American Psychological Association

Journal of Experimental Psychology: Human Perception and Performance

0096-1523/18/$12.00

2018, Vol. 44, No. 10, 1508–1522
http://dx.doi.org/10.1037/xhp0000553

Large Continuous Perspective Change With Noncoplanar Points Enables Accurate Slant Perception

Xiaoye Michael Wang
Indiana University

Mats Lind
Uppsala University

Geoffrey P. Bingham
Indiana University

Perceived slant has often been characterized as a component of 3D shape perception for polyhedral objects. Like 3D shape, slant is often perceived inaccurately. Lind, Lee, Mazanowski, Kountouriotis, and Bingham (2014) found that 3D shape was perceived accurately with perspective changes ≥45°. We now similarly tested perception of 3D slant. To account for their results, Lind et al. (2014) developed a bootstrap model based on the assumption that optical information yields perception of 3D relief structure then used with large perspective changes to bootstrap to perception of 3D Euclidean structure. However, slant perception usually entails planar surfaces and structure-from-motion fails in the absence of noncoplanar points. Nevertheless, the displays in Lind et al. (2014) included stereomotion in addition to monocular optical flow. Because stereomotion is higher order, the bootstrap model might apply in the case of strictly planar surfaces. We investigated whether stereomotion, monocular structure-from-motion (SFM), or the combination of the two would yield accurate 3D slant perception with large continuous perspective change. In Experiment 1, we found that judgments of slant were inaccurate in all information conditions. In Experiment 2, we added noncoplanar structure to the surfaces. We found that judgments in the monocular SFM and combined conditions now became correct once perspective changes were ≥45°, replicating the results of Lind et al. (2014) and supporting the bootstrap model. In short, we found that noncoplanar structure was required to enable accurate perception of 3D slant with sufficiently large perspective changes.

Public Significance Statement

A majority of the studies on slant perception as well as perception of other spatial properties, like shape, have shown perception is inaccurate. This is a problem for the control of actions that are not continuously guided by vision. In this study, we investigated visual information generated by motion, and showed that with sufficiently large continuous perspective change, surface orientation was perceived accurately. Such perspective changes would be generated by the locomotion of observers in everyday environments. This result not only provided additional confirmation for a model developed in a previous study on accurate shape perception, but also showed that the 3D structure of natural environments is important for perception of surface orientations.

Keywords: geographical slant perception, affine geometry, space perception, stereomotion, structure-from-motion

Three-dimensional (3D) slant perception has been treated as a component of 3D shape perception (e.g., Beck & Gibson, 1955; Gibson, 1950; Hoffman & Richards, 1984; Kaiser, 1967; Koffka, 1935; Sakata, Tsutsui, & Taira, 2005; Todd, 2004; Wallach & Moore, 1962; Welchman, Deubelius, Conrad, Bülthoff, & Kourtzi, 2005). A 3D polyhedral shape is composed of a set of local surfaces each of which would be perceived in respect to its 3D orientation. In turn, the perceived 3D orientation of a planar surface can be specified via its slant and tilt. Slant, in particular "local" or "optical" slant was defined by Gibson (1950; Gibson & Cornsweet, 1952) as the angle between a planar surface normal and the line of sight, while tilt is defined as the projected orientation of the surface normal in the image plane relative to a vertical axis (Stevens, 1983; Todd & Perotti, 1999). Tilt is used to define the direction of local slant. The fundamental assumption in this

This article was published Online First June 21, 2018.

Xiaoye Michael Wang, Department of Psychological and Brain Sciences, Indiana University; Mats Lind, Department of Information Technology, Uppsala University; Geoffrey P. Bingham, Department of Psychological and Brain Sciences, Indiana University.

Correspondence concerning this article should be addressed to Xiaoye Michael Wang, Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, IN 47408. E-mail: wang492@indiana.edu

understanding of 3D shape perception is that local perceptual estimates of 2D surface orientations are composed to yield perceptions of global 3D object shapes. This approach has also been elaborated to address the perception of 3D smoothly curved shapes to entail perceptual estimates of local higher order shape specific surface properties (e.g., Perotti, Todd, Lappin, & Phillips, 1998). Accordingly, Todd (2004) argued that 3D shapes were represented through a local map consisting of different neighborhoods that depict different aspects of the local 3D structures. Together, these local neighborhoods specify a particular 3D shape as a whole. Thus, according to these theories, the ability to perceive a 3D shape entails the ability to perceive local 3D structures of the object. In the case of 3D polyhedral shapes, this would entail perception of local surface slant.

Research in 3D shape perception has revealed inaccuracy in judgments of 3D shape, namely that perceived shapes are typically compressed or expanded in depth. Such findings have been interpreted to mean that the geometry of visual space is affine rather than Euclidean or metric (e.g., Todd, Oomes, Koenderink, & Kappers, 2001). See Wagner (2006) Chapter 5 and Wagner (2008) for review. Specifically, metric structure means the distance between any pair of points on a 3D object will be preserved after any arbitrary rigid translation and rotation, whereas 3D affine structure, or more properly, 3D relief structure (see Appendix A) allows a homogeneous stretch or compression of distances in depth, that is, in a direction orthogonal to a frontoparallel plane. The width-to-depth aspect ratio of a 3D object can be used to evaluate whether the perceived 3D shape is metric or relief, where metric structure would yield accurate judgment of the aspect ratio while relief would not. Numerous studies have shown that observers were unable to perceive metric structure, exhibiting judgments that were biased (that is, compressed or expanded in depth) and highly variable (e.g., Perotti et al., 1998; Tittle, Todd, Perotti, & Norman, 1995; Todd, Tittle, & Norman, 1995). Such findings prompt researchers to argue that 3D shape perception is of relief structure, and that a continuous family of linearly scaled depth maps would be required to describe the object's perceived depth (Koenderink & van Doorn, 1991; Lee, Lind, & Bingham, 2013). See Appendix A for description and discussion of different models and approaches relative to the terminology used to describe 3D structure.

Typically, studies of 3D shape perception have involved limited variations in perspective when observers viewed objects. On the other hand, Bingham and Lind (2008) found that with sufficiently large continuous perspective change (≥45°), observers could perceive the metric structure of 3D shapes. In a series of experiments, these investigators asked participants to perform feedforward reaches to touch the sides of upright cylindrical objects with a stylus. The elliptical cross-sections of the objects exhibited different aspect ratios. When the objects were viewed with perspective changes less than 45°, perception of the aspect ratios was poor. However, when participants viewed the object using stereovision and continuous perspective change greater than or equal to 45°, judgments became accurate regardless of whether the motion was generated by object rotation or by participants moving around the object. Moreover, discrete views differing by 90° failed to yield the same effect. This finding was later replicated (Lee & Bingham, 2010; Lee, Lind, Bingham, & Bingham, 2012) and extended to judgments of the aspect ratios of nonsymmetrical polyhedral objects, judged by sizing an image of the cross section of the object to the perceived aspect ratio (Lind et al., 2014).

Lind et al. (2014) proposed a bootstrap model to account for these results. They argued that the optical stereo and structure-from-motion (SFM) information enabled perception of relief shape, that is, shape with an unknown stretch or compression along the line of sight compared with the actual object. Using models published in previous studies (e.g., Koenderink & van Doorn, 1991; Shapiro, Zisserman, & Brady, 1995; Lind, 1996), one is able to recover such structure with small amounts of perspective change or rotation. With continuous perspective change, the orientation of two lines on the object, initially parallel and perpendicular to the line of sight (that is, at 90° to one another), would change. A 45° rotation yields bisection by the line of sight of the angle between the lines. Recognition of this bisection is enabled by relief structure (that is, structure that is affine or better) and, at that point, the unknown scaling factor relating the width and depth of the object can be determined. The authors also provided numerical simulations to demonstrate the effectiveness of the model and the importance of the 45° rotation (the model is reviewed in detail in Appendix B).

In light of this finding, Lee et al. (2013) further argued that failures in perceiving 3D metric shape found in many studies were merely due to a lack of information that specified the metric structure of the objects, rather than being due to any systematic distortion imposed by the perceptual system itself as suggested in previous articles (e.g., Johnston, 1991; Tittle et al., 1995). So, sufficient perspective change when viewing the 3D shapes (i.e., ≥45°) could provide additional information regarding the metric shape that would eliminate such ambiguity.

Likewise, many studies have also reported similar inaccuracies in local slant judgments. For instance, Gibson (1950), using a palm board as a response method for judged slant, found that perceived slants tended to be biased toward the frontoparallel plane, that is, appeared to be less depthy, when the slants were presented using texture gradients. Additionally, Flock (1964) demonstrated that such underestimation could be alleviated when the slanted surface is in lateral motion. In more recent studies, slant perception has been investigated using discrimination tasks (e.g., Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003). In contrast, Todd, Christensen, and Guckes (2010) used a slant-matching task in which participants were asked to adjust the orientation of a line to match the 3D slant of a surface shown in a display of a static texture gradient with a relatively small field of view. Using this method, the authors found that the perceived slant angle was systematically underestimated. They also found that static binocular disparities enabled more accurate perception of slant.

In studies of structure-from-motion, information about object shape is made available by rotating an object around an axis, usually vertically oriented, through its center producing continuous variation in perspective. If the shape is polyhedral, then, naturally, the planar object surfaces would be subject to variation in perspective. However, if slant is defined egocentrically as optical or local slant (that is, as an angle relative to the line of sight), then such changes would yield continuous variations in slant. That is, optical slant would not exhibit constancy under the rotation in SFM that is expected to improve perception. In particular, in respect to perception of Euclidean shape, Lind et al. (2014) found that a larger amount of rotation was required to yield accurate perception.

Perception of local slant would not be expected to improve under such conditions simply because more variation in perceived slant should be introduced by greater changes in perspective, thus effectively perturbing perceived slant rather than improving it. However, Gibson (1950) offered an alternative definition of slant that he called gravitational or geographical slant. This is defined as the angle formed relative to the horizontal and vertical axes determined by gravity (p. 369; see also Gibson & Cornsweet, 1952; Sedgwick & Levy, 1985). Gravitational slant is defined relative to the environment and remains constant over rotation around a vertical axis (see Figure 1 for illustration). In the current study, we defined geographical slant as the angle formed between the slanted surface and a level surface normal to gravity. As a consequence, variations in the corresponding tilt (i.e., geographical tilt), would correspond to the SFM rotation. We would define 0° tilt as the orientation in which the slant is away from the observer. All subsequent references to slant and tilt in this article refer to geographical slant and tilt.

Logically, if 3D shape perception can be improved with large continuous perspective change, and if slant perception is a component of the process yielding 3D shape perception (for polyhedral objects), then we would expect performance in slant perception to improve to become veridical with sufficiently large continuous perspective change. However, there is a potential problem if the bootstrap model is correct. The bootstrap model assumes that the available optical information enables perception of relief structure. The model then applies affine operations with sufficiently large perspective changes to obtain metric or Euclidean structure. The problem is that slant perception usually entails displays of planar surfaces. The structure from motion models require noncoplanar points (Koenderink & van Doorn, 1991; Lind, 1996; Shapiro et al., 1995). 3D shapes under rigid rotation in SFM displays would entail noncoplanar points, and thus, yield relief structure, but planar surfaces rigidly rotating in monocular SFM displays would not.

Additionally, all of the extant studies providing results supporting the bootstrap model have involved stereovision together with the continuous perspective changes in SFM displays. When objects move with stereo viewing, three different types of motion generated information become available including two forms of stereomotion in addition to monocular SFM. The stereomotion information is change-of-disparity-overtime (CDOT) and interocular velocity difference (IOVD; Allen, Haun, Hanley, Green, & Rokers, 2015; Cumming & Parker, 1994; Nefs, O'Hare, & Harris, 2010; Shioiri, Saisho, & Yaguchi, 2000). CDOT consists of the temporal rate of change of stereo disparity as the object and perceiver move relative to one another. IOVD consists of stereo disparities between the monocular optic flows in each eye. Although Lind et al. (2014) used the combination of all three types of visual information in their study, their model only required optical information that would yield relief 3D object structure then used for the bootstrap to Euclidean structure. Either monocular SFM or static stereo disparities should in principle be sufficient to yield relief 3D shape, but not the slant of a planar surface. On the other hand, stereomotion is above first order because it entails more than the mere two views assumed to be available for SFM or stereo. Stereo disparities entail two views, but the disparities themselves transform over time in stereomotion yielding effectively more than two views. Stereomotion information in addition to monocular optic flow for SFM may enable perception of relief structure even with planar surfaces. It remains unclear whether large perspective changes would be effective in perception of the slant of planar surfaces with only stereomotion (either CDOT or IOVD), or with the combination of stereomotion and monocular SFM. It is unlikely to be effective with only monocular SFM information.

In the current study, we investigated several interrelated questions. First, we combined the experimental paradigms of Lind et al. (2014) and Todd et al. (2010) to examine whether 3D slant perception improves to become accurate with large (≥45°) continuous perspective changes. In displays that varied in respect to the amount of perspective change from rotation, participants judged the slant of a visible planar surface by adjusting a line to the perceived slant angle. We first tested if performance in judging the slant of planar surfaces improved as the amount of perspective change was increased to 45° of rotation, beyond which it was

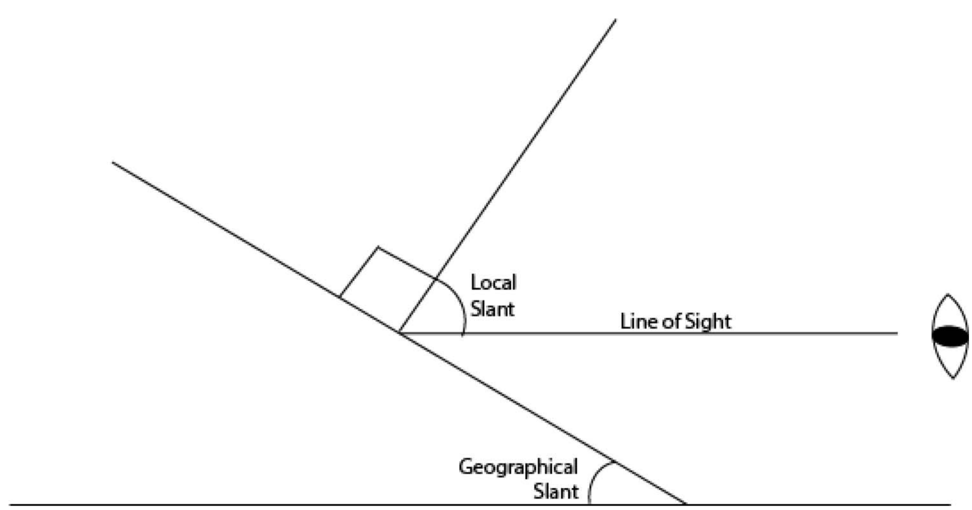

Figure 1. Illustration of local versus geographical slant.

expected to be accurate. Subsequently, we tested if the introduction of additional noncoplanar points to the surface would yield improved performance, as suggested by the bootstrap model and SFM models. A second goal of this study was to explore the role of different motion-generated sources of visual information in perception of 3D slant. In particular, we aimed to explore the effectiveness of stereomotion (CDOT), monocular structure-from-motion (SFM), and the combination of the two (plus IOVD) for accurate perception of 3D slant.

Experiment 1

In Experiment 1, observers judged displays of a slanted planar surface rotating around a vertical axis through its center with different amounts of rotation and different types of visual information. Displays contained one of the three types of visual information, namely, SFM, CDOT, or the combination plus IOVD. Observers adjusted the orientation of a line in the display to match the perceived 3D slant of the surface. To anticipate, we found that 3D slant perception was different from 3D shape perception to the extent that large continuous perspective change did not reproduce the findings of Lind et al. (2014). Rather, we found that the stereomotion and combined conditions exhibited the same results, both exhibiting a trend for veridical judgment without actually achieving it, whereas the monocular SFM condition remained consistently poor.

Method

Participants. Thirteen adults, age between 20 and 30 (three males and 10 females) participated in this experiment, all of who provided their informed consent prior to participating in the experiment with the approval from Indiana University's Institutional Review Board (IRB). They were paid $10/hr. All participants had normal or corrected-to-normal eyesight, and also passed a stereo fly test (Stereo Optical Co., Inc., Chicago, IL) that measured their stereo acuity. Participants had to be able to identify the target circle indexed by a disparity of 80 seconds of arc to be included.

Stimuli and apparatus. Stimuli were presented on a Dell UltraSharp U2312HM 23-in. monitor (51 cm by 29 cm) with a resolution of 1,920 × 1,080 and a refresh rate of 60 Hz. We used MATLAB 2011b (the Mathworks Inc., Natick, MA, 2011) to generate and present the displays. From trial to trial, slant was varied randomly between 27° and 73° by 2° increments yielding 24 different slants. Again, because we were testing geographical slant, the slant angles were all defined in terms of angles formed between the surface and the horizontal plane. Because the center of the surface was placed at eye level, the equivalent optical slant when the surface tilt was 0° (i.e., facing the frontoparallel direction during rotation) would be 90° minus the geographical slant. The range of slant angles was chosen to avoid slants near either 0° or 90°. We presented a slanted surface rotating in the display about a vertical axis through the center of the surface with the rotation centered on a 0° tilt. Similar to the range of rotation used in Lind et al. (2014), there were five different rotation amounts: 25°, 35°, 45°, 55°, or 65°. This range would reveal improvement as rotation approached 45° and then continued accuracy for amounts greater than 45°. For each rotation amount (e.g., 25°), the surface rotated first to one side of 0° tilt by half of the amount (e.g., 12.5°) and then back and to the other side and back by the same amount. The surface rotated at a constant speed of 20°/s This meant that even with 65° of rotation, the surface only rotated 32.5° away from 0° tilt. Displays continued the oscillating rotation until observers completed their judgments and ended the trial. Thus, even though a single speed of rotation would yield different durations for the different rotation amplitudes, this was independent of trial duration and the duration during which information was available to the observer. Todd and Bressan (1990; and subsequently, other researchers) similarly controlled for their variations in frame number in SFM displays by continuously oscillating the display until judgments were completed. The 24 different slants were tested using each of the five different rotation amounts. We also varied the length of the slanted surface while fixing the width at 10 cm. Three surface lengths (8 cm, 10 cm, and 12 cm) yielded three projected heights at each slant.

We used red-blue random dot anaglyphs to generate displays that contained three types of visual information: monocular motion, stereomotion, and the two combined. For all conditions, the dots had the size of one pixel, with a total of 6,000 dots. The displays contained a background, placed 18 cm behind the screen, and a target surface, placed 9 cm behind the screen. The point of observation used to generate the stimuli was placed 76.2 cm in front of the screen, mimicking the actual distance during the experiment. Given the above layout, the background surface spanned approximately 9.10° horizontally and vertically. The horizontal size of the projected slant surface is 6.72°, whereas its vertical size varied as a function of slant angle and surface length, approximately ranging from 2.10° (slant angle 23° and surface length 8 cm) to 7.71° (slant angle 73° and surface length 12 cm). The size of the stimuli therefore enabled approximation of perspective projection using scaled orthographic projection, which is a crucial requirement for the model. Figure 2 is a schematic of this setup.

The displays were generated by first constructing the actual target surface and a background using random dots in a 3D space, and back-projecting the dots onto the screen surface through the point of observation (projection point). The back-projected dots were the ones to be drawn on the screen. Hidden line removal was used to address the occlusion of the background as the slanted target surface rotated.

For both stereomotion and combined conditions, we projected the points through left and right projection points, corresponding to the viewers' left and right eyes, with an interpupillary distance (IPD) of 6 cm. The stereomotion condition only contained CDOT information. To do this, we rerandomized the dot positions every frame while preserving the evolving disparities, so that only binocular disparity could specify the changing 3D structures in the display. The combined condition contained CDOT, IOVD, and monocular SFM information. The random texture was not rerandomized each frame. CDOT was specified through the evolving binocular disparity, whereas IOVD and SFM were specified through the stable monocular structure across frames. Finally, the monocular motion condition only contained structure-from-motion information. For this condition, the texture was not rerandomized each frame and the IPD was set to 0. The color of the dots in this condition was therefore the combination of red and blue. Because of the way our displays were constructed, there was also texture gradient information in the combined and SFM displays.

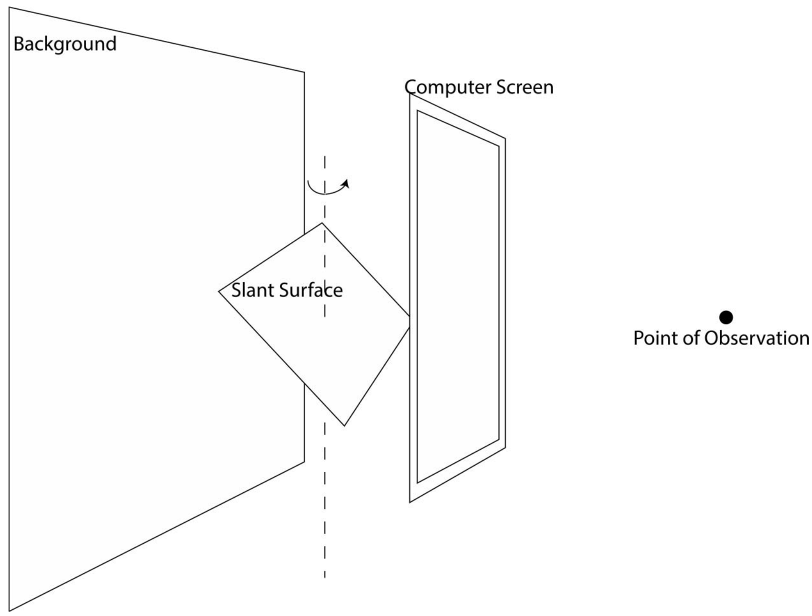

Figure 2. Schematic demonstration of the stimuli setup in Experiment 1. Note: display setup is not up to scale.

Because the combinations of rotation amount and slant angle yielded 120 trials for each of the three visual information conditions, we decided not to fully cross surface heights with slant angles. Instead, we treated three consecutive slant angles as a block (e.g., 17°, 19°, and 21°), and randomly assigned one of the three slant heights to each of the slants in each consecutive block (e.g., 10 to 17°, 12 to 19°, and eight to 21°). The surface height assignment was fixed for each combination of rotation angle and slant angle of each visual information condition (e.g., all participants would see the surface height of 10 for slant angle of 17°, rotation angle of 25, in the monocular condition).

Procedure. After participants provided informed consent and passed the stereo fly test, they sat in front of the computer screen, wearing a pair of red-blue filter glasses and viewing the display binocularly for all three visual information conditions, and were instructed how to perform the task. Participants first observed a rotating planar surface with a certain slant. After the surface completed one cycle of rotation (i.e., the surface rotated back to the starting position at 0° tilt), a 2D response line appeared on the screen to the right of the display. Participants were instructed to use the left and right arrow keys to adjust the orientation of the line so that it matched the slant of the surface in the display (as in Todd et al., 2010). The surface continued to rotate back and forth throughout the trial while participants adjusted the response line and until they hit the space bar to enter the judgment once they were satisfied that the orientation of the line matched the surface slant.

Each participant completed three sessions, one for each visual information condition (120 trials). The order of visual information conditions was randomly determined. Each session was conducted at least 1 day apart from the previous one. Within each session, participants experienced the rotation amounts in order from 25 to 65. We did not randomize the order of rotation amounts to avoid the possibility that large perspective change would calibrate the judgments performed with smaller perspective changes. Within each rotation angle condition, the slant angles were randomly displayed.

Data analysis. For data analysis, we first used multiple regression to examine the effects of surface height on perceived slant. We found such effects to be small, so we averaged across different surface heights in the subsequent analyses. The goal of subsequent analyses was to explore the effects of rotation amount and visual information conditions on perceived slant. We performed simple linear regressions with the actual slant as the independent variable and the perceived slant as the independent variable. We computed the regression slopes, intercepts and r^2 for each participant in each information and rotation amount condition. Then, we used the resulting slopes, intercepts, and r^2 as dependent measures in repeated-measures analyses of variance (ANOVA's). In these ANOVA's, there were two within-subject factors, visual information (three levels) and rotation amount (five levels).

Results and Discussion

We began analysis by using multiple regression to evaluate the effects of surface height on judged slant. We evaluated the effects of three factors (rotation angle, actual slant angle, and surface height) separately within each visual information condition. The dependent variable was perceived slant. For all three visual information conditions, slant angle accounted for the majority of the variance in the model. Surface height accounted for a very small portion of the total variance in each of three visual information conditions (0.8% with Cohen's $f^2 = 0.036$ for CDOT, 2.8% with Cohen's $f^2 = 0.079$ for monocular motion, and 0.4% with $f^2 = 0.019$ for the combined condition). Because changes in r^2 with the inclusion of surface height were exceedingly small, it was not included as a variable in the subsequent analyses.

Next, we performed linear regressions comparing perceived and actual slants for each unique combination of visual information and rotation amount separately for each participant. Accurate judgments would yield slopes of 1 and intercepts of 0. Good precision would yield r^2 approaching 1. Figures 3–5 show the average regression slopes, intercepts, and r^2 and for each unique combination of conditions.

For regression slopes, ANOVA showed that there were significant main effects of visual information, $F(2, 22) = 24.77$, $p < .001$, $\eta_p^2 = 0.69$, and of rotation amount, $F(4, 44) = 7.01$, $p < .001$, $\eta_p^2 = 0.39$. There was also a significant interaction effect between visual information and rotation amount, $F(8, 88) = 2.35$, $p < .05$, $\eta_p^2 = 0.18$. For the significant main effects, a least significant difference (LSD) adjustment for multiple comparisons was used. Among the three visual information conditions, the means for the monocular condition were significantly lower than those for the stereomotion condition ($p <$ .001), and the combined condition ($p < .001$). However, there was no significant difference between the stereomotion and combined conditions ($p > .6$).

Because there was no difference between the stereomotion and combined conditions, we performed a post hoc test comparing the stereomotion and combined conditions. As expected, the ANOVA showed a significant main effect of rotation amount, $F(4, 44) = 7.81$, $p < .001$, $\eta_p^2 = 0.42$, but no difference between the two information conditions ($p > .6$, $\eta_p^2 = 0.023$). LSD adjustment on different levels of rotation amount showed that the 25° condition yielded a significantly greater regression slope than the other conditions. Moreover, the regression slope in the 65° condition was also significantly lower than other conditions. Despite these changes, the regression slope failed to reach 1 at 45° or beyond. We performed a series of one-tailed one-sample t tests to compare values at each rotation amount with 1, which showed that all conditions had regression slopes significantly different from 1 ($p < .001$). A post hoc ANOVA performed on the data in the monocular motion condition yielded no effect of rotation amount ($p > .6$, $\eta_p^2 = 0.051$). The mean slopes were below 1, with an overall mean of 0.89 and a standard error of 0.042. One-sample t test showed that except for 35° rotation ($p = .09$), all other rotation conditions had regression slopes that are significantly different from 1 ($p < .05$). As shown in Figure 3, the mean slopes in the stereomotion and combined conditions started at greater than 1 and, as the rotation amount increased, their values decreased but have never reached 1. On the other hand, in the monocular condition, the mean slopes remained below 1 for all rotation amounts.

An ANOVA on regression intercepts showed that there were significant main effects of both visual information, $F(2, 22) = 38.18$, $p < .001$, $\eta_p^2 = 0.78$, and rotation amount, $F(4, 44) = 7.84$, $p < .001$, $\eta_p^2 = 0.42$, as well as a significant interaction effect between the two, $F(8, 88) = 3.03$, $p < .01$, $\eta_p^2 = 0.22$. A post hoc LSD comparison showed that intercepts for the monocular motion condition were significantly greater than those for the stereomotion condition ($p < .001$), and the combined condition ($p < .001$), whereas there was no significant difference between the two stereo conditions ($p > .8$). To explore the effects of rotation amount on the different visual information conditions, we again performed a post hoc ANOVA on the stereomotion and combined conditions. It yielded a significant main effect only of rotation amount, $F(4, 44) = 10.26$, $p < .001$, $\eta_p^2 = 0.48$. LSD comparison showed that the 25° rotation amount exhibited significantly smaller mean intercepts than all other rotation amounts. As shown in Figure 4, for both the stereomotion and combined conditions, as the rotation amounts increased, regression intercepts also gradually increased and only reached 0 at 65° (where their values were not significantly different from 0, with $p > .05$). An ANOVA on the

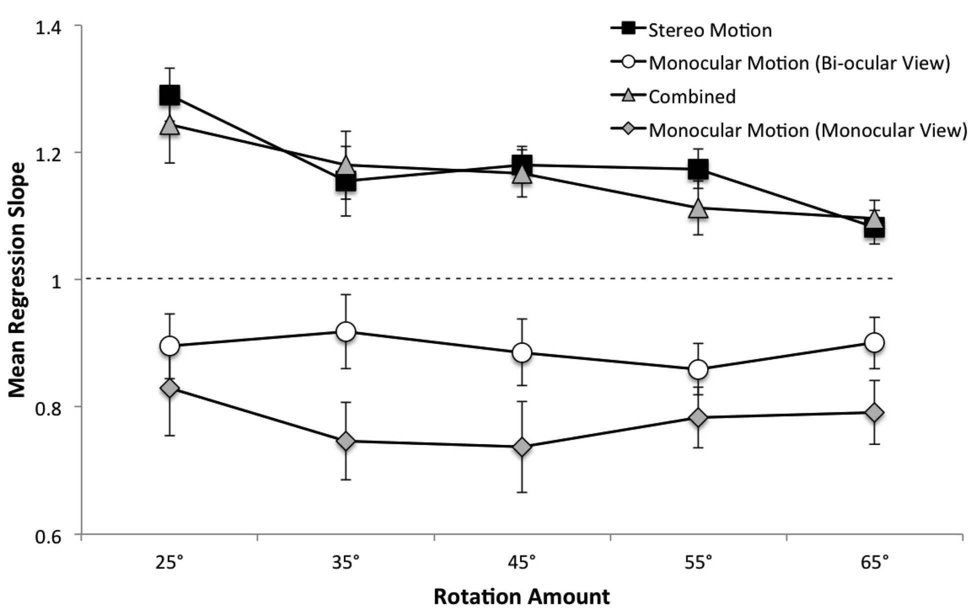

Figure 3. Mean regression slopes in Experiment 1 for the biocularly viewed monocular motion condition (circles), stereo motion (squares), and combined (triangles) conditions, as well as monocularly viewed monocular motion condition (diamonds) plotted by rotation amount. Error bars represent standard errors.

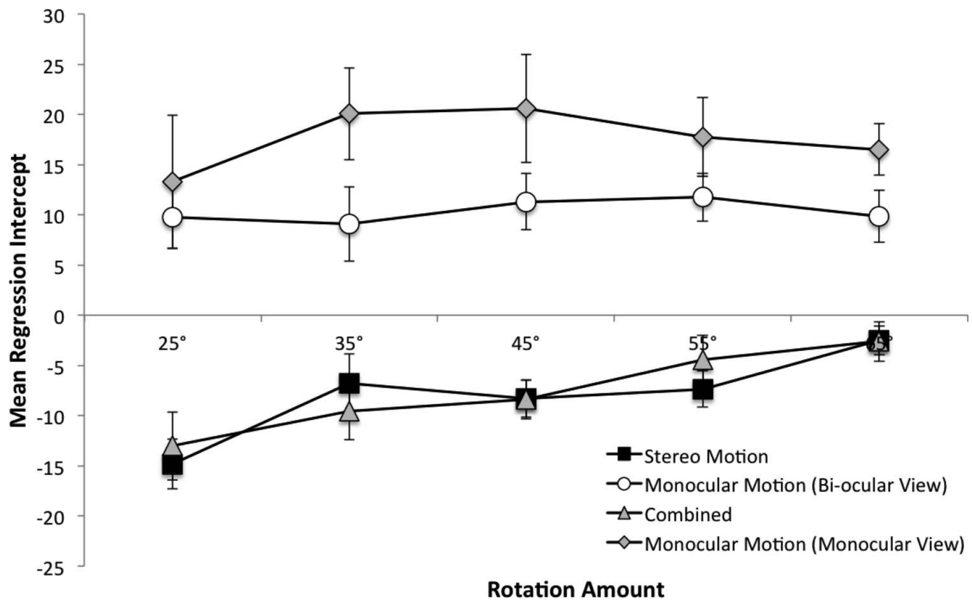

Figure 4. Mean regression intercepts in Experiment 1 for the biocularly viewed monocular motion condition (circles), stereo motion (squares), and combined (triangles) conditions, as well as monocularly viewed monocular motion condition (diamonds) plotted by rotation amount. Error bars represent standard errors.

intercepts in the monocular motion condition yielded no significant effects of rotation amount. However, the mean intercepts for all rotation amounts were above 0 ($p < .05$ for all rotation amounts).

Next, we performed an ANOVA on the r^2 from the simple regressions. This yielded a significant main effect of visual information condition, $F(2, 22) = 28.46$, $p < .001$, $\eta_p^2 = 0.72$, and of rotation amount, $F(4, 44) = 5.15$, $p < .01$, $\eta_p^2 = 0.32$. However, there was no significant interaction effect between the two factors, $F(8, 88) = 1.48$, $p > .1$, $\eta_p^2 = 0.12$. LSD showed that the monocular condition exhibited significantly smaller mean r^2 than both the stereomotion ($p < .001$) and combined conditions ($p < .001$), and that there was no difference between the means in the stereomotion and combined conditions. Similarly, a separate ANOVA performed only on the stereomotion and combined conditions yielded no significant main effects or interactions. More-

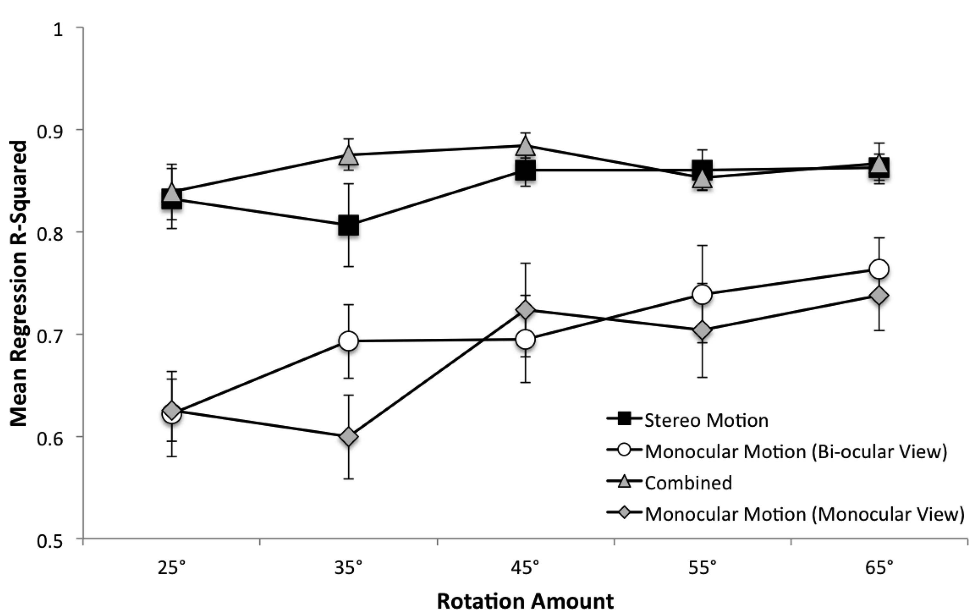

Figure 5. Mean regression R^2 in Experiment 1 for the biocularly viewed monocular motion condition (circles), stereo motion (squares), and combined (triangles) conditions, as well as monocularly viewed monocular motion condition (diamonds) plotted by rotation amount. Error bars represent standard errors.

over, an ANOVA on the monocular condition did yield a significant main effect of rotation amount, $F(4, 44) = 2.91$, $p < .05$, $\eta_p^2 = 0.21$. Post hoc LSD comparison showed that the 25° rotation amount yielded a significantly smaller r^2 than did the 35° and 65° rotation conditions. As Figure 5 shows, r^2 for both the stereomotion and combined conditions remained relatively high and stable at around 0.85, while the r^2 for the monocular motion condition started off relatively low and increased as the rotation amount increased but never reached the level of the stereomotion and combined conditions.

A potential confound in Experiment 1 is that participants viewed the monocular SFM displays biocularly, that is, using two eyes. Although information regarding the slant is specified monocularly, the addition of static stereoptic information specifies the upright display screen, which might affect the judgments of slant. Thus, we also ran a control experiment using the same stimuli with 10 additional naïve participants to perform in the monocular SFM condition monocularly with their dominant eye only. None had participated in the original experiment. Mean regression slopes, intercepts, and r^2 are plotted in Figures 3–5. Data analysis revealed that there was no significant difference in regression slopes ($p > .05$), intercepts ($p > .1$), or r^2 ($p > .5$) between the two monocular information conditions. We conducted a series of one-sample t tests on regression slopes and intercepts in the monocularly viewed monocular motion condition, comparing the former with 1 and the latter with 0. Results showed that regression slopes were significantly different from 1 and regression intercepts were significantly different from 0 for all rotation amounts ($p < .05$). Thus, when viewed monocularly, performance in the monocular motion condition was as bad or worse than when the displays were viewed biocularly.

In summary, performance in the monocular SFM condition was as expected. The lack of noncoplanar points in the rigid motion of the planar surface should yield failure of structure-from-motion. Indeed, performance was inaccurate and accuracy was unaffected by increases in the amount of perspective change. In contrast, the accuracy of judgments made using stereomotion information improved with increases in the amount of perspective change. However, the availability of stereomotion information failed to generate slopes of 1 at 45° of rotation. Instead, for the two stereomotion conditions, regression slopes only gradually decreased toward 1 but they failed to become 1 even at 65° of rotation. Such gradual change was also reflected in the change of intercepts. In the Lind et al. (2014) study, performance rather suddenly became accurate and more precise with 45° of rotation and then remained so with greater amounts of rotation.

Why should perceived slant be different in this way from perceived shape? Certainly, in the monocular SFM condition, the failure to reproduce the results of previous studies could be attributed to the lack of noncoplanar points in the planar surfaces judged in Experiment 1. Although the stereomotion conditions yielded changes in performance in response to increased perspective change, the changes did not replicate those found in previous studies. In particular, judgments failed to become accurate. Our displays of strictly planar surfaces may not have yielded perception of 3D relief structure even with stereomotion. The 3D shapes judged in previous investigations all entailed sets of rigid noncoplanar points. The surfaces judged in Experiment 1 did not. In Experiment 2, we used the same paradigm to test perception of the slant of surfaces containing noncoplanar points.

Experiment 2

In Experiment 2, we examined the possibility that performance with the strictly planar surfaces was due to the lack of 3D surface structure in the display. The 2D planar surfaces might have failed to provide the viewer with the relief structure required by the bootstrap model. We tested this possibility by adding nine small 3D cuboids to the top of the displayed surfaces. Essentially, we cut out sections of the original surface and raised them slightly above the surface without altering the surface orientations, that is, they remained parallel to the original surface. The point was to minimize perturbation of slant that would be confounded with the addition of noncoplanar points.

Because performance in the original stereomotion and combined conditions was almost identical, we only tested the combined condition in addition to the monocular SFM (which was viewed monocularly). All other aspects of the experimental design in Experiment 1 were the same. With noncoplanar points, even the monocular SFM condition would be predicted to yield perception of relief 3D structure and thus, with sufficient perspective change, perception of accurate geographical slant. Results showed improved performance in both visual information conditions that now replicated the previous 3D shape perception results.

Method

Participants. Twenty-four adults, age between 20 and 30 (11 males and 13 females) participated in this experiment, all of who provided their informed consent prior to participating in the experiment with the approval from Indiana University's IRB. They were paid $10/hr. All participants had normal or corrected-to-normal eyesight, and also passed the same stereo fly test. There were 12 participants in each of two information conditions, with five males and seven females in the monocular noncoplanar condition, whereas six males and six females in the combined noncoplanar condition.

Stimuli and apparatus. Experimental apparatus and display manipulations were all the same as in the previous experiment, with the following exceptions: First, only the monocular SFM and combined conditions were tested, and second, nine identical rigid cuboids were added on the top of each planar surface (see Figure 6 for a schematic demonstration). The top surfaces of the cuboids were parallel to the main surface and thus avoided adding noise to the display. Each cuboid was 1 cm in length and width, and 0.55 cm in height above the main surface. Depending on the slant angle, the projections of the top surface of each cuboid had a size ranging from 0.26° (23° slant) to 0.64° (73° slant). The cuboids were placed in a three-by-three grid on the top of the slanted surface, and were separated by 3 cm from one another. In this way, the cuboids spanned the entire surface. The height of the cuboids was selected so that when the surface rotated, the cuboids did not occlude one another. As with Experiment 1, the stimuli were generated using random dot anaglyphs.

Procedure. Unlike Experiment 1, this experiment entailed a between-subjects design to provide a stronger test of each of the

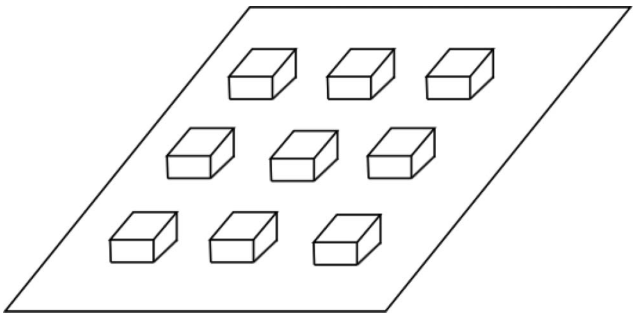

Figure 6. A schematic illustration of the noncoplanar object used in Experiment 2. Note that the actual display still employs random dots as used in Experiment 1.

two information conditions. In the monocular condition, participants viewed the display monocularly with their dominant eye. Two participants' data in the combined condition were discarded due to extremely poor performance as a result of failure to correctly understand the instructions and/or lack of attentiveness during the experiment.

Results and Discussion

To evaluate the effects of the addition of 3D structure to the surfaces, we combined the results from this experiment with those in the monocularly viewed monocular SFM and combined conditions in Experiment 1, treating the type of surface (planar vs. noncoplanar) as a between-subjects factor.

We performed mixed-design ANOVAs on slopes, intercepts and r^2 with two between-subjects factors, the type of surface and visual information condition, and one within-subject factor, the amount of rotation. Figure 7 shows the mean linear regression slopes for noncoplanar surfaces as a function of different rotation amounts for the monocular and combined conditions. As the figure shows, the overall regression slopes decreased with increasing rotation in both the monocular and combined conditions. Changes in both conditions shifted the overall mean slopes toward 1. Analyses on regression slopes showed that there was a significant main effect of rotation amount, $F(4, 160) = 6.74$, $p < .001$, $\eta_p^2 = 0.14$, and of visual information, $F(1, 40) = 15.44$, $p < .001$, $\eta_p^2 = 0.28$, but no significant main effect for the type of surface ($p > .3$). However, there was a significant interaction effect between visual information and the type of surface, $F(1, 40) = 18.00$, $p < .001$, $\eta_p^2 = 0.31$. Taking a closer look at the interaction, we found that there was a significant difference between the monocular and combined conditions for the planar surfaces (95% confidence intervals were between 0.68 and 0.88, and between 1.07 and 1.25, respectively), but not for the noncoplanar surfaces (the respective confidence intervals were between 0.93 and 1.11, and between 0.91 and 1.10). Due to the lack of difference between monocular and combined conditions with the noncoplanar surfaces, we combined the data in a post hoc test to examine the effects of rotation amount. We performed a one-sample *t* test, comparing regression slopes with 1 for each rotation amount. We found that regression slopes were significantly greater than 1 for both 25° and 35° rotation ($p < .01$ and $p < .05$). However, at 45° and beyond, regression slopes were no longer different from 1.

Analyses on regression intercepts revealed a similar trend as the slopes (see Figure 8). There were significant main effects of rotation amount, $F(4, 160) = 8.96$, $p < .001$, $\eta_p^2 = 0.18$, and of visual information, $F(1, 40) = 17.60$, $p < .001$, $\eta_p^2 = 0.31$, but no significant main effect for the type of surface ($p > .3$). However, there was a significant interaction effect between visual information and the type of surface, $F(1, 40) = 13.71$, $p = .001$, $\eta_p^2 = 0.26$, and between rotation amount and visual information, $F(4, 160) = 2.79$, $p < .05$, $\eta_p^2 = 0.07$. Again the interaction between visual information and type of surface showed that intercepts were significantly different between monocular and combined conditions for the planar surface (with confidence intervals above 0 (between

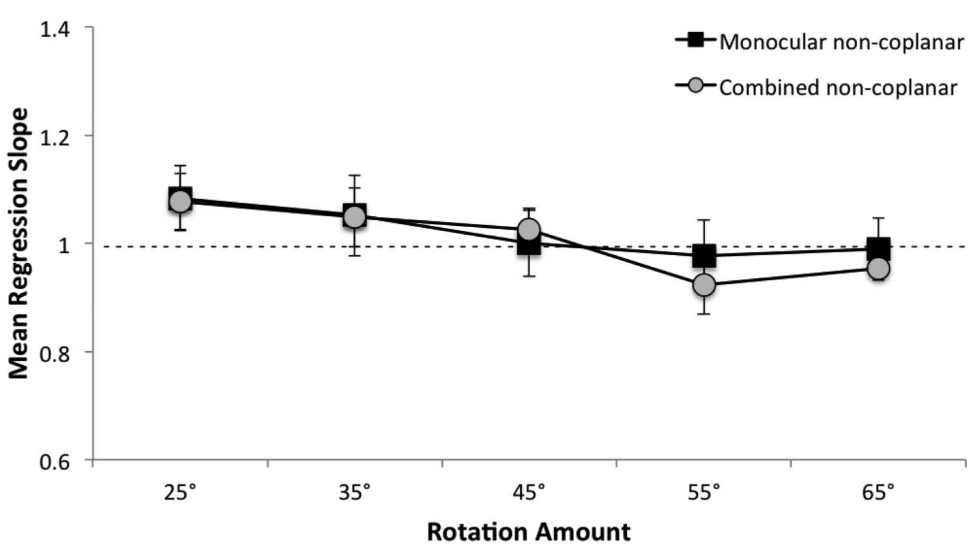

Figure 7. Mean slopes in Experiment 2 for the monocular (squares) and combined (circles) conditions plotted by rotation amount. Dashed line indicates slope of 1. Error bars represent standard errors.

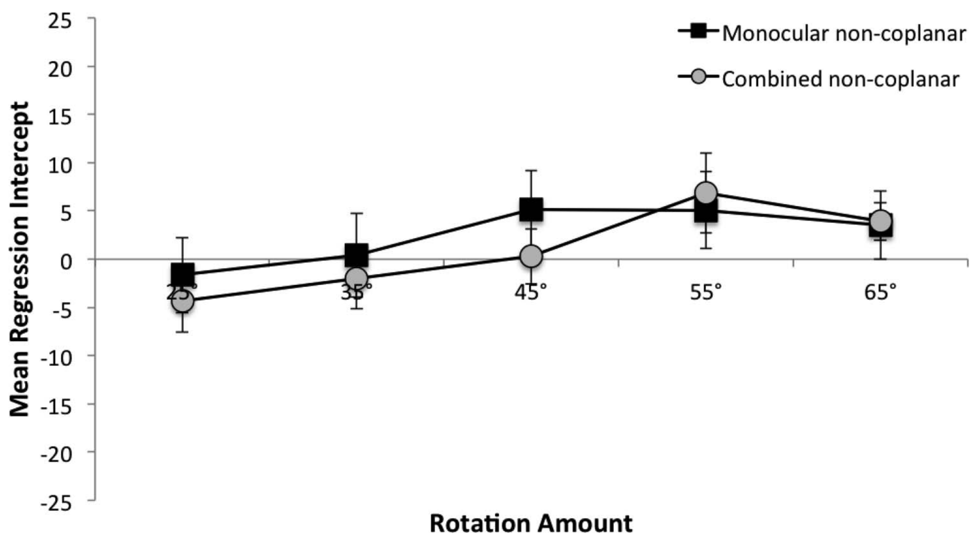

Figure 8. Mean intercepts in Experiment 2 for the monocular (squares) and combined (circles) conditions plotted by rotation amount. Error bars represent standard errors.

10.89 and 24.40) and below 0 (between −13.13 and −2.11, respectively), but not for the noncoplanar surface, with intercepts in both information conditions indistinguishable from 0 (confidence intervals are between −3.00 and 8.02, and between −5.10 and 6.97, respectively). Intercepts with all amounts of rotation for monocular and combined conditions with noncoplanar surfaces, were not significantly different from 0.

An ANOVA on regression r^2 showed significant main effects of rotation amount, $F(4, 160) = 4.37$, $p < .01$, $\eta^2 = 0.099$, and of visual information, $F(1, 40) = 21.52$, $p < .001$, $\eta^2 = 0.35$. There was also a significant interaction effect between rotation amount and visual information, $F(4, 160) = 7.74$, $p < .01$, $\eta^2 = 0.16$, and between visual information and types of surface, $F(1, 40) = 10.30$, $p < .01$, $\eta^2 = 0.21$. As shown in Figure 9, there was a slight increase of r^2 for the monocular condition after introducing noncoplanar surfaces, and a decrease for the combined condition.

Results using noncoplanar surfaces in monocular and combined conditions in comparison with those using planar surfaces revealed that introducing additional 3D structure to planar surfaces enhanced the accuracy of slant judgments. The findings provided support for the Bootstrap model, as they replicated the pattern of results found in Lind et al. (2014) with 3D shapes. In particular, noncoplanar slants in both monocular and combined conditions had regression slopes that were greater than 1 when the rotation amount was less than 45°, but slopes became indistinguishable from 1 and intercepts became indistinguishable from 0 once the rotation reached 45° or beyond. Thus, judgments of geographical

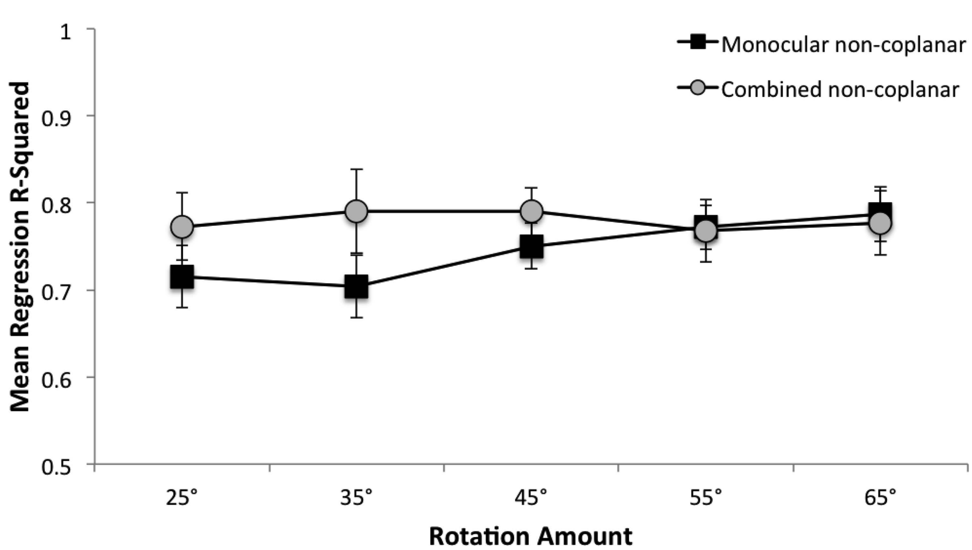

Figure 9. Mean R^2 in Experiment 2 for the monocular (squares) and combined (circles) conditions plotted by rotation amount. Error bars represent standard errors.

slant became accurate with perspective changes ≥45° and noncoplanar structure on the surfaces.

A potential confound in this experiment concerns the necessity of motion. Specifically, it is possible that the addition of 3D cuboids provided static 3D structure that enables observers to perceive metric slant, with or without motion. To address this issue, we conducted a control experiment with stimuli used in the monocular SFM condition with cuboids and 45° of continuous perspective change, as it was when judgment became veridical. The display sampled static frames from the motion sequence when the object is frontoparallel (tilt = 0°), at either end of the perspective change (i.e., −22.5° and 22.5°), as well as the midpoints between the frontoparallel view and the end views (i.e., −11° and 11°). Each frame was showed sequentially (from −22.5° to −11° and to 0° etc.) with a 1-s motion mask inserted between each frame, and participants had to finish viewing an entire cycle of rotation before making judgment. When making judgment, participants could also toggle the static frame using a computer key. Six participants were recruited, viewing the displays monocularly with their dominant eye. The mean regression slope was 0.67 (SE = 0.063), intercept 26.65 (SE = 6.85), and r^2 0.69 (SE = 0.043). A one-sampled t test showed that the regression slopes were significantly smaller than 1 ($p < .01$), and intercepts were significantly greater than 0 ($p < .01$). Comparing with its motion counterpart, we can see that even with a reasonable sample of the static frames, motion, in this case, large continuous perspective change is still required to yield metric judgments.

The above findings indicate that relatively poor performance with planar surfaces was due to the lack of 3D structure on the surfaces preventing observers from obtaining the relief structure that would then be used to derive the accurate slant estimation given increased perspective change. However, with noncoplanar elements added to the planar surface, one could use the emerging relief structure to perform the bootstrap method with 45° rotation. Moreover, we saw that regression slopes did not monotonically decrease as a function of rotation amount. Instead, they remained relatively stable once the rotation amount was greater than 45°, further confirming the bootstrap model.

General Discussion

In this study, we attempted to replicate the 3D shape perception results of Lind et al. (2014) to examine whether large continuous perspective change (≥45°) would enable accurate 3D slant perception. The reason was that perceived slant has often been characterized as a component of 3D shape perception for polyhedral objects. Lind et al. (2014) developed a bootstrap model to account for their results. The model is based on the assumption that the available optical information yields perception of 3D relief structure (e.g., see for instance Koenderink & van Doorn, 1991; Lind, 1996; Shapiro et al., 1995) that is then used in the context of large perspective changes to bootstrap to perception of 3D Euclidean structure. A potential problem arises for application of the model to slant perception because the surfaces in slant perception studies are typically planar. Structure-from-motion fails in the absence of noncoplanar points. On the other hand, the Lind et al. (2014) displays included stereomotion as well as monocular optic flow. Because stereomotion is higher order, it remained possible that the bootstrap model might apply in the case of the perceived slant of planar surfaces. So, we also investigated how different types of motion generated visual information, namely stereomotion (CDOT), monocular structure from motion (SFM), and the combination of CDOT, IOVD, and SFM, would mediate 3D slant perception under large continuous perspective change. Finally, if perspective change failed to generate accurate slant perception with planar surfaces, then we were also interested to see whether the pattern of results would persist when noncoplanar elements were added to the surfaces.

In Experiment 1, we tested slant perception with planar surfaces. Our results yielded a significant difference between the monocular SFM condition (viewed both monocularly and biocularly), and the two stereomotion conditions. As was expected given the need for noncoplanar points for structure-from-motion, the accuracy of judgments in the monocular SFM condition was poor and unaffected by the amount of perspective change. In contrast, judgments in the stereomotion conditions were affected by the amount of perspective change. Despite this, however, results in the stereomotion conditions failed to become accurate with ≥45° of perspective change. Thus, the results failed to replicate those of Lind et al. (2014).

Furthermore, the results in the CDOT and combined conditions were the same. This implied that the stereomotion was the effective source of information in the combined condition. Participants in both conditions tended to overestimate the slant, that is, judging the planar surfaces to be more upright and thus less depthy. With an increasing amount of perspective change, such overestimation became gradually attenuated but performance never became veridical. Thus, it is the perceived depth that increased with more perspective change rather than the accuracy of slant judgments. Reliability of judgments in these conditions remained relatively high and stable over different rotation amounts. In the monocular SFM condition, increased amounts of perspective change failed to produce any significant change in the consistent overestimation of slant. Nonetheless, the reliability of the performance increased as the rotation amount increased. The structure-from-motion models predicted the poor performance in the monocular conditions as a lack of perceived 3D relief structure due to the absence of noncoplanar points. If the bootstrap model of Lind et al. (2014) is correct, the persistent inaccuracy of judgments in the stereo-motion conditions should also be attributed to a lack of perceived 3D relief structure. Experiment 2 was performed to investigate these possibilities.

In Experiment 2, we introduced noncoplanar elements to the surfaces, namely, nine small evenly distributed cuboids and tested judgments of slant with monocular SFM and combined visual information. Results from this manipulation were very different, as performance in both visual conditions exhibited a substantial increase in accuracy similar to the results of previous studies testing perception of 3D metric shape. More specifically, the large difference between monocular and combined conditions when using planar surfaces ceased to exist with noncoplanar elements on the surfaces. Both visual conditions now produced regression slopes greater than 1 when the rotation amount was less than 45°. Judgments became accurate and remained so when the rotation amount was equal to or greater than 45°. This pattern replicated the findings for 3D shape perception in Lind et al. (2014) and indicated the importance of noncoplanarity for those results. This, in turn, provided additional confirmation for the bootstrap model. Once noncoplanar points were available in the rigid rotation of the surfaces, monocular SFM presumably allowed perception of 3D relief structure. With sufficiently large perspective change, application of affine operations allowed bootstrap of perception to the Euclidean structure of the surfaces, yielding accurate perception of 3D slant.

The results also showed that the slant that was accurately perceived in Experiment 2 must be the geographical slant of the surface rather than egocentric optical or local slant. In these displays, the local slant would vary strongly with the perspective changes, as the angle of the line of sight to the surface would inevitably change over the rotation. However, the judged slants in each trial for every participant faithfully reflect the geographical slant specified as the orientation of the surface with regard to the horizontal axis defined by gravity. Moreover, we have also found that sufficiently large perspective changes enable the perception of accurate 3D slant. Logically, if local slant was what was perceived, such large perspective changes would produce large variations in perceived slant, whereas the results indicated that the large perspective variations yielded more reliably accurate perception of slant. Thus, it must be the case that the geographical slant is perceived as invariant or constant throughout the rigid rotation of the surface. The slant remains constant over the changes in tilt that occur as the surfaces rotated rigidly around the vertical axis through the surface. Although there is research investigating perception of geographical slant, for instance studies by Proffitt and colleagues on perception of large scale hills (e.g., Proffitt, Bhalla, Gossweiler, & Midgett, 1995), the majority of the studies that focus on different forms of visual information have been cast as investigations of egocentric optical slant (using small surfaces). Once the study of information for slant includes SFM, the characterization of slant as egocentric rather than geographical no longer makes sense. The current studies have shown that large continuous changes in perspective were required to yield accurate perception of slant. These results suggest that future studies on slant should perhaps shift in focus to the question of information for geographical slant.

Moreover, slant perception has often been characterized as a part of the perception of the 3D shape of polyhedral objects because a polyhedral object consists of a joined set of planar surfaces, each at a different 3D orientation, and enclosing a volume. However, we found that noncoplanar surface structure was required to yield perception of accurate 3D slant. The current study used cuboids to add noncoplanar points while controlling for the potentially perturbing effects of varying surface orientation or simply adding 3D noise points. The results suggest that such 3D structure is essential for perceiving slant accurately. Although many would argue that slant perception precedes shape perception, our results in fact suggest that the latter might necessitate the former, that is, that shape perception is required for accurate slant perception, an intriguing idea that might merit further investigation.

References

Allen, B., Haun, A. M., Hanley, T., Green, C. S., & Rokers, B. (2015). The optimal combination of the binocular cues to 3d motion. *Investigative Ophthalmology & Visual Science, 56,* 7589–7596. http://dx.doi.org/10.1167/iovs.15-17696

Beck, J., & Gibson, J. J. (1955). The relation of apparent shape to apparent slant in the perception of objects. *Journal of Experimental Psychology, 50,* 125–133. http://dx.doi.org/10.1037/h0045219

Bingham, G. P., & Lind, M. (2008). Large continuous perspective transformations are necessary and sufficient for accurate perception of metric shape. *Perception & Psychophysics, 70,* 524–540. http://dx.doi.org/10.3758/PP.70.3.524

Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Research, 34,* 483–495. http://dx.doi.org/10.1016/0042-6989(94)90162-7

Flock, H. R. (1964). Some conditions sufficient for accurate monocular perceptions of moving surface slants. *Journal of Experimental Psychology, 67,* 560–572. http://dx.doi.org/10.1037/h0048481

Gibson, J. J. (1950). The perception of visual surfaces. *The American Journal of Psychology, 63,* 367–384. http://dx.doi.org/10.2307/1418003

Gibson, J. J., & Cornsweet, J. (1952). The perceived slant of visual surfaces-optical and geographical. *Journal of Experimental Psychology, 44,* 11–15. http://dx.doi.org/10.1037/h0060729

Hecht, H., & Savelsbergh, G. J. P. (2004). *Theories of time-to-contact*. Boston, MA: MIT Press.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4,* 967–992. http://dx.doi.org/10.1167/4.12.1

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition, 18,* 65–96. http://dx.doi.org/10.1016/0010-0277(84)90022-2

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31,* 1351–1360. http://dx.doi.org/10.1016/0042-6989(91)90056-B

Kaiser, P. K. (1967). Perceived shape and its dependency on perceived slant. *Journal of Experimental Psychology, 75,* 345–353. http://dx.doi.org/10.1037/h0025039

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43,* 2539–2558. http://dx.doi.org/10.1016/S0042-6989(03)00458-9

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A, Optics, Image and Science, 8,* 377–385. http://dx.doi.org/10.1364/JOSAA.8.000377

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

Lee, Y. L., & Bingham, G. P. (2010). Large perspective changes yield perception of metric shape that allows accurate feedforward reaches-to-grasp and it persists after the optic flow has stopped! *Experimental Brain Research, 204,* 559–573. http://dx.doi.org/10.1007/s00221-010-2323-2

Lee, Y. L., Lind, M., & Bingham, G. P. (2013). Perceived 3D metric (or Euclidean) shape is merely ambiguous, not systematically distorted. *Experimental Brain Research, 224,* 551–555. http://dx.doi.org/10.1007/s00221-012-3334-y

Lee, Y. L., Lind, M., Bingham, N., & Bingham, G. P. (2012). Object recognition using metric shape. *Vision Research, 69,* 23–31. http://dx.doi.org/10.1016/j.visres.2012.07.013

Lind, M. (1996). Perceiving motion and rigid structure from optic flow: A combined weak-perspective and polar-perspective approach. *Perception & Psychophysics, 58,* 1085–1102. http://dx.doi.org/10.3758/BF03206835

Lind, M., Lee, Y. L., Mazanowski, J., Kountouriotis, G. K., & Bingham, G. P. (2014). Affine operations plus symmetry yield perception of metric shape with large perspective changes (≥45°): Data and model. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 83–93. http://dx.doi.org/10.1037/a0033245

Nefs, H. T., O'Hare, L., & Harris, J. M. (2010). Two independent mechanisms for motion-in-depth perception: Evidence from individual differences. *Frontiers in Psychology, 1,* 155. http://dx.doi.org/10.3389/fpsyg.2010.00155

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception & Psychophysics, 60,* 377–388. http://dx.doi.org/10.3758/BF03206861

Proffitt, D. R., Bhalla, M., Gossweiler, R., & Midgett, J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review, 2,* 409–428. http://dx.doi.org/10.3758/BF03210980

Sakata, H., Tsutsui, K., & Taira, M. (2005). Toward an understanding of the neural processing for 3D shape perception. *Neuropsychologia, 43,* 151–161. http://dx.doi.org/10.1016/j.neuropsychologia.2004.11.003

Sedgwick, H. A., & Levy, S. (1985). Environment-centered and viewer-centered perception of surface orientation. *Computer Vision Graphics and Image Processing, 31,* 248–260. http://dx.doi.org/10.1016/S0734-189X(85)80008-6

Shapiro, L. S., Zisserman, A., & Brady, M. (1995). 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision, 16,* 147–182. http://dx.doi.org/10.1007/BF01539553

Shioiri, S., Saisho, H., & Yaguchi, H. (2000). Motion in depth based on inter-ocular velocity differences. *Vision Research, 40,* 2565–2572. http://dx.doi.org/10.1016/S0042-6989(00)00130-9

Stevens, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics, 46,* 183–195. http://dx.doi.org/10.1007/BF00336800

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. http://dx.doi.org/10.1037/0096-1523.21.3.663

Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences, 8,* 115–121. http://dx.doi.org/10.1016/j.tics.2004.01.006

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48,* 419–430. http://dx.doi.org/10.3758/BF03211585

Todd, J. T., Christensen, J. C., & Guckes, K. M. (2010). Are discrimination thresholds a valid measure of variance for judgments of slant from texture? *Journal of Vision, 10,* 1–18. http://dx.doi.org/10.1167/10.2.20

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. http://dx.doi.org/10.3758/BF03194781

Todd, J. T., Oomes, A. H. J., Koenderink, J. J., & Kappers, A. M. L. (2001). On the affine structure of perceptual space. *Psychological Science, 12,* 191–196. http://dx.doi.org/10.1111/1467-9280.00335

Todd, J. T., & Perotti, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception & Psychophysics, 61,* 1577–1589. http://dx.doi.org/10.3758/BF03213119

Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of three-dimensional space in the perceptual analysis of motion and stereo. *Perception, 24,* 75–86. http://dx.doi.org/10.1068/p240075

Wagner, M. (2006). *The Geometries of Visual Space*. New York, NY: Psychology Press.

Wagner, M. (2008). Comparing the psychophysical and geometric characteristics of spatial perception and cognitive maps. *Cognitive Studies, 15,* 6–21.

Wallach, H., & Moore, M. E. (1962). The role of slant in the perception of shape. *The American Journal of Psychology, 75,* 289–293. http://dx.doi.org/10.2307/1419614

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience, 8,* 820–827. http://dx.doi.org/10.1038/nn1461

Appendix A

On Affine Structure

The term "affine structure" has been utilized in a large number of publications dealing with structure from motion during the last three decades. It is important to note, however, that there are two closely related, but not identical, lines of inquiry in which this term has been used. One of them is represented by Koenderink and van Doorn (1991) where it is shown that a large number of visual tasks involving 3D judgments can be performed by only using affine geometry. The article by Todd and Bressan (1990) represents the other line. Here the starting point is a visual system that can measure distances (angles) on a 2D projection screen, and the authors showed that such a visual system can reliably estimate the 3D properties of a viewed scene in terms of a family of solutions. The difference between this family of solutions and the metrically defined[1] 3D properties of the viewed scene is an unknown stretching or compression of the scene along the line of sight. Such a transformation between the viewed 3D world and the recovered structure preserves a number of properties which these authors refer to as "affine properties" (Todd & Bressan, 1990, p. 421). However, they also immediately point out:

> We shall refer to these properties hereafter as *affine structure*, but it is important not to be misled by this nomenclature. The inherent ambiguity in 2-frame apparent motion sequences is restricted to affine stretching transformations along the z-axis, which does not encompass the entire class of possible affine transformations. Thus, two objects that are affine equivalent in the more general sense could still be discriminated from 2-frame apparent motion sequences if they are related by a stretching transformation in any direction that is not parallel to the z-axis. (Todd & Bressan, 1990, p. 421)

[1] Because only relative distances can be recovered by any SFM process, one can also use the term "similarity" in place of metric.

(*Appendices continue*)

The key here is as follows: A visual system that can measure metric distances in the image plane (i.e., angles) is able to discover all properties of a viewed 3D scene that are invariant over affine transformations from two-frame apparent motion sequences, *but also* to discover a larger set of properties of the viewed scene. This is the general difference between the general affine structure as specified in Klein's hierarchy and a mere affine stretching along the line of sight.

Empirical work based on the insights from both lines of inquiry have focused on two cases. One is finding whether observers can reliably judge properties of viewed 3D scenes that are invariant under affine transformations, which they can (e.g., Todd et al., 2001). The other case is whether observers can reliably judge properties of viewed 3D scenes that are only invariant under metric (or similarity) transformations, which they cannot (e.g., Todd & Norman, 2003). However, there are few studies that have investigated whether observers can reliably judge properties of a viewed 3D scene that are preserved under unknown stretching or compression along the line of sight and not invariant under general affine transformations.

Todd and Bressan (1990) used orthographic projection in their analysis. However, such an analysis presumes that the distance between a viewed scene and the observer does not change during the analysis. In real situations, this is not a reasonable assumption and the literature on our ability to judge time-to-contact also shows that it is not (e.g., Hecht & Savelsbergh, 2004). A solution to this is to use scaled orthographic projection that, on top of the orthographic projection, allows for a uniform scaling of the 2D image due to motion toward or away from the observer. When analyzing this case, it can be shown (e.g., Lind, 1996; Shapiro, Zisserman, & Brady, 1995) that the same type of depth map that Todd and Bressan (1990) describe can also be reliably found in this case, that is, a depth map with an unknown scaling along the line of sight. The cost for this is that the analysis requires at least four noncoplanar points instead of the three points needed in the pure orthographic case.

In the work presented in this article, as well as that in Lind et al. (2014), a recovered depth map of the Todd and Bressan (1990) type is presumed and not a general extraction of information invariant under affine transformation only. There is no accepted term for this type of depth map. Koenderink and van Doorn (1991) have used the term "relief" although "a relief" in their work can contain an unknown amount of shear. With the proviso that any such shear is excluded, we here propose the use of the term *relief depth* to refer to the Todd and Bressan (1990) type of recovered 3D information.

Appendix B

The Bootstrap Model

Starting with two temporarily close 2D views of a rigid and moving 3D object, it has been shown that it is possible to reconstruct the 3D relief structure of an object (see for instance Koenderink & van Doorn, 1991; Lind, 1996; Shapiro et al., 1995). If a coordinate system is defined with the line of sight as its z-axis, such 3D relief structure also entails the actual physical x and y coordinates of identifiable points on the 3D object. However, the recovered z-values of these points are not the actual z-values. Instead, these z-values are scaled by a common, unknown factor. The bootstrap model simply further analyzes such recovered 3D relief structure, as the 3D object continues to move, to uncover the metric 3D structure of the object. The general logic behind the bootstrap model is to first analyze pairs of images to recover 3D relief structure and then recover the metric structure by, in a second step, further analyzing the relief structure.

The bootstrap model starts by identifying two points, A and B, on the perceived 3D relief structure of the object, equidistant to the observer. In that the recovered z-coordinates are scaled by a common, albeit unknown, factor, this simply entails choosing any two points sharing the same recovered z-values in the 3D relief structure. Because all recovered z-values are scaled by a common factor, two points that share recovered z-values also share the same actual z-values. By definition, the line AB is orthogonal to the line of sight. Subsequently, a Cartesian coordinate system (x, y, z) is established, where the origin is placed at point A and the line of sight is parallel to the z-axis. These points can then be said to have the following coordinates:

$$A = (0, 0, 0) \tag{1}$$

$$B = (x_b, y_b, 0) \tag{2}$$

Then one more point, C, needs to be identified different from points A and B. This point is described by the coordinates:

$$C = (x_c, y_c, z_{rc}) \tag{3}$$

where index "rc" denotes that it is measured in recovered 3D relief coordinates. The point C should be chosen so that the lines AB and AC are orthogonal. This is ensured by using the dot product and selecting C such that its coordinates fulfill the equation $x_b x_c + y_b y_c = 0$.

(*Appendices continue*)

Then let the object undergo further arbitrary rigid rotation apart from the trivial case of pure rotation around the line of sight. Lines AB and AC on the physical object will remain orthogonal to one another but their world coordinates as well as their respective recovered values in 3D relief coordinates will have changed. By again defining a Cartesian coordinate system with its origin in the new position of point A and its z-axis parallel to the line of sight, the three points will now have these 3D relief coordinates:

$$An = (0, 0, 0) \tag{4}$$

$$Bn = (x_{bn}, y_{bn}, z_{rbn}) \tag{5}$$

$$Cn = (x_{cn}, y_{cn}, z_{rcn}) \tag{6}$$

Because for all three points, the x and y coordinates are identical for the physical and perceived shape also in this new view whereas the z-coordinates in the perceived shape are stretched or compressed by an unknown scaling factor, q, applying the inverse of this factor to z_{rbn} and z_{rcn} makes AB and AC on the perceived shape orthogonal. We have on the perceived shape:

$$AB = [x_{bn}, y_{bn}, q \cdot z_{rbn}]$$

$$AC = [x_{cn}, y_{cn}, q \cdot z_{rcn}]$$

$$\begin{gathered} \therefore \; AB \cdot AC = 0 \\ \Leftrightarrow x_{bn}x_{cn} + y_{bn}y_{cn} + q^2 \cdot z_{rbn}z_{rcn} = 0 \\ \Leftrightarrow q = (+)\sqrt{\frac{-(x_{bn}x_{cn} + y_{bn}y_{cn})}{z_{rbn}z_{rcn}}} \end{gathered} \tag{7}$$

As soon as the object moves to a new position (apart from a few degenerate cases, for instance that points B and A or points C and A are equidistant) one simply can in this manner find the scaling value along the new line of sight that brings the angle between AB and AC on the perceived shape to be 90°. Consequently, the metric structure of the perceived shape can be obtained by finding the value of q. Again, because the bootstrap model starts with the relief structure obtained from visual information according to previous models (i.e., Koenderink & van Doorn, 1991; Lind, 1996; Shapiro et al., 1995), the z values in the denominator of Equation (7) can be directly obtained. Under this conception, the most informative new view of the object would be when the direction of the bisection line of the angle between AB and AC is either parallel or perpendicular to the line of sight, that is, with a rotation of 45°.

Received May 18, 2017
Revision received March 27, 2018
Accepted March 27, 2018 ■