Discerning nonrigid 3D shapes from motion cues

Anshul Jain[1] and Qasim Zaidi[1]

Graduate Center for Vision Research, State University of New York College of Optometry, New York, NY 10036

Edited by Thomas D. Albright, Salk Institute for Biological Studies, La Jolla, CA, and approved December 9, 2010 (received for review October 28, 2010)

Many organisms and objects deform nonrigidly when moving, requiring perceivers to separate shape changes from object motions. Surprisingly, the abilities of observers to correctly infer nonrigid volumetric shapes from motion cues have not been measured, and structure from motion models predominantly use variants of rigidity assumptions. We show that observers are equally sensitive at discriminating cross-sections of flexing and rigid cylinders based on motion cues, when the cylinders are rotated simultaneously around the vertical and depth axes. A computational model based on motion perspective (i.e., assuming perceived depth is inversely proportional to local velocity) predicted the psychometric curves better than shape from motion factorization models using shape or trajectory basis functions. Asymmetric percepts of symmetric cylinders, arising because of asymmetric velocity profiles, provided additional evidence for the dominant role of relative velocity in shape perception. Finally, we show that inexperienced observers are generally incapable of using motion cues to detect inflation/deflation of rigid and flexing cylinders, but this handicap can be overcome with practice for both nonrigid and rigid shapes. The empirical and computational results of this study argue against the use of rigidity assumptions in extracting 3D shape from motion and for the primacy of motion deformations computed from motion shears.

optic-flow | structure-from-motion

All animals, such as fish, insects, birds, and reptiles, and many mechanical objects, such as cranes, earthmovers, harvesters, and bicycles, change shape as they move. Some shape changes (e.g., bending and straightening of legs) create forward propulsion, others such as torso flexes and head bobs serve to maintain balance, whereas others like head turns may be unrelated to object motion. Humans seem to be quite good at disentangling shape changes from object motion and characterizing motion as tumbling, rolling, swaying, stretching, leaping, spinning, flapping, dancing, kicking, bucking, jerking, sliding, gliding, tripping, or shaking. A large number of studies have examined human perception of rigid 3D shapes from motion cues (1–5); however, very few have examined nonrigid shape perception (6–8), and these have not dealt with what shapes are perceived.

Structure from motion models fall into three broad theoretical classes. In the incremental rigidity scheme (9, 10), the visual system assumes that the object is rigid or approximately rigid to extract the 3D shape. The perceptual system creates a 3D model of the object that is continuously updated to minimize nonrigid deformations across frames. This inverse optics approach seeks to compute the exact Euclidian (up to a scale and rotation) structure that would project to the 2D images. However, it is very sensitive to measurement noise and thus, requires accurate frame by frame feature tracking. In addition, often, the human perceptual system does not extract a rigid Euclidian or Affine structure that corresponds to the moving object (11, 12).

The most common retinal motions arise from the movements of observers, and in this case, local retinal velocities are inversely proportional to distances in a given scene (3, 13–15). Similarly, retinal velocities arising from an object translating in front of a stationary observer are inversely proportional to distances of different parts of the object. Models that assume that this principle is used by the visual system to extract depth from relative velocities will be called motion perspective models. Psychophysical (16–18) and physiological (19–22) evidence for relative velocity detectors suggests that they could play an intermediate role in computing 3D shape, and electrophysiological studies have implicated the middle temporal (MT) cortical area, which contains such neurons, as having a significant role in computing 3D structure from motion (23–27). This approach has been shown to be in general agreement with human perception of rigid objects (3, 28, 29) but has not been tested on nonrigid motion.

Under general conditions (i.e., motions of observers and objects that include rotations and shape changes of objects), motion leads to deformations of the pattern of retinal image velocities (optic flow). After translation parallel to the image plane has been factored out, a particularly useful parsing of the optic flow is in terms of the kinematic differential invariants *curl*, *div*, and *def* (30). Of these, only the *def* component carries information about the structure of the scene and has formally been linked to local slants and tilts (30–33). Building on this formal work, experimental studies have shown that the perceived slant is proportional to the *def* component of the optic flow (34), a *def*-based heuristic model has been suggested for computing local slants and tilts of planar stimuli (4, 35), and *def* has been used in a template model for volumetric rigid stimuli rotating along arbitrary axes (36). Physiological studies on nonhuman primates and imaging studies on humans have found neurons in the medio-superior temporal (MST) area (37–39), the temporal polysensory area (STPa) (40, 41), and the superior temporal sulcus (STS) (42) that preferentially respond to the *curl* and *div* components of optic flow but not to the *def* component. Lagae et al. (43) reported a few cells that preferentially respond to the *def* component in area MSTd. In addition, *def* can be composed from a pair of motion shears or compressions, and a number of studies have found cells in area MT that can signal motion shear (19–22).

Nonrigid structures have been studied extensively in one domain: point light depictions of biological motion (44), where humans can identify gender (45), emotion (46), and other attributes from moving lights placed at the joints of articulation. How this is done by the brain is an open question, but some neural models postulate that it is by the use of snapshot neurons that are tuned to different poses of a deforming humanoid shape (47). However, it is unlikely that snapshot neurons exist for very many objects in the large class of deforming shapes.

In machine vision, moving rigid shapes are most often extracted using variants of factorization of the image stream matrix into a 3D shape matrix and an orthonormal rotation matrix (48). For nonrigid shapes, the shape matrix changes on every frame, but because the shapes generally change continuously with smooth movements, the factorization can be extended by either constraining form as a linear combination of a small set of basis shapes derived from the images, or by constraining motion as a linear

Author contributions: A.J. and Q.Z. designed research; A.J. performed research; A.J. and Q.Z. analyzed data; and A.J. and Q.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: qz@sunyopt.edu or anshuljjain@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016211108/-/DCSupplemental.

NEUROSCIENCE

combination of a small set of basis trajectories (e.g., oscillations) (49–52).

The current study measured observers' ability to discern nonrigid shapes from motion cues and examined different models for explaining these capabilities. We began with asking observers to discriminate deviations from perfect circularity for rotating rigid and flexing cylinders (flexing cylinders were similar to a garden hose shaken from a central hold). We found that observers were as sensitive to deviations from circularity for nonrigid shapes as for rigid shapes. A motion perspective model was better than a trajectory basis model in predicting average performance. In addition, we found that, when rotation of symmetric stimuli created asymmetric retinal velocity fields, observers reported asymmetric percepts. Finally, we examined how performance degrades in the presence of multiple nonrigid deformations (i.e., for cylinders that inflated and deflated while flexing in the image plane or depth).

Results

Experiment 1: Discriminating Shapes Using Motion Cues. Observers monocularly viewed perspective projections of rigid and flexing point light cylinders rotated simultaneously about the vertical axis in the image plane and the depth axis orthogonal to the image plane (Fig. 1). The diameter of the cylinder in the image plane was kept constant, but the cylinder's depth was varied to set the aspect ratio at 0.7071, 0.8409, 1, 1.1892, or 1.4142, selected randomly on each trial. Stereo, texture, and density cues were not informative. Observers were instructed to report whether the cross-sections were shallower or deeper than perfect circles; 20 blocks of 45 trials each were spread over two sessions.

Fig. 2*A* shows the average percent of trials perceived as deeper than a perfect circular cross-section for the four types of cylinders (two rigid and two flexing). Points of subjective circularity (PSC) estimated from the smooth fits are shown in Fig. 2*B*. Observers' judgments of aspect ratio were significantly deeper than veridical (PSC < 1.0 with $P < 0.05$) for the cylinder flexing in the image plane and were slightly shallower than veridical for the other three (PSCs from 1.05 to 1.2, significantly larger than 1.0 with $P < 0.01$). The slopes of the psychometric curves show good sensitivity to the stimulus variations, ranging from 0.1 to 0.8 for a twofold increase in the stimulus signal strength. More importantly, the slopes of the four psychometric functions did not differ significantly [$F(3,28) = 0.9435$, $P = 0.43$), showing that nonrigidity does not affect the sensitivity for discriminating shapes extracted on the basis of motion cues. Because rigidity assumptions are not useful in discerning the shapes of flexing cylinders and the similarity between the slopes of the psychometric curves make it unlikely that entirely separate neural processes are used for flexing than for rigid cylinders, these results suggest that percepts of rigidity may arise as a result of a more general prior assumption. Although observers were able to perform this task in a consistent manner (percent perceived deeper varied monotonically with the aspect ratio, and mean PSCs were within 20% of veridical), the sensitivity observed here is substantially lower than that for discriminating elliptical departures from circular shapes in 2D (53).

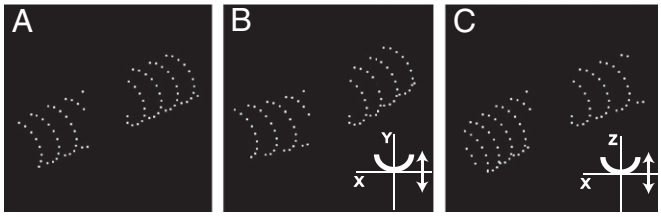

Fig. 1. Sample frames for simultaneous rotation about the vertical and depth axes for (*A*) a rigid cylinder, (*B*) a cylinder flexing in the image plane, and (*C*) a cylinder flexing in depth. The regular grids of dots in the figure are only for illustration purposes; in the experiments, dots were randomly placed after surface generation to remove density and texture cues to shape, and cylinders were presented behind a frame so that the curved edges were not visible to the observer.

Computational analyses. To understand how observers were accomplishing the experimental tasks, we analyzed our stimuli using the global motion perspective approach (3, 13, 14) and the trajectory basis approach (50).

Motion perspective model. Using the heuristic that the image velocity of every point is inversely related to the distance of the point from the optical center of the observer (28), we computed a velocity contrast metric (VCM) for each narrow cross-sectional slice as $VCM = (V_{max}-V_{min})/(V_{max}+V_{min})$ (i.e., the normalized difference between the maximum and minimum velocity along the cross-sectional slice) (Fig. 3*A*). To compute the VCM for the entire trial, the VCM was averaged across each cross-sectional slice on a single frame and then across all the frames.

Trajectory basis model. We applied the algorithm of Akhter et al. (50) to compute the structure of the deforming cylinder. Taking advantage of the fact that the cylinders were deforming smoothly, the 3D path taken by each dot on the cylinder was modeled as a linear combination of simple trajectories formed by discrete cosine transforms. The number of harmonics used as basis trajectories was empirically determined to be three for the two flexing cylinders. After the structure was determined (up to an arbitrary scale and rotation), we aligned the cylinder along its length and computed the aspect ratio on each frame. The aspect ratio for the trial was computed by averaging the aspect ratio across all frames. The schematic in Fig. 3*B* shows how a complex smooth trajectory can be modeled as a linear combination of cosine oscillations. It should be noted that the shape basis approach (51) recovers similar nonrigid structures from our stimuli. We chose the trajectory basis functions because of the greater potential for generalization: the same trajectory basis functions can be used for a wide range of 3D shapes, but a novel shape basis needs to be computed each time for a different shape.

The simple motion perspective model simulates psychometric curves that are roughly parallel, and the order of the curves is similar to the order in the psychophysical data (Fig. 2*C*): the VCM is highest for the cylinder flexing in the image plane and smallest for the faster-rotating rigid cylinder. Interestingly, the faster-rotating rigid cylinder was perceived as shallower than the slower-rotating rigid cylinder, indicating that perceived depth magnitude depends more on velocity contrast than on absolute speed. These simulations suggest that observers rely on relative velocity information to infer 3D shapes, at least for these qualitative judgments. On the other hand, the aspect ratio computed by the trajectory basis model is very similar for all the cylinders and slightly higher than the veridical aspect ratio (Fig. 2*D*). Thus, this model does not explain the perceptual biases observed in the psychophysical data, particularly the bias to perceive cylinders flexing in the image plane as deepest.

Experiment 2: Asymmetric Percepts from Symmetric Cylinders. The cylinders presented in experiment 1 were perfectly symmetric around the central transversal plane orthogonal to the image plane; however, while making observations in the experiment, the authors noted that some cylinders appeared asymmetric. Experiment 2 was designed to measure this systematically. The stimuli were exactly the same as in experiment 1, except that the aspect ratio was fixed at one of these three values (0.7071, 1, or 1.4142) selected pseudo-randomly on each trial. On one-half of the trials, selected pseudo-randomly, the cylinder rotated about both the depth and vertical axes (compound rotation condition as in experiment 1), and on the other one-half, the cylinder rotated only about the vertical axis (simple rotation condition).

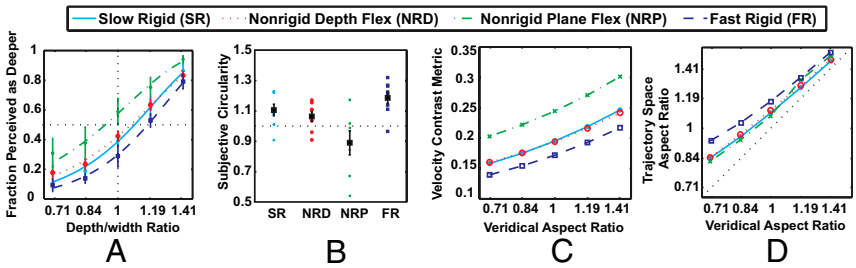

Fig. 2. Psychophysical and simulation results from experiment 1. (*A*) Fraction of trials perceived as deeper than a circle and plotted as a function of aspect ratio for the four conditions, slow rigid (SR), depth flex (NRD), plane flex (NRP), and fast rigid (FR), averaged across eight observers. (*B*) Observers' points of subjective circularity for the four conditions in experiment 1 with the group mean. (*C*) The velocity contrast metric (VCM) computed as a function of the aspect ratios for the four types of stimuli. VCM gives a qualitative measure of the computed depth. (*D*) The aspect ratios of the shapes extracted by the trajectory space-based model for the four types of stimuli.

The observers' task was to report whether the cylinder had a symmetric or asymmetric profile by pressing one of two keys. Observers ran 12 blocks of 54 trials each. The entire session lasted about 1 h.

Fig. 4*A* shows the percentage of trials perceived as asymmetric as a function of aspect ratio for the rigid and flexing cylinders for the two types of motion. Cylinders were perceived as asymmetric more often when undergoing compound rotation than when undergoing simple rotation [$F(1,5) = 22.574$, $P < 0.01$], and shallow cylinders were perceived as asymmetric more often than deep cylinders [$F(2,10) = 4.915$, $P < 0.05$]. Furthermore, there was also a main effect of nonrigidity [$F(2,10) = 7.610$, $P < 0.01$], particularly cylinders flexing in the image plane were perceived as asymmetric more often than the other two cylinders.

To understand the perceived asymmetry, we looked at the velocity profiles generated by the stimuli. Fig. 5 *A* and *B* shows the velocity profiles for rigid cylinders under simple and compound rotation, respectively. As the figure shows, the velocity profiles are symmetric about the horizontal midline for simple rotation but distinctly asymmetric for compound rotation. The combined rotation about the vertical and depth axes is equivalent to a rotation about an oblique axis that passes below the midline of the front surface of the cylinder, and the location of the minimum velocity is the projection of this intersection. When we reversed the relative phase of the two simultaneous rotations, the intersection of the rotation axis and the minimum velocity shifted above the horizontal midline, and this was reflected in the perceived asymmetry.

To judge whether the two models could simulate the perceived asymmetry, we calculated an asymmetry metric (AM) as the mean-squared error between the veridical symmetric shape and the shapes computed by the trajectory basis and motion perspective models. The shapes were normalized to have the same range and magnitude before computing the error. Fig. 4 *B* and *C* shows the computed AM for the two models. Neither of these models explains the psychophysical data perfectly. The motion perspective model shows that asymmetry decreases with increasing aspect ratio, as observed in the data, and also shows higher asymmetry for compound rotation than for simple rotation, as found in the experimental data; however, the asymmetry for cylinders flexing in depth was highest, unlike the experimental data. The trajectory basis model does predict that cylinders flexing in the image plane appear most asymmetric; however, the bias is grossly exaggerated, because the cylinders flexing in depth and the rigid cylinders have very little asymmetry for high-aspect ratios, even for compound rotation. The results of this experiment provide further evidence that observers rely on relative velocity measurements to compute 3D shape to estimate both qualitative properties as in experiment 1 and global shape properties like symmetry. This provides an interesting counterpart to results showing that global properties like symmetry can help to resolve ambiguities in nonrigid motion (54).

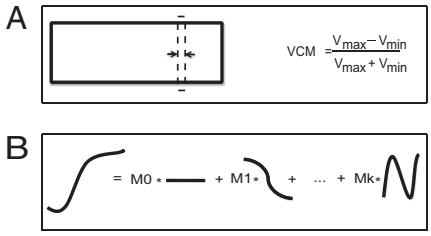

Fig. 3. The two modeling approaches. (*A*) Computation of the velocity contrast metric (VCM) for a cross-sectional slice on the cylinder. (*B*) A complex but smooth feature trajectory modeled as a linear combination of sinusoidal trajectories.

Experiment 3: Detection of Multiple Nonrigidities. In the natural world, objects often undergo multiple deformations simultaneously (e.g., a stalking animal inhales and exhales). To see if observers can detect multiple simultaneous nonrigidities, we examined whether observers could detect inflation and deflation of rigid and flexing cylinders by smoothly varying the aspect ratio of the cylinder during the trial (the cross-section in the image plane was kept constant). Any given trial began with a cylinder with aspect ratio of 1.0 (perfectly circular cross-section), and then, it inflated by a fixed amount in the first one-half of the trial and deflated by the same amount to a perfectly circular cylinder or deflated by a fixed amount followed by an inflation by the same amount. The order of inflation and deflation was randomized across trials. The observers' task was to report whether the trial consisted of an inflation followed by a deflation or vice versa. This experiment was conducted in two parts. In the first part, we used a constant stimulus paradigm using relatively high signal values to assess the performance of observers on the task; only those observers who could perform the task in a consistent manner participated in the second part. In the second part, we used a Bayesian adaptive threshold estimation procedure [QUEST (55)] to determine 82% correct thresholds for the rigid and flexing cylinders. The degree of inflation/deflation was controlled as the percentage change from unit aspect ratio.

Fig. 6*A* shows the percentage accuracy averaged across inflation and deflation for nine naïve observers as a function of signal strength. Only two of nine observers could perform the task consistently, implying that the task is inherently difficult and may require extensive training. The difficulty may be related to problems in detecting instantaneous stretching of objects along the line of sight from motion cues alone (56). We selected these two observers plus an author (A.J.) and another observer, who was extensively trained on the task but was uninformed about the

NEUROSCIENCE

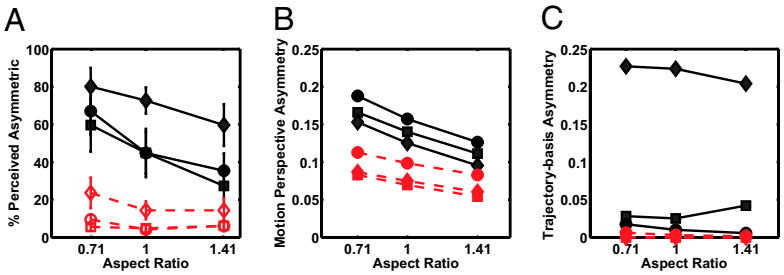

Fig. 4. Asymmetric percepts from symmetric cylinders. (*A*) Percentage perceived as asymmetric as a function of the aspect ratio for the three types of cylinders. Squares, rigid cylinders; diamonds, plane-flex cylinders; circles, depth-flex cylinders. Black solid lines and symbols represent compound rotation, and red dashed lines and hollow symbols represent simple rotation. (*B*) Asymmetry as predicted by the motion perspective model. (*C*) Asymmetry as predicted by the trajectory basis model.

purposes of the experiment, and we measured accuracy thresholds for inflation and deflation independently. Fig. 6*B* shows the 82% accuracy thresholds averaged across the four observers. Overall, thresholds for deflation were higher than thresholds for inflation [$F(1,18) = 5.33$, $P < 0.05$]. More importantly, there was no systematic difference in thresholds between rigid and flexing cylinders [$F(2,18) = 0$, $P = 0.99$], and there was no significant interaction [$F(2,18) = 0.18$, $P = 0.83$].

Fig. 7 shows the performance of the two computational models for the rigid cylinder undergoing inflation and deflation. The motion perspective model follows the variation in aspect ratio quite well (Fig. 7 *A* and *B*), but the trajectory basis model does not (Fig. 7 *C* and *D*). We observed no improvement in the performance of the trajectory basis model by using a larger number of basis trajectories. These results show that detecting changes in relative velocities over time would be a successful strategy for this task. Performance in the task may be limited by the inability to detect accelerations (57, 58). However, after training, four observers could perform the task reliably, which suggests that observers can learn to associate changes in relative velocities with changes in shape.

Discussion

Motions in the retinal image are predominantly caused by movements of the observer's body and head, and in this case, most objects have to be inferred to be rigid and stationary, despite distortions in the retinal images. There is some evidence for a rigidity prior that biases scene percepts during self-motion (59), and it is possible that this prior is generalized to object-generated retinal motions. The most significant findings of the experiments in this study are that humans are equally sensitive at discriminating nonrigid and rigid shapes from motion cues, show similar biases in perceiving departures from nonrigid and rigid symmetries, and are equally sensitive to expansion and contraction of nonrigid and rigid shapes. These results suggest that the human perceptual system uses similar mechanisms for extracting rigid and nonrigid shapes from motion cues, and it is unlikely that a rigidity assumption per se is paramount in this process. As general alternatives to the rigidity assumption, a number of regularization principles have been proposed, such as minimal mapping (9, 60), smoothest motion (60), or motion coherence (61), but a model that generates general nonrigid percepts on the basis of minimization principles remains to be constructed (62).

The computational simulations in this study found that the motion perspective model predicted human performance better than the trajectory space model, suggesting that the human perceptual system uses relative motions to compute 3D structure. In addition, simulated symmetric cylinders were perceived as asymmetric whenever the velocity profile was asymmetric, providing further evidence that the relative velocity profile determined the perceived 3D shape. The motion perspective model is essentially based on the computation of motion shears at different spatial scales. We have presented it as a simple explanation of our empirical results while recognizing that it will need to be elaborated on for general situations. In a simple case like our rotating rigid cylinders, local motions are predominantly parallel as are the shears. In general, variations in speed, as a function of depth and distance from the rotation axis, lead to 2D deformations in the perspective image (32) that can be captured by 2D combinations of local shears corresponding to the *def* component of the optic flow (63, 64). For a translating observer and a rigid stationary object, functions of the *def* can be used to compute the angle of inclination (tilt) and the magnitude of inclination (slant) up to a bas-relief ambiguity (31, 65). In preliminary simulations, we find that, when approximating the smooth cross-sections by

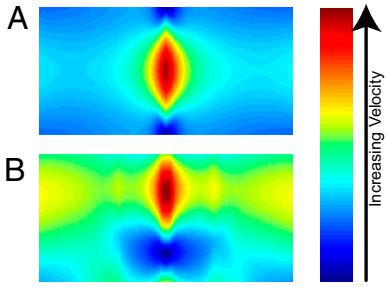

Fig. 5. Velocity profiles for the plane-flex cylinder with an aspect ratio of 0.71 under simple rotation (*A*) and compound rotation (*B*).

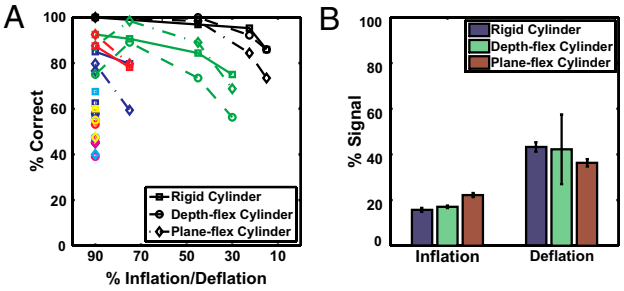

Fig. 6. Results from experiment 3. (*A*) Percentage of correct responses for distinguishing different amounts of inflation and deflation (different colors for nine observers). Observers who did not reach the threshold for 90% deformation were not tested on less extreme deformations. Rigid, depth-flex, and plane-flex cylinders are represented using squares, circles, and diamonds, respectively. (*B*) Detection thresholds measured for inflation and deflation on the three types of cylinder averaged across four observers capable of doing the task.

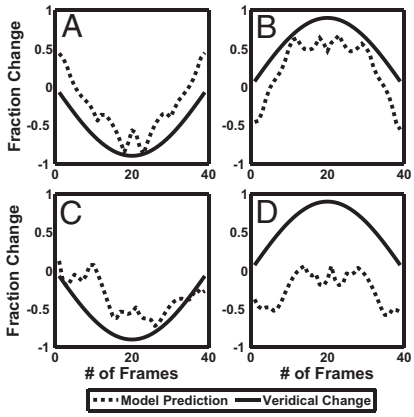

Fig. 7. Simulated performance of the two computational models while following inflation/deflation of the cylinders. *Left* (*A* and *C*) shows model predictions for a rigid cylinder undergoing 90% inflation, whereas *Right* (*B* and *D*) shows model predictions for a rigid cylinder undergoing 90% deflation. *Upper* shows predictions based on the motion perspective model, whereas *Lower* shows the predictions based on the trajectory basis model. Models made similar predictions for flexing cylinders.

locally planar polygons, the relative magnitudes of *def* were consistent with the local slants of the rigid cylinder stimuli, and therefore, this could be a more general alternative to the motion perspective model.

This study has concentrated on possibly the simplest attribute of our stimuli, the cross-section. It is gratifying that percepts of this attribute have general theoretical implications, but it is likely that more complicated aspects, such as the nature of the perceived nonrigidity, will supply much richer theoretical concepts on how nonrigid shapes are constructed in the brain. Hence, we are collecting additional evidence about perceived shapes and rotations. It is possible that the differential invariants will be useful in modeling human performance. For example, motion perspective alone predicts that a rigid cylinder in compound or simple rotation should be seen as flexing in the center, and therefore, 1D divergences that are a signature for object rotation may have been used in the inference of a rigid rotation. In addition, spatial gradients of differential invariants can be formally linked to the Gaussian and mean curvatures of the objects (30), thus going considerably beyond estimating local slant and tilt (e.g., flexing of the cylinder could be signaled by a change in sign of the Gaussian curvature). It is also possible that it will be useful to compute analogs of the differential invariants on extended scales instead of using local directional derivatives (65) (i.e., the outputs of deformation and divergence filters); for example, extended deformation templates can be formulated to signal the difference between rigid and deforming shapes of the class studied by Koenderink and van Doorn (66). There is a small amount of evidence that observers can learn to associate particular motion patterns with rigidity. Some of the evidence is informal (e.g., reports of initially seeing rigid objects as nonrigid), and some is explicit that preexposure to wire-frame shapes promotes rigid percepts (67). We intend to explore whether learning of deformation patterns underlies the perception of different classes of nonrigidity.

Methods

Observers. Eight observers participated in experiment 1, and six observers in experiment 2; 11 observers participated in experiment 3. Both authors participated in experiments 1 and 2; A.J. also participated in experiment 3. Observers other than the authors were informed about the purpose of the experiment only after it was completed. All subjects had normal or corrected to normal vision. The experiments were conducted in compliance with the standards set by the Internal Review Board at the State University of New York-College of Optometry. Subjects gave their informed consent before their inclusion in the study and were paid for their participation.

Apparatus. Stimuli were generated using the CRS Toolbox for MATLAB (The Mathworks) and were displayed on a Sony CRT monitor using Cambridge Research Systems' ViSaGe system (Cambridge Research Systems) controlled by a Dell GX620. The monitor's resolution was set to 1,024 × 768 pixels, and the refresh rate was set to 120 Hz. The experiments were conducted in a dark room. The observers viewed the stimuli monocularly from a distance of 100 cm using a chin rest to stabilize head position.

Stimuli and Tasks. Stimuli consisted of white dots on a black background randomly placed on an opaque black horizontal cylinder after rendering the 3D surface (Fig. 1), thus removing texture and density cues to 3D shape. The cylinders were presented in proper perspective for the position of the observer's eye. The 16.92°-long and 2.12°-wide cylinder was rotated simultaneously at π radians per second about the vertical and at $2\pi/3$ radians per second about the depth axis (the cylinder did not spin on its axis, and 3D shape was not perceived with rotation only in the image plane). The starting orientation was set at $-\pi/6$ roll and $-\pi/4$ yaw for all trials, applied in that order. The cylinder rotated until it reached an orientation with $+\pi/6$ roll and $+\pi/4$ yaw and then rotated back to the initial position. The stimuli lasted for 1.0 s. The rotating cylinder was only visible through a window 6.36° wide so that observers could not see the ends of the cylinder. The central 0.8° section of the cylinder was occluded to prevent observers from using just the approximately rigid central section. The cylinder was either rigid or flexed smoothly in the depth or the image plane using Eqs. **1** and **2**, respectively (Movies S1, S2, and S3):

$$z(k) = z_0 - (x_0/20)^2 * \sin(4\pi k/N) \quad [1]$$

and

$$y(k) = y_0 - (x_0/30)^2 * \sin(4\pi k/N) \quad [2]$$

where k = current frame, N = total frames, and x_0, y_0, z_0 = initial position.

Rotation was applied after computing the flexing. The depth-flex amplitude was approximately 1.2° visual angle (dva), and the image plane flex was 0.53 dva at the extremes for cylinder flexing in the image plane. In pilot measurements, these two flexes were judged to have perceptually similar magnitudes. We also included a condition where the rigid cylinder rotated at a higher speed (1.5×) so that the mean speed of the dots approximately matched the mean speed for the nonrigid cylinders.

ACKNOWLEDGMENTS. This work was supported by National Eye Institute Grants EY07556 and EY13312 (to Q.Z.).

1. Wallach H, O'Connell DN (1953) The kinetic depth effect. *J Exp Psychol* 45:205–217.
2. Sperling G, Dosher BA, Landy MS (1990) How to study the kinetic depth effect experimentally. *J Exp Psychol Hum Percept Perform* 16:445–450.
3. Sperling G, Landy MS (1989) Kinetic depth effect and identification of shape. *J Exp Psychol Hum Percept Perform* 15:826–840.
4. Domini F, Caudek C (1999) Perceiving surface slant from deformation of optic flow. *J Exp Psychol Hum Percept Perform* 25:426–444.
5. Rogers B, Graham M (1979) Motion parallax as an independent cue for depth perception. *Perception* 8:125–134.
6. Braunstein ML, Hoffman DD, Pollick FE (1990) Discriminating rigid from nonrigid motion: Minimum points and views. *Percept Psychophys* 47:205–214.
7. Todd JT (1984) The perception of three-dimensional structure from rigid and nonrigid motion. *Percept Psychophys* 36:97–103.
8. Hogervorst MA, Kappers AM, Koenderink JJ (1996) Structure from motion: A tolerance analysis. *Percept Psychophys* 58:449–459.
9. Ullman S (1979) The interpretation of structure from motion. *Proc R Soc Lond B Biol Sci* 203:405–426.
10. Grzywacz NM, Hildreth EC (1987) Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. *J Opt Soc Am A* 4: 503–518.
11. Domini F, Caudek C, Richman S (1998) Distortions of depth-order relations and parallelism in structure from motion. *Percept Psychophys* 60:1164–1174.
12. Domini F, Braunstein ML (1998) Recovery of 3D structure from motion is neither euclidean nor affine. *J Exp Psychol Human Percept Perform* 24:1273–1295.
13. Helmholtz HV (1910) *Treatise on Physiological Optics* (Optical Society of America, New York), Vol 3.

14. Gibson JJ, Olum P, Rosenblatt F (1955) Parallax and perspective during aircraft landings. *Am J Psychol* 68:372–385.
15. Buracas GT, Albright TD (1996) Contribution of area MT to perception of three-dimensional shape: A computational study. *Vision Res* 36:869–887.
16. Graham M, Rogers B (1982) Simultaneous and successive contrast effects in the perception of depth from motion-parallax and stereoscopic information. *Perception* 11:247–262.
17. Rogers BJ, Graham ME (1983) Anisotropies in the perception of three-dimensional surfaces. *Science* 221:1409–1411.
18. Loomis JM, Nakayama K (1973) A velocity analogue of brightness contrast. *Perception* 2:425–427.
19. Hammond P, Smith AT (1982) On the sensitivity of complex cells in feline striate cortex to relative motion. *Exp Brain Res* 47:457–460.
20. Allman J, Miezin F, McGuinness E (1985) Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci* 8:407–430.
21. Frost BJ, Nakayama K (1983) Single visual neurons code opposing motion independent of direction. *Science* 220:744–745.
22. Snowden RJ, Treue S, Erickson RG, Andersen RA (1991) The response of area MT and V1 neurons to transparent motion. *J Neurosci* 11:2768–2785.
23. Bradley DC, Chang GC, Andersen RA (1998) Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* 392:714–717.
24. Nawrot M, Blake R (1991) A neural network model of kinetic depth. *Vis Neurosci* 6: 219–227.
25. Fernández JM, Watson B, Qian N (2002) Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Res* 42:883–898.
26. Grunewald A, Bradley DC, Andersen RA (2002) Neural correlates of structure-from-motion perception in macaque V1 and MT. *J Neurosci* 22:6195–6207.
27. Orban GA, Sunaert S, Todd JT, Van Hecke P, Marchal G (1999) Human cortical regions involved in extracting depth from motion. *Neuron* 24:929–940.
28. Caudek C, Proffitt DR (1993) Depth perception in motion parallax and stereokinesis. *J Exp Psychol Hum Percept Perform* 19:32–47.
29. Liter JC, Braunstein ML, Hoffman DD (1993) Inferring structure from motion in two-view and multiview displays. *Perception* 22:1441–1465.
30. Koenderink JJ, van Doorn AJ (1975) Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Opt Acta (Lond)* 22(9): 773–791.
31. Longuet-Higgins HC, Prazdny K (1980) The interpretation of a moving retinal image. *Proc R Soc Lond B Biol Sci* 208:385–397.
32. Koenderink JJ (1986) Optic flow. *Vision Res* 26(1):161–179.
33. Koenderink JJ, van Doorn AJ (1976) Local structure of movement parallax of the plane. *J Opt Soc Am* 66(7):717–723.
34. Freeman TCA, Harris MG, Meese TS (1996) On the relationship between deformation and perceived surface slant. *Vision Res* 36:317–322.
35. Domini F, Caudek C (2003) 3-D structure perceived from dynamic information: A new theory. *Trends Cogn Sci* 7:444–449.
36. Fernandez JM, Farell B (2009) A new theory of structure-from-motion perception. *J Vis* 9:1–20.
37. Saito H, et al. (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J Neurosci* 6:145–157.
38. Tanaka K, Saito H (1989) Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J Neurophysiol* 62:626–641.
39. Sugihara H, Murakami I, Shenoy KV, Andersen RA, Komatsu H (2002) Response of MSTd neurons to simulated 3D orientation of rotating planes. *J Neurophysiol* 87: 273–285.
40. Anderson KC, Siegel RM (2005) Three-dimensional structure-from-motion selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey. *Cereb Cortex* 15:1299–1307.
41. Anderson KC, Siegel RM (1999) Optic flow selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey. *J Neurosci* 19:2681–2692.
42. Beer AL, Watanabe T, Ni R, Sasaki Y, Andersen GJ (2009) 3D surface perception from motion involves a temporal-parietal network. *Eur J Neurosci* 30:703–713.
43. Lagae L, Maes H, Raiguel S, Xiao DK, Orban GA (1994) Responses of macaque STS neurons to optic flow components: A comparison of areas MT and MST. *J Neurophysiol* 71:1597–1626.
44. Johansson G (1973) Visual perception of biological motion and a model for its analysis. *Percept Psychophys* 14(2):201–211.
45. Mather G, Murdoch L (1994) Gender discrimination in biological motion displays based on dynamic cues. *Proc R Soc Lond B* 258:273–279.
46. Dittrich WH, Troscianko T, Lea SE, Morgan D (1996) Perception of emotion from dynamic point-light displays represented in dance. *Perception* 25:727–738.
47. Giese MA, Poggio T (2003) Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4:179–192.
48. Tomasi C, Kanade T (1992) Shape and motion from image streams under orthography: A factorization method. *Int J Comput Vis* 9(2):137–154.
49. Torresani L, Hertzmann A, Bregler C (2008) Nonrigid structure-from-motion: Estimating shape and motion with hierarchical Priors. *IEEE Trans Pattern Anal Mach Intell* 30: 878–892.
50. Akhter I, Sheikh YA, Khan S, Kanade T (2009) Nonrigid structure from motion in trajectory space. *Adv Neural Inf Process Syst* 21:41–48.
51. Bregler C, Hertzmann A, Biermann H (2000) Recovering non-rigid 3D shape from image streams. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society Press, Washington, DC), Vol 2, pp 690–696.
52. Brand M (2005) A direct method for 3D factorization of nonrigid motion observed in 2D. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society Press, Washington, DC), Vol 2, pp 122–128.
53. Regan D, Hamstra SJ (1992) Shape discrimination and the judgment of perfect symmetry: Dissociation of shape from size. *Vision Res* 32:1845–1864.
54. Cohen EH, Jain A, Zaidi Q (2010) The utility of shape attributes in deciphering movements of non-rigid objects. *J Vis* 10:29.
55. Watson AB, Pelli DG (1983) QUEST: A Bayesian adaptive psychometric method. *Percept Psychophys* 33:113–120.
56. Norman JF, Todd JT (1993) The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Percept Psychophys* 53:279–291.
57. Todd JT, Bressan P (1990) The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Percept Psychophys* 48:419–430.
58. Hogervorst MA, Eagle RA (2000) The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *J Exp Psychol Hum Percept Perform* 26:934–955.
59. Glennerster A, Tcheang L, Gilson SJ, Fitzgibbon AW, Parker AJ (2006) Humans ignore motion and stereo cues in favor of a fictional stable world. *Curr Biol* 16:428–432.
60. Hildreth EC (1984) *Measurement of Visual Motion* (MIT Press, Cambridge, MA).
61. Yuille AL, Grzywacz NM (1988) A computational theory for the perception of coherent visual motion. *Nature* 333:71–74.
62. Rokers B, Yuille A, Liu Z (2006) The perceived motion of a stereokinetic stimulus. *Vision Res* 46:2375–2387.
63. Meese TS, Harris MG (1997) Computation of surface slant from optic flow: Orthogonal components of speed gradient can be combined. *Vision Res* 37:2369–2379.
64. Meese TS, Harris MG, Freeman TCA (1995) Speed gradients and the perception of surface slant: Analysis is two-dimensional not one-dimensional. *Vision Res* 35: 2879–2888.
65. Cipolla R, Okamoto Y, Kuno Y (1993) Robust structure from motion using motion parallax. *Proceedings of the International Conference on Computer Vision* (IEEE Computer Society Press, Washington, DC), pp 374–382.
66. Koenderink JJ, van Doorn AJ (1986) Depth and shape from differential perspective in the presence of bending deformations. *J Opt Soc Am A* 3:242–249.
67. Sinha P, Poggio T (1996) Role of learning in three-dimensional form perception. *Nature* 384:460–463.