Depth-order violation in structure from motion

Julian Martin Fernandez & Bart Farell

Institute for Sensory Research, Syracuse University, Syracuse, NY, 13244, USA

Humans can recover the structure of a three-dimensional object from motion cues alone. Recovery of structure-from-motion (SFM) from the projected 2D motion field of a rotating object has been studied almost exclusively in the condition where the axis of rotation lies in the frontoparallel plane. Here we assess the ability of humans to recover SFM in the general case, where the axis of rotation may be tilted out of the frontoparallel plane. Using elliptical cylinders whose cross-section was constant along the axis of rotation, we find that, across a range of parameters, subjects accurately matched the simulated shape of the cylinder regardless of how much the axis of rotation is inclined away from the frontoparallel plane. We also find that subjects do not perceive the inclination of the axis of rotation veridically. These results violate a relationship between perceived angle of inclination and perceived shape that must hold if SFM is to be recovered from the instantaneous velocity field. The contradiction can be resolved if the angular speed of rotation is not consistently estimated from the instantaneous velocity field. This in turn predicts that variation in object size along the axis of rotation can cause depth-order violation in object shape perception. This prediction was verified using rotating circular cones as stimuli. Thus, the visual system reduces the complexity of computing structure from motion at the expense of introducing severe distortions in the recovered shape.

The trajectories of points on a surface rotating about a frontoparallel axis are linear and parallel to one another under orthographic projection. The trajectories of points on a surface rotating about an axis tilted out of the frontoparallel plane, however, are curved and do not bear a simple geometrical relationship to each other.

We investigated the ability of humans to recover SFM in this latter case. Subjects viewed a motion-defined elliptical cylinder textured with random dots. They then viewed a cylindrical cross section, which they adjusted to match the profile of the motion-defined cylinder. We define flatness (**F**) as the cylinder's depth-to-width ratio (i.e., **F**=1 for a circular cylinder). Figs. 1a-b show the ratio between perceived ($\mathbf{F_{obs}}$) and simulated ($\mathbf{F_{sim}}$) flatness averaged across the four subjects. $\mathbf{F_{obs}}/\mathbf{F_{sim}}$ does not change with the simulated angle of inclination (Fig. 1a), fluctuating narrowly between the values 0.83 and 0.88. However (Fig. 1b) $\mathbf{F_{obs}}/\mathbf{F_{sim}}$ does depend on the simulated flatness. In all cases, cylinders were perceived as flatter than simulated, but the difference was large for flattened cylinders ($\mathbf{F_{sim}}<1$).

Results shown in Figs. 1 a and b are averages of four different conditions combining short and sustained dot lifetimes and occluded and unoccluded surface boundaries. The

short dot lifetime eliminates potential dot-density cues and the occluder eliminates potential boundary cues. Dot lifetime had little effect on performance (Fig. 1c). The presence of an occluder had a small but noticeable effect, which might be due to the greater uncertainty about the cylinders' width when an occluder is present. Thus, if subjects judge the cylinder to be wider when its boundaries are occluded, then the cylinder's width-to-depth ration will be increased and $\mathbf{F_{obs}/F_{sim}}$ will be reduced. However, the effect of the occluder is small. The important conclusion from Fig. 1c is that subjects' performance reflects pure motion cues, not dot-density or boundary cues.

For rotations around an axis on the frontoparallel plane, our result that perceived shape is not veridical is in agreement with results of previous studies. According to these studies, SFM is recovered from the first-order velocity field, which implies that shape is recoverable only up to a scaling factor in depth[1-4]. Hence, accuracy is low for judgments requiring veridical perception of Euclidean metric structure, like judgments of lengths or angles[5-11]. Conversely, accuracy is high for judgments of an object's affine structure, such as depth order between pairs of points, parallelism between lines defined by pair of points on the surface, and coplanarity among points[8,12]. Violations of affine structure have also been reported[13].

Our data show that recovered depth relative to the axis of rotation does not depend on the angle of inclination. The standard procedure in studies of SFM is to assess recovered depth along the line of sight. Depth along the line of sight is the same as depth relative to the axis of rotation only when the axis of rotation is frontoparallel. Therefore, our result means that the ratio between perceived and simulated depth *along the line of sight* does change with the angle of inclination (see Eq. 1 below). To obtain a quantitative measure of this change, we need to assess the perceived inclination of the angle of rotation.

Thus, in a second experiment, we assessed subjects' ability to recover the inclination of the axis of rotation from the frontoparallel plane. The inclination was always misperceived, the perceived angle being much smaller than the simulated angle (Fig. 2a). The relationship between perceived and simulated inclination is linear, with the slope varying across subjects. This result agrees with previous findings[15] (see also ref. 16), although our slopes are somewhat smaller, presumably due to differences in stimuli and procedure.

The ratio between perceived and simulated depth **d** along the line of sight is

$$\mathbf{d}=(\mathbf{F_{obs}/F_{sim}})(\cos\theta_{sim}/\cos\theta_{obs}), \quad (1)$$

where θ_{sim} and θ_{obs} are the simulated and perceived angle of inclination of the axis of rotation, respectively. Fig. 2b shows this ratio **d**, averaged across the 4 subjects, as a function of θ_{sim} . Perceived depth becomes a smaller fraction of the simulated depth as the angle of rotation deviates from the vertical. The results agree well with those from a previous study[14] using elongated ellipsoids.

There is a major difference between rotations around a vertical axis and rotations around an axis inclined towards the observer. Vertical rotations can be considered a degenerate case of general rotations. They satisfy, under orthographic projection, the following relationship between relative depth and relative speeds[17]:

$$\Delta v = -\Omega\, \Delta Z/Z_0 \quad (2)$$

where Ω is the angular speed of rotation, Z_0 is the distance of the observer to the axis of rotation, and Δv and ΔZ are the differences in retinal velocity and distance, respectively,

between any two points on the object. As mentioned above, the instantaneous velocity field yields no information about Ω and as a consequence object shape is recoverable only up to a scale factor in depth. The visual system is thus free to set Ω, which in general it does non-veridically, resulting in the non-veridical recovery of the object's shape.

When the axis of rotation is not on the frontoparallel plane, motion is no longer rectilinear, and the 3D structure of the object under orthographic projection must satisfy the following two relations:

$$\Delta Z = Z_0 \ \Delta v_x / (\Omega \cos\theta) + Z_0 \ \Delta y \tan\theta \quad (3)$$

and

$$\Omega = -\nabla_x v_y / \sin\theta, \quad (4)$$

so that

$$\Delta Z = (-\Delta v_x / \nabla_x v_y + \Delta y) Z_0 \tan\theta, \quad (5)$$

where, as shown in Fig. 3a, Z_0 is the distance to the object, Ω is the angular speed of rotation, θ is the perceived angle of inclination of the axis of rotation ($0°<\theta<90°$) from the frontoparallel plane, and $\nabla_x v_y$ is the gradient in the horizontal direction of the vertical component of the retinal velocity, v_y. Eq. 5, which implicitly assumes rigidity, gives the difference in depth ΔZ between any two points on the object from their differences in horizontal retinal velocity Δv_x and angular vertical position Δy.

Notice that now Ω can be recovered from the instantaneous velocity field (Eq. 4), contrary to the vertical rotation case, using the now available vertical component of the retinal speed, v_y. However, what it is not observable now from the instantaneous velocity field is the inclination of the axis of rotation θ. Thus again we have a one-parameter family of recoverable shapes that differ by a scaling factor in depth—Ω for vertical rotations and θ for inclined rotations. In principle the visual system can set θ to an arbitrary value.

Thus, our result showing that observers do not perceive θ veridically is expected if they recover SFM only from the instantaneous velocity field. But this failure to perceive θ veridically, together with the fact that flatness was close to veridical for $\mathbf{F_{sim}} \geq 1$, violates Eqs. 3 and 4. This is easily demonstrated; it can be seen intuitively by considering that if shape relative to the axis of rotation is recovered almost veridically (i.e., for $\mathbf{F_{sim}} \geq 1$), then the angle of inclination should have been recovered almost veridically, too. Neither the Euclidean nor the affine structure was recovered, though the instantaneous velocity field carries information about affine structure. Thus, the object's shape is recovered with distortions beyond a simple scaling in depth.

This violation of Eqs. 3 and 4 suggests that Ω, the angular speed of rotation, is not estimated from the instantaneous velocity field (Eq. 4). Let us introduce the factor λ in Eq. 3 to quantify the non-veridical Ω that is estimated by whatever heuristic observers' use in place of the instantaneous velocity field. Then the recovered structure becomes:

$$\Delta Z = (-\lambda \Delta v_x / \nabla_x v_y + \Delta y) Z_0 \tan\theta \ . \quad (6)$$

The factor λ can be seen as a scaling factor for depth relative to the slanted plane that includes the axis of rotation. For instance, if $\lambda=2$, all distances to this plane double (see Fig. 3b). From Fig. 3b it is easy to see that for $\lambda \neq 1$ violations of depth order should be observed: the relative depth of any two points on the bottom half of the object will be perceived as reversed, whereas pairs of points in the top half, where the cylindrical radius

is constant, will be perceived with the correct depth order. Thus, a uniform distortion relative to the axis of rotation results in a non-uniform distortion across the image plane, in this case along the vertical direction. This contrasts with the uniform distortions from depth scaling found for frontoparallel axis of rotation, where relief structure of the object is always conserved. A depth order violation when $\lambda \neq 1$ can only occur when the shape of the object changes along the axis of rotation, as in the examples shown in Fig. 3. In fact, it can be shown that any value of $\lambda \neq 1$ will result in a depth order violation for any pair of points satisfying certain simple geometric relationships.

Thus, for an inclined axis of rotation we have two independent free parameters, θ and λ. A value of $\lambda \neq 1$ implies that the recovered depth structure is not affine. For a frontoparallel axis of rotation, by contrast, we have only one free parameter, namely Ω in Eq. 2, so depth structure can be recovered up to an affine transformation in depth.

From Exps. 1 and 2 it is possible to estimate λ. Values obtained, shown in Fig. 3c, depart greatly from the value of one. This result implies large departures from affinity in the recovered shape. However, the way in which λ was estimated for Fig. 3c is susceptible to systematic errors in the estimation of $\theta_{obs,}$ which could be strongly amplified (see Eq. S11 of Supplementary information), especially for small θ_{obs}.

There is another way to estimate λ that is independent of any measurement of the angle of inclination and provides a more precise estimate. If subjects perceive surface slant free of biases, then, if $\lambda=1$, a surface simulated as vertical must be perceived as vertical regardless of the perceived value of the inclination of the axis of rotation, whereas $\lambda \neq 1$ implies a depth-order violation. In such a case, we can estimate λ from the slant required for the simulated surface to appear vertical. Control experiments are necessary to take into account any intrinsic biases subjects may have in perceiving surface slant.

Subjects viewed a simulated rotating cone attached to a circular cylinder, with an axis of rotation inclined by 45° from the frontoparallel plane. Subjects' task was to indicate the slant of the upper cone's surface relative to the frontoparallel plane (Fig. 4a). Subjects were offered a 2-alternative forced choice: top of the cone near vs. far with respect to the bottom of the cone. Control experiments (Fig. 4a) assessed subjects' capacity to perform the task and their bias in perceiving slant. Control surfaces rotated about a vertical axis and thus were not subject to errors due to non-affinity when $\lambda \neq 1$ (Eq. 6). Any misperception of the surface's slant in this case can be attributed to the observer's intrinsic bias.

Fig. 4b shows that three of the four subjects do not have significant bias for control stimuli; they perceive the surface as vertical when it is simulated to be so. One of the subjects does have a bias, but in the opposite direction needed to explain the experimental result. Obtained λ values, also shown in Fig. 4b, significantly differ from 1.0 for all four subjects, implying depth-order violations. λ exceeds 1.0 in all cases, implying that subjects see the top of the cone as closer than the bottom (i.e., $\gamma>0$). This perceived shape conflicts with the physical shape of the simulated cone, which had the bottom closer. In the control condition, one subject shows a bias to perceive the top of the cone as far relative to the bottom (i.e., $\gamma<0$) and the others show a slight tendency in that direction. Thus, the λ's obtained are not an artifact of the subjects' intrinsic biases. Values for λ obtained from this experiment are smaller than those obtained earlier from the inclination

matching task (Fig. 3c), although their rank across subjects is conserved (i.e., subject S2 has the largest λ and so on). Besides differences in task and stimuli, reasons for the smaller λ values include subjects' bias to perceive inclination in the opposite direction to that of the depth order violation, and, as already mentioned, the potential underestimation of θ_{obs} in Exp. 2.

Depth-order violations have been reported before[13], using hinged planes (open books) rotating around a vertical axis as stimuli, but the cause was different. The two planes were perceived as separate objects rotating with different angular speeds, and the depth-order violation arose because each plane was recovered using a different scaling factor in depth. In our case, the depth-order violation is intrinsic to the object. A single angular speed and depth-scaling factor is used to recover the structure of the whole object.

Structure recovered from rotations around an axis in the frontoparallel plane differs from the more general case of structure recovered from rotations around an arbitrary axis. The former preserves affine structure, while the later, in general, does not. This results from the human's inability to extract rotational speeds from the instantaneous velocity field. A possible reason behind this inability could be reducing the computational complexity of structure from motion computation. The reduction in complexity comes at the cost of severe distortions in the recovered shape.

Methods

The stimuli consisted of moving dark 'dots' (3.4' x 3.4' squares) on a moderately bright background. An attenuator used to boost luminance resolution to approximately 12 bits drove only the monitor's green gun. Thus the color of both the dots and the background was green. Dot and background luminances were 2 cd/m^2 and 29 cd/m^2, respectively. The motion was shown at the refresh rate of the monitor (75 Hz). Stimuli were monocularly viewed at an optical distance of 94 cm, using a chin rest to stabilize head position. No separate fixation point was required, but was present as a 10' x 10' square between trials to guide fixation to the center of the display. Binocular viewing was also tested in pilot experiments and produced no significant differences from monocular viewing.

Perceived depth from SFM displays often increases to an asymptotic value with increases in parameters such as dot density, stimulus' duration, dot lifetime, range of oscillation, and speed of rotation. Pilot experiments were performed to verify that the parameters used in our experiments resulted in perceived depth reaching asymptotic values.

One experienced subject (S1), and three naïve inexperienced subjects (S2 to S4) were used in Experiment 1. Three subjects (S1 to S3) were used in Experiment 2, and four subjects, two experienced (S1 and S5) and two naïve and inexperienced (S2 and S3) were used in Experiment 3. Subject numbering reflects subject identity across experiments. All subjects gave written informed consent for participation in the study, which was approved by the Institutional Review Board of Syracuse University.

Experiment 1. We created 28-frame (373 ms) movies of individual elliptical non-transparent rotating cylinders defined by moving dots (n=300). Cylinders rotated at a constant angular speed (135 °/s). The axis of rotation corresponded with the cylinder's

longitudinal axis. Rotation was confined to an arc of 50°; one cycle of rotation brought the cylinder to its starting position after rotating a half-cycle of 50° in one direction and another half-cycle of 50° in the opposite direction. Test stimuli consisted of a 1.49 s movie, obtained by displaying two cycles of rotation, each cycle formed by playing in forward and reversed order the same 28-frame sequence. The axis of rotation could be tilted away from the frontoparallel plane by an angle of 0°, 20°, 40°, 60° or 80°. However, the axis of rotation was always within the subjects' mid-sagittal plane, that is, its projection onto the frontoparallel plane was vertical and centered on the subject's line of sight. The cylinders cross section was elliptical. The cylinders' path was such that one of the main axes of the ellipse crossed the mid-sagittal plane halfway through the cylinders' half-cycle of rotation. Cylinders' flatness was defined as the ratio between the two main axes of the ellipse. In computing the ratios, the axis used in the numerator was the axis that crossed the mid-sagittal plane halfway through the cylinders' half-cycle of rotation. Flatness values of .5, .75, 1, 1.5 and 2 were used. Initial position of the dots defining the cylinder was random in the projected view and therefore was not random on the 3D cylinder's surface. To eliminate density cues, which become more important the more elongated the cylinders, two different dot conditions were tested: sustained dots vs. finite dot lifetime. For sustained dots, dots remained attached to the surface during the whole movie. For finite dot lifetimes, a dot that disappeared was replaced by a new dot at a random position in the projected view. Dot lifetimes were distributed. Let us define, for a given movie, the maximum dot lifetime, lt_{max}. Then at every frame in the movie we replaced n/lt_{max} dots (rounded to the nearest integer) by new randomly positioned dots, where n=300 is the total number of dots. A different set of dots was replaced on each frame until all dots were used, and the process was started again in the same order until the end of the movie. To keep dot lifetimes inside the range of asymptotically perceived depth, lt_{max} covaried with flatness, being 24, 18, 12, 9, and 6 frames for flatnesses of .5, .75, 1, 1.5 and 2, respectively. These values were low enough at each flatness value to also ensure that no dot density cues were present, as assessed by judging individual frames, which did not allow the identification of the cylinder's shape.

Cylinders were occluded at the top and bottom by dark rectangular maskers (2 cd/m^2, 4°(H) x 2°(V)) to cover these borders. The projected vertical cylinder height between the masks was 4°. Cylinders horizontal size was 4° when located midway through the half-cycle. There were conditions in which side maskers covered the vertical borders. The size of these maskers was made just large enough to hide the cylinders' boundaries so they would not be seen to expand or contract laterally during rotation. Maskers' size thus never exceeded 10% of the cylinders' width.

Subjects' task was to view the rotating cylinder and then to adjust a cylindrical cross-section to match the profile of the cylinder previously seen in the movie. Subjects could repeat the 1.49 s movie as many times as they wanted during the adjustment procedure. The cylindrical cross-section had the same horizontal angular size as the moving cylinder; subjects adjusted the cross section by clicking with a mouse on "+" and "-" symbols located on the screen below the cross section. Each run consisted of 20 trials obtained by randomly selecting without replacement one of the 100 different stimuli obtained by combining the five flatnesses, the five axis inclinations, the two masking conditions and the two dot lifetime conditions. Each subject completed 40 runs, and so adjusted each stimulus 8 times.

Experiment 2. Experiment 2 differed from Experiment 1 in three ways: (1) Only the circular cylinder was used. (2) The tested inclinations of the axis of rotation were 0°, 35°, 45°, 65° and 85°. (3) Subjects were required to adjust the orientation of a line appearing in the screen to match the inclination of the axis of rotation.

Experiment 3. We created 112-frame (1.49 s) movies of rotating geometric structures each defined by 1200 moving dots. The structures were opaque and rotated at a constant angular speed of 33.75 °/s about an axis of rotation along their longitudinal axis. The angular speed of rotation was made slower than in the two previous experiments in order to keep retinal speeds in the same range, as the stimulus was now larger. Rotation was in both directions and confined to a 50° arc. Test stimuli consisted of a 5.96 s movie, obtained by repeating back and forth the 1.49 ms movie twice. The axis of rotation was tilted 45° away from vertical, with the upper part of the axis tilted towards the observer. Initial positions of the dots defining the structure were random in the projected view. Maximum dot lifetime, $lt_{max,}$ was 18 frames, with a vanished dot replaced by a new dot placed at random in the projected view, using the procedure explained in Experiment 1. The structures were viewed through a 8°x8° square window. The structure's cross section was circular, with its half bottom forming a cylinder and its half top forming a cone. For test stimuli, the radius of the cylinder was 6°, and that of the cone increased from 6° at a linear rate such that the cone's surface within the mid-sagittal plane was inclined with angle γ relative to the vertical (Fig. 4a). For control stimuli the only difference was in the top half. Here both the axis of rotation and the longitudinal axis of the cone were vertical, thus bottom and top halves had different axes of rotation. The radius along the vertical axis was variable, increasing or decreasing linearly from 6° so that the cylinder's surface within the mid-sagittal plane was inclined with angle γ relative to the vertical (Fig. 4a). For both test and control stimuli all four boundaries were occluded from view by maskers.

For test stimuli, the psychometric functions recording the proportion of trials in which subjects reported the bottom of the cone as "near", as a function of the parameter β, were obtained using the method of constant stimuli, where $\beta=(\tan\gamma/\tan\theta)+1$ (see Supplementary information, Eq. S13). β was used as a parameter instead of γ because it is more closely related to λ, although this choice is arbitrary and should have no influence on the results. The five β values were individually selected for each subject on the basis of pilot data in order to optimize the range of the psychometric function. A cumulative normal was fit to the psychometric functions by probit analysis, from which the points of perceived verticality were obtained. This point was defined as the point at which the response rate for bottom-seen-as-near was 50%.

Subjects' task was to indicate whether the top or the bottom of the cone appeared as near. To ease this task, two static square markers (10' x 10', 2 cd/m^2) were placed at the upper and lower extremes of the line defined by the intersection of the cone's surface and the mid-sagittal plane. It was this line, at the horizontal midpoint of the cone, that subjects were to judge. Each run consisted of 50 trials obtained by randomly choosing one of the 5 different stimuli that differed in β. Each subject totaled at least 2 runs, so as to judge each stimulus at least 20 times. The same procedure was followed for control

stimuli, but this time using γ as a parameter instead of β, which is not defined when the axis of rotation is vertical.

References

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

2. Todd, J. T. Theoretical and biological limitations on the visual perception of 3D structure from motion. In *High-level Motion Processing—Computational, Neurophysiological and Psychophysical Perspectives* (ed. T. Watanabe) 359-380 (MIT Press, Cambridge, MA, 1998).

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

4. Werkoven, P. & van Veen, H. A. Extraction of relief from visual motion. *Perception & Psychophysics* **57**, 645-656 (1995).

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

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

7. Cornilleau-Peres, V., & Droulez, J. Visual perception of curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics* **46**, 351-364 (1989).

8. Todd, J. T. & Bressan, P. The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics* **48**, 419-430 (1990).

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

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

11. Hogervorst, M., Kappers, A. M. L. & Koenderink, J. J. Perception of metric depth from motion parallax. *Perception* **22**, supplement 101 (1993).

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

13. Domini, F. & Braunstein, M. L. Recovery of 3-D structure from motion is neither Euclidean nor affine. *J. Exp. Psychol.: Human Perception and Performance* **24** (4), 1273-1295 (1998).

14. Loomis, J. M. & Eby, D. W. Perceiving structure from motion: Failure of shape constancy. In *Proceedings of Second International Conference on Computer Vision* (Washington, D.C.: Computer Society of the IEEE, 383-391, 1988).

15. Pollick, F. E., Nishida, S. , Koike, Y. & Kawato, M. Perceived motion in structure from motion: pointing responses to the axis of rotation. *Perception & Psychophysics* **56**, 91-109 (1994).

16. Caudek, C. & Domini, F. Perceived orientation of axis of rotation in structure-from-motion. *J. Exp. Psychol.: Human Perception and Performance* **24**, 609-621 (1998).

17. Fernandez, J. M., Watson, B. & Qian, N. Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research* **42**, 883-898 (2002).

Figure captions

Figure 1. Recovered depth. a) The ratio $\mathbf{F_{obs}/F_{sim}}$ between perceived and simulated flatness, averaged across the four subjects and across the different simulated flatnesses, does not change with the simulated angle of inclination. b) $\mathbf{F_{obs}/F_{sim}}$, averaged across the four subjects and across simulated angles, does depend on the simulated flatness. c) $\mathbf{F_{obs}/F_{sim}}$ as a function of viewing condition (dot lifetime and presence of an occluder). In all figures error bars are individual subject's s.e.m. averaged across subjects and conditions.

Figure 2. Effect of inclination. a) Perceived inclination of the axis of rotation as a function of the simulated one. A linear relationship is followed, in which perceived angles are substantially lower than those simulated. b) Ratio between perceived and simulated depth along the line of sight (**d**), as a function of the angle of inclination of the axis of rotation. In both figures error bars are s.e.m.

Figure 3. Non affine structure. a) Schema showing the nomenclature utilized. b) Side view (sagital cut) from the object shown in (a). A change in λ results in a depth order reversal of the bottom half, as all the distances to the axis of rotation (dotted line) double. c) Variation of λ with the simulated angle of inclination. Error bars are s.e.m.

Figure 4. Depth order violation. a) Experimental and control stimuli used in depth order violation experiments. γ is positive towards the observer (as in the control example). Drawings are not to scale. b) λ recovered from depth order experiments (black bars) and subjects' intrinsic bias (gray bars). Error bars are s.e.m.

Supplementary information

Derivation of Eqs. 3 and 4:

To derive Eq. 3, observe that, for an object rotating around a vertical axis, the trajectories of points are horizontal and the distance between a given point on the object and the frontoparallel plane that includes the axis of rotation is (under orthographic projection):

$$r_{axis} = Z_0 \, v / \Omega \qquad \text{(S1)}$$

where Z_0 is the distance between the object and the observer, **v** is the retinal speed of the given point and Ω is the angular speed of rotation (for a derivation see Fernandez et al, 2002). In what follows, it is assumed that the object is distant enough from the observer so that Z_0 approximates the distances of all points on the object.

Let us consider now the same object when the axis of rotation makes an angle θ with respect to the frontoparallel plane. The distance between a given point on the object and an arbitrary reference frontoparallel plane will be (see Fig. S1a) :

$$Z = Z_1 + Z_2 . \qquad \text{(S2)}$$

From simple trigonometry we have (for $0°<\theta<90°$):

$$Z_2 = r_{axis} / \cos\theta \qquad \text{(S3)}$$

and

$$Z_1 = Y \tan\theta \qquad \text{(S4)}$$

where Y is the distance between the given point and a horizontal plane passing through the line of sight.

In the case of an inclined axis of rotation, the trajectories of points on the surface are no longer horizontal but curved. Let (X', Y', Z') be the coordinates of a given point in a coordinate system in which the axis Y' coincides with the axis of rotation (Fig. S1b). Let us assume that the axis of rotation is inclined by θ relative to the frontoparallel plane, and let (X, Y, Z) be the coordinates in our canonical reference frame in which Y is vertical (Fig. S1b)—note that, for simplicity, we use the same labels for axes and the coordinates of a point relative to those axes. Then a simple rotation of coordinates gives:

$$X = X'$$

$$Y = Y' \cos\theta - Z' \sin\theta$$

$$Z = Y' \sin\theta + Z' \cos\theta .$$

Only the first two equations need to be considered for our purposes. If we differentiate both sides of these equations with respect to time, and then divide by Z_0 (and noticing that $v_{y'} = 0$ for rotations about axis Y'), we obtain:

$$v_x = v_{x'}$$

$$v_y = -v_{z'} \sin\theta .$$

Taking into account that $v_{x'}=v$ (Eq. S1) and $v_{z'} = \Omega x' = \Omega x$ (lower caps refer to angular variables):

$$v_x = v \quad \text{(S5)}$$

$$v_y = -\Omega x \sin\theta . \quad \text{(S6)}$$

From Eq. S6, after differentiation with respect to x, we obtain Eq. 4. Substituting from Eqs. S1, S3, S4 and S5, Eq. S2 becomes:

$$Z = Z_0 v_x / (\Omega \cos\theta) + Z_0 \, y \tan\theta \quad \text{(S7)}$$

where $y=Y/Z_0$ is the angular distance between the given point and a horizontal plane passing through the line of sight. Taking differences between any two points using S7 yields Eq. 3.

Obtaining λ from Exps. 1 and 2:

Let us consider two points on the rotating object, one obtained from the intersection of the line of sight and the object's surface (*p2*), and the other from the intersection of the line of sight and the axis of rotation (*p1*) (Fig. S2a). Let us write Eqs. 5 and 6 for these two points using the simulated axis of rotation and again using the perceived axis of rotation. By definition $\Delta y=0$, so we have:

$$\Delta Z_{sim} = -Z_0 \tan\theta_{sim} \Delta v_x / \nabla_x v_y \quad \text{(S8)}$$

for the simulated axis and

$$\Delta Z_{obs} = -\lambda Z_0 \tan\theta_{obs} \Delta v_x / \nabla_x v_y \quad \text{(S9)}$$

for the perceived axis.

From Eqs. S8 and S9 we obtain λ:

$$\lambda = \tan\theta_{sim} \Delta Z_{obs} / (\tan\theta_{obs} \Delta Z_{sim}) \quad \text{(S10)}$$

Using the relationship $r=\Delta Z\cos\theta$ (Fig. S2a) for simulated and perceived values of θ, and using $\mathbf{F_{obs}}/\mathbf{F_{sim}}=r_{obs}/r_{sim}$, Eq. S10 becomes:

$$\lambda = (\mathbf{F_{obs}}/\mathbf{F_{sim}})(\sin\theta_{sim}/\sin\theta_{obs}) \quad \text{(S11)}$$

Obtaining λ from Exp. 3:

Let us consider two points, *p1* and *p2*, on the rotating object, located on the same sagittal plane, but having different depths and heights relative to the observer (Fig. S2b). The line joining the two points meets with the vertical to form a perceived angle γ given by :

$$\tan\gamma = -\Delta Z / \Delta Y \quad \text{(S12)}$$

Using Eq. 6 this becomes:

$$\tan\gamma = \tan\theta \, (\lambda\beta - 1) \quad \text{(S13)}$$

where $\beta = \Delta v_x / (\nabla_x v_y \Delta y)$ is a quantity that depends only on the physical parameters of the stimulus, and thus can be set by the experimenter.

By definition, the surface is perceived as vertical when $\gamma=0$. In this case we can obtain λ as:

$$\lambda=1/\beta \quad \text{(S14)}$$

We can obtain β (and thus λ) at the point of perceived verticality from a psychometric function in which β is the independent variable, as described in *Methods*. Note that the

value of λ obtained in this way is independent of θ and γ, and does not require their measurement.

A depth-order violation occurs when the sign of the perceived γ is opposite the sign of the simulated γ.

$$\text{sign}(\gamma_{per}) \neq \text{sign}(\gamma_{sim}) \quad \text{(S15)}$$

By definition $\lambda=1$ for the simulated object, so Eq. S15 is equivalent to:

$$\text{sign}(\lambda\beta-1) \neq \text{sign}(\beta-1) \quad \text{(S16)}$$

This inequality gives three cases that result in depth-order violations:

$$\begin{array}{lll} \text{Case 1.} & \lambda<1/\beta & \text{and } \beta<0 \\ \text{2.} & \lambda>1/\beta & \text{and } 0<\beta<1 \\ \text{3.} & \lambda<1/\beta & \text{and } 1<\beta \end{array} \quad \text{(S17)}$$

Thus, any $\lambda\neq1$ will result in a depth-order violation for any pair of points satisfying Eqs. S17.

References

S1. Fernandez, J. M., Watson, B. & Qian, N. Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research* **42**, 883-898 (2002).

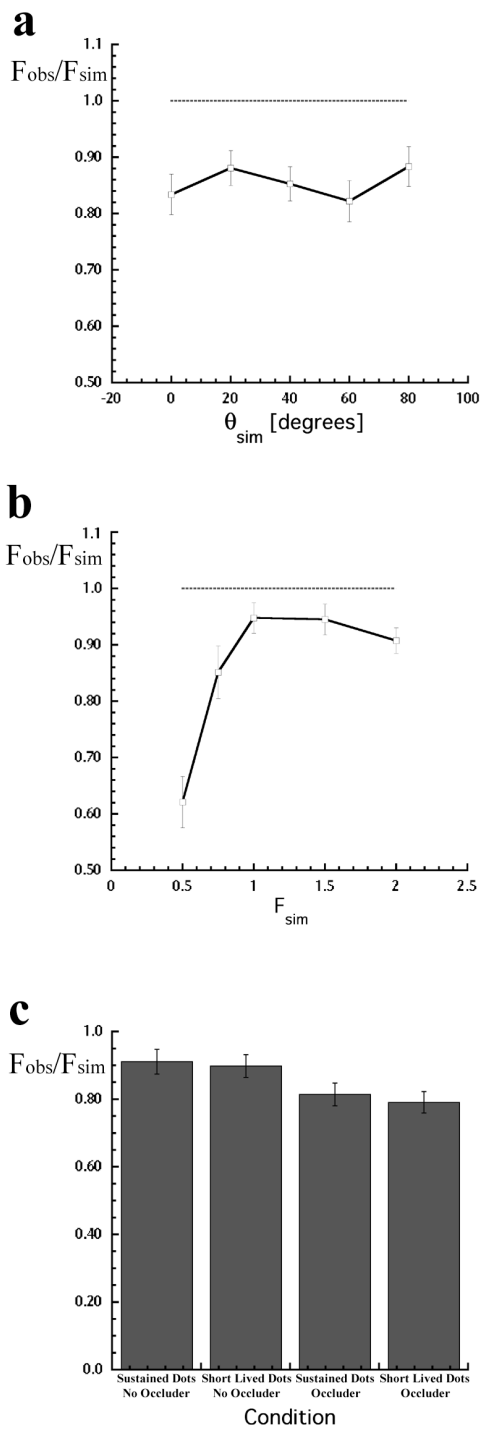

a
Fobs/Fsim
1.1
1.0
0.90
0.80
0.70
0.60
0.50
-20
0
20
40
60
80
100
θ sim [degrees]
b
Fobs/Fsim
1.1
1.0
0.90
0.80
0.70
0.60
0.50
0
0.5
1
1.5
2
2.5
F sim
c
Fobs/Fsim
1.0
0.80
0.60
0.40
0.20
0.0
Sustained Dots No Occluder
Short Lived Dots No Occluder
Sustained Dots Occluder
Short Lived Dots Occluder
Condition

Fig. 1

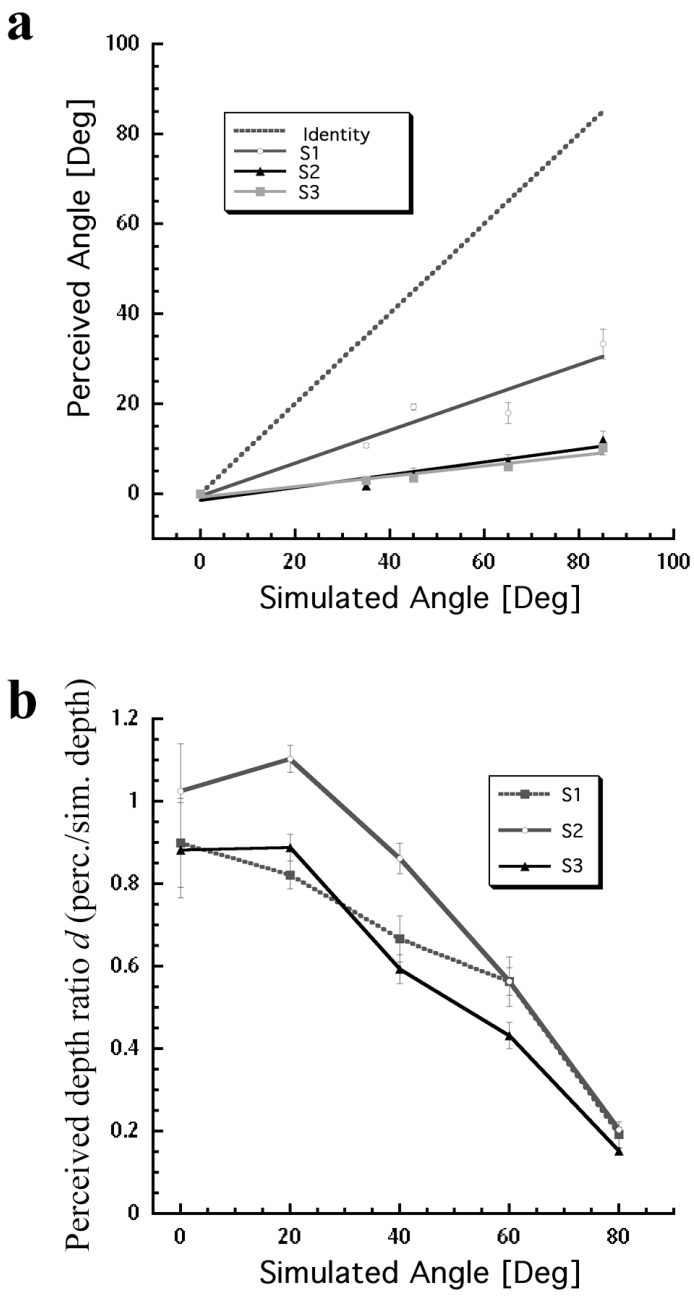

a
Perceived Angle [Deg]
Simulated Angle [Deg]
Identity
S1
S2
S3
0
20
40
60
80
100
b
Perceived depth ratio d (perc./sim. depth)
Simulated Angle [Deg]
S1
S2
S3
0
0.2
0.4
0.6
0.8
1
1.2
0
20
40
60
80

Fig. 2

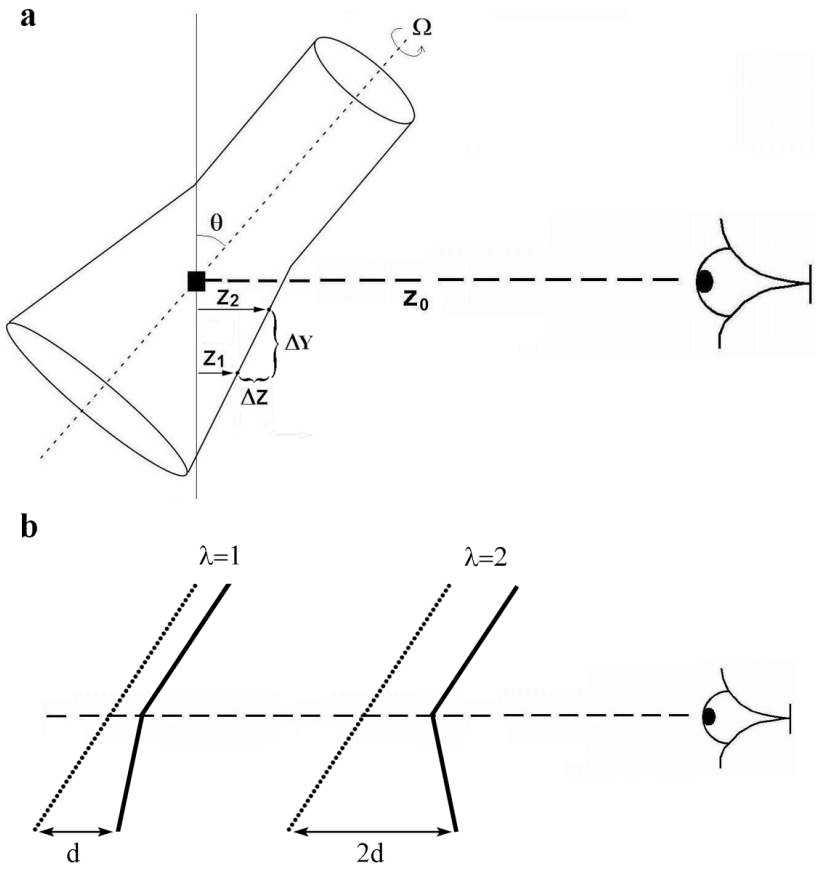
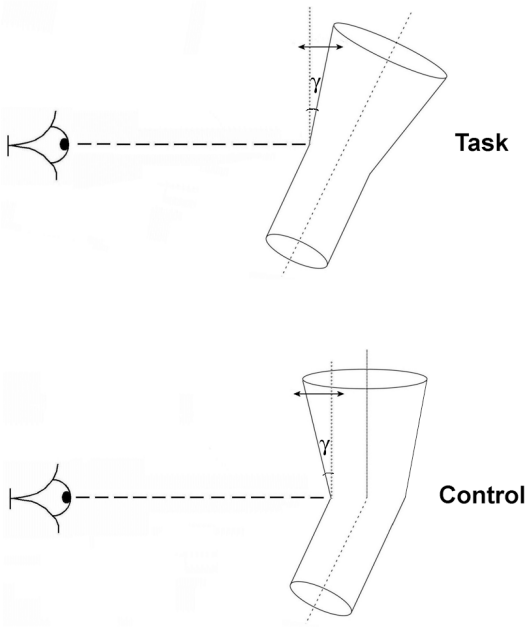

a
Ω
θ
Z2
Z0
ΔY
Z1
ΔZ
b
λ=1
λ=2
d
2d
c
λ
S1
S2
S3
20
15
10
5
0
0
20
40
60
80
Simulated Angle [Deg]

Fig. 3

a


γ
Task
γ
Control

b

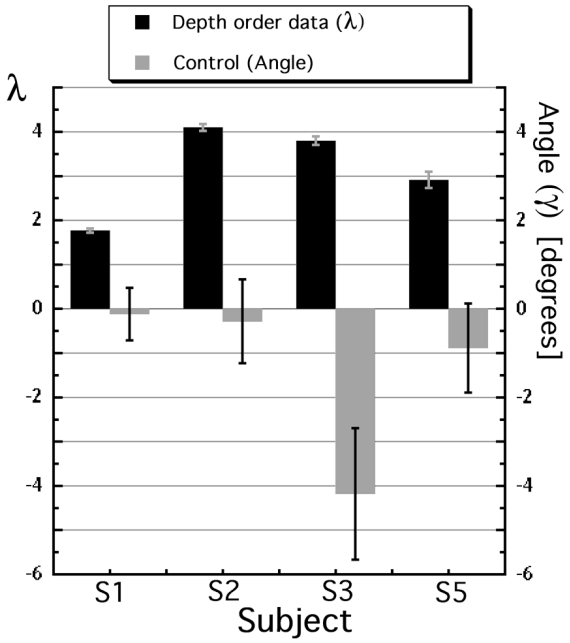

Depth order data (λ)
Control (Angle)
λ
Angle (γ) [degrees]
4
2
0
-2
-4
-6
S1
S2
S3
S5
Subject

Fig. 4

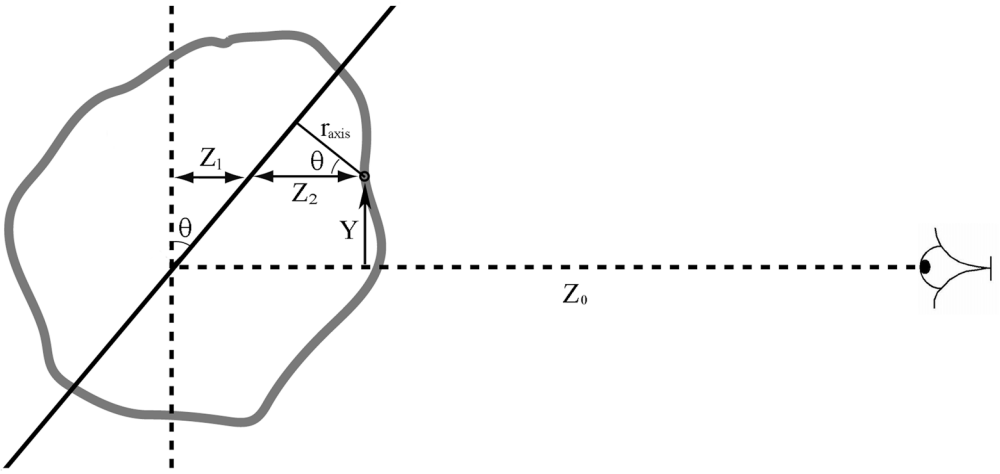


Fig. S1a

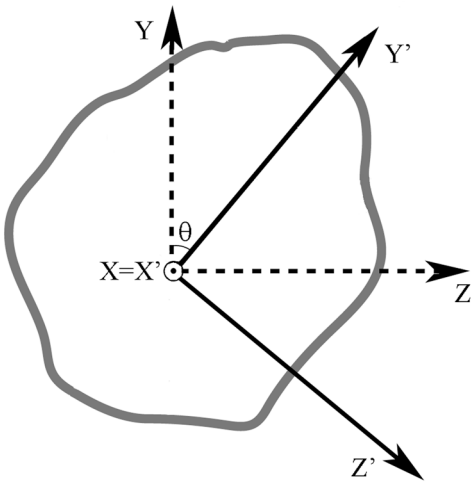


Fig. S1b

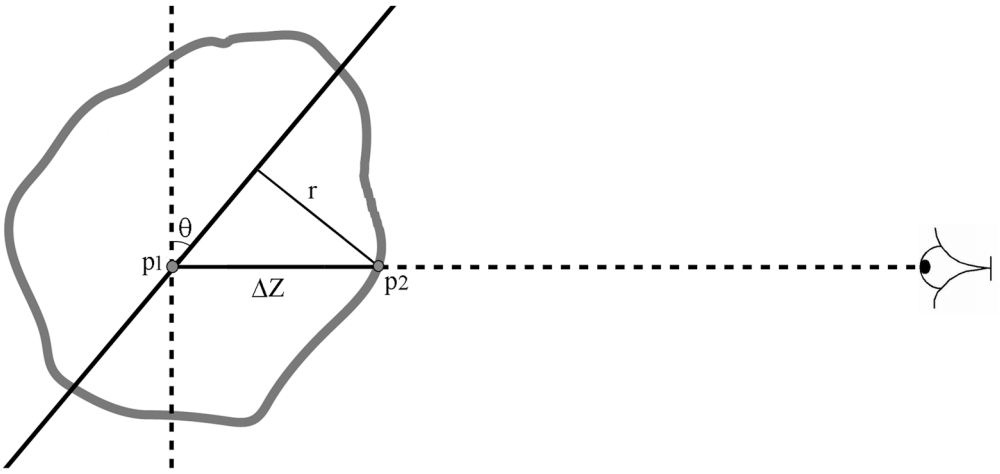


Fig. S2a

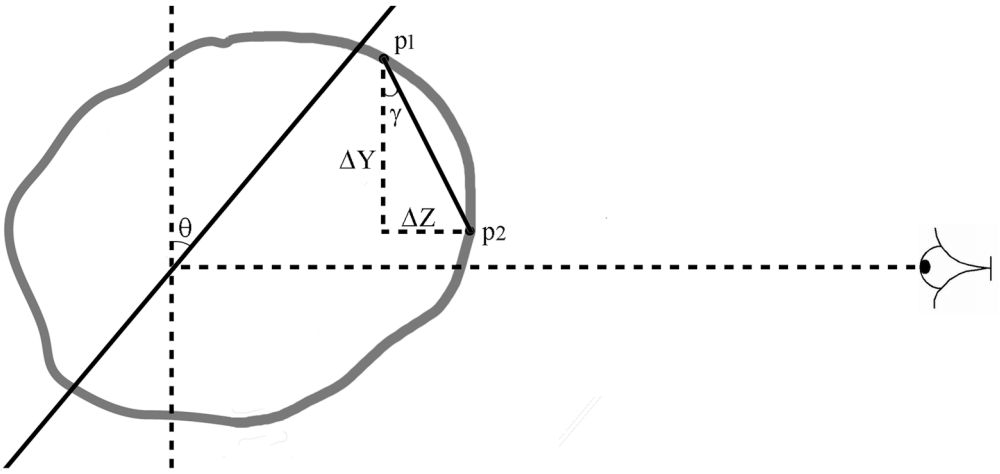


Fig. S2b