Effects of the amount of monocular shape information on stereo scaling problem

HIROAKI SHIGEMASU* *Kochi University of Technology*
TAKAO SATO *University of Tokyo*

Abstract: The perception of 3-D shape from binocular disparity depends on the estimation of viewing distance. Therefore, in the absence of good distance information, an observer misestimates the viewing distance and perceives a wrong depth for an object. While motion depth cues can theoretically overcome this problem, past studies have reported inconsistent results. We examined whether differences in the amount of monocular shape information between the stimuli used in these past studies can resolve the discrepancy. The amount of information was manipulated by changing the presence of object boundaries, dot density, and rotation angles. The results indicate the effect of the amount of monocular shape information and thus suggest its significance in accounting for the discrepancies among past studies. It should also be noted that shape factors interact with viewing distance. Motion cues cannot correct distortion at near viewing distance. This suggests that the cue that gives the larger depth dominates in determining perceived shape.

Key words: depth, cue integration, stereopsis, structure-from-motion, stereo scaling problem.

The veridical depth perception from binocular disparity requires information about viewing distance (the absolute distance between the observer and the object) and interocular separation. Therefore, in the absence of good distance information, an observer misestimates viewing distance and perceives wrong depth (Foley, 1980). The perceived 3-D shape of an object, which is defined as the ratio between depth and height, also becomes distorted when the viewing distance is misestimated, because disparities scale approximately inversely with the square of the viewing distance, while the size (height and width) of the object scales inversely with the viewing distance (Foley, 1980; Johnston, 1991). In contrast, the perceived shape from motion is not affected by the estimation of the viewing distance because depth from relative motion scales linearly with distance.

When both stereo and motion cues are combined, the information of the veridical shape is obtained from motion cue and the estimation of viewing distance can be derived by selecting the distance for which the shape from stereo are in agreement with the veridical shape from motion cue. From the information of shape and viewing distance, veridical metric value of size and depth can be calculated. Therefore, theoretically, when both stereo and motion cues are provided, these cues compensate the information from each other and the stereo scaling problem should be solved (Johnston, Cumming, & Landy, 1994; Landy, Maloney, Johnston, & Young, 1995; Richards, 1985).

However, past studies have reported inconsistent results. Several studies support this hypothesis (Bradshaw, Parton, & Eagle, 1998; Johnston et al., 1994; Richards & Lieberman, 1985) but others do not (Bradshaw, Frisby, & Mayhew, 1987; Brenner & Landy, 1999; Scarfe & Hibbard, 2006; Tittle, Todd, Perotti, & Norman, 1995; Todd & Norman, 2003). A clear contrast was seen between studies by Johnston et al. (1994) and by Tittle et al. (1995), where this issue was examined using similar methods (a shape judgment task of hemi-cylinders), but inconsistent results were reported. Johnston et al. (1994) found no distortion of perceived shape for combined cues, but distortions for the perceived shape from stereo alone. In contrast, Tittle et al. (1995) found distortions of perceived shape for combined cues as well as for shape from stereo alone. We suspect this inconsistency might be from the differences between their stimuli. First, the hemi-cylindrical objects used by Johnston et al. (1994) had object boundaries, whereas Tittle et al. (1995) eliminated the boundaries as a source of shape information. Second, the surface of the stimuli was textured with small spherical patches, which generated clear texture cues such as compression in Johnston et al. (1994), while Tittle et al. (1995) used random dot stereograms with minimal texture cues. Third, the rotation angle of the cylindrical object that provided the motion shape information was

16–30 deg in Johnston et al. (1994) but 10 deg in Tittle et al. (1995). In short, the stimuli used by Johnston et al. (1994) had richer monocular shape cue information than those of Tittle et al. (1995). It is likely that, with the stimuli of Tittle et al. (1995), the information from the monocular shape cues was insufficient to overcome the stereo scaling problem. In this study, we investigated the effects of the amount of monocular shape cue information by manipulating the presence of the object boundary, the number of dots which composed the random dot stereograms, and the rotation angle.

In the present study, we controlled the cues for the estimation of viewing distance. There were cues to perceive veridical viewing distance such as vergence angle and retinal size in past studies. The results of Johnston et al. (1994) should not necessarily be used to conclude that motion information overcame the stereo scaling problem at a near viewing distance, as there was no specific tendency of distortion of perceived shape from stereo cue alone in the first place. We suspect that there was enough information to obtain the veridical distance perception in the condition and it made the shape perception veridical. In the hypothesis of Richards (1985), the perceived shape with motion and stereo cues should be veridical even when there is no viewing distance information. Therefore, to control the viewing distance information, we set the vergence angle of participants to be zero (i.e., the visual axes of both eyes are parallel), although the binocular disparities of the objects were generated with a simulated vergence angle of each viewing distance condition. We also did not change the retinal size of the stimuli manipulating the simulated physical size so as to make the retinal size identical regardless of simulated distance.

Experiment 1 Effects of object boundary

Object boundary is a strong cue by which to perceive the 3-D shape of an object (e.g., Cortese & Andersen, 1991; Koenderink, 1984; Koenderink, van Doorn, Kappers, & Todd, 1997). To avoid contaminations from the boundary cue, Tittle et al. (1995) eliminated object boundaries by simulating displays viewing through a rectangular aperture (Figure 1). In contrast, Johnston et al. (1994) had no consideration on this point, and the stimuli had object boundaries. This difference might cause the inconsistent results between these two studies. In Experiment 1, we examined whether the presence and absence of object boundaries influenced the perceived shape when stereo and motion cues were combined.

Methods

Participants. Four participants, including one of the authors, participated in the experiment. Three of the participants were myopic and wore their optical correction to perform the experiments. All participants were well trained and had good stereopsis. All participants except one of the authors were unaware of the specific aims of the experiment.

Stimuli and apparatus. The stimuli were horizontally oriented cylinders that were composed of bright random dots on a dark background. The position of the dots was determined using the ray-tracing method. An anti-aliasing method was used to produce the subpixel positioning. The highest luminance was 38.3 cd/m^2 and the luminance of the background was less than 1 cd/m^2. The stimuli were generated on a personal computer and presented on a 20-in. CRT monitor (Sony Multiscan 20se) which had a spatial resolution of 1024 × 768 pixels with a temporal refresh rate of 120 Hz.

There were three depth cue conditions (static binocular stereopsis, monocular motion rotating about the vertical axis, and combined motion and stereopsis) and two object boundary conditions (with and without object boundaries). The simulated viewing distances were 57, 114, and 171 cm. The stimuli oscillated 30 deg (±15 deg from a frontoparallel orientation) in the motion and the combined cue conditions. The speed of rotation was randomized from 30 to 60 deg/s. The object boundaries were eliminated using the same way as Tittle et al. (1995) in the no boundary condition. The stimuli subtended 4 deg horizontally and 3 deg vertically, both in the visual angle, when the stimuli were in a frontoparallel orientation. The dot density on the surface of the stimuli was 3%. The dots were uniformly distributed for a frontoparallel view of the surfaces with orthogonal projection. The dot density is defined in terms of this orthogonal projection. The vergence angle of the participant was set to 0 deg regardless of the simulated viewing distance condition and the optical distance was fixed to 114 cm. Thus, the term "viewing distance” in our experiment does not refer to the optical distance but the simulated distance. Stimuli in different viewing distance conditions varied in both horizontal and vertical disparity and texture gradient generated by perspective projection. As the retinal size of the stimuli was small, the vertical disparity had little or no effect on the perceived shape of the cylindrical surfaces (Rogers & Bradshaw, 1995). The experiments were conducted in a completely dark room. A chin-rest was used to minimize head movement and the participants viewed the stimuli through a mirror stereoscope. No fixation point was provided.

Procedure. The participants adjusted the elongation in the depth axis of the cylindrical stimuli using a key press until the cross-section of the stimuli appeared circular (apparently circular cylinder (ACC) task, Johnston 1991). In other words, the participants adjusted the apparent depth of the stimuli equal to one-half of the height of the stimuli. The relative disparities between the center and the upper/lower edge of the simulated circular cylinders were 9.5, 4.7, and 3.1 min for the 57, 114, and 171 cm viewing distance conditions, respectively.

The effects of three variables were examined: (a) depth cue types (static binocular stereopsis, monocular motion, or combined motion and stereopsis); (b) object boundary conditions (object boundary present or not); and (c) viewing distances (57, 114, or 171 cm). Thus there were 18 conditions in this experiment. In the monocular motion cue condition, stimuli were presented to only one of the eyes. Each participant performed 12 trials for each condition and the total number of trials for each participant was 216. The stimuli were presented until the participants made a judgment.

Results

As all participants showed the same tendency, the average data are presented in Figure 2. The left panel shows the results for the boundary present conditions and the right panel shows those for the no boundary conditions. In each panel, the ratio between depth and half-height was plotted as a function of viewing distance. This ratio is an index of 3-D shape perception. As this ratio is inversely related to perceived depth, the values less than one indicate an overestimation of depth and the values greater than one indicate an underestimation of depth.

The perceived shape from stereo alone varied depending on the simulated viewing distance. The perceived shape was overestimated when the simulated distance was near (57 cm) and underestimated when the simulated distance was far (171 cm). These results seem to indicate that the visual system used a “default” viewing distance when there was little cue for distance (Gogel, 1972) and this default value did not change greatly between distance conditions. There was little effect of vergence angle on estimating viewing distances, because the perceived shape was not overestimated much even though vergence was kept at 0 deg, which, by itself, indicates infinite viewing distance. The perceived shape from motion alone was slightly overestimated but close to veridical at all distances.

The perceived shape in the combined cue condition, which is the main focus of this study, showed different results from the stereo and motion cue alone conditions, suggesting effects of the interaction of both cues. The results also showed that the perceived depth was larger when the object boundary was present than not present at 114 cm and 171 cm, suggesting effects of the presence of the object boundary. The perceived shape at 57 cm, however, was overestimated irrespective of the presence of the object boundaries. The amount of overestimation was similar to that found in the stereo alone condition.

The data were analyzed using repeated measures of three-way analysis of variance (ANOVA). The analysis revealed significant effects of depth cue type, $F(2, 6) = 29.77$, $p < .01$, viewing distance, $F(2, 6) = 127.03$, $p < .01$, but no significant effect of object boundary. The analysis also revealed significant effects of two-way interaction between depth cue type and viewing distance, $F(4, 12) = 105.33$, $p < .01$, and three-way interaction among depth cue type, object boundary, and viewing distance, $F(2, 6) = 29.77$, $p < .01$, but no other interaction effects.

The separated simple main effects analysis of depth cue type revealed that when object boundaries were absent and the viewing distance was 171 cm, the perceived shape with combined cues was significantly different from that with motion cue alone, $p < .05$. In contrast, the analysis revealed that when object boundaries were present and the viewing distance was 114 cm or 171 cm, the perceived shape with combined cues was significantly different from that with the stereo cue alone, $p < .05$. The analysis also revealed that when object boundaries were present and the viewing distance was 57 cm, the perceived shape with combined cues was significantly different from that with the motion cue alone, $p < .05$. The separated simple main effects analysis of object boundaries revealed that when the viewing distance was 171 cm and the combined cue was present, the effect of object boundaries was significant, $p < .05$. There was no significant effect of object boundaries in other conditions.

Discussion

When object boundaries were absent, the perceived shape from the combined cue at the far (171 cm) viewing distance was not equal to that obtained from motion alone. This tendency is consistent with the results of Tittle et al. (1995). In contrast, when object boundaries were present, the perceived shape at the far viewing distance was nearly equal to that from motion alone. These results seem to indicate that the shape from motion overcame the depth perception from stereo cue and was consistent with the results of Johnston et al. (1994). As the results of both Tittle et al. (1995) and Johnston et al. (1994) were obtained depending on whether object boundary was present or not, the inconsistency between these two studies should be attributed to, at least partially, the presence or absence of object boundaries. There was no significant effect of object boundaries in the stereo alone and motion alone conditions. As the cylindrical stimuli were in a frontoparallel orientation in the stereo alone condition, the shape cue of the object boundaries was weak and there appeared to be no significant effect. As the perceived shape from the motion cue was nearly veridical even if object boundaries were absent, there appeared to be no significant effect.

It should be noted that even when the object boundary was present, the perceived shape from the combined cue was overestimated and nearly equal to that for stereo alone at the near viewing distance. Thus, at this distance, the motion cue was not enough to readjust the depth determined by the stereo cue. These results are inconsistent with the hypothesis of Richards (1985) and the results of Johnston et al. (1994).

Experiment 2 Effects of dot density

The results in Experiment 1 revealed that object boundaries influenced shape perception from the combined cue. However, at the nearest viewing distance, the perceived shape was not affected by the presence of object boundaries and was nearly equal to that from stereo alone with distortions. Thus, at the nearest viewing distance, we did not obtain results that were consistent with those of Johnston et al. (1994), even when the object boundaries were present. In Experiment 1, we used 3% dot density, which was close to that used by Tittle et al. (1995). In contrast, the stimuli used by Johnston et al. (1994) were textured with small spherical patches and had clear object boundaries. Therefore, the distorted perceived shape from the combined cue at the near viewing distance may have resulted from weak object boundaries generated by the sparse dot density used in Experiment 1. When the dot density increased, the object boundary became clearer, and perceived shape from motion may overcome the perception of overestimated depth by stereo cue at the near viewing distance. We investigated this point by using the stimuli with an increased dot density.

We also investigated whether the shape from motion with increased dot density could overcome the stereo scaling problem even in the no object boundary condition. As the dot density becomes higher, motion energy becomes larger and the amount of motion information increases. Although higher dot density also increases information from stereo cues, the change of dot density may provide enough monocular shape information and cause the perceived shape with the combined cue to be more veridical even in the no object boundary condition.

Methods

The same four participants from Experiment 1 participated in this experiment. The stimulus and the apparatus were identical to those used in Experiment 1, except that the dot density of the stimuli was increased to 10%. The participants' task was the ACC task, as in Experiment 1.

Results

The average data from four participants are presented in Figure 3. The left panel shows the results for the boundary present conditions and the right panel shows those for the no boundary conditions. The perceived shape from stereo alone and motion alone showed similar tendencies as for Experiment 1. That is, the perceived shape from stereo alone was affected by the simulated viewing distance while the perceived shape from motion alone was close to veridical and not affected by the simulated viewing distance. In contrast to the results of Experiment 1, the perceived shape from the combined cue did not change depending on the presence or absence of object boundaries. In both boundary conditions, the perceived shape at 171 cm was nearly equal to that from motion alone and the perceived shape at 57 cm was nearly equal to that from stereo alone.

The data were analyzed using repeated measures of three-way ANOVA. The analysis revealed significant effects of depth cue type, $F(2, 6) = 6.42$, $p < .05$, and viewing distance, $F(2, 6) = 286.92$, $p < .01$, but no significant effect of object boundary. The analysis also revealed a significant effect of two-way interaction between depth cue type and viewing distance, $F(4, 12) = 46.68$, $p < .01$, but no other two-way interaction nor a three-way interaction among depth cue type, object boundary, and viewing distance. The separated simple main effects analysis of object boundaries showed no significant effect.

We also analyzed the data of the perceived shape from the combined cue in both Experiment 1 and 2 to investigate the effect of the dot density. The analysis of three-way ANOVA revealed significant effects of viewing distance, $F(2, 6) = 61.14$, $p < .01$, and a ,three-way interaction among dot density, depth cue type and viewing distance, $F(2, 6) = 19.86$, but no significant effect of dot density or object boundary, nor any two-way interaction effects. The separated simple main effects analysis of dot density revealed that when the viewing distance was 171 cm and object boundaries were not present, there was a significant effect of dot density, $p < .05$.

Discussion

In Experiment 2, the perceived shape from the combined cue at 57 cm was nearly equal to that from the stereo cue alone, regardless of the presence or absence of object boundaries. Similar results had been found in Experiment 1, and these indicate that the dot density had no effect on shape perception at near viewing distances. Thus, even when the object boundary was clearer, the perceived shape at the near viewing distance was still determined using the stereo cue. The most striking difference from the results of Experiment 1 was found at the far viewing distance with no object boundaries. Here, the perceived shape was nearly equal to that from motion alone. This result might arise from larger motion energy and an increased amount of motion information with increased dot density.

The results at the far viewing distance with object boundary or with higher dot density are similar to those of Johnston et al. (1994) and the results without object boundaries and with lower dot density are similar to those of Tittle et al. (1995). The results thus indicate that the difference of the amount of the monocular shape cue information can account for the discrepancy between these two studies. However, the perceived shape from the combined cue at the nearest viewing distance was nearly equal to that from stereo alone even with relatively rich monocular shape information. These results are inconsistent with those of Johnston et al. (1994), as well as the theoretical analysis by Richards (1985). The abundance of monocular shape cue information cannot account for the discrepancy between these two studies at near viewing distances. We will discuss this point further in the general discussion section.

Experiment 3 Effects of rotation angle

In Experiment 3, we examined the inconsistency between Johnston et al. (1994) and Tittle et al. (1995) in terms of the rotation angle. The stimuli used in Johnston et al. (1994) rotated 16–30 deg while the stimuli used in Tittle et al. (1995) rotated 10 deg. For a smaller rotation angle, the motion information should be poorer, and the perceived shape from motion is more susceptible to noise (Eagle & Blake, 1995; Hogervorst & Eagle, 1998). Thus the difference in the amount of motion information caused by the different rotation angles could be a source of inconsistency. The rotation angle of the stimuli in Experiments 1 and 2 was 30 deg. In this experiment, we decreased the rotation angle down to 10 deg (±5 deg from a frontoparallel orientation), which was equal to that used by Tittle et al. (1995). If the rotation angle is one of the reasons for the inconsistency between the two studies, the results for the combined cue condition in this experiment should be close to the result of Tittle et al. (1995). Although the decreasing rotation angle causes less information from the stereo cue as well as the motion cue, if the combined cue condition shows that the stereo scaling problem is not overcome and the effect of the stereo cue becomes larger, the results should occur because of insufficient information from the motion cue.

Methods

The same four participants from Experiments 1 and 2 participated in Experiment 3. The rotation angle of the stimuli was 10 deg. Object boundaries were not eliminated. Other methods were identical to those for Experiment 2. There were two variables: (a) depth cue types (static binocular stereopsis, monocular motion, or combined motion and stereopsis); and (b) viewing distances (57, 114, or 171 cm). The stimuli for the static stereopsis condition were identical to those in Experiment 2. There were nine conditions in this experiment and the participants performed 12 trials in each condition. The total number of trials for each participant was 108.

Results

The average data from four participants are presented in Figure 4. Similar to previous experiments, the perceived shape from stereo alone changed depending on the viewing distance, and the perceived shape from motion alone was close to veridical regardless of the viewing distance. The perceived shape from the combined cue at the far viewing distance was underestimated compared with the results of Experiment 2, in which the rotation angle was larger. In Experiment 2, the perceived shape from the combined cue was nearly equal to that from motion alone at 171 cm, while the perceived shape from the combined cue in Experiment 3, at 171 cm, was underestimated than that from motion alone.

The data in this experiment and the data from the object boundary present condition in Experiment 2 were analyzed using repeated measures of three-way ANOVA. The ANOVA revealed significant effects of viewing distance, $F(2, 6) = 219.99$, $p < .01$, a two-way interaction between the rotation angle and the depth cue type, $F(2, 6) = 15.04$, $p < .01$, and a two-way interaction between the depth cue type and the viewing distance, $F(4, 12) = 52.47$, $p < .01$, but no significant effects of two-way interaction between the rotation angle and the viewing distance, nor a three-way interaction among the rotation angle, the depth cue type, and the viewing distance.

Discussion

When the rotation angle is small (10 deg), the perceived shape from the combined cue at the far viewing distance was not equal to that from motion alone. This tendency is consistent with the results of Tittle et al. (1995). In contrast, when the rotation angle is large (30 deg) as in Experiment 2, the perceived shape at the far viewing distance was nearly equal to that from motion alone, which is consistent with the results of Johnston et al. (1994). As the results from bothTittle et al. (1995) and Johnston et al. (1994) were obtained depending on the extent of the rotation angle, the inconsistency between these two studies should also be accounted for by the rotation angle.

General discussion

The present results revealed that one of the reasons for the inconsistency among previous studies about the interaction of stereo and motion cues is the difference in the amount of information about the monocular shape. When the information about the monocular shape was insufficient, the perceived shape from the combined cue was not equal to that from motion alone. When it was sufficient, the perceived shape from the combined cue was equal to that from motion alone at the far viewing distance. From these results, it can be said that the amount of information about the monocular shape in the stimuli of Tittle et al. (1995) was not sufficient, so that the results they produced were inconsistent with those of Johnston et al. (1994).

However, it should be noted that the perceived shape from the combined cue at the near viewing distance is similar to that from stereo alone even when the information about the monocular shape is abundant. This cannot be explained by the hypothesis of Johnston et al. (1994), that when the stereo and motion cues are combined the veridical shape is perceived using the motion cue. To account for this point, we propose a new hypothesis, which is that the perceived depth of an object is determined using the cue which shows the larger depth estimation of the object when multiple cues are combined. The depth cues which are obtained from the visual scene change from situation to situation and are not necessarily sufficient in some cases (Gepshtein & Banks, 2003; Jacobs, 2002). If the depth is preferentially perceived using the cue which gives larger depth with sufficient information and is not influenced by the cue which gives smaller depth, the constant and veridical depth can be perceived. This aspect of our hypothesis is close to the modified weak fusion model (Landy et al., 1995), which suggests that the weights of the cues are malleable to the reliability and availability of each cue (Hillis, Watt, Landy, & Banks, 2004; Landy et al., 1995). As the perceived shape from the stereo cue at the far viewing distance is underestimated because of the stereo scaling problem, becoming less vulnerable to such a misestimation is effective for perceiving the veridical depth. The present results also indicate that when the information of the cue that provides larger depth is insufficient, such as the motion cue with no object boundary or a small rotation angle, the weight of the motion cue becomes less and the perceived depth is not equal to that from the motion cue alone, which shows larger depth. If the motion information is sufficient, the weight of the stereo cue which causes underestimation becomes zero. Thus, the criteria of sufficiency of the depth cue in this study can be defined whether the cue that gives the larger depth determines the depth perception of an object.

From this hypothesis, the perceived depth would be erroneously determined using the stereo cue, which overestimates the depth derived from the stereo scaling problem at the near viewing distance. However, in the real world, the amount of information for viewing distance is much different from this study, in which there are very few cues for viewing distance. The perceived depth from the stereo cue will be veridical with various viewing distance information at a near viewing distance in the real world. For example, the veridical cue of the vergence angle can be obtained in the real world. The vergence angle is a strong cue and when viewing distance is near the estimation is accurate (Mon-Williams & Tresilian, 2000). Also, in the experiment of Johnston et al. (1994), the viewing distance was physically changed and observers could obtain much of the viewing distance information, such as the vergence angle, accommodation, and retinal size, resulting in veridical shape perception from the stereo cue alone. Thus, at the near viewing distance, it cannot exactly be said that the motion information resolved the stereo scaling problem. Those results in Johnston et al. (1994) can be explained by our hypothesis that the cue which provides the larger depth determines the final depth perception.

The perceived depth at the 114 cm viewing distance with the combined cue was larger than the motion or the stereo cue alone condition if the monocular shape cue is abundant, although there was no significant effect of depth cue type in this viewing distance condition. This cannot be explained by our hypothesis. To explain this result, we propose the additional hypothesis that if the perceived depth with each cue is similar and the cue information is abundant, the effect of each depth cue will be additive rather than averaged.

In the hypothesis of Richards (1985) and Johnston et al. (1994), there is an assumption that shape from motion is unique and veridical. However, as the perceived shape with the combined cue was not equal to that with the motion cue alone when the viewing distance was near or when the information about the monocular shape is poor, these results indicate that shape from motion is unreliable (Norman & Todd, 1993; Tittle et al., 1995; Todd & Bressan, 1990; Todd & Norman, 1991). Thus, when the perceived depth from the stereo cue is larger at a near viewing distance or when the monocular shape information is poor, the motion cue does not overcome the stereo scaling problem, but the shape from motion is conversely influenced by the stereo cue.

In conclusion, when the stereo and motion cues are integrated, the perceived depth depends on the amount of monocular shape information and it is determined by the cue which gives larger depth, at least with an abundant amount of cue information and with poor viewing distance information. In the studies of depth cue interaction, there are many studies that report inconsistent results. These factors could explain the inconsistency in these previous studies.

References

Bradshaw, M. F., Frisby, J. P., & Mayhew, J. E. (1987). The recovery of structure from motion: No evidence for a special link with the convergent disparity mechanism. *Perception*, **16**, 351–357.

Bradshaw, M. F., Parton, A. D., & Eagle, R. A. (1998). The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. *Perception*, **27**, 1317-1331.

Brenner, E., & Landy, M. S. (1999). Interaction between the perceived shape of two objects. Vision Research, 39, 3834–3848.

Cortese, J. M., & Andersen, G. J. (1991). Recovery of 3-D shape from deforming contours. *Perception & Psychophysics*, **49**, 315-327.

Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion. *Vision Research*, **35**, 2927-2941.

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

Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, **13**, 483-488.

Gogel, W. C. (1972). Scalar perceptions with binocular cues of distance. *American Journal of Psychology*, **85**, 477-497.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, **4**, 967–992.

Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 1587-1593.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, **6**, 345-350.

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

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, **34**, 2259-2275.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, **13**, 321-330.

Koenderink, J. J., van Doorn A. J., Kappers, A. M., & Todd, J. T. (1997). The visual contour in depth. *Perception & Psychophysics*, **59**, 828-838.

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

Mon-Williams, M., & Tresilian, J. R. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research*, **133**, 407-413.

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.

Richards, W., & Lieberman, H. (1985). Correlation between stereo ability and the recovery of structure from motion. *American Journal of Optometry and Physiological Optics*, **62**, 111–118.

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

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity Scaling and the Perception of Frontoparallel Surfaces. *Perception*, **24**, 155-179.

Scarfe, P., & Hibbard, P. B. (2006). Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy, *Vision Research*, **46**, 1599-1610.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 663-678.

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.

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

(Received May 6, 2011; accepted November 5, 2011)

Figure 1 A schematic view of the cylinder stimuli in which the object boundary was eliminated.

Figure 2 Depth/half height ratio for cylinders which appeared circular: (a) with object boundary, and (b) no object boundary. The rotation angle was 30 deg in the motion and the combined cue conditions. The dot density was 3%. Error bars indicate standard deviations.

Figure 3 Depth/half height ratio for cylinders which appeared circular: (a) with object boundary, and (b) no object boundary. The rotation angle was 30 deg in the motion and the combined cue conditions. The dot density was 10%. Error bars indicate standard deviations.

Figure 4 Depth/half height ratio for cylinders which appeared circular. The object boundary was present. The rotation angle was 10 deg in the motion and the combined cue conditions. The dot density was 10 %. Error bars indicate standard deviations.

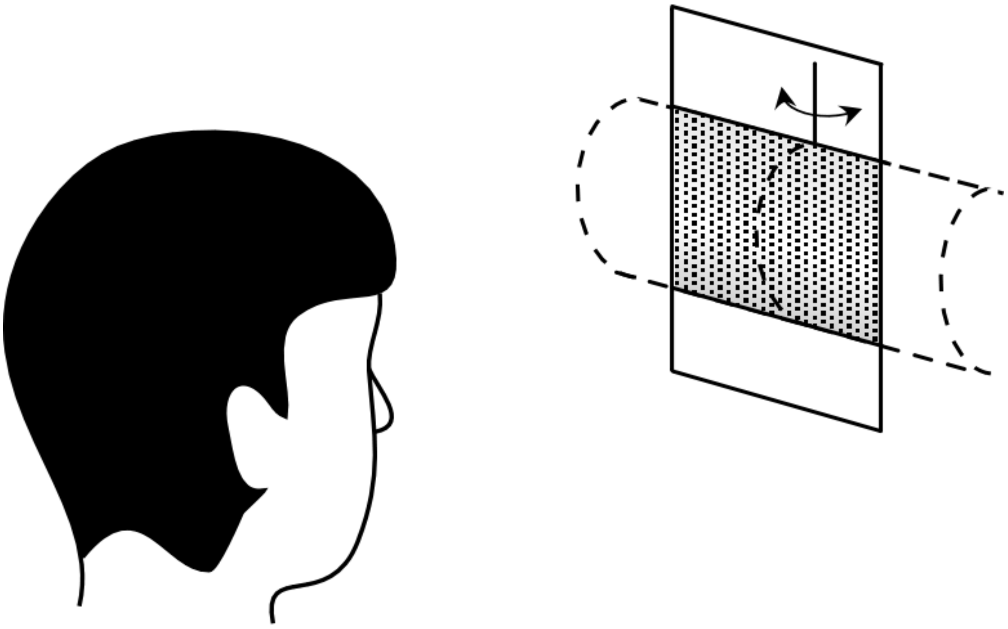

Figure 1

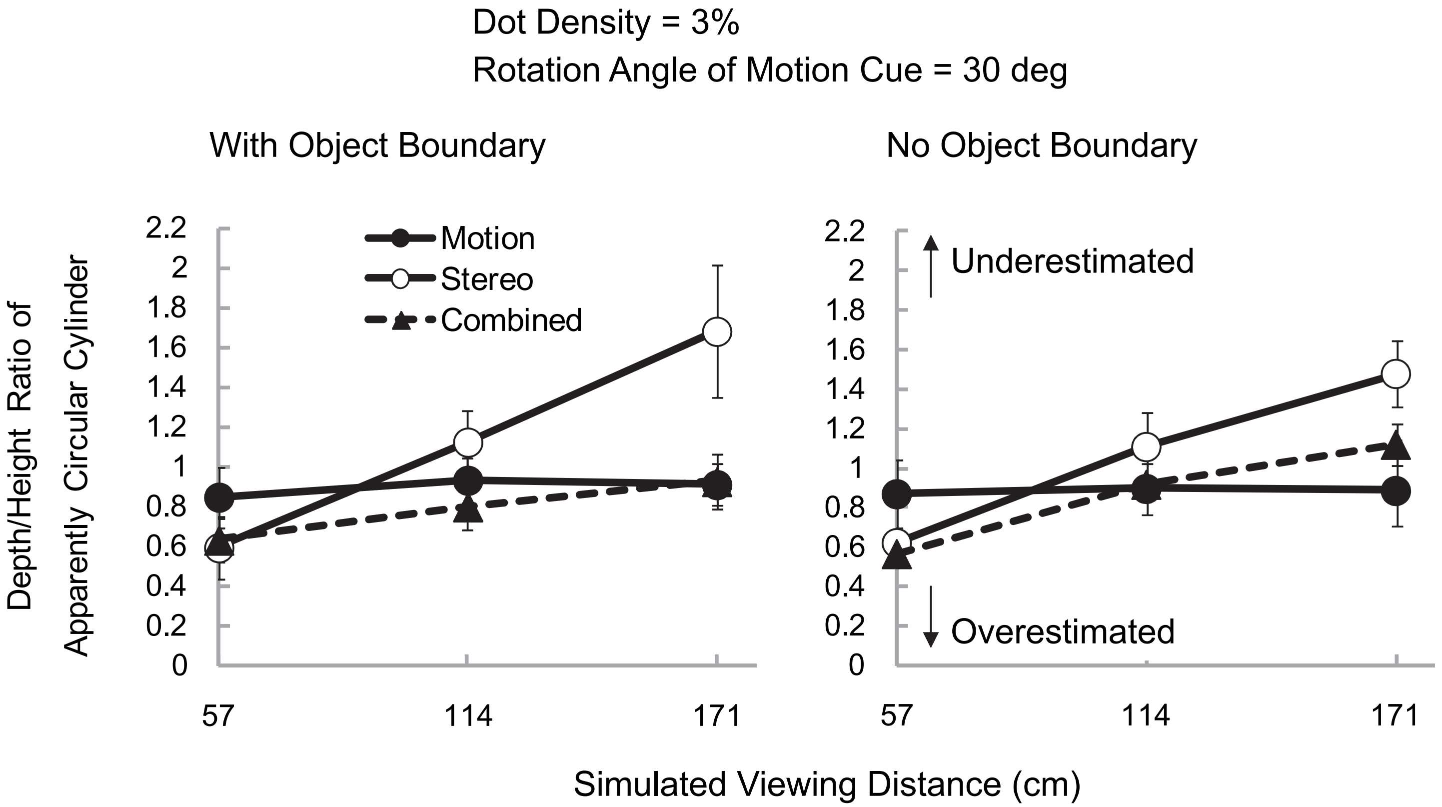
Dot Density = 3%
Rotation Angle of Motion Cue = 30 deg
With Object Boundary
No Object Boundary
Depth/Height Ratio of Apparently Circular Cylinder
Motion
Stereo
Combined
Underestimated
Overestimated
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
57
114
171
Simulated Viewing Distance (cm)

Figure 2

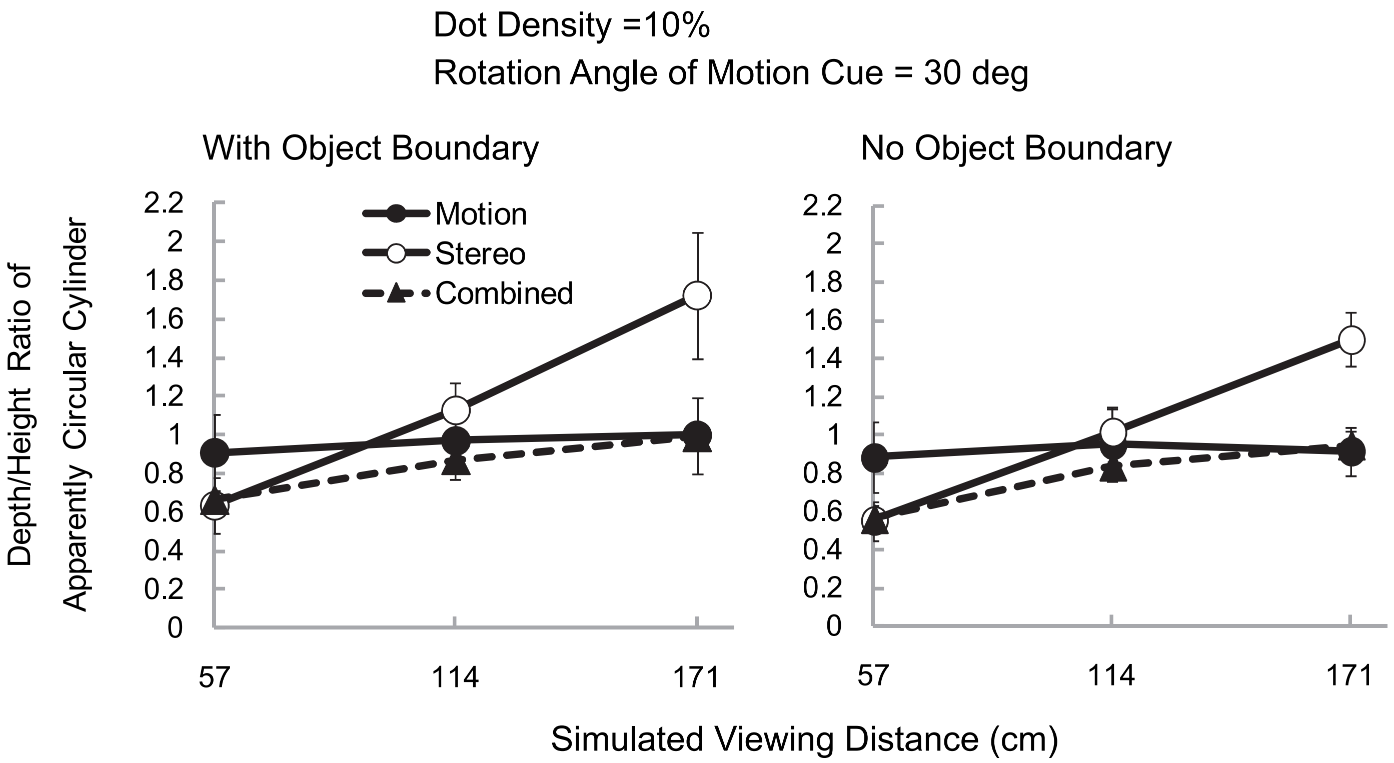
Dot Density =10%
Rotation Angle of Motion Cue = 30 deg
With Object Boundary
No Object Boundary
Motion
Stereo
Combined
Depth/Height Ratio of Apparently Circular Cylinder
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
57
114
171
Simulated Viewing Distance (cm)

Figure 3

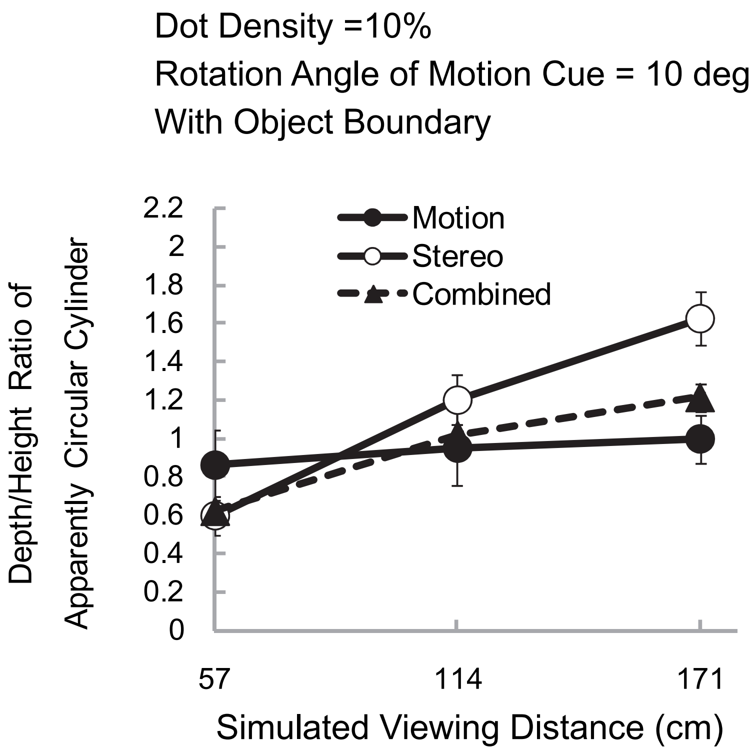
Dot Density =10%
Rotation Angle of Motion Cue = 10 deg
With Object Boundary
Depth/Height Ratio of Apparently Circular Cylinder
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
57
114
171
Simulated Viewing Distance (cm)
Motion
Stereo
Combined

Figure 4