Journal of Experimental Psychology:
Human Perception and Performance
1995, Vol. 21, No. 3, 679-699

Copyright 1995 by the American Psychological Association, Inc.
0096-1523/95/$3.00

Comparing Depth From Motion With Depth From Binocular Disparity

Frank H. Durgin, Dennis R. Proffitt,
and Thomas J. Olson
University of Virginia

Karen S. Reinke
University of Arizona

The accuracy of depth judgments that are based on binocular disparity or structure from motion (motion parallax and object rotation) was studied in 3 experiments. In Experiment 1, depth judgments were recorded for computer simulations of cones specified by binocular disparity, motion parallax, or stereokinesis. In Experiment 2, judgments were recorded for real cones in a structured environment, with depth information from binocular disparity, motion parallax, or object rotation about the *y*-axis. In both of these experiments, judgments from binocular disparity information were quite accurate, but judgments on the basis of geometrically equivalent or more robust motion information reflected poor recovery of quantitative depth information. A 3rd experiment demonstrated stereoscopic depth constancy for distances of 1 to 3 m using real objects in a well-illuminated, structured viewing environment in which monocular depth cues (e.g., shading) were minimized.

It has been pointed out that the geometric information supporting the perception of depth from binocular disparity is actually less determinate than that supporting the recovery of structure from object rotation or motion parallax (Richards, 1985). Stationary binocular stereopsis provides 2 two-dimensional perspectives on a three-dimensional object, whereas object rotation or motion parallax provides a continuous sampling of two-dimensional perspectives in the form of relative motions. The latter situations, which entail having information obtained from more than two views, can theoretically provide a more accurate (Euclidean) geometric recovery of the distal object's three-dimensional configuration, so long as object rigidity is assumed. However, this analysis makes other assumptions about the kinds of information available to an observer that are probably not warranted in practice. The studies presented here show that, within the range of parameters investigated for real and simulated objects, quantitative depth judgments that are based on information from object rotation and motion parallax are much less accurate than those that are based on binocular disparity.

Geometry of Disparity and Relative Motion Information

When an object is rotated about an axis other than the line of sight, the relative motions of features on the object can specify the three-dimensional structure of the object to monocular vision (see the Appendix for a mathematical analysis). A similar recovery of structure can be obtained if the object and the viewer undergo an angular displacement with respect to one another. Although the geometry of displacement differs from that of rotation, in both cases the object is observed from more than one angle. If the observer maintains fixation on the object during angular displacement, the two cases are practically identical for small rotations or displacements. Note that it is the angular rotation of the object with respect to the viewer that is crucial; a simple translation toward or away from a monocular viewer does not provide geometric depth information.[1] Recovery of structure from object rotation is often referred to as the kinetic depth effect (KDE; Wallach & O'Connell, 1953; see also Braunstein, 1962). Depth information specified by viewer or object translation is referred to as motion parallax. Because both KDE and motion parallax with fixation may be shown to involve similar geometric analyses, we refer to both as structure from motion (SFM).[2]

Frank H. Durgin and Dennis R. Proffitt, Department of Psychology, University of Virginia; Thomas J. Olson, Department of Computer Science, University of Virginia; Karen S. Reinke, Department of Psychology, University of Arizona.

This research was supported by National Aeronautics and Space Administration Grants NAG2-812 and NCA2-708. Helpful comments on earlier drafts of this article were provided by J. Todd, J. Lappin, L. Kontsevich, and F. Pollick. We thank Jonathan Midgett and Melissa Salva for assisting in the collection of data. Steve Jacquot programmed the displays for Experiment 1 and helped Frank H. Durgin design the stereo viewer, which was constructed, as were the mechanical apparatuses of Experiments 2 and 3, by Ron Simmons.

Correspondence concerning this article should be addressed to Frank H. Durgin, who is now at the Department of Psychology, Swarthmore College, Swarthmore, Pennsylvania 19081, or to Dennis R. Proffitt, Department of Psychology, University of Virginia, Charlottesville, Virginia 22903. Electronic mail may be sent via Internet to fdurgin1@cc.swarthmore.edu.

[1] This assumes orthographic projection. Under perspective, translation toward the target does provide depth information. However, the information is unstable around the focus of expansion and hence is not likely to be useful.

[2] KDE is typically simulated by orthographic projection, which is of theoretical significance because it eliminates perspective information. The term *SFM* is often reserved for orthographic

The geometry of binocular disparity is similar to that of the two SFM cases. Viewing an object with two eyes provides views at two different angles to the object. Assuming central fixation, the two views of binocular disparity are equivalent to (a) a single eye displaced along the interocular distance or (b) two static frames of object rotation. These geometrical equivalences are illustrated in Figure 1, which shows how three points on an object bear the same relationship in the two represented views of (A) binocular disparity, (B) motion parallax with fixation, and (C) object rotation. If the eye is fixating the object in the motion parallax case (i.e., tracking), then the two views (D) are identical to those of rotation and binocular disparity. Note that our discussion here is geometric rather than psychological. For example, Simpson (1993) has reviewed the evidence that velocity rather than displacement information is used in the perceptual recovery of structure from motion, and other characterizations of the optical information available in these situations are possible (e.g., Lappin, 1990). Included in Figure 1A are the paths a single eye would take to generate KDE-like rotation (curved path) or motion parallax (straight, dashed line). The differences between the intermediate frames of continuous motion parallax and object rotation are fairly slight for small rotations and translations and, as we argue later, unimportant in the recovery of depth. The three points chosen here are not special. The same equivalence of geometry applies to every point on the object.

It is well known that a single pair of frames is insufficient to determine metric shape, unless the transformation between the two frames is known. In Figure 1D, The same relative displacement (or velocity) of the central point with respect to the others could arise from a shallow object with a large change of angular perspective or from a deep object with a small change in angular perspective. Disparity information, in itself, cannot distinguish[3] between a near, shallow object and a far, deep object because the angle of virtual rotation in the binocular disparity case is greater for nearer objects. In the motion cases, the angular rotation is not perfectly correlated with viewing distance, but the same ambiguity about the depth of the object would remain given only two frames of relative displacement (or relative velocity) information: It could be a deep object undergoing a small rotation or a shallow object undergoing a large rotation. The ambiguity we wish to emphasize concerns the shape of the object rather than the scaling factor associated with distance. Independent of stereo or motion, if one has only two views, one needs to know the angle between them. If this angle can be recovered, then so can the depth relative to the two-dimensional projected size of the object.

There are many reasons to believe, as suggested by Richards (1985), that depth information from motion is superior to that from disparity. Static stereopsis is inherently a two-frame problem. On the other hand, one has, with motion, a continuous sequence of frames. As shown later, the presence of this additional information is a necessary condition for recovering angular rotation information in a monocular SFM display. Furthermore, because interocular distance is fairly small, the angular rotation of an object provided by binocular disparity is less than 4° at a fixation distance of 1 m, whereas motion parallax information (e.g., information provided by moving around in an environment) can easily exceed that. However, a closer examination of the problem indicates that neither of these advantages is sufficient to produce accurate recovery of the angle of rotation for SFM with small rotations (e.g., less than about 15°). In the Appendix, we derive equations for determining object shape from relative velocity and from disparity information. Although velocity information is similar to disparity information, the disparity equation requires only knowledge of fixation distance to determine shape and size, whereas the motion cases require information about angular rotation to determine shape (scaled depth) and distance information to determine size.

Recovering Depth From SFM Displays

As just argued, displacement (or instantaneous velocity) information is insufficient to provide quantitative depth information in the absence of information about the object's quantity of rotation or angular velocity. To make this more precise, consider a rigid object centered at the origin and rotating about the y-axis as shown in Figure 2.

Assuming orthographic projection, an arbitrary point $\bar{X}$ will have image x-coordinate $R\cos(\theta)$ and image velocity $\dot{X} = -R\sin(\theta)\,\dot{\theta} = -Z\,\dot{\theta}$. That is, the depth of each point is given by its image velocity scaled by the inverse of the angular velocity. Thus, unless $\dot{\theta}$ is known or other information is available, depth can be known only up to a multiplicative constant.

Quantitative depth can be recovered if additional information is available. Differentiating $\dot{X}$ gives the image acceleration $\ddot{X} = -R\cos(\theta)\,\dot{\theta}^2 = -X\,\dot{\theta}^2$. Because X is known, acceleration determines $\dot{\theta}$ up to a sign reversal. The sign ambiguity can be resolved by perspective analysis or by heuristics that are based on assuming opaque objects. In any event, velocity plus acceleration is sufficient to provide quantitative estimates of object shape.

An alternative to using acceleration information is to observe changes in relative position over a noninfinitesimal interval and make use of some additional information. Con-

projection (usually KDE) so as to distinguish motion information, per se, from the perspective information available from motion parallax. We use the term SFM for both motion parallax and KDE (with or without orthographic projection, in the latter case). Our usage is primarily a matter of expository convenience, but there is *evidence that the underlying perceptual* processing is the same in either case (Braunstein, Liter, & Tittle, 1993).

[3] Vertical disparities, which can theoretically be used to solve for metric depth, are present in stereoscopic images, but this information is weak at points near fixation. Cumming, Johnston, and Parker (1991) have argued that vertical disparities and differential perspectives in stereoscopic images are not used. Rogers and Bradshaw (1993) have provided evidence that this information can be effective for very large displays (~80°), but we consider smaller displays here.

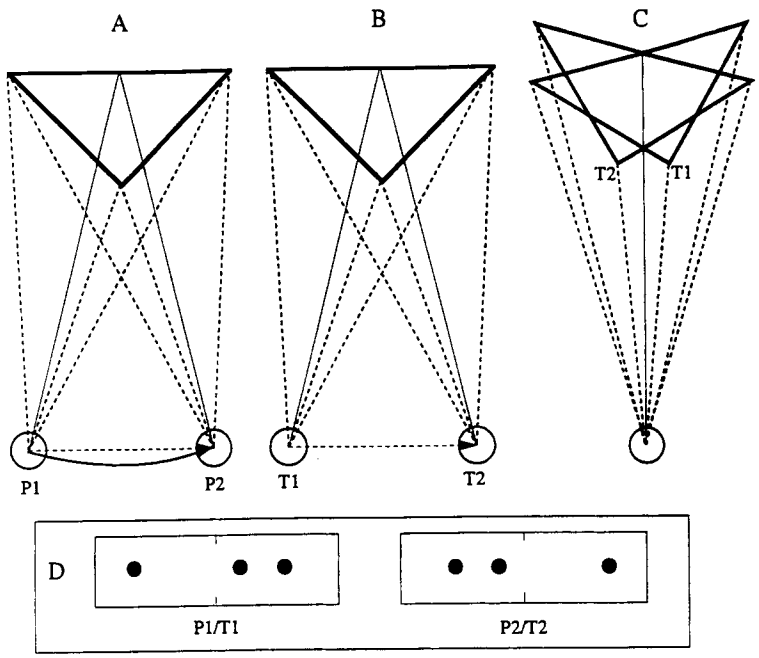


Figure 1. Geometric similarity among disparity, motion parallax, and object rotation. The same two views (D) are provided (A) binocular disparity (two simultaneous views from positions P1 and P2), (B) motion parallax (two successive views with a single displaced eye at times T1 and T2), and (C) object rotation (two monocular views of a rotating object). If a single fixating eye is displaced along a circular arc with a radius equal to the distance to the object, the geometry of the proximal stimulus is similar to the kinetic depth effect (rotation at a fixed distance). Such a curved path is illustrated (A), as is the straight motion parallax path. The line of sight to the center of the base is shown as fixed here, but any stationary point of fixation can be used.

sider two points with a common z-coordinate $\bar{X}_1 = (X_1, Y_1, Z)$ and $\bar{X}_2 = (X_2, Y_2, Z)$. (Such points are identifiable because their instantaneous velocities are the same.) After a rotation through the angle θ, the x-coordinates of the points become $X_1' = X_1 \cos(\theta) + Z \sin(\theta)$ and $X_2 \cos(\theta) + Z \sin(\theta)$, and the change in distance (foreshortening) between the two points is $X_2' - X_1' = (X_2 - X_1)\cos(\theta)$. In their analysis of the stereokinetic effect (SKE; Musatti, 1924), Proffitt, Rock, Hecht, and Schubert (1992) referred to the common $[Z \sin(\theta)]$ component of displacement as the between-contours motion and called the change in relative position $[X_1 \cos(\theta) - X_2 \cos(\theta)]$ the within-contour motion. The former contains all of the information about depth, but that informationis confounded with the rotation angle θ.

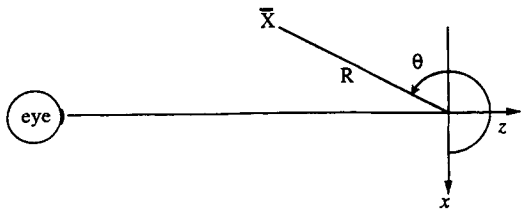


Figure 2. Imaging situation for structure from motion (rotation or motion parallax): A point, $\bar{X}$, at polar position (R,θ) from the intersection of the x axis and the z axis.

The latter (i.e., the foreshortening), in theory, can be used to specify θ and, thereby, to permit quantitative depth judgments.

Although velocity plus either acceleration or foreshortening suffices to determine the angle of rotation of an SFM stimulus, there is substantial evidence that this information is not always used by humans for recovery of metric depth information. For example, humans are quite bad at perceiving and judging acceleration (Gottsdanker, Frick, & Lockard, 1961; Simpson, in press). Moreover, Caudek and Proffitt (1993, 1994) compared depth judgments for SFM stimuli with those obtained with SKE, which differs only in that foreshortening information is absent. They found that, for small effective rotations, judgments were similar and that the presence of foreshortening information in the SFM cases had little effect on perceived shape. This result is consistent with the idea (Todd & Bressan, 1990) that human shape judgments typically do not involve effective use of higher order temporal relations such as acceleration or comparisons of more than two views of the scene.

The case can be restated in terms of pragmatic considerations: For an angular rotation of 15°, the foreshortening of the distance between any two points initially parallel to the projection plane will amount to a decrease in length of less than 4%. Distinguishing a total angular rotation of 10° from an angular rotation of 5° involves distinguishing a compres-

sion factor of 0.985 from a compression factor of 0.996. But whereas these compression factors differ by only 1%, the scaled object depth implied by the two angles (for a given velocity gradient) would differ by a factor of 2. We expect that it would be quite difficult to accurately discriminate this magnitude of compression information. It therefore seems that, for small object rotations or over small increments of rotation, the expected sensitivity of the motion system to quantity of rotation and therefore quantitative depth should not be very good.

Caudek and Proffitt (1993, 1994) have argued that perceived depth in such situations is determined more by heuristics and biases than by an appropriate geometry. Specifically, they reported (a) a bias to see SFM-specified objects as about as deep as they are wide and (b) a heuristic that objects with greater relative motions are deeper.

Before turning to the case of binocular disparity, we should comment on the angular information theoretically available in motion parallax. If an observer is viewing a translating object and knows that it is translating without rotating, then the angle of object rotation is given by the visual angle subtended by the translation. However, motion parallax produced by object translation suffers ambiguity on the grounds that, insofar as the accurate recovery of small angular rotations is difficult, it is pragmatically impossible for an observer to distinguish between an object that is translating and rotating and an object that is translating only. When the viewer is moving and the object is not, there are so many degrees of freedom of observer motion that the angle of rotation may again be difficult to determine visually, although the observer is quite aware of the moment-to-moment egocentric position of the object. Thus, for example, concomitant object rotation is sometimes reported in motion parallax displays, with some (appropriate) reduction in depth (M. E. Ono, Rivest, & Ono, 1986). A number of reports have indicated that depth from motion parallax and depth from object rotation are processed similarly, even when the motion parallax is produced by self-motion (Braunstein, Liter, & Tittle, 1993; Caudek & Proffitt, 1993, 1994; Loomis & Beall, 1993); other findings, however, have indicated some advantage to viewer-produced motion parallax (H. Ono & Steinbach, 1990). We believe that, for small rotations, angular information is not accurately recovered in either case.

Recovering Depth From Disparity

We have noted that disparity information, in itself, is inadequate to recover depth. Because depth specified by disparity varies as the square of distance, the same disparity information could specify a near, shallow object or a far, deep object (Richards, 1985). However, because interocular distance is a fixed quantity, the angle of virtual rotation is fully specified in the case of binocular disparity by knowledge of the distance to the point of zero disparity. Therefore, as shown in the Appendix, if precise information about the *distance to the object is available*, disparity information can be converted into absolute depth perception. Accommodation (the adjustment of the eye's lens necessary to focus images) and vergence (the rotation of the eyes for fixation in depth) have both been established as effective, if imperfect, indicators of distance that are yoked in active vision (see Owens, 1987, for a review). Moreover, there is normally an abundance of other sources of information about egocentric distance to objects, including perspective, familiar size, and so forth. Although using such distance information may seem redundant (if distance can be precisely known, so can depth), consider that the task of identifying fixation distance is different from the task of perceiving object shape. The visual system can accomplish the latter using disparity information if a single value of fixation distance is given. Thus, although the angle of virtual object rotation available to binocular viewing at typical distances is small, the precision and accuracy of depth perception are limited by the precision and accuracy of distance perception. In fact, it has been well established that perceived depth from disparity depends on perceived distance (H. Ono & Comerford, 1977), a fact ironically illustrated by the following.

One observation that has been (mistakenly) taken as support for the idea that SFM is treated as visually superior to binocular stereopsis is the well-replicated finding that stereoscopic judgments of object depth appear to be "recalibrated" after viewing of the rotation of a three-dimensional wire object through a telestereoscope, an apparatus that optically alters interocular separation and thus disparity (Wallach, Moore, & Davidson, 1963). The interpretation originally offered for this effect was that the disparity system was being recalibrated to a new interocular distance on the basis of information from the SFM system, and the intended implication was that SFM was more reliable than stereopsis. However, the use of the telestereoscope incidentally introduced a mismatch between vergence information and actual optical distance (which drives accommodation). The aftereffect can actually be shown to depend on the recalibration of the yoking between the vergence and accommodative systems: It can be produced both in the absence of stereopsis and in the absence of motion but is not produced by the presence of (tele)stereopsis and motion when artificial pupils are used to reduce blur-driven accommodation (Fisher & Ebenholtz, 1986). In other words, contrary to prior claims,[4] the aftereffect depended on the visual system using (distorted) distance information provided by accommodation in its derivation of depth from disparity.

Comparing Perceptions of Depth From Disparity and Depth From Motion

We have argued that the instantaneous geometric information provided by disparity, motion parallax, and object rotation is roughly equivalent: A map of the disparities

[4] Wallach, Moore, and Davidson (1963) had dismissed the notion that depth distortion was due to accommodative misperception of distance because they found no evidence of a concomitant size distortion such as would be produced by misperceived distance. In fact, the expected magnitude of the size distortion would be much smaller than the depth distortion and apparently went undetected.

produced by an object viewed binocularly is similar to a velocity map of that object either rotating or translating through the same angle (with fixation). Although the recovery of veridical depth is possible if more than two frames of motion are used, we believe that there are practical difficulties in extracting such information and that empirical evidence suggests that these difficulties are insurmountable for small rotations.

Proposals that motion information is superior to disparity information are further complicated by the finding that the two extreme frames of an SFM sequence are reported to produce the same subjective depth as a more continuous multiframe case and that affine structure, rather than Euclidean structure, is readily recovered from motion (Todd & Bressan, 1990). Because two-frame motion stimuli suffer the same theoretical inadequacies as the static stereo displays discussed earlier, a natural hypothesis is that information from disparity and motion might be processed by the same algorithms. Indeed, certain similarities between the systems have been noted (e.g., Graham & Rogers, 1982; Kontsevich, 1994; Rogers & Graham, 1982). However, one might predict that the binocular system would have some advantage in deriving object depth because interocular displacement is a fixed quantity within the system. Rotation magnitude is less easily determined when observing object rotations, and the information specifying such magnitude apparently is not used for small rotations (Caudek & Proffitt, 1993, 1994).

If the angle of object rotation is not used in recovering structure from motion, but distance information is used in recovering depth from disparity, then judgments of object depth based on disparity information ought to be superior to those based on recovery of structure from motion for small rotations. The primary purpose of the present studies was to assess this possibility by directly comparing the perceived depth of objects presented with either binocular disparity or monocular motion information. Both computer simulations of objects (Experiment 1) and real objects (Experiment 2) were used. Because of a recent demonstration of a failure of stereoscopic depth constancy (Johnston, 1991) with stereograms viewed in a dimly lit environment, a secondary goal, accomplished in Experiment 3, was to demonstrate stereoscopic depth constancy over a range of distances from 1 to 3 m with real objects in a well-structured viewing environment.

Experiment 1

In the first experiment, comparisons were made between depth perceived from monocular motion and depth perceived from geometrically equivalent binocular disparity information presented by computer simulation. The objects simulated were cones that varied in depth along their principal axes.[5] The cones were represented by five evenly spaced circular contours. In two conditions of the experiment, the depth of the cones was specified either by disparity information or by motion parallax information provided by cone translation.

A third condition was run in which stereokinetic cones were presented that were geometrically matched to those of the other two conditions and that appeared as rotating cones. Proffitt et al. (1992) have shown that the stereokinetic phenomenon may be defined as the subset of information available in a KDE display when contour foreshortening is absent. True cone rotation was not simulated because the appropriate foreshortening for a 2.5° rotation in each direction is less than 1% of the contour diameter, which was close to the resolution of the monitor for even the largest circular contour. This condition was run for comparison with motion parallax, in which the appropriate foreshortening is, in fact, accomplished, but by the change in angle of regard of the observer.

Method

Participants. The participants were 30 University of Virginia undergraduates who were able to identify simple forms in random dot stereograms. Fifteen men and 15 women were divided equally among three experimental conditions. Four additional students were excluded from the study either because they failed to perceive depth in a random dot stereogram or because of experimenter error.

Apparatus. The displays were produced on a Sun 3/60 workstation with a color monitor and were viewed through a stereo viewer built with first-surface mirrors. The left and right eyes received images from the left and right halves of the screen, respectively, but the images were optically rotated[6] such that the apparent screens were 22.9 cm × 14.6 cm (900 × 576 pixels) and were viewed at an optical distance of 72 cm. The mirrors were set so that the vergence angle of the eyes would match the optical distance to the screen for the typical observer. In the monocular conditions, stimuli were presented only to the right eye, and the left channel of the viewer was occluded.

Stimuli. Cones of nine different depths were simulated in each of three conditions: binocular disparity (BD), motion parallax (MP), and stereokinesis (SKE). The cones were represented by five white circular contours presented against a black background.

[5] One might be concerned that these frontal views are "degenerate," as described by Todd and Norman (1991), who argued that SFM information would fail to discriminate between shapes that differ only in affine stretching along the line of sight. However, this is precisely the point. If the cones were shown at oblique angles, then cones of different length would be readily discriminable from each other as affine objects along the line of sight; that is, they would project different families of metric structures. The purpose of the experiment was to test for accuracy of quantitative object depth along the line of sight rather than affine shape discrimination. Note that if the observer may assume (e.g., from the experimental context) that the three-dimensional form being simulated is symmetrical about some axis, then the affine structure derived from the motion of a three-quarter view, for example, could theoretically be constrained to a single metric interpretation because only one member of the projective family will be symmetrical in metric space.

[6] The optics of image rotation may be best illustrated with right angle prisms. Two such prisms, placed face to face at a 90° angle, are used in binoculars to rotate the image (inverted by the objective lens) by 180°. In fact, any amount of rotation may be produced by setting the relative orientation of the two prisms to half the desired rotation (e.g., a 90° image rotation is produced by a relative prism orientation of 45°). The same image rotation was accomplished with two appropriately oriented mirrors in the present apparatus.

The base (largest circle) of each cone was 6 cm (4.8°) in diameter, and the simulated tip-to-base depths were 2.4, 3.0, 3.6, 4.5, 6.0, 9.0, 12, 15, and 18 cm. In the BD condition, the simulated stationary cones pointed directly at the observer. An interpupillary distance of 6.2 cm was assumed,[7] and thus each eye received an image of a cone seen along lines of sight 2.46° from the axis of the cone. The two motion conditions were designed to match the geometric information provided in the disparity condition but were presented monocularly. In the MP condition, the simulated cones moved continuously from left to right and back. The total displacement of the base was 6.2 cm, or 2.46° in either direction, with a speed of 4.13 cm/s. Thus, at either extreme, the projected image was geometrically similar to that presented to one of the eyes in the disparity condition. Likewise, in the SKE condition, motion of the contours corresponded to a (KDE) cone rotation of 2.46° in each direction but without any foreshortening of contours. Because of the absence of foreshortening, SKE cones do not have a true simulated height. We did not attempt to simulate foreshortening information (necessary to specify true KDE) because of the resolution limits of the monitor.

A mouse-adjustable icon was simultaneously presented on the screen. The icon began as a vertical line and could be adjusted by the participant to form an isosceles triangle with the original line as the base. The icon represented a side view of the cone that the student could use to indicate the perceived depth (shape) of the object. The base of the triangle was equal to the diameter of the base of the cone, and the cone's five contour lines were represented as lines on the triangular icon for clarity. Note that correct adjustment of the icon is a test of the perceived depth to base ratio and does not require accurate two-dimensional size perception.

Procedure. Students were pseudorandomly assigned to one of the three conditions. Each student was required to identify a letter presented as a random dot stereogram through the stereo viewer. Students were then instructed that they would be viewing simulated cones through the viewing box and that they were to adjust the mouse-controlled triangular icon on the screen to indicate the apparent depth (shape) of the cones. For students in the monocular viewing conditions, the left channel of the viewer was occluded to prevent light from reaching the left eye.

Four blocks of nine pseudorandomly ordered stimulus trials were presented to the student. Each of the nine cone depths was presented once per block. The first block was considered practice, and the data were not analyzed. Students were not informed about the number of distinct stimuli or about the range of sizes, nor were the experimental blocks distinguished. On each trial, the student had unlimited time to view the stimulus and adjust the mouse-controlled triangular icon. The icon and the cone were both present on the screen until the student pressed a mouse button to indicate a satisfactory match between the icon and the cone. Depth judgments were recorded in pixels (40 pixels/cm) and converted to centimeters. Several students in the BD condition reported difficulty in fusing the tips of the deepest cones. This is consistent with the findings of Ogle (1952, 1953), who showed that there is a limited range of fusible disparities beyond which exists a range of "patent stereopsis" in which some depth is perceived despite a failure of binocular fusion.

Results

Analysis of means. The mean judged depths for each cone in each condition are plotted in Figure 3. It is evident that students in the BD condition were fairly accurate at judging the sizes of the cones, whereas the judgments of the other two groups did not reflect much sensitivity to the geometrically defined depth. Between-subjects variability is indicated by the error bars in Figure 3, which represent the standard error of the mean. A 3 × 9 × 3 (Condition × Simulated Depth × Block) mixed-design repeated measures analysis of variance (ANOVA) revealed reliable main effects of condition, $F(2, 27) = 6.06$, $p < .01$, and simulated depth, $F(8, 20) = 21.2$, $p < .001$. Block was not reliable, $F(2, 26) = 1.15$. Most important to our hypothesis, the interaction between condition and simulated depth was reliable, Wilks's $F(16, 40) = 3.99$, $p < .01$.

Individual regression slopes were computed for each student. Slope × Condition ANOVA revealed a reliable effect of condition, $F(2, 27) = 58.50$, $p < .001$. Planned comparisons revealed that the BD condition slopes ($M = 0.90$) differed reliably from the MP condition slopes ($M = 0.26$) and SKE condition slopes ($M = 0.17$) but that the SKE condition and MP condition slopes did not differ from each other (Ryan-Einot-Gabriel-Welsch Multiple Range Test; REGWQ, $p < .05$). Slopes for the SKE and MP conditions did differ from zero, indicating sensitivity to depth differences, $t(9) = 5.45$, $p < .001$, and $t(9) = 3.24$, $p < .001$. Although it is evident that depth judgments fell short for the deepest cones, the mean slope of the BD group did not differ reliably from 1, $t(9) = 1.76$, *ns*.

Within-subject variability. How consistent were each student's repeated judgments of each stimulus? For purposes of comparison with other experiments, this analysis was performed by first converting all judgments into equivalent retinal disparities (or absolute differences in retinal displacement, which is proportional to velocity if the participant maintains fixation on the object) expressed in minutes of arc. For each student in each condition, the standard deviation was computed from the three judgments of cone depth. This value was divided by the mean for that cone depth in that condition to produce an estimate of within-subject variability.[8] Because the distribution of these scores was skewed, median values (rather than means) are reported in Table 1. It is evident from Table 1 that variability was

[7] Many authors have assumed an interpupillary separation of 6.5 cm. According to the *Handbook of Human Factors* (Salvendy, 1987), the median interpupillary separation is about 6.27 cm (fixating on infinity). At nearer fixations, our value should therefore be closer to the median than is the traditional 6.5 cm, although the difference in perceived depth should be only on the order of 5%.

[8] Because we were working from only three data points for each participant in each cell, it was necessary to ascertain whether our variability estimates were biased. To accomplish this, we ran simulations in which 1,000 sets of three scores were sampled from a normal distribution with a mean of 100 and a standard deviation of 10, 20, 30, or 40. We calculated the sample standard deviation score divided by the mean for each of the 1,000 samples. The resulting scores were roughly normally distributed but underestimated the standard deviation of the population by a factor of about 0.9. Specifically, the means were 0.089, 0.18, 0.27, and 0.38. This should be taken into account when evaluating our reported errors (for samples with a cell size of 2, the simulation estimates were 0.081, 0.16, 0.25, and 0.35, requiring a correction by about 0.8).

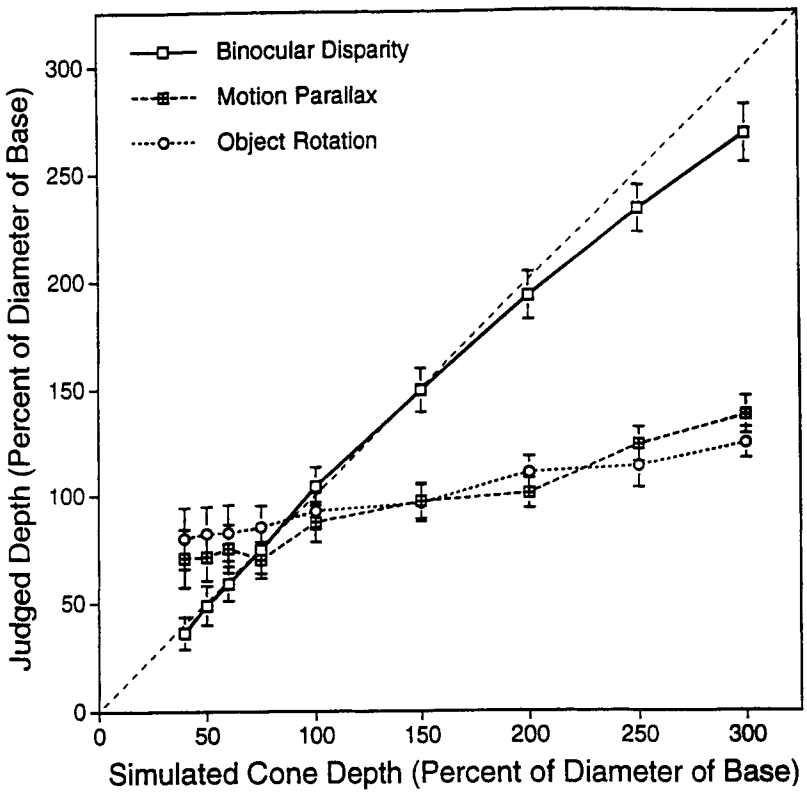


Figure 3. Results of Experiment 1. Mean perceived depth is plotted against simulated depth for each viewing condition. (Depth is expressed as a percentage of the base's diameter, which was 6 cm.) For the stereokinesis condition, simulated depth was defined assuming a rotation of 4.92° to match the other two conditions. The major diagonal dashed line represents accurate performance. Error bars represent standard errors of mean (i.e., between-observers variability).

higher in the MP condition than in the SKE condition. This is consistent with the fact that, although the relative motions presented in the two monocular motion conditions were approximately the same, they were added to a pedestal velocity of translation in the MP condition. Variability appeared to be lowest in the BD condition.

Discussion

The similarity of depth judgments from motion parallax and SKE displays found here is consistent with prior research (Caudek & Proffitt, 1993), in which it was determined that foreshortening information that distinguished motion parallax from SKE was not used for small rotations or displacements. Caudek and Proffitt suggested that perceived depth was mediated by a compactness assumption. This effect is illustrated in Figure 3 by the bias of the motion-based judgments toward a depth equal to the diameter of the base. Note that depth judgments equaling the diameter of the base are indicated by the 100% value in Figure 3 and that depth from motion judgments did not depart much from this default value. According to Caudek and Proffitt, the discrimination of depths may be mediated by a heuristic that greater relative motion signals greater

Table 1
Median Within-Subject Variability as a Function of Condition and Equivalent Retinal Disparity in Experiment 1

	Retinal disparity or displacement (minutes of arc)								
Condition	10	13	16	20	27	42	59	78	98
Binocular disparity	0.14	0.13	0.13	0.13	0.13	0.11	0.13	0.07	0.10
Stereokinesis	0.20	0.16	0.17	0.16	0.12	0.16	0.09	0.14	0.12
Motion parallax	0.23	0.35	0.23	0.22	0.22	0.27	0.17	0.23	0.16

Note. Values are based on standard deviations of individual participants' judgments divided by group cell means (3 judgments per cell per participant and 10 participants per cell).

depth. This heuristic would fit the results of the current situation. The important novel result is the direct evidence that geometrically equivalent depth information does not lead to the same quantitative perception when presented through motion parallax as when presented through binocular disparity.

Considered pragmatically, the disadvantage of motion parallax information is not surprising. The information specifying the angular rotation of the objects in the two monocular conditions was either left out entirely (as in SKE) or potentially ambiguous (in motion parallax). Although the cones in the MP condition were simulated as translating, they were, for practical purposes, indistinguishable from simulations of different-sized cones translating and rotating simultaneously. What should the visual system "conclude"? The recovery of form from relative motion information seems to be coupled with a default assumption regarding the objects' depths, such as the compactness assumption proposed by Caudek and Proffitt (1993). On the other hand, so long as distance is known to a reasonable precision, disparity information is quite determinate. The two frames that determine depth from disparity information are simultaneous and therefore cannot normally be confounded with object motion.

Experiment 2

In Experiment 2, we used real cones in a brightly lit, fully structured environment. There were two SFM conditions and a static binocular viewing condition. In the MP condition, the participant moved his or her head from side to side. Not only does this reduce the ambiguity of the object's actual lack of rotational motion, but optical flow from the surround may help to specify the quantity of head motion to greater precision. This might afford the MP condition an advantage over the second SFM condition, which consisted of a simple rotation of the object that was matched to the quantity of virtual rotation in the MP condition. Except for perspective projection, the object rotation condition was similar to a KDE stimulus. However, the ambiguities of rotational angle were essentially the same as in the SKE condition of the first experiment, because the foreshortening information provided by the very slight rotation used was negligible. As in Experiment 1, a static BD condition was used to assess the accuracy of depth from stereopsis.

Method

Participants. The participants were 36 University of Virginia undergraduates who were able to identify simple forms in random dot stereograms. Eighteen men and 18 women were divided equally among three experimental conditions. Two additional students were excluded from the study because they failed to perceive depth in a random dot stereogram.

Apparatus. Four wooden cones were used. Each was painted flat black. The tip and four equally spaced thin, circular contours were painted white. All cones were 10 cm in diameter at the base. The depths of the cones were 5, 10, 15, and 20 cm. The cones screwed into a vertical metal rod hidden behind a black cloth. The black cloth extended 40 cm to the left and right of the cone center. The rest of the room was left unoccluded so as to provide a structured viewing environment. The cones were illuminated by a floodlight behind and slightly above the student's head and a fill light placed below and in front of the students to minimize shading information on the cones.

The experimenter was able to manually rotate the cones using a known angular displacement. Participants' heads were restrained by a sliding chin rest that enabled side-to-side motion for the MP group, but that was fixed stationary for the other two groups. The chin rest had cheek rests on the side to prevent head rotation. Depth judgments were recorded with a Macintosh SE computer by means of a mouse-controlled icon similar to that used in Experiment 1.

Stimulus geometry. As a means of minimizing the influence of accommodative depth cues that might have been used to discriminate differences in the absolute distance of the tip and base, the cones were viewed from a distance of 2 m (to the base). This is nearly three times the distance used in Experiment 1. Pilot studies revealed that limiting viewer displacement in the MP condition to interocular distance produced little sensation of depth at this distance. As a more generous test of motion parallax, we had students move through a total of 25 cm (4 times the interocular distance). Thus, the observer moved 3.6° to the right and left of the base of the cone in the MP condition. In the object rotation condition, the cones were rotated 3.6° to the left and to the right (as in the MP condition). The angle of convergence was approximately 0.9° for each eye for the BD group.

Procedure. As with Experiment 1, a mixed design was used. Students were randomly assigned to one of three conditions: BD, cone rotation (CR), or viewer-movement-induced MP. Within each condition, students made two judgments of each of the four stimulus cones. The order of presentation was broken into two blocks of four randomly ordered trials.

Students were told that they would be viewing real wooden cones under special viewing conditions and were instructed in the use of the mouse-adjustable icon used to register perceived depth of the cones. They were advised to pay attention to the shape rather than to the absolute size of the cone when adjusting the computer icon. They were not informed of the number of distinct stimuli or of the range of sizes.

For the monocular conditions (CR and MP), students wore an eye patch on whichever eye they chose. The chin rest was adjusted so that the students' exposed eye was in line with the principal axis of the cone when the cone (CR condition) or the eye (MP condition) was at the center of their excursions. Students' heads remained in the chin rest throughout the procedure. The adjustable icon appeared on a screen in front of the student, below the line of vision to the cone, and the students could freely look back and forth between the cone and the icon. In the MP condition, students were encouraged to slide their heads back and forth quite rapidly to enhance the perception of three-dimensional structure. Because students would stop sliding to adjust the icon, they were encouraged to alternate between sliding their heads and adjusting the icon.

A curtain was lowered between the student and the cone apparatus between trials while the experimenter changed cones. Each student made eight judgments and was then assessed for stereo depth perception with a random dot stereogram viewed through a hand-held stereoscope. Some students in the MP condition reported experiencing apparent object rotation.

Results and Discussion

The mean judged heights for each cone in each condition are shown in Figure 4. As in Experiment 1, students in the BD condition were quite accurate at judging the depth of the cones. The judgments of the other two groups reflected much less sensitivity to the geometrically defined height of the stimulus information. A 3 × 4 (Condition × Depth) mixed-design repeated measures ANOVA revealed a highly reliable main effect of depth, $F(3, 31) = 64.60$, $p < .001$, and, more important, a highly reliable interaction between condition and depth, Wilks's $F(6, 62) = 6.40$, $p < .001$. There was no main effect of condition, which indicates, for example, that there is no overall bias to see greater depth when both eyes are open.

As in Experiment 1, individual regression slopes were computed for each student. A Slope × Condition ANOVA revealed a highly reliable difference between conditions, $F(2, 33) = 17.90$, $p < .01$. Planned comparisons between groups revealed that slopes of the BD group ($M = 1.07$), the MP group ($M = 0.60$), and the CR group ($M = 0.31$) all differed reliably from each other (REGWQ, $p < .05$). All slopes differed reliably from zero, and only the slope of the BD group did not differ reliably from 1, $t(11) = 0.73$, *ns*.

Median within-subject variability scores are shown in Table 2 as a function of equivalent retinal disparity. The scores were calculated as in Experiment 1. Note that because these estimates were based on only two judgments per cell per student, they underestimated true variability by a factor of about 0.8 (see Footnote 8). The retinal disparities viewed in the binocular condition were less than those of Experiment 1, and there was a notable increase in the variability for depth judgments of the smallest disparities (shallowest cone). Variability in the MP condition (with equivalent disparities or relative displacements 4 times larger) was comparable to that of the BD condition ($M = 0.12$ in each case).

There were two main results: On one hand, the superiority of binocular disparity was replicated, but on the other, a clear superiority was demonstrated for the MP condition over the CR condition. Thus, in a well-illuminated, structured environment, it appears that the active self-motion of an observer produces depth judgments that are superior in accuracy to that of equivalent object rotation but remain inferior to static binocular viewing. The relatively good performance of observers in the monocular conditions may have been due in part to monocular depth cues such as

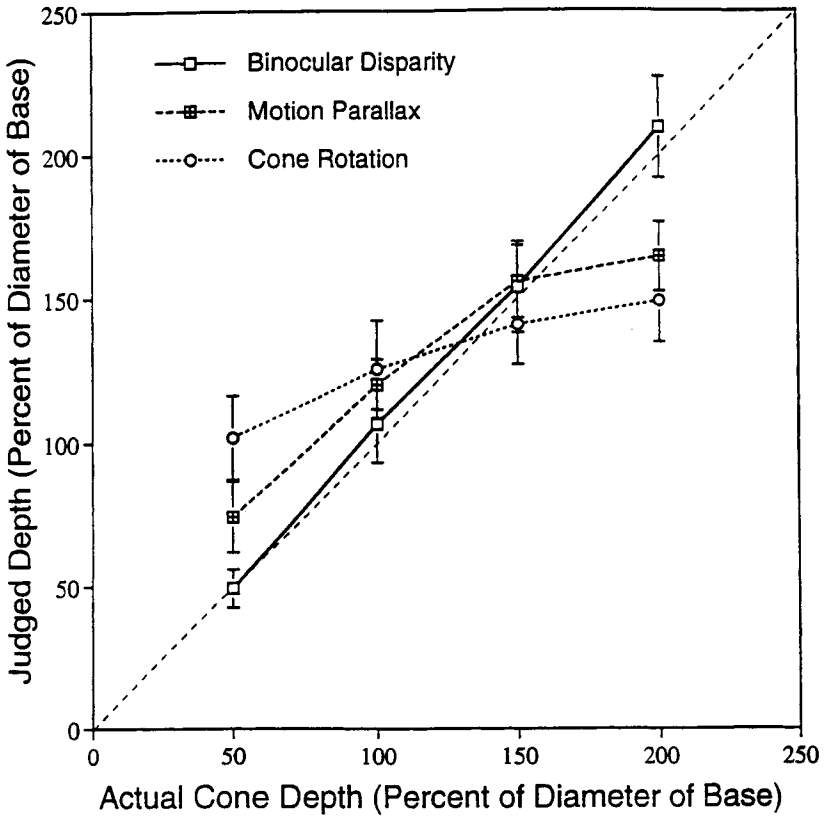


Figure 4. Results of Experiment 2. Mean perceived depth is plotted against actual cone depth for each viewing condition. (Depth is expressed as a percentage of the base's diameter, which was 10 cm.) The major diagonal dashed line represents accurate performance. Error bars represent standard errors of the mean.

Table 2
Median Within-Subject Variability as a Function of Condition and Equivalent Retinal Disparity in Experiment 2

	Retinal disparity or displacement (minutes of arc)							
Condition	2.73	5.61	8.64	11.8	11.0	22.5	34.7	47.5
Binocular disparity	0.18	0.13	0.10	0.06				
Cone rotation					0.12	0.16	0.15	0.13
Motion parallax					0.16	0.12	0.10	0.11

Note. Values are based on standard deviations of individual participants' judgments divided by group cell means (2 judgments per cell per participant and 12 participants per cell).

accommodation, differential luminance reflection, and other information not available in the computer simulations. Using the same cones but different illumination conditions, Caudek and Proffitt (1994) found that observers could discriminate the relative depths of these cones in a static monocular viewing condition. Also, the angular displacements and object rotations were somewhat greater than in Experiment 1. It is probable that improved information about the angle of object displacement was available in the MP condition. However, it is also possible that the improved performance of the MP group was due, at least in part, to a reduction of the ambiguity about object rotation: Some students reported that some objects seemed to rotate, so they assumed these objects were taller than they "looked." We found no evidence to support the notion that depth from motion is equal to or superior to depth from disparity. On the contrary, perceived depth in the binocular disparity situation seemed to be excellent in the present viewing conditions, whereas depth from motion appeared to tend toward the default depth assignment of the base's diameter identified by Caudek and Proffitt (1993).

Is There Stereoscopic Depth Constancy?

Although the present findings support our claim that binocular stereopsis can support more accurate depth recovery than can SFM with geometrically similar displays, there are reports of the failure of depth constancy in the literature that would seem to contradict our conclusions regarding the accuracy of depth from stereopsis generally (e.g., Collett, Schwarz, & Sobel, 1991; Tyler, 1983). Perhaps the clearest case of the failure of depth constancy is that demonstrated by Johnston (1991). She presented observers with random dot stereograms that simulated elliptical cylinders with their principal axis horizontal in the picture plane. The participants' task was to judge whether a presented cylinder appeared "squashed" or "stretched" in depth (relative to an imagined circular cylinder). Such data can be interpreted as a measure of a perceived depth to base ratio. The task was performed at three viewing distances with five cylinder heights (vertical diameters). Johnston's results were straightforward: Perceived depth was exaggerated for the nearest viewing distance and grossly underestimated for the farthest. Such results are consistent with an inadequate use of distance information to scale disparity information. Indeed, Johnston determined that her results were entirely consistent with the interpretation that participants used a default presumed viewing distance of about 80 cm with a change in (perceived) distance corresponding to 26%[9] of the actual change in distance from that default.

Johnston's (1991) methodology and findings are of particular interest because she was able to show that participants did not seem to use purely visual (retinal) sources of information to perceive the shape of her smoothly curved surfaces. Rogers and Cagnello (1989) and Koenderink and Van Doorn (1976) had both suggested analyses of visually available information that could produce accurate depth from stereopsis for smoothly curved surfaces without requiring knowledge about viewing distance. Apparently, observers did not use these kinds of theoretically available, purely visual sources of information.

Johnston's (1991) findings are consistent with the claim that stereoscopic depth judgments are scaled by information about fixation depth but suggest that information about egocentric viewing distance is quite poor, at least under the experimental conditions used. Foley (1980) reviewed evidence that the perception of egocentric distance could be derived from vergence information and concluded that although vergence information is probably used, egocentric distance is not typically recovered accurately. In our Experiment 1, the fixation distance (72 cm) was coincidentally quite close to the distance that appears to be the default in Johnston's data. It therefore should come as no surprise that we found such accurate depth judgments. However, our fixation distance in Experiment 2 (2 m) was similar to the longest fixation distance used by Johnston, and yet we still found apparently veridical stereoscopic perception of depth, whereas she found a marked underestimation of depth (failure of depth constancy). Why should this difference arise?

Our stimuli in Experiment 2 differed from those of Johnston (1991) in several ways. We used real cones specified by clearly demarcated contours and presented in a well-lit, structured environment without a fixation target, whereas she used random dot stereograms of cylinders presented in a darkened environment with a fixation target. Assuming our result at 2 m is not an artifact, it is reasonable

[9] Johnston's figure is probably artificially low because she used images with matched retinal sizes and textures. Stimulus size itself (among similar stimuli) is a cue to distance that interacts with oculomotor cues (Collett, Schwarz, & Sobel, 1991). Foley (1980, 1985) suggested a gain of 0.5.

to suppose that the crucial differences between the situations involve the perception of distance rather than the recovery of disparity information itself, because inadequate recovery of distance information, not disparity information, seems to underlie the failure of stereoscopic depth perception in Johnston's experiments. Thus, we suspect that the source of the difference between Johnston's results and our own was the experimental environment.

There is a great deal of evidence suggesting that distance perception is faulty in low illumination or in the absence of a surrounding context (e.g., Owens & Leibowitz, 1980). Gogel (1969b) and Gogel and Tietz (1973) have shown that, in dim light, most observers tend to underestimate the distance of far objects and overestimate the distance of near objects, just as Johnston's (1991) results suggest. This phenomenon is referred to as the "specific distance tendency," wherein objects in reduced viewing conditions tend to be perceived as being at a default distance (usually about 2 m, according to Gogel, 1982; see also Foley, 1980). Owens and Leibowitz (1976) have shown that oculomotor adjustments to a binocularly viewed target in a dark environment show systematic errors that are broadly consistent with a specific distance tendency. Owens and Leibowitz (1980) measured *dark vergence* and *dark focus* of 60 participants and found mean vergence and accommodative states appropriate to egocentric distances of 1.16 m and 0.76 m, respectively. Note that these distances are quite close to the default found in Johnston's work. It is therefore possible that the failure of depth constancy found in Johnston's experiment, and perhaps in portions of the distance perception literature in general, is due to the low level of ambient illumination that is typically used to provide reduced viewing conditions. Distance perception may critically depend on viewing conditions that are more normal. Thus, although Johnston's results represent a crucial refutation of certain more elaborate models of depth from disparity (i.e., Koenderink & Van Doorn, 1976; Rogers & Cagnello, 1989; but see Rogers & Bradshaw, 1993), they are, as she herself argued, entirely consistent with the thesis that depth from disparity is scaled by information about viewing distance.

Experiment 3

To demonstrate that stereoscopic depth constancy can indeed exist in a nonreduced viewing situation such as that used in Experiment 2, one must systematically vary viewing distance. Experiment 3 was a partial replication of the BD condition of Experiment 2 with a systematic variation of viewing distance from 1 to 3 m by 0.5-m increments. Static monocular control conditions were also tested at 1 and 2 m. A failure of depth constancy would be indicated both by changes in the slope between judged and actual depth and by absolute depth values at different distances. Because the proximal disparity produced by the same distal object decreases with greater viewing distance, if distance information is not taken adequately into account, judgments of depth from disparity should decrease as viewing distance increases.

Method

Participants. The participants were 56 University of Virginia undergraduates divided evenly among seven experimental conditions. Three additional students were excluded from the study because they failed to follow instructions or failed to perceive depth from disparity.

Design. A mixed design was again used. Students in binocular conditions were assigned to one of five viewing distances (1.0, 1.5, 2.0, 2.5, or 3.0 m, measured to the base of the cones). There were also two static monocular conditions in which students viewed cones from a distance of either 1.0 or 2.0 m. Each student made single depth judgments (side-view triangles) of each of four cones presented in random order in each of three blocks. From the student's point of view, there were simply 12 trials without break. The first block was considered practice and was not analyzed.

Apparatus. The room used for testing was not the same as that used in Experiment 2, but the apparatus was otherwise quite similar. The same four wooden cones were used, and the same black cloth extended 40 cm to the left and right of the cone center. As before, the rest of the room was left unoccluded so as to provide a structured viewing environment, and the cones were illuminated by a floodlight behind and slightly above the students' heads and a fill light placed below and in front of the students to minimize shading information on the cones. Students' heads were positioned in a chin rest. Depth judgments were recorded with a Macintosh SE computer by means of a triangular mouse-controlled icon similar to those used in the previous experiments.

Procedure. Students were told that they would be viewing real wooden cones under special viewing conditions and were instructed in the use of the mouse-adjustable icon used to register the depth of the cones. As in previous experiments, no information was provided to the students about the number of distinct stimuli or about the range of sizes. They were told to represent how each cone would look from the side but were advised to pay attention to the shape rather than to the absolute size of the cone when adjusting the computer icon. The adjustable icon appeared on a screen in front of the student, below the line of vision to the cone, and the students could freely look back and forth between the cone and the icon. Both the cones and the observers were stationary in all conditions. Students in the monocular condition wore an eye patch over one eye. Each student made 12 judgments, of which the first 4 were considered practice. A barrier was raised between the student and the cone apparatus between trials while the experimenter changed cones.

Results

The mean judged heights for each cone in each of the seven viewing conditions (five binocular and two monocular) are shown in Figure 5. Individual regression slopes were computed for each student. Slope × Condition ANOVA revealed a highly reliable main effect of viewing condition, $F(6, 49) = 31.20, p < .0001$. Planned comparisons revealed no differences within the binocular or monocular conditions but indicated that the mean slopes in the monocular conditions ($M = 0.02$) were significantly lower than those in the binocular conditions ($M = 1.33$; REGWQ, $p < .05$). The mean slope for the binocular conditions was reliably greater than 1, $t(39) = 5.35, p < .0001$. The mean slope for the monocular conditions did not differ reliably from 0, $t(15) = 0.36$, *ns*. The monocular and binocular conditions were considered separately in further analyses.

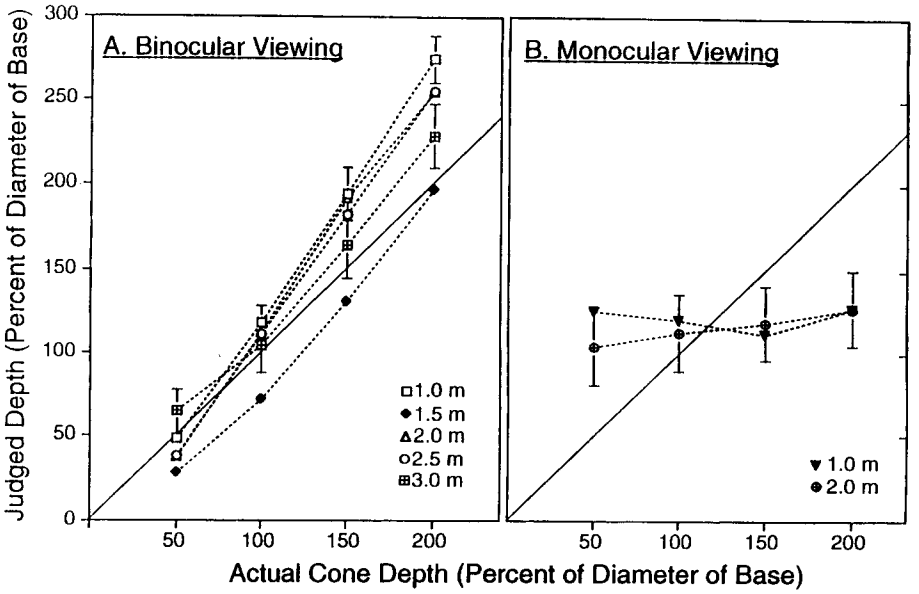


Figure 5. Results of Experiment 3. Mean perceived depth of real cones viewed either (A) binocularly or (B) monocularly as a function of actual depth and viewing distance. (Depth is expressed as a percentage of the base's diameter, which was 10 cm.) In (A), standard errors are shown for the 1 m and 3 m conditions. In (B), standard errors are shown for the 2 m condition.

Monocular conditions. In our static monocular conditions, many students commented that they saw no definite depth or even depth order. The judgments they made should therefore be understood as "best bet" guesses. A 2 (viewing distance) × 4 (cone depth) × 2 (block) repeated measures ANOVA of the monocular conditions revealed a reliable Block × Viewing Distance interaction, as well as reliable main effects of block, $F(1, 14) = 8.94$, $p < .01$, and cone depth, $F(3, 12) = 4.59$, $p < .05$. Because of the interaction of viewing distance and block, 4 × 2 (Cone Depth × Block) repeated measures ANOVAs were performed on data from each of the two conditions individually. For the 1 m monocular condition, there was no effect of block, $F(1, 7) = 0.13$, *ns,* or of cone depth, $F(3, 5) = 1.19$, *ns.* However, for the 2 m monocular condition, the depth judgments in the last block ($M = 13.1$ cm) were significantly larger than those in the first analyzed block ($M = 9.4$ cm), $F(1. 7) = 19.10$, $p < .01$. There was also a marginal effect of cone depth, $F(3, 5) = 5.00$, $p = .058$, suggesting some sensitivity to different depths. The mean judged depths for the four cones in the monocular 2 m condition (smallest to deepest actual cone) were 10.5, 11.3, 11.8, and 12.7 cm (recall that the cones were actually 5, 10, 15, and 20 cm in depth and had a base diameter of 10 cm). The depth judgments in these monocular conditions suggest that when there was very little monocular information available to specify object depth in our viewing situation, students used a judgmental default consistent with a compactness assumption.

Binocular conditions. A 5 (viewing distance) × 4 (cone depth) × 2 (block) repeated measures ANOVA of the binocular conditions revealed a highly reliable main effect of cone depth, $F(3, 33) = 181.00$, $p < .0001$. Apparently because of the lower depth judgments in the 1.5 m binocular condition, there was also a marginal main effect of viewing distance in the binocular conditions, $F(4, 35) = 2.38$, $p = .07$. The mean judged depth in the 1.5 m condition was 10.7 cm, whereas the means in the 1.0, 2.0, 2.5, and 3.0 m conditions were 15.9, 14.7, 14.9, and 14.0 cm, respectively.[10]

The within-subject variability between the two adjustments made for each cone was calculated as in previous experiments. Because perceived distance may be expected to remain constant for each observer, within-subject variability is expected to be modulated primarily by the quantity of disparity information available, which covaries with depth and viewing distance. It can be seen in Figure 6 that the variability was greatest for displays in which the maximum disparity present was less than about 2 minutes of arc.

Discussion

Using real objects in a well-illuminated, structured viewing environment, we tested whether adequate distance information would be provided to support stereoscopic depth constancy. Indeed, the results are indicative of such constancy, because they do not reflect any systematic distor-

[10] To determine whether these marginal differences would replicate, we conducted a follow-up experiment in which the 1.5 and 2.0 m conditions were repeated with 8 students each. The marginal differences did not replicate. The mean depth for new students in the 1.5 m condition was 13.2 cm, which did not differ reliably from the 12.0-cm mean of the new students in the 2.0 m condition, $t(15) = 1.18$, *ns.* In addition, the mean slope between judged and actual depth at 1.5 m (1.30) did not differ reliably from that at 2 m (1.14), $t(15) = 0.92$, *ns.*

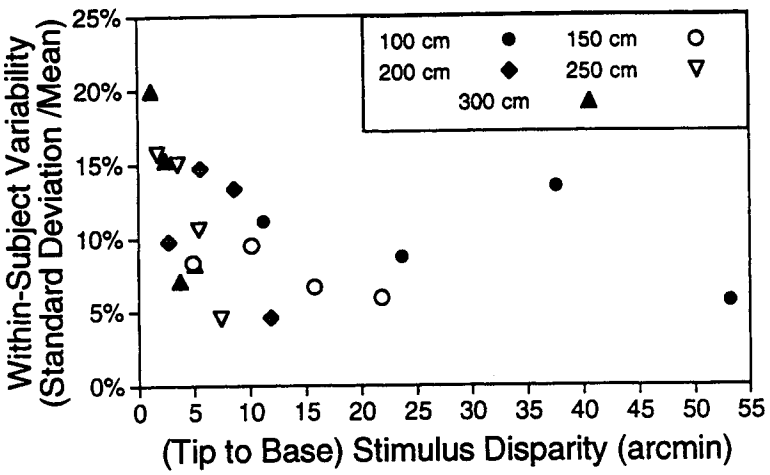


Figure 6. Median within-subject error in Experiment 3 as a function of retinal disparity. Within-subject error was computed by taking the root-mean-square error for each participant's judgments of each cone and dividing by the group mean. Each plot point represents the median of eight such ratios (corrected, here, by a factor of 0.8; see Footnote 8). Because error was computed separately for each participant, between-subjects variability in perceived viewing distance is not represented. As expected, variability was greatest for extremely small disparities. Note that the abscissa represents the maximum retinal disparity in each stimulus. Four points are shown for each viewing distance, corresponding to the four cones used in the experiment. arcmin = minutes of arc.

tions of depth with changes in viewing distance from 1 to 3 m such as those found by Johnston (1991) in a reduced viewing environment. Thus, the claim from Experiment 2 that depth from disparity information can be scaled by viewing distance has received important confirmation.

The study of depth constancy is not new, but the basis for its sometime failure is poorly understood. Johnston (1991) argued that her experiments indicated that the cues of accommodation and convergence were inadequate to support depth constancy. However, as reviewed earlier, there is some dispute in the literature regarding oculomotor adjustments and the perception of distance in dim light (e.g., Foley, 1980; Owens & Leibowitz, 1980). Johnston took care to establish that the vergence angle through the stereo viewer was correct for the distance, and she used a cross-hair fixation device to ensure that participants were accommodating at the right distance. Given the elegance and care with which her experiment was carried out, it may seem unreasonable to doubt the general validity of its conclusion. However, the techniques used by Johnston to establish the failure of stereoscopic depth constancy and, presumably, of distance perception entailed conditions (dim illumination) in which vergence and accommodation typically fail (Leibowitz, Hennessy, & Owens, 1975; Owens & Leibowitz, 1980). In such conditions, vergence and accommodation, even when correct, may receive little weight in the evaluation of distance. It should not be concluded from her experiment that there is no depth constancy, or even that oculomotor cues are entirely ineffective, because she found evidence of some distance compensation. Indeed, the effects of oculomotor cues on perceived stereoscopic depth have also been documented in a number of studies conducted by Wallach and colleagues (O'Leary & Wallach, 1980; Wallach & Zuckerman, 1963) and other investigators (e.g., Collett et al., 1991; Cumming, Johnston, & Parker, 1991), although there are no reports of full stereoscopic depth constancy from oculomotor cues alone.

In general, perfect stereoscopic depth constancy depends critically on the veridical perception of egocentric distance. Using disparity afterimages, Cormack (1984) has demonstrated that stereoscopic depth constancy holds quite well for far viewing distances (e.g., 2,000 m) if one takes into account the predictable biases and failures of distance perception outlined by Gilinsky (1951). O'Leary and Wallach (1980) have shown that linear perspective cues, even when put in conflict with otherwise effect (but deceptive) oculomotor cues to distance, can affect perceived stereoscopic depth in a manner consistent with distance scaling. Moreover, Collett et al. (1991) have demonstrated that relative size effects, which affect distance perception, also affect disparity scaling. In short, sources of visual and oculomotor information that affect perceived distance seem also to influence the scaling of binocular disparity information.[11]

[11] O'Leary and Wallach (1980) also reported that the deceptive manipulation of the distance cue of familiar size (a shrunken dollar bill) affected perceived stereoscopic depth in a manner predicted by constancy. However, their own data and those of other investigators (Predebon, 1993) indicate that only linear (not quadratic) scaling occurs, as if perceived depth is scaled proportionally to judged size, and indeed that the effect of familiar size on distance perception (as compared with distance judgment) is fairly inconsequential (Gogel, 1969a; Gogel & Da Silva, 1987; Predebon, 1992).

It appears that participants did use distance information to scale depth from disparity in the present experiment. However, there are two important features of the data that suggest that depth from stereopsis was imperfect: (a) The data showed evidence of an overestimation of depth for the deeper cones, and (b) there was evident variability in the depth judgments between observers. One probable source of error arguably external to the perception of depth might involve individual differences or difficulties, or both, in transforming three-dimensional perceptions into two-dimensional side-view responses. One could easily have a metric scale of depth along the line of sight without having a very good means of mentally transforming distance along one axis to distance along another. An even more pernicious source of variability and bias, however, may have arisen from participants' interpretations of the experimental instructions (Brunswick, 1956; Carlson, 1977; Leibowitz & Harvey, 1969). Carlson (1977) reviewed evidence that instructions to report on the objective size of an object in a size-constancy experiment, for example, will tend to produce overconstancy (a kind of overcompensation for distance), whereas instructions to report apparent size may result in a closer approximation to constancy. Because viewing distance was varied between participants in the present experiment, one would expect to find no "overconstancy" with respect to viewing distance, and none was found. However, the fact that the mean slopes in the binocular conditions were uniformly greater than 1 is consistent with a view that participants may have been similarly compensating for what they perceived to be a reduced viewing condition (static viewing). In other words, the participants may have exaggerated differences in perceived depth in their responses.

A Geometry of Visual Space?

Studies of distance perception clearly implicate a compressive scale of distance over large distances (e.g., Gilinsky, 1951; Luneburg, 1947), although the apparent compression in near space (less than 2 or 3 m) may be relatively small in nonreduced conditions. In an outdoor full-cue viewing environment, Loomis, Da Silva, Fujita, and Fukusima (1992) recently replicated the basic finding that perceived spatial intervals in depth are increasingly compressed relative to perceived frontal intervals as viewing distance increases from 4 to 12 m; however, they also demonstrated that performance on a blind walking task (after visual inspection of the goal) is linearly related to distance (and accurate) for these same distances. Wagner (1985; see also Toye, 1986) examined the geometry of visual space in a somewhat larger outdoor environment and concluded that the amount of compression of perceived space (i.e., of depth relative to frontal distances) was constant over the distances he sampled (5–70 m) and equal to about 0.5 (depth intervals are perceived as half as large as equivalent frontal intervals viewed at the same distance). Nonetheless, in comparing his findings with those of other investigators, Wagner concluded that the "geometry of visual space itself appears to be a function of stimulus conditions" and that "the visual world approaches the Euclidean ideal of veridical perception as the quantity and quality of perceptual information increases" (1985, p. 493; see also Baird, 1970; Suppes, 1977).

The scaling of binocular disparity information requires accurate egocentric distance but does not otherwise entail any particular "geometry of visual space" as assessed by depth interval studies. Because the physical world is (for most behaviorally relevant purposes) Euclidean, one would hope that the visual system would not reject Euclidean interpretations if it could easily get at them; however, there is ample evidence that people's perceptions are frequently biased by viewpoint-dependent information. In reduced environments, there is a well-demonstrated failure of accurate distance perception, even for near objects, and a concomitant failure of disparity scaling (Foley, 1980). We have suggested that the perception of near distances will approach veridicality in well-illuminated, fully structured environments (see also Johansson, 1973). Note that, even in good illumination, there are reports of failures in the perception of near depth intervals. For example, Baird and Biersdorf (1967; see also Todd & Norman, 1994) found that the length of a 20.3-cm × 1.25-cm rectangular strip of posterboard viewed at a distance of 2 m in a well-lit visual alley was compressed by a factor of about 0.85 relative to its appearance at a distance of 60 cm.[12]

In contrast even to this fairly slight compression observed by Baird and Biersdorf (1967), our results demonstrate no failure of depth constancy whatsoever. We can only speculate about the basis for this difference, but there are several differences between our stimuli and procedure and theirs.

[12] Baird and Biersdorf (1967) used a well-illuminated visual alley with a black floor and green walls in which they presented strips of white posterboard that were 1.25 cm wide and varied in length. The strips were viewed binocularly from one end of the alley, and size matches were made between standard strips (20.3 cm in length) presented at several distances (from 61 cm to 549 cm) and comparison strips presented at the extreme distances. The authors were the observers, and they tried to make "objective" matches for pairs of strips presented frontally or lying flat on the floor of the visual alley (46 cm below eye level). Consider the comparison of the near strip (presented at 61 cm) with an intermediate strip (presented upright at 204 cm and flat at 206 cm). For the frontally viewed strips, Baird and Biersdorf found slight overconstancy (consistent with objective instructions): A strip at 61 cm distance needed to be 20.9 cm, on average, to appear equal to one of 20.3 cm presented at 204 cm. On the other hand, underconstancy was observed for the strips laid flat: A flat strip at 61 cm needed to be only 18.0 cm to appear equal in length to one of 20.3 cm viewed at 206 cm (compression by 0.89). Moreover, even at 309 cm, an object of 20.3 cm was matched to one of 18.3 cm viewed at 61 cm (compression by 0.90). If the overconstancy factor of frontal size is taken into account, these results can be interpreted as evidence of roughly 0.85 compression of a depth interval along the surface of the alley with a roughly threefold increase in absolute viewing distance. Although depth and frontal views are not truly segregated because of the elevation of the eye above the surface, this amount of compression is much less than that observed at greater viewing distances (e.g., Wagner, 1985).

First, our participants made shape judgments about volumetric solids, whereas Baird and Biersdorf compared lengths of rectangular strips. Second, our surrounding experimental setting was more elaborately structured than their visual alley and may have provided better peripheral information about viewing distance (e.g., differential perspective; Rogers & Bradshaw, 1993; Tyler, 1991). Third, because the relevant depth axes of our stimuli were all aligned directly with the participant's cyclopean line of sight, conflicting size cues (i.e., from changes in angle of regard) were not confounded with viewing distance. Finally, our participants were not individually asked to compare sizes at different viewing distances (distance was manipulated between subjects).

In summary, our concern has not been with assessing the geometry of visual space (if such a thing exists) but rather with examining how well human observers can recover the metric shape of objects viewed binocularly in near space. Thus, participants were required not to explicitly compare spatial intervals but rather to report on three-dimensional shapes. Moreover, we used real solid objects that were textured with widely spaced contours in a manner that reduced any binocular correspondence difficulties, and we had participants view these objects in a well-lit, structured environment. The methods we used differed from those of traditional constancy experiments in that each participant viewed the objects at only a single distance and different distances were compared between participants. This strategy was used to avoid problems of recognizing the objects but may have had the advantage of reducing attention to distance itself. We found that performance of our task was quite good and have suggested that this was probably a result of access to accurate distance information from oculomotor cues as well as optical distance cues such as linear perspective and, possibly, differential perspective.

General Discussion

In the experiments presented here, judgments of object depth that were based on binocular disparity information were found to be substantially more accurate than judgments based on monocular motion information. This was true both for contour-specified computer simulations and for real objects viewed in a structured environment. The superiority of binocular disparity was maintained when the geometric information available from motion was equivalent or theoretically superior to that available from disparity. These findings are consistent with other evidence that information about angle of rotation in SFM displays is not recovered for small rotations (Caudek & Proffitt, 1993, 1994; Loomis & Beall, 1993). In Experiment 2, an advantage was also found for viewer-induced motion parallax over object rotation in a situation in which the local visual information was essentially equivalent.

Our findings are interpretable within a view of perception as a heuristic process (Braunstein, 1976). We would argue that the derivation of a single definite percept from the plethora of information sources available should proceed according to pragmatic considerations. For example, suppose that the accurate perception of depth from disparity is possible because the visual system can be calibrated to reliably determine the distance to the point of zero disparity from vergence, accommodation, and other cues in normal viewing environments. There is evidence that oculomotor sources of information are fairly accurate within a range of 25–200 cm (Leibowitz, Shina, & Hennessy, 1972), although they may fail in dim illumination. Obviously, distance determination from oculomotor cues will deteriorate with increasing distance; within the range of space normally relevant to object manipulation, however, binocular disparity should be quite reliable. On the other hand, the crucial parameters required to make SFM accurately determine object depth are often difficult to obtain and are likely to be less precise in any case.

A number of models suggest that SFM might generate a set of structures all related by a single parameter (essentially an affine structure; Aloimonos & Brown, 1989; Bennett, Hoffman, Nicola, & Prakash, 1989; Huang & Lee, 1989; Koenderink & Van Doorn, 1981). Derivation of a definite object depth from such models requires either the overt specification of the free parameter or some other constraint on the visual interpretation. From our point of view, however, the apparent success of affine derivations of structure (e.g., Todd & Bressan, 1990) may simply result from the fact that it is typically impossible to precisely specify crucial parameters of the Euclidean equations (see Appendix) from within the local visual flow field; the relative velocities or displacements necessary to specify such parameters are too difficult to discriminate. For example, Eagle and Blake (1994) found that when thresholds for relevant visual information are matched, performance is as good for Euclidean tasks as for affine ones. In cases of reduced angular rotation or translation, the perceptual system should accept only an affine (or weaker) interpretation of motion information because the visual system cannot reliably extract more. The findings of Todd and his colleagues (Norman & Todd, 1993; Todd & Bressan, 1990; Todd & Norman, 1991) are therefore consistent with our argument.

Norman and Todd (1993) have shown that instantaneous stretching of an SFM-specified object along the line of sight is essentially undetectable, which is entirely consistent with the view that the angle of rotation is not easily recovered and that depth is often assigned by default assumptions. Pollick (1994) provided evidence suggesting that these "unperceived" mutations of form may often instead be perceived as changes in angular velocity of rotation, a Euclidean interpretation approximately consistent with the two-dimensional projections involved. Stretching in other directions can also be undetectable, however, suggesting that there is a fair amount of tolerance in the system. For example, SKE stimuli simulate a different kind of stretching that is nonetheless perceived as rigid (Proffitt et al., 1992). On the other hand, Loomis and Eby (1988) have shown that simulations of rigid elongated objects appear to stretch and contract (with their two-dimensional projection) as they rotate, and, more recently, Caudek and Proffitt (1994) have shown that a simulation of a rotating wire form appears

most rigid if it actually stretches and contracts along an axis rotating with the object so that its projected two-dimensional size remains fairly constant. All of these findings suggest that there are limits to the amount of depth the visual system is willing to recover from motion and that (perceived) violations of rigidity may by primarily determined by marked changes in the two-dimensional envelope in which the stimulus is projected.

This last finding points toward an explanation of the specific depth assignments found in SFM research. The relative depth assignments produced in the current experiments reflect an appropriate discrimination of differences: Displays simulating deeper objects are judged to simulate deeper objects, and displays of shallower objects are judged to be shallower. However, such discriminations appear to be driven by differences in the two-dimensional displacement of near object features relative to the base because they also occurred with illusory depth in the case of SKE (see also Caudek & Proffitt, 1993, 1994). Included in these depth judgments, however, is a central tendency that is not explained by models of geometric SFM. The depth of the shallowest objects is overestimated on the basis of motion information, whereas that of the deepest ones is underestimated. Cuadek and Proffitt have attempted to explain this phenomenon in terms of a compactness assumption. A compactness assumption represents a default assumption that the particular view one has of an object is randomly selected among those possible. This particular assumption should come into play when depth is insufficiently determined. According to this view, in the absence of other information, the best estimate of the depth of a solid object is provided by its apparent width (width should be used preferentially to height because objects are often oriented relative to the gravity-specified vertical). We believe that this heuristic may be regarded as one of many within a biological perceptual system that can take into account both the content and the reliability of the information available to it.

The SFM stimuli used in our experiments involved small angles of rotation, and we have argued that, for such small angles, the foreshortening (or acceleration) information available to specify the angle of rotation to an observer is understandably difficult to recover (cf. Simpson, in press). This is not to say that angular information is never used. For larger angles of rotation (>15°), there is clear evidence that compression information affects judgments of depth in the appropriate direction (Braunstein et al., 1993; Caudek & Proffitt, 1994; Proffitt et al., 1992). For example, Braunstein et al. (1993) have shown that when additional compression information was added to a motion parallax display (indicating rotations of 20° and 28.4° out of the picture plane), the apparent depth of a monocularly viewed horizontal dihedral angle was decreased. This is consistent with an increase in perceived object rotation.

Moreover, there is reason to believe that there is better information about the effective angle in self-produced motion parallax when an enriched viewing environment is available. In Experiment 2, our MP group demonstrated a greater sensitivity to depth than did our equivalent object rotation group. We believe that this difference can be understood as arising from richer sources of information about effective object rotation in the MP condition. This information could be given by sensitivity to the change in angle of regard, which might be supported by optic flow information in well-illuminated, structured environments.

A convergent interpretation of how depth from motion parallax might be scaled more accurately when supporting information is given has been put forth by M. E. Ono et al. (1986). They provided an analysis of motion parallax (defined as differential projected motions produced by self-motion) as an analog of binocular disparity in which they expressed the quantity of parallax-induced "disparity" relative to a unit of head motion that was equal to interocular distance. From this analysis, they argued that viewing distance must be taken into account, in the same manner as in stereopsis, to retrieve depth from motion parallax. To test this conjecture, they used random dot motion parallax displays that were yoked to head motion, like those of Rogers and Graham (1982), but the displays were viewed from several distances in a well-illuminated environment. In one study, their stimuli simulated vertical sinusoidal forms with matched retinal characteristics of size, peak-to-peak horizontal distance, and movement-induced parallax at each of two distances (40 and 80 cm). With a doubling of distance, equivalent parallax (like equivalent disparity) signals a change in depth by a factor of roughly 4, whereas a failure of depth constancy would predict no change in perceived depth. Ono et al. found a mean change in perceived depth by a factor of about 3.5, consistent with nearly complete depth constancy from motion parallax. This result is consistent with the current analysis of information about the effective angular rotation of objects: At a greater viewing distance, the same horizontal motion of a viewer produces less (effective) angular rotation of the object. Thus, an observer in Ono et al.'s situation who successfully extracted information about ego-relative object velocity (angular rotation) would appear to implicitly have taken distance into account.[13] At present, Ono et al.'s account in terms of distance

[13] Rivest, Ono, and Saida (1989) have demonstrated that depth judgments from observer-produced motion parallax displays are affected by information about apparent distance. However, their results are even more consistent with the recovery of angle information. First, they did not find any effect of manipulating participants' vergence angle. Second, they did find differences in perceived depth from motion parallax for displays presented alongside dollar bills of which the size was manipulated (as in O'Leary & Wallach, 1980); however, the depth differences were also quantitatively consistent with simple size scaling (i.e., were roughly linear rather than quadratic). Finally, the only condition in which the change in judged depth was quadratically related to a change in apparent distance was that in which apparent distance to a display was manipulated by means of an induction screen placed in front of the display and the change in angular position of the center of the motion parallax display (through the induction aperture) was made consistent with the intended "false" distance. In this case, the change in angular regard was consistent with the nearer false distance. It therefore seems that angular information, rather than distance per se, may be most important for scaling depth from motion parallax.

and our own in terms of angle appear to be notational variants, but it is clear from their results and strongly suggested by our Experiment 2 that information necessary to specify the effective angular displacement of objects can indeed be used in the recovery of depth from motion parallax.

The heuristic view of perception has implications for understanding the interaction of various kinds of depth information. Recent research has attempted to examine the quantitative combination of depth information from different sources, including SFM and binocular disparity (Bulthoff & Mallot, 1988; Clark & Yuille, 1990; Dosher, Sperling, & Wurst, 1986; Johnston, Cumming, & Parker, 1993; Landy, Maloney, & Young, 1991; Rogers & Collett, 1989; Tittle & Braunstein, 1991, 1993). Tittle and Braunstein, for example, found that motion facilitates depth from disparity. Although this is partly an incidental facilitation in helping to solve the stereo correspondence problem (a problem that is most pronounced for transparent random dot–specified surfaces of the kind used for their research), they argued that differential velocity is an important factor. The present studies suggest that modeling the combination of cues involved in perceiving depth may involve including inherent default biases, such as the compactness assumption, and specific distance tendencies. Moreover, the result of having multiple kinds of depth information is unlikely to be adequately modeled by a combination of independent quantitative depth estimates from different sources. We would argue that, much as disparity information must be combined with distance information to produce depth, the SFM system does not provide, in itself, a definite depth estimate except in combination with other information that includes (somewhat ineffective) default assumptions. Tittle and Braunstein (1991) reported data on the combination of different levels of disparity-specified and KDE-specified depth that are quite consistent with our findings: The primary determinant of perceived depth (Euclidean shape) is binocular disparity, and motion information seems to additively increment the perceived depth in a manner consistent with the SKE heuristic of Caudek and Proffitt (1993, 1994).

In general, efforts to study depth cue combination quantitatively must recognize that disparity information, per se, does not provide metric depth information without taking distance into account. Insofar as different sources of distance information can be used, a comprehensive study of depth cue combination involving binocular disparity would probably need to explicitly manipulate the kinds and quality of distance information available. The manner in which disparity information combines with other kinds of cues to three-dimensional shape may depend on the reliability of the distance information necessary to support stereoscopic depth perception. Upon reduced viewing conditions, with a static observer, that reliability may be low.

With respect to investigations involving stereoscopic depth (and motion parallax), our findings suggest that computer simulations in a dimly lit or otherwise unstructured environment may represent a very limited scenario for investigations of metric form perception (for a noteworthy exception, see Lappin & Love, 1992). Although experiments in reduced environments can help to adjudicate between theories that do not suggest a role for a supporting environment, a full theory of stereoscopic depth perception seems likely to depend on the distance information that may be available in well-illuminated, fully structured environments. Proffitt and Kaiser (1986) have pointed out a number of potential limitations introduced by computer simulation in general and have recommended that natural objects also be used, when possible, to provide convergent information with computer simulation. Indeed, there has recently been a report (Buckley & Frisby, 1993) of a situation in which stereoscopic depth perceptions involving real objects differed from those involving stereograms. Although results from studies using reduced viewing conditions and computer-simulated stimuli are not, in themselves, invalid, they may be limited, and the present study demonstrates the usefulness of having additional data involving not only real objects but also a well-illuminated, structured environment. We believe that not only the "reality" of the target stimuli themselves, but also that of the environment in which they are presented, is an important factor to consider in investigations of stereoscopic depth perception by itself or in combination with other information about depth.

A new complementarity of SFM and binocular stereopsis has recently been proposed by Olson (1991). Although some stereoscopic depth can be recovered even when disparities are so large that fusion fails (Ogle, 1952), the range of retinal disparities that can be precisely discriminated is quite limited (McKee, Levi, & Bowne, 1990). Thus, disparity information for near fixations is most informative within a fairly small region about the fixation plane (Ogle, 1953; Westheimer & McKee, 1978). Although the proportional depth of field from small disparities increases with distance, the accuracy of depth from stereopsis for fixation distances of more than a few meters may be limited by the precision or accuracy of distance perception.[14] On the other hand, motion parallax can function over the entire visual field and at least recover a rough sense of global structure as the observer moves through the environment (e.g., Cutting, 1986). Thus, the two systems appear to have different domains of maximum utility (see also McKee et al., 1990). We believe that this view of the complementarity of motion and disparity information has promise for guiding future research.

Conclusion

The accurate perception of depth from binocular disparity information entails the use of accurate distance information, which, in nonreduced circumstances, may be precise enough to specify object depth within the range of distances normally available for human manipulation. With computer simulations and with real objects, we found that depth judgments based on static binocular viewing were quite

[14] This statement must be qualified by consideration of results (e.g., Loomis, Da Silva, Fujita, & Fukusima, 1992) demonstrating fairly precise egocentric distance information available for blind walking to targets up to 12 m distant.

accurate in this range. Conversely, the recovery of depth from relative motions produced by object rotation or motion parallax entails obtaining angular displacement information that may be difficult to determine precisely in many instances. Our studies confirm that depth is not accurately recovered even in the case of motion parallax, in which angular displacement information can be fairly well specified. For small rotations, depth magnitude perceived from relative motions appears to be mediated by heuristic assumptions and biases (Caudek & Proffitt, 1993, 1994). We conclude that, for near objects in a fully structured environment, the recovery of depth from static binocular viewing is normally well determined, whereas depth from motion is poorly determined for small rotations and angular displacements.

References

Aloimonos, J., & Brown, C. M. (1989). On the kinetic depth effect. *Biological Cybernetics, 60,* 445–455.

Baird, J. C. (1970). *Psychophysical analysis of visual space.* Oxford, England: Pergamon Press.

Baird, J. C., & Biersdorf, W. R. (1967). Quantitative functions for size and distance judgments. *Perception & Psychophysics, 2,* 161–166.

Ballard, D. H., & Ozcandarli, A. (1988). Eye fixation and early vision. In *Proceedings of the 2nd International Conference on Computer Vision* (pp. 524–531). Washington, DC: Institute of Electrical and Electronic Engineers, Computer Society Press.

Bennett, B. M., Hoffman, D. D., Nicola, J. E., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America, 6A,* 1052–1069.

Braunstein, M. L. (1962). The perception of depth through motion. *Psychological Bulletin, 59,* 422–433.

Braunstein, M. L. (1976). *Depth perception through motion.* New York: Academic Press.

Braunstein, M. L., Liter, J. C., & Tittle, J. S. (1993). Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 598–614.

Brunswick, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley: University of California Press.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research, 33,* 919–933.

Bulthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the American Optical Society, 5A,* 1749–1758.

Carlson, V. R. (1977). Instructions and perceptual constancy judgments. In W. Epstein (Ed.), *Stability and constancy in visual perception* (pp. 217–254). New York: Wiley.

Caudek, C., & Proffitt, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 32–47.

Caudek, C., & Proffitt, D. R. (1994). *Perceptual biases in the kinetic depth effect.* Manuscript submitted for publication.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems.* Boston: Kluwer.

Collett, T. S., Schwarz, U., & Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception, 20,* 733–754.

Cormack, R. H. (1984). Stereoscopic depth perception at far viewing distances. *Perception & Psychophysics, 35,* 423–428.

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical disparities and the perception of three-dimensional shape. *Nature, 349,* 411–413.

Cutting, J. E. (1986). *Perception with an eye for motion.* Cambridge, MA: MIT Press.

Dosher, B. A., Sperling, G., & Wurst, S. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973–990.

Eagle, R. A., & Blake, A. (1994). 2-D limits on 3-D structure-from-motion tasks [Abstract]. *Investigative Ophthalmology & Visual Science, 35,* 1277.

Fisher, S. K., & Ebenholtz, S. M. (1986). Does perceptual adaptation to telestereoscopically enhanced depth depend on the recalibration of binocular disparity? *Perception & Psychophysics, 40,* 101–109.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review, 87,* 411–434.

Foley, J. M. (1985). Binocular distance perception: Egocentric distance tasks. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 133–149.

Gilinsky, A. (1951). Perceived size and distance in visual space. *Psychological Review, 58,* 460–482.

Gogel, W. C. (1969a). The effect of object familiarity on the perception of size and distance. *Quarterly Journal of Experimental Psychology, 21,* 239–247.

Gogel, W. C. (1969b). The sensing of retinal size. *Vision Research, 9,* 1079–1094.

Gogel, W. C. (1982). Analysis of the perception of motion concomitant with a lateral motion of the head. *Perception & Psychophysics, 32,* 241–250.

Gogel, W. C., & Da Silva, J. A. (1987). Familiar size and the theory of off-sized perceptions. *Perception & Psychophysics, 41,* 318–328.

Gogel, W. C., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics, 13,* 284–292.

Gottsdanker, R. M., Frick, J. W., & Lockard, R. B. (1961). Identifying the acceleration of visual targets. *British Journal of Psychology, 52,* 31–42.

Graham, M. E., & Rogers, B. J. (1982). Simultaneous and successive contrast effects in the perception of depth from motion-parallax and stereoscopic information. *Perception, 11,* 247–262.

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

Johansson, G. (1973). Monocular movement parallax and near-space perception. *Perception, 2,* 135–146.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31,* 1351–1360.

Johnston, E. B., Cumming, G. G., & Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research, 33,* 813–826.

Koenderink, J., & Van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics, 21,* 29–35.

Koenderink, J., & Van Doorn, A. J. (1981). Exterospecific component of the motion parallax field. *Journal of the Optical Society of America, 71,* 953–957.

Kontsevich, L. L. (1994). The depth-scaling parameter for two-frame structure-from-motion matches the default vergence angle

[Abstract]. *Investigative Ophthalmology & Visual Science, 35,* 1277.

Landy, M. S., Maloney, L. T., & Young, M. J. (1991). Psychophysical estimation of the human depth combination rule. In P. S. Schenker (Ed.), *Proceedings of the Society of Photo-Optical Instrumentation Engineers: Vol. 1838. Sensor fusion III: 3-D perception and recognition* (pp. 247–254). Bellingham, WA: Society of Photo-Optical Instrumentation Engineers.

Lappin, J. S. (1990). Perceiving the metric structure of environmental objects from motion, self-motion and stereopsis. In R. Warren & A. H. Wertheim (Eds.), *Perception and control of self-motion* (pp. 541–578). Hillsdale, NJ: Erlbaum.

Lappin, J. S., & Love, S. R. (1992). Planar motion permits perception of metric structure in stereopsis. *Perception & Psychophysics, 51,* 86–102.

Leibowitz, H. W., & Harvey, L. O., Jr. (1969). Effect of instructions, environment, and type of test object on matched size. *Journal of Experimental Psychology, 81,* 36–43.

Leibowitz, H. W., Hennessy, R. T., & Owens, D. A. (1975). The intermediate resting position of accommodation and some implications for space perception. *Psychologia, 18,* 162–170.

Leibowitz, H. W., Shina, K., & Hennessy, R. T. (1972). Oculomotor adjustments and size constancy. *Perception & Psychophysics, 12,* 25–29.

Loomis, J. M., & Beall, A. C. (1993). *Visual processing of depth from motion is similar for object translation and object rotation.* Unpublished manuscript.

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 906–921.

Loomis, J. M., & Eby, D. W. (1988). Perceiving structure from motion: Failure of shape constancy. In *Proceedings of the 2nd International Conference on Computer Vision* (pp. 383–391). Washington, DC: Institute of Electrical and Electronic Engineers: Computer Society Press.

Luneburg, R. K. (1947). *Mathematical analysis of binocular vision.* Princeton, NJ: Princeton University Press.

McKee, A. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30,* 1763–1779.

Musatti, C. L. (1924). Sui fenomeni stereocineti [On the stereokinetic phenomenon]. *Archivio Italiano di Psicologia, 3,* 105–120.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53,* 279–291.

Ogle, K. N. (1952). On the limits of stereoscopic vision. *Journal of Experimental Psychology, 44,* 253–259.

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America, 40,* 627–642.

O'Leary, A., & Wallach, H. (1980). Familiar size and linear perspective as distance cues in stereoscopic depth constancy. *Perception & Psychophysics, 27,* 131–135.

Olson, T. J. (1991). Stereopsis for fixating systems. In *Proceedings, Institute of Electrical and Electronic Engineers International Conference on Systems, Man, and Cybernetics* (pp. 7–12). Charlottesville, VA: IEEE.

Ono, H., & Comerford, J. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception* (pp. 91–128). New York: Wiley.

Ono, H., & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception & Psychophysics, 48,* 179–187.

Ono, M. E., Rivest, J., & Ono, H. (1986). Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 331–337.

Owens, D. A. (1987). Oculomotor information and perception of three-dimensional space. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 215–248). Hillsdale, NJ: Erlbaum.

Owens, D. A., & Leibowitz, H. W. (1976). Oculomotor adjustments in darkness and the specific distance tendency. *Perception & Psychophysics, 20,* 2–9.

Owens, D. A., & Leibowitz, H. W. (1980). Accommodation, convergence, and distance perception in low illumination. *American Journal of Optometry & Physiological Optics, 57,* 540–550.

Pollick, F. E. (1994). *The perception of motion and structure in structure-from-motion: Comparisons of affine and Euclidean tasks.* Manuscript submitted for publication.

Predebon, J. (1992). The role of instructions and familiar size in absolute judgments of size and distance. *Perception & Psychophysics, 51,* 344–354.

Predebon, J. (1993). The familiar-size cue to distance and stereoscopic depth. *Perception, 22,* 985–995.

Proffitt, D. R., & Kaiser, M. K. (1986). The use of computer graphics animation in motion perception research. *Behavior Research Methods, Instruments, & Computers, 18,* 487–492.

Proffitt, D. R., Rock, I., Hecht, H., & Schubert, J. (1992). The stereokinetic effect and its relations to the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 3–21.

Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America, 2A,* 343–349.

Rivest, J., Ono, H., & Saida, S. (1989). The roles of convergence and apparent distance in depth constancy with motion parallax. *Perception & Psychophysics, 46,* 401–408.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature, 361,* 253–255.

Rogers, B. J., & Cagnello, R. B. (1989). Disparity curvature and the perception of three-dimensional surfaces. *Nature, 339,* 135–137.

Rogers, B., & Collett, T. S. (1989). The appearances of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology, 41A,* 697–717.

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22,* 261–270.

Salvendy, G. (Ed.). (1987). *Handbook of human factors.* New York: Wiley.

Simpson, W. A. (1993). Optic flow and depth perception. *Spatial Vision, 7,* 35–75.

Simpson, W. A. (in press). Temporal summation of visual motion. *Vision Research.*

Suppes, P. (1977). Is visual space Euclidean? *Synthese, 35,* 397–421.

Tittle, J. S., & Braunstein, M. L. (1991). Shape perception from binocular disparity and structure-from-motion. In P. S. Schenker (Ed.), *Proceedings of the Society of Photo-Optical Instrumentation Engineers: Vol. 1838. Sensor fusion III: 3-D perception and recognition* (pp. 225–234). Bellington, WA: Society of Photo-Optical Instrumentation Engineers.

Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics, 54,* 157–169.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional

affine structure from minimal apparent motion sequences. *Journal of Experimental Psychology: Human Perception and Performance, 48,* 419–430.

Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics, 50,* 509–523.

Todd, J. T., & Norman, J. F. (1994). The visual perception of 3D length in natural vision [Abstract]. *Investigative Ophthalmology & Visual Science, 35,* 1329.

Toye, R. C. (1986). The effect of viewing position on the perceived layout of space. *Perception & Psychophysics, 40,* 85–92.

Tyler, C. W. (1983). Sensory processing of binocular disparity sensitivity. In C. M. Schor & K. J. Ciuffreda (Eds.), *Vergence eye movements: Basic and clinical aspects* (pp. 199–295). Boston: Butterworth.

Tyler, C. W. (1991). The horopter and binocular fusion. In D. Regan (Ed.), *Binocular vision* (pp. 19–37). Boca Raton, FL: CRC Press.

Wagner, M. (1985). The metric of visual space. *Perception & Psychophysics, 38,* 483–495.

Wallach, H., Moore, M. E., & Davidson, L. (1963). Modification of stereoscopic depth-perception. *American Journal of Psychology, 76,* 191–204.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45,* 205–217.

Wallach, H., & Zuckerman, C. (1963). The constancy of stereoscopic depth. *American Journal of Psychology, 76,* 404–412.

Westheimer, G., & McKee, S. P. (1978). Stereoscopic acuity for moving retinal images. *Journal of the Optical Society of America, 68,* 450–455.

Appendix

A Mathematical Comparison of Stereopsis and Structure From Motion

In this appendix, we explore the relationship between depth and optical flow in kinetic depth and motion parallax and compare it with the relationship between disparity and depth in stereopsis. We assume visual fixation in each case as follows: In the kinetic depth case, the target object rotates around the fixation point, and in motion parallax the observer translates while keeping the gaze fixed on the target. In stereopsis, the optic axes of the eyes converge at a point on the target. The analysis to follow shows that recovering depth in the two motion cases requires information that is more difficult to obtain than the information required to recover depth from stereo. The derivations are based on those of Olson (1991) and of Ballard and Ozcandarli (1988).

Figure A1 shows the imaging situation for the two motion cases. We assume perspective projection with the focal point of the imaging system at the origin. The image plane is parallel to the X-Y plane and passes through $(0, 0, f)$, so that a world point (X, Y, Z) projects to $(x, y) = f(X/Z, Y/Z)$ in the image plane. For the case of motion parallax, the observer translates along the x-axis at a speed of $\dot{X}$, maintaining fixation on point F. This induces an instantaneous rotation of object points about F with angular velocity $\dot{X} = \dot{X}/Z_F$. For the kinetic depth case, the observer is stationary ($\dot{X} = 0$) and the target rotates about F with a constant speed of $\dot{\theta}$.

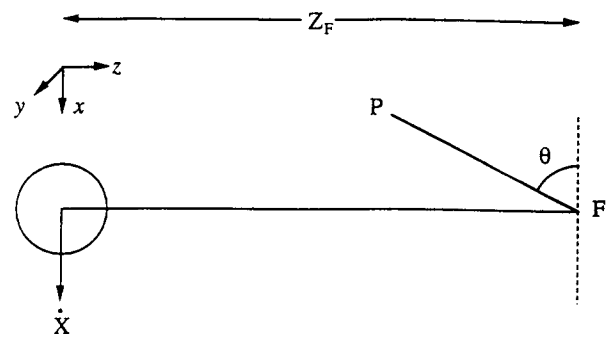


Figure A1. Imaging situation for structure from motion (rotation or motion parallax). A point P at angle θ from the image phase is viewed while fixating F at a distance Z_F and translating at a rate of $\dot{X}$ (motion parallax). For rotation, θ changes.

Consider object point P as a distance r from the fixation point. P can be expressed as $(Xp, Yp, Zp) = (-r\cos\theta, Yp, Zf - r\sin\theta)$, so the x-coordinate of this image plane projection is given by

$$x = -f\frac{r\cos\theta}{Z_F - r\sin\theta}. \tag{A1}$$

Differentiating with respect to time gives

$$\dot{x} = f\dot{\theta}\left[\frac{r\sin\theta}{Z_F - r\sin\theta} - \left(\frac{r\cos\theta}{Z_F - r\sin\theta}\right)^2\right]. \tag{A2}$$

Solving for Z and combining with the expressions for P yield the depth:

$$Z = \frac{Z_F}{1 + \dfrac{x^2}{f^2} + \dfrac{\dot{x}}{f\dot{\theta}}}. \tag{A3}$$

For the motion parallax case, this can be written in terms of the observer's velocity:

$$Z = \frac{Z_F}{1 + \dfrac{x^2}{f^2} + \dfrac{Z_F\dot{x}}{f\dot{X}}}. \tag{A4}$$

The case of stereopsis under binocular fixation is essentially a discrete version of the observer translation case. Figure A2 shows the assumed geometry. In this situation, it can be shown (Olson, 1991) that the depth of an arbitrary world point is given by

$$Z = Z_F\left[\frac{(f - x_L\tan\theta_L)(f + x_R\tan\theta_R)}{f^2 + x_Lx_R + \dfrac{f(x_L - x_R)}{\tan(\theta_L + \theta_R)}}\right]. \tag{A5}$$

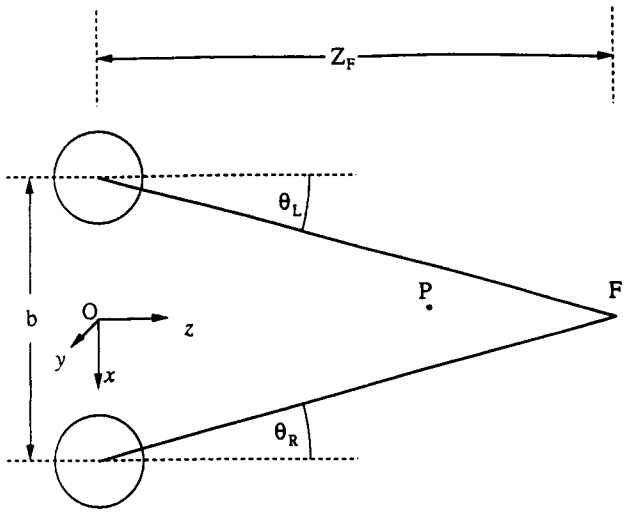


Figure A2. Imaging situation for binocular stereopsis (two views with a baseline separation, *b*, and fixation point, *F*, regarding point with vergence angles θ_R and θ_L).

When the gaze direction is forward (i.e., $\theta_L \approx \theta_R$), the preceding expression can be well approximated, for the region of central vision, by

$$Z \approx \frac{Z_F}{1 + \frac{x_L x_R}{f^2} + \frac{Z_F}{2bf}(x_L - x_R)}. \quad \text{(A6)}$$

The depth equations for the three cases are similar in form, and all require knowledge of the fixation depth Z_F to determine absolute depth. For reasonably close fixations, Z_F can be estimated from accommodation and (in the case of stereopsis) vergence angle. Aside from Z_F, the stereo depth equation depends only on quantities that can be measured in the images (disparity and position of feature points) and on intrinsic properties of the imaging system (baseline and focal length). The motion equations, on the other hand, require an additional extrinsic parameter, the angular velocity $\dot{\theta}$. For the motion parallax case, this is, in principle, computable from the observer's translational velocity and the fixation depth or simply from the angular velocity required to track the fixation point. For the kinetic depth case, the depth is underdetermined unless the angle of rotation can be extracted. In the introduction, we showed how angular rotation information could be derived from foreshortening information, although, for small rotations or small increments of rotation, this information is poorly specified.

In summary, all of the parameters required to determine depth from stereoscopic disparity are readily available. In the case of motion parallax, an additional parameter (the angular velocity) must be estimated, and it must be assumed that the target does not have a concomitant rotation. If the target moves (as in the kinetic depth case), depth is underdetermined. The fact that recovering structure from motion requires more assumptions and has an additional free parameter suggests that obtaining a complete three-dimensional solution in the motion cases may not be worth the effort. This might explain why, in the experiments described in this article, shape judgments based on stereo information proved far more accurate than judgments based on either kinetic depth or motion parallax.

Received July 2, 1993
Revision received July 18, 1994
Accepted August 26, 1994 ■