Effects of surface reflectance and 3D shape on perceived rotation axis

Katja Doerschner
Department of Psychology, Bilkent University, Ankara, Turkey
National Magnetic Resonance Research Center (UMRAM), Ankara, Turkey

Ozgur Yilmaz
ASELSAN, MGEO, Ankara, Turkey
Department of Computer Engineering, Turgut Ozal University, Ankara, Turkey

Gizem Kucukoglu
Department of Psychology, New York University, New York, NY, USA

Roland W. Fleming
Department of Psychology, University of Giessen, Germany

Surface specularity distorts the optic flow generated by a moving object in a way that provides important cues for identifying surface material properties (Doerschner, Fleming et al., 2011). Here we show that specular flow can also affect the perceived rotation axis of objects. In three experiments, we investigate how three-dimensional shape and surface material interact to affect the perceived rotation axis of unfamiliar irregularly shaped and isotropic objects. We analyze observers' patterns of errors in a rotation axis estimation task under four surface material conditions: shiny, matte textured, matte untextured, and silhouette. In addition to the expected large perceptual errors in the silhouette condition, we find that the patterns of errors for the other three material conditions differ from each other and across shape category, yielding the largest differences in error magnitude between shiny and matte, textured isotropic objects. Rotation axis estimation is a crucial implicit computational step to perceive structure from motion; therefore, we test whether a structure from a motion-based model can predict the perceived rotation axis for shiny and matte, textured objects. Our model's predictions closely follow observers' data, even yielding the same reflectance-specific perceptual errors. Unlike previous work (Caudek & Domini, 1998), our model does not rely on the assumption of affine image transformations; however, a limitation of our approach is its reliance on projected correspondence, thus having difficulty in accounting for the perceived rotation axis of smooth shaded objects and silhouettes. In general, our findings are in line with earlier research that demonstrated that shape from motion can be extracted based on several different types of optical deformation (Koenderink & Van Doorn, 1976; Norman & Todd, 1994; Norman, Todd, & Orban, 2004; Pollick, Nishida, Koike, & Kawato, 1994; Todd, 1985).

Introduction

Motion is a natural source of information for recognizing objects (Balas & Sinha, 2008; Vuong & Tarr, 2004). However, estimating the properties of a moving object from optical flow is a challenge because the image sequence results from a complex combination of three-dimensional (3D) shape, object motion (rotation/translation/looming), surface reflectance, and illumination. This is especially so when the moving object deforms nonrigidly or when surface material contributes its own reflectance-specific image motion, as is the case for shiny surfaces (Doerschner, Fleming, et al., 2011; Doerschner, Kersten, et al., 2011). Solutions for the recovery of 3D shape and observer motion for rigid, matte-textured objects have been proposed (Koenderink & Van Doorn, 1992; Longuet-Higgins & Prazdny, 1980), and previous research using random dot displays (which assume trackable surface markings) demonstrated that the visual system is able

Citation: Doerschner, K., Yilmaz, O., Kucukoglu, G., & Fleming, R. W. (2013). Effects of surface reflectance and 3D shape on perceived rotation axis. *Journal of Vision*, *13*(11):8, 1–23, http://www.journalofvision.org/content/13/11/8, doi:10.1167/13.11.8.

doi: 10.1167/13.11.8 Received November 23, 2012; published September 10, 2013 ISSN 1534-7362

to use optic flow information to estimate rigid (e.g., Bradley, Chang, & Andersen, 1998; Koenderink & Van Doorn, 1991; Landy, 1987; Landy, Dosher, Sperling, & Perkins, 1991; Ullman, 1979; Wallach & O'Connell, 1953) and nonrigid 3D shape (Domini, Caudek, & Proffitt, 1997; Jain & Zaidi, 2011; Ullman, 1984). However, how object shape, motion trajectory, and surface reflectance jointly affect the estimation of 3D structure from motion (SfM) has not been studied extensively.[1]

One crucial implicit computational step in SfM is the estimation of the axis of rotation. Previous work using random dot displays (Caudek & Domini, 1998), showed that perceived slant of the rotation axis of an object can be predicted by global measures of first-order optic flow. In fact, the authors suggest that human perception of SfM may be limited to an analysis of first-order optic flow properties (Caudek & Domini, 1998). However, this would assume that all changes in first-order optic flow would arise solely from rigid or nonrigid shape transformations and densely textured, matte surfaces. This is not the case, however. For moving specular objects, for example, shape and surface reflectance–specific optic flow patterns are intermingled. In recent work, we showed that optic flow properties of shiny and matte, textured objects are indeed significantly different, and these differences are used by human observers in recognizing material properties (Doerschner, Fleming, et al., 2011). Thus, the question arises whether specular SfM would be different from matte-textured SfM and, if so, how this difference would manifest itself perceptually. Assuming identical objects, there are three possibilities: (a) shiny and matte, textured objects differ in perceived shape; (b) they differ in perceived rigidity; and (c) they differ in perceived object motion. We will focus here on (c), specifically, how specular flow affects the perceived object rotation axis.[2] As an example, consider Figure 1. The objects in the upper and lower parts of the display have the same shape and rotate around the same vertical axis (purple) at the same angular velocity. The objects differ, however, in their motion-defined surface reflectance properties: The upper one is matte and textured, the lower one is specular. Specifically, the patterns visible in the surface of the lower object slide over the surface exactly as specular reflections do, whereas for the top object, the same patterns were rigidly attached to the surface to ensure that they rotate with the object like texture markings, as described in Doerschner et al. (2011) and Hartung and Kersten (2002). These differences significantly alter the optical flow patterns generated by the two objects, leading to a change in apparent motion. Most observers would report that these objects have the same 3D shape, but at the same time, they would report that the perceived axis of rotation (yellow) differs markedly between the two objects.

Why might this be? The image motion generated by the silhouette of the shiny and matte, textured pots is identical. However, silhouette motion can at best provide ambiguous information about an object's rotation axis and at worse give rise to nonrigid percepts (Sinha and Poggio, 1996; Wallach & O'Connell, 1953; Weiss & Adelson, 2000; but see also Norman & Todd, 1994; Todd, 1985). Therefore, one might suspect that by supplementing silhouette motion with optic flow arising from the object's surface reflectance, sufficient information should be available to disambiguate the object rotation axis. However, whereas the image motion from rotating diffusely reflecting, textured objects is dictated primarily by the motion of the object, image motion generated by specular objects also greatly depends on the object's 3D curvatures (Adato et al., 2007; Doerschner, Kersten, & Schrater, 2011; Koenderink & Van Doorn, 1980; Vasilyev, Adato, Zickler, & Ben-Shahar, 2008). Therefore, we make the following predictions: (a) The optic flow generated by an object's surface reflectance influences the perceived rotation axis. (b) This effect is shape dependent. (c) Given that a rigid 3D shape is perceived in Figure 1—regardless of surface reflectance—it is reasonable to assume that SfM mechanisms may in part account for differences in perceived rotation axis of shiny and matte, textured objects. We tested these three hypotheses in three experiments described below.

General methods

Overview

We conducted three behavioral experiments to investigate the effect of surface reflectance–specific image motion on the perceived rotation axis of objects. The main variable of interest was *surface material* as defined below, in particular, potential differences between shiny and matte, textured objects. We also included two control conditions: object silhouettes and uniform albedo–Lambertian reflectance. These two generate considerably different optical flow patterns than the previous two material types. Silhouettes can at best provide ambiguous information, thus serving as a general baseline. Uniform shading, however, generates image motion that strongly depends on the complexity and mesoscale structure of the 3D shape of the object (Koenderink & Van Doorn, 1980). As a shaded object rotates relative to a fixed light source, the shading patterns deform and move relative to the surface. Smooth shading, as created by simple curved surfaces,

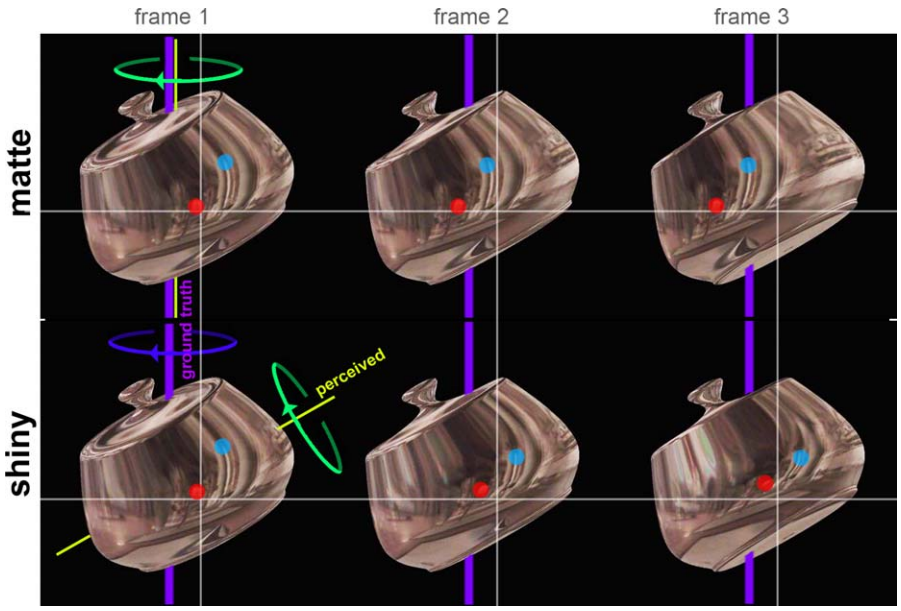

Figure 1. Perceived rotation axis orientation is affected by surface reflectance. The patterns on the top object are rigidly attached to the surface, so that although they look like reflections in any single static frame, when seen in motion, the patterns move with the surface, leading to a matte appearance. The perceived rotation axis (yellow) for the matte object (upper row) differs markedly from that of the shiny object (bottom row). This may be due to the differences in image motion generated by these two objects, illustrated by the motion trajectory of two surface features (red and blue dots) across three consecutive frames. Whereas matte features move coherently from left to right, the trajectories of corresponding specular features differ markedly from this pattern in speed and configuration. See http://gandalf.psych.umn.edu/users/kersten/kersten-lab/demos/1_S_0001/1_S_0001.mov for the original demonstration.

creates few trackable features. Thus, it constitutes a valuable baseline measurement of the contributions of those cues, as well as the contributions of shading per se, to the perceived rotation axis direction (Norman & Todd, 1994; Todd, 1985). The second variable of interest was 3D shape complexity. This was varied within Experiment 1 as well as across Experiments 1 and 2. In Experiment 2, in addition to the object material, we also systematically varied elevation and azimuth of the true rotation axis direction. In Experiment 3, we compared observers' percepts of object rotation axis to SfM-based model predictions. Remaining procedural and computational details are given below, as well as in the respective Experiment sections.

Observers

Four volunteer observers participated in Experiment 1, including two of the authors. Seven volunteer observers, including one author, participated in Experiments 2 and 3. Observers were different in Experiments 1, 2, and 3, except for one author (K. D.), who participated in all of the experiments. All observers had normal or corrected-to-normal vision. Observers gave their written consent to participation prior to the beginning of the experiment.

General stimuli

Stimuli were real-time–rendered (OpenGL) rotating[3] sinusoidally modulated spheres (Experiment 1) and isotropic surfaces of revolution (Experiments 2 and 3). Objects were either rendered with 100% specular reflectance (*shiny*), diffuse reflectance and textured (*texture*), diffuse reflectance with uniform albedo 0.5 (*uniform*), or could be displayed as dark gray silhouettes (*silhouette*; see Figure 2 for material samples). *Shiny* objects could reflect one of three possible environments: "Grace" and "Campus" as well as the desaturated and phase-scrambled (in spherical harmonics) version of the Debevec (2002) "Grace" map (Figure 2a). *Textured* objects were mapped with one of four possible 2D textures (Figure 2b), which were designed with the goal of providing rich visual structure. Note that the environments reflected and the matte, 2D textures were randomly chosen on a given material trial. This was done to minimize object recognition during the course of the experiment. In addition, to prevent such learning effects, the object's

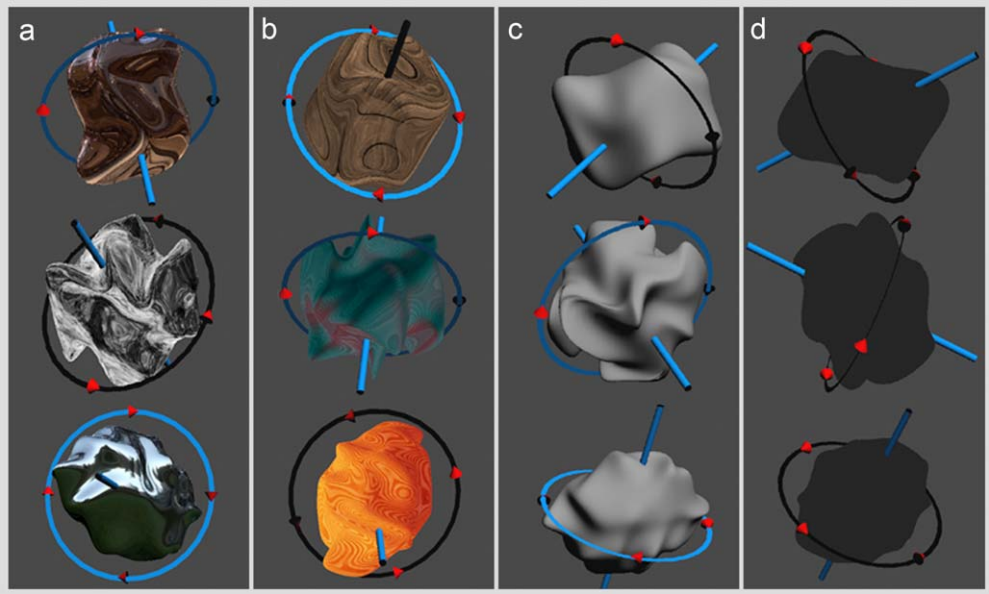

Figure 2. Experiment 1 stimuli. Shown are sample trial snapshots, including the circle and stick probe used in the experiment. Object shape was varied from simple (top) to more complex (bottom) by adjusting the frequency and amplitude of the sinusoidal modulation. We used a spherical basis shape to minimize dominant object orientations. Columns a–d show different surface material conditions: (a) *shiny*, (b) *texture*, (c) *uniform*, and (d) *silhouette*.

intrinsic orientation (relative to the rotation axis) was randomly perturbed at the beginning of every trial. Experiment-specific stimuli parameters are given in the respective sections below.

General procedure

Observers viewed rotating objects monocularly on an LCD screen (1680 × 1050) that was placed at 60-cm distance. Stimuli were approximately 11.5 cm in diameter thus subtended about 11° visual angle. Observers were instructed to estimate the axis around which the object is rotating and to adjust a stick probe (see Figure 2) such that it would be aligned with this perceived rotation axis. The orientation of the stick probe was chosen randomly at the beginning of each trial and could be adjusted by moving the mouse. Before completing a trial by pressing the "space" bar, observers were also asked to indicate the direction of rotation by setting one of two possible directions of a set of arrows (Figure 2). This setting was needed to disambiguate the tilt of the rotation axis in Experiment 1 (but not in Experiments 2 and 3). The arrow directions could be toggled via the "f" key on a computer keyboard. Observers found this task to be intuitive and were allowed to practice (with a *textured* object not used in the experiments) prior to the beginning of the experiments. Sample video clips of the task can be found at http://bilkent.edu.tr/~katja/orientation.html. The experimental software was written by us using Psychtoolbox routines (Brainard, 1997; Pelli, 1997).

General analysis

We computed the angular error between ground truth rotation axis direction o and the observer's estimated rotation axis direction $\hat{o}$:

$$\varepsilon = \arccos\left(\frac{o \cdot \hat{o}}{||o||\,||\hat{o}||}\right). \quad (1)$$

In Experiment 1, we analyze the effects of object *shape* and *surface material* on ε as well as on estimated rotation axis elevation $\hat{\theta}$ and azimuth $\hat{\phi}$ using 3 × 4 (shape × material) two-way analyses of variance (ANOVAs). In Experiment 2, we examine the effects of *surface material* and *true rotation axis direction* on ε, $\hat{\theta}$, and $\hat{\phi}$ 5 × 12 × 4 (elevation θ × azimuth ϕ × material) three-way ANOVAs. In Experiment 3, we introduce an SfM model that allows us to predict perceived rotation axes directions for a set of isotropic *shiny* and *textured* objects.

Experiment 1

To investigate how the surface material and shape interact to affect the perceived axis of rotation, we

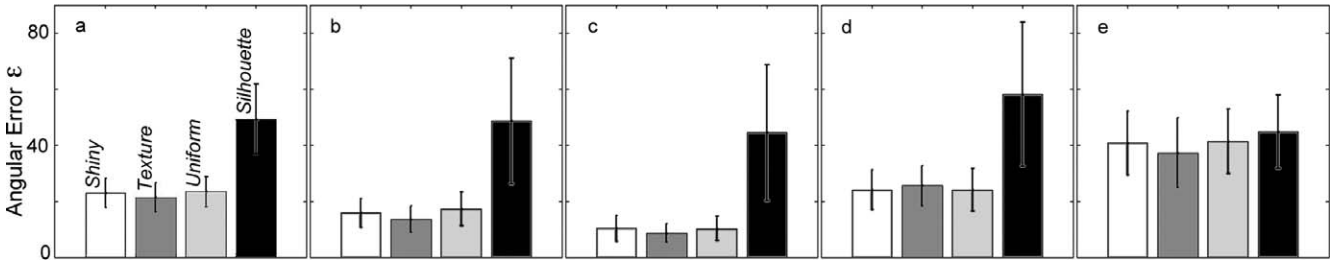

Figure 3. Experiment 1 results. (a) Mean data of the four observers; b–e show individual data. Other than a significant difference between the *silhouette* conditions and other material conditions, we did not find any significant effects of surface material (e.g., *shiny* vs. *textured*) on rotation axis estimation error. We attributed this to abundant shape complexity and addressed this issue further in Experiment 2. Mean angular errors ε for each object and material type can be found in Table 1. Error bars are two times the standard error of the mean.

used three unfamiliar irregularly-shaped objects of varying shape complexity. We analyze observers' patterns of errors in a rotation axis estimation task under four surface material conditions: *shiny*, *textured*, *uniform*, and *silhouette*. Although we were primarily interested in the differences between the *shiny* and *texture* conditions, we also included *uniform* and *silhouette* conditions to provide a baseline for comparison. *Silhouettes* served as a general baseline for the contributions of contour motion, and *uniform* objects provided information about the contributions of shading and the complexity and mesoscale structure of the 3D shape to the perceived rotation axis.

Stimuli

Object shape was varied from simple to complex (Figure 2, top to bottom row, respectively) by adjusting the frequency and amplitude of the randomly oriented sinusoidal modulations that were applied to the shape. We used a spherical basis shape to avoid dominant object orientations and resulting motion biases (Mulholland, 1956).

Ten rotation axes were sampled randomly (continuous random variable) from the unit hemisphere for each session. This resulted in 120 trials per observer (3 shapes $\times$ 4 materials $\times$ 10 rotation axes) per session. Objects rotated around an axis through their center point at a speed of 60°/s.

Procedure

Observers performed the rotation axis estimation task as described above. Each observer completed two sessions (240 trials), each lasting about 35 min.

Results

Angular error

The 3 $\times$ 4 two-way ANOVA revealed that there was no statistically significant effect of object shape, $F(2, 948) = 0.73$, $p = 0.48$, but a statistically significant effect of surface material on rotation axis orientation estimation error, $F(3, 948) = 48.41$, $p < 0.0001$ (Figure 3). The interaction between shape and material was not significant, $F(6, 948) = 0.74$, $p = 0.62$, $\alpha = 0.01$. Post hoc analysis of the main effect of "material" indicated that estimation errors in the *silhouette* condition were significantly larger than all other material conditions at the $\alpha = 0.05$ level. The remaining pairwise comparisons yielded no significant differences. We adjusted for multiple comparisons using Scheffe's S procedure. Table 1 shows mean angular errors for all conditions.

Given that responses could fall anywhere on the hemisphere of possible orientations, with a minimum angular error of 0° and a maximum angular error of 108°, random settings would lead to an average error of 90°. We find that all conditions for all subjects were significantly better than chance performance. Observers' average estimation errors ε for each material condition were $\varepsilon_{shiny} = 22.76$; $\varepsilon_{texture} = 21.29$; $\varepsilon_{uniform} = 23.25$; $\varepsilon_{silhouette} = 49.13$ ($t_{shiny}[239] = -52.01$, $p < 0.0001$; $t_{texture}[239] = -52.75$, $p < 0.0001$; $t_{uniform}[239] = -50.03$, $p < 0.0001$; $t_{silhouette}[239] = -13.19$, $p < 0.0001$, one sample t tests, two-tailed, Bonferroni corrected $\alpha = 0.0125$). Thus, observers were able to perform the task even under the most ambiguous condition (*silhouette*).

Elevation settings

There were no statistically significant main effects of surface material and object shape on perceived rotation axis elevation $\hat{\theta}$, $F(3, 948) = 0.61$, $p < 0.610$; $F(2, 948) = 0.113$, $p = 0.893$. There was no statistically significant interaction between shape and material, $F(3, 948) = 0.74$, $p = 0.817$, $\alpha = 0.01$.

	Surface material			
Shape	Shiny	Texture	Uniform	Silhouette
Smooth	18.88 $_{SE=1.73}$	15.23 $_{SE=1.61}$	16.64 $_{SE=2.04}$	45.17 $_{SE=6.37}$
Medium	15.46 $_{SE=1.80}$	14.89 $_{SE=1.71}$	15.89 $_{SE=1.71}$	48.39 $_{SE=6.77}$
Complex	15.80 $_{SE=1.74}$	17.74 $_{SE=1.96}$	19.07 $_{SE=1.92}$	57.99 $_{SE=7.47}$

Table 1. Mean angular error. *Notes:* Mean angular errors ε and SEs are shown for three objects and four material types. Note that the mean angular error for smooth and medium modulated *shiny* and *uniform* objects tended to be slightly larger than for *textured* objects. This trend is consistent with the argument that material-specific image motion may affect only the perceived rotation axis when the shape under consideration is sufficiently simple. To test this idea, we employed isotropic objects in Experiment 2.

Azimuth settings

There were no statistically significant main effects of surface material and object shape on perceived rotation axis azimuth $\hat{\phi}$, $F(3, 948) = 0.830$, $p = 0.478$; $F(2, 948) = 0.613$, $p < 0.542$. There was no statistically significant interaction between shape and material, $F(3, 948) = 0.252$, $p = 0.958$, $\alpha = 0.01$.

Discussion

Silhouettes

We found the angular error for the *silhouette* condition to be significantly larger than for any other material type. This was not surprising, given that all the information is restricted to contour-generated motion. Such stimuli have been shown to cause—in the best case—ambiguous perception of the direction of rotation, and can—in the worst case—be perceived as deforming nonrigidly (Sinha and Poggio, 1996; Wallach & O'Connell, 1953). The variability in settings for the *silhouette* condition was also substantially larger than for the other conditions. Interestingly, we find that although the other three conditions have unimodal error distributions, the error distribution for the *silhouette* condition appears to be bimodal (Figure 4). This presumably reflects an inherent ambiguity about the direction of rotation for completely homogeneous silhouettes: Subjects could estimate the axis of rotation broadly correctly but in some cases confused clockwise and anticlockwise rotations (which are equivalent to flipping the axis of rotation by 180°). The finding that observers are able to estimate the rotation axis for *silhouette* objects is consistent with earlier results by Norman et al. (2004), Norman and Todd (1994), and Todd (1985), who showed that observers can estimate 3D shape from optical deformations that violate the correspondence constraint,[4] including occluding boundaries and smooth shading gradients. In fact, if it were not for the sign of the rotation confusion, observers' performance would have been rather similar to the remaining conditions.

Shiny and textured objects

Surprisingly, despite the compelling demonstration in Figure 1, we did not find the expected difference in angular error between *shiny* and *textured* objects. This might be explained by the fact that our shapes—even the smoothest one—had prominent regions of high curvature. Specular features tend to stick to these regions; thus, the image motion that these points generate is very similar to that generated by texture markings. Consequently, the resulting optic flow

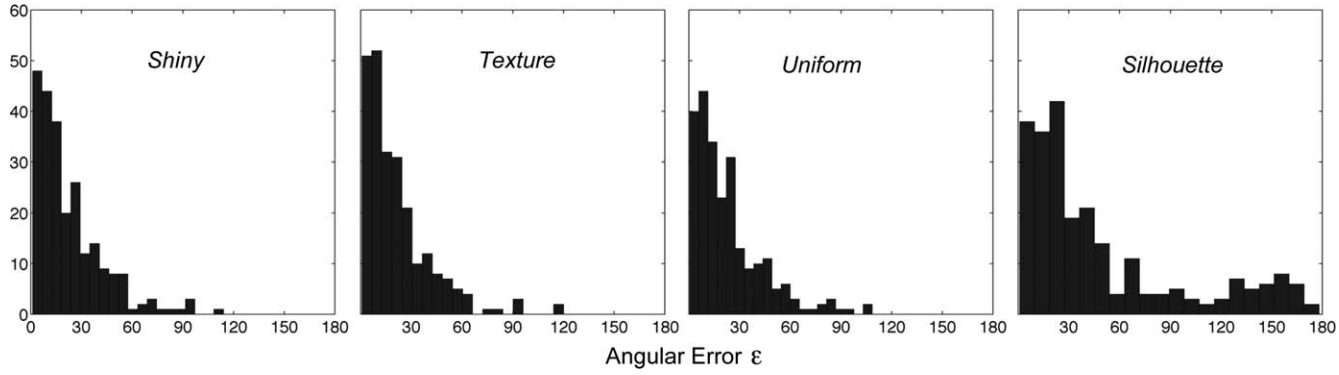

Figure 4. Experiment 1 distribution of angular errors. In contrast to all other material conditions, the error distribution for *silhouettes* is bimodal, a pattern that would be consistent with an observer that estimates the axis of rotation correctly while judging the direction of rotation wrong. This is in line with previously observed bistability of rotation direction of silhouette objects, as seen, for example, in the well-known Spinning Dancer Illusion by Nobuyuki Kayahara.

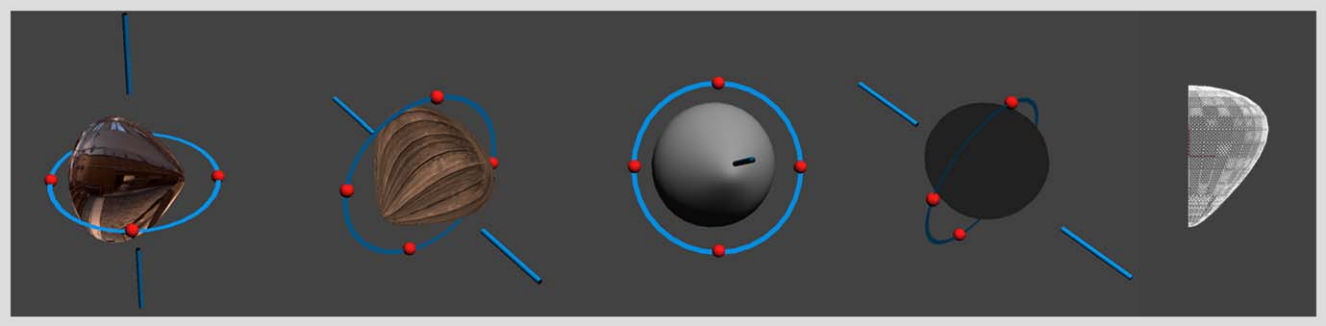

Figure 5. Experiment 2 stimuli. Shown are sample trial snapshots, including the noninvasive probe, across different surface material conditions for the *top*-shaped isotropic object: *shiny*, *texture*, *uniform*, and *silhouette*, from left to right. The right-most panel shows the cross section of the object.

patterns may provide a sufficient number of trackable features to establish corresponding features and thus to support the estimation of the rotation axis. Although these features are sparse, the visual system may be able to infer global motion from their coherent motion, in much the same way as a sparse set of sparse, but coherently moving, texture markings can support a global interpretation of object rotation (Koenderink & Van Doorn, 1980, 1982; Todd, 1985).

Uniform objects

The above argument would be consistent with the fact that we found no difference in ε between the *uniform* and the *textured* objects. Although it has been shown that 3D shape can be estimated from the optical deformations of smooth shading generated by rotating *uniform* objects (Koenderink & Van Doorn, 1979, 1980, 1982; Todd, 1985), it is also true that the motion of texture elements that are stuck to the object surface does provide important additional cues to the motion of the object. To take an extreme example, the rotation of a perfectly *uniform* sphere about its vertical (or any) axis would be invisible because there would be no optic flow created by the object motion. However, if mesoscale geometrical features or texture markings were added to the *uniform* sphere, the motion of the resulting patterns would unambiguously reveal the rotation axis and other characteristics such as the sign and speed of the rotation.

Given these considerations, one might argue that material-specific image motion, especially the differences between *shiny* and *textured* motion patterns, may affect only the perception of object rotation axis when the shape under consideration is simple enough such that the 2D texture would provide observers with a clear advantage (less ambiguity) over the *uniform* case. A class of 3D shapes that meets this requirement are surfaces of revolution, such as the pot in Figure 1. We conducted a second experiment using this type of object.

Probe design

An additional factor contributing to the lack of difference in ε between conditions might have been the somewhat invasive nature of the probe we used: It intersected with the object boundaries, which might have provided additional cues to the observers, allowing them to perform the task well and somewhat independent of the material category. To eliminate this possibility, we improved the probe so that it did not intersect with the actual object in Experiment 2 (Figure 5).

Elevation and azimuth settings

We found no effect of *material* on estimated rotation axis elevation and azimuth ($\hat{\theta}$, $\hat{\phi}$), yet the *material* effect is present when inspecting observers' angular error patterns. This result implies that rotation axis direction estimations were not systematically different between materials but were simply more variable for some conditions (e.g., for silhouette objects). However, this does not mean that these systematic differences do not exist. Material-specific differences in $\hat{\theta}$ and $\hat{\phi}$ may depend on the true rotation axis direction, and by averaging results from randomly sampled rotation axis directions, such a pattern would not emerge. Therefore, a systematic study of how the true rotation axis elevation and azimuth modulate the effect of *material* on estimated rotation axis direction would be a more sensible design, which we explore in Experiment 2.

Experiment 2

The absence of an effect of *material* on perceived rotation axis in Experiment 1 suggests that material-specific image motion differences may affect only the perceived axis of rotation of objects whose 3D shape is simple enough that a 2D *texture* map would substantially disambiguate the object motion—relative to a

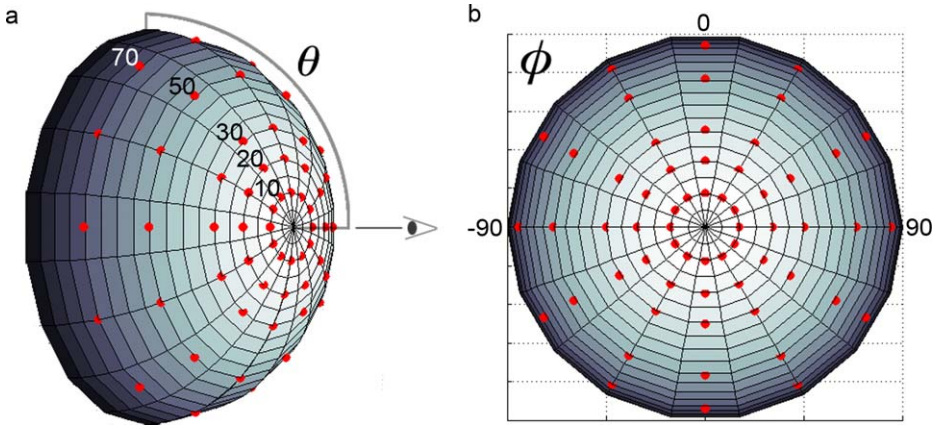

Figure 6. Experiment 2 rotation axis samples. We systematically measured the effect of surface material on perceived rotation axis across 60 locations (red dots) on the unit hemisphere. (a) Shows a perspective view of the sampled rotation axis, illustrating the θ dimension (elevation). (b) A top-down view onto the rotation axis samples, illustrating the ϕ dimension (azimuth). One should visualize the rotating object to be located at the center of the sphere, with the rotation axis aiming through the object's center coordinate.

uniform albedo (*uniform*) version of the same object. Thus, we repeated Experiment 1 using a rotationally symmetric (isotropic) object (Figure 5). Under these conditions (i.e., without easily trackable surfaces features in the *uniform* case), we expected the effect of surface material on perceived rotation axis to emerge. Moreover, we expected that the effect may be measurable for only a certain range of rotation axes. For example, rotations around an axis along the line of sight ($\theta = 0$) might yield similar settings across materials, due to the unambiguous contour motion cues, whereas for more oblique $0 < \theta < 90$ values, the material-specific motion cues may yield different, material-dependent rotation axis estimates. We examined the effect of surface material on perceived rotation axis systematically for 60 rotation axis directions (Figure 6), in the range from $10 < \theta < 70$, for which we expect greatest ambiguities. In this experiment, the probe did not intersect with the object (Figure 5), and we restricted the object rotation to cycling through 20°, to eliminate the bimodal nature of the error in the *silhouette* condition, thus making it more comparable to the other materials. This adjustment changes the computation of ε slightly:

$$\varepsilon = argmin\left(\arccos\left(\frac{o \cdot \hat{o}}{||o||\,||\hat{o}||}\right), \pi - \arccos\left(\frac{o \cdot \hat{o}}{||o||\,||\hat{o}||}\right)\right). \quad (2)$$

Stimuli

We choose a simple, isotropic object "*top*" (Figure 5). Sixty rotation axes were sampled from the unit hemisphere that varied in azimuth ϕ: −180, −150, −120, −90, −60, −30, 0, 30, 60, 90, 120, 150 and elevation θ: 10, 20, 30, 50, 70 (see Figure 6). These axes were chosen to be most informative (i.e., away from $\theta = 0$ and $\theta = 90$). The remaining stimuli details were as in Experiment 1.

Procedure and observers

The procedure was the same as in Experiment 1. Seven observers (six naive, one author [K. D.]) completed 240 trials (5 elevations × 12 azimuths × 4 materials) in four sessions. Each session lasted about 15 min.

Results

Angular error

Figure 7 shows the average error across all trials for each material condition. Unlike Experiment 1, we see clear differences in accuracy between materials, with *texture* yielding the lowest angular errors and *silhouette* and *shiny* yielding the worst performance. Interestingly, performance in the *shiny* condition was essentially as bad as just seeing the silhouette on its own. However, as we discuss below, this was not due to lack of information (for example, because subjects could not estimate flow for the specular surfaces, as in the *silhouette* condition) but because they systematically misinterpreted the flow when trying to estimate the axis of rotation.

In Figure 8, we break down the errors as a function of azimuth and elevation, to work out which orientations were most problematic for the observers. Darker

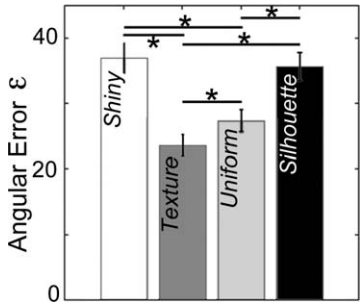

Figure 7. Experiment 2 results. Mean data of seven observers. Using a simple isotropic object, an interesting pattern emerges: *Shiny* objects tend to elicit a larger ε than *textured* and *uniform* ones. Also, ε for *uniform* objects was significantly larger than for *textured* shapes. Error bars are two times the standard error of the mean.

shades indicate worse performance. The fact that the center of the plots are brighter for all four materials indicates that observers found low elevation angles (i.e., axes that are close to pointing at the observer, along the line of sight) easier to estimate than ones that were more slanted. This makes sense because at zero slant, the outline of the object rotates rigidly in the 2D view plane, providing relatively unambiguous information about the rotation axis.

Interestingly, the pattern of errors is somewhat different for *shiny* and *silhouette* conditions, despite similar average performance. For the *shiny* condition, error falls off rapidly with increasing elevation but is not as bad as the *silhouette* condition for large values of elevation. By contrast, for the *silhouette* stimuli, performance is good for a wider range of low-elevation angles but falls off precipitously beyond elevations of $\theta > 30°$ to substantially lower accuracy than in the *shiny* condition. This, again, indicates that poor performance in the *shiny* condition is not due to lack of information; otherwise, we would expect a similar pattern of errors as in the *silhouette* condition.

The $5 \times 12 \times 4$ three-way ANOVA indicated that there was a statistically significant main effect of elevation θ, $F(4, 1440) = 248.83$, $p < 0.0001$; no significant main effect of azimuth ϕ, $F(11, 1440) = 2.01$, $p = 0.019$, $\alpha = 0.01$; and a significant main effect of surface material on angular error, $F(3, 1440) = 70.71$, $p < 0.0001$. Further, there was a significant two-way interaction of θ and material, $F(12, 1440) = 4.11$, $p < 0.0001$. The remaining two-way and the three-way interactions were not significant.

Post hoc analysis of the main effect "material" indicated that the estimation error in the *shiny* and *silhouette* conditions was significantly larger than in the other conditions at the $\alpha = 0.05$ level, adjusted for multiple comparisons using Scheffe's S procedure. The average angular error for the *textured* object was significantly smaller than that of all other material conditions, and the average angular error for *uniform* objects was larger than for textured ones (see Figure 7).

Observers' average estimation errors ε for each material condition ($\varepsilon_{shiny} = 36.71$, $\varepsilon_{texture} = 23.47$, $\varepsilon_{uniform} = 27.18$, $\varepsilon_{silhouette} = 35.39$) were each significantly smaller than chance performance ($\varepsilon_{chance} = 45^5$, $t_{shiny}[419] = -7.45$, $p < 0.0001$; $t_{texture}[419] = -26.31$, $p < 0.0001$; $t_{uniform}[419] = -21.45$, $p < 0.0001$; $t_{silhouette}[419] = -9.01$, $p < 0.0001$, one sample t-tests, two-tailed, Bonferroni corrected $\alpha = 0.0125$).

Post hoc analysis of the main effect "elevation" revealed that the estimation error systematically and significantly increased with increases in theta, at the $\alpha = 0.05$ level, adjusted for multiple comparisons using Scheffe's S procedure. Observers' average estimation error ε for each θ condition ($\varepsilon_{\theta=10} = 13.91$, $\varepsilon_{\theta=20} = 22.44$, $\varepsilon_{\theta=30} = 29.05$, $\varepsilon_{\theta=50} = 41.43$) were each significantly better than chance performance ($\varepsilon_{chance} = 45$; $t_{\theta=10}[335] = -42.51$, $p < 0.0001$; $t_{\theta=20}[335] = -27.78$, $p < 0.0001$; $t_{\theta=30}[335] = -21.45$, $p < 0.0001$; $t_{\theta=50}[335] = -4.12$, $p <$

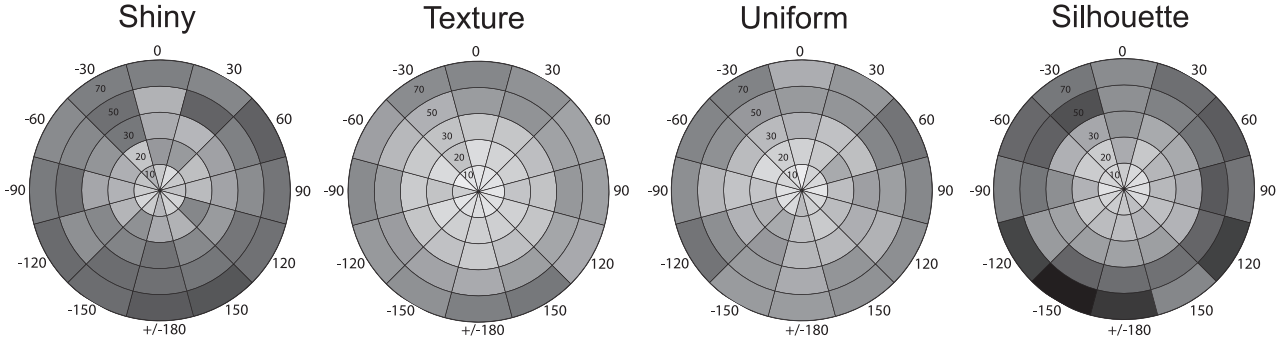

Figure 8. Experiment 2 angular error. *Shiny*, *textured*, *uniform*, and *silhouette* normalized ε averages from seven observers are plotted as a function of true θ and ϕ. Darker shades of gray indicate larger error. It is evident that *shiny* and *silhouette* conditions yielded the largest ε. Across conditions, larger deviations from θ were associated with a larger ε. There was no systematic difference in ε across ϕ.

	True θ				
Material	10	20	30	50	70
Shiny	18.57	29.69	37.82	46.64	50.83
Texture	9.74	16.08	20.05	32.65	38.81
Uniform	12.78	21.23	26.10	35.15	40.66
Silhouette	14.55	22.78	32.25	51.31	56.07

Table 2. Material × Angular Error ε interaction. *Notes:* Marginal ε means across material and θ conditions. Values indicate that this interaction is mainly driven by the differential changes in average angular error ε for each material category as a function of change in θ true. For example, ε_{shiny} is about twice as big as $\varepsilon_{texture}$ for θ_{10}, but this proportional difference is quite different for θ_{70}.

0.0001 one sample t tests, two-tailed, Bonferroni corrected $\alpha = 0.0125$). The average estimation error for the $\varepsilon_{\theta=70} = 46.59$ was not different from chance performance, $t_{\theta=70}(335) = 1.2$, $p = 0.201$.

Following up the interaction θ of and material on ε with a two-way ANOVA yielded significant main effects of θ, $F(4, 1660) = 249.13$, and material, $F(3, 1660) = 70.79$, $p < 0.0001$, on $\hat{\theta}$ and a significant interaction, $F(12, 1660) = 4.12$, $p < 0.0001$, $\alpha = 0.01$. This interaction occurred due to the differential changes in average angular error ε for each material category as a function of change in true θ (see Table 2).

Although the amount of error that observers would make as a function of material category was our primary variable of interest, the systematic variation of

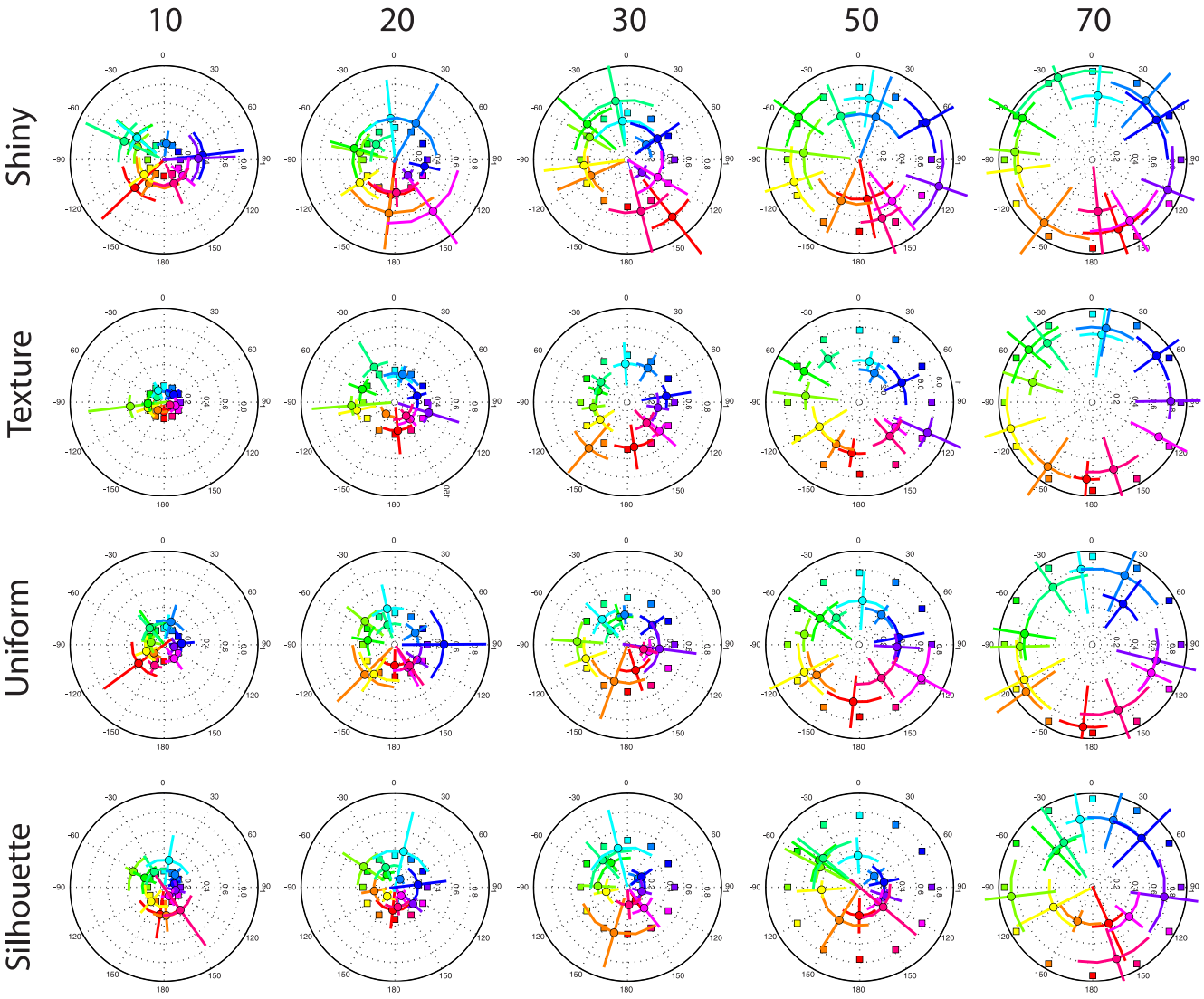

Figure 9. Experiment 2 estimated rotation axis directions. Shown are the average settings of seven observers. Rows denote material conditions; columns denote θ conditions. Square symbols indicate the ground truth; circles are average observer data. The azimuth in each polar plot is color coded. To assess correspondence between the ground truth and observer setting, one has to locate the same-color square and circular symbols. Error bars are two times the standard error of the mean. See the text for details.

the true rotation axis direction in Experiment 2 allowed assessment of how observers' settings varied in elevation θ and azimuth ϕ across material conditions. Thus, we conducted three-way ANOVAs also for $\hat{\theta}$ and $\hat{\phi}$.

Elevation settings

Figure 9 plots estimated rotation axis directions as a function of true θ and ϕ for each material condition. In all four materials, there is a tendency to underestimate large values of elevation, especially when $\theta = 50°$. The errors in the estimation of elevation do not appear to be clustered around specific azimuthal angles. For example, there is no clear cardinal axis effect, suggesting that subjects are relying more on image information than priors for these stimuli. There was a statistically significant main effect of elevation θ, $F(4, 1440) = 166.06$, $p < 0.0001$; no significant main effect of azimuth ϕ, $F(11, 1440) = 1.51$, $p = 0.12$; and a statistically significant main effect of surface material, $F(3, 1440) = 27.52$, $p < 0.0001$, $\alpha = 0.01$, on $\hat{\theta}$ settings. In addition, there was a significant two-way interaction of θ and material, $F(12, 1440) = 2.27$, $p = 0.008$. The remaining two-way and three-way interactions were not significant.

Post hoc analysis of the main effect "elevation" indicated that $\hat{\theta}$s varied systematically with changes in true θ. The differences between levels of $\hat{\theta}$ were significant except between $\theta = 20°$ and $\theta = 30°$. Observers' average "elevation" estimations were $\hat{\theta}_{10} = 13.16$; $\hat{\theta}_{20} = 18.91$; $\hat{\theta}_{30} = 22.00$; $\hat{\theta}_{50} = 31.32$; and $\hat{\theta}_{70} = 49.52$. With the exception of $\hat{\theta}_{20}$, all of these estimates were significantly different from the respective ground truth values of θ ($t_{\theta=10}[335] = 3.75$, $p < 0.0001$; $t_{\theta=30}[335] = -7.59$, $p < 0.0001$; $t_{\theta=50}[335] = -14.14$, $p < 0.0001$, one sample t tests, two-tailed, Bonferroni corrected $\alpha = 0.01$).

Post hoc analysis of the main effect "material" indicated that $\hat{\theta}$ for *shiny* objects ($\hat{\theta}_{shiny} = 33.31$) was significantly larger than for any of the other material conditions and that $\hat{\theta}$ in the *silhouette* condition ($\hat{\theta}_{silhouette} = 20.68$) was significantly smaller than in any of the other material conditions, at the $\alpha = 0.05$ level, adjusted for multiple comparisons using Scheffe's S procedure. The average $\hat{\theta}$ for *textured* and *uniform* objects was not significantly different ($\hat{\theta}_{texture} = 27.47$, $\hat{\theta}_{uniform} = 26.47$).

Following up the interaction of θ and material on $\hat{\theta}$ with a two-way ANOVA yielded significant main effects of θ, $F(4, 1660) = 166.03$, $p < 0.0001$, $\alpha = 0.01$, and material, $F(3, 1660) = 27.52$, $p < 0.0001$, on $\hat{\theta}$ and a significant interaction, $F(12, 1660) = 2.27$, $p = 0.008$, $\alpha = 0.01$. This interaction occurred due to the differential changes in $\hat{\theta}$ for each material category as a function of change in true θ (see Table 3).

	True θ				
Material	10	20	30	50	70
Shiny	19.60	24.62	28.06	40.32	53.95
Texture	8.19	18.47	24.00	32.68	54.05
Uniform	11.79	19.46	19.80	31.28	50.02
Silhouette	13.07	13.12	16.14	20.99	40.07

Table 3. Material $\times$ θ interaction on $\hat{\theta}$. *Notes:* Marginal $\hat{\theta}$ means across material and θ conditions. Values indicate that this interaction is due to the differential changes in $\hat{\theta}$ for each material category as a function of change in true θ. For example, $\hat{\theta}$ is about twice as large as $\hat{\theta}$ for θ_{10}, but they are approximately the same for θ_{70}.

Azimuth settings

There was no statistically significant main effect of elevation θ, $F(4, 1440) = 2.55$, $p = 0.037$; a statistically significant main effect of azimuth ϕ, $F(11, 1440) = 1007.22$, $p < 0.0001$; and no significant main effect of material, $F(3, 1440) = 2.50$, $p = 0.053$, $\alpha = 0.01$, on $\hat{\phi}$ settings. There was a significant two-way interaction of θ and material, $F(12, 1440) = 2.99$, $p < 0.0001$. The remaining two-way and the three-way interactions were not significant.

Although not all paired comparisons were significant, post hoc analysis of the main effect "azimuth" indicated that $\hat{\phi}$ increased systematically and significantly with increases in ϕ. The differences between levels of $\hat{\phi}$ were significant except between pairs $\phi = -90$ and $\phi = -60$ and $\phi = -60$ and $\phi = -30$, $\alpha = 0.05$, adjusted for multiple comparisons using Scheffe's S procedure. $\hat{\phi}$ means (-180.43, -143.84, -115.01, -56.64, -37.83, -5.24, 29.34, 67.94, 153.59) were not significantly different from the respective true ϕ values (-180, -150, -120, -60, -30, 0, 30, 60, 150). Azimuth estimates for $\phi = -90$, 90, and 120 were significantly different from true ϕ values, $\hat{\phi} = -78.38$, ($t[139] = 4.29$, $p < 0.0001$), $\hat{\phi} = 108.64$, ($t[139] = 5.68$, $p < 0.0001$), $\hat{\phi} = 131.57$, ($t[139] = 3.407$, $p < .001$), (one sample t tests, two-tailed, Bonferroni corrected $\alpha = 0.0042$).

Following up the interaction of θ and material on $\hat{\phi}$ (see Table 4) with a two-way ANOVA yielded no significant main effects of θ and material on $\hat{\phi}$ and no significant interaction.

Discussion

Angular error

We observed significant differences in rotation axis estimation errors ε between all material conditions except between *shiny* and *silhouette*. In general, the emergence of this effect in Experiment 2 suggests that it depends crucially on the 3D shape complexity of the object. In our case, the object had, in fact, only positive

	True θ				
Material	10	20	30	50	70
Shiny	−14.26	−7.15	−2.85	−9.81	−5.61
Texture	−23.73	−11.36	−16.08	−11.74	−12.22
Uniform	3.46	−4.54	−25.69	−9.48	−15.98
Silhouette	−10.95	4.13	−10.86	−8.12	−17.66

Table 4. Material × θ interaction on $\hat{\phi}$. *Notes:* Marginal $\hat{\phi}$ means across material and θ conditions. Values suggest that this interaction is due to the differential changes in $\hat{\phi}$ for each material category as a function of change in true θ. For example, $\hat{\phi}_{shiny}$ becomes less negative from θ_{10} to θ_{70}; the opposite pattern occurs for $\hat{\phi}_{uniform}$. Note, however, that this interaction was not significant in the follow-up analysis.

(but not constant) curvature, resulting in a purely convex occluding boundary. This constitutes a particular difficult condition for the extraction of 3D shape from object boundaries during object rotation (Koenderink & Van Doorn, 1976; Todd, 1985). Consequently, the largest errors are made in the *silhouette* condition. However, as in Experiment 1, observers performed significantly better than chance in this condition, further supporting the argument that 3D structure and the rotation axis direction can be estimated from this class of stimuli (Koenderink, 1984; Norman & Todd, 1994; Norman et al., 2004; Todd, 1985).

Interestingly, errors made for *shiny* objects were just as large as for silhouettes, suggesting that the motion of specular reflections does not disambiguate occluded boundary motion. However, this does not imply that observers made the *same* estimation errors in these two conditions. In fact, a quick inspection of Figure 9 shows that the pattern of errors is quite different for these two conditions. Figure 10 illustrates this more clearly: Whereas the difference between ε for *shiny* and *textured* objects is largest for small values of θ, the opposite pattern occurs for the difference between *silhouettes* and *textured* objects. This was also in part responsible for the observed interaction of *material* and θ in our analysis. We will return to the root of this difference below when discussing *material* dependent differences in rotation axis elevation and azimuth estimation.

Textured objects yielded the smallest ε (about 23°; Figure 7), which confirms that trackable features and texture gradients indeed provided rich cues to 3D structure and rotation axis direction estimation. Note, however, that ε systematically increased with increases in θ for all four material conditions, suggesting that at oblique (in slant) rotation axis, directions are more likely to be misestimated. Because we recorded elevation and azimuth, we will be able to characterize the nature of these increases in ε more precisely below.

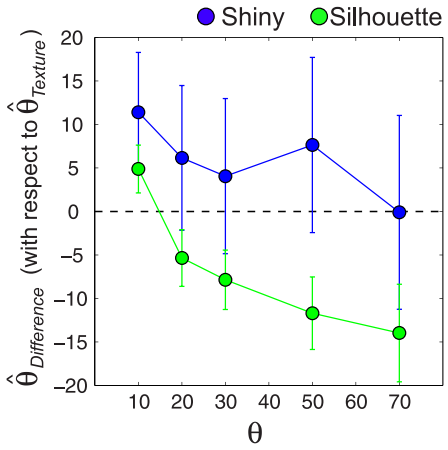

Figure 10. Experiment 2 differences in $\hat{\theta}$ for *shiny* and *silhouette* objects. We plot average $\hat{\theta}_{shiny} - \hat{\theta}_{texture}$ (blue) and $\hat{\theta}_{silhouette} - \hat{\theta}_{texture}$ (green) as a function of θ. $\hat{\theta}_{shiny} - \hat{\theta}_{texture}$ is larger for small and smaller for large θ; however, the opposite pattern occurs for $\hat{\theta}_{silhouette} - \hat{\theta}_{texture}$. This plot highlights that the source of the angular estimation error was rather different between *shiny* and *silhouette* conditions, even though the overall error magnitude was the same. Because we were ultimately interested in differences between materials and not veridicality of settings, we chose $\hat{\theta}_{texture}$ as the baseline to compare against. Error bars are one standard error of the mean.

Interestingly, ε for *uniform* objects was only slightly but significantly larger (about 27°; Figure 7) than for *textured* objects and far smaller than for *shiny* objects or *silhouettes*. This finding is different from the results by Norman et al. (2004), who found no difference in shape discrimination threshold for rotating textured and shaded objects. This discrepancy might be explained, however, by the fact that their objects were much more complex than our isotropic shapes, providing a richer set of cues in the shading condition (Koenderink & Van Doorn, 1980, 1982). In general, these results confirm that 3D shape, and the rotation axis direction, can be extracted from the deformation of shading gradients during object motion (Norman et al., 2004; Norman & Todd, 1994; Todd, 1985).

Across materials, results indicate that the magnitude of estimation errors depended critically on θ (elevation), not on ϕ azimuth. Although the majority of previous studies have in fact manipulated elevation (or slant, e.g., Norman & Todd, 1994; Pollick et al., 1994; Todd, 1985) and found substantial estimation errors, tilt estimation has not received much attention. Caudek and Domini (1998) found that tilt estimation was not different from ground truth. Our results seem to support this finding, namely, that azimuth contributed little to the rotation axis estimation error. Because we found no interaction of material type and azimuth levels, the estimated azimuth $\hat{\phi}$ must have been close to veridical across materials and variations in θ, a

suggestion that we will examine below. This latter result is somewhat of a surprise, given the demonstration in Figure 1, which clearly suggest misestimation of ϕ, not θ. We will discuss below (Azimuth section) how this might be explained.

Elevation

Shiny objects: Figure 9 shows observers' rotation axis direction estimates as a function of both θ and ϕ. Several things become apparent: The $\hat{\theta}$ patterns closely mirror those of ε in Figure 8 (i.e., large ε for *shiny* or *silhouette* objects tend to translate into large deviations of $\hat{\theta}$ from θ).

Consider, for example, ε for shiny and silhouette objects (Figure 7). Inspecting Figure 10, it becomes apparent that the reason of larger ε is quite different across material category: Whereas ε for *shiny* objects is large due to an overestimation of θ, a larger ε for *textured* objects is mostly due to a systematic underestimation of θ.

In previous studies that used silhouettes or smoothly shaded rotating objects, underestimation of θ (slant) has been a prevalent finding (e.g., Pollick et al., 1994; Todd, 1985, when adding unconstrained noise). Although our findings for *textured*, *uniform*, and *silhouette* objects agree with this, the slant estimation of the rotation axis for *shiny* objects does not follow this pattern, at least for small values of θ (e.g., $\hat{\theta}_{10} = 19.6$ and $\hat{\theta}_{20} = 24.6$).

What might cause this pronounced overestimation? As outlined in the Introduction, rotating specular objects combine boundary and specular image motion. Although the former, and the associated optical deformations, produces image velocities that largely depend on the rotation speed of the object, the latter produces image velocities that are inversely related to 3D curvature magnitude (Koenderink & Van Doorn, 1980). Moreover, the direction of specular feature motions is critically related to the shape, moving toward high surface curvature points, causing large translational displacement across the extent of the object. For *textured* objects, such translation image patterns are more typical for rotations around the axis with a larger slant; however, for *shiny* objects, and depending on the 3D structure, these motion patterns may also occur at small θ, and this may bias the observer to overestimate the slant. This possibility should be examined in future work.

Textured and uniform objects: *Textured* objects produced the lowest error rates. Figure 9 illustrates that $\hat{\theta}$s for *textured* and *uniform* objects are very similar. This observation was confirmed by our statistical analysis that found no significant differences between these two conditions, confirming that observers are able to use different types of optical deformations to estimate the rotation axis (and 3D shape) and not only those that satisfy projective correspondence (Norman et al., 2004; Norman & Todd, 1994; Pollick et al., 1994; Todd, 1985).

Silhouettes: Although, on average, all material conditions yielded underestimated θ, *silhouettes* tended to have the most pronounced underestimation (Figure 9). To the best of our knowledge, this has not been reported explicitly before. For any two successive time points during the *silhouette* rotation, any given point on the contour will not belong to the same point on the object's surface, which makes estimation of 3D shape problematic to begin with, if, for example, trying to establish projective correspondence. Moreover, partial rotations (less than 360°) would theoretically not allow recovery of rotation axis slant (Giblin, Pollick, & Rycroft, 1994). Given these considerations, it is surprising that observers were able to perform the task at all. Low-complexity silhouettes, as the object in Experiment 2, are more likely to produce nonrigid percepts (Todd, 1985), that is, 2D motion, yet we asked to observer to perform a 3D task. Two-dimensional and 3D motion patterns are in agreement when a silhouette rotates around the viewing axis. Given the ambiguity of the contour point loci, the predominantly 2D motion appearance, and the forced 3D judgment, observers' perceptions might have been biased toward smaller slant angles θ (i.e., rotation axes closer to the view direction).

Across all materials, we found that the variability of $\hat{\theta}$ gets larger for larger values of θ.

Azimuth

Azimuth settings $\hat{\phi}$ were tightly clustered around ground truth values (Figure 9), and there was no difference in settings between material conditions and across elevation levels. One could be tempted to conclude that the effects of surface material on rotation axis estimation seem to be limited to the θ component of the rotation axis direction; however, the demonstration in Figure 1 seems to suggest otherwise, that is, that the specular feature motion should bias rotation axis azimuth just as much as it biased the elevation. Why did we not see this effect? It might have to do with the design of our experiment: As discussed above, the direction and velocity of specular feature motion are critically related to the shape of the object. Because the intrinsic orientation of the object was randomly changed on every trial, so would be the visible 3D geometry and the associated specular feature motion. Therefore, any systematic effect of azimuth might have been washed out. Consequently, if one presented the same object with the same view, one should be able to detect these postulated effects of specular feature motion on $\hat{\phi}$, a possibility we will examine below in Experiment 3.

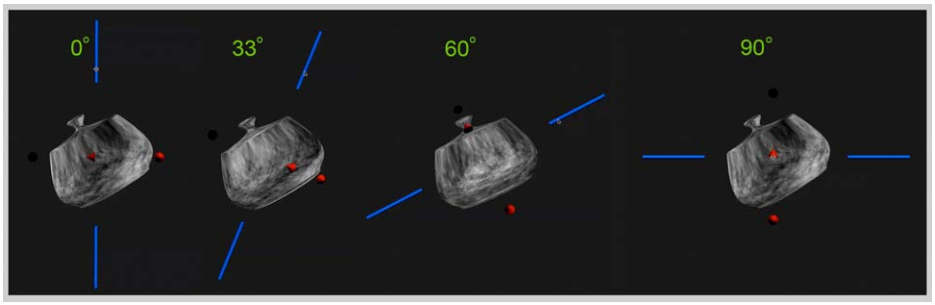

Figure 11. Experiment 3 stimuli. Stimuli were the *shiny* and *textured* (sticky reflections) versions of a *pot*. The object rotated back and forth through 20° at 24°/s around one of four possible rotation axes, 0°, 33°, 63°, and 90°, as indicated by the probe orientation.

Overall, the near-veridical values of $\hat{\phi}$ across conditions were surprising given earlier results by (Pollick et al., 1994), who found quite large misestimations of ϕ, especially for naive observers. This might be due to differences in experimental design: In those experiments, observers used their finger to indicate the direction of the rotation axis.

Taken together, the results of Experiment 1 and 2 imply close coupling of surface material–specific optic flow, 3D shape, and rotation axis estimation. Although we have so far obtained and discussed results from four material conditions, the focus of this article is on the differences between *shiny* and *textured* conditions (Figure 1), especially how differences in perceived rotation axis might be related to the extraction of 3D shape for each material category. Thus, in Experiment 3, we limited our investigation to these two material categories.

Experiment 3

When estimating the orientation of the rotation axis, the visual system conjointly extracts surface material properties, object motion, and 3D shape. Thus, in principle, a model based on SfM should be able to predict observers' percepts. There are a number of caveats, however: for typical SfM algorithms to successfully recover shape, the object under question has to be rigid, must be of sufficient complexity, and/or have a rich 2D texture (to establish corresponding points) and should reflect light diffusely (Huang & Netravali, 1994). Because *shiny* objects violate these assumptions (note that smooth, *uniform* objects and *silhouettes* violate these assumptions as well), we might expect SfM to be problematic. In fact, Doerschner et al. (2011) and Swaminathan, Kang, Szeliski, Criminisi, and Nayar (2002) recently showed that under generic conditions (object complexity and motion), violation of epipolar geometry (a prerequisite to recover SfM) is highly diagnostic for specular surfaces. However, rigidly moving specular objects of noncomplex geometry may not necessarily violate epipolar geometry (Doerschner, Fleming et al., 2011); thus, we expect SfM to be possible,[6] and in that case, we could extract the camera motion between frames. For nonshaded, matte, textured objects, camera (or observer) motion is analogous to object motion if the object is rotating around a single axis, as is the case in our experiment. Thus, in principle, an SfM-based model could compute the object rotation axis from the extracted camera motion, as we propose below. To test this idea, we compare computational rotation axis estimates derived from an SfM algorithm to observers' percepts for *shiny* and *textured* objects.

Unlike *shiny* objects, which have a dense (specular) texture suitable for feature tracking, *uniform* (simple) objects and *silhouettes* do not have a rich texture that allows us to establish projective correspondence. Therefore, an SfM-based model would most likely fail for these material categories. We realize that this limits the generalizability of our proposed model. However, the point we wish to make is that the estimation of rotation axis direction and 3D shape are closely related and that both the shape and rotation axis estimates can be biased by surface reflectance. Regardless of how 3D structure is extracted from image motion (through correspondence or by some other mechanism, e.g., Koenderink & Van Doorn, 1979), this relationship should not only be true for *shiny* and *textured* but also for *uniform* and *silhouette* objects. Our data from Experiment 2 support this reasoning.

Stimuli

We chose to test our algorithm on the original illusion illustrated in Figure 1. Stimuli were the *shiny*, *textured* version of the pot in Figure 1 rotating back and forth through a total excursion of 20° at 24°/s around one of four possible rotation axes in the frontoparallel plane ($\theta = 90$, $\phi = 0$, 33, 63, 90; Figure 11).[7] We phase-scrambled the Debevec "Grace" environment map (Debevec, 2002) using spherical harmonics and used it as a texture map for *textured* stimuli and as an environment map for *shiny* stimuli.

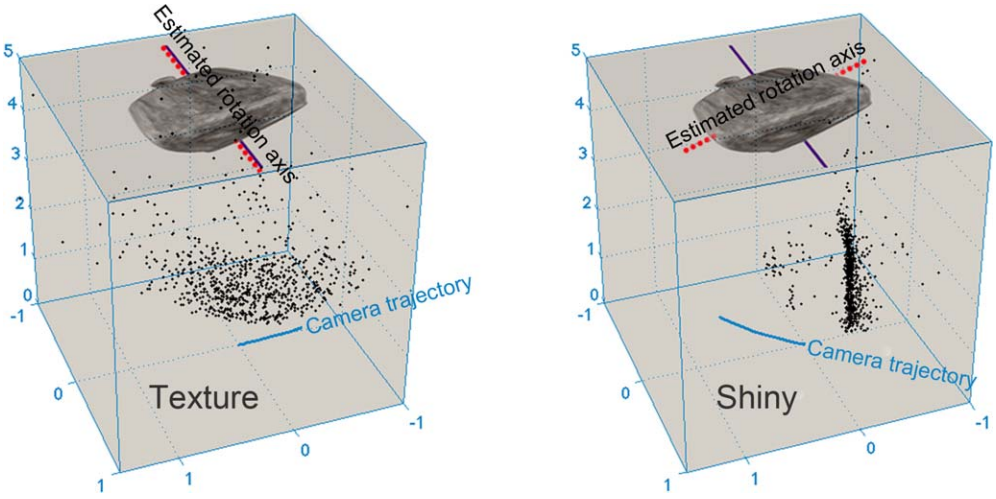

Figure 12. Model. The rotation axis of an object can be estimated from camera motion for *textured* and *shiny* objects of noncomplex geometry using our SfM model. The dots are trackable image features on the basis of which 3D shape is reconstructed and camera trajectory is estimated (blue). Overlaid is one frame from the corresponding matte and shiny sequences with the ground truth ϕ orientation of the rotation axis indicated by the purple line. Note the angular difference in estimated camera trajectory and rotation axis (red, dotted line) between the *textured* and *shiny* object.

Procedure and observers

The same procedure as in Experiment 1 was used. Seven observers completed eight trials (2 materials × 4 rotation axes).

Model

The model involves tracking of corresponding features (SIFT features; Lowe, 2004) between frames (Beis & Lowe, 1997), obtaining the 3D coordinates of a few features and estimating the camera position from these. Given the known camera positions, it is possible to extract the 3D shape and the rotation vector for each camera position, from which the object rotation axis can be computed. The motivation and computational details of the model are given in the Appendix. Figure 12 shows the estimated rotation axis for $\phi = 0°$.

Analysis

We compared observers' estimates $\hat{\theta}_{shiny}$, $\hat{\phi}_{shiny}$, $\hat{\theta}_{texture}$ $\hat{\phi}_{texture}$ to θ and ϕ as well as to the predictions made by the SfM-based model. For completeness and to explicitly test the limitations of our model, we obtained rotation axis estimates for *uniform* and *silhouette* versions of the stimuli. Given the feature-tracking nature of our algorithm, however, we expected to not get very reliable estimates for these classes of stimuli.

Results

Elevation settings

There were no statistically significant main effects of surface material and ϕ on $\hat{\theta}$ (2 × 4 (material × rotation axis) two-way ANOVA, $F(1, 48) = 0.29$, $p = 0.594$), $F(3, 48) = 0.89$, $p = 0.453$, and no significant interaction between factors, $F(3, 48) = 0.62$, $p = 0.61$, $\alpha = 0.01$. Suggesting that $\hat{\theta}$ was more or less the same across variations in ϕ (average $\hat{\theta} = 85.36$, $SD = 9.37$) was consistent with our initial observation in Figure 1.

Azimuth settings

The 2 × 4 (material x rotation axis) two-way ANOVA showed that there was a statistically significant main effect of surface material, $F(1, 48) = 12.67$, $p < 0.001$, and a significant main effect of ϕ, $F(3, 48) = 27.41$, $p < 0.0001$, on $\hat{\phi}$, as well as a significant interaction between factors, $F(3, 48) = 21.33$, $p < 0.0001$, $\alpha = 0.01$. Inspecting the observer means of $\hat{\phi}$ across conditions in Table 5 reveals that the interaction was driven by the material-dependent changes in $\hat{\phi}$. Whereas the perceived azimuth for the *shiny* object remained fairly stable across ϕ manipulations, $\hat{\phi}$ varied with changes in level of ϕ of the *textured* object.

Model predictions

We next compared observers' $\hat{\phi}$ to model predictions. Figure 13 illustrates that observers' estimates of ϕ show good correspondence to ϕs predicted by our model. One-sample t tests reveal that observers'

Material	0	33	63	90
$Shiny_{Observer}$	58.57 $_{SE=7.17}$	62.27 $_{SE=9.75}$	70.3 $_{SE=3.83}$	61.13 $_{SE=6.42}$
$Shiny_{model}$	62.49	65.4	55.1	66.02
$Texture_{Observer}$	10.39 $_{SE=4.68}$	27.43 $_{SE=5.99}$	68.19 $_{SE=0.99}$	91.69 $_{SE=2.89}$
$Texture_{model}$	3.97	34.13	59.2*	85

Table 5. Observer and model estimates for ϕ. *Notes:* Observers' estimates of $\hat{\phi}$ show good correspondence to s predicted by our model. One-sample t tests reveal that observers' estimates $\hat{\phi}$ are not significantly different from the model predicted ϕ ($\alpha = 0.063$, after Bonferroni correction). *Significant difference between model observer estimates. Figure 13 shows the corresponding angular plots.

estimates $\hat{\phi}$ were not significantly different from the values of ϕ predicted by the model ($\alpha = 0.0063$, after Bonferroni correction). Note that θ estimates of the model were close to veridical.

The ϕ values estimated by the model for the *uniform* and *silhouette* versions are overlaid in Figure 13 as red and orange arrows, respectively. Because we did not measure $\hat{\phi}$ for these materials experimentally, we cannot statistically test the prediction quality; however, we believe that there is quite a good agreement with perception, and we invite the reader to check this for himself or herself by inspecting the corresponding stimuli at http://bilkent.edu.tr/~katja/orientation.html.

Discussion

As predicted, we found that the motion of specular features across an object can also systematically bias the perceived rotation axis azimuth, whereas the perceived azimuth for *textured* objects was closer to veridical. The emergence of this effect highlights that the axis of object symmetry and the axis of rotation of conjointly affect the visible geometry and object motion and thus specular flow. The specific nature of this interaction should be examined in future work.

Because a crucial implicit computational step in SfM is the estimation of the object rotation, and given the known differences of matte, textured, and specular optic flow (Hartung & Kersten, 2002; Wendt, Faul, Ekroll, & Mausfeld, 2010; Zang, Doerschner, & Schrater, 2009), we asked whether specular SfM would be different from matte, textured SfM. In particular, we explored how *shiny* and *matte* objects differ in perceived rotation axis. This was in part fueled by the observation that the object in the seminal demonstration by Hartung and Kersten (2002) appeared not to change not only its surface material but also its axis of rotation as it transitioned from shiny to matte, even though neither the true axis nor its perceived 3D shape actually changed (Figure 1).

We postulated that structure from motion mechanisms may in part account for the rotation axis estimation errors in the *shiny* condition in Experiments 2 and 3, and we tested whether a model based on SfM could predict observers' percepts of rotation axis orientation for a single rotating object of noncomplex geometry with specular reflectance or matte texture. We find good agreement between the model's prediction of rotation axis tilt and observers' percepts and find that both the model's and observers' $\hat{\phi}$ estimates varied as a function of surface reflectance of the object. To the best of our knowledge, this is the first time that a close link between SfM models and human perception has been explicitly demonstrated.

General discussion

Summary

Three-dimensional object structure can be inferred even when image motion is the only available cue. However, in the real world, there are several distinct sources that conjointly contribute to optic flow, including object shape, motion trajectory, and surface reflectance. Therefore, recovering these parameters is a mathematically underconstrained problem. Despite this, the visual system simultaneously estimates shape, object motion, and surface material with ease—although the solution that it comes up with may not necessarily reflect the physical reality. To systematically relate percepts to the visual input and ground truth has the potential of revealing the mechanisms by which the visual system solves this problem.

Here, we explored in three experiments how object shape, motion trajectory, and surface reflectance jointly affect the estimation of 3D structure from motion. We measured observers' angular errors as well as rotation axis elevation and azimuth settings in a rotation axis direction estimation task for irregular (Experiment 1) and isotropic objects (Experiments 2 and 3), under four material conditions: *shiny, textured, uniform*, and *silhouette*. In general, we found that adding a reflectance parameter to the boundary motion reduces the estimation error (Experiment 1 and 2); however, this effect varied across material categories (Experiment 2)

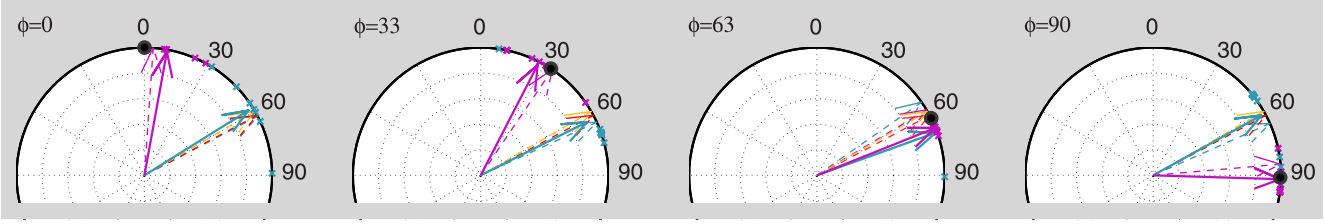

Figure 13. Observer and model estimates of ϕ of the rotation axis. We measured the observer's ϕ estimates for four orientations: 0°, 33°, 63°, and 90°. Black circles indicate ground truth values for ϕ, solid lines indicate mean observer estimates of ϕ, and dashed lines model estimates. The x symbols show individual settings. Teal colors denote *shiny* material, purple *texture*, red *uniform*, and orange *silhouettes*. Two things become apparent: (a) model and observer estimates show close correspondence and (b) both the model and the observers make similar perceptual errors when estimating ϕ for shiny objects. This suggests that the perceived rotation axis orientation depends on global properties of the optic flow generated by the object (i.e., not just contour flow but also surface material–dependent flow characteristics). Table 5 shows the corresponding values for shiny and textured objects. Note that we did not test observers in *uniform* and *silhouette* conditions but invite the reader to inspect the corresponding movies at http://bilkent.edu.tr/~katja/orientation.html.

and was dependent on the 3D shape complexity of the object. In Experiment 2, we found that material category systematically affected perceived rotation axis slant but not tilt. For smaller θ, observers tended to overestimate the rotation axis slant of shiny objects relative to other material categories, and the general underestimation of rotation axis slant was most pronounced for *silhouette* objects. We offered a potential explanation for these patterns. In Experiment 3, we showed that *shiny* and *textured* objects differed in perceived rotation axis tilt and demonstrated that a structure from a motion-based model can account for the observed differences.

Taken together and in line with earlier findings by Norman et al. (2004), Norman and Todd (1994), Pollick et al. (1994), and Todd (1985), our results show that observers can extract 3D structure from a wide range of optical deformations, including those resulting from the motion of occluding boundaries, smooth shading, and specular features. Yet as our data show, each one of these optical deformation biases the rotation axis estimate differently. How these observed biases can be precisely explained by the respective image motion remains the subject of future study.

Previous work

As discussed above, there have been several studies with objectives related to this article, in particular, the work of Norman et al. (2004), Norman and Todd (1994), Pollick et al. (1994), and Todd (1985). Todd (1985) investigated the validity of the projective correspondence assumption by measuring the perceived slant of rotating surfaces that were defined through the motion of dots. He varied the degree and character of the added noise (constrained, unconstrained) as well as the density of the dots defining the surface and found that observers are able to recover slant even for stimuli that have very low correspondence. Thus, he concludes that projective correspondence, which is the fundamental assumption of most SfM algorithms, does not appear to be a necessary requirement for the human visual system to recover structure from motion. This conclusion is further supported by his qualitative experiment on the perceived motion of self-occluding boundaries and smoothly shaded ellipsoids. Although he found that only the yoked motion of silhouettes of two rotating 3D ellipses are perceived as rotating rigidly in 3D (not the motion of individual silhouettes), smoothly shaded ellipsoids were perceived as rigidly rotating in depth when presented in isolation. Note that both of these stimuli violate the correspondence assumption, yet observers were able to extract 3D structure.

Our results are in line with the conclusions by Todd (1985). Our observers are able to estimate the rotation axis for stimuli for which projective correspondence is violated (e.g., the *shiny*, *uniform*, and *silhouette* conditions). However, in Experiment 2, we find interesting differences between these material categories: Whereas estimation errors for *shiny* and *silhouette* objects are high, they are rather low (and comparable with the *texture*) for the *uniform* object. One might conclude that adding specular reflectance to the object's boundary motion is about as detrimental to 3D shape perception as removing all shading texture cues from the object. However, as we have discussed above, it is not as simple as this, because these errors originate from systematic biases of the estimated rotation axis direction, and these biases are different for each material category. How optical deformation biases the estimated rotation axis, for example, to precisely account for the systematic over- and underestimation of θ in our experiments, remains to be investigated.

Norman and Todd (1994) further investigated the necessity of boundary complexity (Koenderink & Van Doorn, 1976; Todd, 1985) in the perception of rigid structure. They measured observers' discrimination thresholds for identifying rigid and nonrigid motion and found that thresholds for intersecting ellipsoids—which generated a more complex silhouette—were much lower than those for nonintersecting ellipsoids. This tendency for singularities in moving silhouettes (due to the underlying complex 3D structure) to facilitate rotation axis judgments might in part explain the differences in results between our Experiments 1 and 2. The complex object contours and the 360° rotation of the objects in Experiment 1 might have provided rich information on which observers relied when estimating the rotation axis direction, including those conditions when the object shape was tainted with specular reflectance. In contrast, if the contour is smooth, convex, and simple, it provides little information about the 3D structure of the object. Under these conditions, the reflectance-specific motion cues combine with or dominate the boundary motion, thus giving rise to the observed material effects in Experiment 2.

Pollick et al. (1994) measured rotation axis estimation across a variety of azimuths and tilts by asking observers to align their index finger with the perceived rotation axis of SfM (dot) displays that varied in frame number and silhouettes of rotating ellipsoids. They compared slant and tilt estimates as well as angular error for naive and experienced subjects. In the SfM condition, naive participants had trouble estimating tilts correctly for slant values smaller than 72°. Interestingly, although all but one subject in our study were naive, we did not find such a tendency. In fact, tilt estimations seemed to have been unaffected by manipulations of material and slant (θ). This difference might be accounted for by the differences in design. Our probe was in the immediate vicinity of the object and might have thus provided additional cues. Although our probe did not intersect with the object in 3D (Experiment 2), the projected images of object and probe did (depending on the direction of the probe; see Figure 5), which might have provided additional information for doing the task. One can examine these factors by designing a probe that is located outside the object and investigate whether this affects the results systematically. The difference in probe design might also account for the different results (Pollick et al., 1994) obtained for silhouettes: In contrast to our study, their naive observers overestimated slant, whereas in our Experiment 2, participants tended to underestimate it. However, these different findings might also be explained by the 3D object that was used; although simple, our stimulus was slightly more complex than theirs, potentially providing more cues to 3D structure. Taken together, these perceptual biases that Pollick et al. (1994), we, and others found strongly suggest that the representation of 3D structure in the human visual is not veridical in the sense of corresponding to Euclidean space (Pollick et al., 1994).

Although not studying the perceived rotation axis directly, Norman et al. (2004) have shown that observers are able to discriminate 3D shapes under a range of surface reflectances (highlights, shading, occlusion, texture, individually, and combined) and motion conditions (single frame, stereo, and multiple frame). Although we cannot compare results directly, as our task was not 3D shape recognition, it might be worth pointing out that in contrast to our study, Norman et al. (2004) found consistently low discrimination thresholds with the addition of specular highlights, whereas we found that specularities produced the largest errors in Experiment 2, which might, again, be attributable to differences in shape complexity between studies. We do not believe that the addition of stereo might have altered the results in our experiments significantly, as motion already provides multiframe information (also, Norman et al. [2004] found no difference between motion and stereo conditions), and we have observed that stereo does not change reflectance-based motion illusions (Doerschner, Fleming et al., 2011). Stereo might have slightly improved the impression of shininess, however (Wendt et al., 2010).

Perceptual quality, heuristics, and optic flow

Glossy objects (shading and highlights) and marble (texture, shading, and highlights) represent interesting mixed categories of those used in our experiments; thus, one might wonder how the information from these two categories combined affect rotation axis estimates. In order to do so, one has to have an approximate idea of how the optic flow is used by the visual system to extract 3D motion. Several strategies and heuristics have been proposed, and before we add another one to these, we invite the reader to inspect the movies, showing the mixed categories as well as their constituents at http://bilkent.edu.tr/~katja/orientation.html.

Perception is often based more on qualitative aspects of 3D structure, such as ordinal or topological relations, than on strict Eucledian metric (Todd & Norman, 2003). Koenderink and Van Doorn (1976) suggested how these qualitative aspects could be used to recover 3D structure from the motion of occluding boundaries. Perhaps it is qualities such as these that observers use to perform our tasks. For example, a source of information could be the relation of boundary to within-boundary motion. It is the within-boundary motion that defines the object's surface

reflectance (Doerschner, Kersten et al., 2011) and possibly also influences its rigidity. For example, for matte, textured surfaces, globally unidirectional optic flow across the object (but within boundaries) suggests a large slant angle, and the tilt angle would be roughly perpendicular to the within-boundary average flow vector (in analogy to in-depth rotating dots). As the slant angle decreases, more and more points would trace elliptical paths changing the within-boundary global optic flow field and thus perceived slant and tilt.

Earlier work by Caudek and Domini (1998) suggests that human perception of SfM, including the axis of rotation, may involve only an analysis of first-order optic flow properties (see their equation 1). Specifically, they showed that for motion dot displays, the slant of the rotation axis was well predicted by the ratio of the global *def* and the component parallel to the image plane of the global velocity vector ρ (Caudek & Domini, 1998). Although the perceived rotation axis slant could be manipulated by manipulating this ratio, they found that the perceived rotation axis tilt was always in agreement with ground truth. In our experiments, we find that not just slant but also tilt can be misestimated by observes. Our model accounts for the perceived tilt of richly textured shiny and matte objects, and it does so without relying on the assumption of affine image transformations or homogenous optic flow (Caudek & Domini, 1998). However, a limitation of both models is their reliance on projected correspondence; thus, they have difficulty in accounting for the perceived rotation axis of smooth shaded objects and silhouettes.

Rigidity and 3D shape

In Experiment 1, we observed that for many stimuli, the visual system preferred a rigid interpretation at the expense of a perceived stable rotation axis. Observers experienced a piecewise stable rotation axis that changed its orientation smoothly, several times during the actual 360° rotation of the object, rather than a nonrigidly deforming shape and a constant rotation axis. Although a rigidity prior has been debated (Jain & Zaidi, 2011), our observation suggests that under some conditions, it may exist. Boundaries and prior knowledge have been shown to influence perceived rigidity (Sinha & Poggio, 1996; Ullman, 1984). When the impression of a coherent, bounded object is degraded (e.g., by masking object contours), the rigidity assumption might also play a weaker role, and the visual interpretation might be dominated by other priors (also see Domini et al., 1997; Norman & Todd, 1994). Even for unfamiliar, irregularly-shaped objects, contours are very powerful cues to rigidity (Koenderink & Van Doorn, 1976).

We did not explicitly test the accuracy of 3D shape reconstructions generated by the model, as this was not our primary interest; however, informal inspection of *shiny* stimuli in Experiment 2, as well as previous results by Doerschner et al. (2011) and Swaminathan et al. (2002) concerning the difficulty of recovery of epipolar geometry from moving specular objects[8] suggest that specular structure from motion may be more problematic if not impossible under certain conditions. The nature of these conditions as well as the neural basis of these computations is the subject of future study.

Outlook

We investigated how object shape, motion trajectory, and surface reflectance jointly affect the estimation of 3D structure from motion (SfM). It may be that in order to estimate the rotation axes of objects, the visual system extracts some kind of measure of 3D shape coherence, such as the one proposed by Doerschner et al. (2011), and the strength of this index—signaling the degree of "objectness" or "rigidity"—determines how first-order optic flow properties are factored into rotation axis estimation. This idea could be explored by systematically degrading 3D shape cues (e.g., by removing the object contour) and measuring how this alters rotation axis tilt and slant.

Keywords: rotation axis, optic flow, surface reflectance, material qualities, object recognition, structure from motion

Acknowledgments

This work was supported by a Marie Curie International Reintegration Grant (239494) within the 7th European Community Framework Programme awarded to K. D. R. F. was funded as part of the research program of the Bernstein Center for Computational Neuroscience, Tubingen, Germany, funded by the German Federal Ministry of Education and Research (BMBF; FKZ: 01GQ1002). Supported by the EU Marie Curie Initial Training Network "PRISM" (FP7-PEOPLE-2012-ITN, Grant Agreement: 316746) to R. W. F. and K. D. K. D. was also supported by a grant of the Scientific and Technological Research Council of Turkey (TUBITAK 1001 grant 112K069), and by a fellowship by the Turkish Academy of Sciences. We thank the reviewers for their helpful comments.

Commercial relationships: none.
Corresponding author: Katja Doerschner.
Email: katja@bilkent.edu.tr.

Address: Department of Psychology, Bilkent University, Ankara, Turkey.

Footnotes

[1]There is some theoretical work on shape from specular flow in the computer vision literature (Adato, Vasilyev, Ben Shahar, & Zickler, 2007); however, this assumes known camera motion.

[2]We are currently exploring possibilities (a) and (b) in ongoing work.

[3]Object rotation reveals 3D structure much more effectively and comprehensively than translational or looming trajectories.

[4]Most structure from motion approaches necessitate that a visual system needs to establish the correspondence of a point in successive frames to compute 3D structure.

[5]Since we did not need to estimate the direction of rotation in Experiment 2 (the object rotated back and forth), the largest possible angular error ε between any two vectors could be 90°. Specifically, we defined the angular error ε to be the smaller one of the two angles at the intersection point. Consequently, chance performance would correspond to $\varepsilon = 45°$.

[6]This is consistent with cases of failure of epipolar geometry to predict surface material; for example, there were no significant deviations of epipolar geometry measured for a shiny cube-like object in Doerschner, Fleming et al. (2011).

[7]Because of a small miscalculation in the rendering, θ_0 and θ_{90} were 90°, whereas θ_{30} and θ_{60} were 100°. This difference was perceptually not detectable and did not cause any systematic differences in the perception of theta across these conditions. See the "Results" section.

[8]Because of the appearance distortion in specular flow, the corresponding (across frames) specular features often violate epipolar geometry, which renders SfM problematic (Doerschner, Fleming et al., 2011; Swaminathan et al., 2002).

[9]These models are highly successful and often more accurate than human observers because of the available extensive memory and processing resources.

References

Adato, Y., Vasilyev, Y., Ben Shahar, O., & Zickler, T. (2007). Toward a theory of shape from specular flow. In *Proceedings of the IEEE International Conference on Computer Vision,* pp. 1–8.

Adelson, E., & Movshon, J. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*, 523–525.

Andersen, R. (1997). Neural mechanisms of visual review motion perception in primates. *Neuron, 18,* 865–872.

Andersen, R., Bradley, D., & Shenoy, K. (1996). Neural mechanisms for heading and structure from motion perception. *Cold Spring Harbor Symposia on Quantitative Biology, 61,* 15–25.

Andersen, R., & Siegel, R. (2005). Three-dimensional structure-from-motion selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey. *Cerebral Cortex, 15,* 1299–1307.

Balas, B., & Sinha, P. (2008). Observing object motion induces increased generalization and sensitivity. *Perception, 37*, 1160–1174.

Beardsley, P., Zisserman, A., & Murray, D. (1997). Sequential updating of projective and affine structure from motion. *International Journal of Computer Vision, 23*, 235–259.

Beis, J., & Lowe, D. (1997). Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference*, pp. 1000–1006.

Bradley, D., Chang, G., & Andersen, R. (1998). Encoding of three-dimensional structure-from-motion by primate area mt neurons. *Nature, 392*, 714–716.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Caudek, C., & Domini, F. (1998). Perceived orientation of axis rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 609–621.

Debevec, P. (2002). Image-based lighting. *IEEE Computer Graphics and Applications,* 26–34.

Doerschner, K., Fleming, R., Yilmaz, O., Schrater, P., Hartung, B., & Kersten, D. (2011). Visual motion and the perception of surface material. *Current Biology, 21*, 2010–2016.

Doerschner, K., Kersten, D., & Schrater, P. (2011). Rapid classification of specular and diffuse reflection from image velocities. *Pattern Recognition, 44*, 1874–1884.

Domini, F., Caudek, C., & Proffitt, D. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1111–1129.

Fischler, M., & Bolles, R. (1981). Random sample consensus: a paradigm for model fitting with

applications to image analysis and automated cartography. *Communications of the ACM, 24*, 381–395.

Giblin, P., Pollick, F., & Rycroft, J. (1994). Recovery of an unknown axis of rotation from the profiles of a rotating surface. *Journal of the Optical Society of America A, 11*, 1976–1984.

Hartley, R., Gupta, R., & Chang, T. (1992). Stereo from uncalibrated cameras. In *Computer Vision and Pattern Recognition, 1992. Proceedings cvpr'92, 1992 IEEE Computer Society Conference*, pp. 761–764.

Hartung, B., & Kersten, D. (2002). Distinguishing shiny from matte. *Journal of Vision, 2*(7):551, http://journalofvision.org/2/7/551, doi:10.1167/2.7.551. [Abstract]

Huang, T., & Netravali, A. (1994). Motion and structure from feature correspondences: A review. *Proceedings of the IEEE, 82*, 252–268.

Jain, A., & Zaidi, Q. (2011). Discerning nonrigid 3D shapes from motion cues. *Proceedings of the National Academy of Sciences USA, 108*, 1663–1668.

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

Koenderink, J., & Van Doorn, A. (1991). Affine structure from motion. *Journal of the Optical Society of America A, 8*, 377–385.

Koenderink, J., & Van Doorn, A. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics, 32*, 211–216.

Koenderink, J., & Van Doorn, A. (1992). Second-order optic flow. *Journal of the Optical Society of America A, 9*, 530–538.

Koenderink, J., & Van Doorn, A. (1982). The shape of smooth objects and the way contours end. *Perception, 11*, 129–137.

Koenderink, J., & Van Doorn, A. (1976). The singularities of the visual mapping. *Biological Cybernetics, 24*, 51–59.

Koenderink, J., & Van Doorn, A. (1980). Photometric invariants related to solid shape. *Optica Acta, 27*, 981–996.

Landy, M. (1987). Parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A, 4*, 864–877.

Landy, M., Dosher, B., Sperling, G., & Perkins, M. (1991). The kinetic depth effect and optic flowii. first-and second-order motion. *Vision Research, 31*, 859–876.

Longuet-Higgins, H., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 208,* 385–397.

Lourakis, M., & Argyros, A. (2004). The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm. *Institute of Computer Science-FORTH, Heraklion, Crete, Greece, Tech. Rep, 340*.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*, 91–110.

Mulholland, T. (1956). Motion perceived while viewing rotating stimulus-objects. *American Journal of Psychology, 69*, 96–99.

Nadler, J., Angelaki, D., & DeAngelis, G. (2008). A neural representation of depth from motion parallax in macaque visual cortex. *Nature, 452*, 642–645.

Norman, J., & Todd, J. (1994). Perception of rigid motion in depth from the optical deformations of shadows and occlusion boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 343–356.

Norman, J., Todd, J., & Orban, G. (2004). Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science, 15*, 565–570.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437–442.

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

Sampson, P. (1982). Fitting conic sections to. *Computer Graphics and Image Processing, 18*, 97–108.

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature, 384*, 460–463.

Stoner, G., & Albright, T. (1992). Neural correlates of perceptual motion coherence. *Nature, 358*, 412–414.

Swaminathan, R., Kang, S., Szeliski, R., Criminisi, A., & Nayar, S. (2002). On the motion and appearance of specularities in image sequences. *Lecture Notes in Computer Science, 2350*, 508–523.

Todd, J. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance, 11*, 689–710.

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

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 203*, 405–426.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-d structure from rigid and rubbery motion. *Perception, 13,* 255–274.

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J., et al. (2002). Extracting 3D from motion: Differences in human and monkey intraparietal cortex. *Science, 298*, 413–415.

Van Essen, D., & Gallant, J. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*, 1–10.

Vasilyev, Y., Adato, Y., Zickler, T., & Ben-Shahar, O. (2008). Dense specular shape from multiple specular flows. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. cvpr,* 2008, pp. 1–8.

Vuong, Q., & Tarr, M. (2004). Rotation direction affects object recognition. *Vision Research, 44*, 1717–1730.

Wallach, H., & O'Connell, D. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205–217.

Watanabe, T., & Cole, R. (1995). Propagation of local motion correspondence. *Vision Research, 35*, 2853–2861.

Weiss, Y., & Adelson, E. (2000). Adventures with gelatinous ellipses: Constraints on models of human motion analysis. *Perception, 29,* 543–566.

Wendt, G., Faul, F., Ekroll, V., & Mausfeld, R. (2010). Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision, 10*(9): 7, 1–17, http://www.journalofvision.org/content/10/9/7, doi:10.1167/10.9.7. [PubMed] [Article]

Zang, D., Doerschner, K., & Schrater, P. (2009). Rapid inference of object rigidity and reflectance using optic flow. In *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns,* p. 888.

Appendix

Motivation

Object versus ego motion

There are two sources for 3D structure to arise from motion: One is motion (rotation) of the object, and the other is motion of the visual sensor (eye or camera). Although the former scenario is more commonly used to study the kinetic depth effect (Wallach & O'Connell, 1953) in humans, the latter is the prevalent scenario to estimate structure from motion in computer vision. For textured, nonshaded objects, these two phenomena are obviously related (as are stereo vision and motion parallax) as in both, image features need to be tracked over time (or frames), and from these features, the 3D structure is computed. In fact, in cases where there is a single object and a textureless background (as in our experiments), the two are computationally identical (note, however, that object and camera would be clearly distinct for smoothly shaded objects). Thus, we base our model on a classic structure from the motion paradigm in which both the 3D structure of the scene and the motion of the camera (egomotion) are estimated simultaneously. We can easily compute *object motion* under the assumption of a stationary camera. We propose below that measuring the change in rotation angle of the camera as a function of time is equivalent to estimating the object's orientation axis from object rotation. Note that our model does not rely on any special assumptions about projection type, object shape, surface reflectance, or motion trajectories.

SfM and ego motion in the human visual system

Because it is the basis of our model, the question arises as to what extent ego motion plays a role in human SfM estimation. Although the neural substrate of structure from motion has been studied (e.g., Andersen & Siegel, 2005; Bradley et al., 1998; Vanduffel et al., 2002), the neural mechanisms of 3D structure and ego-motion estimation from image motion are still elusive (Andersen, Bradley, & Shenoy, 1996; Nadler, Angelaki, & DeAngelis, 2008). Algorithmic approaches in computer vision exploit both accurate localization and tracking of feature points and nonlinear minimization of a high-dimensional error function. These two issues are closely related to correspondence assignment and hierarchical information processing in the motion pathway of the visual cortex (Adelson & Movshon, 1982; Andersen, 1997; Stoner & Albright, 1992; Van Essen & Gallant, 1994; Watanabe & Cole, 1995).

We suggest that although the architecture and design principles of information processing differ between computer vision models and the visual cortex, structure-from-motion solutions in computer vision may be useful for understanding ego-motion–based SfM in general, because this class of models uses all of the motion information available in an image sequence (as do human observers), including object boundary- and surface material–based image motion.[9] Our results

support this argument in which we find that the structure-from-(ego) motion–based model makes the same perceptual errors as the observers in our experiments.

To go from ego-motion SfM to rotation axis orientation is a sequential process in our model. However, if and how these computations could be done by the visual system are not known. Global integration may involve cortical areas that have been associated with the computation of structure from motion (e.g., V5+, hMT; Bradley et al., 1998; Vanduffel et al., 2002).

Model

Step 1: Feature tracking

The goal in this step is to identify the fundamental matrix F such that $xFx' = 0$, where $x = [x_0 y_0 1].'$ We first find putative matches (Beis & Lowe, 1997) corresponding to points between frames based on similarity in appearance (SIFT features; Lowe, 2004). To do so, at least three frames are required. To eliminates features that are inconsistent with the fundamental matrix estimate, we perform epipolar filtering using a random sample consensus model (RANSAC; Fischler & Bolles, 1981) to find F with an eight-point direct linear transform (DLT) model (Hartley, Gupta, & Chang, 1992). We then used Sampson error (Sampson, 1982) to label inconsistent features as outliers. Consistent features were tracked from one frame to the next until matching becomes impossible.

Step 2: Stereo reconstruction

We used a standard incremental reconstruction (Beardsley, Zisserman, & Murray, 1997). To obtain the 3D coordinates of a few features, we perform initial stereo reconstruction for only two frames. These 3D points serve as an initial set for subsequent steps.

Step 3: Pose estimation

An eight-point DLT model (Hartley et al., 1992) with RANSAC (Fischler & Bolles, 1981) elimination was used to estimate the rotation and translation (pose) of the camera with respect to a predefined coordinate system. The pose estimation model requires 3D points and their 2D image correspondents. The 2D points are the extracted feature points. The 3D points are the reconstructed points in Steps 2 and 4. Given the known 2D position of the features, one can estimate the camera pose of a given frame. Note that in order to do so, one needs at least three features with valid 3D coordinates.

Step 4: 3D reconstruction

Once the camera pose is known, new 3D points can be reconstructed using triangulation and sparse bundle adjustment (Lourakis & Argyros, 2004). Pose estimation and 3D reconstruction steps are executed iteratively.

Step 5: Rotation axis estimation

We apply the Rodrigues formula to extract the rotation vector for each camera position. We then compute the *change in rotation angle* for all frames. To obtain the final rotation axis, we take the arctan of the estimated rotation angles in elevation and azimuth. Figure 12 shows the estimated rotation axis for $\phi = 0°$.