Shape Perception of Thin Transparent Objects with Stereoscopic Viewing

JIANHUI CHEN and ROBERT S. ALLISON, York University

Many materials, including water surfaces, jewels, and glassware exhibit transparent refractions. The human visual system can somehow recover 3D shape from refracted images. While previous research has elucidated various visual cues that can facilitate visual perception of transparent objects, most of them focused on monocular material perception. The question of shape perception of transparent objects is much more complex and few studies have been undertaken, particular in terms of binocular vision.

In this article, we first design a system for stereoscopic surface orientation estimation with photo-realistic stimuli. It displays pre-rendered stereoscopic images and a real-time S3D (Stereoscopic 3D) shape probe simultaneously. Then we estimate people's perception of the shape of thin transparent objects using a gauge figure task. Our results suggest that people can consistently perceive the surface orientation of thin transparent objects, and stereoscopic viewing improves the precision of estimates. To explain the results, we present an edge-aware orientation map based on image gradients and structure tensors to illustrate the orientation information in images. We also decomposed the normal direction of the surface into azimuth angle and slant angle to explain why additional depth information can improve the accuracy of perceived normal direction.

Categories and Subject Descriptors: I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Perceptual reasoning*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Stereo*

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: 3D shape perception, transparent, shape-from-texture, stereoscopic, material perception

ACM Reference Format:

Chen, J. and Allison, R. S. 2013. Shape perception of thin transparent objects with stereoscopic viewing. ACM Trans. Appl. Percept. 10, 3, Article 15 (August 2013), 15 pages.
DOI: http://dx.doi.org/10.1145/2506206.2506208

1. INTRODUCTION AND PREVIOUS WORK

Perceiving the orientation of visible surfaces in the environment is thought to be a key step in spatial vision, object recognition and visually guided action [Barrow and Tenenbaum 1981; Marr 1982]. Since light passes through transparent objects, one not only sees the surfaces of the transparent object, but also the scene behind the object. With these complex images, how can the brain obtain the shape of an object? For opaque surfaces an important cue to surface orientation is binocular disparity. The stereoscopic views provide additional information about binocular depth based on the disparities between the

This work is supported by NSERC (Canada) Grant #RGPIN/239072-2010.
Authors' address: J. Chen, R. S. Allison, 4700 Keele St, Toronto, ON, Canada, M3J1P3, Centre for Vision Research, Department of Computer Science and Engineering, York University; email: {jianhui, allison}@cse.yorku.ca.

DOI: http://dx.doi.org/10.1145/2506206.2506208

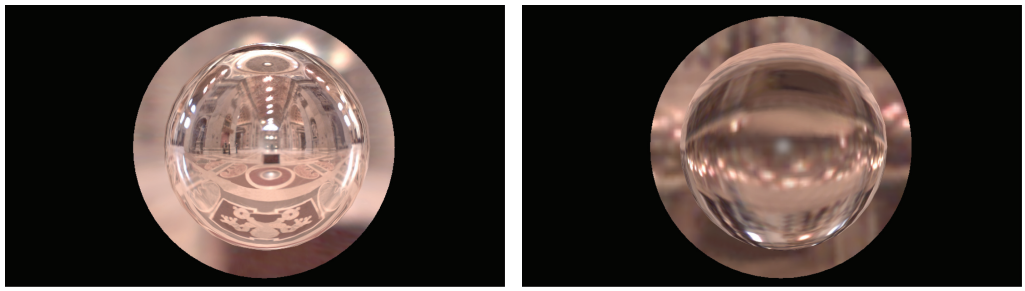
images in the left and right eyes. Surface orientation is encoded by gradients of disparity and surface curvature by higher-order disparity transforms. Even with stereoscopic depth, the multiple layers of transparent objects pose a formidable problem for vision. How does the visual system solve a problem that seems computationally intractable? We attempt to address this by studying transparent surface perception.

Early researchers [Gibson 1950] argued for a direct correspondence between perception and some measurable property of the optical stimulation. However, the mapping from the world into the image is many-to-one, and hence there is no unique solution to the problem; however it is often true that 3D interpretations of given image are highly constrained. One possibility is that human visual system exploits regularities of the natural environment to select an interpretation that is statistically most likely. For instance, the light source position is assumed to be in front of and above the object [Sun and Perona 1998], the world is stable and the pattern of optic flow indicates structure and self motion [Gibson 1950], and the pattern of texture on a physical surface has some form of underlying regularity [Todd et al. 2004].

When light from a scene is reflected by highly reflective objects and then reaches people's eyes, people perceive the luminance/colors as arising from the scene instead of as surface markings on the mirrored object. A similar situation happens when light is refracted by transparent objects and reaches people's eyes. In both cases, the image depends on the pattern of the light in the environment as well as the shape of the surface. Thus, study of the influence of the environment map is very important for the understanding of shape perception of reflective/refractive objects. Fleming et al. [2003] studied the statistics of real world illumination that facilitates the perception of a reflective surface. They designed a surface reflectance matching task to measure the accuracy of human surface reflectance estimation. They demonstrated that subjects could match surface reflectance properties reliably and accurately in the absence of context, as long as the illumination was realistic. Their work also suggested that subjects use stored assumptions about the statistics of real-world illumination when estimating surface reflectance. Illumination maps derived from very different scenes in the world share certain statistical regularities in structure. The authors argued that these regularities allow the visual system to make certain assumption in interpreting images of surfaces. This analysis and the supporting experiments indicated that real-world illumination is necessary for the study of perception of reflective objects.

Due to the similarity between mirror reflectivity and transparent refraction, the most relevant work to ours involves shape perception of reflective objects [Fleming et al. 2004]. Fleming et al. argued that, specular reflections can provide reliable and accurate constraints on 3D shape based on an experiment that had subjects match a gauge figure to the perceived surface normals of reflective objects with photo-realistic rendering. Subjects were generally good at estimating the shapes of perfectly mirrored surfaces, even though the stimuli were presented without any context to specify the scene surrounding the object. They concluded that the visual system can use the "orientation fields" across the image to recover 3D shape. Although they obtained the perceptual error of tilt and slant angles, they do not analyze the error quantitatively.

However, there are two obvious differences between the specular and transparent cases. First, if we use the tangential plane of the local surface as a reference plane, the directions of the reflected/refracted ray are on the opposite side. Second, the refractive angle depends on the IOR (index of refraction) of the material as well as the incident angle, while the reflective angle only depends on the incident angle. These differences make the perception of the shape of transparent objects much more difficult than the mirror reflective objects, even for the case of a single refraction/reflection. Moreover, there are potential ambiguities between reflection and refraction when we mask the environment. In other words, subjects may not be able to discern which object is reflective and which is refractive given two images such as in Figure 1. This is because of the similarity of light paths between reflective



(a) (b)

Fig. 1. A reflective object and refractive object rendered in an indoor scene. The scene was masked so that object appears in isolation.

objects and refractive objects. From the perspective of geometry of light, refraction can flip the environment map left to right and upside down, while reflection maintains the relative order of left/right and up/down. So people are able to "fake" a refractive image of a reflective object, and vice versa, by modifying the environment map. Visibility of the scene helps to resolve the visual ambiguity between reflection and refraction.

Since the light passes multiple surfaces through transparent objects, the visual system must recover the properties of two or more surfaces simultaneously to see veridically. Kersten et al. [2006] presented an experiment using stereoscopic volume rendering of purely absorptive media to study enhancement of depth perception with stereoscopic viewing. The subjects were required to determine the direction of rotation of a cylindrically orientated volume. The results strongly suggested that relative depth can be perceived using stereoscopic viewing. Although their research considered only purely absorptive media (i.e. no refraction), it demonstrated the advantages of stereoscopic viewing in shape perception of transparent objects.

Fleming et al. [2011] highlighted the utility of the distortion field produced by thick transparent objects to the perception of the material properties of transparent objects. They used matching experiments to study the influence of IOR, distance to the rear wall, and thickness on material perception. They concluded that the distortion field is a midlevel cue that involves comparing the relative scale of texture elements seen through the transparent object with the elements seen directly.

This study focuses on the problem of extracting surface orientation from visual information, especially from stereoscopic views. This article makes the following contributions.

(1) We present a photo-realistic image-generation system for experiments on surface orientation estimation. It combines stereo image display and real-time rendering of a stereoscopic 3D shape probe.

(2) We demonstrate that people can consistently perceive the shape of thin transparent objects, and that stereoscopic viewing can improve accuracy.

(3) We present an edge-aware orientation map to illustrate the orientation information specifying shape in images, and analyze the geometrical relationship between local dominant orientation and surface normal direction.

2. STEREOSCOPIC SHAPE ESTIMATION SYSTEM

The system displayed an S3D gauge that subjects could adjust to indicate the surface normal direction at specified locations in photo-realistic images.

Fig. 2. Apparatus and experimental setup with subject. The subject viewed the image through shutter glasses and adjusted an overlaid gauge figure to match the surface normal direction by moving the mouse. The experiment was conducted in a dark room to control extraneous features and stray light.

2.1 Hardware

The system was implemented and tested on a desktop PC (Intel duo core 2.4GHz, 2GB memory and Nvidia GeForce GT 240). Figure 2 shows the experimental setup. The display was a 23.6 inch LCD monitor (Acer GD235HZ) with ability to present 1920 × 1080 resolution images at 120 Hz. For stereoscopic viewing, we used the Nvidia 3D vision kit. It consisted of an IR (infrared) emitter and wireless glasses. The wireless glasses consisted of a pair of liquid crystal shutters that can block vision selectively in the left or right eye to synchronize with the presentation of the left and right image in the display. These images were alternated at 120 Hz to provide independent 60 Hz video to each eye.

2.2 Software

2.2.1 *Object Modeling and Stimulus Rendering.* We modeled hatlike randomly shaped thin objects based on an algorithm developed by Norman and Todd [1996]. Figure 3 shows four objects used in the experiment. We used randomly generated objects since the observers could use prior knowledge to judge the surface shape of familiar objects. From this perspective, these objects have been widely used in shape discrimination [Lee and Saunders 2011] and shape perception [Fleming et al. 2004].

There are a number of reasons for choosing thin objects. First, thin objects represent many important refractive phenomena in the natural world such as water surfaces and thin lenses. Second, in the thick objects, most of the light is refracted more than once and the environment is severely distorted so that subjects may totally lose the visual shape cues. On the other hand, thin objects simplify the light path and make shape judgement possible.

The stimuli were rendered by a ray tracing engine in Blender [Flavell 2010]. The material of the object was perfectly transparent without diffuse and mirror specular properties. The IOR was set to 1.4. We adopted a light probe [Debevec 1998] to simulate the environment. It was a 1500 × 1500 HDR (high dynamic range) image with dynamic range of 200000:1. It allowed us to render perfectly transparent surfaces that yield highly realistic images. The ray tracing depth was set to 4 so that the light will be refracted at most 4 times. Since the object has a planar back surface, most of the light was refracted only once. The over sampling number was set to 16 to minimize the rendering artifacts. The images were tone mapped for the stereoscopic display.

We used a stereoscopic camera model [Schneider 2012] to generate the left and right eye views. We set the camera separation to 0.1 Blender units and zero-parallax at 6 Blender units from the

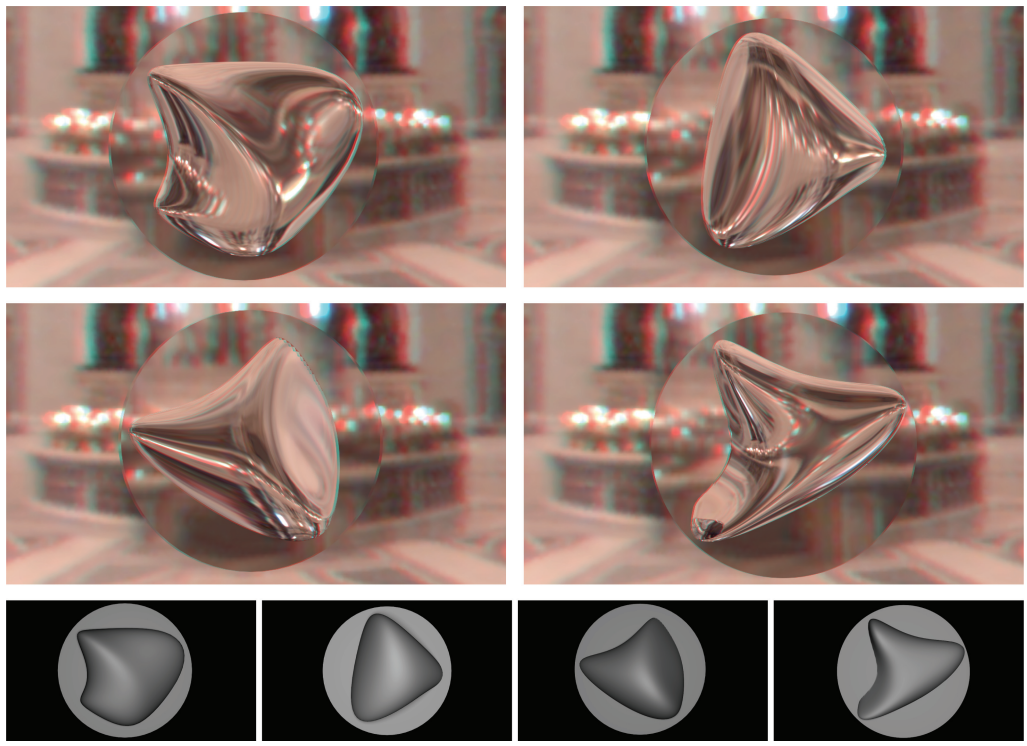


Fig. 3. The four objects used in the experiment (here with red-cyan anaglyphs). In the third row, the objects were rendered as Lambertian to illustrate their shapes.

viewpoint. The max parallax angle was set to 1°. The distance between viewer and display screen was set to 50.8 cm and PPI (Pixel Per Inch) was set to 96. The transparent object was placed behind the zero-parallax plane so that it was located behind the screen plane in the experiment. These settings located the objects inside of the comfortable zone to eliminate stereoscopic discomfort and artifacts.

Besides the stereo images, we also stored the focal length f_c, camera sensor width s_w, and two additional images to recover the 3D position and normal vector for each pixel. The first image was a depth map that stored the depth value for each pixel. The second image was a normal map that stored the normal vector for each pixel. Both the depth value and normal value were specified in camera coordinates.

2.2.2 *Stereo Image Display and User Interface.* The display system utilized DirectX 3D to render stereoscopic images and a shape probe (gauge). The gauge was used to indicate the local normal direction. It had a disk-like base and a stick perpendicular to the disk base. Since the images were rendered offline and the probe was rendered in real time, the probe position needed to be aligned with the surfaces in the pre-rendered images. The key requirement was to attach the center of the probe precisely on the surface, neither in front of nor behind the object surface.

We recovered the surface position p (where we wish to place the center of gauge) from the depth map by ray casting (see Figure 4). For a pixel position (x, y), its corresponding position on the surface is:

$$p = \frac{d}{f_c} \times \left(\frac{s_w}{w} \times (x - w/2), \frac{s_w}{w} \times (h/2 - y), f_c \right), \quad (1)$$

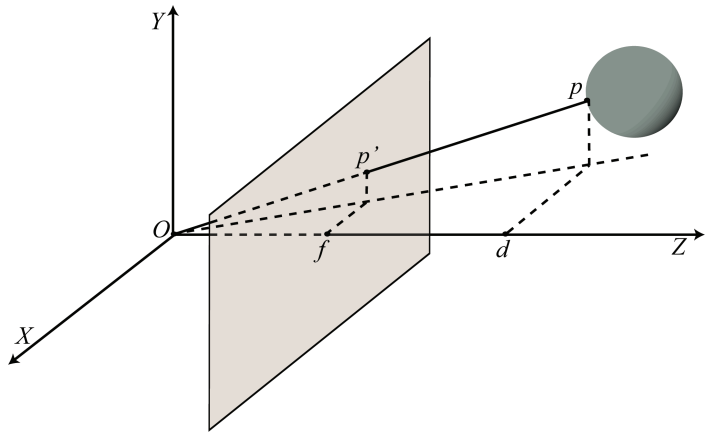


Fig. 4. Surface position recovery.

where w and h are the resolution of the image, s_w is sensor width, f_c is focal length, and d is the depth value of p. Thus, we aligned the prerendered image and real-time rendered gauge in 3D space.

The gauge was first rendered in the monocular frustum using DirectX 3D. When the stereo setting was turned on, the left image and right image of the gauge was automatically generated in the left and right frustum by the Nvidia 3D vision kit. To put the disk center on the surface of the object, we matched the disparities of the disk and the pre-rendered image by adjusting the 3D vision kit parameters. Thus, the system fixed the position of the disk center at the surface so that subjects could only adjust the gauge orientation.

When the subjects moved the mouse, the mouse displacements in the horizontal and vertical directions were mapped to rotation of the gauge about the vertical axis and horizontal axis in 3D space, respectively. With these user interface settings, the subjects could easily adjust the gauge to the desired orientation.

3. PERCEPTION EXPERIMENT

3.1 Stimuli and Conditions

Figure 3 shows the stimuli used in the experiment, generated as described in Section 2.2.1. In the stimuli, each object has a randomly generated surface on which the estimates of surface normal were made. This surface is referred as the object surface. Each object was rendered from three parallel views: left camera, center camera and right camera. Each stimulus object was presented under two viewing conditions: a non-stereoscopic view displayed from the center camera (monocular cues) and a stereoscopic view displayed from the left and right cameras (stereoscopic cues). Therefore, there were 8 object-by-view stimulus combinations (4×2) for each subject. All the images were viewed binocularly through the shutter glasses.

3.2 Subjects

Six male and two female subjects (mean age 26.3, SD 3.9) participated in the experiment. Subjects were students or research fellows at the university. All subjects had normal or corrected-to-normal vision; six wore glasses during the experiment. All of them were naive to the purpose of the study and provided their informed consent. Three of the subjects had previous experience with S3D psychophysical tests, and others had little or no experience. We verified all subjects' ability to see stereopsis by the

Randomdot Stereotest (Stereo Optical Company, Inc. Chicago, IL). All of them had good stereo acuity (<40 seconds of arc).

3.3 Procedure

We used a gauge figure adjustment task [Koenderink et al. 1992] to estimate people's perception of the shape of the objects. Subjects were shown a series of gauges placed at different surface points in random order and asked to adjust each gauge to appear normal to the local surface.

Before the experiment, subjects practiced two tasks with an additional stimulus. First, subjects were required to put the mouse at an arbitrary place on the object surface after which the program would display a gauge with ground truth orientation. The intent was to give subjects understanding of the concept of surface orientation and in how to interpret the direction of the gauge figure. Second, subjects performed the gauge figure task with a reference gauge. When subjects confirmed their estimation, the program calculated the angular difference between the estimated angle and ground truth angle. If the difference was larger than a threshold (15° in practice), the program displayed a green color reference gauge to show the surface orientation. The subjects were required to adjust the gauge until the angular difference was less than the threshold. The practice lasted about 10–20 minutes.

For each object, 80 orientation estimates were sampled evenly crossed the entire object surface and tested in random order. All the orientation estimates for a given object-by-view stimulus combination were made in a block, and then the next combination was measured. We separated the 8 blocks into 2 sessions. Each session was tested on different days to avoid potential eye and/or hand strain. There was no time limit for each trial, but subjects were encouraged to do the test with consistent speed so that they would not spend too much time on a particular trial. It took subjects about 30–45 minutes to finish each session.

4. RESULTS AND DISCUSSION

4.1 How Well Do People Perceive the Shape of Transparent Objects, and Does a Stereoscopic View Improve Accuracy?

We used the 3D angular difference between perceived and ground truth normal directions as the perceptual error. Mean errors for every subject and every object under both viewing conditions were calculated to describe the accuracy of perceived normal direction. We found that the potential improvement assuming a bas-relief compared to Euclidean model [Koenderink et al. 2001; Belhumeur et al. 1999] was very small (about 1°), so we used the original (uncorrected by bas-relief) data in this article.

Figure 5 shows mean errors for each subject and viewing condition. We found that subjects could perceive the normal direction of transparent surfaces consistently: mean error across all trials was 22.7°. Moreover, perceived errors from stereoscopic view were smaller than those from nonstereoscopic view (21.1° versus 24.4°). A repeated-measures ANOVA on estimation error by view condition and subject found a significant main effect of view condition ($F(1, 7.0) = 106.8$, $p < 0.001$).

Figure 6 shows mean errors for each object and viewing condition. A repeated-measures ANOVA showed that the effect of object was significant ($F(2.5, 17.49) = 3.66$, $p = 0.039$). The perceived normal directions from stereoscopic view were more accurate than those from non-stereoscopic view for each object. The interaction between view condition and object was not significant ($F(3, 21) = 2.56$, $p = 0.082$).

We quantified the reliability of the estimation by comparing the estimated direction against the median vector of estimated normal directions [Cole et al. 2009]. We randomly chose 1000 gauge settings and plotted their errors relative to ground truth and the median vector in Figure 7. The figure shows that the errors were clustered towards the center of each plot. Figure 8 shows the cumulative

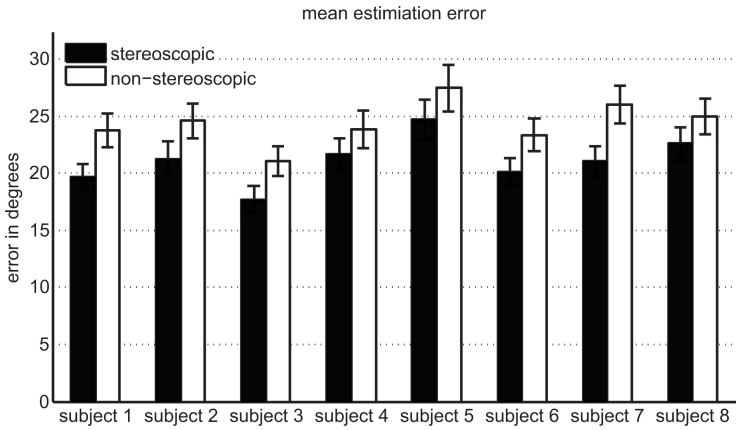


Fig. 5. Mean estimation errors for each subject and viewing condition (the error bars show confidence intervals (95%)).

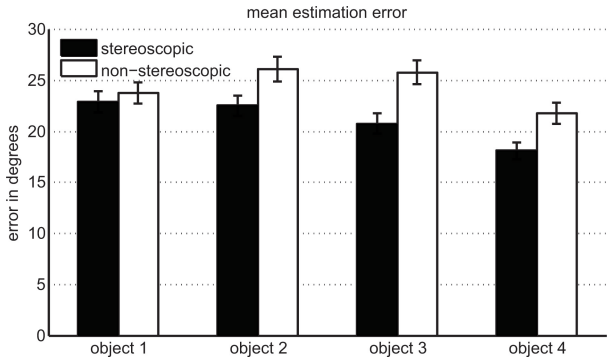


Fig. 6. Mean estimation errors for each object and viewing condition (the error bars show confidence intervals (95%)).

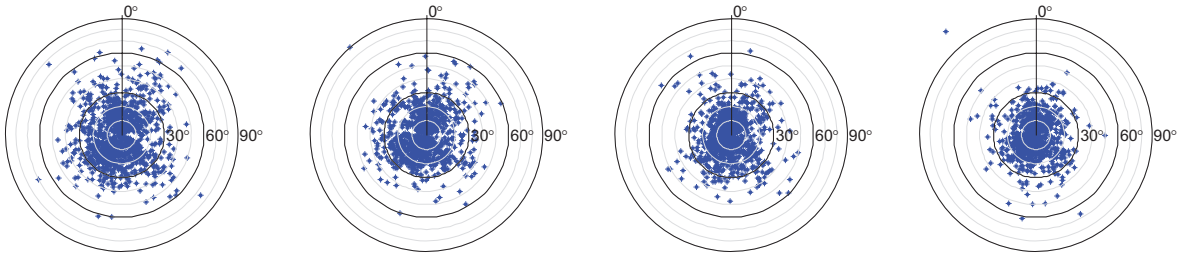


Fig. 7. Scattering of error from 1000 randomly selected trials. (a) non-stereo error from ground, (b) stereo error from ground, (c) non-stereo error from median, (d) stereo error from median. Radial distance is the magnitude of the error, compass direction is the direction on the screen.

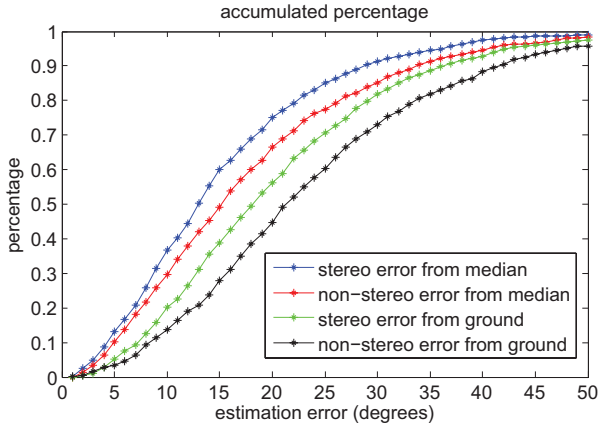


Fig. 8. Cumulative distribution (empirical) of the estimation error.

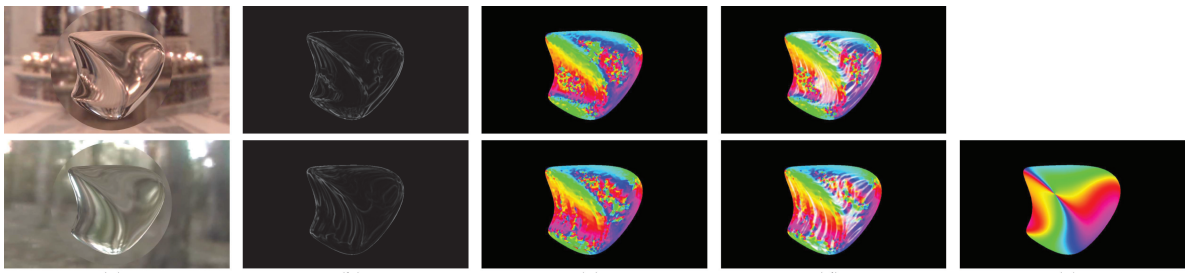
(a) 16 largest errors (non-stereo) (b) 16 largest errors (stereo)

Fig. 9. Example of localized errors. The gauges show median vectors of estimated normal direction for the positions with the largest errors.

distribution of errors as a function of increasing estimation error. We found that errors from ground truth were larger than those from the median vector, and errors from non-stereoscopic were larger than those from stereoscopic views. Compared with chance performance (mean error of 90°) or the estimation error from smooth shaded images (mean error of 24° reported by Cole et al. [2009]), both non-stereoscopic and stereoscopic view provided reliable estimation, and estimates from stereoscopic view were more reliable than those from non-stereoscopic view.

4.2 Is Error from Ground Truth Localized?

Since we had densely sampled test points we were able to analyze if error was localized in a few important areas. For each image condition, we used the median vector to represent the estimated normal direction at each location and sorted the estimation error in nondecreasing order. We selected the 16 (out of 80 in total) surface locations with the largest errors and rendered their median vector as the gauge figure in the original image. Figure 9 shows one of these examples. We found that the highest errors were localized in two particular types of areas. One was the ridge areas where the normal directions changed too fast to be estimated accurately. The other was smoothly textured areas, which will be further analyzed in Section 4.3.



(a) (b) (c) (d) (e)

Fig. 10. Orientation maps. (a) an object rendered in two scenes, (b) edge map, (c) orientation map, (d) edge-aware orientation map, (e) orientation ground truth.

4.3 Edge-Aware Orientation Map

Inspired by the similarity between mirror reflection and transparent refraction and previous work of [Fleming et al. 2004, 2009], we used an edge-aware orientation map to analyze how people are able to perceive the shape of transparent objects with simple refraction geometry (single refractive interface or a "thin lens").

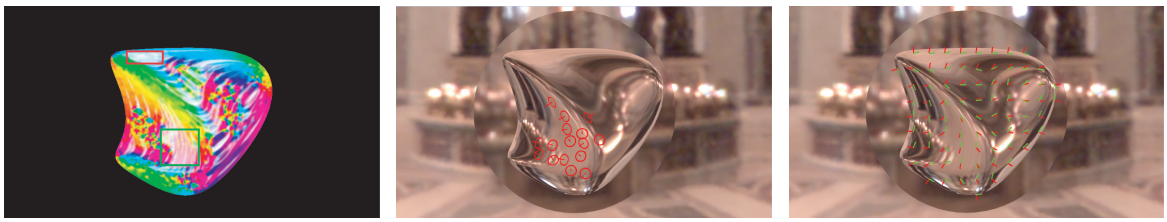
To visualize optical information from images and 3D models, we rendered the same object with two different environment maps (Figure 10(a)) and presented three images to illustrate the visual information in the object surface areas. The first image was an edge map. It was obtained by calculating gradient magnitudes at each pixel using the Sobel operator in OpenCV [Bradski and Kaehler 2008]. Figure 10(b) shows that an object can have similar edges under very different environment maps. Most of these edges were located in boundaries and ridge areas of the object. Because edges often arise from object structure that is relevant to the goal of visual perception, they can provide basic shape information, such as boundary and concavity/convexity.

The second image was an orientation map. We first calculated eigenvalues and eigenvectors at each pixel from the structure tensor [Knutsson et al. 2011], which has been widely used in edge detection and feature extraction [Harris and Stephens 1988]. Specifically, the structure tensor is:

$$s = \sum_u \sum_v w(u, v) \begin{bmatrix} {I_x}^2 & I_x I_y \\ I_x I_y & {I_y}^2 \end{bmatrix}, \tag{2}$$

where I_x and I_y are the partial derivatives of image intensity I with respect to x and y, and w is a Gaussian smoothing operator. We calculated the structure tensor from a gray scale image in a neighborhood, u and v (7×7 pixels in practice), surrounding a given pixel. We adopted the direction of least curvature (the one whose corresponding eigenvalue has the smallest absolute value) as the local orientation (range in [0, 180) degrees). We found that edge-aware orientation maps of a given object in different environments could be very similar (Figure 10(c)). This similarity may underlie the ability of people to consistently perceive shapes of transparent objects in different scenes. Fleming et al. [2004] reached a similar conclusion when they studied reflective objects.

The third image was an edge-aware orientation map combining the orientation map with a confidence map. We used the edge map for the confidence metric because (1) the visual system is more sensitive to the variation in intensity than the intensity itself [Van Nes et al. 1967] and (2) the orientation of a constant or slowly varying intensity region is ill-defined. There is a clear ecological utility to the visual system's sensitivity to variations over space and time. Being able to accurately sense changes in the environment is crucial to our survival [Thompson et al. 2011]. Figure 10(d) shows an



(a) (b) (c)

Fig. 11. Example of areas in an edge-aware orientation map with little orientation information. (a) Two areas (denoted by two rectangles) with little local orientation information in the image (ground truth orientation is in Figure 10(e)). (b) The 16 gauge settings with the largest error for this stimulus (non-stereoscopic view). (c) Azimuth angles (red) and orientation angles (green).

example of edge-aware orientation maps of a transparent object rendered in two different scenes. The orientation is coded by hue and the confidence by saturation; if the color of pixels is more desaturated, then this indicates that these pixels have less salient orientation information.

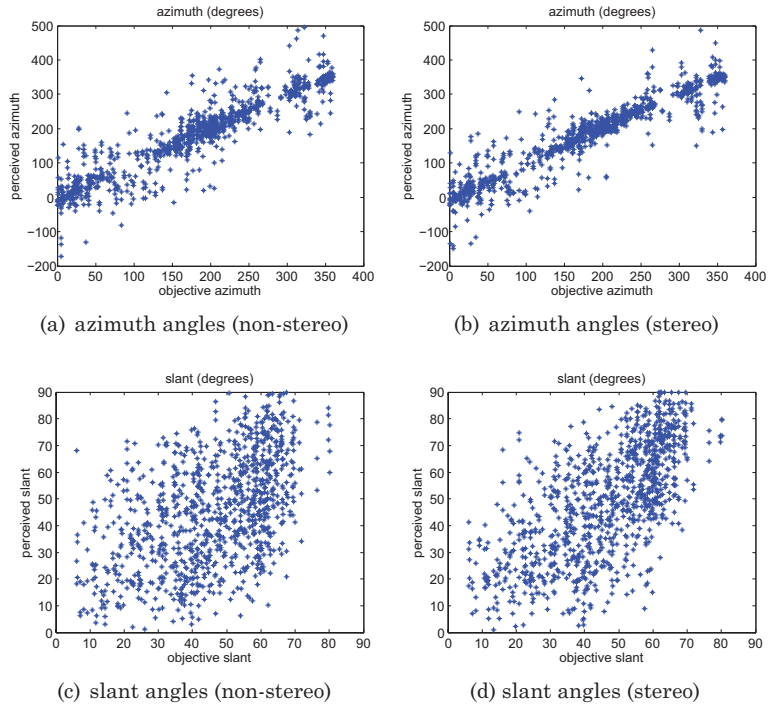
For each object, we obtained the orientation ground truth by projecting surface normals to the image plane. We aligned the image plane with the $X-Y$ plane, and set the direction normal to the image and towards the observer as the Z direction. We decomposed the normal direction into two angles: a slant angle ϕ, and an azimuth angle θ:

$$\phi = arccos(z), \theta = arctan\left(\frac{y}{x}\right), \tag{3}$$

where $[x, y, z]^T$ is the normalized normal vector.

We found that the direction of the azimuth angle is perpendicular to the direction of minimum curvature in the depth map (for proof see appendix A). Since the dominant orientation angle is aligned with the direction of minimum curvature, we calculated the orientation angle of the model by rotating the azimuth angle 90°. Compared with the estimating minimum curvature from the depth map, obtaining the orientation angle from the azimuth angle is more robust because we already have accurate value of normals for each pixel from the rendering process. This orientation map (for example, see Figure 10(e)) was regarded as the ground truth. Note that the pixel colours are saturated since these are ground truth data (confidence is high). Figure 11(c) shows the azimuth angles (red) and orientation angles (blue) of a stimulus. If the true azimuth angle and the orientation in the image are perpendicular to each other in a local area, this area may provide accurate visual cues for the surface orientation.

Figure 11(a) shows two areas with little orientation information in an image. The first area (denoted with a red rectangle) is on the top left of the object. It is close to the object boundary and most of its nearby areas have accurate orientation information. On the other hand, the second area (denoted with a green rectangle) is relatively far from the object boundary and its nearby areas have large orientation errors. Figure 11(b) draws the 16 (out of 80 in total) gauge settings with the largest error for this stimulus. We found that the sampling points in the green rectangle area had larger error than those in the red rectangle areas even though both areas lacked local orientation information. How did people have very different estimation accuracy in two areas that both had little local orientation information? One possible explanation is that people tend to interpolate the orientation information from nearby areas when the local orientation information is impoverished or ambiguous. The shape of the first area was perceived relatively more accurately than that of the second area since the subjects could interpolate "correct" orientation information from nearby areas. This figure also demonstrates the predictive capability of our edge-aware orientation map.



(a) azimuth angles (non-stereo) (b) azimuth angles (stereo)

(c) slant angles (non-stereo) (d) slant angles (stereo)

Fig. 12. Plot of azimuth and slant angles from 1000 randomly selected trials.

4.4 How Does Stereoscopic View Improve Accuracy?

In developing the edge-aware orientation maps we noted that the projection of the normal direction to image space was perpendicular to the ground truth of the orientation map. The normal direction could be estimated from the structure tensor of the depth map and more accurate depth information could help to determine the azimuth (tilt) and slant angles of the surface normal [Stevens 1983]. It is well known that stereospsis can provide robust relative depth information for surface orientation estimation [Kersten et al. 2006; Bridge and Cumming 2008; Allison et al. 2009]. Therefore, the accuracy of perceived azimuth angles could be improved by more accurate depth information from stereoscopic view.

However, disparities of features in a transparent object seldom correspond to the depth of the surface itself and often violate the epipolar constraint (i.e. have vertical disparities and skew projections of corresponding points into the scene). For specular objects, these disparity signatures have long been used by computer vision stereo algorithms to help distinguish Lambertian and specular features and to aid surface reconstruction [Blake and Brelstaff 1988]. A recent study suggests that similar processes may operate for specular perception in humans [Muryy et al. 2012]. It is possible that stereopsis may improve performance by providing depth information directly in areas of reliable stereoscopic cues as well as by helping to identify areas corresponding to refractive specularities.

Figure 12 shows the azimuth and slant angles from 1000 randomly selected trials compared with the corresponding ground truth angles. For azimuth angle, the mean error with non-stereoscopic view was 28.7°, and with stereoscopic view was 23.5°. For slant angle, the mean error with non-stereoscopic view was 15.0°, and with stereoscopic view was 13.0°. Azimuth angle estimation error was 18% smaller for stereoscopic view compared to nonstereoscopic view. This result confirms our

prediction that stereoscopic 3D indeed improves the perceived accuracy of azimuth angle, and finally improves the accuracy of perceived surface orientations.

5. CONCLUSION

5.1 Main Findings

First, people are able to consistently perceive the normal direction of thin transparent objects. This was demonstrated by experiments using the gauge task and theoretical analysis using edge-aware orientation maps. The task is difficult and substantial errors remain; however, performance is similar to previous reports for specular and Lambertian surface perception [Fleming et al. 2004; Cole et al. 2009].

Second, stereoscopic viewing improves the precision of surface orientation perception. This was demonstrated by comparing the perceptual accuracy from stereoscopic and non-stereoscopic viewing conditions. The relationship (perpendicularity) between the orientation angle ground truth and the azimuth angle of normal directions provides a geometric basis for this improvement.

Third, perceptual errors were concentrated in particular areas of the refractive surface. We found observers had larger perceptual errors in two major areas. The first area was the ridge area in which normal directions change too fast to be perceived accurately. The second area was smoothly textured areas surrounded by "incorrect" orientation information in the image.

5.2 Generality of the Result

For transparent and other objects, there is a large space of variation in shapes and materials and these objects can be placed in many types of environments under diverse lighting conditions. The shape can be varied from regular geometries such as cubes and spheres to highly complicated forms. Also the object can have all kinds of colors, translucency and reflection/refraction properties. Although we only use 4 relatively simple objects as the stimuli and future work is required to determine whether the findings generalize to larger classes of objects, since the objects were randomly generated and had variation in surface curvature they can, to some extent, represent many potential shapes.

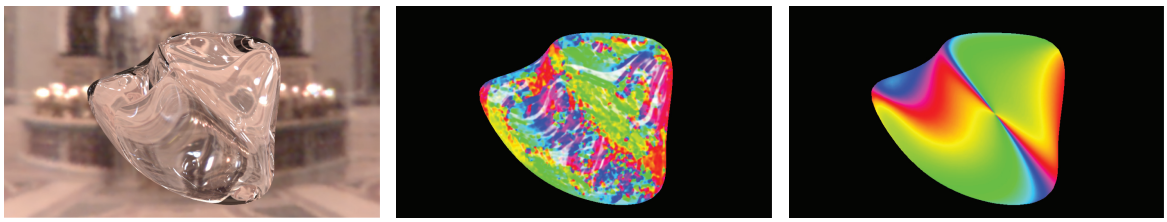
While we tested only 8 subjects we have a considerable amount of data for each due to the repeated-measures design. Combined with the fact that the experiments took place in a well-controlled environment we have good estimates for individual observers. This combined the fact that intersubject variability was modest lead us to believe the results will generalize to a larger sample.

As previously discussed the results do not necessarily generalize to thick transparent surfaces. However, example images suggest that the orientation field approach may have utility here, but this remains to be tested.

5.3 Limitations and Future Work

Our work has several limitations that will be addressed in future work.

First, the objects were designed so that most of light was refracted only once. Although they represent several important natural phenomena, many transparent objects violate this constraint; for example, when light crosses multiple optical interfaces such as the front and back of a thick glass object or when there is internal reflection. The light paths in those objects are much more complex than those that we studied in this article. However, we could extend this study to thick transparent objects. Figure 13 shows an example of a thick transparent object and its orientation maps. The figure suggests that the edge-aware orientation map might still be able to provide correct shape information in some areas.



(a) original image (b) edge-aware orientation map (c) ground truth orientation map

Fig. 13. Edge-aware orientation map of a thick transparent object.

Second, the current experiment used only one environment map to generate stimuli, which may not represent the diversity of natural scenes. While empirical evidence from previous work [Fleming et al. 2004] and theoretical analysis in Section 4.3 showed that shape from reflective/refractive objects mainly depends on the orientation map rather than information from the scene, it is still necessary to test subjects in different natural scenes and test the generality of the findings. While natural scenes have wide variety but well-known statistical regularity, the use of orientation information in transparent shape perception suggests there may be differences between object perception in natural environments and carpentered environments with dominant horizontals and verticals.

Finally, we found that particular areas of objects were difficult to perceive even with stereoscopic viewing. It would be interesting to investigate what types of objects are most challenging for shape perception.

APPENDIX

A. ORIENTATION FROM PROJECTED NORMAL DIRECTION

For a position (x, y) in a depth map I, the structure tensor (with block size of 1×1) is:

$$\begin{bmatrix} {I_x}^2 & I_x I_y \\ I_x I_y & {I_y}^2 \end{bmatrix}, \tag{4}$$

where I_x and I_y are the partial derivatives I with respect to x and y. The secondary eigenvector (the one whose corresponding eigenvalue has the smallest absolute value) of the structure tensor is:

$$\vec{e} = [-I_y/I_x \;\; 1]^T \tag{5}$$

and its direction can be considered as the orientation of the 3D model [Fleming et al. 2009].

PROOF. Without loss of generality, we will show that the projection of a normal direction (in a local space) to the image plane is perpendicular to the secondary eigenvector of the structure tensor.

Given a pixel position (x, y) with the gradient $[I_x \;\; I_y]^T$ in the depth map I, the normal of the local patch is:

$$\vec{n} = [-I_x \;\; -I_y \;\; 1]^T. \tag{6}$$

Its projection to image plane is:

$$\vec{n}_p = [-I_x \;\; -I_y]. \tag{7}$$

Because $\vec{e} \cdot \vec{n}_p = 0$ where $\cdot$ is the dot operator, proof done. □

REFERENCES

ALLISON, R. S., GILLAM, B. J., AND PALMISANO, S. A. 2009. Stereoscopic discrimination of the layout of ground surfaces. *J. Vision 9*, 12.

BARROW, H. G. AND TENENBAUM, J. M. 1981. Computational vision. *Proc. IEEE 69*, 5, 572–595.

BELHUMEUR, P. N., KRIEGMAN, D. J., AND YUILLE, A. L. 1999. The bas-relief ambiguity. *Int. J. Comput. Vision 35*, 1, 33–44.

BLAKE, A. AND BRELSTAFF, G. 1988. Geometry from specularities. In *Proceedings of the IEEE International Conference on Computer Vision*. 394–403.

BRADSKI, G. AND KAEHLER, A. 2008. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media.

BRIDGE, H. AND CUMMING, B. G. 2008. Representation of binocular surfaces by cortical neurons. *Curr. Opin. Neurobiol. 18*, 4, 425.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S., AND SINGH, M. 2009. How well do line drawings depict shape? *ACM Trans. Graph. 28*, 3, 28.

DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 189–198.

FLAVELL, L. 2010. *Beginning Blender: Open Source 3D Modeling, Animation, and Game Design*. Apress.

FLEMING, R., TORRALBA, A., AND ADELSON, E. 2004. Specular reflections and the perception of shape. *J. Vision 4*, 9.

FLEMING, R. W., DROR, R. O., AND ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *J. Vision 3*, 5.

FLEMING, R. W., JÄKEL, F., AND MALONEY, L. T. 2011. Visual perception of thick transparent materials. *Psych. Sci. 22*, 6, 812–820.

FLEMING, R. W., TORRALBA, A., AND ADELSON, E. H. 2009. Shape from sheen. Tech. rep., MIT, CSAIL. 10.

GIBSON, J. J. 1950. The perception of visual surfaces. *Amer. J. Psych. 63*, 3, 367–384.

HARRIS, C. AND STEPHENS, M. 1988. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*. Vol. 15.

KERSTEN, M. A., STEWART, A. J., TROJE, N., AND ELLIS, R. 2006. Enhancing depth perception in translucent volumes. *IEEE Trans. Visual. Comput. Graphics 12*, 5, 1117–1124.

KNUTSSON, H., WESTIN, C.-F., AND ANDERSSON, M. 2011. Representing local structure using tensors ii. In *Image Analysis*, Springer, 545–556.

KOENDERINK, J. J., VAN DOORN, A. J., AND KAPPERS, A. M. 1992. Surface perception in pictures. *Percept. Psychophysics 52*, 5, 487–496.

KOENDERINK, J. J., VAN DOORN, A. J., KAPPERS, A. M., AND TODD, J. T. 2001. Ambiguity and the 'mental eye' in pictorial relief. *Perception-London 30*, 4, 431–448.

LEE, Y. AND SAUNDERS, J. 2011. Stereo improves 3d shape discrimination even when rich monocular shape cues are available. *J. Vision 11*, 9.

MARR, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt, New York, NY.

MURYY, A. A., FLEMING, R. W., AND WELCHMAN, A. E. 2012. Binocular cues for glossiness. *J. Vision 12*, 9, 869–869.

NORMAN, J. AND TODD, J. 1996. The discriminability of local surface structure. *Perception-London 25*, 4, 381–398.

SCHNEIDER, S. 2012. Stereoscopic cameras in blender 2.6. http://www.noeol.de/s3d/.

STEVENS, K. A. 1983. Slant-tilt: The visual encoding of surface orientation. *Biol. Cybern. 46*, 3, 183–195.

SUN, J. AND PERONA, P. 1998. Where is the sun? *Nature Neurosci. 1*, 3, 183–184.

THOMPSON, W. B., FLEMING, R. W., CREEM-REGEHR, S. H., AND STEFANUCCI, J. K. 2011. *Visual Perception from a Computer Graphics Perspective*. AK Peters Limited.

TODD, J. T., OOMES, A. H., KOENDERINK, J. J., AND KAPPERS, A. M. 2004. The perception of doubly curved surfaces from anisotropic textures. *Psych. Sci. 15*, 1, 40–46.

VAN NES, F. L., BOUMAN, M. A., ET AL. 1967. Spatial modulation transfer in the human eye. *J. Opt. Soc. Am. 57*, 3, 401–406.

Received June 2013; accepted July 2013