Journal of Experimental Psychology:
Human Perception and Performance
1995, Vol. 21, No. 3, 663–678

Systematic Distortion of Perceived Three-Dimensional Structure From Motion and Binocular Stereopsis

James S. Tittle, James T. Todd, Victor J. Perotti, and J. Farley Norman
Ohio State University

The geometric relation between physical and perceived space as specified by binocular stereopsis and structure from motion was investigated. Four experimental tasks were used, each of which required a different aspect of three-dimensional (3-D) structure to be performed accurately. To examine whether the transformation between physical and perceptual space preserved the 3-D structural properties required to perform each of our tasks, the constancy of judged shape over changes in a depicted object's viewing distance or orientation was examined. Our results reveal that observers' judgments of 3-D shape from binocular stereopsis and motion contained systematic distortions: Perceived 3-D shape from motion was not invariant over orientation change and perceived 3-D structure from stereo, and motion and stereo in combination was not invariant over changes in viewing distance.

Everyday experience reveals that the human visual system obtains sufficient information about three-dimensional (3-D) structure to safely navigate around and interact with objects in the environment. The ease with which people seem to perform such tasks obscures the fact that the relation between retinal images and 3-D configurations is one to many. Consequently, most recent theories of 3-D structure have concentrated on what additional constraints or assumptions are necessary to define the mapping between a particular pattern of optical stimulation and the 3-D configuration that gave rise to it. For example, Ullman (1979) showed that by using a rigidity constraint, one could recover the 3-D structure (up to a reflection about the image plane) and motion of an object from three views of four noncoplanar points.

For binocular stereopsis, the problem is also underconstrained because a particular horizontal relative disparity specifies depth only up to a scaling factor and does not uniquely define a depth interval. Thus, researchers have proposed that recovering shape from stereopsis requires the use of additional information, such as the perceived distance to fixation (Foley, 1980) or vertical disparity (Mayhew & Longuet-Higgins, 1982). Recent evidence, however, strongly suggests that people cannot use information about fixation distance to perceive shape veridically from stereopsis (Foley, 1980; Johnston, 1991). In addition, vertical disparities have been shown to influence the depth obtained from horizontal disparities only for visual angles greater than 30° (Fox, Cormack, & Norman, 1987; Rogers & Bradshaw, 1992).

Another means of resolving the stereo scaling problem, proposed by several researchers, is to combine stereo with structure from motion (Richards, 1985). According to this view, stereo resolves the reflection ambiguity, but depth is scaled according to the motion information. Although such an approach could work, in theory, recent evidence indicates that observers do not perceive veridical shape in structure-from-motion displays (Liter, Braunstein, & Hoffman, 1993; Loomis & Eby, 1988; Norman & Todd, 1993; Todd & Bressan, 1990; Todd & Norman, 1991). Furthermore, Rogers and Collett (1989) and Tittle and Braunstein (1993) found that when stereo and motion simulated different shapes, the judged shape was determined almost entirely by the stereo information.

Although there have been numerous reports in the literature that the visual perception of 3-D structure can be surprisingly inaccurate and unreliable—even in conditions in which there are multiple sources of information available (e.g., see Baird & Biersdorf, 1967; Loomis, Da Silva, Fujita, & Fukusima, 1992; Wagner, 1985)—there have been many other reports of experimental tasks involving the perception of 3-D structure for which observers' judgments are remarkably precise. For example, research on binocular vision has shown that observers can reliably detect depth differences from disparities as small as a few seconds of arc (Ogle, 1952). Similarly, in perceiving structure from motion, observers can accurately detect certain types of nonrigid deformations or determine whether a configuration of points is coplanar (e.g., Norman & Lappin, 1992; Todd & Bressan, 1990).

It would appear from these seemingly conflicting reports that observers can accurately judge certain aspects of 3-D structure but not others. In an effort to develop a possible theoretical explanation for why some perceptual tasks are easy and others are relatively hard, it is useful to consider the possible mathematical relations between physical space

---

James S. Tittle, James T. Todd, Victor J. Perotti, and J. Farley Norman, Department of Psychology, Ohio State University.

This research was supported in part by Ohio State University Seed Grant 221465 and Air Force Office of Scientific Research Grant F49620-93-1-0116. Portions of this research were presented at the annual meeting of the Association for Research in Vision and Ophthalmology held in Sarasota, Florida, May 1993.

Correspondence concerning this article should be addressed to James S. Tittle, Department of Psychology, Ohio State University, 201 Iazenby Hall, Columbus, Ohio 43210. Electronic mail may be sent via Internet to jtittle@magnus.acs.ohio-state.edu.

Φ and perceived space Ψ. One such possibility, for example, is that these two spaces are congruent to one another, that is, $\Psi = \Phi$. An observer's perceptual judgments in that case would be perfectly veridical. Another possibility, however, is that the relation between physical and perceived space involves some type of systematic distortion, such that $\Psi = f(\Phi)$. Is there some particular function $f(\Phi)$ that can explain the specific patterns of success and failure exhibited by human observers in the visual perception of 3-D form? The research described in this article was designed to address this issue.

Our general approach to this problem was based on the hierarchy of geometries first proposed by the German mathematician Felix Klein in a speech at Erlangen University in 1872. Klein noted that different geometries can be categorized by the different types of transformations they allow and the structural properties of objects that remain invariant under those transformations. For example, euclidean geometry allows arbitrary translations and rotations, which preserve the distance between any pair of points on an object. Affine geometry, by contrast, also allows arbitrary stretching transformations, which do not leave distance invariant but do preserve a variety of other properties, such as the sign of Gaussian curvature or the parallelism of a pair of line segments.

If one considers the mapping between physical and perceived space as a geometric transformation, then it follows from Klein's analysis that some properties of 3-D structure will be systematically distorted and others will remain invariant. Those properties that are invariant over this mapping should be perceived with a high degree of accuracy, whereas judgments of properties that are not invariant should result in systematic errors. Conversely, by analyzing the specific perceptual tasks that yield accurate or inaccurate performance, it should be possible to identify the particular transformation by which physical and perceived space are mathematically related. Consider some of the possible candidates, which are depicted in Figure 1.

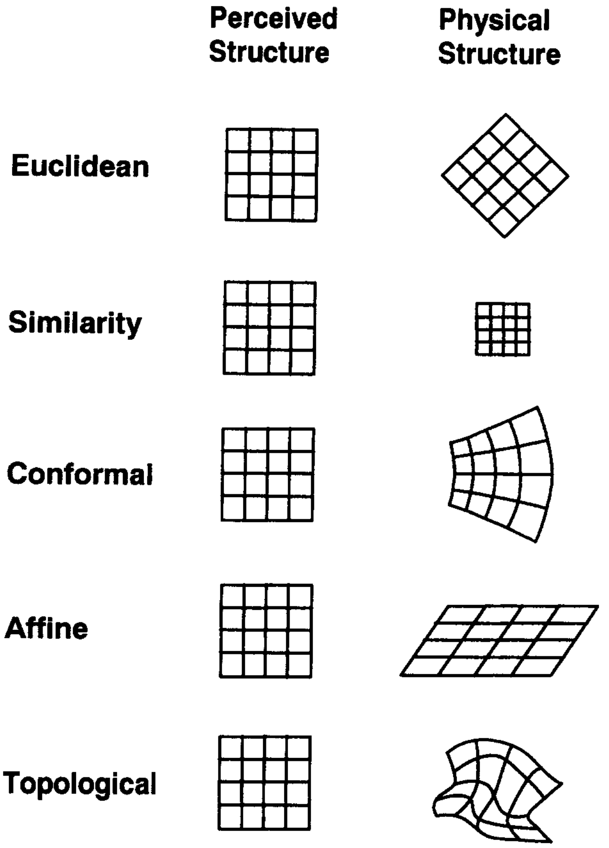

Figure 1. Some possible geometric transformations between physical and perceptual space.

A *similarity transformation* is perhaps the simplest possible distortion that could exist between physical and perceived space. This type of mapping would preserve almost all aspects of 3-D structure except for absolute size. If $\Psi = f(\Phi)$ were governed by a similarity transformation, observers would be capable of making accurate judgments of object shape or the relative length ratios or angles of line segments oriented in different directions.

Another possible type of perceptual distortion called a *conformal transformation* is one that preserves local angles. This particular type of mapping between physical and perceived space was implicit in the analysis of binocular vision proposed by Luneburg (1947, 1950) and Blank (1953, 1958), in which it was argued that the perceived metric structure of the environment is best described as a homogeneous Riemannian space of constant negative curvature. According to this analysis straight lines in the environment can appear curved, but within sufficiently small local neighborhoods, euclidean distance and angle relations would be preserved.

Other investigators have suggested that physical and perceived space are related by an *affine transformation,* such that distances in depth appear to be homogeneously stretched relative to the horizontal and vertical (e.g., Todd & Bressan, 1990; Todd & Norman, 1991; Wagner, 1985). Affine transformations preserve distance ratios in any given direction, which would make it possible, for example, to perform accurately on a distance bisection task (e.g., Lappin & Fuqua, 1983; Purdy & Gibson, 1955). However, when line segments oriented in different directions are subjected to an affine transformation, their relative lengths and angles can be systematically altered.

An even more severe pattern of perceptual distortion would be produced by *topological transformations,* which lie at the bottom of the Klein hierarchy. The primary invariant of continuous topological transformations is the preservation of local neighborhoods and patterns of connectivity. If perceived space were topologically related to physical space, then observers would still be able to reliably identify certain important aspects of 3-D structure such as surface smoothness or the presence of a discontinuity.

The research reported in this article was designed to investigate which of these different classes of transformation is most characteristic of the perceptual capabilities and limitations of human observers. We used four experimental tasks, each of which focused on a different aspect of 3-D structure. In Experiment 1, observers were required to adjust the eccentricity of a cylindrical surface until its cross section in depth appeared to be circular. Note that the overall size of the cylinder need not be perceived veridically to accurately perform this task. Its structure need only be specified up to a similarity transformation. In Experiment 2, observers adjusted the angle between two connected planes until they appeared to be perpendicular to one another. This task required that the 3-D structure of the surfaces be perceptually specified up to a conformal transformation. In Experiment 3, observers adjusted the angle between two separated planes until they appeared to be parallel to one another, which required that affine properties be preserved in the geometric mapping between physical and perceived space. Finally, in Experiment 4, observers adjusted the angle between two connected planes until there was no detectable discontinuity at their points of intersection. Although the planes also would be parallel in that case, this task could still be performed on the basis of the topological properties of a surface even if its affine properties were systematically distorted.

There are two other important aspects of the overall design of this research. First, in addition to examining how performance varied for different types of perceptual judgments, we also investigated how these judgments would be influenced by different sources of optical information including motion and stereo, both individually and in combination. Second, we investigated the extent to which these judgments would exhibit constancy over changes in viewing distance or the orientation of an object in depth.

Experiment 1

Method

Participants. Six experienced psychophysical observers served as the participants in this experiment. These observers included the four authors and two additional participants who were unfamiliar with the specific hypotheses under investigation.

Design. We examined the influence of four variables on judgments of shape: (a) simulated distance (70, 114, and 170 cm); (b) average slant out of the frontoparallel plane (0°, 15°, and 30°); (c) stimulus display type (monocular motion, static binocular stereopsis, or combined motion and stereopsis); and (d) angular velocity (1.0° or 1.5° per frame). This last manipulation can be applied only to the motion and combined display types; thus, there were a total of 45 conditions in this experiment.

Stimuli and apparatus. The stimuli were generated and presented on a Silicon Graphics Iris (4D/310 VGX) workstation. Displays were presented within an approximately 34 × 27 cm region of the CRT screen that had a spatial resolution of 1,280 × 1,024 pixels. However, hardware anti-aliasing enabled us to use subpixel positioning to make the effective resolution 16 times greater than the physical resolution of the screen. The edges of the CRT were covered with black felt, and the room was completely dark except for the luminous dots on the display screen.

The stereoscopic stimuli were presented using LCD shuttered glasses that were synchronized with the display refresh rate. The left and right views of a stereo pair were displayed at the same position on the screen, but they were temporally offset by 8.3 ms. The left and right lenses of the LCD glasses shuttered synchronously with the display, so that each view of the stereo pair was seen only by the appropriate eye. The CRT was capable of refreshing at 120 Hz, thus, each view in the stereo pair was updated at 60 Hz. Because running the CRT at 120-Hz refresh effectively halved the vertical resolution, the monocular motion stimuli also were displayed at 120 Hz, so that there would be no resolution difference between the monocular and binocular conditions.

The stimuli in this experiment simulated perspective projections of horizontally oriented cylinders rotating about a vertical axis (see Figure 2). The cylindrical surfaces were composed of bright dots positioned such that they had a uniform distribution on the display CRT for a frontoparallel view of the surface. The dots were four square pixels in size and subtended a visual angle of 0.03° at the 114-cm viewing distance used throughout the experiment. Preliminary observations indicated that the blue and green phosphors of our CRT had sufficiently long persistence to cause leakage between the left and right views of our LCD glasses. Thus, red dots

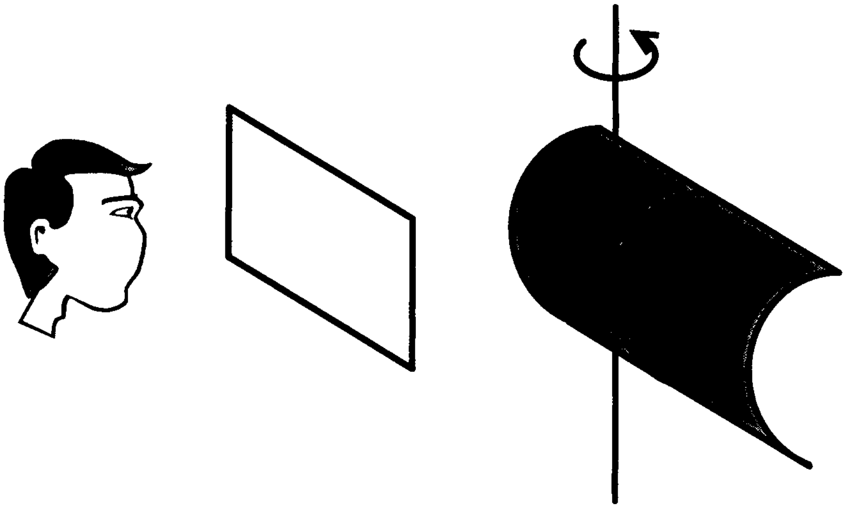

Figure 2. A schematic view of the cylinder stimuli used in Experiment 1.

were used for all stimuli. To eliminate the presence of object boundary deformations as a potential source of shape information, the displays simulated viewing through a rectangular window or aperture (see Figure 3 for a sample stereogram). Although observers were always positioned 114 cm away from the display CRT, we simulated three viewing distances: 70, 114, and 170 cm. For the monocular motion conditions, this was accomplished by generating displays with a perspective consistent for each of the viewing distances. For the stereo and combined motion–stereo conditions, simulated distance was manipulated by changing both the perspective and the amount of convergence needed to achieve global fusion of the display. The horizontal extent of the aperture was constant throughout the experiment and subtended a visual angle of 11.8°. This area was equivalent to the region of the display containing matched points between the left and right views of a stereo pair. Because the cylindrical object had a constant vertical size (in 3-D) throughout the experiment, the vertical two-dimensional size of the stimulus display decreased as simulated distance increased. When the surface was accurately adjusted to be a hemicylinder, for the 70-, 114-, and 170-cm distances, its vertical extent subtended 9.0°, 5.5°, and 3.7°, respectively. Because the size of dots (on the CRT) did not vary as a function of distance, the density values for the three distances and slants were chosen by the experimenters to maintain constant perceived surface density. The minimum number of dots on the cylindrical surface occurred in the 170-cm distance, 0° slant condition (7,600), and the maximum number of dots occurred in the 70-cm distance, 30° slant condition (23,000).

The perception of the 3-D shape of the cylindrical object was created through motion (rotation about a vertical axis that touched the front surface of the object), binocular stereopsis, or both in combination. For the motion conditions, the object oscillated ±5° about the average slant out of the frontoparallel plane (0°, 15°, or 30°). Although the displays were refreshed at 60 Hz, each view of the motion sequence was visible for 100 ms. Preliminary observations indicated that this stimulus onset asynchrony led to the smoothest perceived motion (Todd, Akerstrom, Reichel, & Hayes, 1988). Binocular disparities were generated by presenting different views, separated by a simulated interocular distance of 6.1 cm, to the right and left eyes. We could present motion and binocular disparity information separately or in combination, but when they were presented together motion and binocular disparity always provided consistent information about 3-D shape. In addition, a fixation cross subtending 0.25° of visual angle was positioned on the surface, where it intersected the axis of rotation. The purpose of this cross was to help observers retain global fusion throughout the duration of a trial.

Observers adjusted the shape of the cylindrical surface by adjusting the workstation mouse. The 3-D horizontal extent of the surface was constant throughout a trial, but movement of the mouse caused the object to be rescaled along its depth axis (which was aligned with the z-axis of the viewing geometry when the cylinder was in a frontoparallel orientation); because of perspective effects, this also caused the vertical extent of the surface to vary.

A mathematical analysis of the motion and disparity fields for these displays is given in the Appendix. For the cylinders in a frontoparallel orientation (i.e., those with zero slant), all of the horizontal disparities were uncrossed. When adjusted veridically, they varied with position over a range from 0 to 6 min at the farthest convergence distance of 170 cm and from 0 to 27 min at the nearest convergence distance of 70 cm. These ranges increased dramatically when the cylinders were slanted in depth. For example, at a 30° slant the horizontal disparities ranged from −6 min (crossed) to 14 min (uncrossed) at the far distance and from −16 to 47 min at the near distance. Despite the large magnitude of these disparities in the most peripheral regions of the displays, all observers reported that the objects appeared as fused single images.

The presence of vertical disparities in these displays is important because they provided potential information about the convergence distance independently of any extraretinal information about the actual positions of the eyes (e.g., Longuet-Higgins, 1981). The maximum vertical disparities were relatively unaffected by changes in slant, although they did decrease significantly with increasing convergence distance from approximately 2.4 to 0.4 min.

The horizontal and vertical extents of moving elements in the image (over the entire 10° rotation) were qualitatively similar to the pattern of disparities, but their magnitudes were larger. For example, when the surfaces were viewed at a convergence distance of 70 cm and an average slant of 30°, the displacement of dots in the image varied with position over a range of −15 to 106 min. The maximum vertical displacements attributable to perspective ranged from approximately 8.5 min at simulated distances of 70 cm to approximately 4 min at simulated distances of 170 cm.

Procedure. Observers were seated 114 cm away from the

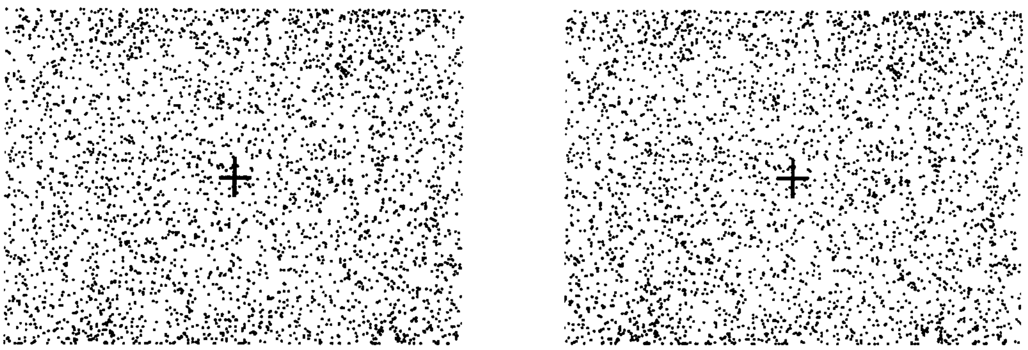

Figure 3. A sample stereogram of a frontoparallel cylindrical surface viewed through an aperture similar to those used in Experiment 1. Observers were required to adjust the eccentricity of the depicted surface so that its cross-section appeared circular. Actual displays in all experiments consisted of bright dots against a dark background.

display CRT in a darkened room and viewed the stimulus displays with their head held still by a chin rest. The experiment was divided into five sessions, run on separate days, each consisting of three blocks. The motion, binocular stereopsis, and motion–stereo combined conditions were each presented as a separate block. However, to provide a clear test of perceptual constancy, we presented all variations of the distance, slant, and velocity manipulations in each block. The order of the three blocks was counterbalanced in each session, and every block contained two replications of each combination of the slant, distance, and velocity variables. Thus, the total number of trials per session was 90. Observers were presented with approximately five practice trials and instructed to adjust the shape of the visible surface until its cross section appeared circular. Before each trial, a nonius marker appeared to ensure that observers were properly converged for the simulated viewing distance on that particular trial. This fusion stimulus was visible until the observer pushed a button on the mouse, after which the cylinder stimulus appeared and remained visible for as long as the participant needed to make the adjustment. In addition, the cylinder stimulus contained a fixation marker to help ensure that observers were converged at the appropriate distance throughout a trial. At the beginning of each new trial, the initial shape of the cylindrical object was randomly assigned within a range ±5 cm of the value for a true circular cylinder.

Results and Discussion

In Experiment 1, observers adjusted the eccentricity of a cylindrical object until it appeared to have a semicircular cross section. To make this adjustment, they needed to estimate the ratio of two nonparallel depth intervals, so that accurate performance required that the relation between physical and perceptual space was at least a similarity mapping. In performing this task, all observers complained that it was difficult. Although their adjustments produced large changes in the perceived depth of the depicted surfaces, they did not have a clear sense in the absence of feedback of the precise setting that should correspond to a circular cross section.

This uncertainty is reflected in the overall reliability of their judgments. To provide a quantitative measure of reliability, we calculated the standard deviation as a percentage of the mean for the 10 repeated judgments made by each observer for each of the 45 different combinations of stimulus parameters. The results reveal that there was considerable variability when a given observer judged the same stimulus over multiple occasions. For the stereo and combined displays, the average standard deviation (as a proportion of the mean) across all of the conditions was approximately 18%, whereas in the motion-alone condition it was approximately 25%. These values were similar for all individual observers, and there were no systematic differences as a function of slant or distance.

Despite this relatively large intertrial variability, however, there were clear and consistent effects of slant and distance in the overall pattern of means. The average adjusted eccentricity values, over all 6 observers, are presented in Figure 4. These data are scaled such that an adjusted cylinder eccentricity of 1.0 represents a veridical judgment, values less than this indicate overestimation of depth, and those greater than 1.0 indicate an underestimation. The data were analyzed using a one-way analysis of variance (ANOVA), followed by a set of eight orthogonal planned comparisons. The ANOVA revealed that the differences in the observers' judgments across the different display conditions were highly significant, $F(45, 220) = 4.57$, $p < .001$, and the error term from this one-way analysis was used as the denominator for each of the individual planned comparisons. Three of these turned out to be significant, and those three accounted for approximately 87% of the total treatment variance.

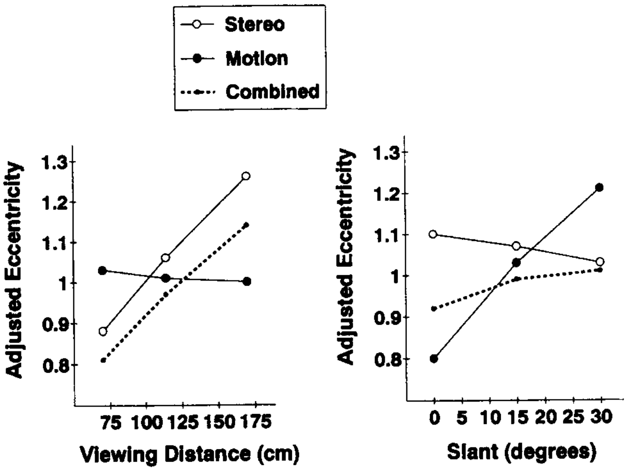

Figure 4. Average adjusted cylinder eccentricity for the stereo, motion, and combined conditions as a function of simulated viewing distance and surface slant. An adjusted eccentricity of 1.0 indicates veridical performance.

As can be seen in Figure 4, there were large effects of convergence distance for the static stereo displays, $F(2, 220) = 18.24$, $p < .01$, which resulted in an overestimation of depth at the near distance (the average adjusted cylinder was 0.88), a slight underestimation at the intermediate distance (1.06), and a large underestimation at the far distance (1.26). A significant effect of convergence distance was also obtained for the stereo–motion combined displays, $F(2, 220) = 26.8$, $p < .01$, with a similar overall pattern of overestimation at the near distance (0.81), a small overestimation at the intermediate distance (0.97), and underestimation at the far distance (1.14). The variation of surface slant did not have a significant effect on the observers' judgments in either of these stereoscopic conditions.

One possible interpretation for the effects of convergence distance is that the perceptual distortions occurred because convergence and accommodation were not always consistent in our experimental configuration. However, note that Johnston (1991) manipulated distance by changing the physical distance between the observer and the CRT display (thus equating convergence and accommodation) and found the same pattern of results reported here: Judged depth from binocular stereopsis is overestimated at near distances and underestimated at far distances. Furthermore, this same pattern of results also has been reported with judgments of real objects viewed in a natural environment under full illumination (Baird & Biersdorf, 1967; Loomis et al., 1992). In that case, clearly there would be ample opportunity for accommodation to play a role in relative distance perception, and yet the results are consistent with those reported here.

For the monocular motion displays, the pattern of results was much different from those reported earlier for the stereo and combined conditions, in that the observers' adjustments were significantly influenced by changes in surface slant, $F(2, 220) = 42.5, p < .01$. Note that in Figure 4 the judged depth of the motion displays was systematically overestimated when the average slant was in the frontoparallel plane (0.79), slightly underestimated when the average slant was 15° (1.03), and underestimated by a large amount when the average slant was 30° (1.22). These motion-alone displays also contrasted with the stereoscopic conditions in that there were no effects of convergence distance.

The finding that judged depth from binocular stereopsis was overestimated at near distances and underestimated at far distances is not surprising and has been reported by other researchers (e.g., Baird & Biersdorf, 1967; Johnston, 1991). However, our observation that judged 3-D shape from motion was overestimated for rotations near frontoparallel (0° slant) and underestimated for rotations with a 30° average slant has not previously been reported, to our knowledge.

This result is particularly interesting because of the presence of perspective information in our displays. Previous investigations of structure from motion have suggested that the perception of 3-D structure in orthographic displays is based on those properties that remain invariant under an affine transformation (Norman & Todd, 1993; Todd & Bressan, 1990; Todd & Norman, 1991). These findings are consistent with the idea the visual system uses only the first-order (i.e., velocity) information available in a motion sequence to specify 3-D structure up to a one-parameter family (Bennett, Hoffman, Nicola, & Prakash, 1989; Koenderink & van Doorn, 1991; Liter et al., 1993; Todd & Bressan, 1990; Ullman, 1983). However, there are two special cases in which 3-D structure can be uniquely recovered from only two frames of a motion sequence: (a) rotations confined to a single plane (Hoffman & Flinchbaugh, 1982); and (b) rotations viewed under a strong polar perspective (Longuet-Higgins, 1981). The perspective information identified in Longuet-Higgins's analysis causes the image of a square to undergo a trapezoidal deformation as it rotates out of the frontoparallel plane, which is analogous to the role of vertical disparities in binocular stereopsis. This type of perspective information is not available in translations perpendicular to the line of sight, and thus its role has not been investigated in the numerous studies of depth perception from motion parallax (e.g., Braunstein & Tittle, 1988; Caudek & Proffitt, 1993; Rogers & Graham, 1979).

Evidence for a lack of shape constancy in perspective rotation displays has been reported by Loomis and Eby (1988). However, it is difficult to assess fully the role of perspective in specifying 3-D structure from their study *because of the presence of deforming occluding contours* and relatively low resolution (320 × 200) of the displays. The displays used in this experiment, by contrast, had a spatial resolution of 1,280 × 512 and were anti-aliased so that the effective resolution was even higher. Thus, the systematic distortions in the observers' perceptions, especially at the near viewing distance (where the effects of perspective were highest), provide strong evidence that for objects subtending 11.8° of visual angle, observers are unable to make use of the vertical components of disparity or motion to determine an object's euclidean metric structure.

To summarize, the results of this experiment indicate that visual information from motion and binocular stereopsis produces distinctly different patterns of perceptual deformation. For motion displays, the perceived eccentricity of a circular cylinder decreases with its slant relative to the line of sight but is unaffected by viewing distance. However, for stereoscopic displays, the perceived eccentricity decreases with viewing distance but is unaffected by slant. When motion and stereo are presented in combination, the results are similar to those obtained from stereo alone.

Experiment 2

The overall pattern of results for the cylinder adjustment task indicates that when observers have to equate depth intervals in nonparallel directions, their judgments are neither accurate nor constant. These findings support the idea that physical and perceptual space are not related by a similarity transformation (which would preserve distance ratios in nonparallel directions but not overall size). A conformal transformation, by contrast, need not preserve lengths in nonparallel directions, but it does preserve local angles. Thus, in Experiment 2, we examined whether physical and perceptual space would be related by a conformal transformation by having observers adjust the magnitude of a dihedral angle to 90°. If local angles are preserved in the relation between physical and perceptual space, we would expect dihedral angle judgments to be constant across variations in convergence distance and average slant.

Method

Participants. The same 6 observers from Experiment 1 participated in this experiment.

Design. The design was identical to that used in Experiment 1.

Stimuli and apparatus. The apparatus was identical to that used in the first experiment. However, for Experiment 2, our stimuli consisted of 7,500 bright dots projected onto the surface of two slanted planes meeting at the center of the display to form a dihedral angle (see Figure 5). The dots were distributed on the surface such that they had a uniform distribution on the display CRT for a frontoparallel view of the dihedral angle. Unlike in the previous experiment, the top and bottom edges of the surface were not visible; thus, the dihedral angle displays had both a constant height (13.3°) and width (6.3°) inside the region of matched points from the left and right views of the stereo pair. Once again, the shape of the surface was presented by motion (vertical axis rotation), binocular stereopsis, and motion–stereo combined (see Figure 6 for a sample stereogram from the frontoparallel slant condition). Also, as in the first experiment, the shape of the surface, the

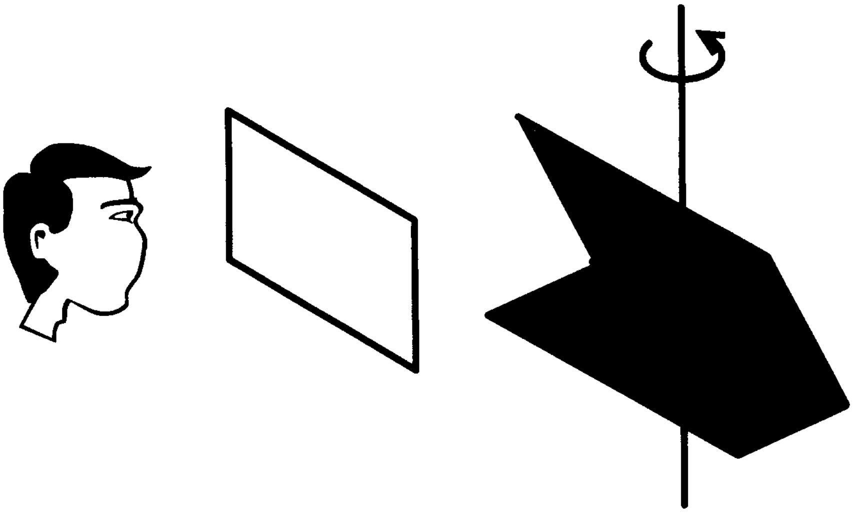

Figure 5. A schematic view of the dihedral angle stimuli used in Experiment 2.

angle between the top and bottom planes, was manipulated by adjusting the mouse.

For the dihedral angles in a frontoparallel orientation (i.e., those with zero slant), all of the horizontal disparities were crossed (see the Appendix for an analytic description of the disparity and motion fields). When adjusted veridically, they varied with position over a range from 0 to −15 min at the farthest convergence distance of 170 cm and from 0 to −36 min at the nearest convergence distance of 70 cm. These ranges were increased when the objects were slanted in depth. For example, at a 30° slant the horizontal disparities ranged from −21 min (crossed) to 5 min (uncrossed) at the far distance and from −52 to 11 min at the near distance. Again, these largest disparities occurred only in the most peripheral regions of the displays, and observers had no difficulty fusing the patterns into a single coherent object. The maximum vertical disparities were relatively unaffected by changes in slant, although they did decrease significantly with increasing convergence distance from approximately 2 to 0.8 min.

The horizontal and vertical extents of moving elements in the image (over the entire 10° rotation) were qualitatively similar to the pattern of disparities, but once again were larger in magnitude. For example, when the surfaces were viewed at a convergence

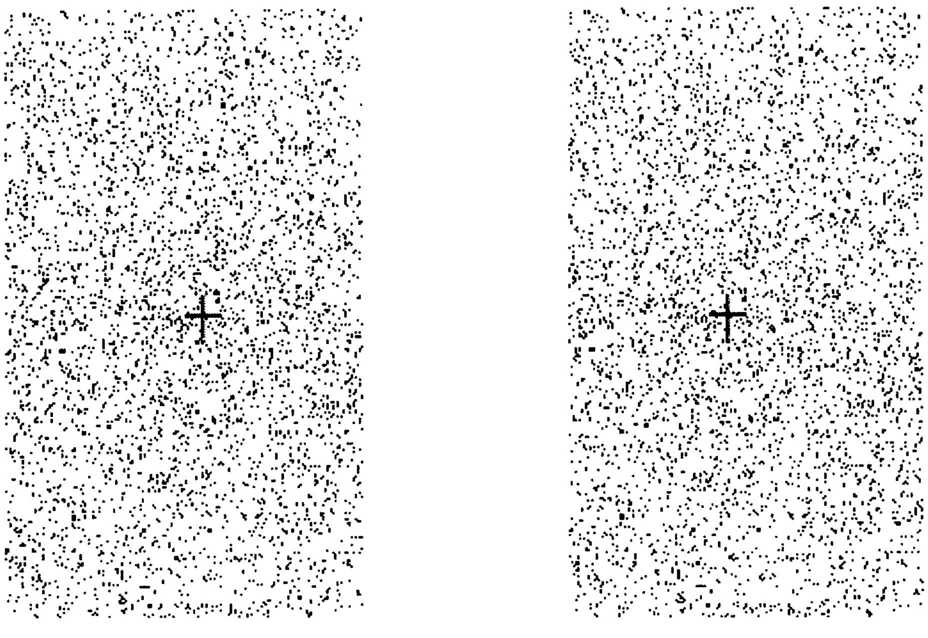

Figure 6. A sample stereogram of a frontoparallel dihedral angle similar to those used in Experiment 2. Observers were required to adjust the upper and lower surfaces so that they appeared to form a right angle.

distance of 70 cm and an average slant of 30°, the displacement of dots in the image varied with position over a range of −15 to 106 min. The maximum vertical displacements attributable to perspective were approximately 4 min for all of the displays (i.e., they did not vary with average slant or convergence distance).

Procedure. The procedure was identical to that used in the first experiment, except that observers used the mouse to adjust the magnitude of a dihedral angle until it appeared to be 90°.

Results and Discussion

The observers' subjective impressions of this task were similar to those they reported for the cylinder adjustments of Experiment 1. That is, although their adjustments produced perceptually compelling changes in the 3-D structure of the depicted surfaces, they did not feel confident in the precise setting that should correspond to a right angle. This uncertainty was again reflected in the overall reliability of their judgments, which were highly variable, even when a given observer judged the same stimulus over multiple occasions. For the stereo and combined displays, the average standard deviation (as a proportion of the mean) for the 10 repeated adjustments in each individual condition was approximately 14%, whereas in the motion-alone condition it was approximately 20%. These values were similar for all 6 observers, and there were no systematic differences as a function of slant or distance.

The average adjusted dihedral angle magnitudes collapsed over observers are presented in Figure 7. Keep in mind that the data in this figure are plotted as the magnitude of the adjusted dihedral angle rather than the magnitude of adjusted eccentricity as in Figure 4 and that an adjustment of 180° indicates the perception of a perfectly flat surface. Because the observers' task in this experiment was to adjust the angle until it appeared to be 90° (standard), adjusted angular values greater than the standard represent overestimation of depth and those less than the standard represent underestimation.

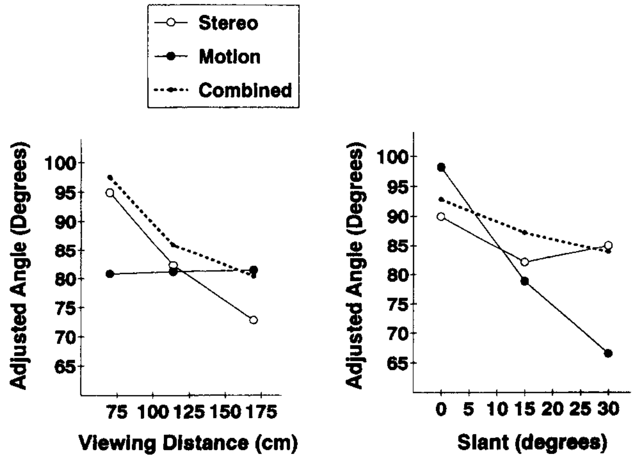

Figure 7. Adjusted dihedral angle as a function of surface slant and simulated viewing distance. An adjusted angle of 90° indicates veridical performance.

As in Experiment 1, the data were analyzed using a one-way ANOVA with eight orthogonal planned comparisons. The ANOVA revealed that the differences in the observers' judgments across the different display conditions were highly significant, $F(45, 220) = 17.26, p < .001$, and the error term from this one-way analysis was used as the denominator for each of the individual planned comparisons. Five of these comparisons were significant, and they accounted for more than 90% of the total treatment variance.

As in the first experiment, there were significant effects of convergence distance for both the stereo, $F(2, 220) = 48.9, p < .01$, and the combined stereo–motion, $F(2, 220) = 59.2, p < .01$, conditions, such that the overall depth of the displays was systematically overestimated at the near convergence distance and systematically underestimated at the far distance. The variations of surface slant had no detectable influences on the observers' judgments in the static stereo condition, although there was a small significant effect of slant in the combined condition, $F(2, 220) = 15.18, p < .01$, which accounted for 4% of the total treatment variance.

Although there were no effects of convergence distance in the motion-alone condition, there were large effects of slant, $F(2, 220) = 203.4, p < .01$, similar to those obtained in Experiment 1. Judged depth was systematically overestimated when the surfaces were oriented at a 0° slant and decreased progressively as the slant was increased to 15° and 30°. Unlike in the first experiment, the effect of average velocity was significant, $F(1, 220) = 30.23, p < .01$, but it accounted for only 4% of the treatment variance, as opposed to more than 53% that resulted from variations in slant.

The results from Experiment 2 provide strong evidence that the transformation between physical space and perceptual space does not preserve local angles and is therefore not conformal, because the preservation of local angles is the defining characteristic of a conformal transformation. One potential criticism of this conclusion that needs to be considered concerns the adequacy of the available information for making the required adjustments. Although motion and binocular disparity are known to be compelling sources of information about an object's 3-D form, they are certainly not the only sources of information available in natural vision. For example, when objects are depicted in pictorial displays such as stereograms, there are no gradients of accommodative blur as typically occur when real objects are viewed directly (see Buckley & Frisby, 1993). Similarly, the displays used in the current experiments did not depict an intervening ground surface, which could, under more natural conditions, provide additional information about egocentric distance that could be used to more accurately scale binocular disparities.

Is it possible that the systematic distortions of perceived 3-D form revealed by Experiments 1 and 2 were simply an artifact of having limited sources of information and that real objects were perceived veridically when observed in a sufficiently rich environment? To address this question, consider some of the classic research on distortions of perceptual space that was conducted in the 1950s and 1960s.

Although much of this research was concerned with the structure of stereoscopic space under reduced cue conditions (see Foley, 1980), there is also a substantial body of literature from that period on the perceived metrical structure of real surfaces in a natural environment under full-cue conditions.

When observers are asked to compare the magnitudes of physically identical length intervals at different distances, the overall pattern of results depends critically on how those intervals are oriented in space. For lengths that are aligned in the frontoparallel plane, there is a general tendency under full-cue conditions to overcompensate for the optical effects of perspective, such that perceived size increases with viewing distance (e.g., Baird & Biersdorf, 1967; Carlson, 1962; Epstein, 1963; Gilinsky, 1955; Holway & Boring, 1941). A much different pattern of results is obtained, however, if the intervals to be compared are aligned in depth. The tendency in that case is for perceived size to become systematically compressed with increasing viewing distance (e.g., Baird & Biersdorf, 1967; Gilinsky, 1951; Gogel, 1960, 1964; Harway, 1963; Loomis et al., 1992; Toye, 1986; Wagner, 1985).

These effects are measurable over surprisingly small regions of visual space. For example, Baird and Biersdorf (1967) measured the constancy of perceived length intervals oriented in depth or in the frontoparallel plane over a range of viewing distances from 0.5 to 5.0 m, using real objects placed on a table top with normal indoor illumination. The perception of frontal size increased over this range of distances by 14%, whereas the perceived intervals in depth decreased by 21%. In a related experiment by Loomis et al. (1992), observers were presented with a horizontal target interval marked on the ground in an open field and were required to identify an interval in depth at the same viewing distance whose length was perceptually identical to that of the target. Over a range of viewing distances from 4 to 12 m, the perceived compression of intervals in depth relative to the horizontal increased from 30% to 50%. At distances beyond 20 m, Wagner (1985) showed that this compression remained constant at approximately 50% (see also Harway, 1963), such that the mapping between physical and perceived space for intermediate-to-far distances was governed by a homogeneous affine transformation.

The key thing to note in considering these earlier investigations is that the same basic pattern of perceptual distortion obtained in the current experiments can also be observed when real objects are viewed directly in a natural environment. Although the relative magnitude of these effects may vary depending on the particular types of optical information that are available in any given context, the qualitative form of this distortion is highly similar over a wide range of stereoscopic conditions. That is, there is a systematic compression of perceived intervals in depth in near visual space that reaches an asymptote at intermediate viewing distances. A much different pattern of distortion occurs, however, when objects are viewed monocularly in motion. In that case, there is a systematic expansion of perceived intervals in depth relative to those in the frontoparallel plane (see also Braunstein, Liter, & Tittle, 1993; Todd & Bressan, 1990; Todd & Norman, 1991), and the relative spatial extents in these different directions appear to remain invariant over changes in viewing distance. Whether this pattern persists with additional monocular sources of information such as shading, texture, or accommodative blur is an interesting question that remains unanswered.

Experiment 3

The perceived relative extents of physically identical length intervals in different orientations is a particularly important property for categorizing the mapping from physical to perceived space from the perspective of a Kleinian analysis. If perceived extents in different directions remain in fixed proportions to one another within a given region of space, then the perceptual mapping in that region can be categorized as locally affine. The available evidence suggests that this is the case for objects viewed monocularly in motion or for objects viewed stereoscopically at sufficiently long distances. If, on the other hand, perceived relative extents in different directions change with position, as appears to occur in near stereoscopic space, then the perceptual mapping must involve a more complicated nonaffine transformation.

One simple way of evaluating whether this condition is satisfied is to use a surface orientation matching task. In regions of space where the mapping is affine, physically planar surfaces also should appear perceptually planar. However, if the perceived extents in different directions undergo a detectable change with position, then physically planar surfaces should be perceived as curved, and, conversely, surfaces that are perceived as planar should be physically curved. Experiment 3 was designed to test this property for moving and stationary surfaces viewed both monocularly and stereoscopically.

Method

Participants. The participants for this experiment were the four of us.

Design. Three independent variables were manipulated: (a) simulated distance (70, 114, or 170 cm); (b) stimulus display type (monocular motion, static binocular stereopsis, or combined motion and stereopsis); and (c) angular velocity (1.0° or 1.5° per frame). This last manipulation could be applied only to the motion and combined display types; thus, there were a total of 15 conditions in this experiment.

Stimuli and apparatus. The apparatus was identical to that used in the first 2 experiments. The stimuli were similar to those used in Experiment 2, subtending the same visual angle in both the vertical and horizontal directions, except now the observers' task was to adjust the angle between the two planes until it was 180°. Each display contained two planar surfaces with a shared horizontal edge at the center of the display screen, which, when adjusted to be coplanar, had a vertical slant of 20° (see the top of Figure 8). To ensure that the observers could not perform the task solely on the basis of the presence of a visible orientation discontinuity, a dark occluding region was added to the center of the display,

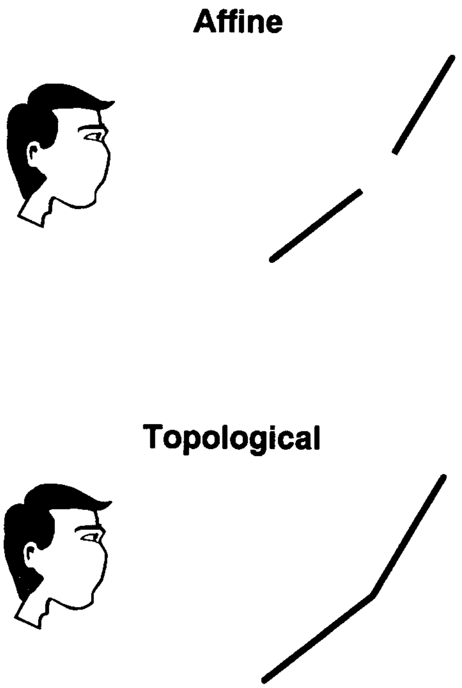

Figure 8. A schematic view of the stimuli used for the affine task in Experiment 3 and the topological task in Experiment 4.

whose vertical extent subtended 1.3° of visual angle. We also wanted to be sure that observers could not use texture density cues to match the orientation of the top and bottom surfaces. Thus, the two planes were populated independently with a different number of dots that varied randomly between 5,000 and 10,000 (see Figure 9 for a sample stereogram).

When adjusted correctly, the overall pattern of horizontal disparities ranged from -5 min (crossed) to 6 min (uncrossed) at the far convergence distance and from -13 to 16 min at the near convergence distance. The maximum vertical disparities at those distances were 0.8 min and 2 min, respectively. For the motion displays, the total horizontal displacements of dots in the image ranged from -23 to 27 min, whereas their maximum vertical displacements attributable to perspective were approximately 4 min.

Procedure. The experimental layout and general procedure were similar to those used in the first 2 experiments. The motion, stereo, and motion–stereo combined conditions were presented in separate blocks that were completed on different days. Each block contained 10 replications of all combinations of the distance and velocity (for the motion and combined conditions) variables. This led to a total of 150 trials for each observer: 60 for each of the motion and combined motion–stereo blocks and 30 for the static stereo block.

Results and Discussion

In describing their subjective impressions of this task, all of the observers noted that they felt much more confident in adjusting a surface to be coplanar than they did for the cylinder or dihedral angle adjustments of Experiments 1 and 2. Their judgments were also more reliable. The average standard deviation for the 10 repeated adjustments in each individual condition was 3° for the motion-alone displays, 2.4° for the static stereo displays, and 1.4° for the combined stereo–motion displays.

As in the first 2 experiments, the means of these adjustments were analyzed using a one-way ANOVA with a set of planned orthogonal comparisons. The average adjusted angles between the two planes in the different conditions are shown in Figure 10, in which a value of 180° represents perfect performance. The adjusted angle for the motion condition was not only constant over variations in simulated distance, but was also close to veridical. As in the first 2 experiments, there were no significant effects of simulated distance or angular velocity in the motion-alone condition. Note in Figure 10, however, that the adjusted angle for both the stereo-alone and stereo–motion conditions was less than 180° for the near convergence distances and approached veridicality only at the far convergence distance. Although these effects of viewing distance were relatively small, they were statistically significant for both the static stereo, $F(2, 42) = 7.17, p < .01$, and the combined stereo–motion, $F(2, 42) = 3.42, p < .05$, conditions. This finding indicates that for displays containing binocular disparities, observers consistently saw concave surfaces as if they were planar.

It is important to remember that when comparing the results of Experiments 1–3, the different procedures were designed to explore distinctly different aspects of the systematic distortions between physical and perceived space. In Experiments 1 and 2, it was possible to assess perceived relative extents in different directions for objects presented at any given viewing distance or orientation. Experiment 3 was designed, by contrast, to measure the perception of affine structure. To determine whether a surface is planar, it does not matter if the relative extents in different directions are perceived veridically. All that is required is that they stay in fixed proportion to one another within an appropriately large neighborhood of visual space. If not, then a physically planar surface will appear curved, which is exactly what occurred when the stereoscopic surfaces were observed at close viewing distances.

These findings are consistent with other empirical results reported in the literature. The evidence suggests that the mapping between physical and perceived space is affine for objects viewed monocularly in motion (e.g., Todd & Bressan, 1990; Todd & Norman, 1991) or for objects viewed stereoscopically at sufficiently long distances (e.g., Wagner, 1985) but that the mapping involves a more complicated nonaffine transformation when objects are viewed stereoscopically in near visual space (e.g., Baird & Biersdorf, 1967; Harway, 1963).

Experiment 4

An important methodological property of the displays used in Experiment 3 was that there was no visible edge at the intersection of the two planes the observers were required to adjust. In the absence of this control, it would have been possible to perform the task on the basis of the perception of the surface topology rather than its affine struc-

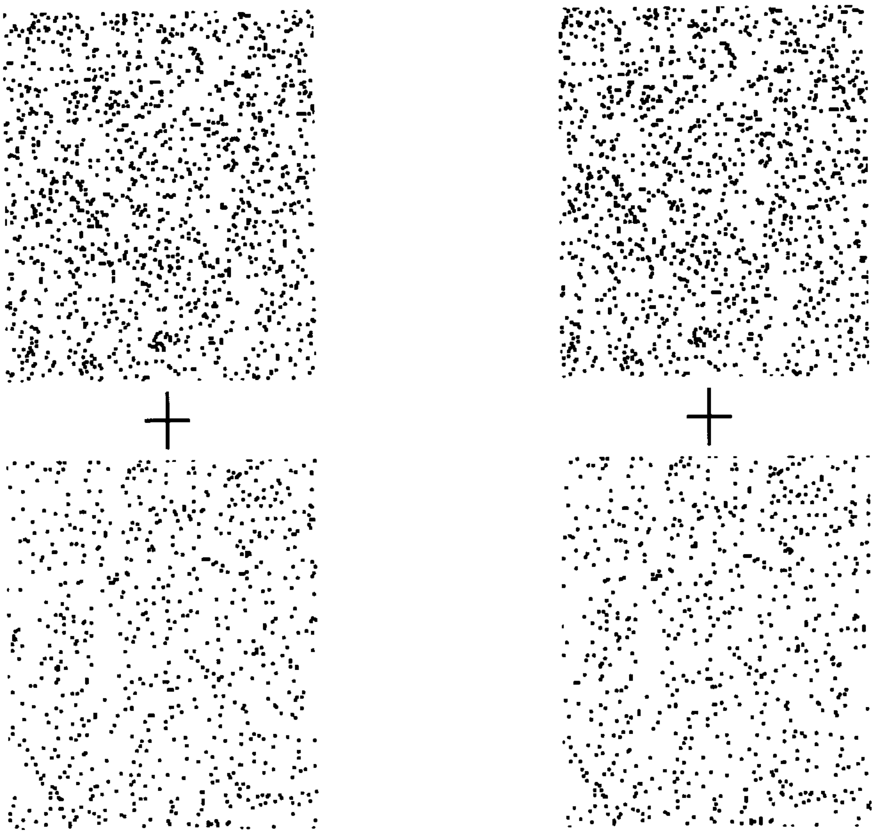

Figure 9. A sample stereogram similar to those used for the affine task in Experiment 3. Observers were required to adjust the angle between the upper and lower surfaces so that they appeared coplanar.

ture by adjusting the two planes until there was no visible discontinuity at their shared border. This would still be possible even though the surface might appear somewhat curved at the precise level of adjustment at which it appeared perfectly smooth. Experiment 4 was designed to test the relative perceptual salience of this topological structure by repeating the procedure of Experiment 3 without occluding the intersecting edge.

Method

The methods in this experiment were identical to those used in Experiment 3 in every way, except that the central occluding strip was removed so that observers could view the intersection of the top and bottom planes (see Figure 11 for a sample stereogram).

Results and Discussion

Experiment 4 required the information preserved from only the most general type of relationship between physical and perceptual space, because observers could perform accurately simply by adjusting the two planes until there was no orientation discontinuity between them. Under these conditions, observers exhibited slightly better performance than that obtained in Experiment 3. The average adjusted angles in each condition are shown in Figure 12 (once again, 180° indicates a veridical adjustment). A one-way ANOVA with orthogonal planned comparisons revealed a pattern of results similar to those found in Experiment 3. For the motion-alone displays, the observers' judgments were highly accurate, and there were no detectable effects of convergence distance. There was a small systematic effect of viewing distance for the static stereo condition, $F(2, 220) = 7.6$, $p < .01$, but this effect did not reach significance for the combined motion–stereo condition. In comparing these results with those shown in Figure 10, it is clear that the presence of an intersecting edge facilitated performance.

All of the observers noticed while performing this task in the stereo and combined conditions that the two criteria for making the adjustment were not in perfect agreement with one another. That is, when the displays were adjusted appropriately so that the visible discontinuity was removed, the overall surface could appear slightly curved rather than coplanar. This effect was particularly noticeable at near viewing distances.

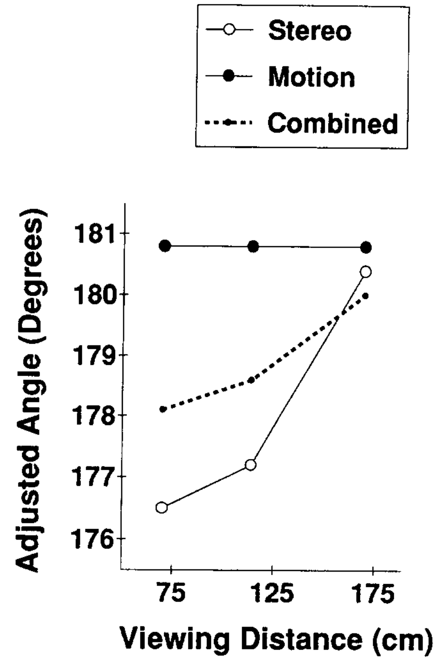

Figure 10. Adjusted angle as a function of viewing distance. An adjusted angle of 180° indicates veridical performance.

General Discussion

The research described in this article was designed to investigate the geometric relationship between physical and perceived space. The experimental design was motivated in part by Felix Klein's insight that any geometric transformation will always leave at least some of an object's properties unchanged and that different types of transformations can be categorized hierarchically by the specific configurations of properties they leave invariant. Klein's analysis provides a convenient method for identifying the mapping between physical and perceived space by measuring observers' judgments to a variety of different object properties. Depending on the particular transformation by which these spaces are formally related, some object properties should appear systematically distorted, whereas others that are invariant over this transformation should be perceived accurately. In our experiments, we examined four different types of object properties requiring an analysis of 3-D structure up to an arbitrary similarity transformation in Experiment 1, up to a conformal transformation in Experiment 2, up to an affine transformation in Experiment 3, and up to a continuous topological transformation in Experiment 4.

On the basis of previous investigations of the perception of 3-D structure from orthographic projections of rotation (Norman & Todd, 1993; Todd & Bressan, 1990; Todd & Norman, 1991), it might be tempting to conclude that the geometric transformation between physical and perceived space is essentially affine. A closer examination of the available evidence reveals, however, that although this hypothesis is well supported for the perception of structure from motion, it cannot explain certain perceptual distortions that are known to exist in stereopsis.

To better understand the nature of these inconsistencies, it is useful to consider how orientation and depth constancy would be affected by affine distortions of visual space. Suppose, for example, that an observer underestimates distances in depth by a factor of 2 relative to distances in the frontoparallel plane (e.g., Toye, 1986; Wagner, 1985) and is asked to judge the perceived 3-D structure of an object in different positions and orientations. Note that the perceived structure of the object in this context would be invariant under arbitrary rigid translations but would vary dramatically with changes in orientation. In other words, an affine distortion of perceptual space would produce good depth constancy, but not orientation constancy.

When the stimuli in our experiments were presented with motion alone, the results did indeed conform to this general pattern. The errors in the observers' judgments were consistent with a systematic affine expansion of perceptual space in depth, as has been reported previously by Todd and Bressan (1990), Todd and Norman (1991), and Braunstein et al., (1993); however, Loomis and Eby (1988) obtained systematic compressions in depth. When considered as a whole, these findings provide strong support for an affine analysis of perceived structure from motion.

A much different pattern of results was obtained, however, when the displays were presented with stereo alone or when motion and stereo were presented in combination. For all of the different response tasks used in these experiments, the effects of orientation in the stereo-alone or stereo–motion conditions were negligible, but there were large effects of viewing distance, the opposite of what would be expected if the mapping between physical and perceived space were affine. Although these results contrast sharply with those obtained from motion alone, they are highly consistent with previous reports of distortions of stereoscopic space. The expansion of depth at near distances and the contraction of depth at far distances has been reported previously by Johnston (1991) using computer-generated stereograms and by numerous other investigators for real objects viewed under full illumination in a natural environment (e.g., Baird & Biersdorf, 1967; Gilinsky, 1951; Gogel, 1960, 1964; Harway, 1963; Loomis et al., 1992; Toye, 1986; Wagner, 1985). Similarly, the misperception of affine structure exhibited in Experiments 3 and 4 is consistent with the classic literature on distance bisection, the apparent frontoparallel plane, and the adjustment of parallel alleys (see Foley, 1980, for an excellent review).

There are two obvious conclusions that follow from these observations: First, they suggest that the processes for determining 3-D structure from motion and stereo are much different, in that they are differentially affected by changes in orientation or viewing distance. Second, they also indicate that distortions of stereoscopic space are more complex than can easily be accounted for by a simple affine model. The precise geometric nature of these distortions remains to be determined, but there is some evidence to suggest that their nonaffine components may be confined to relatively close viewing distances or that they may be attenuated by

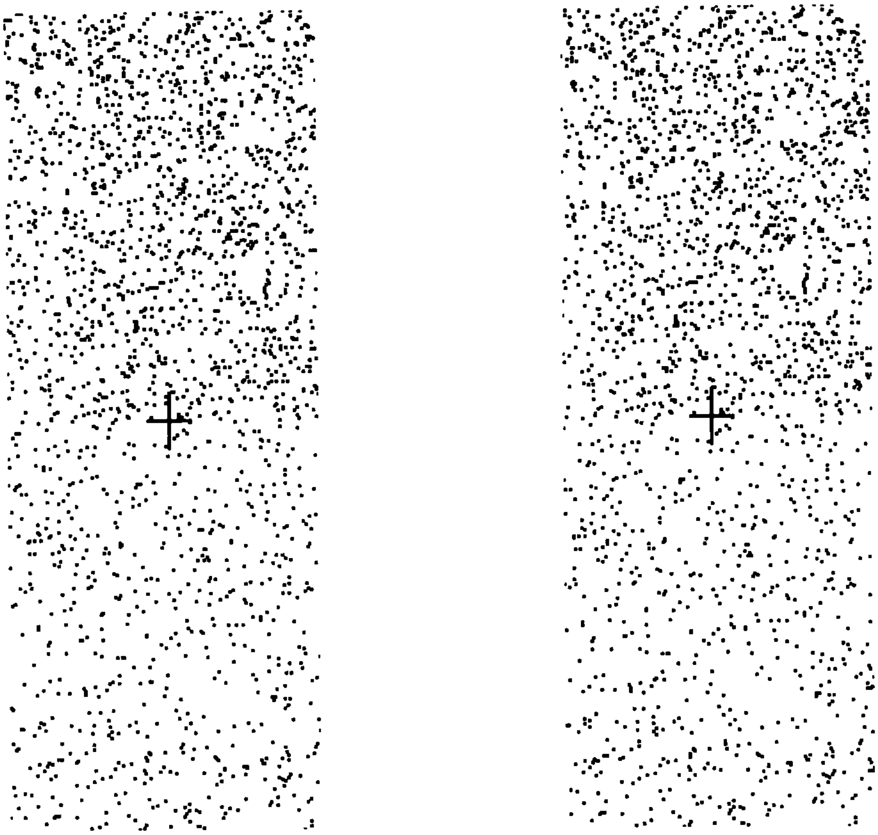

Figure 11. A sample stereogram similar to those used for the topological task in Experiment 4. Observers were required to adjust the angle between the upper and lower surfaces so that they appeared coplanar.

other sources of optical information. In a study by Wagner (1985), for example, observers were asked to judge the distances and angles between wooden stakes that were viewed directly in a large open field (see also Toye, 1986). The results revealed a constant affine compression in depth by a factor of 2 relative to distances in the frontoparallel plane, and there was no evidence to indicate the systematic bending of visual space that has been reported previously for judgments of stereoscopic stimuli in the absence of other sources of information.

All of the discussion of perceptual space presented thus far has been primarily focused on its extrinsic structure relative to the physical environment. From an extrinsic viewpoint, the structure of perceptual space is determined by its formal relation to physical space. Thus, we could conclude that perceptual space is affine, for example, if the perceived 3-D structure of objects is systematically distorted by an affine transformation.

It also is possible, however, to consider the intrinsic structure of perceptual space without making any comparisons to the corresponding structure of the external environment. Suppose, for example that an observer is asked to estimate the three angles in a triangle. If the intrinsic structure of perceptual space is euclidean, then the sum of these three angles should always equal 180°; if it is elliptic, then the sum should be greater than 180°; and if it is hyperbolic, then the sum should be less than 180°. There have been several experiments reported in the literature on the intrinsic structure of stereoscopic space, and the results have generally indicated that its structure is hyperbolic (e.g., Blank, 1961; Foley, 1972).

It is especially interesting to note in this regard that it is mathematically possible for the intrinsic structure of perceptual space to be euclidean, whereas the relationship between physical and perceived space is noneuclidean. Indeed, that is a model that has recently been proposed by Wagner (1985). According to this view, our perception of 3-D form could involve an explicit representation of the euclidean distances and angles between visible points and lines, but these quantities need not be represented accurately.

There is clearly some truth in Wagner's (1985) hypothesis. The fact that observers can understand instructions about judging lengths and angles in 3-D space provides compelling evidence that these concepts exist at some level of representation, but are they the primitive units of people's perceptual awareness, or do they only exist as generalized abstractions at a more cognitive level of analysis? There are

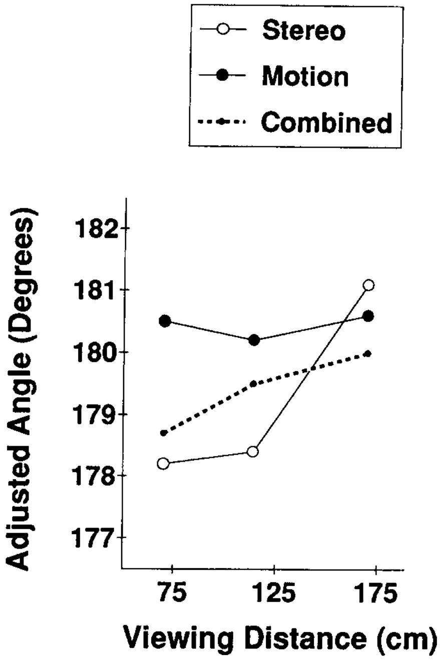

Figure 12. Adjusted angle as a function of viewing distance. An adjusted angle of 180° indicates veridical performance.

several sources of evidence to suggest that the latter of these characterizations may be the most appropriate. One important result that supports this view is the observers' lack of confidence when performing tasks such as cylinder or dihedral angle adjustments. Several observers in our experiments expressed concern that they did not have a clear idea of how a circular cylinder or a 90° dihedral angle should appear at different orientations in depth. This was in striking contrast to their high levels of confidence on the planar adjustment tasks of Experiments 3 and 4. A distorted representation of euclidean properties could explain variations in accuracy for these tasks, but it would not explain why the tasks should vary in their subjective difficulty.

There is another subtle aspect of Wagner's (1985) hypothesis that is noteworthy. Assume that people's perceptual representations of physical space are systematically distorted, as has been indicated by almost all of the existing research on this topic performed over the past century. If euclidean properties are the primitive components of people's perceptual representations and are systematically distorted relative to physical space, then objects in the world should appear to distort nonrigidly as they move relative to the observer. If it is the case, for example, that apparent distances in depth relative to the horizontal can be compressed by a factor of 2, as reported by Wagner, then visible objects should appear to expand and contract by that amount as they rotate in depth. The fact that this does not occur allows only two possible explanations: Either people's perceptual representations of physical space are veridical, which is not supported by the available psychophysical data, or their immediate awareness of rigid motion cannot be based on an explicit representation of euclidean properties.

We believe that the latter of these two hypotheses is the most plausible given the large body of evidence on distortions of perceptual space. According to this view, people's perceptual representations of 3-D form are based primarily on ordinal or topological properties. If required to make judgments about euclidean properties, observers may adopt ad hoc heuristics to satisfy the task demands, but these tasks will typically produce large systematic distortions of judged 3-D structure. It is our contention that most of the perceptual judgments required in natural vision (e.g., object recognition) do not require an explicit knowledge of euclidean metric structure and can be performed accurately on the basis of ordinal or topological relations.

References

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

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

Blank, A. A. (1953). The Luneburg theory of binocular visual space. *Journal of the Optical Society of America, 43,* 717–727.

Blank, A. A. (1958). Analysis of experiments in binocular space perception. *Journal of the Optical Society of America, 48,* 911–925.

Blank, A. A. (1961). Curvature of binocular visual space: An experiment. *Journal of the Optical Society of America, 51,* 335–339.

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.

Braunstein, M. L., & Tittle, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 582–590.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research, 33,* 919–934.

Carlson, V. R. (1962). Size constancy adjustments and perceptual compromise. *Journal of Experimental Psychology, 63,* 68–73.

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

Epstein, W. (1963). Attitudes of judgment and the size-distance invariance hypothesis. *American Journal of Psychology, 76,* 257–265.

Foley, J. (1972). The size-distance relation and intrinsic geometry of visual space: Implications for processing. *Vision Research, 12,* 323–332.

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

Fox, R., Cormack, L., & Norman, J. F. (1987). The effect of vertical disparity on depth scaling. *Investigative Ophthalmology and Visual Science, 28*(Suppl. 4), 293.

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

Gilinsky, A. S. (1955). The effect of attitude on the perception of size. *American Journal of Psychology, 68,* 173–192.

Gogel, W. C. (1960). The perception of a depth interval with binocular disparity cues. *Journal of Psychology, 50,* 257–269.

Gogel, W. C. (1964). The visual perception of spatial extent. *Journal of the Optical Society of America, 54,* 411–416.

Harway, N. I. (1963). Judgment of distance in children and adults. *Journal of Experimental Psychology, 65,* 385–390.

Hoffman, D., & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics, 42,* 195–204.

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology, 54,* 21–37.

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

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America, A8,* 377–385.

Lappin, J. S., & Fuqua, M. A. (1983). Accurate visual measurement of three-dimensional moving patterns. *Science, 221,* 480–482.

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multi-view displays. *Perception, 22,* 1441–1465.

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

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 From the Second International Conference on Computer Vision* (pp. 383–391). Washington, DC: Institute of Electrical and Electronics Engineers.

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

Luneburg, R. K. (1950). The metric of binocular visual space. *Journal of the Optical Society of America, 40,* 627–642.

Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature, 297,* 376–378.

Norman, J. F., & Lappin, J. S. (1992). The detection of surfaces defined by optical motion. *Perception & Psychophysics, 51,* 386–396.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 3,* 279–291.

Ogle, K. N. (1952). On the limits of stereoscopic vision. *Journal of Experimental Psychology, 44,* 252–259.

Purdy, J., & Gibson, E. J. (1955). Distance judgment by the method of fractionation. *Journal of Experimental Psychology, 50,* 374–380.

Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America, A2,* 343–349.

Rogers, B. J., & Bradshaw, M. F. (1992). Differential perspective effects in binocular stereopsis and motion parallax. *Investigative Ophthalmology and Visual Science, 34*(Suppl. 4), 1333.

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

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8,* 125–134.

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., Akerstrom, R. A., Reichel, F. D., & Hayes, W. (1988). Apparent rotation in 3-dimensional space: Effects of temporal, spatial and structural factors. *Perception & Psychophysics, 43,* 179–188.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 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.

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

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

Ullman, S. (1983). Computational studies in the interpretation of structure and motion: Summary and extension. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 459–480). San Diego, CA: Academic Press.

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

(*Appendix follows on next page*)

Appendix

Computing the Field of Projected Image Velocities or Binocular Disparities

Consider a point p in a Cartesian coordinate system (x, y, z) where the x-axis is aligned horizontally, the y-axis is aligned vertically, and the z-axis is aligned in depth relative to the point of observation at the origin. Let p rotate continuously in depth at an angular velocity ω about a vertical axis, which is a distance z_0 from the origin along the z-axis. The instantaneous position of p at each moment in time t can be defined by the following parametric equations:

$$x = r\sin(\omega t) \tag{1}$$

$$z - z_0 = r\cos(\omega t), \tag{2}$$

where r is the euclidean distance of the point from the axis of rotation. If we differentiate these equations with respect to time and perform appropriate rearrangements and substitutions, we obtain the following:

$$\frac{dx}{dt} = \omega(z - z_0) \tag{3}$$

$$\frac{dy}{dt} = -\omega x. \tag{4}$$

Now consider the polar projection p' of this point onto a planar surface with a Cartesian coordinate system (x', y') at a unit distance from the point of observation. The projective mapping from p to p' is given by

$$x' = \frac{x}{z} \tag{5}$$

$$y' = \frac{y}{z}. \tag{6}$$

Note that the coordinates x' and y' in this projection are the tangents of the horizontal and vertical visual angles. If we again differentiate and perform appropriate rearrangements and substitutions, we obtain the instantaneous field of projected velocities in both the horizontal and vertical directions:

$$\frac{dx'}{dt} = \omega\left(1 - \frac{z_0}{z} + x'^2\right) \tag{7}$$

$$\frac{dy'}{dt} = \omega x'y'. \tag{8}$$

Note that the vertical component of velocity provides potential optical information about the object's angular velocity, which could, in principle, be used to scale the horizontal components to compute its veridical three-dimensional structure (see Longuet-Higgins, 1981).

The depicted objects in our experiments were all Monge surfaces of the form $z = f(x, y)$, which can also be expressed in image coordinates as $z = f(x', y')$. For the dihedral angle stimuli, the flow fields are computed most easily from the image coordinates (x', y'). The horizontal image velocity components for the dihedral angles in Experiments 2 and 3, respectively, are described in Equations 9 and 10:

$$\frac{dx'}{dt} = \omega\left[x'^2 - \frac{|y'|}{\cos(\alpha)\tan(\phi)} - x'\tan(\alpha)\right] \tag{9}$$

$$\frac{dx'}{dt} = \omega\left[x'^2 + \frac{y'\tan\left[90^\circ - \frac{|y'|}{y'}(\phi - 20^\circ)\right]}{\cos(\alpha)} - x'\tan(\alpha)\right], \tag{10}$$

where α is the instantaneous slant of the dihedral edge and ϕ is one half the adjusted dihedral angle. This technique is less convenient for the cylindrical surfaces used in Experiment 1, because the projected vertical extents of those surfaces varied with slant and viewing distance. The overall range of projected velocities in that case is most easily determined by computing the horizontal components along the x-axis (i.e., where $y' = 0$), as described in Equation 11:

$$\frac{dx'}{dt} = \omega[x'^2 - x'\tan(\alpha)] \tag{11}$$

and the horizontal and vertical components along the top or bottom cylinder occlusion boundary, as described in Equations 12 and 13:

$$\frac{dx'}{dt} = \omega\left\{1 + x'^2 - \frac{z_0[\cos(\alpha) + x'\sin(\alpha)]}{z_0\cos(\alpha) + \epsilon R}\right\} \tag{12}$$

$$\frac{dy'}{dt} = \frac{\omega x'R[\cos(\alpha) + x'\sin(\alpha)]}{[z_0\cos(\alpha) + \epsilon R]}, \tag{13}$$

where R is one half the vertical height of the cylinder (i.e., 5.5 cm), α is its instantaneous slant about the vertical axis of rotation, and ϵ is the adjusted eccentricity.

Remember that, when using this analysis, image flow field components specified in Equations 8–13 are the instantaneous temporal derivatives for objects undergoing continuous motion about a vertical axis. However, these same equations can also provide a good approximation for the field of binocular disparities or the flow fields produced by objects in apparent motion, provided that the angular displacements in each case are sufficiently small. To perform the calculations for discrete rotations, the parameter ω should be interpreted as the convergence angle for binocular disparity or the extent of angular rotation for motion.

Received December 13, 1993
Revision received June 14, 1994
Accepted July 14, 1994 ■