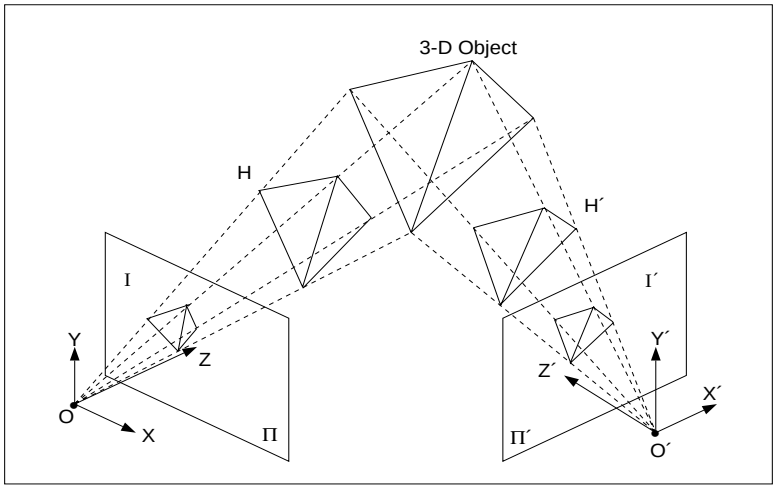


Fig. 5.2. Geometry of a binocular fixating viewing system.

5.1.1 Computing a Relative Depth Ordering

Monocular shape reconstruction of a 3-D object involves computing a vector $\mathbf{z}$, which contains the z coordinate of each point. The x and y coordinates of a 3-D point are computed by the rules of perspectivity from the z coordinate. The relative depth ordering is performed in the beginning and is used to provide more stable monocular reconstruction, and to reduce the computational complexity.

Let $\mathbf{U}$ $(U_xU_yU_z)$ and $\mathbf{V}$ $(V_xV_yV_z)$ be two 3-D points in H, and $\mathbf{u}$ (u_xu_y) and $\mathbf{v}$ (v_xv_y) be their image coordinates respectively. Similarly, let $\mathbf{U}'$ $(U'_xU'_yU'_z)$ and $\mathbf{V}'$ $(V'_xV'_yV'_z)$ be two 3-D points in H$'$, and $\mathbf{u}'$ $(u'_xu'_y)$ and $\mathbf{v}'$ $(v'_xv'_y)$ be their image coordinates. Note that $\mathbf{u}'$, $\mathbf{v}'$ and $\mathbf{u}$, $\mathbf{v}$ are corresponding image points. A relative depth ordering of two points with respect to $\mathbf{O}'$ (and $\mathbf{O}$) is determined as follows:

$U'_z < V'_z$ (and $U_z < V_z$), if $disparity(\mathbf{u}) - disparity(\mathbf{v}) \geq \epsilon$, and vice versa.

Note that $disparity(\mathbf{u}) = u_x - u'_x$, and ϵ is an arbitrary small positive number, which accounts for errors resulting from noisy image coordinates. In the process of monocular reconstruction, the depth ordering established by the above rule is not violated.

5.1.2 Generating a Shape Hypothesis

Some of the constraints that were described in Section 4.2 had been implemented by computer vision researchers. Marill used the minimum standard deviation of angles (MSDA) constraint to reconstruct a shape from one orthographic view of an object [Mar91]. The algorithm begins by assigning the same z coordinate to every image point. Then, the z coordinates are changed to minimize the standard deviation of angles. The MSDA constraint favors 3-D as opposed to 2-D objects. Note that MSDA requires the presence of sufficiently strong topological constraints. For example, applying MSDA to a polygonal line (conditions 3 and 9 in Pizlo and Stevenson's experiment) would result in an infinitely large family of solutions. Each solution would have right angles at the "L" junctions, and all these solutions would have the same minimal value of MSDA equal to zero. Recall that the performance of human subjects in conditions 3 and 9 was close to chance level. Clearly, both MSDA and human subjects require the presence of strong topological constraints. Marill pointed out that his algorithm gives rise to shapes that resemble perceptual interpretations of human observers. Leclerc and Fischler extended Marill's work [LF92]. Specifically, they showed that using planarity of surface contours in addition to MSDA produces more stable and accurate shapes. They also pointed out that their informal observations show that their algorithm is psychologically more plausible.

We modified the algorithm developed by Leclerc and Fischler [LF92] to obtain shape hypotheses. Specifically, we assume perspective rather than orthographic projection, and we use a different cost function. Below, we describe the generation of the left shape hypothesis, H (H′ is obtained in an analogous way). The objective is to find $\mathbf{z}$ that minimizes the following function:

$$E_{mono}(\mathrm{H}) = \lambda_m[SDA(\mathrm{H}) + DS(\mathrm{H})] + (1 - \lambda_m)DP(\mathrm{H}) \tag{5.1}$$

where $SDA(\mathrm{H})$ is the standard deviation of all interior angles of each face of H, $DP(\mathrm{H})$ is a measure of departure from planarity for all faces of H, $DS(\mathrm{H})$ is a measure of asymmetry of H, and λ_m is a scalar which ranges from 0 to 1.

Let F be the number of faces in H, E_i be the number of edges (the number of angles) in the i^{th} face, and θ_{ij} be the j^{th} angle in face i, then

$$DP(\mathrm{H}) = \sum_{i=1}^{F}[(E_i - 2)\pi - \sum_{j=1}^{E_i}\theta_{ij}]^2 \tag{5.2}$$

Note that $DP(\mathrm{H})$ is a planarity measure for convex faces. In the case of non-convex faces, a more general planarity measure is used [LF92]. The measure is based on the observation that the normals defined by pairs of consecutive pairs of lines of a planar polygon have the same direction.

$DS(\mathrm{H})$ is computed as a sum of squared differences of each pair of symmetric angles in H (we assume that the correspondence of these angles is known). This term did not exist in Leclerc and Fischler's algorithm. We added it because it improves the reconstruction performance, although this improvement is rather modest.

Following Leclerc and Fischler, we apply a continuation method [WTK87] [Lec89] to minimize $E_{mono}(\mathrm{H})$. This method involves a sequence of descent steps applied to $E_{mono}(\mathrm{H})$ in which λ_m is decreased in each step. In the algorithm developed by Leclerc and Fischler [LF92], λ_m was suggested to start at some value that is close to 1. This means that the first descent step is strongly influenced by the $SDA(\mathrm{H})$ and $DS(\mathrm{H})$ terms, which give the initial shape of H. As λ_m decreases, $DP(\mathrm{H})$ becomes the dominant factor, which enforces planarity of the faces. This method does not necessarily yield a global minimum, but the reconstruction is quite accurate and fast when the relative depth ordering constraint is enforced.

5.1.3 Correcting the Shape Hypothesis

At this stage, H (and H′) is still an initial "guess" of the 3-D shape of the object. We must incorporate information from the second image in order to obtain a more

accurate shape. In other words, a similarity measure of H and I′ is needed, where I′ is the right image of the 3-D object. Here, we measure the squared differences $SHI(\mathrm{I}', \mathrm{H})$ of the angles in I′ and the angles of H projected on the right image plane Π′. In order to project H on Π′, it is necessary to compute the extrinsic camera parameters. This is done by using the 8 point algorithm with a binocular fixation constraint added to it [CPC99]. Note that if the extrinsic parameters and H are reconstructed correctly, then $SHI(\mathrm{I}', \mathrm{H})$ is zero.

We could have chosen a similarity measure involving Euclidean distances between the points in I′ from the projected points of H on Π′. We decided to use angles, instead of position of points because angles are more robust in the presence of errors in camera orientation. Note that if the camera image were a sphere (like in the case of biological systems), with the center of projection coinciding with the center of the sphere, the image angles would have been completely insensitive to errors in camera orientation. In such a case, the orientation of one camera relative to the other, would have been irrelevant.

We perform binocular shape reconstruction by correcting only some features of the shape hypothesis. This will simplify the computations. Specifically, we modify H without changing the planarity of contours. As a result, the cost function does not have to contain the planarity constraint. This is done by performing the correction in a projective space. We first show how to transform H in a projective space, and then we show how to use $SHI(\mathrm{I}', \mathrm{H})$ to correct H.

Let $(U_x U_y U_z U_w)$ be the homogeneous coordinates of a 3-D point $\mathbf{U}$ belonging to H, $(u_x u_y u_w)$ the homogeneous coordinates of its image point $\mathbf{u}$, $(\hat{U}_x \hat{U}_y \hat{U}_z \hat{U}_w)$ the homogeneous coordinates of $\hat{\mathbf{U}}$ which represents $\mathbf{U}$ after a 3-D projective transformation. Note that U_w can be any non-zero scalar; here we put $U_w = 1$.

We first express the relationship between $\mathbf{u}$ and $\mathbf{U}$ in homogeneous coordinates as follows:

$$\begin{bmatrix} u_x \\ u_y \\ u_w \end{bmatrix} = \mathbf{M} \begin{bmatrix} U_x \\ U_y \\ U_z \\ 1 \end{bmatrix} \tag{5.3}$$

note that $\mathbf{M}$ is a 3x4 matrix, which encapsulates the projective geometry of the camera system [Fau93] [Har97]. Since H is expressed in the coordinate system associated with the camera which obtained image $\mathbf{u}$, $\mathbf{M}$ takes the following form:

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & 0 \\ 0 & m_{22} & m_{23} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The non-zero coefficients of $\mathbf{M}$ represent the intrinsic camera parameters (i.e. coordinates of the principal point, focal length, pixel's aspect ratio and the skew parameter). In order to transform $\mathbf{U}$ in a projective space, we pre-multiply it by a 4x4 matrix $\mathbf{K}$, which represents a general projective transformation:

$$\begin{bmatrix} \hat{U}_x \\ \hat{U}_y \\ \hat{U}_z \\ \hat{U}_w \end{bmatrix} = \mathbf{K} \begin{bmatrix} U_x \\ U_y \\ U_z \\ 1 \end{bmatrix} \tag{5.4}$$

This projective transformation $\mathbf{K}$ must keep $\hat{\mathbf{U}}$ on the visual ray emanating from $\mathbf{u}$ and passing through $\mathbf{U}$. This means that the image of $\hat{\mathbf{U}}$ must be the same as the image of $\mathbf{U}$. This is expressed by the following:

$$\begin{bmatrix} u_x \\ u_y \\ u_w \end{bmatrix} = \mathbf{M} \begin{bmatrix} \hat{U}_x \\ \hat{U}_y \\ \hat{U}_z \\ \hat{U}_w \end{bmatrix} \tag{5.5}$$

We substitute Equation 5.4 into Equation 5.5 to obtain:

$$\begin{bmatrix} u_x \\ u_y \\ u_w \end{bmatrix} = \mathbf{MK} \begin{bmatrix} U_x \\ U_y \\ U_z \\ 1 \end{bmatrix} \tag{5.6}$$

Equations 5.3 and 5.6 imply

$$\mathbf{M} = \mathbf{MK} \tag{5.7}$$

The coefficients of the matrix $\mathbf{K}$ can be solved by equating the corresponding coefficients of $\mathbf{M}$ and $\mathbf{MK}$. It is easy to show that $\mathbf{K}$ must have the following form:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ k_1 & k_2 & k_3 & k_4 \end{bmatrix} \tag{5.8}$$

where k_1, k_2, k_3, and k_4 are unknown scalars. Note that a general projective transformation in 3-D has 16 degrees of freedom (only 15 are independent); however, only 4 degrees of freedom are allowed if the points are restricted to remain on their visual rays. In order to compute $(\hat{U}_x\hat{U}_y\hat{U}_z\hat{U}_w)$, we substitute Equation 5.8 into Equation 5.4:

$$\begin{bmatrix} \hat{U}_x \\ \hat{U}_y \\ \hat{U}_z \\ \hat{U}_w \end{bmatrix} = \begin{bmatrix} U_x \\ U_y \\ U_z \\ k_1U_x + k_2U_y + k_3U_z + k_4 \end{bmatrix} \tag{5.9}$$

The Euclidean coordinates of $\hat{\mathbf{U}}$ are obtained by dividing $(\hat{U}_x\hat{U}_y\hat{U}_z)$ by $\hat{U}_w$. Clearly, dividing the x, y, z coordinates of a point by the same number would move the

point along its projecting line; thus, its image(x/z, y/z) would not change. If the constant has the form $k_1x + k_2y + k_3z + k_4$ for all the points of the object, then the transformation is a 3-D projectivity.

In order to correct H using $SHI(\mathrm{I}', \mathrm{H})$, we compute k_1, k_2, k_3, and k_4 that minimize the following function:

$$E_{bino}(\mathrm{I}', \mathrm{H}) = SHI(\mathrm{I}', \mathrm{H}) + \lambda_b[SDA(\mathrm{H}) + DS(\mathrm{H})] \tag{5.10}$$

Note that the planarity measure $DP(\mathrm{H})$ is not needed in the cost function because the planar surface contours are maintained under projective transformation. The final shape is selected from the two shape hypotheses H and H$'$ as the one that has a smaller value of $E_{bino}(\cdot)$.

We performed a large number of experiments using different values for λ_b in the range from 0.1 to 3.0. The performance did not vary significantly with λ_b.

Note that unified methods for monocular and binocular shape reconstruction have been proposed in the past. For example, Fua and Leclerc [FL95] [FL96] provided a method for surface reconstruction, which involves iterative minimization a cost function expressed as a sum of weighted terms such as smoothness constraint, monocular depth cues, and binocular cues. In their method, a set of user-defined weighting factors that specifies the relative importance of these terms are required at the start of the algorithm. These weighting factors are then adjusted and normalized in each iteration to determine how much weight each term should contribute to the reconstruction process. For example, if a surface patch is richly textured, then the binocular term of this patch is assigned more weight. Note that if binocular disparity is weighted too much, it may lead to large topological errors which may be difficult to correct in subsequent iterations. In our method, binocular disparity is used only to correct some features of the shape hypothesis; thus, it is less likely to cause large topological errors in 3-D.

5.2 Simulations

We tested our new algorithm on the stimuli that were used in Pizlo and Stevenson's psychophysical experiment [PS99]. The stimuli were enclosed in a unit cube. The viewing parameters were as follows:

1) **Viewing distance:** the objects were placed at the distances of 4, 8 or 16 from the center of projection $\mathbf{O}'$. These distances corresponded to vergence angles 8, 4, and 2 degrees.

2) **Noise:** uncorrelated Gaussian noise was imposed on each image coordinate with zero mean and standard deviation of 1%, 2%, 3%, and 5% of the diameter of the object's image. Note that 1.4% of noise for an individual point leads to about 2% of uncorrelated noise for the distance between two points. This level of noise is known to exist in the human visual system [Wat87] [DLN+90]. ϵ of 0.005 was used in all simulation experiments unless stated otherwise.

One hundred objects were generated randomly. The accuracy of the reconstruction of the similarity structure was evaluated by a normalized ratio, which was computed as follows:

$$\mathbf{ratio}_i = \frac{\frac{\|\mathbf{U}'_i, \mathbf{U}'_j\|}{BASE'}}{\frac{\|\mathbf{U}i, \mathbf{U}j\|}{BASE}} \tag{5.11}$$

where i $\neq$ j, $\|\mathbf{U}i, \mathbf{U}j\|$ denotes the Euclidean distance between the points $\mathbf{U}_i$ and $\mathbf{U}_j$, $\mathbf{U}_i$ represents the i^{th} original 3-D point, $\mathbf{U}'_i$ represents the i^{th} reconstructed point. $BASE$ is the Euclidean distance representing the length of the object, and $BASE'$ is the corresponding reconstructed length.

In each simulation experiment, we plotted the frequency distribution of the ratios. A total of 11,900 data points were produced from 100 reconstructions. A perfect reconstruction is represented by a **ratio** of one. The **ratio**s that are greater than or equal 4.0 are treated as outliers, and they are not shown in the graphs. The means and standard deviations of these frequency distributions, and the number of outliers

are shown in tables. The means and standard deviations were computed without the outliers.

5.2.1 Simulation Experiment I

Figure 5.3 shows the histograms computed for the shape hypotheses, corrected shapes, and final shapes obtained with our new algorithm, and for the shapes obtained with the binocular fixating algorithm of Chan et al. [CPC99]. Table 1 shows the means, standard deviations, and the numbers of outliers of these histograms. In this simulation experiment, vergence was 8 degrees, and the noise was 2%.

Table 5.1 Means and standard deviations of the ratios, and the numbers of outliers obtained with our algorithm and the binocular fixating algorithm. Vergence was 8 degrees, and noise standard deviation was 2%. Outliers have ratios greater than or equal 4.0.

	Mean	Std. Dev.	Outliers
Left shape hypothesis	1.0972	0.3043	55
Right shape hypothesis	1.1008	0.3022	57
Left corrected shape	1.0353	0.2132	13
Right corrected shape	1.0416	0.2457	31
Final shape of the new algorithm	1.0389	0.2228	26
Binocular fixating algorithm	1.0546	0.5831	663

It is seen that the new algorithm is quite robust in the presence of noise. In fact, already the monocular shape reconstruction is substantially more stable than the binocular fixating algorithm. Specifically, the standard deviations of the reconstructions representing the shape hypotheses are about one half of the standard deviation of the reconstructions obtained with the binocular fixating algorithm. Moreover, the number of outliers obtained in shape hypotheses is an order of magnitude smaller than the number of outliers obtained with the binocular fixating algorithm. These results suggest that in the presence of noise, constraints are more useful than binocular

disparity. The correction involving the second image results in even more accurate reconstruction. This shows that our new algorithm can successfully use both the information in the images and constraints.

5.2.2 Simulation Experiment II

Figures 5.4 - 5.6 show the histograms computed for the reconstructions with different amounts of noise. The means, standard deviations, and the numbers of outliers of the histograms representing the final reconstructions of our new algorithm, and of the binocular fixating algorithm are shown in Table 2. In this experiment, the vergence was 8 degrees, and the noise standard deviation was 1%, 3% and 5%.

Table 5.2 Means and standard deviations of the ratios, and the numbers of outliers obtained with our algorithm and the binocular fixating algorithm. Vergence was 8 degrees, and noise standard deviation was 1%, 3%, and 5% noise. Outliers have ratios greater than or equal 4.0.

	Mean	Std. Dev.	Outliers
New Algorithm (1% noise)	1.0211	0.1465	1
Binocular Fixating Algorithm (1% noise)	1.0859	0.5487	408
New Algorithm (3% noise)	1.0524	0.2390	56
Binocular Fixating Algorithm (3% noise)	1.1249	0.6181	700
New Algorithm (5% noise)	1.1245	0.4273	185
Binocular Fixating Algorithm (5% noise)	1.1302	0.6357	1289

It is seen that our method is more stable than that of the binocular fixating algorithm at all levels of noise. Specifically, at 1% of noise, which is similar to the amount of noise in the human visual system, our algorithm produced only one outlier (ratio greater than or equal to 4.0). At the same time, the binocular fixating algorithm obtained 408 outliers. The average errors and standard deviation of the reconstructions obtained with our algorithm are about 2% and 15% respectively, which are about four times better than those obtained with the binocular fixating algorithm. More

importantly, the performance of our algorithm is similar to the performance of human subjects in a shape matching task (see Section 4.1) [CPC99]. At 3% of noise, the performance of our algorithm suffers slightly; however, it is still substantially more accurate and stable than the binocular fixating algorithm. Finally, at 5% of noise, the performance of our algorithm deteriorates. It is seen that its mean and standard deviation are close to those obtained with the binocular fixating algorithm. However, the number of outliers obtained with our algorithm is still one order of magnitude smaller than that of the binocular fixating algorithm.

Overall, monocular reconstruction obtained with our algorithm is already more accurate and stable than that of the binocular fixating algorithm. The performance obtained with the correction involving the second image results in even more accurate reconstruction. This shows that, in the presence of a large amount of noise, constraints are still more useful, and the new algorithm can still combine monocular cues related to 3-D topology, constraints, and binocular disparity in order to reconstruct accurate 3-D shapes.

5.2.3 Simulation Experiment III

Figures 5.7 - 5.8 show the histograms computed for reconstructions with vergence angles 4 and 2 degrees respectively. The means, standard deviations, and the number of outliers of these histograms are shown in Table 3. Note that the results obtained in Section 5.2.1 (with 8 degrees vergence) are also shown in Table 3 for comparison. In this simulation experiment, the noise was 2% The ϵ was set to 1.0 in this experiment to relax the relative depth ordering constraint.

It is seen that the performance of our algorithm is more stable and the numbers of outliers are substantially smaller than those obtained with the binocular fixating algorithm. Note that binocular disparity is not very useful when the vergence angle is small because the differences between the left and right images produced by the difference in the viewing directions are small as compared to the differences produced by noise. The reasons that our algorithm is able to obtain relatively stable shape

Table 5.3 Means and standard deviations of the ratios, and the numbers of outliers obtained with our algorithm and the binocular fixating algorithm. Noise standard deviation was 2%, vergence was 2, 4, and 8 degrees. Outliers have ratios greater than or equal 4.0.

	Mean	Std. Dev.	Outliers
New Algorithm (8 degrees)	1.0389	0.2228	26
Binocular Fixating Algorithm (8 degrees)	1.0546	0.5831	663
New Algorithm (4 degrees)	1.1151	0.3622	117
Binocular Fixating Algorithm (4 degrees)	1.1192	0.6623	1008
New Algorithm (2 degrees)	1.1546	0.4534	283
Binocular Fixating Algorithm (2 degrees)	1.2026	0.7103	1236

reconstruction under small vergence angles are: (a) the shape was established by means of monocular reconstruction, in which binocular disparity is not used; and (b) the correction stage of our algorithm can maintain the monocularly reconstructed shape, if λ_b in Equation 5.10 is large. In this simulation, λ_b of 2.0 was used.

5.2.4 Simulation Experiment IV

For visualization purpose, we display two images of one original object and of the reconstructions provided by the new algorithm and the binocular fixating algorithm. In this simulation experiment, vergence was 8 degrees, and the noise was 2%. Figure 5.9 shows the images of a stimulus used in our simulation experiment, and Figure 5.10 shows the images of a house, which is an object on which the algorithm has not been tried before. It is seen that the shapes reconstructed by the new method are quite similar to the original objects. On the other hand, the shapes reconstructed by the binocular fixating algorithm are quite inaccurate.

5.2.5 Experiment with Real Images

We also tested our algorithm on real images. Viewing parameters were similar to those in the simulation (Section 5.2.4). The stereo images of a wooden object and its reconstructions obtained with our new algorithm and the binocular fixating algorithm are shown in Figure 5.11. The image center was assumed to be the principal point, and the focal length was obtained by a calibration procedure. The skew parameter was assumed to be zero, and the aspect ratio was assumed to be one. It is seen that the reconstructed shape obtained with our new algorithm is more similar to the original object as compared to that of the binocular fixating algorithm.

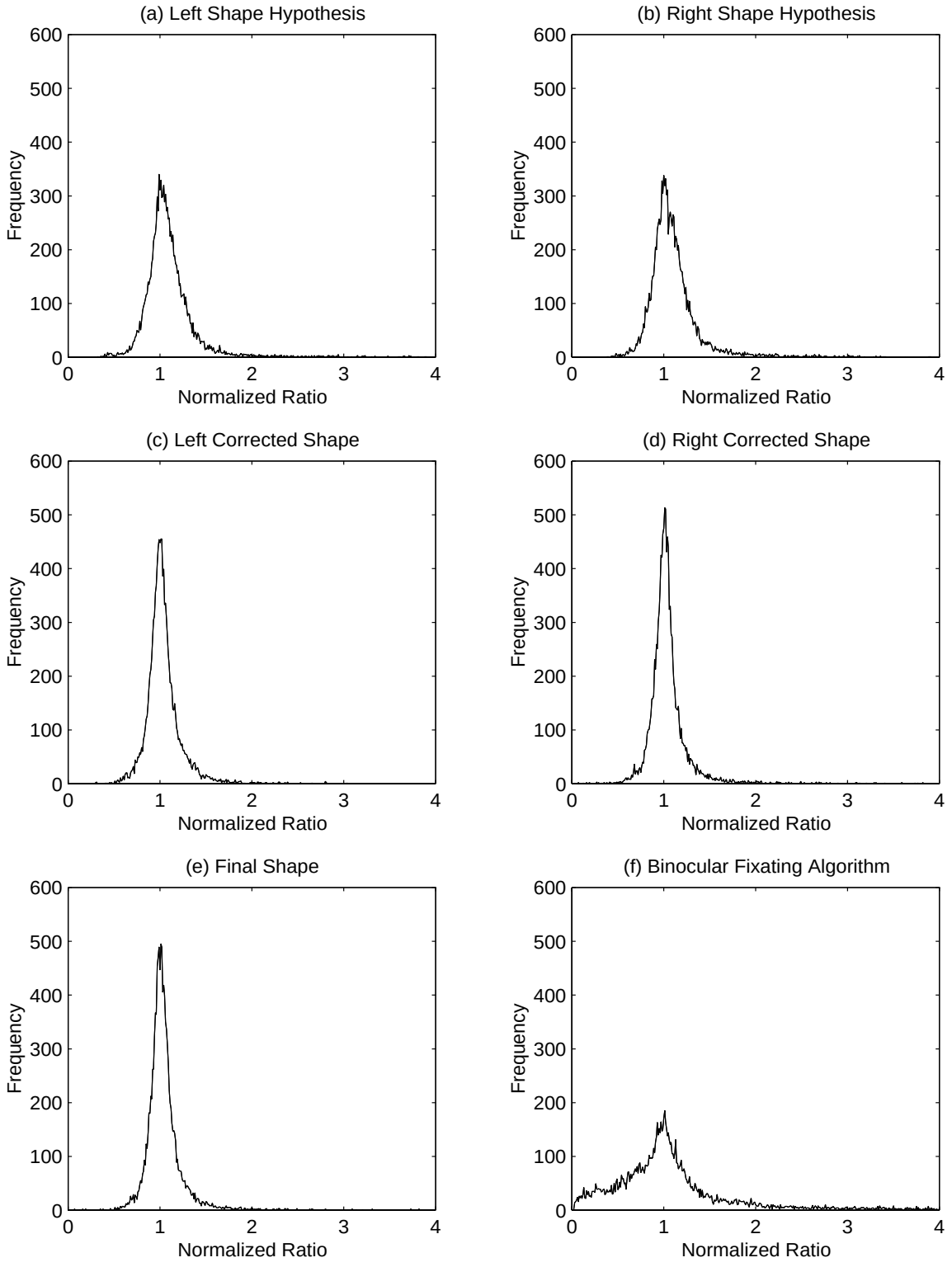


Fig. 5.3. Histograms of the normalized ratios obtained with our new algorithm and the binocular fixating algorithm. Vergence was 8 degrees, noise standard deviation was 2%. The histogram is truncated at the normalized ratio of 4.0.

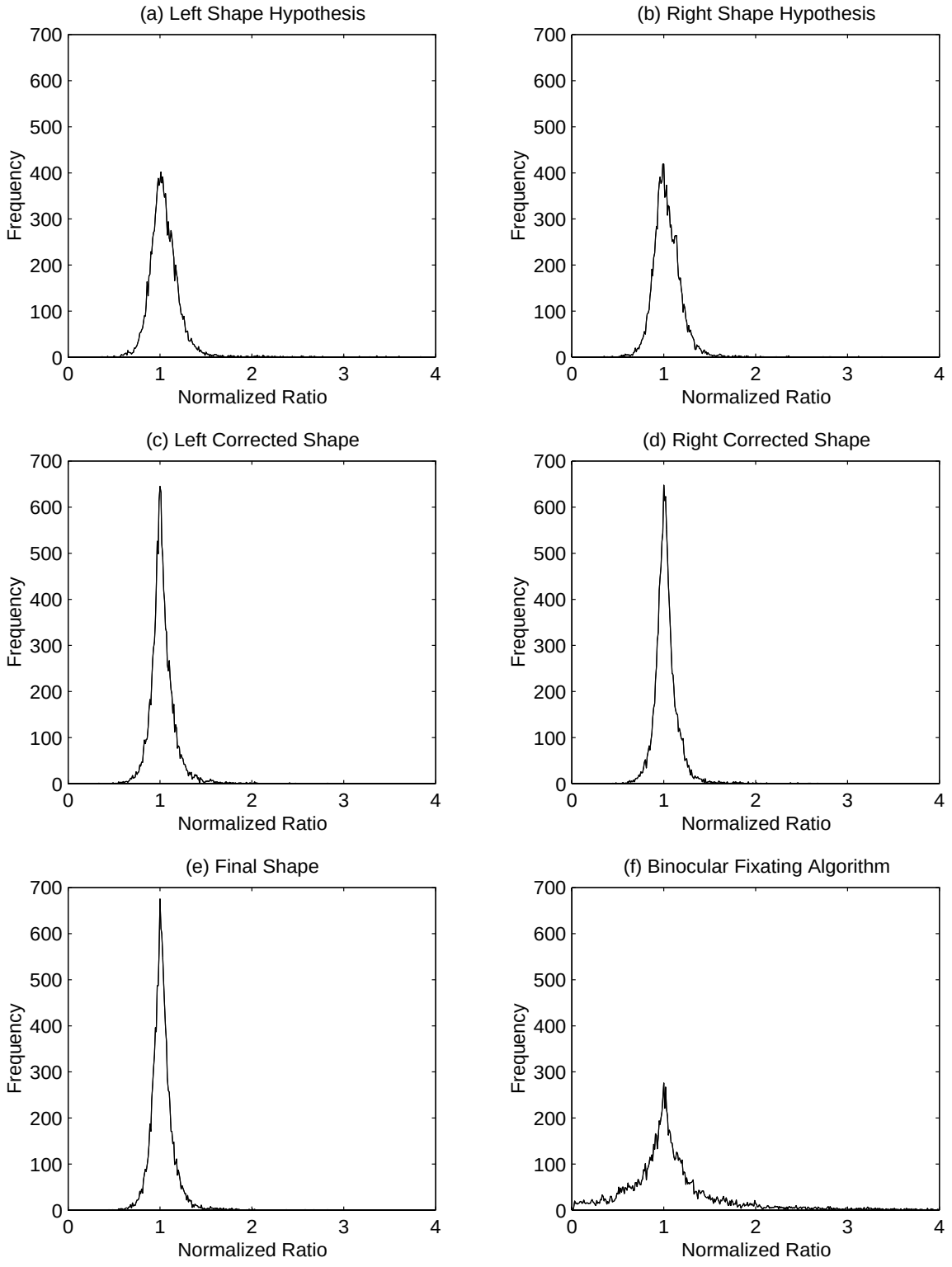


Fig. 5.4. Histograms of the normalized ratios obtained with our new algorithm and the binocular fixating algorithm. Vergence was 8 degrees, and noise standard deviation was 1%. The histogram is truncated at the normalized ratio of 4.0.

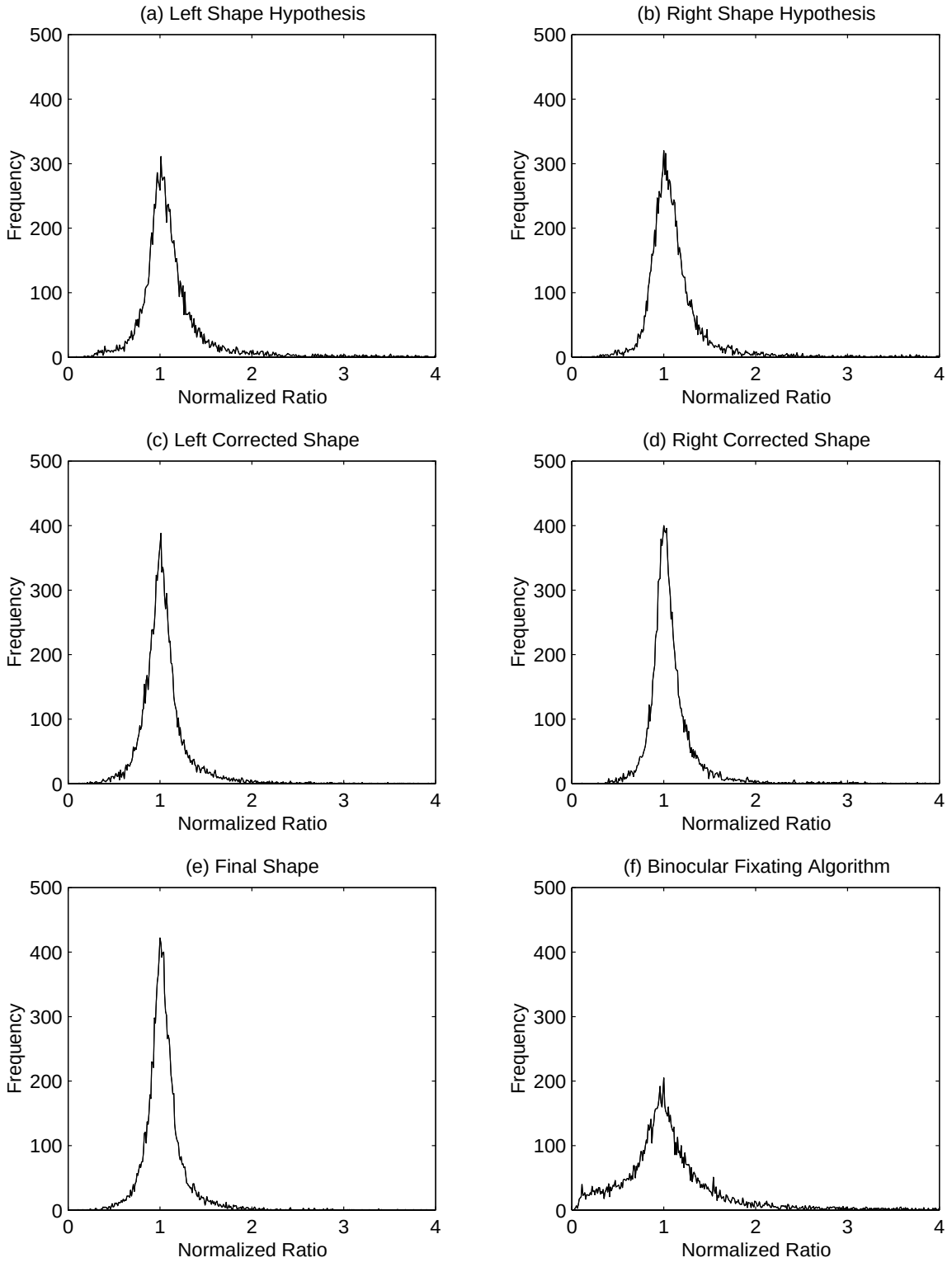


Fig. 5.5. Histograms of the normalized ratios obtained with our new algorithm and the binocular fixating algorithm. Vergence was 8 degrees, and noise standard deviation was 3%. The histogram is truncated at the normalized ratio of 4.0.

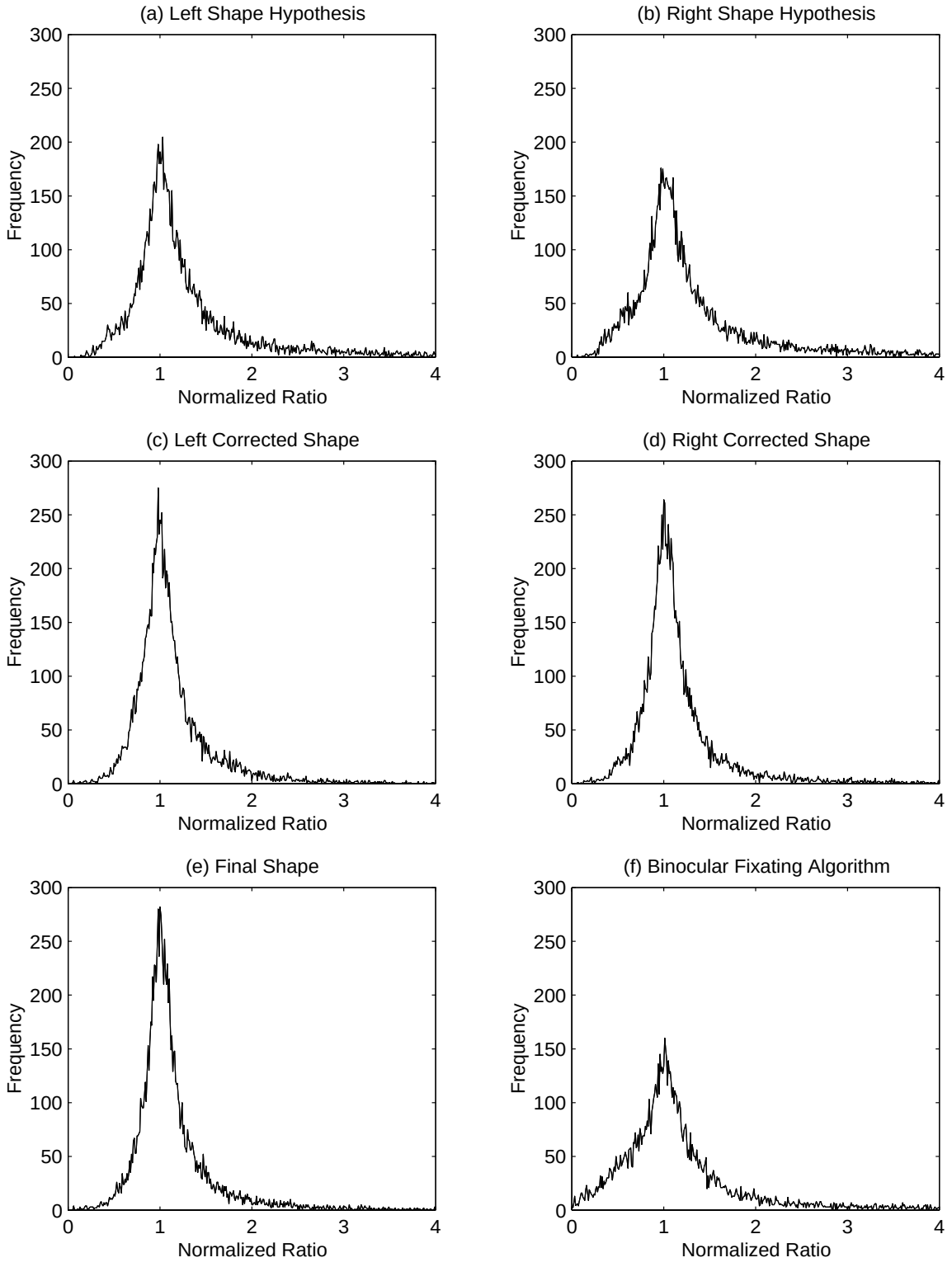


Fig. 5.6. Histograms of the normalized ratios obtained with our new algorithm and the binocular fixating algorithm. Vergence was 8 degrees, and noise standard deviation was 5%. The histogram is truncated at the normalized ratio of 4.0.

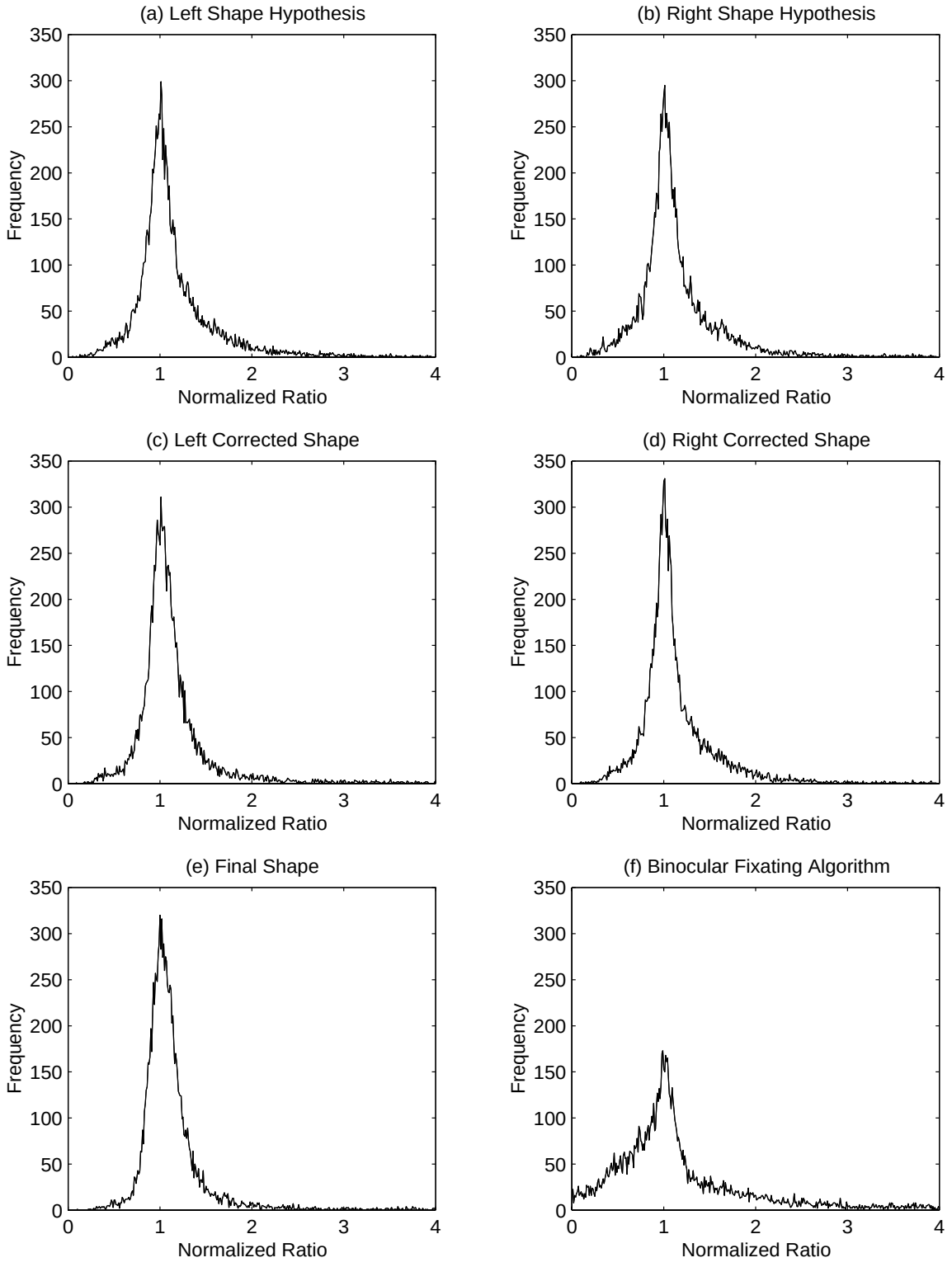


Fig. 5.7. Histograms of the normalized ratios obtained with our new algorithm and the binocular fixating algorithm. Noise standard deviation was 2%, and vergence was 4 degrees. The histogram is truncated at the normalized ratio of 4.0.

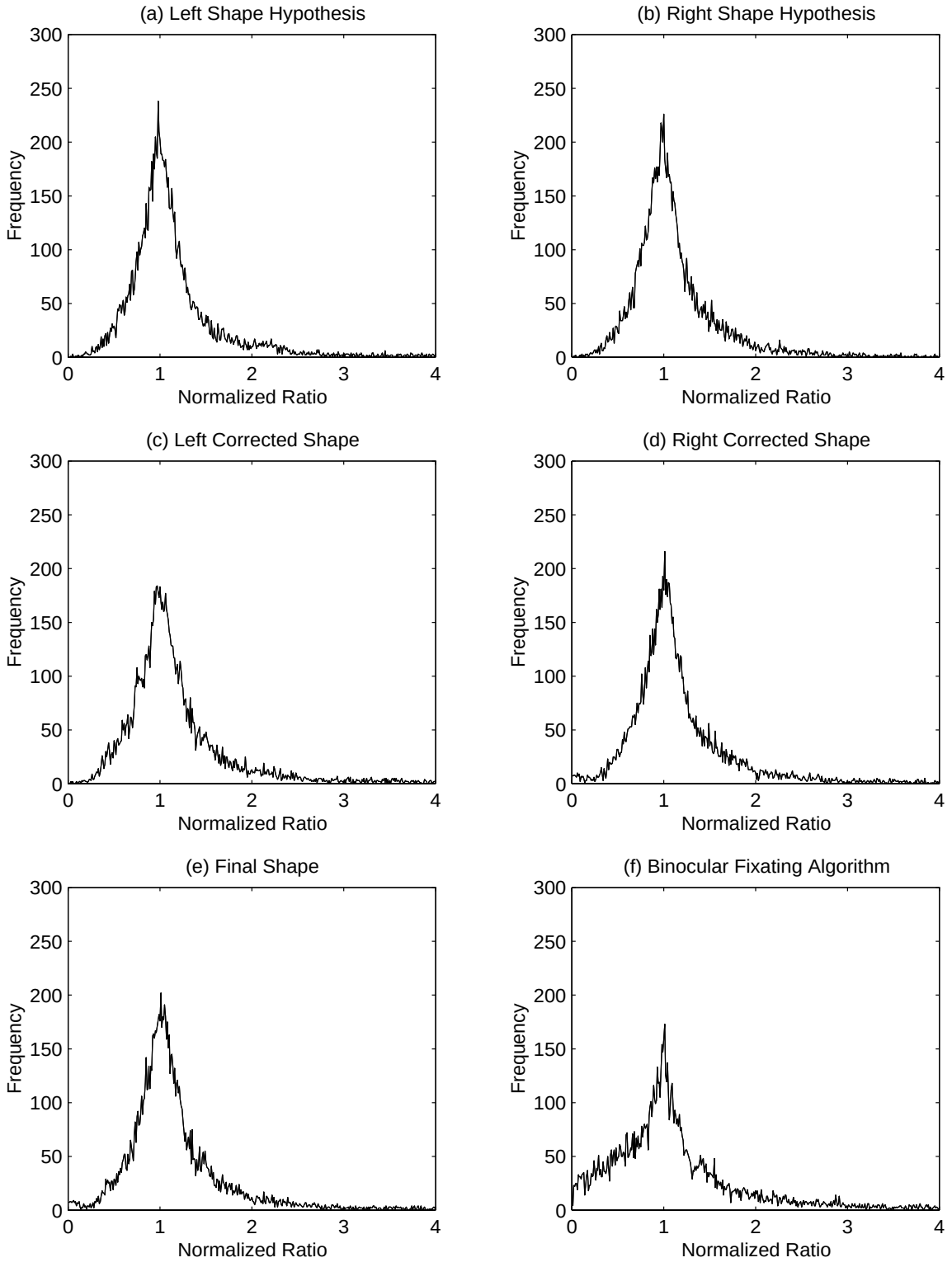


Fig. 5.8. Histograms of the normalized ratios obtained with our new algorithm and the binocular fixating algorithm. Noise standard deviation was 2%, and vergence was 2 degrees. The histogram is truncated at the normalized ratio of 4.0.

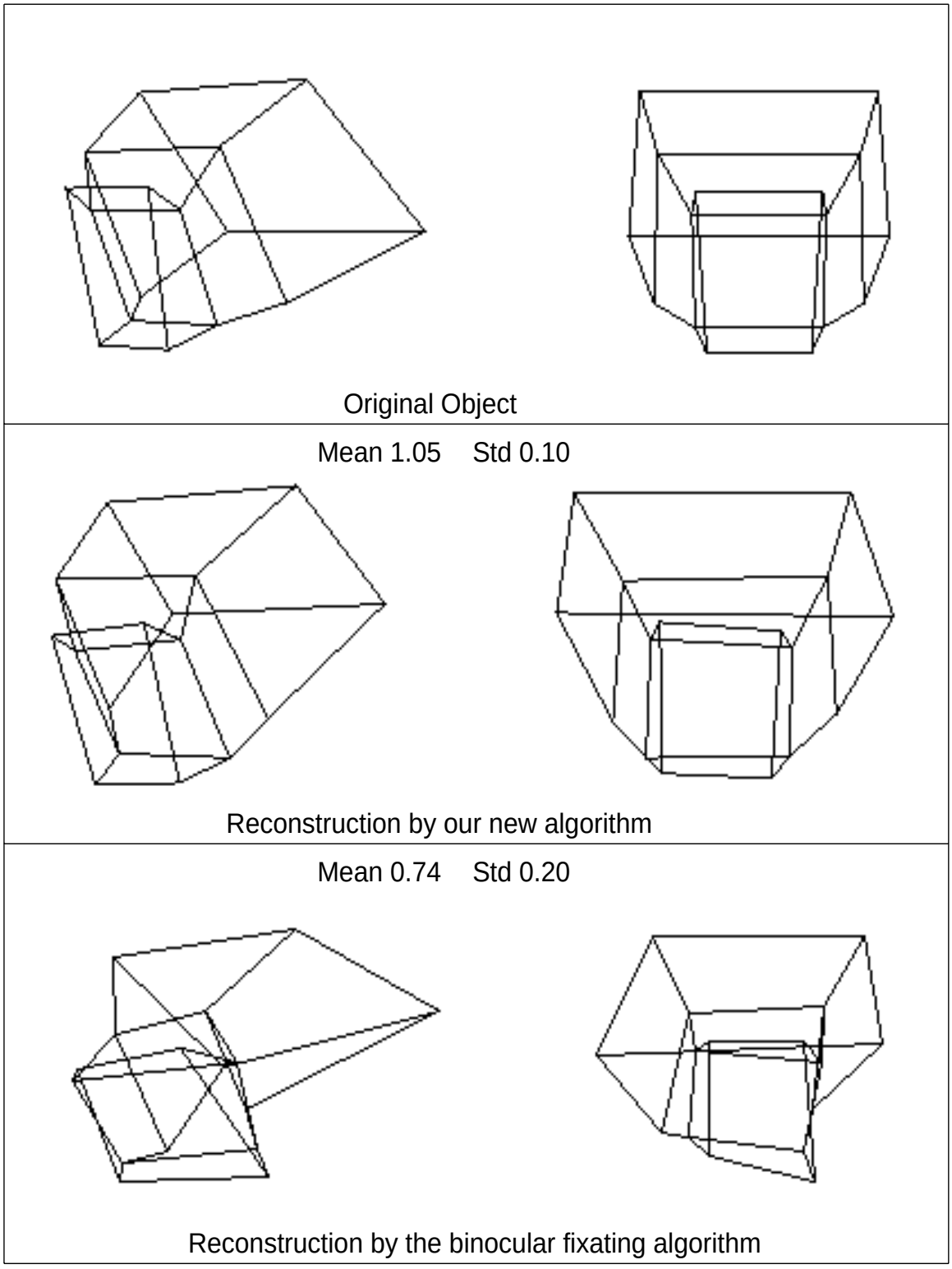


Fig. 5.9. Two different views of the original object, and its reconstructions are shown. The means and the standard deviations of the normalized ratios are provided.

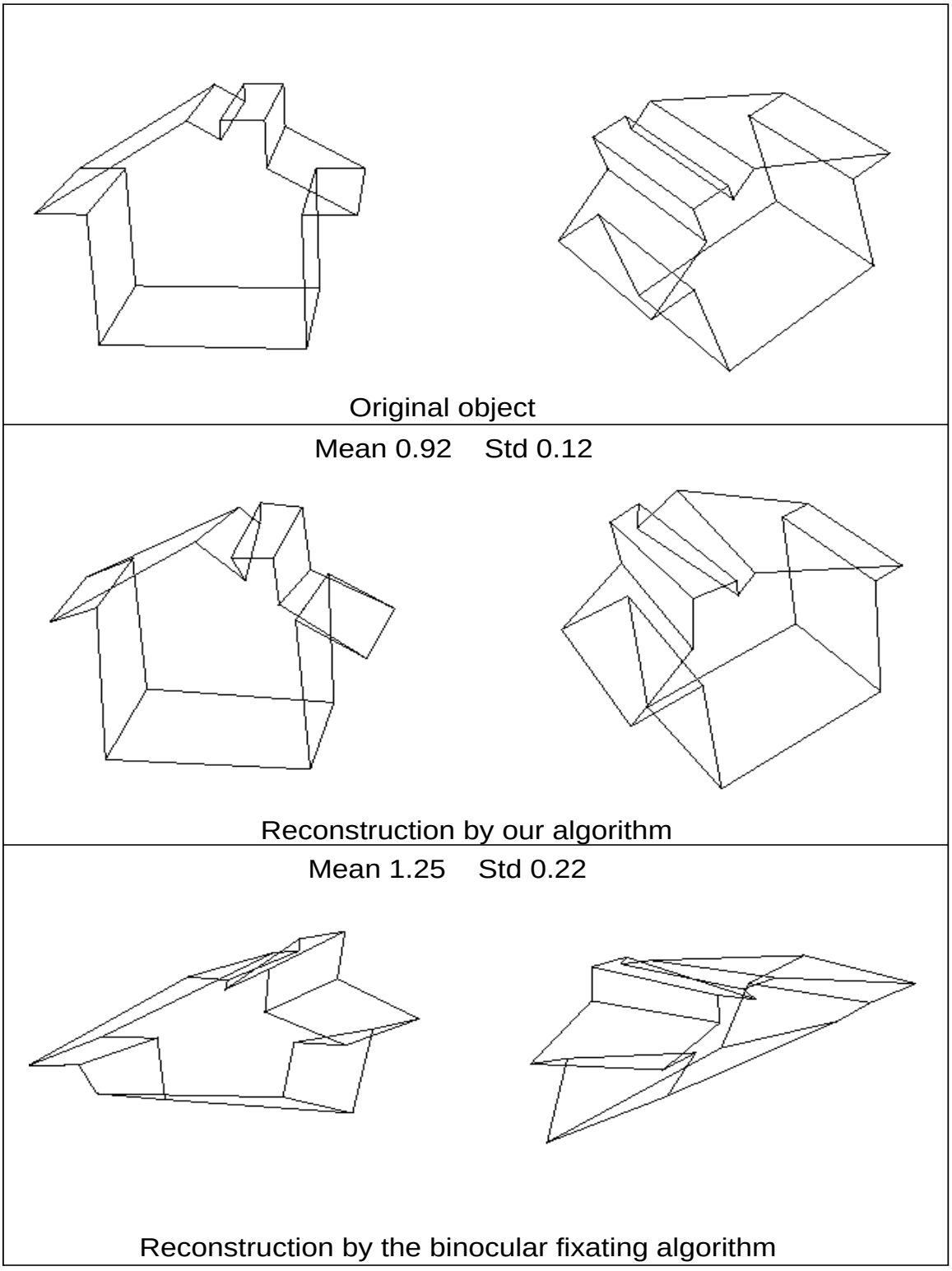


Fig. 5.10. Two different views of the original object, and its reconstructions are shown. The means and the standard deviations of the normalized ratios are provided.

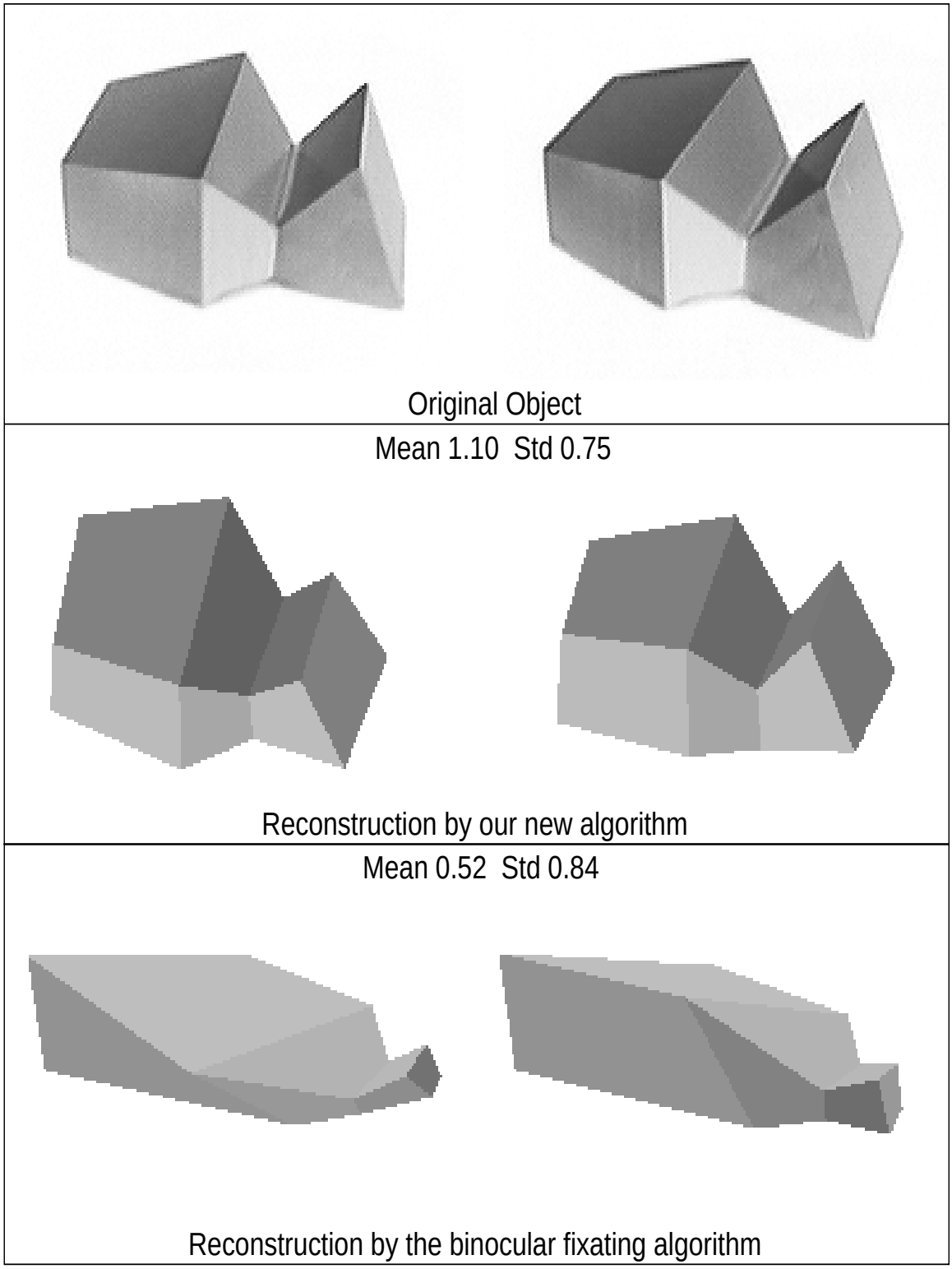


Fig. 5.11. Real stereo images of a wooden object, and two different views of its reconstructions.

6. PSYCHOPHYSICAL TEST OF THE NEW METHOD

Based on the results of prior research on human monocular shape perception and perception of shape from motion, we formulated a new algorithm for binocular shape reconstruction. To verify if human binocular shape reconstruction is related to this algorithm, we performed a psychophysical experiment that is analogous to that described in Section 4.2. If this algorithm is a model of human binocular shape perception, then using stimuli that do not provide constraints that were assumed by the algorithm should harm binocular shape perception.

6.1 Method

6.1.1 Subjects

Four subjects participated. They all had corrected to normal vision. ZP and AS received extensive practice, MC (the author) did not receive extensive practice but was experienced, AM did not receive extensive practice and was not experienced. The subject's head was supported by a chin-forehead rest placed $50cm$ away from the monitor.

6.1.2 Stimuli

We used the same stimuli that were used by Pizlo and Stevenson in their psychophysical experiment (see page 53 for a description of individual conditions). The stimuli were stationary and displayed by means of a stereoscopic display. The left and right images were computed using the same method as described in Section 2.1.2. The simulated viewing distance D was $50cm$, and the simulated interocular distance I was $7cm$.

6.1.3 Procedure

The procedure was the same as that used in the experiment on shape from motion (see Section 4.2 [PS99]). Each session consisted of 200 "same" and 200 "different" trials, presented in a random order. The subject's performance was evaluated by the discriminability measure d′ and its standard error. A higher d′ represents more accurate performance.

6.2 Results and Discussion

Figures 6.1 - 6.4 show the results obtained with conditions 1-3. The average results of these four graphs are shown in Figure 6.5. Overall, the performance was the best in condition 1. Condition 2 and 3 led to much worse performance than condition 1. This shows that binocular shape perception is more accurate if the objects contain a number of constraints, and monocular cues related to 3-D topology.

Figure 6.6 shows the average results of ZP and AS obtained with conditions 1, 4, 5, and 6. Conditions 4 and 5 led to slightly worse performance than condition 1. At the same time, condition 6 led to performance that was much worse than that in condition 1. This shows that planarity of surface contours and symmetry constraints are important in binocular shape perception.

Figure 6.7 shows the average results of ZP and AS obtained with conditions 1, 7, 8, and 9. It is seen that conditions 7-9 led to much worse performance than condition 1. This shows that topological constraints are also very important in binocular shape perception.

These results are essentially identical to those described in the shape from motion experiment. This shows that human binocular shape perception uses the same constraints that are used in shape perception from a single image and from motion.

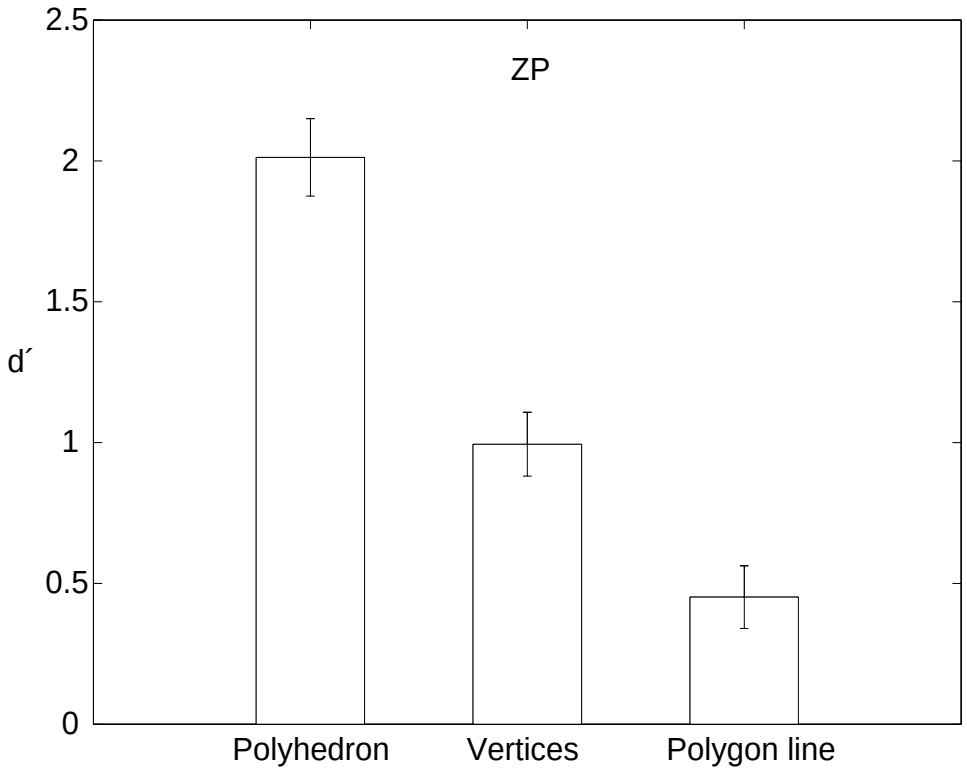


Fig. 6.1. Results of the psychophysical experiment with conditions 1) polyhedron, 2) vertices, and 3) polygonal line. The ordinate shows d′, which is a discriminability measure in signal detection theory. The error bars show +/- standard deviation of the mean.

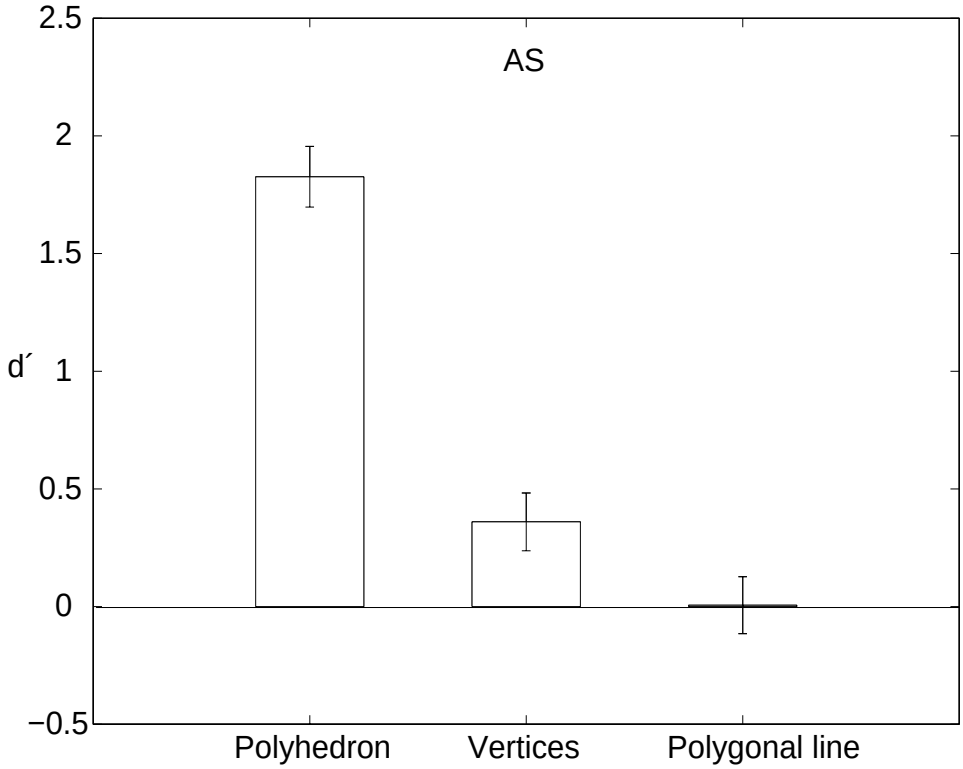


Fig. 6.2. Results of the psychophysical experiment with conditions 1) polyhedron, 2) vertices, and 3) polygonal line. The ordinate shows d′, which is a discriminability measure in signal detection theory. The error bars show +/- standard deviation of the mean.

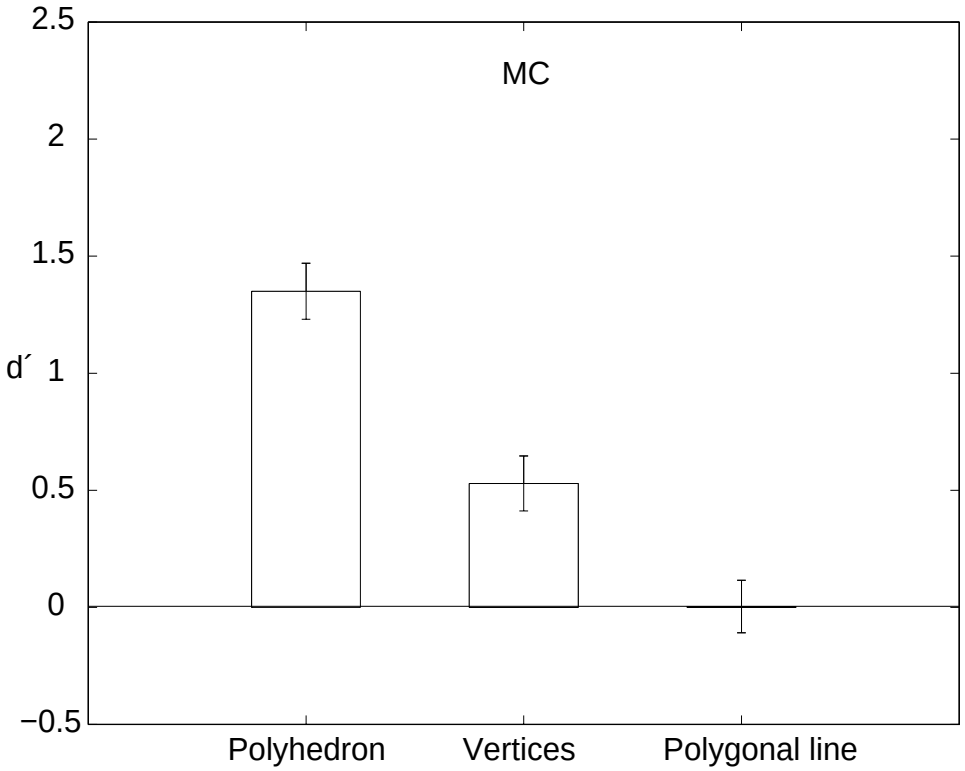


Fig. 6.3. Results of the psychophysical experiment with conditions 1) polyhedron, 2) vertices, and 3) polyhedron line. The ordinate shows d′, which is a discriminability measure in signal detection theory. The error bars show +/- standard deviation of the mean.

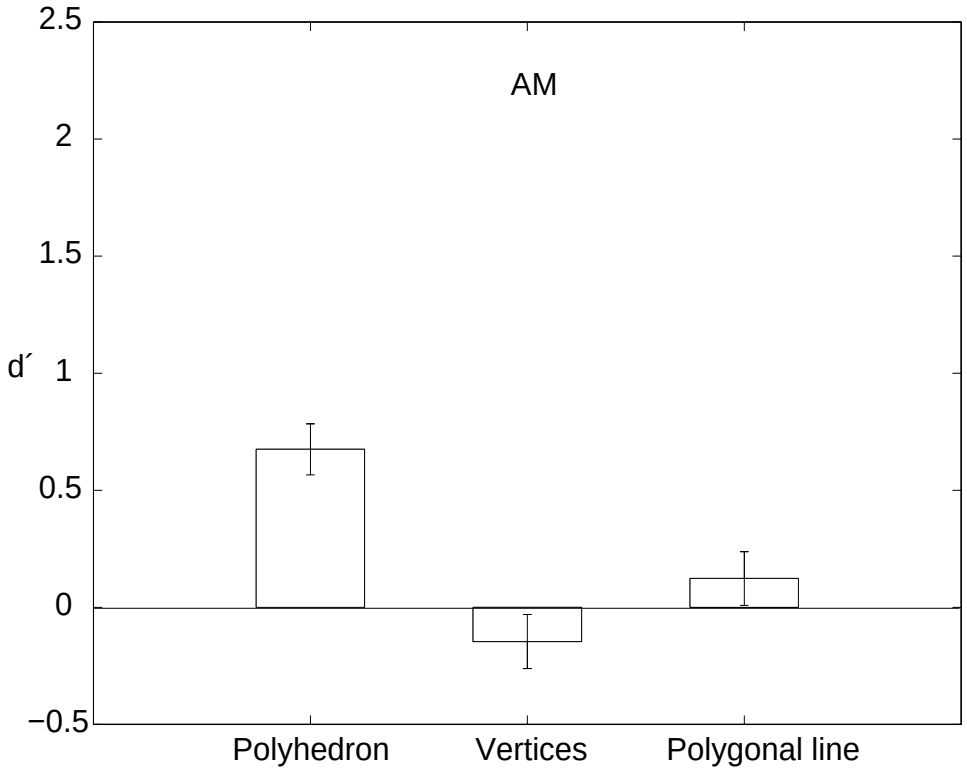


Fig. 6.4. Results of the psychophysical experiment with conditions 1) polyhedron, 2) vertices, and 3) polyhedron line. The ordinate shows d′, which is a discriminability measure in signal detection theory. The error bars show +/- standard deviation of the mean.

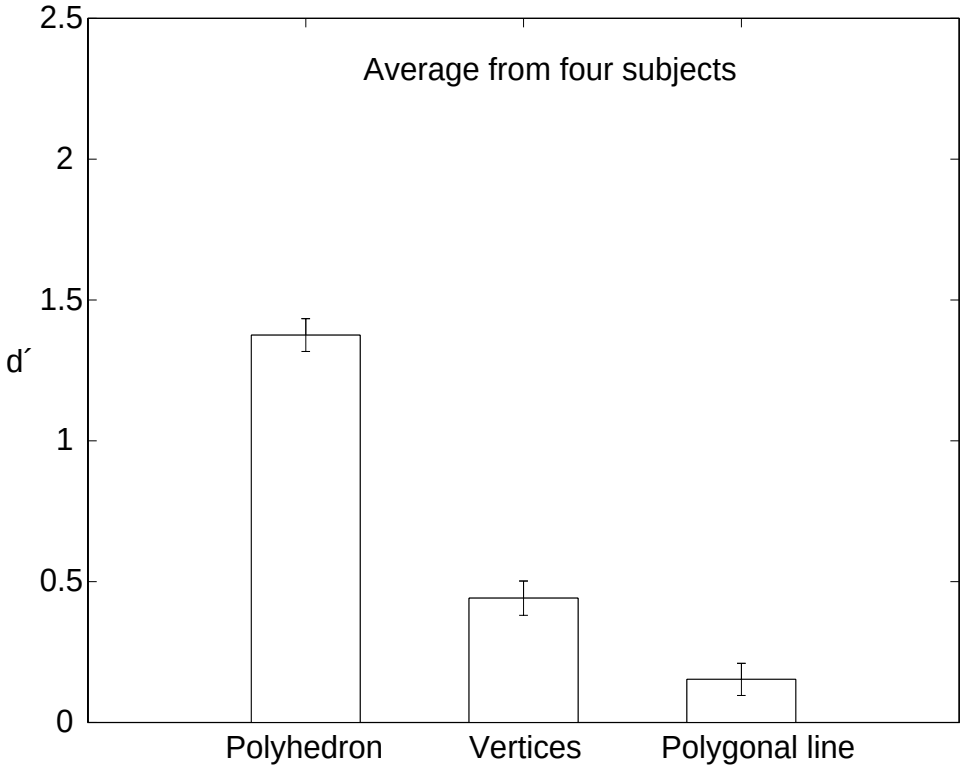


Fig. 6.5. Results of the psychophysical experiment (average from four subjects) with conditions 1) polyhedron, 2) vertices, and 3) polygonal line. The ordinate shows d′, which is a discriminability measure in signal detection theory. The error bars show +/- standard deviation of the mean.

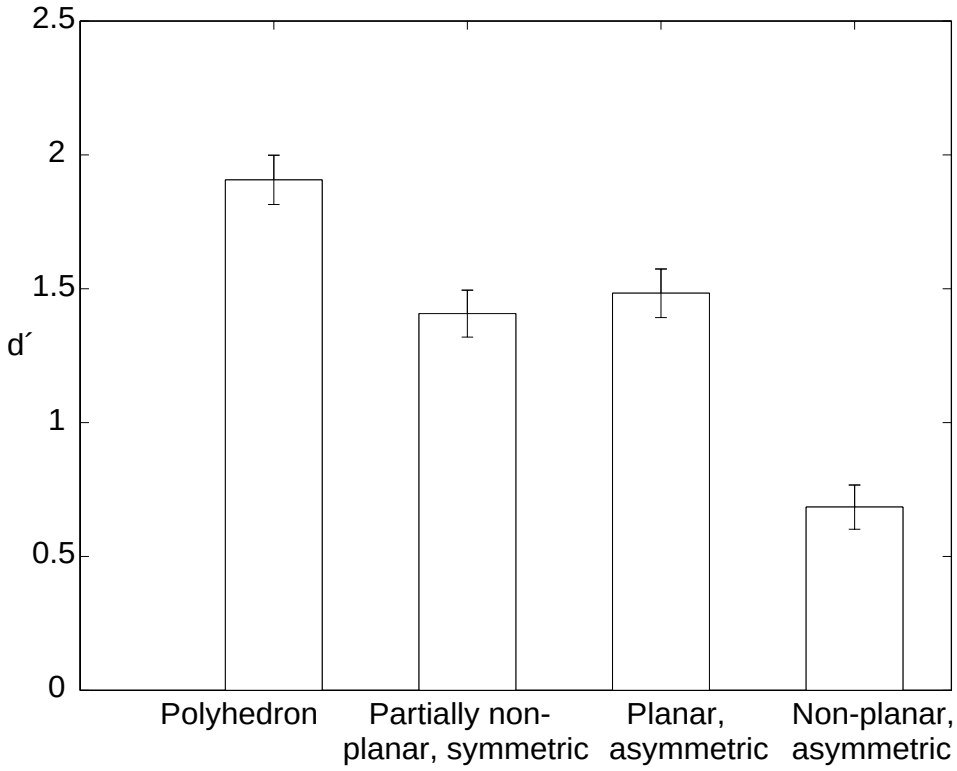


Fig. 6.6. Results of the psychophysical experiment (average from ZP and AM) with conditions 1) polyhedron, 4) partially non-planar, symmetric polyhedron, 5) planar, asymmetric polyhedron, and 6) non-planar, asymmetric polyhedron. The ordinate shows d′, which is a discriminability measure in signal detection theory. The error bars show +/- standard deviation of the mean.

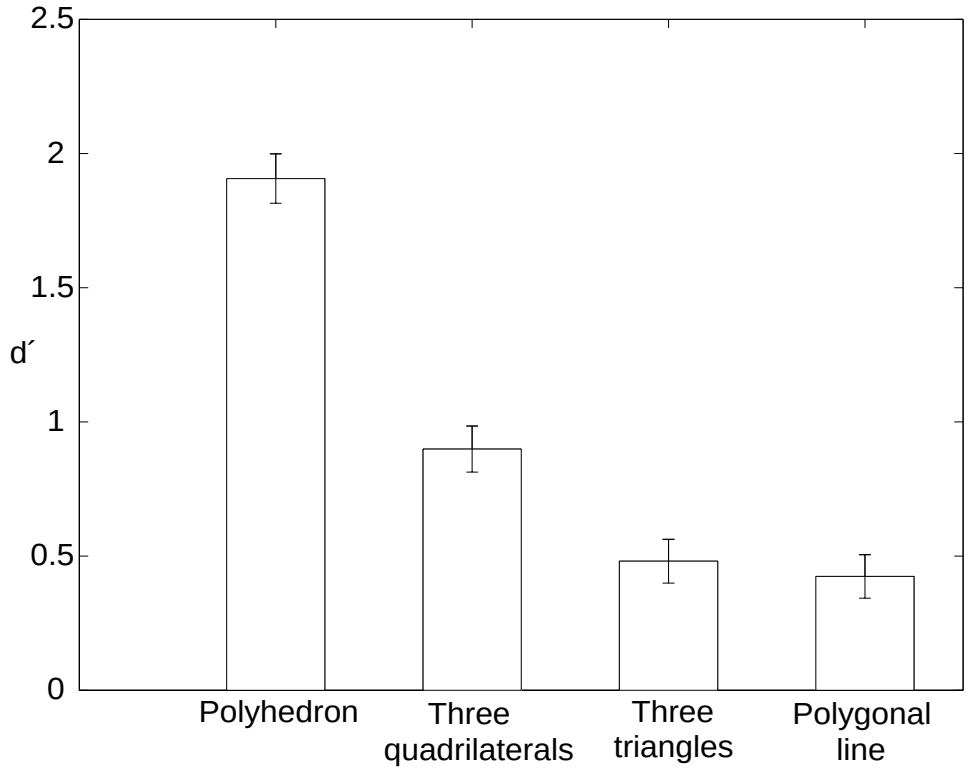


Fig. 6.7. Results of the psychophysical experiment (average from ZP and AM) with conditions 1) polyhedron, 7) three quadrilaterals, 8) three triangles, and 9) polygonal line. The ordinate shows d′, which is a discriminability measure in signal detection theory. The error bars show +/- standard deviation of the mean.

7. CONTRIBUTIONS AND FUTURE WORK

7.1 Summary of Contributions

We have evaluated the psychological plausibility of the 8 point algorithm, which is one of the best known computer vision algorithm for binocular shape reconstruction. Specifically, we compared the performance of the 8 point algorithm to the performance of human subjects in a task of binocular shape reconstruction. We found that the 8 point algorithm is very inaccurate and very unstable when a small amount of noise is added to the image coordinates. This suggests that constraints must be used to stabilize the perceptual solution. We considered two types of constraints: (1) constraints of the viewing system; and (2) constraints of the geometrical properties of the object.

We modified the 8 point algorithm by adding a binocular fixation constraint (system constraint) to it. The average performance of this modified algorithm, which we call binocular fixating algorithm, is substantially better than that of the 8 point algorithm. However, it still is very unstable as compared to human subjects. These results show that 3-D shape reconstructions obtained directly from binocular disparity are very unreliable, and imply that geometrical constraints of the objects must be used to restrict the family of solutions.

It is important to note that the idea of using constraints in binocular shape reconstruction is not new in computer vision literature. However, existing computer vision algorithms for binocular shape reconstruction use only smoothness constraint. This constraint is not appropriate to model surface discontinuity and wire frame objects. It is known from prior psychophysical experiments that topological constraints and constraints such as planar surface contours and symmetry of the object are critical

in shape perception from a single image [AF69] [HB60] [Per72], and shape perception from motion [PS99]. We implemented these constraints in our computer algorithm. Moreover, we used a new approach to avoid the instability caused by binocular disparity. Namely, we begin 3-D shape reconstruction monocularly, and only then use binocular disparity to improve the accuracy of shape reconstruction.

We tested this new algorithm using synthetic and real stereo images. The performance of this algorithm is quite robust in the presence of noise. It is shown in Figure 5.4 that 1% of noise is very harmful to the shapes reconstructed by the binocular fixating algorithm. In fact, the errors are so large that they often lead to topological reversals in 3-D. But the performance of the monocular shape reconstruction is quite stable with 1% of noise in the images. This result is quite surprising because it is commonly assumed that shape reconstruction is more accurate and more stable under binocular viewing as opposed to monocular viewing. Here we show that shape reconstruction that uses monocular cues related to 3-D topology and geometrical constraints is more stable than reconstruction directly from binocular disparity. Furthermore, the correction involving the second image results in even more accurate reconstruction. This shows that our new algorithm can successfully use both the information in the images and constraints.

To evaluate the psychological plausibility of our new algorithm, we performed psychophysical experiments to verify if the constraints that are used in our algorithm are indeed used by the human visual system under binocular viewing. Results show that objects that contain a number constraints lead to better performance. This implies shows that constraints are involved in human binocular shape perception. This result is new: existing theories of human binocular shape reconstruction use purely bottom - up (data driven) approach (see [HR95c] for a comprehensive review of human binocular vision).

Finally, the fact that our new algorithm uses monocular cues related to 3-D topology and constraints such as planar surface contours and symmetry of the object has interesting implications. It suggests that the same set of constraints is used in both

monocular and binocular shape reconstruction, and that the mechanisms involved in monocular shape reconstruction are similar to those of binocular shape reconstruction. Interestingly, in human vision literature, it has been conjectured that binocular vision begins with "elaborated uniocular sensations" [She47] [Ram86] [SB87]. In other words, human binocular system was assumed to start with reconstructing two shapes monocularly, and then combining them to improve the reconstruction accuracy. To our knowledge, our algorithm is the first computational implementation of this conjecture.

7.2 Future Work

We provide suggestions that can be used to further improve the performance of this algorithm, and discussions on how this algorithm can be modified to handle curved surfaces.

1) Our algorithm uses binocular disparity to obtain a relative depth ordering as a starting point of a minimization process, in which a monocular shape reconstruction is found. This relative depth ordering is also used as a constraint to reduce the computational time during the monocular shape reconstruction. In the case of small vergence angles and noise in the images, large errors can be obtained in the relative depth ordering. This causes the performance of the monocular shape reconstruction to drop slightly. To improve the performance of monocular shape reconstruction, one can develop a better algorithm to obtain a relative depth ordering, or use a computationally more expensive minimization technique such as simulated annealing to obtain a monocular shape reconstruction.

2) In order to use the information from the second image to correct a shape hypothesis, we applied a projective transformation with 4 degrees of freedom, in which the 3-D points of the shape are enforced to remain on their visual rays, and the planarity of the surface contours of the shape is preserved. This transformation is involved in the process of minimizing a sum of weighted terms, which includes constraints and binocular disparity. It is shown in our simulation experiments that

this transformation can be used to correct an object which contains as many as 22 points (Figure 5.10). However, this does not imply that this projective transformation with 4 degrees of freedom can be used to correct any type of objects. To correct a more complicated object, one can segment the object into parts, and then correct the parts individually.

3) At the final stage where the corrected shape hypotheses are combined, our algorithm selects a final shape from two shape hypotheses. To obtain a more stable shape reconstruction, it is possible to extend our current algorithm to generate more than two shape hypotheses, and then combine them. This would further reduce the errors obtained in these shape hypotheses. One way to generate more shape hypotheses is to use a sequence of images (motion).

4) Finally, we point out that constraints such as MSDA and planarity of contours are likely to be useful in reconstructing curved surfaces as well. First, instead of minimizing the standard deviation of angles of a polyhedral object, one can minimize the standard deviation of curvature (MSDC) along the contours of a curved object. Second, it has been suggested that if a smoothly curved object contains planar contours, then it often leads to accurate shape perception. Specifically, Stevens [Ste86] pointed out that humans assume that surface contours are planar geodesics. Although, this claim was not directly verified by psychophysical experiments, it is consistent with prior results [Piz94]. The question remains open as to whether MSDC and planarity of surface contours are sufficient for reliable reconstruction.

LIST OF REFERENCES

[AB89] J. Aloimonos and C. M. Brown. On the kinetic depth effect. *Biological Cybernetics*, 60:445–455, 1989.

[AF69] F. Attneave and R. Frost. The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psychophysics*, 6:391–396, 1969.

[BdAHB96] M. J. Brooks, L. de Agapito, D. Q. Huynh, and L. Baumela. Direct methods for self-calibration of a moving stereo head. In *Proceedings, European Conference on Computer Vision, University of Cambridge, England*, pages 415–426, 1996.

[Bel96] P. N. Belhumeur. A bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19:237–260, 1996.

[BF82] S. T. Barnard and M. A. Fischler. Computational stereo. *Computing Surveys*, 14:553–572, 1982.

[BPT88] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76:869–889, 1988.

[BS93] C. Bouman and K. Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on Image Processing*, 2:296–310, 1993.

[CF85] R. Cormack and R. Fox. The computation of retinal disparity. *Perception and Psychophysics*, 37:176–178, 1985.

[CN95] R. Chung and R. Nevatia. Use of monocular groupings and occlusion analysis in a hierarchial stereo system. *Computer Vision and Image Understanding*, 62:245–268, 1995.

[CPC96] M. W. Chan, Z. Pizlo, and D. M. Chelberg. Shape reconstruction by a binocular fixating system. *Purdue University TR 96-18*, 1996.

[CPC99] M. W. Chan, Z. Pizlo, and D. M. Chelberg. Binocular shape reconstruction: Psychological plausibility of the 8 point alogirthm. *Computer Vision and Image Understanding, in press*, 1999.

[CPS99] M. W. Chan, Z. Pizlo, and A. K. Stevenson. A new regularization method for binocular shape reconstruction. *In preparation*, 1999.

[CY90] J.J. Clark and A.L. Yuille. *Data fusion for sensory information processing systems.* Kluwer, Boston, MA, 1990.

[DA89] U. R. Dhond and J. K. Aggarwal. Structure from stereo - a review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19:1489–1510, 1989.

[DLN+90] K. K. DeValois, V. Lakshminarayanan, R. Nygaard, S. Schlussel, and J. Sladky. Discrimination of relative spatial position. *Vision Research*, 30:1649–1660, 1990.

[DPOR95] F. H. Durgin, D. R. Proffitt, T. J. Olson, and K. S. Reinke. Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance*, 21:679–699, 1995.

[EW87] R. D. Eastman and A. N. Waxman. Using disparity functionals for stereo correspondence and surface reconstruction. *Computer Vision, Graphics and Image Processing*, 39:73–101, 1987.

[FA93] C. Fermüller and Y. Aloimonos. The role of fixation in visual motion analysis. *International Journal of Computer Vision*, 11:165–186, 1993.

[Fau92] O. D. Faugeras. What can be seen in 3-D with an uncalibrated stereo rig? In *Proceedings, European Conference on Computer Vision, Santa Margherita, Italy*, pages 563–578, 1992.

[Fau93] O. Faugeras. 3-D *Computer Vision: A Geometric Viewpoint*, chapter 3. MIT Press, Cambridge, MA, 1993.

[FL95] P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16:35–56, 1995.

[FL96] P. Fua and Y. G. Leclerc. Taking advantage of image-based and geometry-based constraints to recover 3-d surfaces. *Computer Vision and Image Understanding*, 64:111–127, 1996.

[FM90] O. D. Faugeras and S. Maybank. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4:225–246, 1990.

[GRB96] A. Glennerster, B. J. Rogers, and M. F. Bradshaw. Stereoscopic depth constancy depends on the subject's task. *Vision Research*, 36:3441–3456, 1996.

[Gri81] W. E. L. Grimson. A computer implementation of a theory of human stereo vision. *Philosophical Transactions of the Royal Society of London*, B 292:217–253, 1981.

[Gri85] W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:17–34, 1985.

[HA89] W. Hoff and N. Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:121–136, 1989.

[Har92] R. I. Hartley. Estimation of relative cameras position for uncalibrated cameras. In *Proceedings, European Conference on Computer Vision, Santa Margherita, Italy*, pages 579–587, 1992.

[Har95] R. I. Hartley. In defence of the 8-point algorithm. In *Proceedings, International Conference on Computer Vision, MIT, MA*, pages 1064–1070, 1995.

[Har97] R. Hartley. Self-calibration of stationary cameras. *International Joural of Computer Vision*, 22:5–23, 1997.

[HB60] J. Hochberg and V. Brooks. The psychophysics of form: Reversible-perspective drawings of spatial objects. *The American Journal of Psychology*, 73:337–355, 1960.

[HGC92] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings, Computer Vision and Pattern Recognition, Champaign, IL*, pages 761–764, 1992.

[HNH90] R. J. Holt, A. N. Netravali, and T. S. Huang. Experience in using homotopy methods to solve motion estimation problems. In *Proceedings, SPIE, Curves and Surfaces in Computer Vision and Graphics, Bellingham, WA*, volume 1251, pages 219–226, 1990.

[HR95a] I. P. Howard and B. J. Rogers. *Binocular Vision and Stereopsis*, chapter 11. Oxford University Press, New York, 1995.

[HR95b] I. P. Howard and B. J. Rogers. *Binocular Vision and Stereopsis*, chapter 10. Oxford University Press, New York, 1995.

[HR95c] I. P. Howard and B. J. Rogers. *Binocular Vision and Stereopsis*. Oxford University Press, NY, 1995.

[HS97] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68:146–157, 1997.

[HWJ88] D. E. Hinkle, W. Wiersma, and S. G. Jurs. *Applied Statistics for the Behavioral Sciences*, chapter 13. Houghton Mifflin Company, Boston, MA, 1988.

[Joh91] E. B. Johnston. Systematic distortions of shape from stereopsis. *Vision Research*, 31:1351–1360, 1991.

[Kru13] E. Kruppa. Zur ermittlung lines objektes aus zwei perspektiven mit innerer orientierung. *Sitz. -Ber. Akad. Wiss., Wien, Math, Naturw. Kl., Abt. Ila.*, 122:1939–1948, 1913.

[KvD91] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *Journal of the Optical Society of America A*, 8:377–385, 1991.

[LBW96] F. Li, M. Brady, and C. Wiles. Fast computation of the fundamental matrix for an active stereo vision system. In *Proceedings, European Conference on Computer Vision, University of Cambridge, England*, pages 157–166, 1996.

[Lec89] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[LF92] Y. G. Leclerc and M. A. Fischler. An optimization-based approach to the interpretation of single line drawings as 3-D wire frames. *International Journal of Computer Vision*, 9:113–136, 1992.

[LF96] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: Theory, algorithm, and stability analysis. *International Journal of Computer Vision*, 17:43–76, 1996.

[LH81] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[LH84] H. C. Longuet-Higgins. The reconstruction of a scene from two projections - configurations that defeat the 8-point algorithm. In *Proceedings, The First Conference on AI Applications, Denver, CO*, pages 395–397, 1984.

[Lip91] L. Lipton. *The Crystaleyes Handbook*. StereoGraphics Corporation, CA, 1991.

[LMY90] M. Landy, L. Maloney, and M. Young. Psychophysical estimation of the human depth combination rule. In *Proceedings, SPIE, Sensor Fusion III: 3-D Perception and Recognition, Bellingham, WA*, volume 1383, pages 247–254, 1990.

[Low87] D. G. Lowe. 3-D object recognition from single 2-D images. *Artificial Intelligence*, 31:355–395, 1987.

[Mar88] R. March. Computation of stereo disparity using regularization. *Pattern Recognition Letters*, 8:181–187, 1988.

[Mar89] R. March. A regularization model for stereo vision with controlled continuity. *Pattern Recognition Letters*, 10:259–263, 1989.

[Mar91] T. Marill. Emulating the human interpretation of line-drawings as 3-D objects. *International Journal of Computer Vision*, 6:147–161, 1991.

[Mat92] L. Matthies. Stereo vision for planetary rovers: Stochastic modeling to near real time implementation. *International Journal of Computer Vision*, 8:71–91, 1992.

[ML89] L. Maloney and M. Landy. A statistical framework for robust fusion of depth information. In *Proceedings, SPIE, Visual Communications and Image Processing (Part 2)*, pages 1154–1163, 1989.

[ML92] H. Maître and W. Luo. Using models to improve stereo reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:269–277, 1992.

[MLB90] S. P. McKee, D. M. Levi, and S. F. Bowne. The imprecision of stereopsis. *Vision Research*, 30:1763–1779, 1990.

[MLH82] J. E. W. Mayhew and H. C. Longuet-Higgins. A computational model of binocular depth perception. *Nature*, 297:376–378, 1982.

[MMP87] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82:76–89, 1987.

[Neg90] S. Negahdaripour. Multiple interpretations of the shape and motion of objects from two perspective images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:1025–1039, 1990.

[NTPT96] J. F. Norman, J. T. Todd, V. J. Perotti, and J. S. Tittle. The visual perception of 3-D length. *Journal of Experimental Psychology: Human Perception and Performance*, 22:173–186, 1996.

[Per72] D. N. Perkins. Visual discrimination between rectangular and nonrectangular parallelopipeds. *Perception and Psychophysics*, 12:396–400, 1972.

[Piz94] Z. Pizlo. A theory of shape constancy based on perspective invariants. *Vision Research*, 34:1637–1658, 1994.

[PR92] Z. Pizlo and A. Rosenfeld. Recognition of planar shapes from perspective images using contour-based invariants. *Computer Vision, Graphics and Image Processing: Image Understanding*, 56:330–350, 1992.

[PS98] Z. Pizlo and M. R. Scheessele. Perception of 3-D scenes from pictures. In *Proceedings, SPIE*, 1998.

[PS99] Z. Pizlo and A. K. Stevenson. Shape constancy from novel views. *Perception and Psychophysics, in press*, 1999.

[PSG95] Z. Pizlo and M. Salach-Golyska. 3-D shape perception. *Perception and Psychophysics*, 57:692–714, 1995.

[PT84] T. Poggio and V. Torre. Ill-posed problems and regularization analysis in early vision. *Artificial Intelligence Laboratory Memo, MIT*, 773, 1984.

[PTK85] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.

[Ram86] V. Ramachandran. Capture of stereopsis and apparent motion by illusory contours. *Perception and Psychophysics*, 39:361–373, 1986.

[RB93] B. J. Rogers and M. F. Bradshaw. Vertical disparity, differential perspective and binocular stereopsis. *Nature*, 361:253–255, 1993.

[RCF95] C. Rothwell, G. Csurka, and O. Faugeras. A comparison of projective reconstruction methods for pairs of views. In *Proceedings, International Conference on Computer Vision, MIT, MA*, pages 932–937, 1995.

[RH94] D. Raviv and M. Herman. A unified approach to camera fixation and vision-based road following. *IEEE Transactions on Systems, Man, and Cybernetics*, 24:1125–1141, 1994.

[rL94] J. Gårding and T. Lindeberg. Direct estimation of local surface shape in a fixating binocular vision system. In *Proceedings, European Conference on Computer Vision, Stockholm, Sweden*, pages 365–376, 1994.

[rPMF95] J. Gårding, J. Porrill, J. E. W. Mayhew, and J. P. Frisby. Stereopsis, vertical disparity and relief transformation. *Vision Research*, 35:703–722, 1995.

[SA89] M. E. Spetsakis and J. Aloimonos. Optimal motion estimation. In *Proceedings, Workshop for Visual Motion, Irvine, CA*, pages 229–237, 1989.

[SB87] K. A. Stevens and A. Brookes. Depth reconstruction in stereopsis. In *Proceedings, International Conference on Computer Vision, London, England*, pages 682–686, 1987.

[She47] C. S. Sherrington. *The Integrative Action of the Nervous System*. Yale University Press, New Haven, CT, 1906/1947.

[SP96] S. Soatto and P. Perona. Motion from fixation. In *Proceedings, Computer Vision and Pattern Recognition, San Francisco, CA*, pages 817–824, 1996.

[SP97] I. Shimshoni and J. Ponce. Recovering the shape of polyhedra using line-drawing analysis and complex reflectance models. *Computer Vision and Image Understanding*, 65:296–310, 1997.

[Sta45] B. K. Stavrianos. The relation of shape perception to explicit judgments of inclination. *Archives of Psychology*, 296:1–94, 1945.

[Ste86] K. Stevens. *From Pixels to Predicates. A. Pentland (Ed)*, chapter Inferring Surfaces from Contours Across Surfaces. Ablex Publishing Corporation, Norwook, NJ, 1986.

[Sug86] K. Sugihara. *Machine Interpretation of Line Drawings.* MIT Press, Cambridge, MA, 1986.

[TA77] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* Winston and Sons, Washington, DC, 1977.

[Taa92] M. A. Taalebinezhaad. Direct recovery of motion and shape in the general case by fixation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:847–853, 1992.

[Ter86] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:413–424, 1986.

[TH84] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of 3-D motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

[Tho31] R. H. Thouless. Phenomenal regression to the real object. II. *British Journal of Psychology*, 22:1–30, 1931.

[Tho59] E. H. Thompson. A rational algebraic formulation of the problem of relative orientation. *Photogrammetric Record*, 3:152–159, 1959.

[VC95] V. Venkateswar and R. Chellappa. Hierarchical stereo and motion correspondence using feature groupings. *International Journal of Computer Vision*, 15:245–269, 1995.

[Wat87] R. J. Watt. Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A*, 4:2006–2021, 1987.

[WHA89] J. Weng, T. S. Huang, and N. Ahuja. Motion and structure from two perspective views: Algorithm, error analysis, and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:451–476, 1989.

[Whe38] C. Wheatstone. Contributions to the physiology of vision - part the first. On some remarkable and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128:371–394, 1838.

[WTK87] A. Witkin, D. Terzopoulos, and M. Kass. Signal matching through scale space. *International Journal of Computer Vision*, 1:133–144, 1987.

VITA

Moses W. Chan received his BSCEE and MSEE degrees from Purdue University in 1991 and 1993 respectively. He is currently pursuing his Ph.D. degree in Electrical and Computer Engineering. While at Purdue, he has served as president, treasurer, and student representative of the Electrical and Computer Engineering Graduate Students Association. He is a recipient of the 1997-1998 Magoon Award for Excellence in Teaching, and the 1999 Purdue University Graduate Student Award for Outstanding Teaching. His research interests include computer vision, image processing, computer graphics, and visual psychophysics.

Mr. Chan also interned with Motorola Incorporated during the summers of 1995-1997, where he developed English, Spanish, and Chinese text compression protocols of paging products in Boynton Beach, Florida.