



# Perception of three-dimensional shape from texture is based on patterns of oriented energy

Andrea Li \*, Qasim Zaidi

*State University of New York, State College of Optometry, 100 E 24th St., New York, NY 10010, USA*

Received 18 September 1998; received in revised form 12 February 1999

## Abstract

This paper presents empirical support for a new observer model of inferring three-dimensional shape from monocular texture cues. By measuring observers' abilities to estimate the relative three-dimensional curvature along a textured surface from two-dimensional projected images, and concurrently examining the local spectral changes occurring in the projected image for various texture patterns, we have found that correlated changes in oriented energy along lines corresponding to the lines of maximum and minimum curvature of the surface are crucial for conveying the three-dimensional shape of the surface. Energy along these lines of maximum and minimum curvature can be used to compute the orientation of local surface patches. Texture patterns consisting of simple and complex sinusoidal gratings and plaids, and filtered noise were drawn onto a surface that was corrugated sinusoidally in depth about the horizontal axis and projected in perspective onto an image plane. The perceived relative surface curvature was reconstructed from measurements of local ordinal depth around a central fixation point at 12 different phases of the corrugation. Our results show that: (1) it is neither necessary nor sufficient to identify individual texture elements or texture gradients in order to extract the shape of the surface; (2) one-dimensional frequency modulation is insufficient for conveying complex three-dimensional shape. (3) Veridical ordinal depth is seen only when the projected pattern contains changes in oriented energy along lines corresponding to projected lines of maximum curvature of the surface. (4) For a surface corrugated in depth about the horizontal axis, this pattern of oriented energy arises from energy along the vertical direction in the global Fourier transform of the pre-corrugated pattern. (5) Local orientation changes across lines of minimum curvature can be also critical for conveying shape. (6) These correlated orientation changes along lines of maximum and minimum curvature are entirely lost in parallel projection. Hence texture is a useful cue for shape if the image is a perspective projection. (7) Only some natural textures will provide sufficient monocular cues to support veridical shape inferences, and this can be predicted from their global Fourier transforms. © 2000 Elsevier Science Ltd. All rights reserved.

*Keywords:* Shape from texture; Perspective; Contour; Curvature; Oriented energy

## 1. Introduction

We present support for a new observer model for inferring three-dimensional shape from monocular texture cues. In this model the observer infers the three-dimensional orientation of local patches of the test surface from the projections of the surface markings. The local orientation of the surface, or equivalently the local normal, is derived from the projections of the local axes of maximum and minimum curvature of the surface. When these axes are physically present as

surface markings, in perspective projection, they form systematic local changes in orientation. These orientation changes form visible contours across the image and the corresponding changes in oriented spatial energy can be measured in local spectral analyses. The pattern of oriented energy can be visible as contours that are solid and continuous, or as contours that are discontinuous and noisy. Hence three-dimensional shape inference does not require the identification of individual texture elements and gradients such as those used in previous shape-from-texture studies (Cutting & Millard, 1984; Todd & Akerstrom, 1987; Cummings, Johnston & Parker, 1993; Knill, 1998a,b,c). Support for the theory is provided by empirical measurements showing

\* Corresponding author. Fax: +1-212-780-5009.  
E-mail address: ali@sunyopt.edu (A. Li)

that shape inference is veridical only when the requisite pattern of oriented energy changes is present and visible across the projected image of the textured surface.

Gibson (1950) described changes in projected images of textured surfaces phenomenologically as texture gradients. These gradients provide potential cues for the three-dimensional form of the surface. Studies examining how changes in individual texture element features affect percepts of depth and shape have shown that changes in compression (i.e. the ratio of  $y/x$  extent of the texture elements) are most effective at conveying surface curvature, while changes in perspective are more effective for conveying depth and distance along flat surfaces (Cutting & Millard, 1984; Todd & Akerstrom, 1987; Cummings et al., 1993; Knill, 1998c). These studies generally assume that the visual system is able to isolate individual elements before element-specific gradients can be used to extract shape. This assumption can not be generally made in the real world where many natural textures (e.g. tree bark, grass) cannot be easily segmented into individual texture elements. The results in our study show that the identification of texture elements or gradients is not necessary to extract shape from texture. In addition, our results show that adding a filter-based element extraction front-end (e.g. Voorhees & Poggio, 1988) to gradient models still does not lead to an observer model that can distinguish between patterns that convey veridical three-dimensional shape and those that do not.

Other studies have suggested that the visual system can use changes in spatial frequency across the image in lieu of texture element gradients as a cue to surface shape (Bajcsy & Lieberman, 1976; Turner, Gerstein & Bajcsy, 1991; Sakai & Finkel, 1993). Spectral approaches have in fact been implemented in models used to extract shape from texture in natural texture patterns which are not easily segmented into individual texture elements (Malik & Rosenholtz, 1994; Krumm & Shafer, 1994). However, stimuli in spectrally based studies have generally been limited to surfaces that are flat (Turner et al., 1991) or convex and cylindrical under parallel projection (Sakai & Finkel, 1993), whereas natural objects often contain both convex and concave curvatures, sometimes even at the same point, for example in saddles. The results in our study show that one-dimensional changes in spatial frequency are not sufficient for conveying the veridical shape of surfaces containing multiple curvatures even when viewed under perspective projection.

The pictures in Fig. 1 all consist of gray-level patterns on flat surfaces. However, observers tend to see all three as corrugated in depth. To explain this percept, the observer model in this paper makes the same fundamental assumption as all other observer models of shape-from-texture, namely that changes in two-dimensional patterns on flat surfaces do not generally

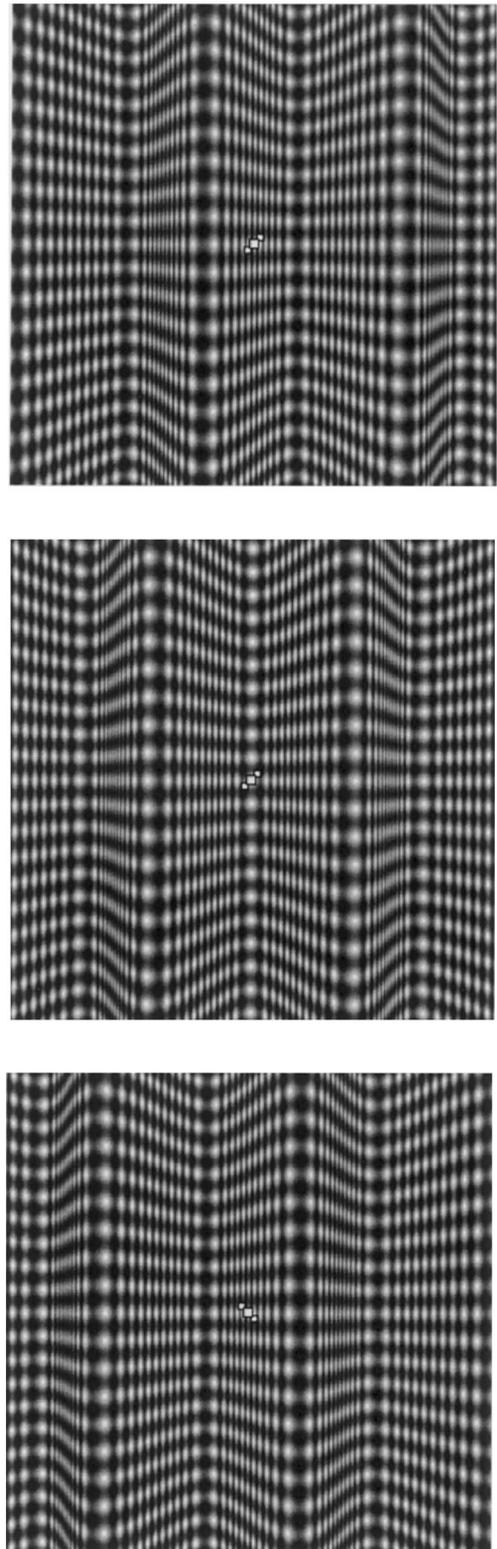


Fig. 1. Examples of a projected horizontal-vertical plaid at three different phases:  $-\pi/2$  (top), 0 (middle) and  $+\pi/2$  (bottom). Observers fixated the central fixation and indicated which of the two flanking test dots appeared closer to them in depth. Test dots were  $0.2^\circ$  away from the fixation and could lie along the  $+45^\circ$  diagonal (top and middle examples) or the  $-45^\circ$  diagonal (bottom example).

mimic changes in perspective, and therefore changes in the pattern should be attributed as much as possible to changes in the surface shape (Witkin, 1981). The proposed model further assumes that observers exploit perspective geometry. In the domain of shape outlines, there is strong evidence that the visual system interprets retinal projections as perspective projections of three-dimensional objects (Ames, 1951; Ittelson, 1952; Shepard, 1990; Griffiths & Zaidi, 1999). The model in this paper assumes that the visual system interprets retinal projections of surface markings as perspective projections of surface markings on three-dimensional objects (see discussion of Fig. 17).

Perspective projections of surface textures have generally been parsed into components like density and foreshortening, and ideal observer and generic models have been constructed for each of these changes as cues to surface shape (Blake, Bulthoff & Sheinberg, 1993; Knill, 1998a,b,c). Empirical evidence presented in this paper suggests that the connection between surface texture and three-dimensional shape may be more direct. For corrugated shapes that possess concavities and convexities, observers perceive three-dimensional shape veridically when the local frequency spectrum across the image contains a pattern of changes in oriented energy that are compatible with the projection of the local axes of maximum curvature. The model thus postulates that the outputs of orientation selective cells in the visual cortex are interpreted in terms of perspective projection, and obviates the need for recognizing texture elements or gradients.

Stevens (1981) showed that local surface orientation can be computed from assumptions about the relationship between the geometry of the surface and the surface markings. For any smooth surface, the Gaussian curvature at any point on the surface is the product of the maximum and minimum curvatures at that point. These are called the principal directions at that point and are always perpendicular to one another (Gray, 1998). Vectors in the projected image that are locally perpendicular to each of these two directions, i.e. the cross product of the two principal directions considered as vectors from the point of intersection, provide the range of possible angles of the surface normal at that point. The greater the curvature of the surface at that point, the greater the foreshortening of the angle between the principal directions and the more the angle will deviate from  $90^\circ$ . The range of possible normals will therefore be most restricted for angles that are most foreshortened, i.e. for locations on the surface that are most curved. At these locations, Stevens used the bisection of the range as an unbiased estimate of the normal. The normal at restricted locations can then be used to estimate the normal at locations with less constrained ranges. We borrow from Stevens the result that if at any point in the image of the surface the

projections of the axes of maximum and minimum curvature are given, then the normal to the surface can be estimated fairly well (Mamassian & Landy, 1998). The model however generalizes Stevens' work in two ways. First, the model uses perspective information as opposed to the orthographic projections considered by Stevens, and second, the axes of maximum and minimum curvature are inferred from a space/spatial-frequency analysis and thus apply to the general cue of surface markings or textures, and not just to parallel rulings.

In our study, simple grating and plaid patterns, complex plaids containing energy at multiple orientations and frequencies, and filtered noise patterns were corrugated in depth and projected in perspective onto an image plane. Spectral analyses were performed both locally across the projected images, and globally on texture patterns before corrugation and projection to examine the types of spectral energy that give rise to the local spectral changes in the projected image. Space/spatial-frequency analysis has previously been applied to texture segmentation (e.g. Reed & Wechsler, 1990, 1991) but not to inferring shape from texture.

We used a new method to measure observers' perceptions of three-dimensional shape from projected images. Measuring the exact perceived shape of a surface is a difficult task. Shape-from-texture studies have generally used global shape estimation tasks in which observers are asked to rate or match the magnitude of the slant or curvature of a surface (Vickers, 1971; Cutting & Millard, 1984; Todd & Akerstrom, 1987; Cummings et al., 1993; Sakai & Finkel, 1993). This kind of task requires comparisons with templates of either internal (memory) or external (template matching) nature, and can subsequently influence the observer's response by limiting the number of possible responses. In our study, observers made a series of local depth judgments in which they indicated which of two locations on the surface around a central fixation appeared closer to them in depth. We reconstructed the global shape of the surface from these local relative depth judgments. The test locations were spaced close together along the surface and were randomly interleaved across trials with a limited presentation time. Observers were therefore forced to make a local relative depth judgment as opposed to a global judgment of the shape of the entire surface.

## 2. Methods

### 2.1. Image generation

High contrast patterns were drawn onto a surface that was corrugated in depth ( $z$ ) as a sinusoidal function of the horizontal axis ( $x$ ). The textured surface was

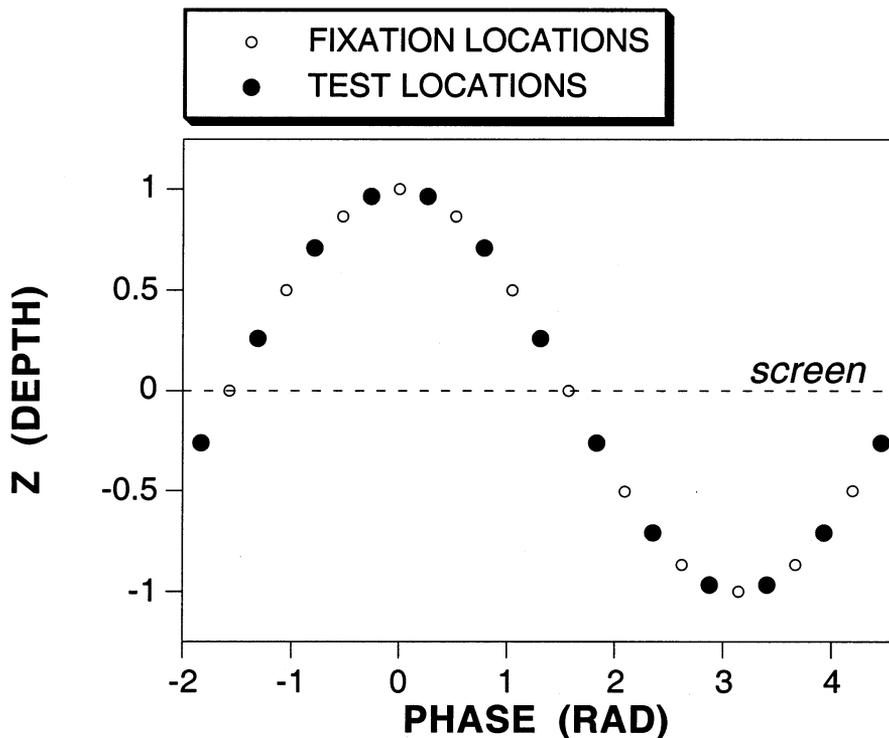


Fig. 2. Fixation and test locations along a single cycle of the corrugation. Depth along the surface ( $z$ ) is plotted as a function of phase in radians. Depths beyond the screen are positive and depths in front of the screen negative. The central fixation was set at one of 12 different phases (open symbols). Each image was projected in perspective with respect to that phase. There were 13 possible test locations (filled symbols), the first and last of which fell at the same phase along the corrugation.

then projected in perspective onto an image plane (Appendix A). Each image was  $381 \times 381$  pixels and subtended  $8.8 \times 8.8^\circ$  of visual angle. The center of the image was defined as the origin and the depth along the surface was calculated as:

$$z = \cos(2\pi fx + \varphi) \quad (1)$$

where  $\varphi$  is the phase of the corrugation at the center of the image. Each image was projected in perspective relative to the phase at the central fixation point. Phases of  $-\pi/2$ , 0, and  $+\pi/2$  are shown from top to bottom in Fig. 1. At the viewing distance of 1 m the corrugation was computed to span 19 cm in depth from peak to trough. Each projected image contained three full cycles of the depth corrugation.

Projections were computed using Matlab and presented on a BARCO 7651 color monitor with a  $736 \times 550$  pixel screen running with refresh rate of 100 frames/s. Images were presented using a Cambridge Research Systems Video Stimulus Generator (CRS VSG2/3). Through the use of 12-bit DACs, after gamma correction, the VSG2/3 is able to generate 2861 linear levels for each gun. The mean luminance of the screen was  $25 \text{ cd/m}^2$ .

## 2.2. Psychophysical method

We measured perceived three-dimensional shape using local ordinal depth judgments. Each projected image contained a central square fixation flanked by two smaller test dots positioned to the left and right of the fixation along a  $-45^\circ$  or  $+45^\circ$  diagonal. The test dots were  $0.2^\circ$  visual angle away from the fixation and abutted it at its diagonal corners. Observers were asked to fixate the central fixation and indicate which of the two locations on the surface as indicated by the two test dots appeared closer to them in depth on the surface, or if they appeared at about equal depths. In the top and middle examples of Fig. 1, the test dots lie along the  $+45^\circ$  diagonal, in the bottom example, they lie along the  $-45^\circ$  diagonal. In the top example the left test location appears closer in depth than the right, in the middle example the two locations appear at about equal depths and in the bottom example the right location appears closer.

Observers used a mouse to make their responses; the left and right mouse buttons corresponded to the left and right test locations, and the middle button was pressed if the two appeared at the same depth. To

sample both convex and concave portions of the surface, the central fixation location was set at one of 12 different phases along a single cycle of the corrugation. Each projection was computed in perspective relative to the phase of the central fixation. The 12 fixation locations are indicated by the open symbols in Fig. 2. Trials using the 12 different fixation locations were randomly interleaved. For each fixation location a comparison was made between the adjacent test locations. The 13 possible test locations are shown in Fig. 2 as solid symbols. The first and last test locations fell at the same phase of the corrugation.

One texture pattern was used for each session. A single session consisted of two periods separated by a brief pause: in the first period we presented five randomly interleaved trials for each of the 12 phases for test dots along the  $+45^\circ$  diagonal, and in the second period we presented the same for test dots along the  $-45^\circ$  diagonal. Images were each presented against a black background for 1 s followed by a gray mask of the same size as the projected image containing the central fixation. The mask remained on until the observer made a response at which time the next stimulus was presented. Observers' heads were fixed in a headrest and viewing was monocular in a dark room. Each session lasted approximately 5 min.

### 2.3. Surface reconstruction

Each relative depth judgment is a measure of the sign of the local slope of the surface. That is, if the left test location appears closer, the surface appears to be locally sloping away from the observer from left to right. Similarly if the right location appears closer, the surface appears to be sloping towards the observer from left to right; and if both are at equal depths the surface is locally flat. The relative curvature of the surface was reconstructed by integrating these local signs cumulatively from left to right in the following way. Initially, the estimate of perceived relative depth at each test location was set to zero (i.e. equivalent to the plane of the screen). Distances beyond the plane of the screen were indicated by positive numbers, distances in front of the screen by negative numbers (see Fig. 2). From the results of each trial, the estimated perceived depth of the right test location was incremented and decremented according to the observer's response: if the right test location appeared closer, the depth estimate at that location was decremented by 1, if the left location appeared closer (i.e. the right location appeared farther), the depth estimate at the right test location was incremented by 1. Depth estimates were left unchanged if the two locations were reported at equal depths. Since there were five trials for each test dot configuration at each location, resulting relative depth estimates ranged from  $-5$  to  $5$ . The final array of relative depths

was then added from left to right, starting at 0. As an example, the final relative depth array of a surface seen veridically is  $[5, 5, 5, 0, -5, -5, -5, -5, -5, 0, 5, 5]$  which yields a veridical reconstruction of  $[5, 10, 15, 15, 10, 5, 0, -5, -10, -10, -5, 0]$ . Results were averaged across the two diagonals.

This reconstruction provided an estimate of the global percept of surface shape based on local depth estimates. Since the paradigm only measures *relative* depth between test locations (i.e. depth estimates were always changed by a constant  $\pm 1$ ), the reconstruction provides an estimate of the perceived ordinal depths along the surface. The perceived shapes of a triangular waveform and a flattened sinusoid could yield the same reconstruction, as could perceived shapes of different amplitudes of depth modulation. The sampling we chose was fine enough so that if depth judgments were all made veridically, the reconstructed depth of the last test location correctly equaled that of the first. The method was designed to reveal all perceived convexities and concavities in the surface, and as the results will show, it was successful in this aim.

There are several advantages to this method. Test locations were spaced closely enough and the different phases were randomly interleaved so that in any given trial, it was difficult for observers to use any sort of memory about the global shape to influence their judgment. The task was extremely local; although we did not monitor eye movements, the presentation time was limited so that observers were forced to look only in the central few degrees of the image where the fixation and test dots lay. Observers found the task to be extremely easy.

### 2.4. Average absolute error

To quantify the veridicality of observers' perceived ordinal depth for each pattern, we computed the average absolute error of perceived ordinal depth per trial. This error value is a measure of the difference between the measured relative depth array and the relative depth array of a surface seen veridically. To simplify the explanation of this computation we will number the test locations from 1 to 12 from left to right, starting at the second location, since the first and last of the 13 test locations fall at the same phase of the corrugation (see Fig. 2). For each of Locations 1–3 and 11–12 the veridical relative depth is  $+5$ , that is, observers should respond for all five trials at each of these locations that it appears farther in depth than the location to its left. Observed relative depth at these locations can vary between  $-5$  and  $+5$ , which yields a maximum error of 10 at each location. Similarly, the veridical relative depth for Locations 5–9 is  $-5$ , that is, observers should respond for all five trials at each of these locations that it appears closer in depth than the loca-

# HORIZONTAL-VERTICAL PLAID

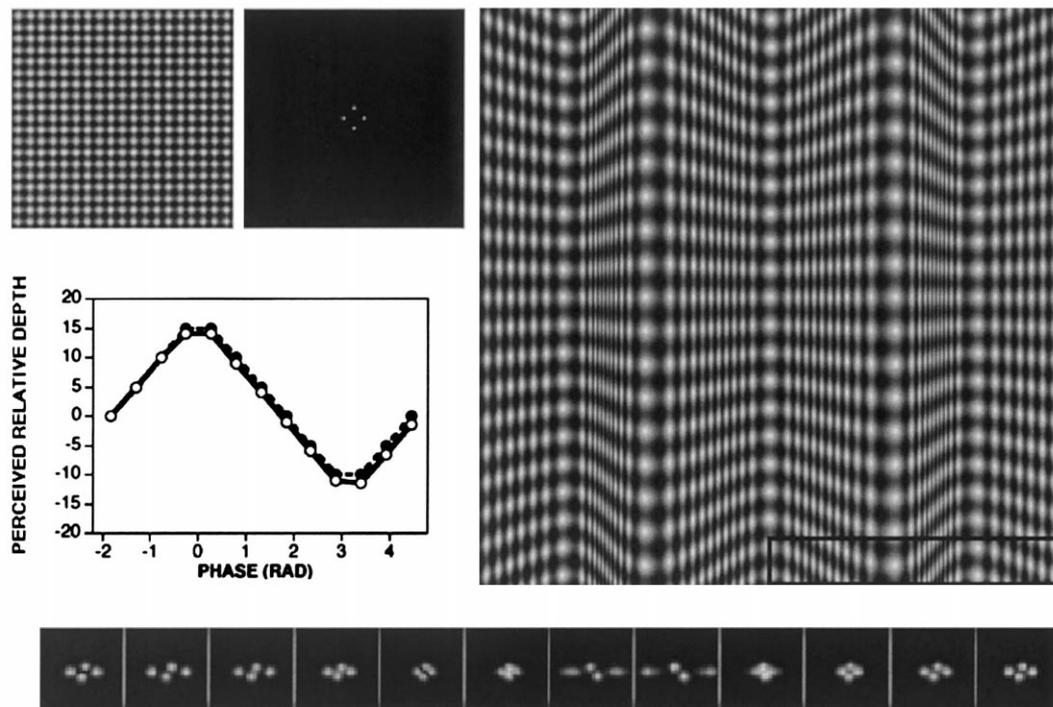


Fig. 3. Horizontal-vertical grating pattern (upper left) and its global FFT to its right. The pattern after corrugation and projection in phase = 0 is shown on the right. Local spectral changes are shown below in FFTs taken of a row of twelve  $32 \times 32$  patches in the lower right-hand portion of the projected image, from the center of the image out to the rightmost edge (outlined in black). Perceived relative depth for two observers are plotted as a function of phase on the left (AL in filled symbols with dashed lines, JR in open symbols and solid lines).

tion to its left. This again yields a maximum error of 10 at each of these locations. At Locations 4 and 10, the veridical relative depth is 0, that is, observers should always see each of these locations as appearing at the same depth as the location to its left. Since responses can only vary between 0 and +5 or 0 and -5, the maximum error will only be 5 at these two locations. This yields a total maximum error of 110 for all locations. Dividing by the number of trials gives a maximum average absolute error per trial of 1.83. The error metric provides a measure of the perceived ordinal depth in addition to the overall perceived shape of the surface; an error of 0 indicates that the observer sees the ordinal depth along the surface veridically, an error of 1.83 indicates that the observer sees the corrugated surface as  $180^\circ$  out of phase with that of the true surface. If the observer sees the surface as either flat or rectified in depth, error values will be close to 1.0.

## 2.5. Global and local spectral analyses

The format of all subsequent stimulus figures will follow that of Fig. 3. In the upper left panel, we show on a reduced scale the texture pattern before corruga-

tion and projection. Its global Fourier transform (FFT) is shown to its right. The zero frequency (dc) lies at the center of the FFT. On the right is the projected image of the corrugated pattern, shown on a larger scale. Projected images will all be shown with the phase ( $\varphi$ ) set to 0. To examine the spectral changes occurring across each projected image, we computed local FFTs on a row of twelve  $32 \times 32$  pixel patches along the lower right-hand edge of the image (outlined in black in the projected image in Fig. 3). The FFTs are shown at the bottom of the figure. The patches overlapped by 14 pixels. We filtered each patch with a Gaussian window before the FFT was computed. For this reason the local FFTs are slightly blurred compared to the global FFTs. Data for two observers are plotted on the left below the flat pattern and global FFT.

## 2.6. Observers

AL (one of the authors) and JR served as observers in the experiment; the latter was naive about the purposes of the study. Both had corrected-to-normal vision.

# VERTICAL GRATING

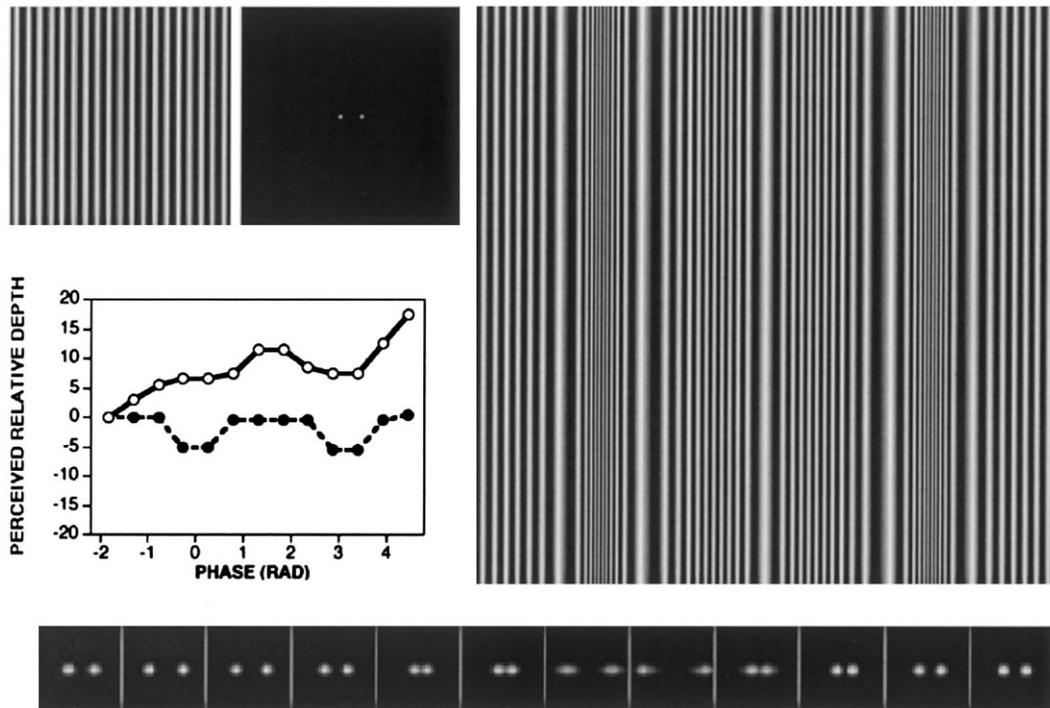


Fig. 4. Vertical grating. The local FFTs of the projected image show local changes in frequency. Observer AL sees the surface as rectified in depth and observer JR sees the peak of the corrugation as flattened and slightly slanted.

### 3. Horizontal–vertical plaid

Fig. 3 shows a horizontal–vertical plaid before (left) and after corrugation and projection (right). The spatial frequency of its grating components before projection was 2 cpd. Data for the two observers is plotted in the lower left of the figure. For this and all subsequent plots, data for observer AL are plotted in filled symbols with dashed lines and data for observer JR in open symbols with solid lines. Perceived relative depth is plotted as a function of phase of the projected image. For both observers the concave and convex portions of the sinusoid are at the correct locations and the slant between them is the proper sign, indicating that both observers perceived the three-dimensional shape veridically for this pattern. The average error was 0 for observer AL and 0.03 for observer JR.

To examine the different types of changes that occur in the projected image, we looked at the projections of the individual vertical and horizontal grating components. The vertical component of the horizontal–vertical plaid is shown in Fig. 4. Corrugation and perspective projection of this component results in frequency changes across the horizontal extent of the image. This is clearly shown in the local FFTs below. The vertical contours in the projected image correspond

to lines of minimum curvature of the surface, i.e. the depth of the surface does not change along any of these contours. The data show that this image appeared rectified in depth rather than sinusoidally corrugated for observer AL; for observer JR the peak of the corrugation was flattened and the entire surface appeared slightly slanted. The perceived rectification and flattening are reflected in the average errors of 0.81 and 0.58 for AL and JR, respectively.

Local changes in one-dimensional frequency in the projected image have been characterized for flat surfaces as changes in scale or compression, and have been shown to be relatively ineffective at conveying surface slant (Braunstein & Payne, 1969; Gillam, 1970; Rosinki & Levine, 1976; Cutting & Millard, 1984; Todd & Akerstrom, 1987; Gillam, 1995; Knill, 1998c). Sakai and Finkel (1993) on the other hand found that one-dimensional frequency modulation did carry shape for a convex cylindrical surface. They used a simulated parallel projection by filtering two-dimensional white noise with filters elongated according to the local curvature of the surface. Our results show that one-dimensional frequency changes even in perspective projection are not sufficient for conveying the shape of a corrugated surface, specifically, they are not sufficient for distinguishing convex and concave curvatures.

# HORIZONTAL GRATING

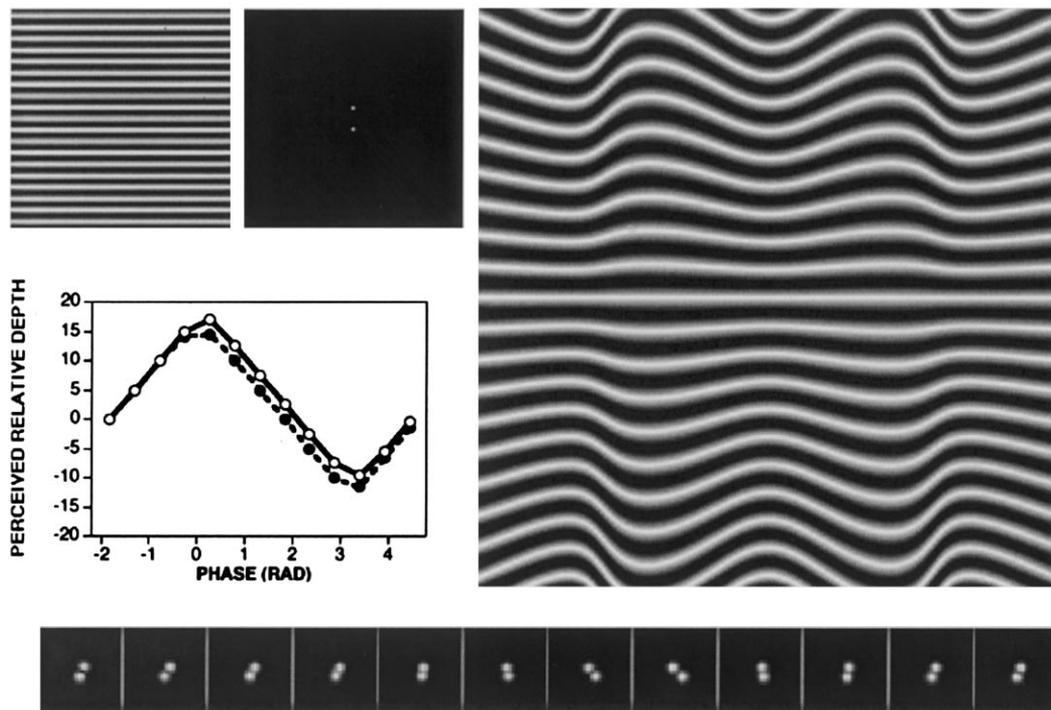


Fig. 5. Horizontal grating pattern. The local FFTs of the projected image show local changes in orientation. Both observers see veridical ordinal depth.

The horizontal grating component of the horizontal–vertical plaid is shown in Fig. 5. Perspective projection of this pattern results in local changes in orientation, as shown in the local FFTs below. Frequency remains fairly constant across the projected image. The changes in orientation that vary systematically about the axis of depth modulation result in a series of oriented contours that are the perspective projections of the lines of maximum curvature of the surface. The data show that both observers perceive three-dimensional shape correctly in this pattern. The average error was 0.06 and 0.09 for AL and JR respectively.

For flat slanted surfaces, these local changes in orientation have been termed ‘linear perspective’ (Gillam, 1995). Perspective gradients have been shown to be more effective than compression gradients in conveying the slant of flat surfaces (Braunstein & Payne, 1969; Rosinski & Levine, 1976; Cutting & Millard, 1984; Todd & Akerstrom, 1987; Gillam, 1995; Knill, 1998c).

While lines of both maximum and minimum curvature were present in the projected horizontal–vertical plaid (Fig. 3), only lines of maximum curvature are present in the projected horizontal component and only lines of minimum curvature are present in the

projected vertical component. Perceived ordinal depth was near veridical for the horizontal–vertical plaid and the horizontal grating, though slightly more so for the former. Ordinal depth was perceived incorrectly for the vertical grating. These results suggest that although contours along the lines of maximum curvature alone can convey veridical ordinal depth, the addition of contours along the lines of minimum curvature increases the veridicality of the percept, but that contours along lines of minimum curvature are not sufficient to convey ordinal depth along the surface.

Note in the FFTs at the bottom of Fig. 3 that the local spectra of the horizontal–vertical plaid pattern are simply sums of the local spectra of the horizontal grating and the vertical grating (Figs. 4 and 5). If the component grating patterns contributed linearly to the overall percept of surface shape, we would expect the perceived surface shape of the projected horizontal–vertical plaid pattern to be the sum of the percepts obtained for the projected vertical and horizontal grating. What happens instead when the two patterns are added together is that the orientation changes around the vertical frequency axis, which represent projections of the lines of maximum curvature, dominate as the cues to surface shape.

# OCTOTROPIC PLAID

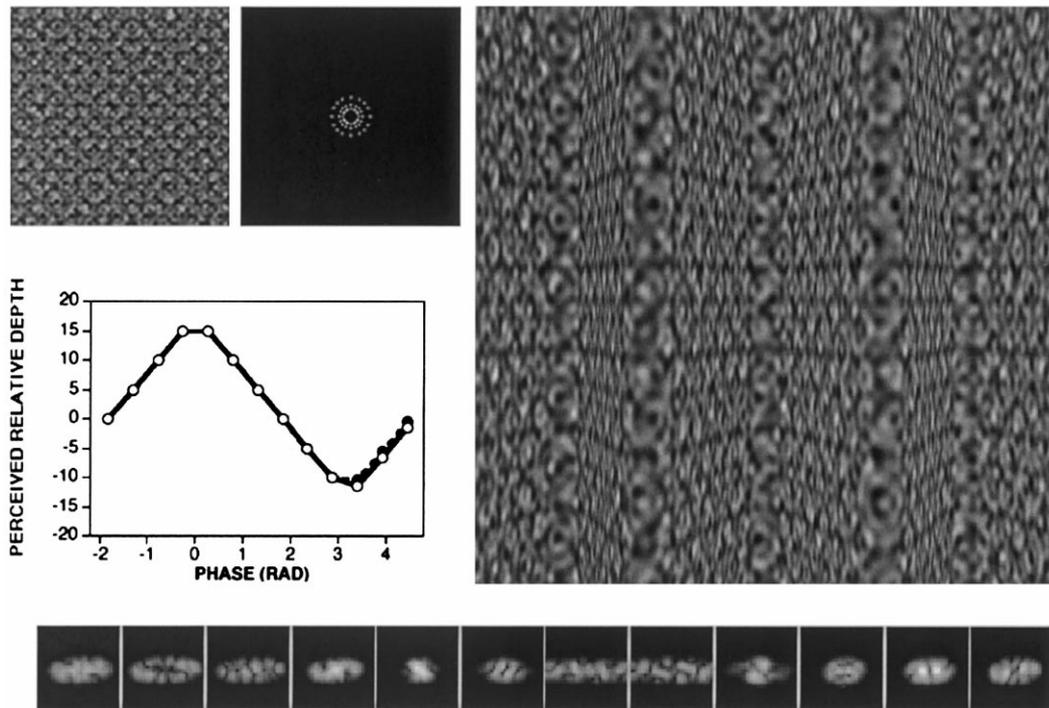


Fig. 6. Octotropic plaid consisting of eight oriented compound gratings, each the sum of three gratings of different frequencies (see text for details). Both observers see veridical ordinal depth.

## 4. Octotropic plaid

It is possible that the contours along lines of maximum curvature that appear to be critical for conveying shape arise only from projection of regular, periodic patterns such as sinusoids. We decided to try a pattern that was spectrally more complex and somewhat less regular to see if the salience of the oriented contours could be generalized. Fig. 6 shows a pattern we will refer to as an octotropic plaid. This pattern is the sum of eight components oriented at intervals of  $22.5^\circ$ ; each component was a sum of three sinusoids at 1.4, 2.1 and 3.5 cpd and at random relative phases along each axis (the horizontal component is shown in Fig. 8). The resulting pattern is less regular than the plaid and gratings examined earlier, although it is far from random. Before this pattern is corrugated and projected, straight noisy contours running along all eight directions appear equally visible (upper left). Because the surface is corrugated in depth as a function of horizontal position, after corrugation and projection the contours corresponding to the horizontal component of the pattern become more salient than the others, and appears as oriented contours falling along projected lines of maximum curvature of the surface. The contours are noisy, and energy along

them would be difficult to extract from the FFTs below, however the visual system is perfectly able to pick them out and use them to carry the percept of the curved surface as shown by the experimental results. The average error was 0.01 and 0.03 for AL and JR, respectively.

When we subtract the horizontal component from the octotropic plaid, the texture pattern appears as in the upper left panel of Fig. 7. The projected image contains all the relevant gradients and changes in frequency across the pattern as the original octotropic plaid (Fig. 6). The contours along lines of maximum curvature, however, are now absent. The surface appeared incorrectly rectified to both observers. Interestingly, the sign of the rectification was different for the two observers: AL saw the surface as rectified towards her and JR saw it as rectified away from him. The average error was 0.83 and 0.79 for AL and JR, respectively. This example provides a strong counter-example to all shape-from-texture theories that do not explicitly consider oriented energy changes corresponding to lines of maximum curvature. When we subtract both the horizontal *and* vertical components from the octotropic plaid, the pattern is almost indistinguishable from that shown in Fig. 7 and will not be shown here. Observers also viewed this pattern as rectified in depth.

# OCTOTROPIC PLAID MINUS HORIZONTAL COMPONENT

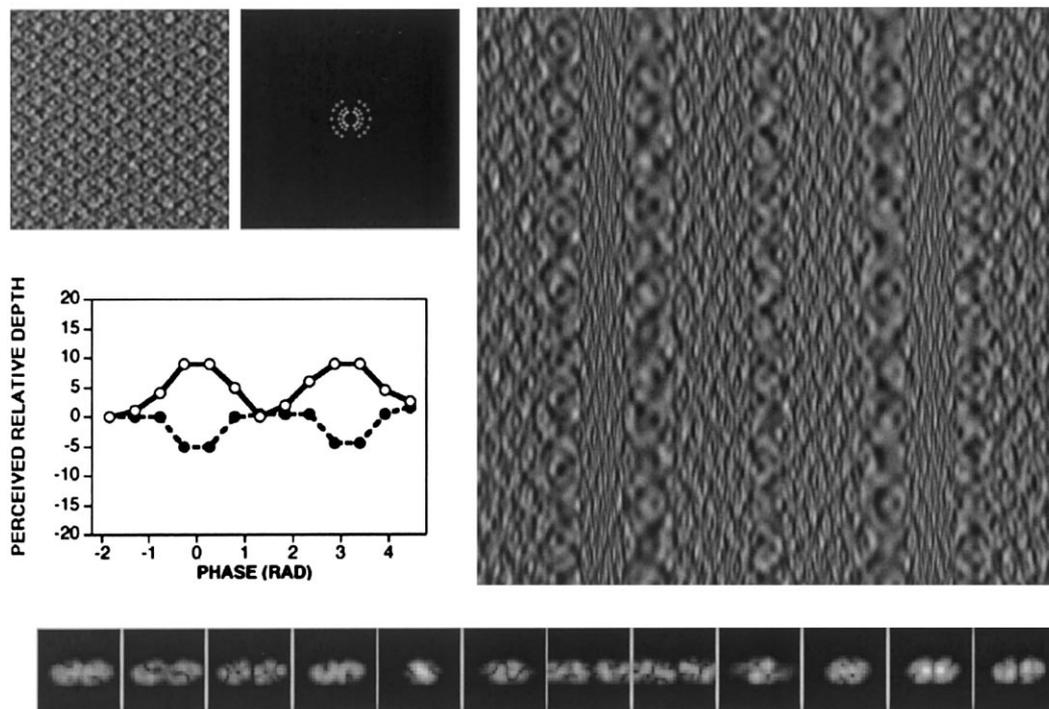


Fig. 7. Octotropic plaid minus the horizontal component (see text for details). Both observers see the surface as rectified in depth, with different signs of rectification.

Fig. 8 shows the horizontal component of the octotropic plaid in isolation. Again, due to the presence of the contours representing projections of the lines of maximum curvature of the surface, observers perceive veridical ordinal depth. The average error was 0.13 and 0.08 for AL and JR, respectively. Note once again that the pattern of energy along these contours (bottom panel) matches those found in the projected horizontal grating.

For visibility, the local FFTs in Figs. 6–8 are shown as logs of the original FFTs. The FFTs of the original octotropic plaid (Fig. 6) are the sum of the local FFTs in Fig. 7 and the FFTs in Fig. 8. The differences between Figs. 6 and 7 are particularly visible in the first three and last three panels. Despite the fact that there is energy at multiple orientations at every location across the octotropic plaid, observers are well able to extract the critical oriented components corresponding to projected lines of maximum curvature. To extract the critical local orientation at each position across the image requires a system that can simultaneously extract energy along many different orientations at each point in the image. This is something that the visual system is well equipped to

do, given that neurons in each cortical column vary systematically in orientation preferences.

## 5. Isotropic broadband noise

In response to the work of Malik and Rosenholtz (1994), the question has been raised as to whether white noise can help convey shape. To address this question we corrugated and projected a pattern of white noise that was broadly band-pass filtered (Fig. 9). The filter passed frequencies out to approximately 10 cpd. The pattern was filtered in this way so that corrugation and projection would cause spectral changes in the resulting image. The energy in this filtered noise pattern is isotropic and continuous over a range of frequencies, i.e. it contains energy at all orientations. The projected image did not contain visible contours along lines of maximum curvature and appeared rectified in depth for observer AL and significantly flattened for observer JR. The average error was 0.74 and 0.47 for AL and JR, respectively. This again suggests that perceived contours along the lines of maximum curvature and the

# HORIZONTAL COMPONENT OF OCTOTROPIC PLAID

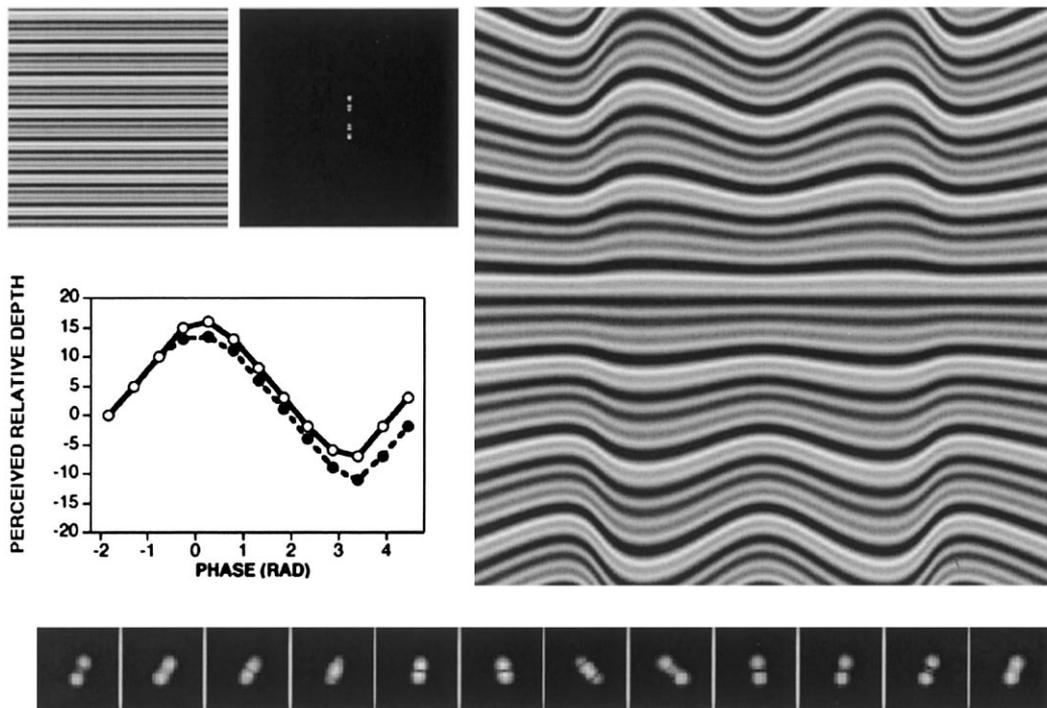


Fig. 8. Horizontal component of octotropic plaid (see text for details). Both observers see veridical ordinal depth.

corresponding oriented energy changes are critical for conveying shape.

## 6. Global spectral analysis

Thus far we have consistently found that projected patterns containing contours lying along lines of maximum curvature of the surface are critical for conveying shape. If these contours are not visible in the projected image, observers are unable to estimate ordinal depth along the surface correctly. For example, the projected octotropic plaid pattern, in which these contours were readily visible, effectively carried depth for both observers while the isotropic broadband noise did not. To better understand the kinds of spectral energy that give rise to these contours, we examined the global spectra of texture patterns before corrugation and projection.

Consider the spectral differences between the octotropic plaid and isotropic broadband noise patterns. The spectrum of the octotropic plaid before corrugation and projection is shown in the upper left corner of Fig. 10. This pattern is discrete in both orientation and frequency content. The spectrum of the noise is shown in the lower right hand corner of this figure. This pattern is broadband in both orientation and fre-

quency. (Notice also that the noise pattern contains higher frequencies than does the octotropic plaid. We also examined noise patterns with frequency cut-offs near the highest frequency of the octotropic plaid — 3.5 cpd — however these patterns appeared rectified in depth just as the original noise pattern did, indicating that the higher frequencies in the original noise pattern are not responsible for observers' inaccurate depth judgments. We will return to this point in the next section.)

The question we ask here is whether it is the discreteness in orientation, or frequency, or both that gives rise to visible contours along lines of maximum curvature after corrugation and projection. We know from the projected horizontal plaid and the octotropic plaid patterns that energy along the horizontal orientation at discrete frequencies carries shape. It is not obvious, however, whether a broadband frequency profile along this orientation will also convey shape. What is equally unclear is whether energy along a continuum of orientations, including the horizontal orientation, at discrete frequencies will carry shape. To answer these questions, we generated two new patterns: one that was discrete in orientation and broadband in frequency, and one that was broadband in orientation and discrete in frequency.

# ISOTROPIC NOISE

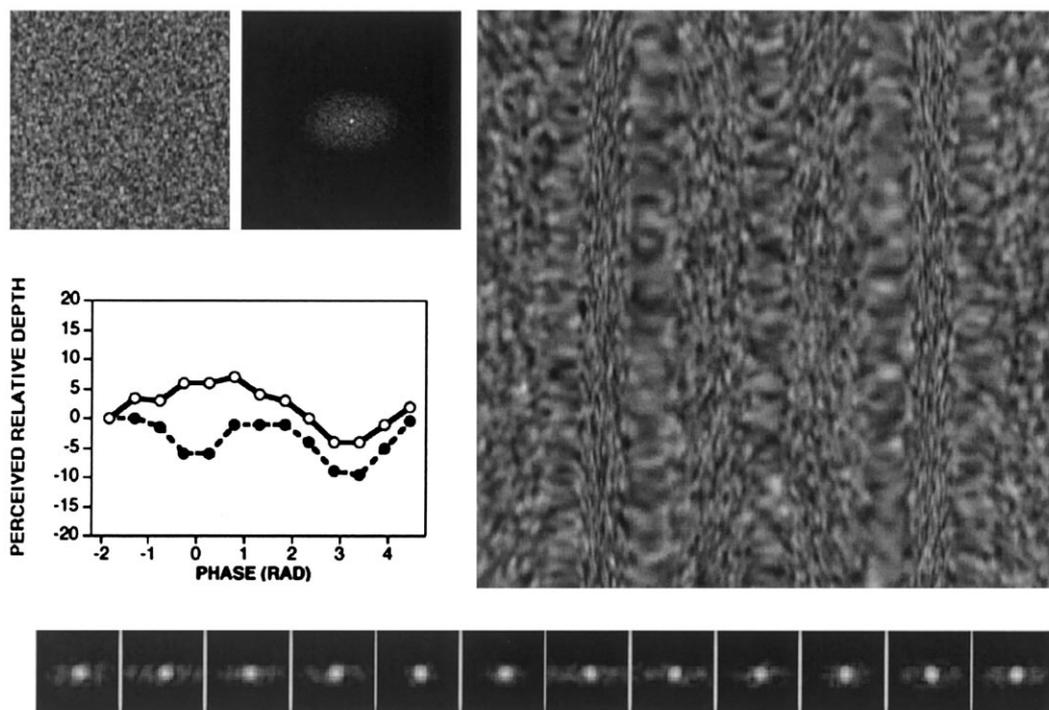


Fig. 9. Isotropic noise generated by white noise passed through a broadband filter (see text for details). Observer AL sees the surface as rectified in depth, observer JR sees the peak as almost completely flattened.

## 6.1. Discrete orientation, broadband frequency

We know that energy at a discrete number of frequencies at a discrete number of orientations yields a pattern that contains contours along lines of maximum curvature, e.g. in the octotropic plaid. What if we extend the frequency content of this pattern so that it is broadband? We generated an image containing energy along the same eight orientations as those found in the octotropic plaid, but that contained energy along a range of spatial frequencies at each orientation. To generate the image we filtered white noise with a filter shown on the upper right of Fig. 10. Since we are filtering white noise, the FFT of the resulting image looks almost identical to the filter. This filter passes energy along the eight equally spaced directions. Energy at each orientation is passed out to about 10 cpd, the same range passed in the filtered white noise. When this filter is applied to white noise, the resulting pattern looks like Fig. 11, left.

Since each of the eight oriented components of this pattern is equivalent to one-dimensional noise, we call this pattern an *octotropic noise* pattern. The pattern contains visible horizontal contours that become systematically oriented along lines of maximum curvature when the pattern is corrugated and projected (right). Observers estimated ordinal depth veridically for this

pattern. The average error was 0.1 and 0.08 for AL and JR, respectively. This suggests that it is not necessary for energy along the horizontal orientation to be discrete in frequency in order for the pattern to convey shape; veridical depth estimates are also obtained when the frequency content along this orientation is broadband.

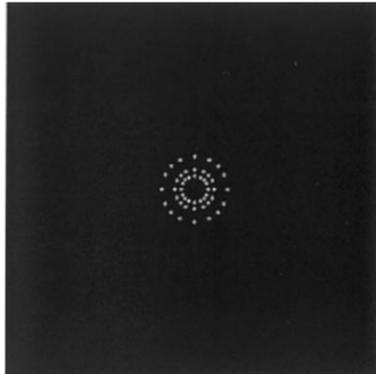
We return to the point about high frequencies here. Both the isotropic broadband noise and the octotropic noise patterns contain higher frequencies than the octotropic plaid (see Fig. 10). However the octotropic noise pattern conveys veridical shape while the isotropic broadband noise does not. This suggests once again that the presence of these higher frequencies is not correlated with the veridicality of the perceived shape.

## 6.2. Broadband orientation, discrete frequency

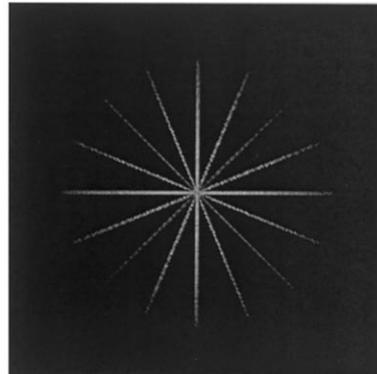
So far our results suggest that discreteness in orientation is sufficient for the critically oriented component to convey shape regardless of the frequency profile. Observers judged ordinal depth veridically for both the octotropic plaid and octotropic noise patterns. We asked the flip question for the frequency domain: will a pattern that is broadband in orientation but discrete in frequency convey shape? We generated an image containing energy at all orientations at the same three

# GLOBAL FFTs

**OCTOTROPIC PLAID**



**OCTOTROPIC NOISE**



**ISOTROPIC  
TRI-SCALE NOISE**



**ISOTROPIC  
BROADBAND NOISE**

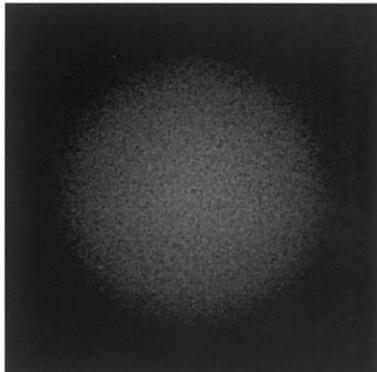


Fig. 10. Global FFTs: the octotropic plaid (upper left) is discrete in both orientation and frequency, the octotropic noise (upper right) is discrete in orientation and broadband in frequency, the isotropic tri-scale noise (lower left) is broadband in orientation and discrete in frequency, and the isotropic broadband noise (lower right) is broadband in both orientation and frequency.

frequencies as those found in the octotropic plaid (1.4, 2.1 and 3.5 cpd). The filter used to create this image is shown on the lower left of Fig. 10. Again since we are filtering white noise, the FFT of the resulting image looks nearly identical to the filter. When this filter is applied to white noise, the resulting pattern looks like Fig. 12, upper left.

This pattern is isotropic and contains energy only at three frequencies so we will refer to it as *isotropic tri-scale* noise. The pattern does not contain salient horizontal contours and therefore contours along lines of maximum curvature are not formed when the pattern is corrugated and projected (Fig. 12, right). Observer AL sees this pattern as rectified in depth, and observer JR sees it as slanted out of the fronto-parallel plane. The average error was 0.82 and 0.83 for AL and JR, respectively. This suggests that isotropic patterns

will in fact not convey veridical depth for a corrugated surface because there are no contours along lines of maximum curvature in the projected pattern; when all orientations are present in the pattern, observers are unable to estimate ordinal depth along the surface correctly. Note that all the energy in the octotropic plaid is also contained in the isotropic tri-scale noise. The crucial difference for the percept is whether the visual system can extract the pattern of oriented energy changes across the image.

Table 1 summarizes our findings. Observers correctly estimated ordinal depth along the surface for both the octotropic plaid and the octotropic noise patterns. It appears that in order for the projected image to convey shape, it is sufficient for the pattern to be discrete in orientation content; frequency can be discrete or broadband along each orientation. After corrugation and

# OCTOTROPIC NOISE

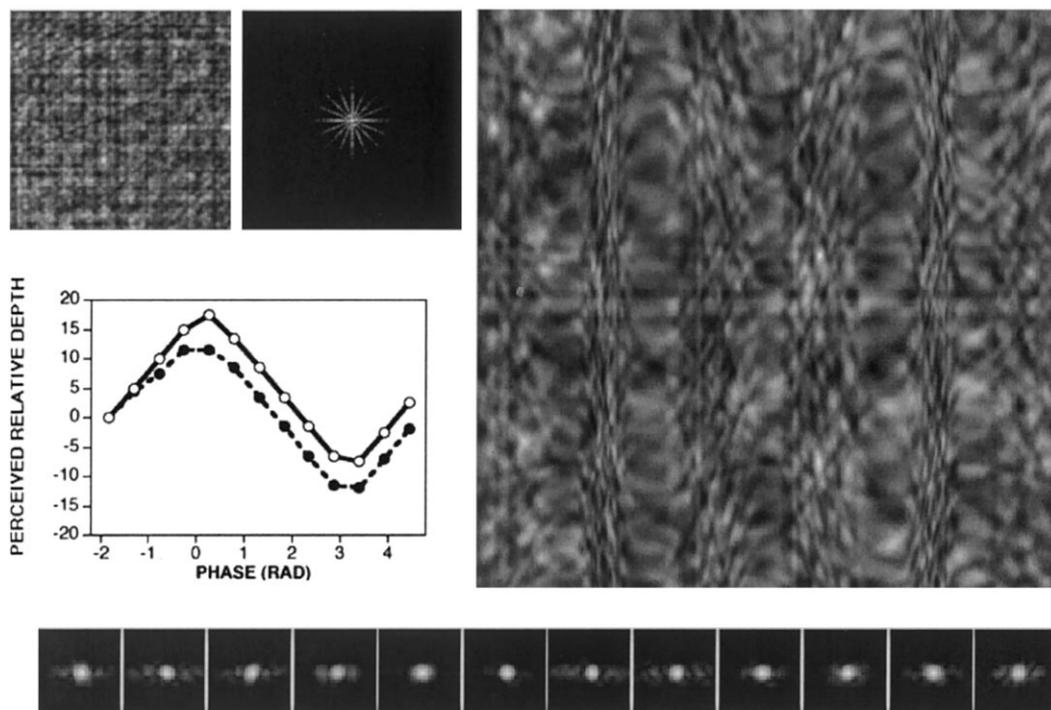


Fig. 11. Octotropic noise pattern generated by filtering white noise with the filter in the upper right panel of Fig. 10. Both observers see veridical ordinal depth.

projection, both of these patterns contain contours along lines of maximum curvature of the surface that arise from energy along the horizontal orientation in the pre-corrugated pattern. Isotropic patterns do not convey shape; observers were unable to estimate ordinal depth correctly for either the broadband noise or the tri-scale noise. Although there is energy along the horizontal orientation in both these patterns, it appears to be masked by energy at other orientations. Contours along lines of maximum curvature are absent in the projected images of both these patterns.

### 6.3. Resolution of orientations

Contours along lines of maximum curvature were not visible in either the broadband noise or the tri-scale noise. Both patterns do contain energy along the horizontal orientation before corrugation, however they also contain energy at many other directions. It is possible that energy along the horizontal direction is being masked by energy along neighboring directions. If this is the case, we would expect that adding more oriented components to the octotropic noise pattern would further reduce the visibility of the contours along lines of maximum curvature in the projected image, and in turn reduce the amount with which the pattern conveys shape. We generated a pattern with energy at

16 equally spaced orientations. This pattern was generated in the same way that the octotropic noise was generated except that the filter applied to white noise now passed energy at 16 orientations spaced  $11.25^\circ$  apart (Fig. 13). Compare this pattern to the octotropic noise pattern in Fig. 11. Notice that the horizontal contours in the precorrugated image (left) and the contours along lines of maximum curvature in the projected image (right) are less visible than in the octotropic noise pattern. For this pattern, shown in Fig. 13, observer AL saw a half-rectified surface with an average error of 0.54; observer JR however, saw veridical ordinal depth with an average error of 0.05.

We similarly generated a pattern containing energy at 32 orientations spaced  $5.625^\circ$  apart, shown in Fig. 14. (Due to contrast limitations, the global FFT contains some smearing.) This pattern resembles the pattern in Fig. 13, but the contours in the pattern are even less visible. As the oriented energy becomes more densely packed, the spectrum of the pattern begins to resemble that of the isotropic broadband noise, which is indeed what the noise of 32 orientations resembles. Observer AL now saw the surface as fully rectified with an average error of 0.77. Observer JR saw the trough of the cycle slightly flattened with an average error of 0.26. The increased error value from the 16 orientation noise condition indicates that the shape percept is beginning

# ISOTROPIC TRI-SCALE NOISE

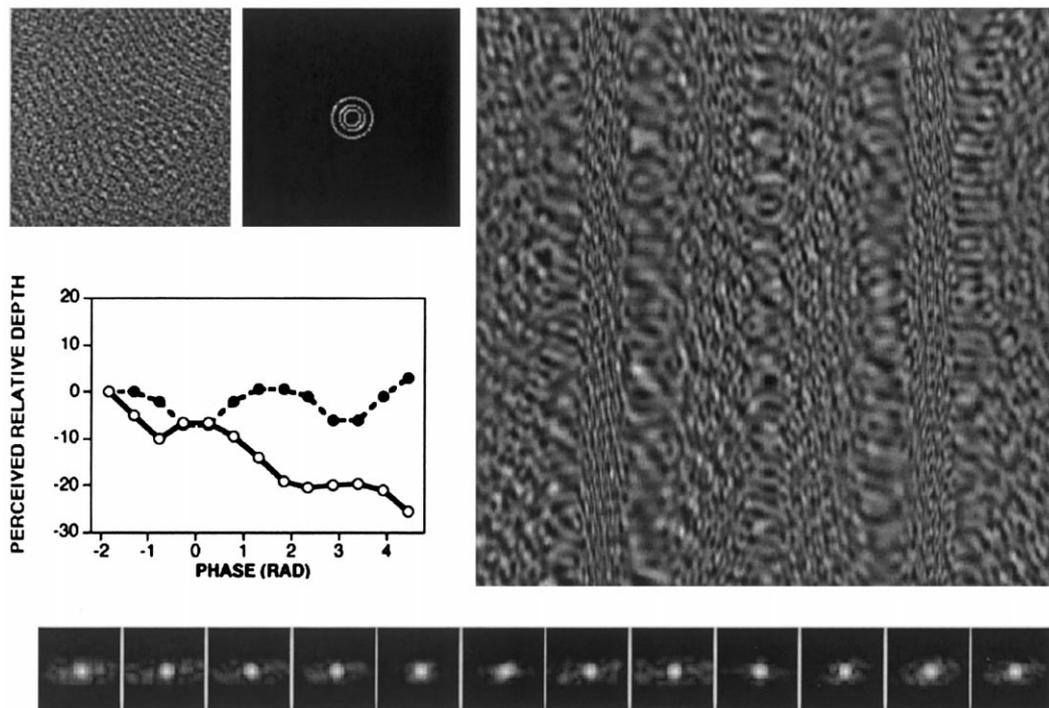


Fig. 12. Isotropic tri-scale noise pattern generated by filtering white noise with the filter in the lower left panel of Fig. 10. Both observers see the surface as rectified in depth.

to deteriorate, and we attribute this to the decreased visibility of the contours along lines of maximum curvature in the projected patterns, and equivalently to masking of oriented energy changes.

The resolution of three-dimensional shape as a function of number of orientations differs somewhat for the two observers. However the trend is the same: as the number of orientations is increased, the contours along lines of maximum curvature in the projected pattern become less visible and shape estimates become less veridical. In order for a noise pattern of this sort to convey veridical shape, energy in the pre-corrugated pattern must be spaced at intervals greater than  $11^\circ$  for observer AL. Observer JR still sees a shape that is close to veridical with noise at this orientation spacing, though the perceived trough is clearly flattened from the perceived trough in the 16 orientation noise condition. We believe that noise at an even finer spacing for this observer ( $< 5^\circ$ ) would render an even less veridical perceived shape. At intervals smaller than these limits, the energy at neighboring orientations begins to mask the energy along the horizontal. As a consequence, the visibility of the contours along lines of maximum curvature is reduced and the shape is no longer seen veridically.

## 6.4. Visibility of oriented energy

It is not necessary for energy along projected lines of maximum curvature to be in the form of continuous contours in order for the projected image to convey veridical shape, as was demonstrated by the noisy contours in Figs. 11 and 13. This is demonstrated even more clearly by the pattern in Fig. 15. This pattern was generated by filtering white noise with an elliptical filter oriented along the vertical direction in frequency space, i.e. allowing higher frequencies along the horizontal orientation to pass relative to the vertical orientation. The filter is shown in Fig. 15 to the right of the pre-corrugated pattern. The ratio of variances of the filter in frequency space along the vertical to horizontal orientations was 5:1. Veridical ordinal depth was perceived as long as the ratio exceeded this. The pre-corrugated pattern contains horizontal streaks—none of which are continuous from one side of the image to the other. After corrugation and projection, the pattern contains energy along lines of maximum curvature of the surface in the form of short streaks rather than continuous contours, somewhat akin to brushstrokes in an oil painting. These are qualitatively different from the noisy contours of the multi-orientation patterns. The

FFT shows that the energy is *continuous* over a neighborhood of orientations surrounding the critical orientation, unlike in the FFTs of the multi-orientation patterns in which energy at the critical orientation was discrete from energy at neighboring orientations. Even though the streaks in the projected image are not continuous, observers see veridical ordinal depth with average errors of 0.15 and 0.03 for AL and JR, respectively. This emphasizes that rather than the presence of actual contours it is really changes in oriented energy consistent with projections of the lines of maximum curvature that is critical for conveying shape. As long as the pattern of changes is visible, whether in the form of continuous or noisy contours or discontinuous streaks, observers can extract veridical ordinal depth.

### 7. Local spectral analysis: correlated changes in orientation

So far we have established that it is the energy along lines of maximum curvature arising from energy along the horizontal orientation that is crucial for conveying the shape of this corrugated surface. On a local scale, as shown in previous figures, the energy along these lines changes in orientation across the horizontal extent of the projected image. We now consider the two-dimensional configuration of these projected lines of maximum curvature across the entire image. For all projected patterns that conveyed veridical shape for observers, the *pattern* of curvature along these pro-

Table 1  
Summary of results for patterns in Fig. 10: patterns that are discrete in orientation convey veridical shape regardless of the frequency content; isotropic patterns do not convey veridical shape

		SPATIAL FREQUENCY	
		DISCRETE	BROADBAND
ORIENTATION	DISCRETE	octotropic plaid 	octotropic noise 
	BROADBAND	isotropic tri-scale noise 	isotropic broadband noise 

jected lines of maximum curvature itself changes across the vertical extent of the image, i.e. along projected lines of minimum curvature.

Fig. 16 shows local FFTs for the entire lower right-hand quadrant of the projected horizontal grating (Fig. 5). The bottom row matches the FFTs shown in Fig. 5. The local orientation along the contours changes as we traverse each row of patches from left to right (except for the top row). These changes correspond to emergent contours consistent with perspective projection of lines of maximum curvature along the surface. The local orientation *also* changes as we move from the top of the quadrant (middle of the projected image) down to the bottom, i.e. along projected lines of minimum curvature. Unlike the changes along projected lines of maximum curvature, these local changes in orientation do *not* correspond to emergent contours along lines of minimum curvature, but to local oriented energy differences across corresponding points along different emergent contours.

The pattern of oriented energy across the entire projected image in Fig. 5 can be derived by reflecting the quadrant pattern in Fig. 16 across the horizontal and vertical axes of the image. The pattern of orientation changes along the vertical extent of the image conveys information about the orientation of the corrugated surface with respect to the observer, i.e. that it lies in the fronto-parallel plane. Deviation from the pattern shown in Fig. 16 reflects a change in the orientation of the surface from the fronto-parallel plane.

In Fig. 17 we have extracted one of the contours close to the top of the projected image of the corrugated horizontal grating, and repeated it at equal distances across the vertical extent of the image. The resulting pattern contains changes in orientation along projected lines of maximum curvature (the horizontal direction), such that the local FFTs would look exactly like those in Fig. 5. However there are no changes in orientation along projected lines of minimum curvature (the vertical direction). Although the surface conveyed by this pattern appears corrugated in depth, the surface appears tilted forward out of the fronto-parallel plane. Similar three-dimensional percepts are conveyed in various works by artists such as Bridget Riley (Riley, 1995).

The curvatures of the lines in Fig. 17 are identical. However, observers agree that the depth curvatures in the surface appear to increase in width with increasing height in the image. This is an illusion created by the visual system under perspective projection. The only three-dimensional surface that will give rise to this set of parallel curved lines in the retinal image is one that contains depth curvatures increasing in width with the entire surface tilted forward out of the fronto-parallel plane. This is another example of the visual system

# NOISE WITH 16 ORIENTATIONS

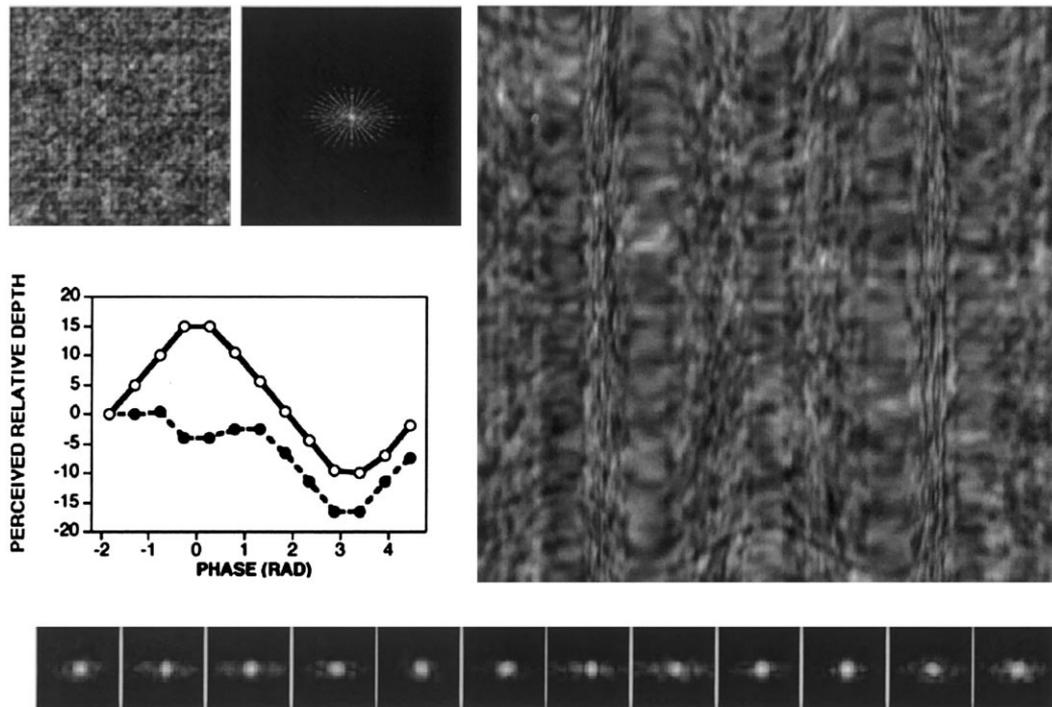


Fig. 13. Noise containing broadband energy at 16 orientations spaced at  $11.25^\circ$  intervals. Observer AL sees the surface as half-rectified, observer JR sees veridical ordinal depth.

# NOISE WITH 32 ORIENTATIONS

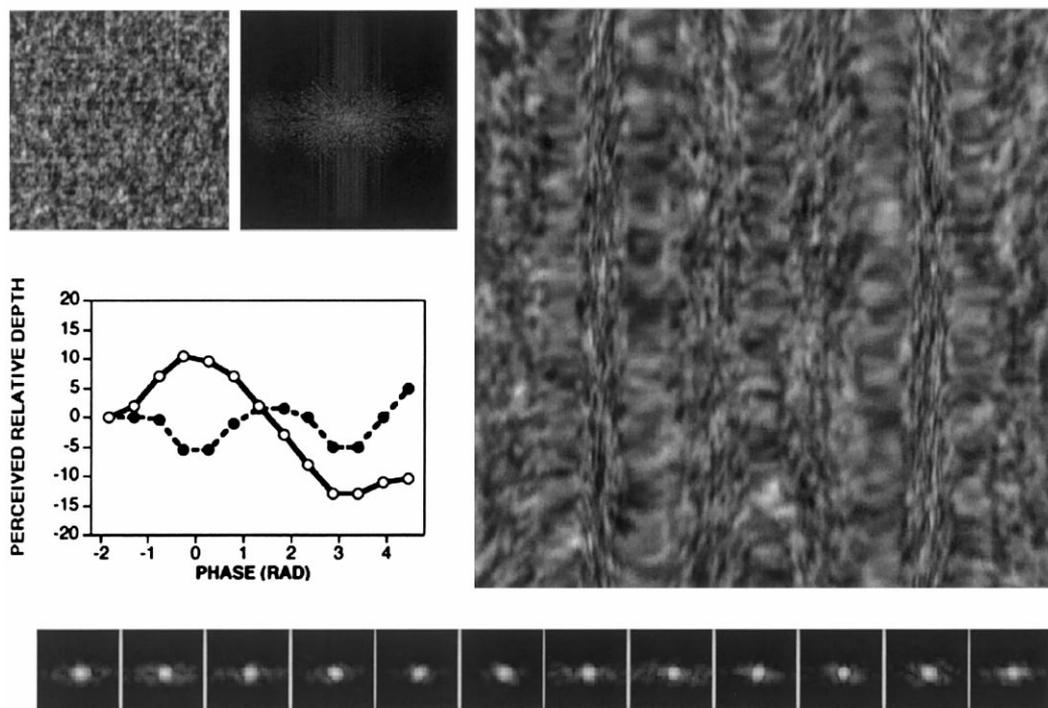


Fig. 14. Noise containing broadband energy at 32 orientations spaced at  $5.625^\circ$  intervals. Observer AL sees the surface as fully rectified, observer JR sees the trough of the cycle as flattened.

# HORIZONTAL STREAKS

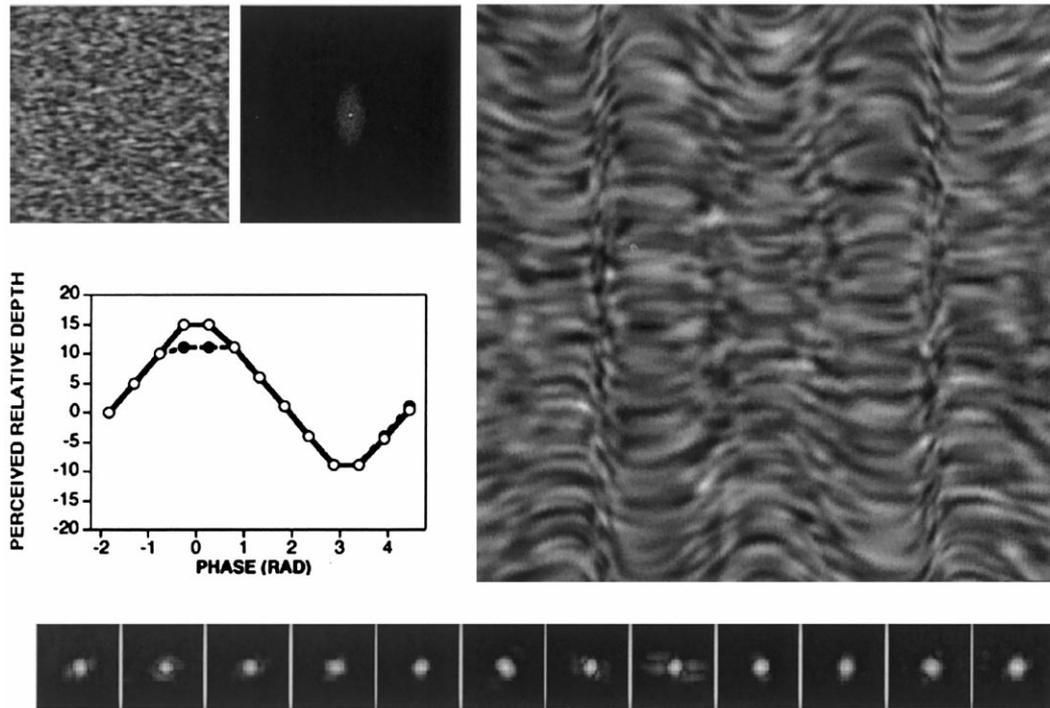


Fig. 15. Noise filtered with an elliptical filter (shown on left) with 5:1 ratio of vertical to horizontal variances. The resulting pattern contains horizontal streaks, none of which are continuous from one side of the image to the other. Both observers see veridical ordinal depth.

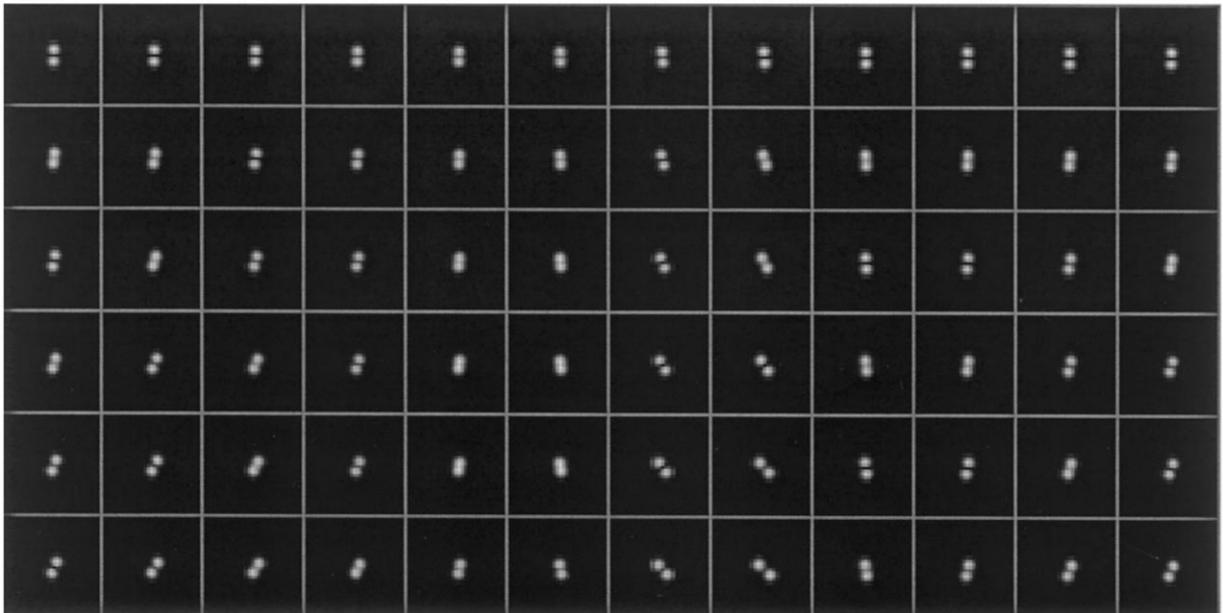


Fig. 16. Local FFTs of lower right-hand quadrant of the projected horizontal grating pattern in Fig. 5. Orientations change across the horizontal and vertical extents of the image.

interpreting retinal images as perspective projections of three-dimensional objects.

## 8. Parallel versus perspective projection

When the surface is viewed at large distances, changes in depth along the surface are small relative to the viewing distance. Under these conditions, the retinal image is projected approximately in parallel rather than in perspective. When we project the corrugated horizontal–vertical plaid shown in Fig. 3 (left) in parallel (see Appendix A for projection computation), the resulting image looks like that shown in Fig. 18. (Compare with the projected image for the same pattern under perspective projection in Fig. 3.) While the corrugation still creates changes in frequency across the vertical grating component, the horizontal grating component is unchanged. As a result there are no changes in orientation along either the horizontal or vertical extent of the projected image. If the contours in the projected image are assumed to lie along lines of maximum and minimum curvature of the surface, then an analysis such as that in Stevens (1981) would predict that the normals to the surface would all point in the same direction and therefore the surface should appear flat, which is indeed what observers see. Observers saw the surface as nearly (AL) or completely flat (JR). The average error was 0.83 for both observers.

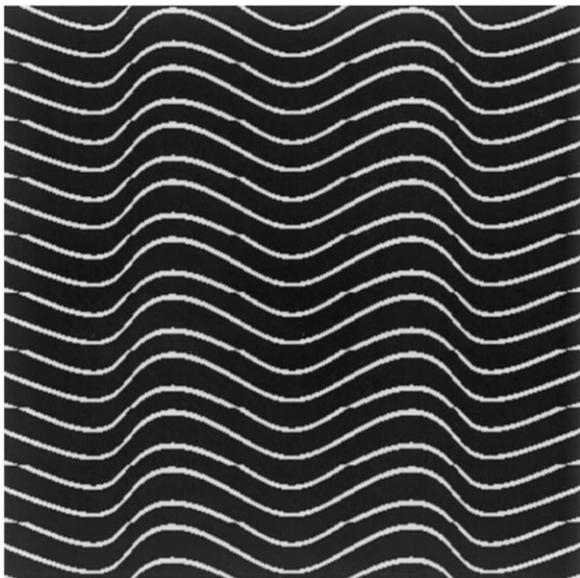


Fig. 17. A pattern in which orientations change across the horizontal extent of the image, but not the vertical extent. To create the pattern, one of the curved contours from the projected horizontal grating in Fig. 5 was replicated at equal distances across the vertical extent of the image. Although the surface appears corrugated in depth, it appears tilted forward out of the fronto-parallel plane.

It appears that texture will not be a good cue for shape unless the image is a perspective projection. Under perspective projection, the horizontal component of the pattern becomes systematically oriented along lines of maximum curvature of the surface, giving rise also to changes in oriented energy along lines of minimum curvature. It is these correlated changes in oriented energy that are crucial for veridical depth estimation.

## 9. Discussion

### 9.1. Global shape percepts from local depth estimates

Our method for measuring perceived shape is an advance over previous methods, but still has some limitations. In previous studies, observers' judgments of three-dimensional shape have generally been based on global percepts. For example, a typical task has involved judging whether a surface is flat or curved (Cutting & Millard, 1984) and/or judging the amount of curvature or depth of a surface (Vickers, 1971; Cutting & Millard, 1984; Todd & Akerstrom, 1987; Cummings et al., 1993; Sakai & Finkel, 1993). However it is difficult to extract the exact perceived shape of the surface based on these kinds of measures. For example, rank-ordering the curvature of a surface (e.g. Cutting & Millard, 1984; Sakai & Finkel, 1993) involves comparing it to some internal template, the shape of which is not measured in the task. While matching to visible templates (e.g. Todd & Akerstrom, 1987) may provide a more direct estimate of the percept, it limits the number of possible responses. Our study shows that a global percept of relative surface curvature can be easily reconstructed from multiple local depth estimates taken across the surface. Since the different phases are randomly interleaved, on any given trial observers make simple relative depth judgments without needing to know the global shape of the surface. Test locations are placed very close to the fixation so that it is actually detrimental for observers to look around the image to try to get a more global percept. Reichel and Todd (1990) used a similar local depth judgment task in examining depth reversals in stimuli with multiple curvatures as defined by shading and surface contours. However rather than reconstruct the perceived shape of the surfaces, they presented the percentage of responses consistent with the local mathematically defined depth. Also, they used parallel projection in their study and placed the test probes at various locations in a single image. Our images were projected in perspective with respect to the central phase so that observers were always fixated at the center of the image no matter what phase was presented.

# HORIZONTAL VERTICAL PLAID IN PARALLEL PROJECTION

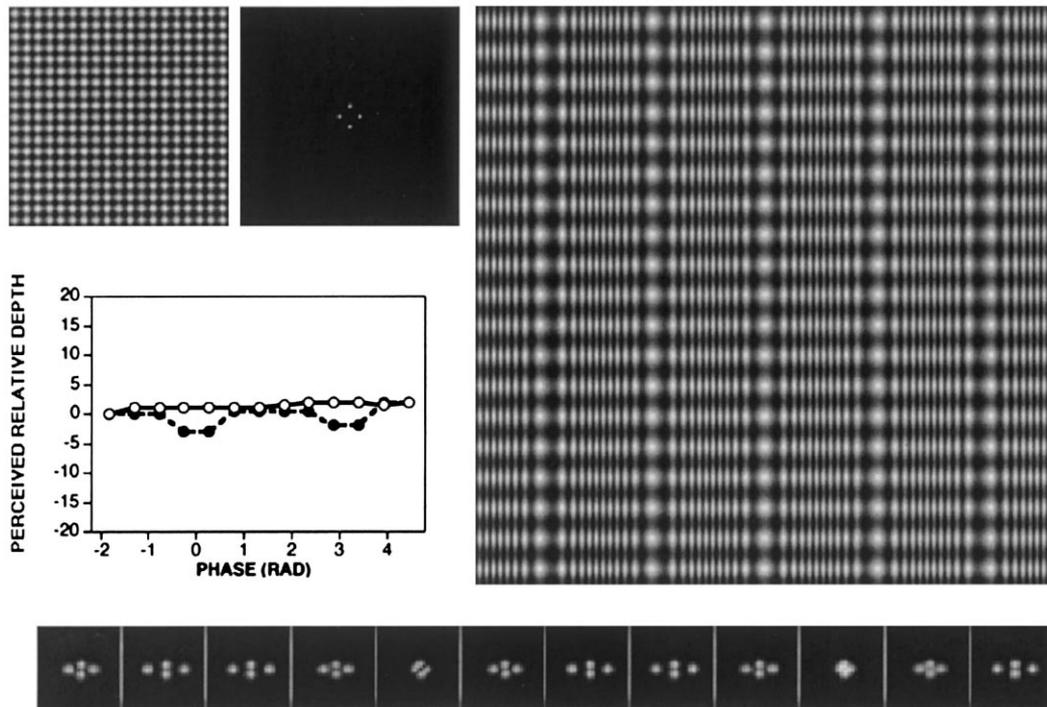


Fig. 18. Horizontal–vertical plaid under parallel projection. Compare with the same pattern under perspective projection in Fig. 3. The surface appears nearly flat for observer AL and completely flat for observer JR.

In principle and practice, this new method can be used to measure all perceived concavities and convexities in a surface of arbitrary complexity. However, there are some limitations to the method. The local ordinal depth judgments seem to be very consistent if perceived ordinal depth is above threshold as in Fig. 3. In such cases, the judgments essentially provide the sign of the slope (derivative) of the surface and not its magnitude. The reconstruction procedure is akin to a cumulative integration operation on the sign of the first derivative. As a result the method barely distinguishes between the triangular depth percept of the horizontal–vertical plaid in Fig. 3 and the flattened sinusoidal appearance of the horizontal streak pattern in Fig. 15. We intend to create elaborated versions of this method to overcome these limitations, but for the purposes of this paper, the results obtained were sufficiently critical.

## 9.2. Multiple depth curvatures

Our results strongly suggest that in order to fully understand how changes in texture are used as cues for surface shape we must utilize stimuli that contain multiple curvatures. Although Sakai and Finkel (1993) showed that one-dimensional frequency modulations

were effective at conveying shape from convex cylindrical surfaces, our results show that they are not sufficient for distinguishing convexities from concavities. This holds true for parallel and perspective projections.

The corrugated surface we have used in these experiments, together with flat and cylindrical surfaces, falls into the category of *developable* surfaces (for review see Stevens, 1981). Developable surfaces are curved along at most one direction in three-dimensional space, and can thus be formed by bending and twisting a piece of paper without stretching or crumpling it. By definition, the Gaussian curvature, i.e. the product of the maximum and minimum curvature, of a developable surface is zero. We plan to examine surfaces of non-zero Gaussian curvature in future experiments, but for the purposes of this paper sinusoidal corrugations have proven to be sufficient.

## 9.3. Perspective projection

The critical pattern of correlated changes in orientation along projected lines of maximum and minimum curvature is absent in parallel projections. This strongly suggests that, for a curved surface lying in the fronto-parallel plane with respect to the observer, texture will

be a useful cue for shape *only* when the retinal image is a perspective projection. Our results emphasize that to study shape from texture properly, one needs to project surfaces in perspective, despite the complexities of calculating the projected image. Previous studies have examined flat surfaces in perspective (Vickers, 1971; Rosinki & Levine, 1976; Cutting & Millard, 1984) and curved surfaces in parallel (Cummings et al., 1993; Blake et al., 1993). In parallel projection, convex and concave curvatures are mathematically ambiguous and therefore the two cannot be distinguished when presented simultaneously. Todd and Akerstrom (1987) reported a slight (but significant) decrease in perceived depth for spherical surfaces viewed under parallel projection compared to the same surfaces viewed under perspective projection. They used a simulated parallel projection by increasing the viewing distance in perspective projection ninefold; it is likely that had they used true parallel projection (see Appendix A), their depth estimates would have been more greatly affected.

#### 9.4. Orientations with respect to surface curvature

The oriented contours along projected lines of maximum curvature result from perspective projection of visible *horizontally* oriented components in the pre-corrugated pattern. By horizontal, we mean that the contours of the component (e.g. the bars of a horizontal sinusoid running along  $x$ ) run *parallel* to the depth modulation ( $z = f(x)$ ). That is, the luminance modulations of the pattern ( $\text{lum} = f(y)$ ) run *orthogonal* to the surface curvature. Likewise, in projected *vertically* oriented patterns, contours run *orthogonal* to the surface curvature (depth is constant along a single contour), and luminance variations run *parallel* to the surface curvature. Visible horizontal components will give rise to the critical oriented contours *only* for surfaces whose depth varies as a function of  $x$ . If the surface were corrugated as a function of  $y$ , the contours would arise from vertical components. The differences that occur between these horizontal and vertical components have been phenomenologically described as perspective and compressive gradients in the context of depth and distance along flat surfaces slanting out of the fronto-parallel plane (Gillam, 1995) and observers have been found to respond differentially to the two types of gradients (Rosinki & Levine, 1976; Goodenough & Gillam, 1997). However to our knowledge the differences have never been spectrally described as correlated orientation and frequency changes that can be picked up by low-level mechanisms in the visual pathway; nor have they been studied in this context in examining shape from texture for curved surfaces. More importantly, since observers perceive three-dimensional shapes from texture in the absence as well as the presence of perceived gradients, in the observer model

suggested by this paper, texture gradients, ‘linear perspective’, and emergent contours are all as much a result of the neural processes of perception as is the resultant shape percept. The underlying process involves the detection and utilization of patterns of oriented energy in the image.

#### 9.5. Spectral composition of textures

In this paper we used synthetic textures consisting of discrete and continuous components in frequency and orientation. One of the main results is that three-dimensional shape is perceived veridically only when the visual system can extract oriented energy changes across the image that correspond to projections of the lines of maximum curvature of the three-dimensional shape. It is well established that the detection of oriented patterns can be masked by the presence of patterns of similar orientations (Graham, 1989; Ross & Speed, 1991). As a result we find that three-dimensional shape is conveyed only by textures that contain variations in a direction orthogonal to the direction of the corrugation, but that this component must be detectable, which happens when either the global frequency spectrum of the texture pattern is discrete in orientations, or is elongated with an aspect ratio of 5 or greater in favor of the critical orientation.

One of the main contributions of this paper is that it suggests that only some natural textures will convey impressions of three-dimensional shape, and that which textures do so can be predicted on the basis of the global FFT of the texture pattern. We have recently analyzed all the natural textures presented by Brodatz (1966) (Li & Zaidi, submitted). Results show that textures whose oriented energy is discrete or dominant along the critical direction of maximum curvature, e.g. herringbone weaves, convey veridical shape, whereas those that are continuous and isotropic, or discrete or dominant along non-critical directions, e.g. pebbles, do not.

Turner et al. (1991) showed that frequency modulations convey depth for flat surfaces, and that slant judgments were increasingly under-estimated as the spectra of the image increased in complexity. The octotropic plaid and octotropic noise patterns used in our experiments were spectrally more complex than any of the plaid/grating patterns, but observers were well able to judge ordinal depth for the curved surface. Sakai and Finkel (1993) also found that spectrally complex noise patterns conveyed cylindrical shapes just as ‘vividly’ as gratings. While observers did fail to see veridical ordinal depth for the broadband noise and the isotropic tri-scale noise patterns in our study, we attribute this to the paucity/absence of visible contours corresponding to the projection of lines of maximum curvature.

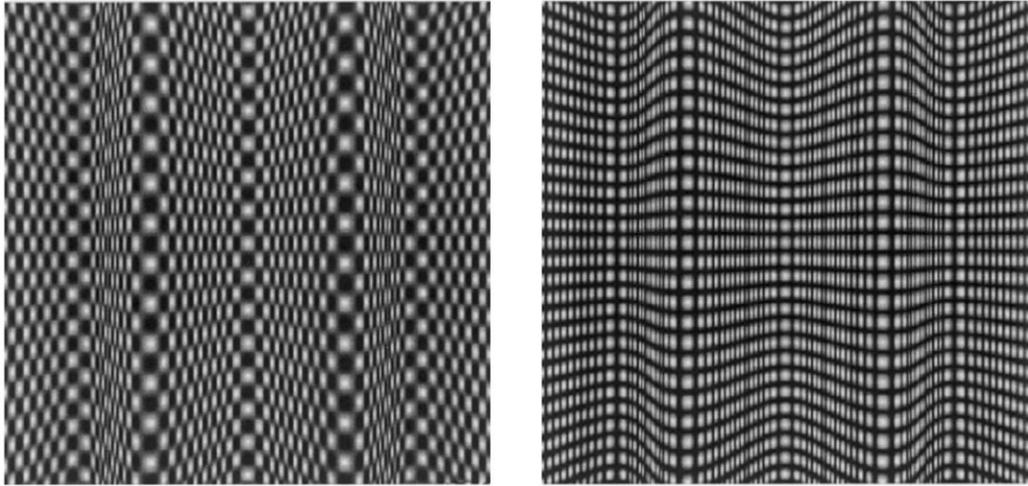


Fig. 19. Left: Diagonal plaid patterns composed of  $\pm 45^\circ$  gratings. Projected lines of maximum curvature are visible as ‘illusory’ contours that do not contain real energy. Despite this, observers see veridical ordinal depth. Right: After the pattern is luminance-rectified, contours along projected lines of maximum curvature contain real energy.

### 9.6. Mechanisms responsible for shape-from-texture

Previous studies have defined stimuli in the context of individual elements such as density, compression and perspective (Cutting & Millard, 1984; Todd & Akersstrom, 1987; Cummings et al., 1993; Blake et al., 1993; Knill, 1998a,b,c). Blake et al. (1993) and Knill (1998a,b,c) have formulated ideal observer and generic observer models for line-element and extended-area element texture gradients of various sorts. These models require the observer to make a number of ancillary assumptions, e.g. isotropy, and compare perceived gradients to physically present gradients. The results in this paper show that the extraction of gradients is neither necessary nor sufficient, and that these models will not be able to predict which patterns will convey veridical shape because they don’t extract lines of maximum curvature. For example, similar texture gradients are present in Figs. 6, 7 and 12 whereas observers report veridical three-dimensional percepts only for the pattern in Fig. 6. Our model is in a sense more direct because it assumes that the observer directly infers local three-dimensional orientation by comparing patterns of oriented energy in an image with knowledge of perspective projections of lines of maximum curvature.

Although models with filter based front-ends (e.g. Voorhees & Poggio, 1988) may be able to extract texture elements from many of the patterns used in this paper, our model does not require this stage. We suggest that instead of extracting texture elements and computing perspective changes in these elements, it would be more generally useful and biologically plausible to define stimuli in the context of, and subsequently postulate mechanisms that are sensitive to, correlated changes in orientation and frequency across the pro-

jected image. In future work we intend to specify this model in an explicit mathematical form, so that it can predict veridical and non-veridical shape percepts. The scope of this paper is limited to providing empirical support that can justify constructing such a model.

The observer model will be able to build on two streams of work. First, the visual cortex contains neurons tuned to many different orientations at every point in retinal space. These neurons will extract all orientations present at every point in the projected image. Such neurons (both simple and complex) have been used in models of texture segmentation (e.g. Bergen & Adelson, 1986; Bergen & Landy, 1991; Sagi, 1995; Graham & Sutter, 1998). However, unlike in texture segmentation, global filtering of an image by such mechanisms will not extract the shape, instead the outputs of local filters tuned to different orientations will need to be kept separate and fed to a higher level module. Second, psychophysical and computational studies examining contour integration and detection in noise (e.g. Field, Hayes & Hess, 1993; Dakin & Hess, 1998) have suggested that an associative linking mechanism between local mechanisms tuned to similar orientations is required for contours to be detected. An explicit model for extracting contours has not yet been created and has been proving elusive (Gallagly & Geisler, 1998). A model for a system to infer shape from texture cues will be more complex in two ways than just detecting a single contour. First, in all the contour integration experiments, the stimulus consisted of elements of a single orientation at any point in the image, whereas in images like Figs. 6 and 11, the visual system has to extract the relevant oriented energy in the presence of energy at many other orientations at every point. Second, as shown in Figs. 16 and 17, it is the

two-dimensional pattern of oriented energy across the image rather than a one-dimensional contour that has to be extracted for a higher level module.

### 9.7. Shape from 'illusory' contours

Lastly, there is a class of patterns that will be discussed in a future paper (Li & Zaidi, in preparation) that is the only exception to our theory that we have discovered. As shown in Fig. 19 (left), this is a projected diagonal plaid pattern consisting of two 45° gratings. Neither of the projected component gratings follows projected lines of maximum curvature, and hence the projected plaid does not contain real energy along these lines. Despite this, observers see veridical shape in this image. Although they do not contain real energy, projected lines of maximum curvature are visible in the form of 'illusory' contours formed by zero-crossings between light and dark regions. Energy along these contours can be extracted by non-linear operations on the luminance values of the image, e.g. rectification or half-squaring. Complex cells in visual cortex perform similar operations (e.g. Heeger, 1992) and thus it has been implemented in many texture segmentation models as a second stage of processing following linear

filtering (Bergen & Adelson, 1986; Bergen & Landy, 1991; Sagi, 1995; Graham & Sutter, 1998). Full-wave rectification of the projected diagonal plaid pattern results in an image that contains real energy along the lines of maximum curvature (Fig. 19, right). This suggests that extraction of the relevant energy may not be confined to a single level in the visual pathway. This pattern further emphasizes that the requisite pattern of energy can be visible in many forms. However, these 'illusory' contours are a special case, and are not likely to be found in patterns containing a larger range of spectral components, such as the pattern in Fig. 7 which, along with oriented components along the two 45° diagonals, also contains components at several other orientations. Unlike the diagonal plaid, there is no formation of 'illusory' contours along the lines of maximum curvature when this pattern is corrugated and projected, nor are the contours formed when we luminance rectify the projected image.

### Acknowledgements

Many thanks to Joseph Rappon for his time and diligence as an observer. This work was presented in

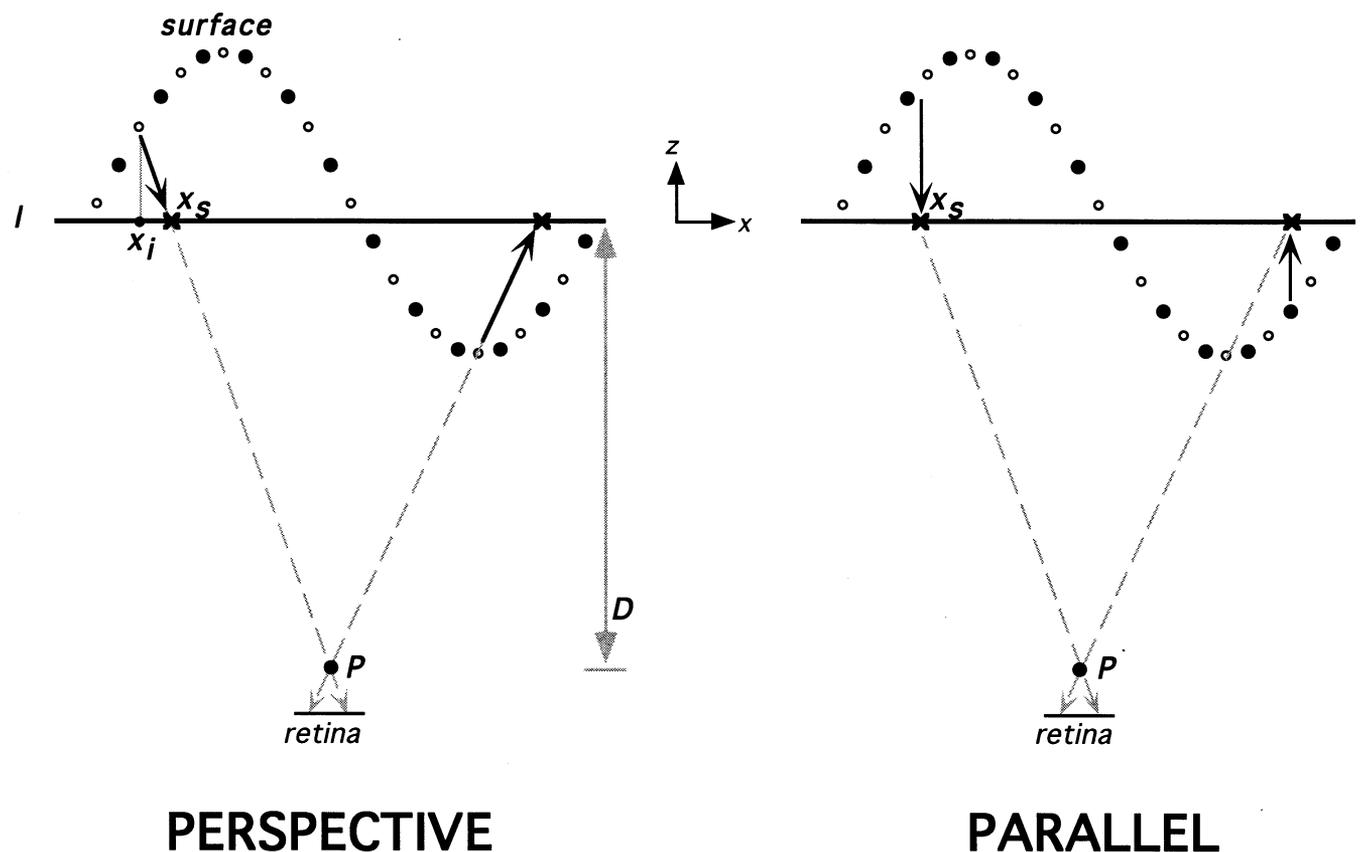


Fig. A1. Perspective and parallel projections in the  $x$ - $z$  plane for a surface corrugated in  $x$ - $z$ . In perspective projection, each location on the surface is projected onto the image plane ( $I$ ) by connecting it with  $P$  (the nodal point of the eye). In parallel projection, locations are projected by parallel rays (perpendicular to the image plane).



$$\frac{Dx_i}{x_s} = A \cos(2\pi f x_i + \varphi) = D \quad (\text{A1})$$

where  $A$  is the amplitude,  $f$  is the frequency, and  $\varphi$  is the phase of the depth corrugation. The value of  $x_i$  was then used to compute the length traversed along the surface using the integral:

$$L = \int_0^{x_i} \sqrt{1 + \left(\frac{dz}{dx}\right)^2} dx \quad (\text{A2})$$

where

$$\frac{dz}{dx} = -2\pi f A \sin(2\pi f x_i + \varphi) \quad (\text{A3})$$

$L$  was used as an index along  $x$  into the (flat) luminance pattern. We then computed the index along  $y$  into the pattern. The depth of the corrugated surface varied only as a function of  $x$  (and not  $y$ ). Therefore for each value of  $x_s$  the depth of the surface ( $z$ ) was constant for all values of  $y$  (see Fig. A2). For each value of  $y_s$ , we use  $x_i$  (computed in Eq. (A1)) to compute  $y_i$ , the  $y$ -coordinate of the intersection of the projecting ray connecting  $P$  and  $y_s$ :

$$y_i = \frac{Dy_s}{D + z} \quad (\text{A4})$$

where

$$z = A \cos(2\pi f x_i + \vartheta) \quad (\text{A5})$$

$(L, y_i)$  now provided an index into the luminance pattern: if the pattern was defined as a continuous function (e.g. grating or plaid) then  $(L, y_i)$  was plugged in to compute the luminance value at that location; if it was not a continuous function (e.g. noise), we first generated a large (flat) image of the pattern and then used  $(L, y_i)$  as an index into the image. The luminance value at  $(L, y_i)$  was then drawn in the image plane at  $(x_s, y_s)$ .

### A.3. Computing parallel projection

In parallel projection, each location across the image ( $x_s$ ) was used directly to compute  $L$  (see Fig. A1, right), by replacing  $x_i$  with  $x_s$  in Eq. (A2) and Eq. (A3), above. Under parallel projection,  $y_i = y_s$  (see Fig. A2, right), therefore  $(L, y_s)$  was used as the index into the luminance pattern.

## References

- Ames, A., Jr. (1951) Visual perception and the rotating trapezoidal window. *Psychological Monograph*, 65 (7), Whole No. 324.  
 Bajcsy, R., & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing*, 5, 52–67.  
 Bergen, J. R., & Adelson, E. H. (1986). Visual texture segmentation

- based on energy measures. *Journal of the Optical Society of America A*, 3, 98.  
 Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. In M. S. Landy, & J. A. Movshon, *Computational models of visual processing*. Cambridge: MIT Press.  
 Blake, A., Bulthoff, H. H., & Sheinberg, D. (1993). Shape from texture: ideal observers and human psychophysics. *Vision Research*, 33(2), 1723–1737.  
 Braunstein, M. L., & Payne, J. W. (1969). Perspective and form ratio as determinants of relative slant judgments. *Journal of Experimental Psychology*, 81(3), 584–590.  
 Brodatz, P. (1966). *Textures: a photographic album for artists and designers*. New York: Dover.  
 Cummings, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, 33, 827–838.  
 Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology General*, 113(2), 196–216.  
 Dakin, S. C., & Hess, R. F. (1998). Spatial-frequency tuning of visual contour integration. *Journal of the Optical Society of America A*, 15(6), 1486–1499.  
 Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local ‘association field’. *Vision Research*, 33(2), 173–193.  
 Gallogly, D. P., & Geisler, W. S. (1998). Contour integration occurs along various stimulus dimensions. *Investigative Ophthalmology & Visual Science (Supplement)*, 39(4), S847.  
 Gibson, J. J. (1950). The perception of visual surfaces. *American Journal of Psychology*, 63, 367–384.  
 Gillam, B. (1970). Judgments of slant on the basis of foreshortening. *Scandinavian Journal of Psychology*, 11, 3134.  
 Gillam, B. (1995). The perception of spatial layout from static optical information. In W. Epstein, & S. Rogers, *Perception of space and motion*. New York: Academic Press.  
 Goodenough, B., & Gillam, B. (1997). Gradients as visual primitives. *Journal of Experimental Psychology: Human Perception and Performance*, 23(2), 370–387.  
 Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford.  
 Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non Fourier) texture channels. *Vision Research*, 38(2), 231–257.  
 Gray, A. (1998). *Modern differential geometry of curves and surfaces with mathematica* (2nd ed.). New York: CRC Press.  
 Griffiths, A. F., & Zaidi, Q. (1999). Perceptual assumption and projective distortions in a three-dimensional shape illusion. *Perception* (in press).  
 Heeger, D. J. (1992). Half-squaring in responses of cat striate cells. *Journal of Neuroscience*, 9, 427–443.  
 Ittelson, W. H. (1952). *The Ames demonstrations in perception*. Princeton, NJ: Princeton University Press.  
 Knill, D. C. (1998a). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research*, 38, 1655–1682.  
 Knill, D. C. (1998b). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research*, 38, 1683–1711.  
 Knill, D. C. (1998c). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, 38, 2635–2656.  
 Krumm, J., & Shafer, S. A. (1994). Segmenting textured 3D surfaces using the space/frequency representation. *Spatial Vision*, 8(2), 281–308.  
 Li, A., & Zaidi, Q. Three-dimensional shapes conveyed by natural textures. *Journal of the Optical Society of America A* (submitted).  
 Li, A. & Zaidi, Q. Three-dimensional shape from real and illusory contours (in preparation).

- Malik, J., & Rosenholtz, R. (1994). A computational model for shape from texture. *Ciba Foundation Symposium*, 184, 272–282.
- Mamassian, P., & Landy, M. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research*, 38, 2817–2832.
- Reed, T. R., & Wechsler, H. (1990). Segmentation of textured images and Gestalt organization using spatial/spatial-frequency representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), 1–12.
- Reed, T. R., & Wechsler, H. (1991). Spatial/spatial-frequency representations for image segmentation and grouping. *Image and Vision Computing*, 9(3), 175–194.
- Reichel, F. D., & Todd, J. T. (1990). Perceived depth inversion of smoothly curved surfaces due to image orientation. *Journal of Experimental Psychology*, 16(3), 653–664.
- Riley, B. (1995). In R. Kudielka, *Bridget Riley: Dialogues on Art*. London: Zwemmer.
- Rosinski, R. R., & Levine, N. P. (1976). Texture gradient effectiveness in the perception of surface slant. *Journal of Experimental Child Psychology*, 22, 261–271.
- Ross, J., & Speed, H. D. (1991). Contrast adaptation and contrast masking in human vision. *Proceedings of the Royal Society of London Series B*, 246, 61–69.
- Sagi, D. (1995). The psychophysics of texture segmentation. In T. V. Papathomas, *Early vision and beyond*. Cambridge: MIT Press.
- Sakai, K., & Finkel, L. H. (1993). Characterization of the spatial-frequency spectrum in the perception of shape from texture. *Journal of the Optical Society of America A*, 12(6), 1208–1224.
- Shepard, R. (1990). *Mind sights*. New York: W.H. Freeman.
- Stevens, K. A. (1981). The visual interpretation of surface contours. *Artificial Intelligence*, 17, 47–73.
- Todd, J. T., & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal Experimental Psychology, Human Perception and Performance*, 113(2), 221–224.
- Turner, M. R., Gerstein, G. L., & Bajcsy, R. (1991). Underestimation of visual texture slant by human observers: a model. *Biological Cybernetics*, 65, 215–226.
- Vickers, D. (1971). Perceptual economy and the impression of visual depth. *Perception and Psychophysics*, 10(1), 23–27.
- Voorhees, H., & Poggio, T. (1988). Computing texture boundaries from images. *Nature*, 333, 364–367.
- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17, 17–45.