

The effects of phase on the perception of 3D shape from texture: Psychophysics and modeling [☆]

Lore Thaler ^{*}, James T. Todd, Tjeerd M.H. Dijkstra

Department of Psychology, The Ohio State University, Columbus, OH 43210, United States

Received 5 July 2006; received in revised form 29 September 2006

Abstract

Two experiments are reported in which observers judged the apparent shapes of elliptical cylinders with eight different textures that were presented with scrambled and unscrambled phase spectra. The results revealed that the apparent depths of these surfaces varied linearly with the ground truth in all conditions, and that the overall magnitude of surface relief was systematically underestimated. In general, the apparent depth of a surface is significantly attenuated when the phase spectrum of its texture is randomly scrambled, though the magnitude of this effect varies for different types of texture. A new computational model of 3D shape from texture is proposed in which apparent depth is estimated from the relative density of edges in different local regions of an image, and the predictions of this model are highly correlated with the observers' judgments.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Visual perception; Depth perception; 3D shape; Texture; Psychophysics; Modeling; Fourier phase; Fourier amplitude

1. Introduction

In an influential series of articles that was first published over 50 years ago, James Gibson identified a new source of visual information about 3D surface structure that he referred to as gradients of optical texture. A compelling example of the perception of 3D shape from texture can be obtained by examining the image presented in Fig. 1, which depicts a smoothly curved object that is covered with a random pattern of circular polka dots. Because of the effects of perspective, the optical projections of these polka dots have variable sizes and shapes that are determined by their relative distances and orientations with respect to the point of observation. It is the pattern of these systematic variations (i.e. the gradients) that produces the perceptual appearance of a 3D surface.

Although numerous computational analyses have been developed for determining 3D shape from texture (e.g., Aloimonos, 1988; Bajcsy & Lieberman, 1976; Blake, Buett-hoff, & Sheinberg, 1993; Gårding, 1992, 1993; Knill, 1998, 2001; Malik & Rosenholtz, 1997; Stevens, 1881), they are all based on a fundamental assumption that variations in reflectance on a visible surface have a highly constrained statistical distribution. Most models assume that the distribution is isotropic (i.e. invariant over rotation), though there are some that are based on a somewhat less restrictive assumption that the distribution is homogeneous (i.e. invariant over translation). Whenever these constraints are satisfied, any distortions of isotropy or homogeneity within the optical projection of a surface can be attributed to variations in the surface geometry, such as depth, slant or curvature.

It is interesting to note that existing computational models for determining 3D shape from texture can be separated into two general categories based on the specific procedures that are employed for measuring optical texture. One popular approach to this problem is to first perform a Fourier

[☆] This research was supported by a grant from NSF (BCS-0546107).

* Corresponding author. Fax: +1 614 292 5601.

E-mail address: thaler.11@osu.edu (L. Thaler).

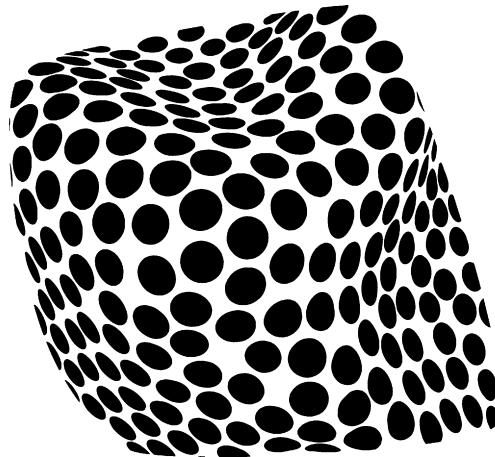


Fig. 1. A pattern of optical texture that is perceptually interpreted as a smoothly curved 3D surface.

transform on various local neighborhoods of an image. The surface geometry depicted in each image region is then estimated either from properties of its local amplitude spectrum (e.g. Bajcsy & Lieberman, 1976; Brown & Shvayster, 1990; Kanatani & Chou, 1989; Krumm & Shafer, 1992; Ribeiro & Hancock, 2000; Sakai & Finkel, 1995; Super & Bovik, 1995), or from systematic changes between amplitude spectra in neighboring regions (Malik & Rosenholtz, 1994, 1997). When considering this class of models, it is important to recognize that by focusing exclusively on the amplitude spectrum, all information that could potentially be provided by the phase spectrum is effectively ignored.

An alternative approach for the measurement of optical texture is to extract the edges within various local neighborhoods of an image. The surface geometry depicted in each region can then be estimated either from the distribution of edge orientations (Aloimonos, 1988; Blake & Marinos, 1990; Blostein & Ahuja, 1989; Marinos & Blake, 1990; Witkin, 1981), or from systematic changes in the distributions of edges across neighboring regions (Gårding, 1993). A fundamental difference between this approach and the analysis of local amplitude spectra is that the extraction of edges is critically dependent on aspects of image structure that are represented in the Fourier domain by the phase spectrum (Oppenheim & Lim, 1981; Piotrowski and Campbell, 1982).¹

In order to illustrate the effects of phase on an image's apparent edge structure more clearly, it is useful to consider-

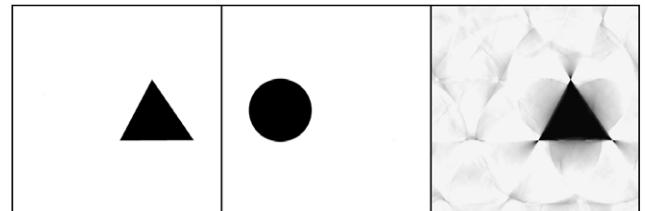


Fig. 2. The relative effects of the amplitude and phase spectra on the appearance of image edges. The left panel shows a white square with black triangle on the right. The middle panel shows a white square with black circle on the left. The right panel shows a composite image that has the amplitude spectrum from the middle image and the phase spectrum from the left image.

er the images in Fig. 2. The left panel of this figure shows a white square with a black triangle on the right. The middle panel shows a white square with a black circle on the left. The right panel was created by combining the amplitude spectrum from the image in the middle panel with the phase spectrum from the image on the left. The result shows clearly that the phase spectrum is what primarily determines the apparent pattern of edges.

Which class of models is most consistent with the analysis of 3D shape from texture in human perception? One possible way of addressing this issue would be to independently manipulate the information provided by the amplitude and phase spectra of local image regions. If the perception of 3D shape from texture is based entirely on the amplitude spectra, as in many computational models, then independent manipulations of the phase spectra should have no measurable effect on performance.

In an effort to test this hypothesis, Li and Zaidi (2001) compared the accuracy of observers' sign of curvature judgments for sinusoidally corrugated surfaces with two different types of texture. On half of the displays, the surfaces were textured using a set of natural patterns that were originally photographed by Brodatz (1966). On the remaining displays, the surface textures were created by combining the amplitude spectra of the Brodatz patterns with random phase spectra.² The results revealed no detectable differences between observers' judgments for the two types of displays, thus suggesting that phase information may indeed be irrelevant for the perception of 3D shape from texture.

However, there are two important reasons to question the generality of this finding. First, the sign of curvature judgments obtained by Li and Zaidi (2001) are a relative crude measure of apparent 3D shape, which could not have

¹ The phase spectrum of an image determines its higher order correlation function and therefore its higher order statistics, such as skew and kurtosis (Yellott, 1993; Thomson, 1999; Thomson, Foster, & Summers, 2000). Skew measures the symmetry in the distribution of pixel intensities, whereas kurtosis measures how much the distribution clusters around certain values. Previous research has demonstrated that these higher order statistics provide perceptually useful information for texture discrimination (Yellott, 1993) and natural scene identification (Thomson, 1999; Thomson et al., 2000).

² For patterns of surface markings that are homogeneous, scrambling the phase spectrum of a physical texture has no effect on amplitude spectra for local regions of its optical projection. Note that this would not be true, if the same manipulation was applied globally to the projection of a textured surface like the one shown in Fig. 1. In this case, phase scrambling would significantly alter the amplitude spectra within individual local image regions, although the global amplitude spectrum (by definition) remains unchanged.

revealed any potential effects of phase on the overall magnitude of perceived relief. Second, the viewing conditions employed in their study were less than optimal, in that the level of perspective³ for many of the texture patterns was well below the threshold required for accurately determining the sign of surface curvature (see Todd et al., 2005). Thus it is possible that the failure to obtain significant effects of phase was due primarily to a floor effect.

Because of these concerns, the research described in the present article was designed to investigate the possible effects of phase on the magnitude of perceived depth for surfaces presented with a variety of different textures and with varying degrees of perspective. The stimuli depicted convex elliptical cylinders whose major axis lengths and orientations varied across trials. Observers indicated the apparent cross section in depth of each stimulus surface by adjusting the shape of an elliptical arc presented on a separate monitor. The key aspect of the experimental design is that the set of possible stimulus images was organized into matched pairs that had the same local amplitude spectra, but whose local phase spectra were quite different. If the perception of 3D shape from texture is based entirely on local amplitude spectra, as argued by Li and Zaidi (2001), then orthogonal variations in the local phase spectra should have no significant effect on performance.

2. Experiment 1

2.1. Methods

2.1.1. Subjects

Six observers participated in the experiment, including two of the authors (LT and JT), and four others who were naïve about the issues being investigated. All subjects had normal or corrected to normal visual acuity.

2.1.2. Apparatus

The experiment was conducted using a Dell Dimension 8300 PC with an ATI Radeon 9700 PRO graphics adapter. Stimulus images were presented at a viewing distance of 83 cm on a standard CRT with a spatial resolution of 1280×1024 pixels and a temporal resolution of 75 Hz. The dimensions of the active display area were 36.7×29.3 cm, and the size of each stimulus image was 29.3×29.3 cm. Thus, each image subtended 20° of visual angle and matched the camera angle with which it was rendered. The displays were viewed monocularly with an eye patch, and a chin rest was used to constrain head movements.

2.1.3. Stimuli

The stimuli depicted upright elliptical cylinders whose occlusion contours were constrained to be located at the left and right boundary of each image. The center of the elliptical major axis was located at a distance of 200 cm from the observers' eye. Given the occlusion and center constraints, an elliptical cylinder can be uniquely defined by the x and z coordinates of the "near point" along a horizontal cross-section that is closest to the observer in depth (see Appendix A). Nine different surfaces were created with varying extensions in depth, and the optical projections of their near points were all located at one of 3 possible eccentricities relative to the center of the image. Fig. 3 shows the basic scene geometry and sample stimulus images for each of these shapes. Mirror reflected versions of these images about a vertical axis were also used on half the trials in order to avoid directional biases. The elliptical arc above each sample image shows a horizontal cross-section of the depicted surface as seen from above. The depth of each surface was defined as the distance between its near point and the farthest occlusion boundary, and these varied in magnitude between 28 and 81 cm.

Each of the different surface shapes was rendered with 6 possible textures, which are shown in Fig. 4. The top row of the figure shows the texture patterns with their original Fourier phase spectra, and the middle row shows these same patterns with randomly scrambled phase spectra. From left to right the textures are denoted as 'dots', 'lines' and 'flagstones', respectively, and their amplitude spectra are shown in the bottom row. The phase scrambled textures were generated by combining the amplitude spectrum of the original texture pattern with a random phase spectrum that was sampled from a uniform distribution between zero and 2π . Because phase scrambling alters the maximum and minimum luminance values within an image, it can create values that fall outside the monitor's possible intensity range of 0–255, and the intensity distribution in that case must be clipped or rescaled. In order to avoid that, the contrasts of the original textures were adjusted so that the intensity values subsequent to phase scrambling would all fall within the acceptable range. As a result of this control, the original textures and their corresponding phase scrambled versions had identical mean luminance and root mean squared luminance contrast.

It is important to keep in mind that phase scrambling was performed on the surface textures, not on the projected images of elliptical cylinders that were used as stimuli. When a texture is mapped onto a 3D surface and then optically projected onto the image plane, the effects of perspective will cause systematic variations in the amplitude and phase spectra within different local image regions. In order to assess the effects of phase on human perception it is essential that the local amplitude spectra of the actual stimulus images are appropriately matched in the scrambled and unscrambled

³ The perspective of a scene is typically defined as a ratio between the depth of its nearest point and the depth of its farthest point (Braunstein & Payne, 1969). For any surface region in an image that is slanted in depth, the magnitude of perspective varies systematically with its depicted field of view (see Todd, Thaler, & Dijkstra, 2005).

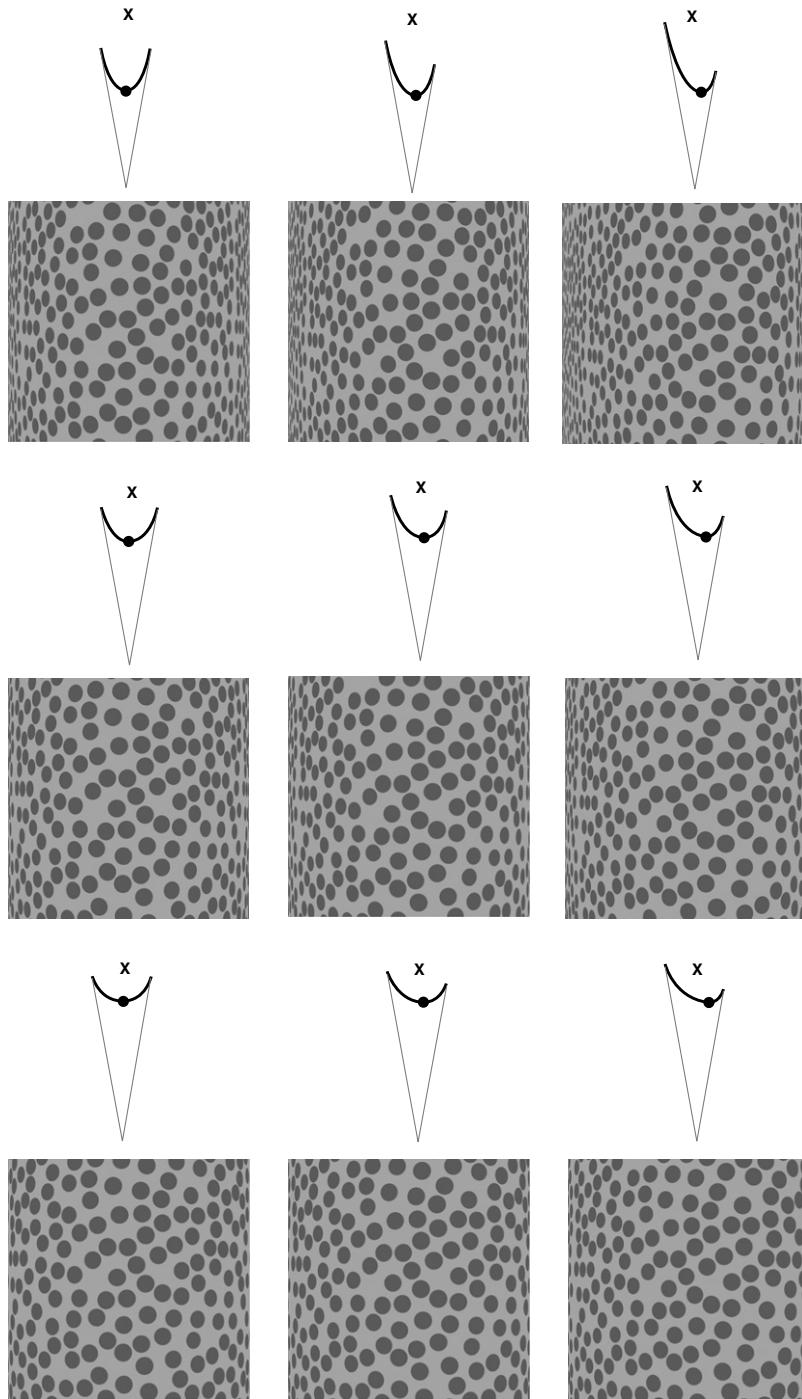


Fig. 3. Some examples of the stimulus images used in Experiment 1. The elliptical arc on top of each image shows a horizontal cross-section of the depicted surface as viewed from above. Thin straight lines depict the 20° viewing cone. For illustration purposes the point on the generating curve nearest to the image plane is denoted with a dot.

conditions. To confirm if that condition was satisfied, each of the possible stimulus images was partitioned into an 8 × 8 grid of local square regions, and their local amplitude and phase spectra were computed individually. For each combination of base texture and surface shape, a regression analysis was performed to compare the local amplitude spectra between the images rendered with

scrambled and unscrambled textures. To provide a baseline for comparison, regression analyses were also performed between the local amplitude spectra of the unscrambled stimuli and those obtained from a duplicate set of images of the same surfaces whose textures were repositioned relative to how they were mapped in the experimental stimuli. Thus, the experimental and

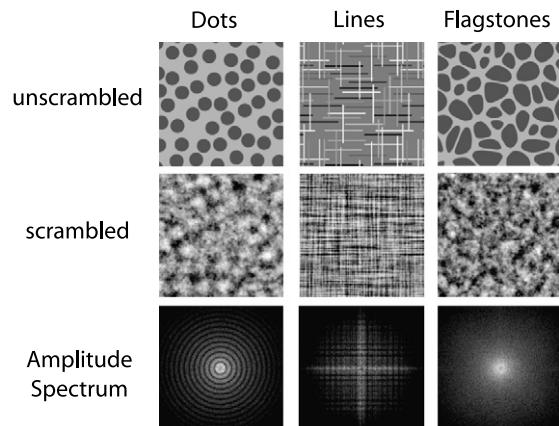


Fig. 4. The textures used in Experiment 1. The top row shows the three original textures before any phase manipulations were performed. The middle row shows the same three textures with randomly scrambled phase spectra, and the bottom row shows their amplitude spectra.

Table 1

The results of a regression analysis from Experiment 1 that compared the local amplitude spectra in corresponding image regions with scrambled and unscrambled textures, and corresponding regions with unscrambled textures in different surface positions (see text for details)

		Unscrambled vs. scrambled		Unscrambled vs. unscrambled	
		r^2	Slope	r^2	Slope
Dots	Mean	0.97	0.99	0.97	0.98
	Stdev	0.01	0.05	0.01	0.04
Lines	Mean	0.97	0.99	0.97	0.99
	Stdev	0.01	0.05	0.01	0.03
Flagstones	Mean	0.97	0.98	0.97	0.98
	Stdev	0.01	0.05	0	0.03

duplicate images shared the same overall patterns of texture, but the positions of their image contours were independent of one another.

To summarize the regression analysis, there were 576 correlations performed between the scrambled and

unscrambled versions of each texture (9 depicted shapes by 64 local regions per image), and 576 correlations between the experimental and duplicate versions of each unscrambled texture. The results are presented in Table 1, which shows the average r^2 and slope for each texture, as well as their standard deviations. Let us first consider the results from the baseline comparisons between the experimental and duplicate versions of the unscrambled textures. The results in Table 1 confirm that repositioning a homogeneous texture pattern on a surface has a negligible effect (less than 3%) on the local amplitude spectra within its optical projection. This is indeed the primary underlying assumption for the gradient based model of 3D shape from texture proposed by Malik and Rosenholtz (1997). Note in addition that the effects of phase scrambling on the local amplitude spectra are equally small. Thus, if the perception of 3D shape from texture is based entirely on local amplitude spectra, as is the case for several computational models, there should be no significant differences in observers' shape judgments for surfaces depicted with scrambled or unscrambled textures.

2.1.4. Procedure

On each trial, an image of an elliptical cylinder was presented on a computer monitor that was directly in front of the point of observation. A second monitor was located off to the side of the main display that contained an adjustment figure, which observers could manipulate with a hand held mouse to match the apparent cross section in depth of the depicted surface. The overall design of the adjustment screen is shown in Fig. 5. The adjustment figure contained an elliptical arc segment whose shape could be manipulated by controlling the position of its near point with horizontal and vertical motions of the mouse (see Appendix A). The shape of the adjustment figure followed the same geometry as the stimulus, but was reduced in scale to fit on the display screen – i.e., 1 cm in stimulus space corresponded to 1 mm in adjustment space. Observers were instructed to manipulate the shape of the adjustment figure so that it matched the apparent cross-section in depth of the depicted

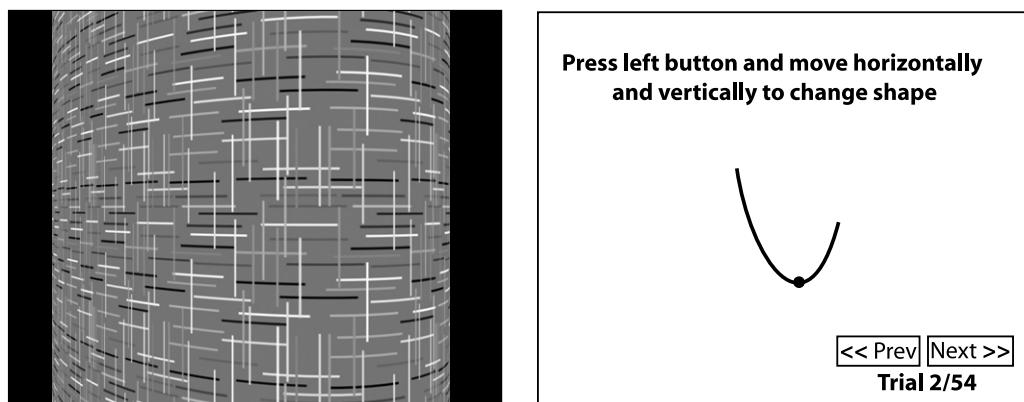


Fig. 5. An example of a stimulus image that was presented directly in front of the observer and the adjustment display was presented to the right of the primary monitor. Observers were instructed to manipulate the shape of the adjustment figure so that it matched the apparent cross-section in depth of the depicted textured surface as closely as possible.

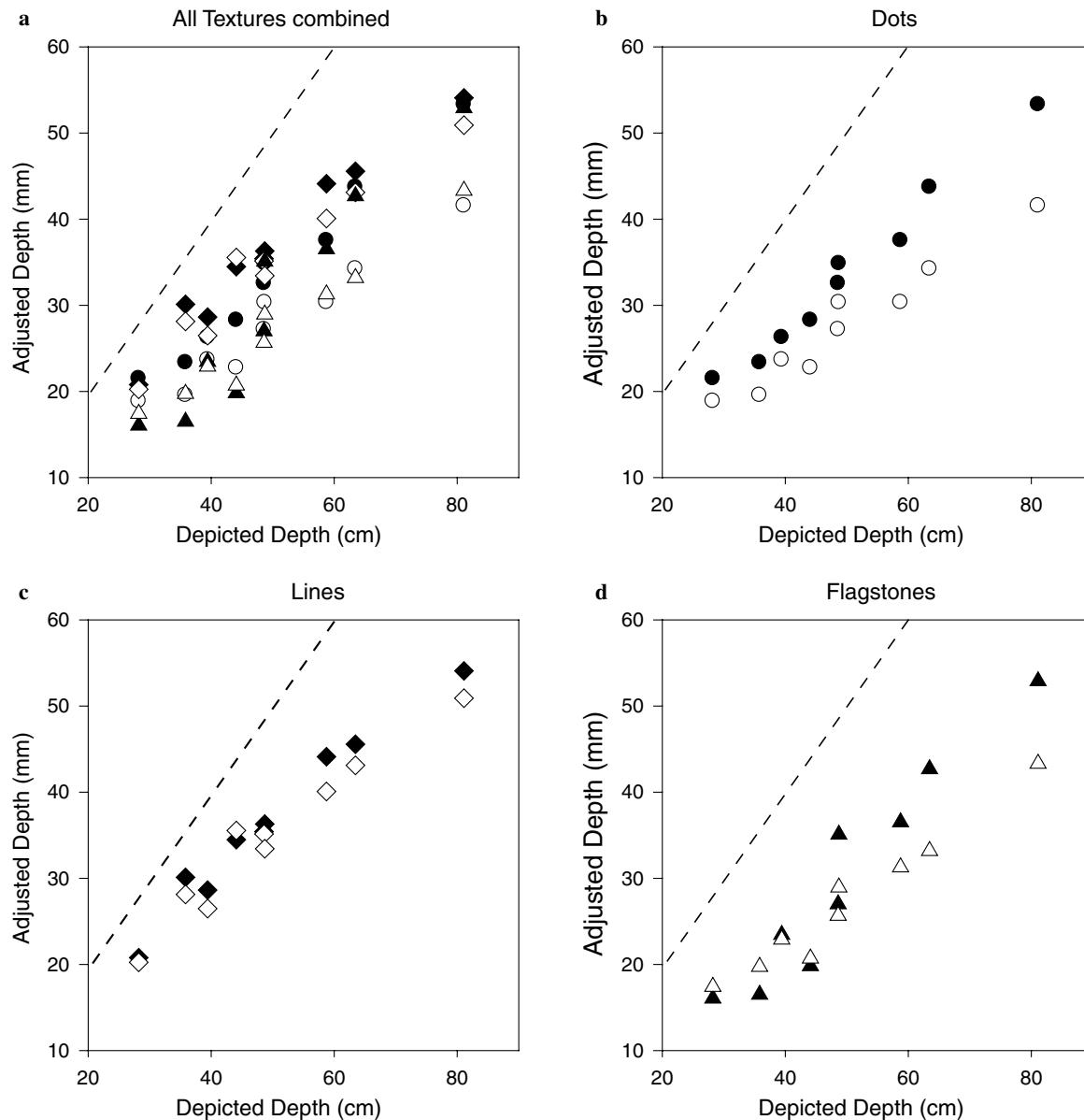


Fig. 6. The average depth settings in Experiment 2 as a function of the depicted stimulus depth for all textures combined (a) and each texture individually (b–d). The scrambled and unscrambled conditions are represented by open and closed symbols, respectively, and the dashed line represents veridical performance. The different units on the horizontal and vertical axes are due to the fact that the adjustment stimulus was reduced in scale to fit on the screen (1 cm in stimulus space = 1 mm in adjustment space).

textured surface as closely as possible. Once observers were satisfied with their settings they could move to the next trial by clicking on a button that was labeled 'next'. It was also possible to move backwards in the sequence to modify a previous response, though none of the observers reported making use of that option. All observers agreed that the task felt quite natural and that they had high degree of confidence in their judgments.

2.1.5. Design

To summarize the overall experimental design, there were 54 experimental conditions defined by the three texture types (dots, lines, or flagstones), two types of phase

(scrambled or unscrambled) and nine surface shapes. Within a given experimental session, each of these conditions was presented once in a random sequence, and each observer participated in eight sessions.

2.2. Results

Fig. 6a shows the average depth settings plotted against the ground truth for all of the different textures. Separate plots for the scrambled and unscrambled versions of each individual texture are presented in **Figs. 6b–d**. The dashed line in each graph represents veridical performance. It is clear from these data that the apparent depth of the surfaces

Table 2

Average within-observer standard deviations (in mm) for the depth settings in Experiment 1

	Unscrambled	Scrambled
Dots	6.58	6.47
Lines	6.67	6.76
Flagstones	7.16	6.88

Averages were computed by collapsing over depth conditions and observers.

es varied linearly with the ground truth in all conditions, and that the overall magnitude of surface relief was systematically underestimated. These findings are consistent with those obtained in several previous investigations (e.g., see Todd et al., 2005). For the surfaces textured with lines, there were no significant differences in judged depth between the scrambled and unscrambled phase spectra. That is not the case, however, for the surfaces that were textured with dots or flagstones. Note in Figs. 6b and 6d that the phase scrambled versions of these textures produced psychometric functions with smaller slopes than those produced by the original unscrambled versions. The significance of this effect was confirmed by an analysis of variance, which revealed a significant depth by phase type interaction for both the dots, $F(8,40) = 4.252$, $p < 0.001$, and flagstones, $F(8,40) = 6.070$, $p < 0.001$. It is important to keep in mind that the projected images of surfaces with scrambled and unscrambled textures had the same local amplitude spectra in corresponding image regions. Thus,

these results show clearly that the observers' judgments were influenced by some source of information that is contained within the local phase spectra.

To estimate the reliability of these data, we calculated the standard deviation of adjusted depth among the eight repeated judgments of each observer for each possible combination of surface shape and texture. The results of this analysis are presented in Table 2, which shows the average standard deviation for the scrambled and unscrambled versions of each texture. Note that there were no systematic differences in reliability among the different conditions.

The results presented thus far have focused exclusively on the overall judged depth of each depicted surface, but the observers were also required to indicate the apparent horizontal position of the surface point that was closest in depth (i.e., the near point). Fig. 7 shows the average error in the judged horizontal positions of the near point as a function of judged depth. The left panel in this figure shows the subset of conditions, in which the depicted near point was located in the center of the display. Note that the observers' judgments in those conditions were quite accurate. The right panel shows the remaining conditions in which the depicted near point had an off-center location. The pattern of results in those conditions was more complex, in that the errors were linearly correlated with the overall apparent depth. The apparent near point was shifted toward the center of the display for surfaces that appeared relatively flat, and toward the outer edge of the display for surfaces that had the greatest apparent depth.

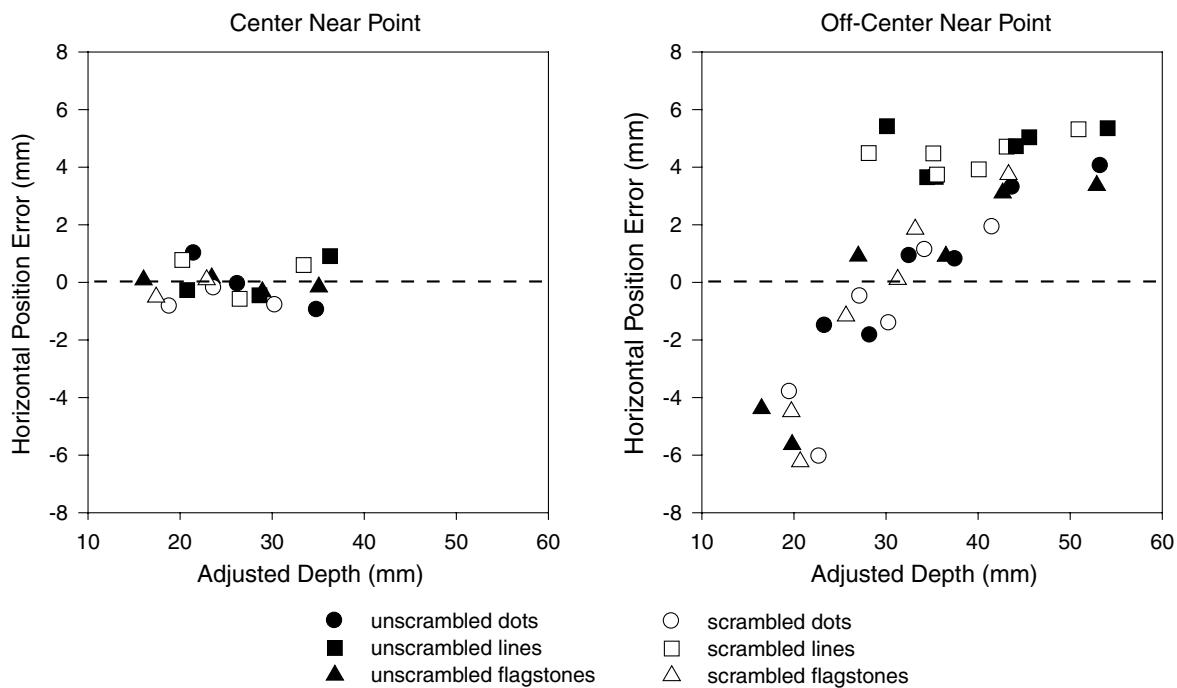


Fig. 7. The average error in the judged horizontal positions of the near point as a function of judged depth for the center near point conditions (left panel) and the off-center near point conditions (right panel) of Experiment 1. Errors toward the center are indicated by negative values, and errors toward the outer edge are indicated by positive values. The dashed line represents veridical performance.

3. Experiment 2

Why should the perception of 3D shape from texture be influenced by the phase spectra of some texture patterns but not those of others? In an effort to shed some more light on this issue, Experiment 2 was designed to investigate the perceptual effects of phase scrambling using a wider variety of textures.

3.1. Methods

The apparatus and general procedure were identical to those used in Experiment 1. The surface shapes in this study were the four from Experiment 1 that had the greatest extension in depth (see Fig. 8), because those were the

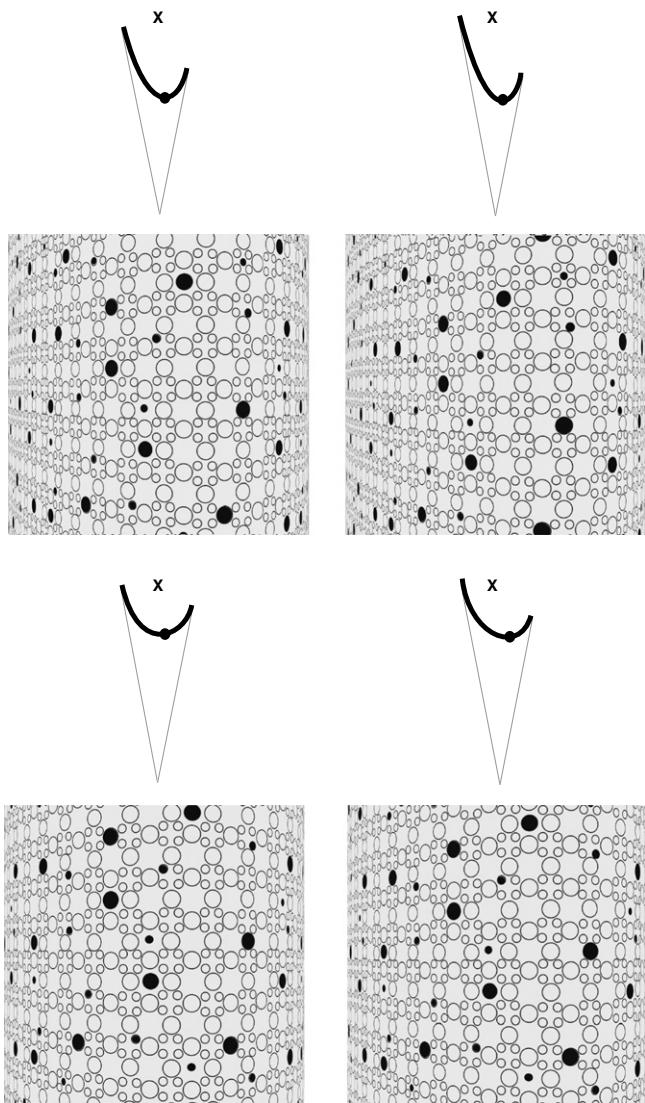


Fig. 8. Some examples of the stimulus images used in Experiment 2. The elliptical arc on top of each image shows a horizontal cross-section of the depicted surface as viewed from above. Thin straight lines depict the 20° viewing cone. For illustration purposes the point on the generating curve nearest to the image plane is denoted with a dot.

ones that produced the largest differences in performance between the scrambled and unscrambled phase spectra. Each of the different surface shapes was rendered with 10 possible textures, which are shown in Fig. 9. The left column of the figure shows the texture patterns with their original Fourier phase spectra, and the middle column shows these same patterns with randomly scrambled phase spectra. From top to bottom the textures are denoted as ‘open circles’, ‘filled squares’, ‘open squares’, ‘diagonal lines’ and ‘wing lattice’, respectively. Their amplitude spectra are shown in the right column.

A regression analysis like the one described for Experiment 1 was performed in order to determine how the local amplitude spectra of these stimuli are affected when the texture on a surface is repositioned or presented with scrambled phase. The results are presented in Table 3, which shows the average r^2 and slope for each texture, as well as their standard deviations. These findings demonstrate that the local amplitude spectra within the optical projection of a textured surface remain highly invariant (within 3%) over changes in the positions or phase spectra of these textures.

Six observers participated in the experiment, including two of the authors (LT and JT), and four others who were naïve about the issues being investigated. Three of these subjects had not participated in Experiment 1, and all had normal or corrected to normal visual acuity.

3.2. Results

Fig. 10a shows average depth settings plotted against the ground truth for all of the different textures. Separate plots for the scrambled and unscrambled versions of each individual texture are presented in Figs. 10b–f. The dashed line in each graph represents veridical performance. An analysis of variance revealed that there were significant effects of phase for the open circles, $F(1,5) = 12.089$, $p < 0.05$, the open squares, $F(1,5) = 31.675$, $p < 0.005$, and the diagonal lines, $F(1,5) = 6.762$, $p < 0.05$, and significant depth by phase interactions for the open squares, $F(3,15) = 7.138$, $p < 0.005$, and diagonal lines, $F(3,15) = 4.843$, $p < 0.05$. The main effects of phase and the depth by phase interactions were not statistically significant for the closed square and wing lattice textures.

In evaluating the perceived depths of Experiments 1 and 2 (see Figs. 3, 6, 9 and 10), it appears to be the case that phase scrambling has the smallest effect for textures composed of parallel lines or texture elements that are organized in a line-like fashion (e.g. the lines and wing grid); it has an intermediate effect for textures composed of non-overlapping solid shapes (e.g. the dots and flagstones); and it has the greatest effect for textures composed of open shapes (e.g. the open squares and open circles). This categorization is not perfect, however, because there was an effect of phase for the diagonal lines textures and an insignificant effect for the filled squares.

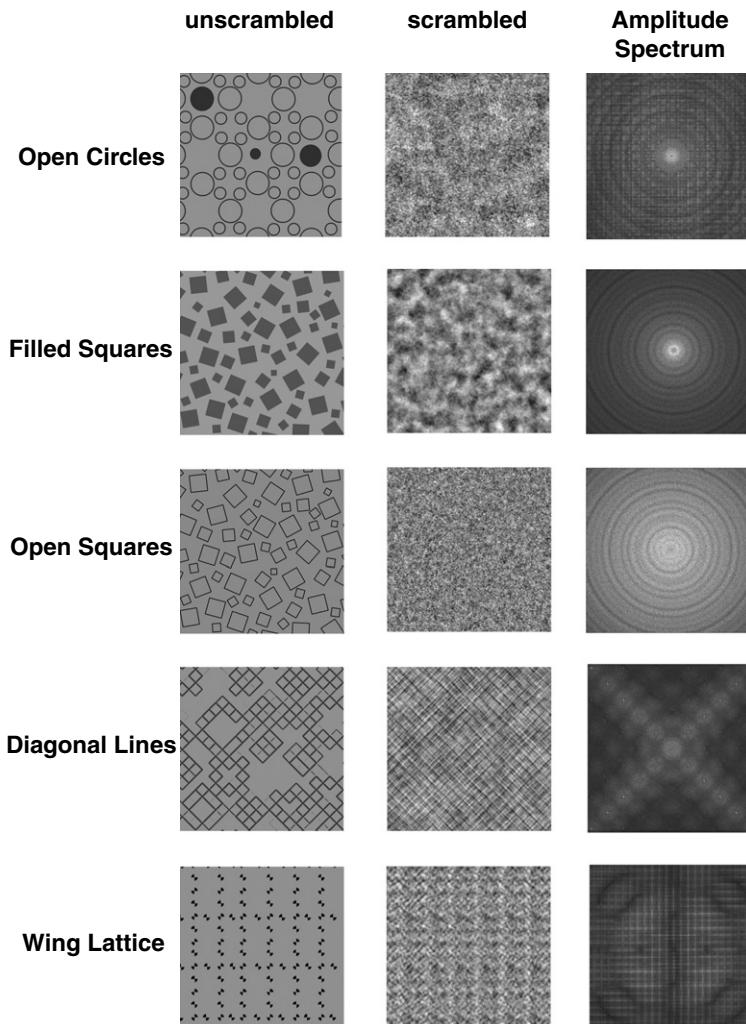


Fig. 9. The textures used in Experiment 2. The left column shows the five original textures before any phase manipulations were performed. The middle column shows the same three textures with randomly scrambled phase spectra, and the right column shows their amplitude spectra.

Table 3

The results of a regression analysis from Experiment 2 that compared the local amplitude spectra in corresponding image regions with scrambled and unscrambled textures, and corresponding regions with unscrambled textures in different surface positions (see text for details)

	Unscrambled vs. scrambled		Unscrambled vs. unscrambled	
	r^2	Slope	r^2	Slope
Open circles	Mean	0.98	0.98	0.98
	Stdev	0.01	0.02	0.04
Filled squares	Mean	0.97	0.98	0.97
	Stdev	0.01	0.03	0.03
Open squares	Mean	0.97	0.98	0.98
	Stdev	0.01	0.01	0.03
Diagonal lines	Mean	0.98	0.98	0.98
	Stdev	0.00	0.02	0.03
Wing lattice	Mean	0.98	0.97	0.99
	Stdev	0.00	0.01	0.02

As in Experiment 1, we measured the reliability of these data by calculating the standard deviation of adjusted depth for each observer in each condition. The results of this analysis are presented in Table 4, which shows the average standard deviation for the scrambled and unscrambled versions of each texture. Note that these results are similar to those obtained in Experiment 1 in that there were no systematic differences in reliability among the different conditions.

Fig. 11 shows the average error in the judged horizontal positions of the near point as a function of judged depth. (Because all of the stimuli in this study had off-center near points, the data are all presented on a single graph). These results confirm that these errors were linearly correlated with the overall apparent depth of a surface. The apparent near point was shifted toward the center of the display for surfaces that appeared relatively flat, and toward the outer edge of the display for surfaces that had the greatest apparent depth. Unlike Experiment 1, there were some systematic differences between the scrambled and unscrambled

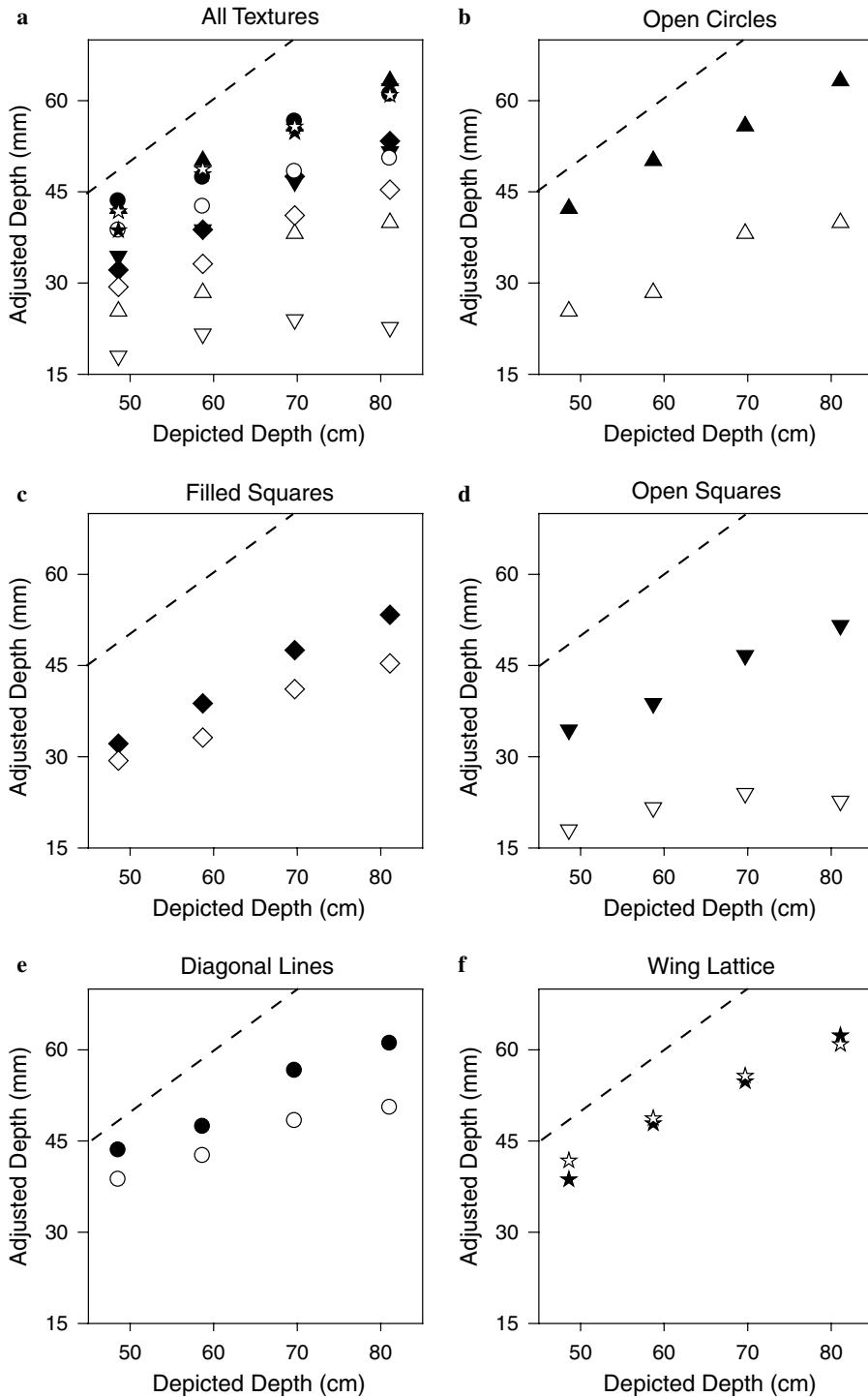


Fig. 10. The average depth settings in Experiment 2 as a function of the depicted stimulus depth for all textures combined (a) and each texture individually (b–f). The scrambled and unscrambled conditions are represented by open and closed symbols, respectively, and the dashed line represents veridical performance. The different units on the horizontal and vertical axes are due to the fact that the adjustment stimulus was reduced in scale to fit on the screen (1 cm in stimulus space = 1 mm in adjustment space).

textures, but this is most likely due to the fact that several of the scrambled displays appeared almost completely flat.

To summarize briefly, the findings from Exp. 2 confirm the results obtained in Experiment 1: The visual perception of 3D shape from texture can be influenced by the phase spectra of some texture patterns but not by those of others.

4. Modeling the results from Experiments 1 and 2

What is the information that determines apparent shapes of these stimuli? The empirical findings from Experiments 1 and 2 demonstrate quite clearly that the local amplitude spectra within the optical projection of a surface

Table 4

Average within-observer standard deviations (in mm) for the depth settings in Experiment 2

	Unscrambled	Scrambled
Open circles	10.29	9.87
Filled squares	10.87	9.97
Open squares	10.69	8.47
Diagonal lines	11.16	11.73
Wing lattice	9.99	9.85

Averages were computed by collapsing over depth conditions and observers.

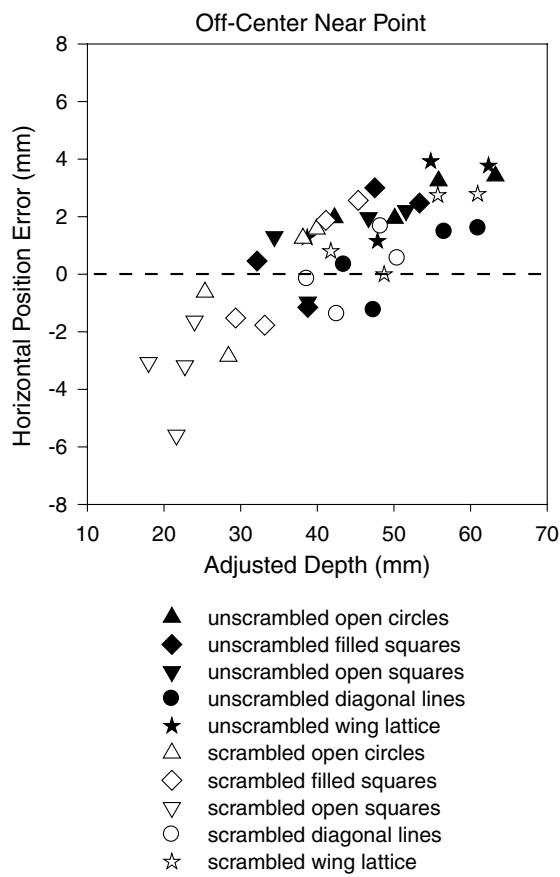


Fig. 11. The average error in the judged horizontal positions of the near point as a function of judged depth in Experiment 2. Errors toward the center are indicated by negative values, and errors toward the outer edge are indicated by positive values. The dashed line represents veridical performance.

cannot be the sole source of information for the perception of 3D shape from texture, and this rules out one popular class of algorithms as potential candidates for modeling human perception (e.g. Bajcsy & Lieberman, 1976; Brown & Shvayster, 1990; Kanatani & Chou, 1989; Krumm & Shafer, 1992; Malik & Rosenholtz, 1994, 1997; Ribeiro & Hancock, 2000; Sakai & Finkel, 1995; Super & Bovik, 1995). An alternative possibility is that observers' perceptions of 3D shape from texture may be based on the distributions of edges within different local neighborhoods of an

image (e.g., Aloimonos, 1988; Blake & Marinos, 1990; Blostein & Ahuja, 1989; Gårding, 1993; Marinos & Blake, 1990; Witkin, 1981). Within the image processing literature, there are two general approaches for the extraction of edges: One that is based on local luminance gradients (e.g. Canny, 1986; Marr & Hildreth, 1980) and another that is based on local Fourier phase congruence (e.g. Kovesi, 1999; Morrone & Burr, 1988) – see Hesse and Georgeson (2005) for an excellent review. Although we have employed both of these approaches in an effort to model the observers' performance in the present experiments, the analyses described below involved a gradient based method of edge detection that was inspired by Hesse and Georgeson (2005), and relied heavily on the MATLAB code in 'gaborconvolve.m' made publicly available by Peter Kovesi at <<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>>.

A bank of six quadrature pairs of logarithmic gabor filters was used to extract the edges in each image. These were all tuned to a wavelength of 3.2 pixels (i.e., 16 cycles/degree in our stimulus images at the experimental viewing distance of 83 cm) and they had six possible orientations of 0°, 30°, 60°, 90°, 120° or 150° relative to vertical. For ease of computation, the convolutions were all performed in the frequency domain and the results were then transformed back into the spatial domain using the inverse Fourier transform. In the frequency domain, the two separable filter components are given by Eqs. (1) and (2),

$$G(\omega) = e^{\frac{-(\log(\omega/\omega_0))^2}{2(\log(\sigma_\omega))^2}} \quad (1)$$

$$G(\theta) = e^{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}} \quad (2)$$

where ω_0 is the center frequency of the filter, θ_0 is its orientation angle, (ω, θ) are the polar coordinates of its position in the amplitude spectrum, and σ_ω and σ_θ are the standard deviations of its Gaussian envelope in the radial and tangential directions. The value of σ_ω was set to 0.65 and the value of σ_θ was set to 0.31.

In the spatial domain, the response of each quadrature pair forms a response vector that is defined by Eq. (3),

$$[e(x, y), o(x, y)] = [I(x, y) * M^e, I(x, y) * M^o] \quad (3)$$

where I is the luminance pattern at a given image location, M^e and M^o are the even-and odd-symmetric filter components, and $e(x, y)$ and $o(x, y)$ are the responses of those components. As suggested by Hesse and Georgeson (2005) based on psychophysical data for edge localization, the final filter output was determined by the largest of the two component responses as defined by Eq. (4).

$$F(x, y) = \max(|e(x, y)|, |o(x, y)|) \quad (4)$$

Using this procedure, edge maps were computed for each of the 94 stimuli that were employed in Experiments 1 and 2. To provide a specific example, Fig. 12 shows a surface with a dot texture from Experiment 1 together with its corresponding edge map. Note that the edges are sparsely

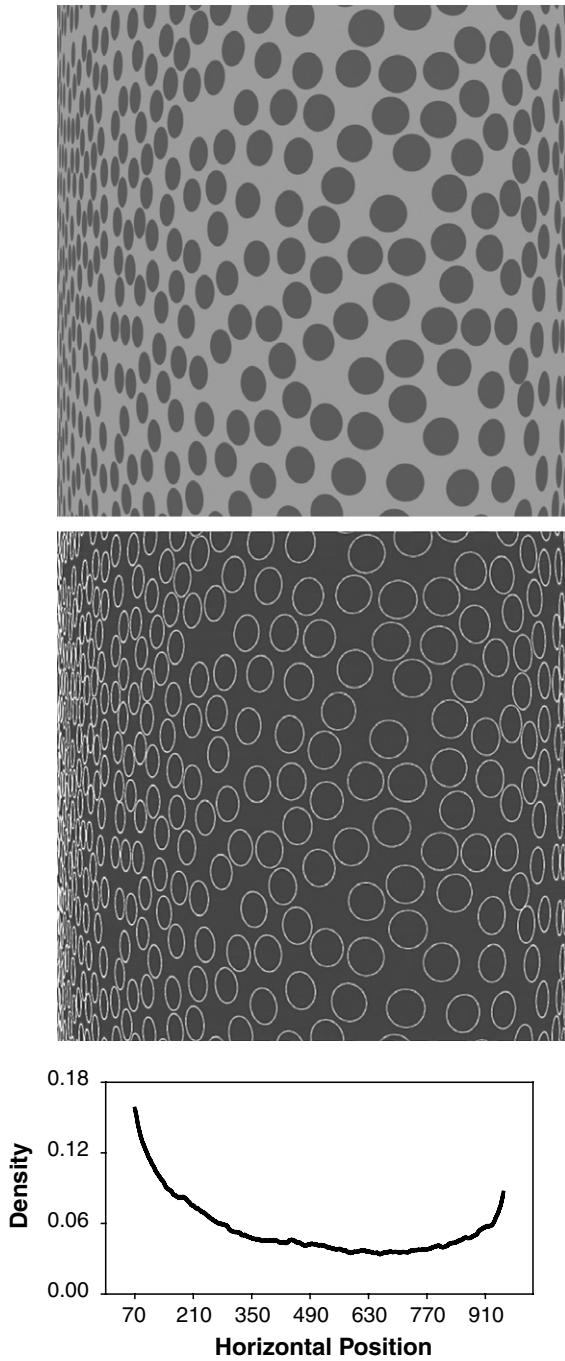


Fig. 12. The computation of image contours. The top panel shows the image of an elliptical cylinder with a polka dot texture. The middle panel shows the edge map that was generated from this image. The lower panel shows how the edge density varies as a function of the horizontal position. Edge density was computed as the average filter output within a sliding window whose dimensions were 1024 (rows) \times 140 (columns).

distributed in the center of this image, and that the density increases moving outward toward the left and right edges due to the effects of perspective and foreshortening. The lower panel of Fig. 12 shows the edge density plotted as a function of horizontal position. Each point on this curve represents the average filter output within a rectangular image region that was 140 pixels wide and 1024 pixels high. In

order to summarize the magnitude of these density changes, it is useful to define a single global measure of density contrast,

$$\text{Density contrast} = \frac{D_{\max} - D_{\min}}{D_{\max} + D_{\min}} \quad (5)$$

where D_{\min} is lowest value of the density function and D_{\max} is the highest. For the stimuli in these experiments, D_{\min} and D_{\max} were located at the projected horizontal positions of the surface regions that were closest and farthest in depth, respectively, relative to the point of observation.

Because Todd et al. (2005) have shown that the apparent surface relief for textured dihedral angles is highly correlated with global changes in texture density, we wondered whether a similar measure could account for the results of the present experiments with textured elliptical cylinders. Figs. 13 and 14 show how the density contrast in an image varies as a function of depicted surface relief for all of the textures that were employed in Experiments 1 and 2. Note that these graphs are qualitatively similar to the psychometric functions presented in Figs. 6 and 10. That is to say, the density contrast increases linearly with the depicted surface depth, and it is significantly attenuated for most textures when their phase spectra are randomly scrambled.⁴

The left panel of Fig. 15 shows the mean adjusted depth as a function of density contrast for all of the 94 possible combinations of surface shape and texture that were used in the present experiments. A regression analysis of these data revealed that density contrast accounts for 61% of the variance in the observers judgments. However, upon closer inspection of this figure it is clear that the residuals of the regression are not distributed randomly about the best fitting linear function. Rather, the data seem to cluster into three different groups that individually have high linear correlations, but that differ from one another with respect to their slopes and intercepts. In particular, the scrambled versions of the dots, flagstones, filled squares and wing lattice and the unscrambled versions of the open squares and circles produced more apparent depth relative to the other textures than is predicted by pure density contrast, and the scrambled and unscrambled line textures produced less apparent depth. It is possible to eliminate these differences by applying a multiplicative scaling to the density contrasts within each of the outlier groups, as shown in the right panel Fig. 15. Thus, with the addition of these two free parameters, the model can account for 89% of the variance among the 94 possible experimental conditions.

⁴ It is interesting to note that the edge density measure described in this paper is positively correlated with local spatial frequency. However, there is an important difference between these measures that deserves to be highlighted. Edges were extracted at high frequencies using a non-linearity, i.e. the max-operator (Hesse & Georgeson, 2005) or the local energy operator (Morrone & Burr, 1988), which makes the measure sensitive to phase. Because local spatial frequency is independent of phase, it cannot account for differences between the scrambled and unscrambled textures that were obtained in the present experiments.

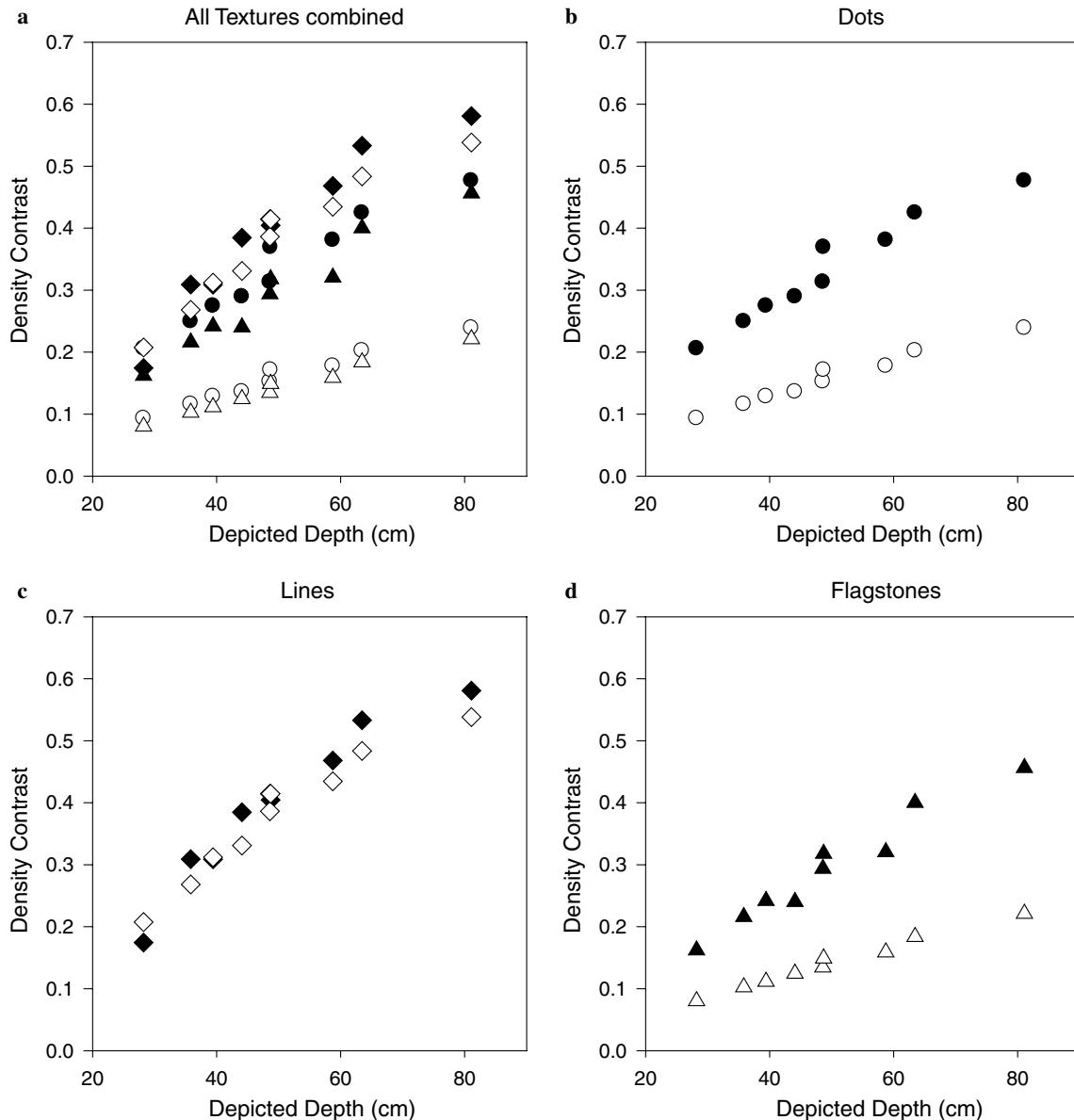


Fig. 13. The density contrast measure for the stimuli in Experiment 1 as a function of the ground truth for all textures combined (a) and each texture individually (b–d). The scrambled and unscrambled conditions are represented by open and closed symbols, respectively.

One possible reason why the density contrast measure does not fully capture the relative perceptual gain for the different types of texture is that the human visual system may employ some sort of edge enhancement mechanism. The method of edge detection used in the present model is sensitive to local contrast of an edge, which is generally reduced for most textures when their phase spectra are randomly scrambled. One way of eliminating the effects of local contrast is to threshold the filter outputs to produce a binary response. We have tried that with several different threshold settings, but it does not significantly improve the overall goodness of fit for the density contrast measure. More sophisticated methods of edge enhancement involve competitive and cooperative interactions among neighboring filters that

are capable of filling in edges in image regions where no local luminance gradients are present (e.g., Grossberg & Mingolla, 1985, 1987). This type of mechanism would undoubtedly alter the overall density contrast in an image, but it remains to be seen whether that can explain the different perceptual gains that are shown in the left panel of Fig. 15 (see also Todd & Akerstrom, 1987).

5. Conclusions

The research described in the present article was designed to compare two different types of algorithm as potential candidates for modeling the perception of 3D shape from texture in human observers. One popular

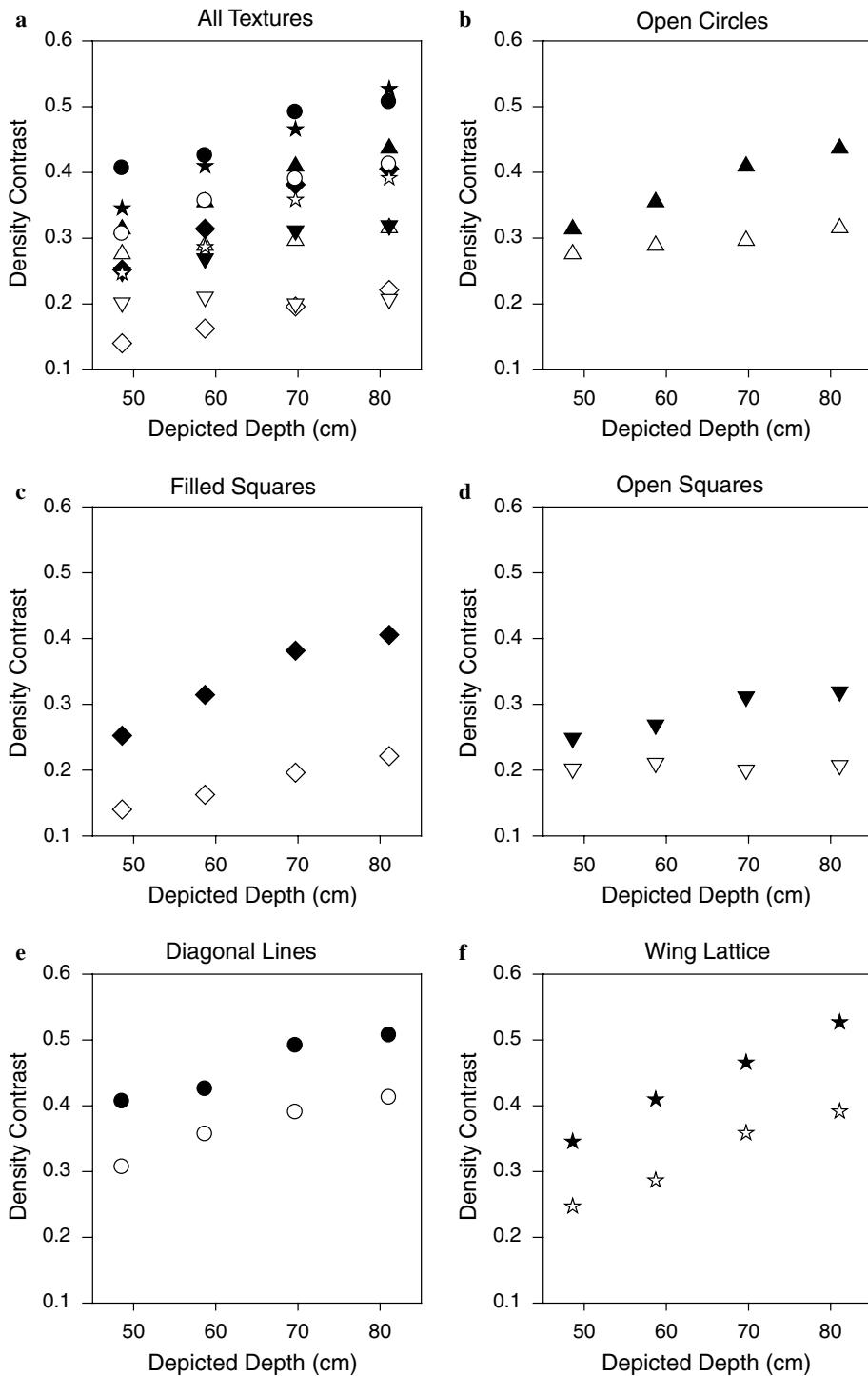


Fig. 14. The density contrast measure for the stimuli in Experiment 2 as a function of the depicted stimulus depth for all textures combined (a) and each texture individually (b–f). The scrambled and unscrambled conditions are represented by open and closed symbols, respectively.

approach to this problem is to first perform a Fourier transform on various local neighborhoods of an image, and then estimate the surface geometry depicted in each region from properties of its local amplitude spectrum (Bajcsy & Lieberman, 1976; Brown & Shvayster, 1990; Kanatani & Chou, 1989; Krumm & Shafer, 1992; Ribeiro & Hancock, 2000; Sakai & Finkel, 1995; Super &

Bovik, 1995), or from systematic changes between amplitude spectra in neighboring regions (Malik & Rosenholtz, 1994, 1997). A fundamental characteristic of this class of models is that performance is unaffected by the structure of local phase spectra in an image. Thus, if this type of mechanism were employed for the visual perception of 3D shape from texture, then manipulations of the

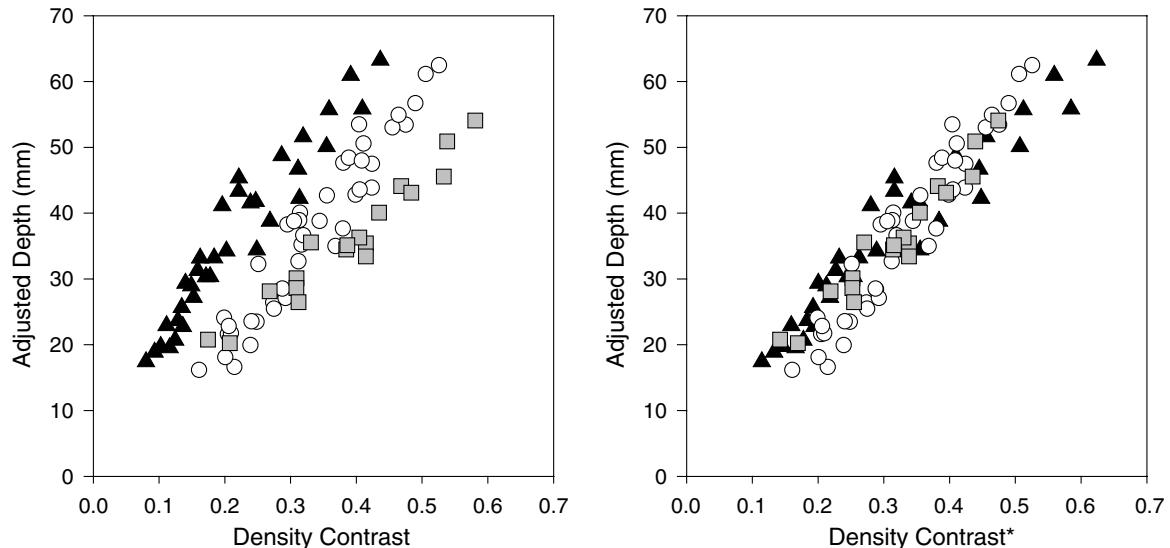


Fig. 15. The average depth settings in Experiments 1 and 2 as a function of density contrast (left panel). The black triangles (group 1) represent all stimuli with scrambled versions of the dots, flagstones, filled squares and the wing lattice and the unscrambled versions of the open squares and circles. The white circles (group 2) represent all stimuli with scrambled versions of the open squares, open circles and diagonal lines, and the unscrambled versions of the dots, flagstones, filled squares, diagonal lines and the wing lattice. The gray squares (group 3) represent all stimuli with scrambled or unscrambled lines. The right panel shows the same data with a multiplicative scaling applied to the density contrasts of group 1 and 3 to minimize the differences between the three groups.

phase spectra should have a negligible influence on the apparent 3D shape of a surface. The results of the present experiments show clearly that this is not the case. For most of the textures employed in these studies random rearrangements of the local phase spectra caused significant reductions in the overall magnitude of perceived relief.

An alternative technique for analyzing 3D shape from texture is to extract the edges within various local neighborhoods of an image, and then estimate the surface geometry depicted in each region from the distribution of edge orientations (Aloimonos, 1988; Blake & Marinos, 1990; Blostein & Ahuja, 1989; Marinos & Blake, 1990; Witkin, 1981), or from systematic changes in the distributions of edges across neighboring regions (Gårding, 1993). The model proposed in the present article is a variant of this general approach. The central assumption of this model is that changes in depth are optically specified by systematic variations in edge density (see also Grossberg & Mingolla, 1987; Todd & Akerstrom, 1987). Based on that assumption, we introduced a new measure called density contrast to estimate the overall magnitude of perceived surface relief, and we implemented a filter based computational procedure for measuring the density contrast in each of the 94 images used in our experiments. Much like observers' shape judgments, the density contrast measure increases linearly with the depicted depth of a surface, and it is significantly attenuated when the phase spectrum of a texture is randomly scrambled. However, in its current implementation, the model cannot fully explain the variations in perceptual gain that were observed among different groups of textures.

Acknowledgments

We thank Delwin Lindsey and Martin Nikolov for their helpful suggestions during the course of this research. This research was supported by a grant from NSF (BCS-0546107).

Appendix A

Here we give details of the geometry of the elliptical cylinders. As described in the introduction, the geometrical constraints on the elliptical cross-section of each surface were motivated by the desire to have as big a texture gradient as possible for a given field of view. This means that the two lines that delimit the field of view are tangent to the ellipse. Put another way, the occluding contour of the ellipse fits snugly in the field of view. A second constraint is that it should be intuitive for participants to manipulate the test ellipse (see Fig. 5, right panel, in the main text). Intuition is best served when the participant is able to manipulate a point on the ellipse. For mathematical convenience, we took as this point the location where the tangent vector to the ellipse is horizontal (equivalent to the tangent plane being fronto-parallel for the elliptical cylinder). As final constraint we fixed the depth of the center of the ellipse, a choice that is mathematically convenient.

Summarizing the constraints and introducing the notation we have as input to the calculation:

1. Field of view ϕ .
2. Location of tangent (x_t, y_t) .
3. Depth of center y_0 .

The output is an ellipse parametrized with scale factor λ , aspect ratio r , orientation of long axis α and center (x_0, y_0) . The mathematical problem is to recover the values of x_0 , α , λ and r (denoted black in Fig. A1) given ϕ , x_t , y_t and y_0 (denoted gray in Fig. A1).

As a first step in the solution we exploit the field of view constraint. To this end we express both the viewing cone with field of view ϕ and the tangent cone to the ellipse as an implicit equation in homogeneous coordinates. The viewing cone is given by:

$$x^2 - \tan^2(\phi/2)y^2 = 0 \quad (\text{A1})$$

which is equivalent to a pair of straight lines with slopes $\tan\phi/2$ and $-\tan\phi/2$.

The tangent cone to the ellipse is a bit more complicated to derive. We follow the same path as Blinn (IEEE Computer Graphics and Applications, Jan 1995, pp. 78–83). We express the ellipse in homogeneous coordinates as:

$$p^t Q p = 0$$

with

$$p = (x, y, 1)^t$$

$$Q = \Phi^t \Lambda \Phi$$

and

$$\Phi = \begin{pmatrix} \cos \alpha & \sin \alpha & -\cos \alpha x_0 - \sin \alpha y_0 \\ -\sin \alpha & \cos \alpha & \sin \alpha x_0 - \cos \alpha y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} (r\lambda)^{-2} & 0 & 0 \\ 0 & \lambda^{-2} & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

The tangent cone from vantage point $O = (0, 0, 1)$ equals (Blinn p. 82):

$$p^t (Q O^t O Q - O Q O^t Q) p = 0$$

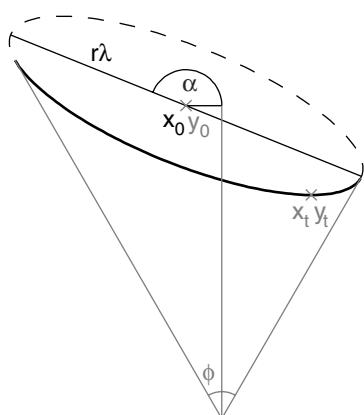


Fig. A1. Examples of the notation used in the derivation of the ellipse geometry. Symbols in gray denote the input to the derivation, while those in black denote the output.

which leads to:

$$(\lambda^2(\cos^2 \alpha + r^2 \sin^2 \alpha) - y_0^2)x^2 + (\lambda^2(1 - r^2) \sin 2\alpha + 2x_0y_0)xy + (\lambda^2(r^2 \cos^2 \alpha + \sin^2 \alpha) - x_0^2)y^2 = 0 \quad (\text{A2})$$

Comparison of the viewing cone (A1) with the tangent cone (A2) leads to the following two constraint equations:

$$\begin{aligned} \lambda^2(1 - r^2) \sin 2\alpha + 2x_0y_0 &= 0 \\ \frac{\lambda^2(r^2 \cos^2 \alpha + \sin^2 \alpha) - x_0^2}{\lambda^2(\cos^2 \alpha + r^2 \sin^2 \alpha) - y_0^2} &= \frac{-\tan^2(\phi/2)}{1} \end{aligned} \quad (\text{A3})$$

We now use the tangent constraint. Straightforward algebra leads to:

$$\left(\begin{array}{c} x_t - x_0 \\ y_t - y_0 \end{array} \right) = \frac{\lambda}{\sqrt{\cos^2 \alpha + r^2 \sin^2 \alpha}} \left(\begin{array}{c} 1/2(r^2 - 1) \sin 2\alpha \\ \cos^2 \alpha + r^2 \sin^2 \alpha \end{array} \right) \quad (\text{A4})$$

Multiplying the two equations, we get:

$$(x_t - x_0)(y_t - y_0) = \lambda^2(1/2(r^2 - 1) \sin 2\alpha) = -x_0y_0$$

where we used the top equation of (A3). Simplifying:

$$x_0 = \frac{x_t(y_0 - y_t)}{y_t}$$

Next, we solve for α by squaring the second equation of (A4) and substitution in the second equation of (A3):

$$\lambda^2(r^2 \cos^2 \alpha + \sin^2 \alpha) = \tan^2(\phi/2)(y_0^2 - (y_t - y_0)^2) + x_0^2 \quad (\text{A5})$$

Squaring the second equation of (B4) and subtracting it to the result above leads to:

$$\begin{aligned} \lambda^2(r^2 - 1) \cos 2\alpha &= \underbrace{\tan^2(\phi/2)(y_0^2 - (y_t - y_0)^2) + x_0^2}_{A} - (y_t - y_0)^2 \\ &= A - (y_t - y_0)^2 \end{aligned} \quad (\text{A6})$$

Combining this result with the first equation of (A3) leads to:

$$\tan 2\alpha = \frac{2x_0y_0}{A}$$

Adding the square of the first equation in (A3) with the square of (A6) leads to:

$$(\lambda^2(r^2 - 1))^2 = A^2 + (2x_0y_0)^2$$

Squaring the second equation of (A4) and adding it to (A5) leads to:

$$\lambda^2(r^2 + 1) = \underbrace{\tan^2(\phi/2)(y_0^2 - (y_t - y_0)^2) + x_0^2}_{B} + (y_t - y_0)^2$$

Combining the last two equations, we obtain:

$$\lambda = \sqrt{\frac{B + \sqrt{A^2 + (2x_0y_0)^2}}{2}} r = \frac{B}{\lambda^2} - 1 \quad (\text{A7})$$

The four boxed equations represent the solutions.

References

- Aloimonos, J. (1988). Shape from texture. *Biological Cybernetics*, 58, 345–360.
- Bajcsy, R., & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics Image Processing*, 5, 52–67.
- Blake, A., Buelthoff, H. H., & Sheinberg, D. (1993). Shape from texture: ideal observers and human psychophysics. *Vision Research*, 33, 1723–1737.
- Blake, A., & Marinos, C. (1990). Shape from texture: estimation, isotropy and moments. *Artificial Intelligence*, 45, 323–380.
- Blinn, J. F. (1995). How to draw a sphere. *IEEE Computer Graphics and Applications*, 15, 78–83.
- Blostein, D., & Ahuja, N. (1989). Shape from texture: integrating texture-element extraction and surface estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 1233–1251.
- Braunstein, M. L., & Payne, J. W. (1969). Perspective and form ratio as determinants of relative slant judgments. *Journal of Experimental Psychology*, 81, 584–590.
- Brodatz, P. (1966). *Texture: A photographic album for artists and designers*. New York: Dover.
- Brown, L. G., & Shvayster, H. (1990). Surface orientation from projective foreshortening of isotropic texture autocorrelation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 584–588.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 679–698.
- Gårding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, 2, 327–350.
- Gårding, J. (1993). Shape from texture and contour by weak isotropy. *Artificial Intelligence*, 64, 243–297.
- Gibson, J. J. (1950a). The perception of visual surfaces. *American Journal of Psychology*, 63, 367–384.
- Gibson, J. J. (1950b). *The perception of the visual world*. Boston: Houghton Mifflin.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries and emergent segmentations. *Perception & Psychophysics*, 38, 141–171.
- Grossberg, S., & Mingolla, E. (1987). Neural dynamics of surface perception: boundary webs, illuminants, and shape from shading. *Computer Vision, Graphics, and Image Processing*, 37, 116–165.
- Hesse, G. S., & Georgeson, M. A. (2005). Edges and bars: where do people see features in 1-D images? *Vision Research*, 45, 507–525.
- Kanatani, K., & Chou, T. C. (1989). Shape from texture: general principle. *Artificial Intelligence*, 38, 1–48.
- Knill, D. (1998). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research*, 38, 1655–1682.
- Knill, D. (2001). Contour into texture: information content of surface contours and texture flow. *Journal of the Optical Society of America A*, 18, 12–35.
- Kovesi, P. (1999). Image features from phase congruency. *Videre: Journal for Computer Vision Research*, 1(3), 1–28.
- Krumm, J., & Shafer, S. A. (1992). Shape from periodic texture using the spectrogram. *Proceedings of the 1992 IEEE computer society conference on computer vision and pattern recognition (CVPR '92)*, 284–289.
- Li, A., & Zaidi, Q. (2001). Verdicality of three-dimensional shape perception predicted from amplitude spectra of natural textures. *Journal of the Optical Society of America A*, 18, 2430–2447.
- Malik, J., & Rosenholtz, R. (1994). A computational model for shape from texture. In G. R. Bock & J. A. Goode (Eds.), *Higher order processing in the visual system* (pp. 272–286). Chichester: John Wiley & Sons.
- Malik, J., & Rosenholtz, R. (1997). Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 23, 149–168.
- Marinos, C., & Blake, A. (1990). Shape from texture: the homogeneity hypothesis. *Proceedings of the 3rd international conference on computer vision*. Osaka, Japan, 350–353.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society B*, 207, 187–217.
- Morrone, C. M., & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London. Series B, biological sciences*, 235, 221–245.
- Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69, 529–541.
- Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, 11, 337–346.
- Ribeiro, E., & Hancock, E. R. (2000). Estimating the 3D orientation of texture planes using local spectral analysis. *Image and Vision Computing*, 18, 619–631.
- Sakai, K., & Finkel, L. H. (1995). Characterization of the spatial frequency spectrum in the perception of shape from texture. *Journal of the Optical Society of America A*, 12, 1208–1224.
- Stevens, K. A. (1881). The information content of texture gradients. *Biological Cybernetics*, 42, 95–105.
- Super, B. J., & Bovik, A. C. (1995). Planar surface orientation from texture spatial frequencies. *Pattern Recognition*, 28(5), 729–743.
- Thomson, M. G. A. (1999). Visual coding and the phase structure of natural scenes. *Network*, 10, 123–132.
- Thomson, M. G. A., Foster, D. H., & Summers, R. J. (2000). Human sensitivity to phase perturbations in natural images: a statistical framework. *Perception*, 29, 1057–1069.
- Todd, J. T., & Akerstrom, R. A. (1987). The perception of three dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 242–255.
- Todd, J. T., Thaler, L., & Dijkstra, T. M. H. (2005). The effects of field of view on the perception of 3D slant from texture. *Vision Research*, 45, 1501–1517.
- Yellott, J. I. (1993). Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture. *Journal of the Optical Society of America A*, 10, 777–793.
- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17, 17–45.