

# Shape from Texture: Integrating Texture-Element Extraction and Surface Estimation

DOROTHEA BLOSTEIN, MEMBER, IEEE, AND NARENDRA AHUJA, SENIOR MEMBER, IEEE

**Abstract**—A perspective view of a slanted textured surface shows systematic changes in the density, area, and aspect-ratio of texture elements. These apparent changes in texture element properties can be analyzed to recover information about the physical layout of the scene. However, in practice it is difficult to identify texture elements, especially in images where the texture elements are partially occluded or are themselves textured at a finer scale. To solve this problem, it is necessary to integrate the extraction of texture elements with the recognition of scene layout. We present a method for identifying texture elements while simultaneously recovering the orientation of textured surfaces. A multiscale region detector, based on measurements in a  $\nabla^2 G$  (Laplacian-of-Gaussian) scale-space, is used to construct a set of candidate texture elements. True texture elements are selected from the set of candidate texture elements by finding the planar surface that best predicts the observed areas of the candidate texture elements. Results are shown for a variety of natural textures, including waves, flowers, rocks, clouds, and dirt clods.

**Index Terms**—Integration, multiscale structure, natural textures, perspective view, region detection, shape from texture, surface orientation, texture elements, texture gradients, texture homogeneity, three-dimensional vision.

## I. INTRODUCTION

TEXTURE variation due to projective distortion provides important cues for recovering the three-dimensional structure of the surfaces visible in an image [11]. A uniformly-textured surface undergoes two types of projective distortions during the imaging process. Firstly, an increase in the distance from the surface to the viewer causes a uniform compression of increasingly large areas of surface onto a fixed area of image. Secondly, foreshortening (due to the angle between the surface and the image plane) causes an anisotropic compression of the texture. These texture variations provide information about the relative distances and orientations of the textured surfaces in an image.

A primary goal of the work reported in this paper is to demonstrate the feasibility of extracting useful measures of texture gradients from images of natural scenes. A ma-

jor challenge in texture analysis is to identify texture scale consistently. Natural surfaces exhibit a rich hierarchy of textures and subtextures. All texture measurements are prone to distortion due to the presence of subtexture, since the imaging process captures more subtexture details for close texture samples than for distant ones. As discussed in Section II-A-1, existing shape-from-texture algorithms do not address the problem of scale. The algorithms presented here provide good surface-orientation estimates even in the face of significant sub- and supertexture.

## A. Texels and Texture Gradients

The term *texel*, short for *texture element*, denotes the repetitive unit of which a texture is composed. "Texel" refers to the physical texture element in the real world as well as to the appearance of the texture element in the image. In cases where the distinction must be made, we use the phrases *physical texel* versus *image texel*. Distance and foreshortening changes alter the image texel, but not the physical texel.

Projective distortion affects many texture features, and hence gives rise to a variety of texture gradients. Consider first the idealized texture of Fig. 1(a): a planar surface covered with nonoverlapping circular disks of constant size. The disks project as ellipses in the image. The major axis of each ellipse is perpendicular to the tilt,<sup>1</sup> whereas the minor axis is parallel with the tilt. Scanning the image from bottom to top (in the direction of tilt), the apparent size of the major axes decreases linearly, due to increasing distance from the viewer (the *perspective gradient*). However, the apparent size of the minor axes decreases quadratically: in addition to the distance scaling, the minor axes are foreshortened. Thus the eccentricity of the ellipses increases in the tilt direction (the *aspect-ratio gradient*). Similarly, the area of the ellipses decreases fastest in the direction of tilt (the *area gradient*). This is accompanied by an increase in the density of the ellipses (the *density gradient*). In this idealized texture, the uniformity in the size, shape and placement of the texture elements leads to pronounced texture gradients.

Manuscript received September 15, 1988; revised January 30, 1989. Recommended for acceptance by A. K. Jain. This work was supported by the Air Force Office of Scientific Research under Grant AFOSR 86-0009 and by the Eastman Kodak Company.

D. Blostein was with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801. She is now with the Department of Computing and Information Science, Queen's University, Kingston, Ont. K7L 3N6, Canada.

N. Ahuja is with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois, Urbana, IL 61801.

IEEE Log Number 8929956.

<sup>1</sup>We express surface orientation in terms of two angles, *slant* and *tilt* [23]. Slant, ranging from 0° to 90°, is the angle between the surface and the image plane. Tilt, ranging from 0° to 360°, is the direction in which the surface normal projects in the image; a tilt of 0° indicates that distance to the viewed surface increases fastest toward the right side of the image. The synthetic textures shown in Fig. 1(a) illustrate the definition of slant and tilt.

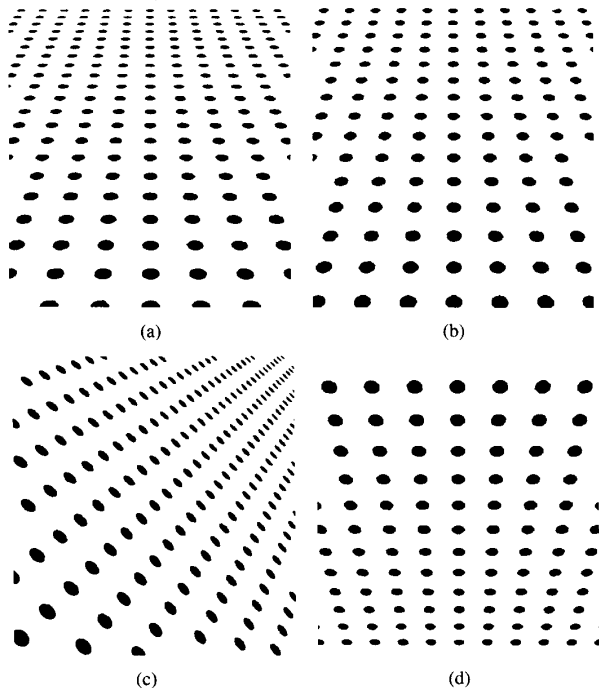


Fig. 1. Synthetic textures illustrating various slants and tilts. Slant is the angle between the textured surface and the image plane. Tilt is the direction in which the surface normal projects in the image. (a) Slant 60°, tilt 90°. (b) Slant 50°, tilt 90°. (c) Slant 60°, tilt 45°. (d) Slant 45°, tilt 270°.

Natural textures are much less regular than the idealized texture of Fig. 1(a); therefore the texture gradients are not as easily observed. Natural textures display considerable variability of texel size, shape, coloration and density. Physical texels are typically three-dimensional, unlike the disks portrayed in Fig. 1(a). Highlights, shadows, and occlusions between texels result. Also, physical texels have a complex structure. In contrast to a uniform synthetic disk, a physical texel changes in appearance as the distance to the camera decreases: more subtexture is visible for the nearby texels than for the distant texels. Supertexture regions arise when small image texels, corresponding to distant physical texels, blur into larger regions of relatively uniform gray level. These factors make it difficult to identify texture elements and extract texture gradients from real images.

The importance of various texture gradients has been studied extensively in the psychology literature (see, for example, [7], [19], [21], [22]). Vickers [24] was among the first to advocate an approach involving accumulation of evidence from multiple texture gradients. Cutting and Millard [9] attempt to quantify the relative importance of various texture gradients. They test human subjects on synthetically generated textures, which are designed to contain only a subset of the normally-occurring texture gradients. Experimental results show that for slant judgments of flat surfaces the perspective and density gradients are more important than the aspect-ratio gradient,

whereas in the perception of curved surfaces the aspect-ratio gradient is dominant, with perspective and density gradients having little impact.

## II. THE INFERENCE OF SURFACE SHAPE FROM TEXTURE GRADIENTS

We now turn to a discussion of the basic requirements for a system that infers surface-shape from texture gradients. In Section II-A we argue that correct measurement of texture gradients requires explicit identification of image texels, especially when textures show three-dimensional relief, when texels exhibit significant subtexture, or when it is unknown *a priori* which texture gradients carry surface-shape information. In Section II-B we address the problem of texel identification. Texture elements cannot be identified in isolated image areas since texels are defined only by the repetitive nature of the texture as a whole. Therefore, the identification of texture elements is best done in parallel with the estimation of the shape of the textured surface.

### A. The Importance of Texel Identification

The extraction of texels is an essential step in measuring texture gradients, because it permits correct analysis of textures containing subtexture. Explicit texel identification also provides the basis for a unified treatment of the various texture gradients (area gradient, density gradient, aspect-ratio gradient) that may be present in an image. Previous researchers have avoided texel identification because it is quite difficult to do in real images.<sup>2</sup> Instead, indirect methods are used to estimate texel features. We give below several examples of such methods, and indicate why these methods may give erroneous results.

1) *Previous Work:* Most previous shape-from-texture algorithms use indirect methods to estimate texel features, by making some assumptions about the nature of texture elements. For example, texel density may be estimated by measuring edge density, under the assumption that all detected edges correspond to the borders of texture elements [1]<sup>3</sup>, [2], [14], [20]. Alternatively, texture elements may be assumed to have uniform edge direction histograms; surface orientation can then be estimated from any deviations from isotropy observed in the distribution of edge directions [10], [13], [26]. However, the directional-isotropy assumption is very restrictive; for example, it does not hold true in images containing elongated texels such as waves. Texture coarseness and directionality may be characterized using Fourier domain measurements [3], ignoring the effect of super- and subtextures. Various researchers [12], [15], [16] have developed

<sup>2</sup>Ohta *et al.* [18] use the observed areas of pairs of texels to obtain vanishing points. However, the method has been tested only on synthetic texture images. The problem of extracting texels from natural images is not addressed.

<sup>3</sup>The theoretical analysis in this paper is based on texel area. In application to real images, edge-density is used instead, under the assumption that the edge detector finds texel boundaries.

algorithms to analyze textures containing parallel and perpendicular lines. Most natural textures are too irregular to be analyzed in this way.

All of these methods may encounter problems when applied to complex natural textures seen under natural lighting conditions. Since texels are not identified and explicitly dealt with, it becomes difficult to distinguish between responses due to texels and those due to other image features, such as subtexture. It appears to be necessary to recognize the texture elements before the various measures can be computed as intended.

Consider, for example, methods based on measuring edge density. If these algorithms are applied to edges produced by an edge-detector, the measurements are made inaccurate by contributions from subtexture and supertexture edges. Fig. 2(a)–(c), which shows the response of an edge detector to several texture images,<sup>4</sup> illustrates that it would be incorrect to interpret all of the detected edges as boundaries of texture-elements. Additional edges result from subtexture; these edges are not artifacts of this particular edge detector, since they are clearly present in the original images. Many natural textures have a hierarchical physical structure that causes observed edge density to be nearly constant throughout the image: edges from subtexture and subsubtexture are observed to whatever detail the camera resolution permits.

In order to measure edge-density as intended by [2] and [14], it is necessary to eliminate subtexture edges. This cannot simply be done by applying a global threshold, since the contrast of texels far from the camera is comparable to the contrast of subtexture features in the foreground. Aloimonos [2] distinguishes between a “strong segmentation” (finding texels) and a “weak segmentation” (finding edges, where the edges are supposed to be texel boundaries), and states that weak segmentation is easy to obtain (apply any general-purpose edge detector). We argue that correct weak segmentation is not possible without simultaneously performing a strong segmentation: in order to eliminate all edges except those that arise from texel boundaries, one has to in effect identify the texels.

2) *Multiple Texture Gradients*: Explicit texel identification offers an additional advantage: texels provide a unifying framework for examination of the various texture gradients (such as gradients of apparent texel area, aspect ratio, density etc.) that may be present in an image. A given image may exhibit a combination of texture gradients. In general, the accuracy of the surface information obtainable from these gradients varies from image to image.<sup>5</sup> Since it is not known in advance which texture gra-

dients are useful for determining the three-dimensional orientation of surfaces, a shape-from-texture system should evaluate the information content of different types of gradients in a given image, and use an appropriate mix of these gradients for surface estimation.

### B. Integration of Texel Identification and Surface-Shape Estimation

Texel identification is difficult because texels have tremendously varied shapes, sizes and gray-level characteristics. A texel cannot be identified in isolation, since texels are only defined by the repetitive nature of the texture as a whole. In order to determine if an image region is a texel, it is necessary to test if the region has properties consistent with the properties of many other image texels, i.e. whether the image region is part of a *texture field*.

We use the term *texture field* (or *field of texels*) to denote a collection of image texels that exhibit one or more consistent texture gradients. Consistency is defined with respect to a perspective view of a given surface. It is not uncommon for a single image to contain several texture fields. First, many images are composed of closely associated bright and dark texture fields which arise from lighting effects. For example, the aerial view of houses in Fig. 5(a) contains a field of bright texels composed of the houses and a field of dark texels composed of the shadows cast by the houses. Second, it is possible for physically separated textured surfaces to be spatially interleaved in an image. This is strikingly illustrated by the birds-over-water image shown in Fig. 7(a). Finally, the same physical surface may contain different texture fields: an aerial view of a residential neighborhood shows one texture field consisting of houses and another texture field consisting of trees.

Texels can only be identified in the context of a texture field, where consistent texture gradients must exist across the whole field. The consistency of a texture gradient can only be evaluated for a particular surface shape and orientation. Thus, texel identification must be combined with surface estimation.

### C. Overview

Motivated by the above discussion, we now summarize the requirements for an ideal shape-from-texture algorithm, and the extent to which our work meets these requirements. Many open problems remain.

<sup>4</sup>We use an edge operator described by Nevatia and Babu [17]. Six 5-by-5 edge masks at different orientations are used; the mask giving the highest output at each pixel is recorded. The edges are thinned by suppressing non-maxima perpendicular to the edge directions.

<sup>5</sup>This may be illustrated by the following examples. It is common for physical texels to be fairly uniform in size and shape, but for the gaps between the texels to be much less uniform [Figs. 7(a), 19(a), 23(a), and 25(a)]. In these images, it is more accurate to infer a three-dimensional

surface from the area and aspect-ratio gradients than from the density gradient or the gradient of spacings between texels. As a second example, the potential accuracy of the aspect-ratio gradient is higher in textures where the physical texels are separated by gaps than in textures where the physical texels overlap and occlude one another [the lily pads in Fig. 25(a) show a much better aspect-ratio gradient than do the rocks in Fig. 3(a)]. Thirdly, for the water hyacinths of Fig. 29(a), the random three-dimensional arrangement of the leaves makes the aspect-ratio gradient very weak, while the area gradient is still quite significant. Finally, in images with partial occlusions [Figs. 13(a) and 15(a)], the perspective gradient (length of the unforeshortened texel dimension) is more accurate than the area gradient: if only part of a texel is occluded, the apparent texel area is decreased, whereas the complete unforeshortened dimension may remain in view.

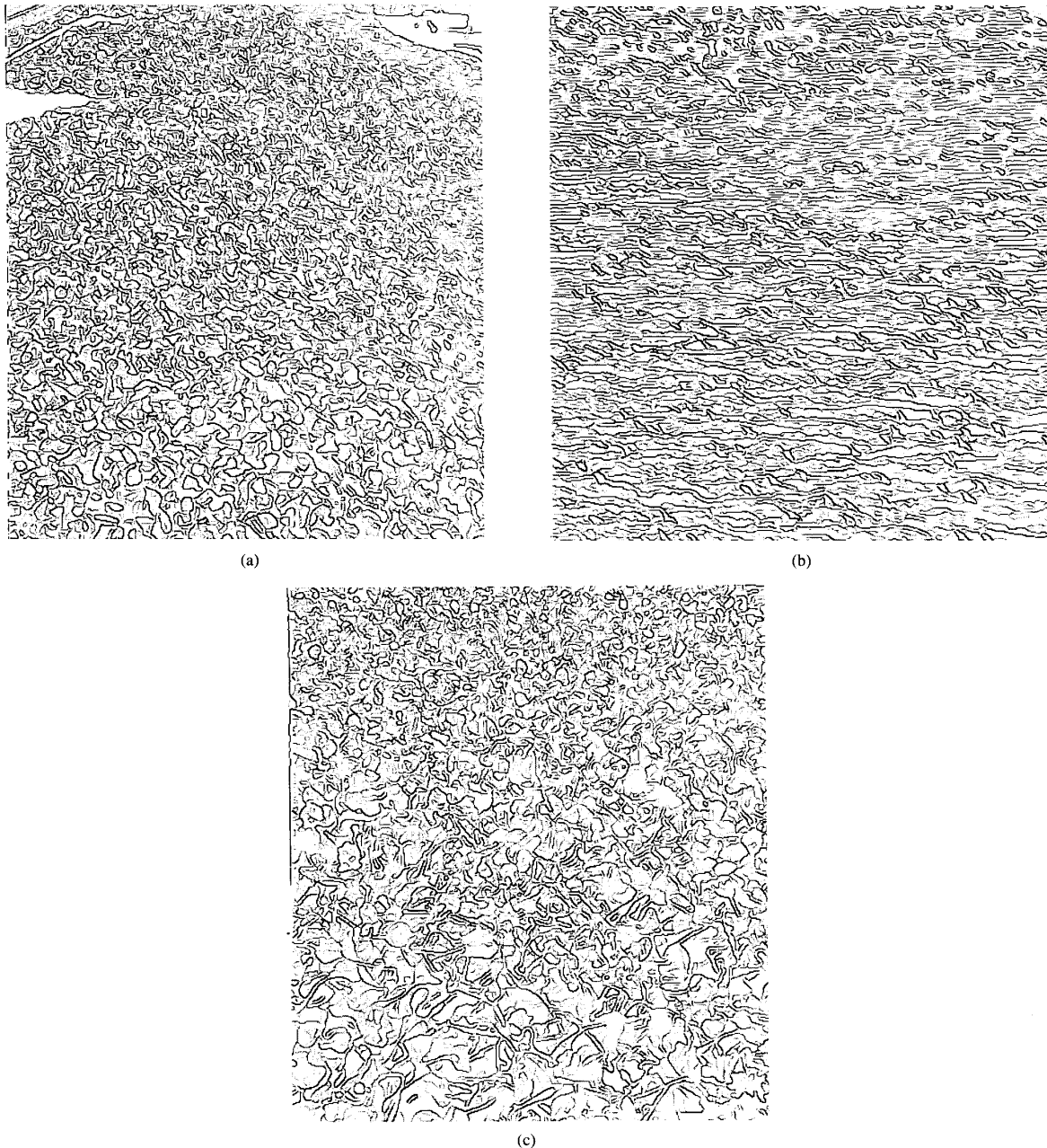


Fig. 2. Edges extracted from three texture images. Only a subset of the detected edges are boundaries of texture elements. If edge density is to be effective in capturing the texture gradient, all edges that do not correspond to texel boundaries must be removed. Such edge removal cannot be accomplished without, in effect, performing an identification of texture elements. (a) Edges from the rock-pile image shown in Fig. 3(a). (b) Edges from the image of birds flying over water, shown in Fig. 7(a). (c) Edges from the water hyacinths image shown in Fig. 29(a).

*Texel Identification:* As discussed in Section II-A, texel identification is important for correct shape-from-texture analysis. In general, physical texels can give rise to complex gray-level patterns. However, it is difficult to test for

repetitive patterns of arbitrary gray-level configuration. In our implementation, we restrict image texels to be regions that have small gray-level variation relative to a neighborhood of their size. Under this restriction, a physical

texel can give rise to several image texels: often the physical repetitive unit of a texture contains both bright and dark regions [for example, the sunflowers in Fig. 15(a)]. We treat the bright and dark image texels as two texture fields, which we analyze separately.

*Texture Gradients:* A shape-from-texture system should test each image for the presence of various texture gradients (area gradient, aspect-ratio gradient, density gradient), and combine these various sources of information to produce a surface estimation. In our current implementation, the only texture gradient we test for is a gradient in texel area.<sup>6</sup> However, we do independent analyses of the area-gradients in positive-contrast and negative-contrast image regions; these two types of regions may correspond to foreground and background, or to different portions of the physical texels. Some images contain bright texels on dark backgrounds, other images contain dark texels on bright backgrounds, and yet other images have no visible "background" region because texels are densely placed. Measurements of foreground regions are often more accurate than measurements of background regions, since texel area tends to be less variable than texel spacing.

*Surface Estimation:* Ideally, a system tests for texture gradients produced by a variety of surface shapes (planar, cylindrical, spherical etc.), and is able to locate discontinuities in depth and surface orientation. Much work remains to be done in this area. We restrict ourselves to fitting a single planar surface to the entire image. This is a common restriction in current implementations of shape-from-texture algorithms, although a theoretical treatment of the nonplanar case has been performed by [1], and has been tested on synthetic images.

*Three-Dimensional Texel Effects:* For accurate image analysis, it is necessary to model the effects of three-dimensional relief on observed texture gradients. We do not address this problem in our current implementation: the equations we use for expected area-gradients are derived under the assumption that texels do not have three-dimensional relief, i.e., that they are "painted" on the textured surface. For textures with relief, this assumption results in a redefinition of texels: ideally, only those parts of physical texels that are parallel to the underlying surface are recognized as defining a texture field. The detected consistency is reduced if the texels have parts that are not parallel to the surface. If the texel relief is regular (so that the texels are mutually parallel), the detected gradient may still be significant and nearly correct recovery of surface orientation may be possible. For example, we obtain satisfactory results on the sunflower image of Fig. 15, where the texels are not parallel to the surface.

Although some theoretical treatment of the foreshortening of textures with relief exists (e.g., [15]), no one has

addressed how in-plane texels could be distinguished from out-of-plane texels in real images.

*Multiple Textures:* A general shape-from-texture system must be able to handle images containing multiple textures; the system must perform texture segmentation as well as shape-from-texture estimation. To solve this problem it is necessary to separate texture variations due to distance and foreshortening effects from texture variations due to a boundary between different physical textures. Our current implementation does not address this problem: each of our images contains only a single texture. Ongoing research is aimed at extending the method to apply to images containing multiple textures.

The rest of the paper describes our two-step algorithm for texel identification and surface estimation. In the first step, we use a multiscale region detector to construct a set of candidate texels (Section III). In the second step we use surface-fitting to identify the true texels from among the candidates, while simultaneously constructing an approximation to the shape of the textured surface (Section IV). The second step thus enforces perspective viewing constraints to select texels. Section V presents results for a variety of images of textured natural scenes.

### III. IDENTIFYING CANDIDATE TEXELS: MULTISCALE REGION DETECTION

We now turn to a description of the multiscale region detector used to construct the set of candidate texels. The set of candidate texels includes all image regions that have small gray-level variation relative to a neighborhood of their size. These image regions may be of any shape and size, and they may be nested, since there is no *a priori* way to distinguish texture regions from subtexture and super-texture regions.

To simplify the problem of extracting regions of arbitrary shapes and sizes, we assume that each region can be represented as a union of overlapping circular disks. Large disks define the rough shape of a region, with overlapping smaller disks capturing finer shape details such as protrusions and concavities. In Section III-A we derive a method of extracting all circular image regions of relatively uniform gray level. Section III-B discusses how sets of overlapping disks are used to form candidate texels.

The region detector is based on the image response to convolution with  $\nabla^2 G$  filters over a range of scales. Related work includes [27] (a scale-space representation of  $\nabla^2 G$  zero-crossings) and [8] (a representation of  $\nabla^2 G$  peaks and ridges over a range of scales<sup>7</sup>). We find circular image regions of uniform gray level by convolving the image with  $\nabla^2 G$  masks over a range of scales, and comparing the convolution output to that expected for an ideal circular disk of constant gray level. Here we present a brief summary of the region detection algorithm; a more detailed discussion may be found in [6].

<sup>6</sup>Our extraction of texel shape is not accurate enough to permit useful measures of aspect-ratio. We also cannot measure texel density accurately because we do not extract all of the texels. Ongoing research into improved texel-extraction will permit the analysis of several texture gradients.

<sup>7</sup>Crowley and Parker use a difference-of-Gaussian operator, which is a discrete approximation to  $(\partial/\partial\sigma)G$  and hence to  $\nabla^2 G$ . By the diffusion equation,  $\nabla^2 G = (1/\sigma)(\partial/\partial\sigma)G$ .

### A. A Closed Form Expression for the $\nabla^2 G$ Response of a Disk

The algorithm for uniform-region extraction is based on calculations of the  $\nabla^2 G$  and  $(\partial/\partial\sigma)\nabla^2 G$  responses of a disk image. Given a function  $I(x, y)$  which specifies the intensity of an image, the  $\nabla^2 G$  response of this image at  $(x, y)$  is given by the following convolution:

$$\begin{aligned} \nabla^2 G(x, y) * I(x, y) \\ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{2\sigma^2 - (u^2 + v^2)}{\sigma^4} e^{-(u^2 + v^2)/2\sigma^2} \\ \cdot I(x - u, y - v) du dv. \end{aligned} \quad (1)$$

Mathematical analysis of the response of the  $\nabla^2 G$  filter to most images is difficult because the convolution integrals of (1) do not have closed form solutions. However, a closed-form solution can be derived for the center point of a circular disk of constant intensity. We analyze the  $\nabla^2 G$  response at the center of an ideal circular disk in the continuous domain; to generate the  $\nabla^2 G$  convolution of digitized images, we sample the  $\nabla^2 G$  filter values and perform a discrete convolution. The image of a disk of diameter  $D$  and contrast  $C$  is defined by

$$\text{disk image: } I(x, y) = \begin{cases} C & \text{if } x^2 + y^2 \leq D^2/4 \\ 0 & \text{elsewhere.} \end{cases} \quad (2)$$

Using this definition of  $I(x, y)$  in (1), and setting  $x$  and  $y$  to zero, we find [6] that at the disk center

$$\nabla^2 G \text{ response} = \frac{\pi C D^2}{2\sigma^2} e^{-D^2/8\sigma^2} \quad (3)$$

$$\frac{\partial}{\partial\sigma} \nabla^2 G \text{ response} = \frac{\pi C D^2}{2} \left( \frac{D^2}{4\sigma^5} - \frac{2}{\sigma^3} \right) e^{-D^2/8\sigma^2}. \quad (4)$$

Dividing these expressions, we solve for the diameter  $D$  and contrast  $C$  of the disk:

$$\begin{aligned} D &= 2\sigma \sqrt{\sigma \left( \frac{\partial}{\partial\sigma} \nabla^2 G * I \right) / (\nabla^2 G * I) + 2} \\ C &= \frac{2\sigma^2}{\pi D^2} e^{D^2/8\sigma^2} (\nabla^2 G * I) \end{aligned} \quad (5)$$

where the convolutions are evaluated at the center of the disk.

### B. Extracting Candidate Texels in Real Images

We construct an approximation of image texels by first fitting disks to uniform image regions, and then forming unions of connected disks. (An alternative approach is presented by [25]. They extract texture elements by convolving the image with a  $\nabla^2 G$  filter and then selecting components of above-threshold pixels that have suitable geometrical properties, such as compactness.) For the first step, we use (5) to estimate disk diameter and disk contrast from the  $\nabla^2 G * I$  and  $(\partial/\partial\sigma)\nabla^2 G * I$  values at the

center of a region. Disks are fit at the extrema of the  $\nabla^2 G * I$  images. The disks fit to local maxima have positive contrast (regions brighter than the surround), whereas the disks fit to local minima have negative contrast (regions darker than the surround).

We use a range of filter sizes. For a region  $R$  in image  $I$ , local extrema in  $\nabla^2 G * I$  occur at the center of  $R$  when the  $\nabla^2 G$  filter size approximately matches the region-diameter. Thus, to fit disks as accurately as possible, we accept a disk only if the computed diameter  $D$  is close to the  $\nabla^2 G$  filter size used to detect the disk.

Parts (b) of Figs. 3–30 illustrate the result of this disk-fitting for the positive-contrast and negative-contrast regions of each image. Implementation details for the disk-fitting are as follows.

1) Compute the convolutions  $\nabla^2 G * I$  and  $(\partial/\partial\sigma)\nabla^2 G * I$  for the following six  $\sigma$  values:  $\sqrt{2}$ ,  $2\sqrt{2}$ ,  $3\sqrt{2}$ ,  $4\sqrt{2}$ ,  $5\sqrt{2}$ , and  $6\sqrt{2}$ . (The center lobes of the six  $\nabla^2 G$  filters have diameters of 4, 8, 12, 16, 20, and 24 pixels, respectively.) To compute  $\nabla^2 G * I$  for a particular  $\sigma$  value, the image is convolved with a mask whose coefficients are taken from

$$\frac{2\sigma^2 - r^2}{\sigma^4} e^{-r^2/2\sigma^2}.$$

To compute  $(\partial/\partial\sigma)\nabla^2 G * I$  for a particular  $\sigma$  value, the image is convolved with a mask whose coefficients are taken from

$$\frac{6r^2\sigma^2 - r^4 - 4\sigma^4}{\sigma^7} e^{-r^2/2\sigma^2}.$$

2) Mark the locations where disks will be fit. To analyze the positive-contrast regions of the original image, mark all local maxima in the  $\nabla^2 G * I$  images. To analyze the negative-contrast regions of the original image, mark all local minima in the  $\nabla^2 G * I$  images. Local maxima and minima are computed relative to a  $3 \times 3$  neighborhood.

3) At each marked location, use the measured  $\nabla^2 G * I$  and  $(\partial/\partial\sigma)\nabla^2 G * I$  values in (5) to compute a disk diameter and disk contrast. Accept the disk only if  $w - 2 \leq D \leq w + 2$ , where  $w = 2\sqrt{2}\sigma$  is the diameter of the center lobe of the  $\nabla^2 G$  filter. Form a single set of disks by taking the union of the disks detected at the various filter sizes.

After the disk-fitting is completed, we form unions of overlapping disks to construct candidate texels. Overlapping disks form concavities. There is no *a priori* way to tell whether a set of disks should be split at a concavity or not: some concavities arise at the border between two neighboring texels; at other times the concavities are part of the shape of an individual texel. Thus both possibilities are included in the list of candidate texels. The implementation details are as follows:

To form the list of candidate texels, extract all subsets of disks that are spatially connected and contain no concavities greater than  $90^\circ$ . If a concavity is in

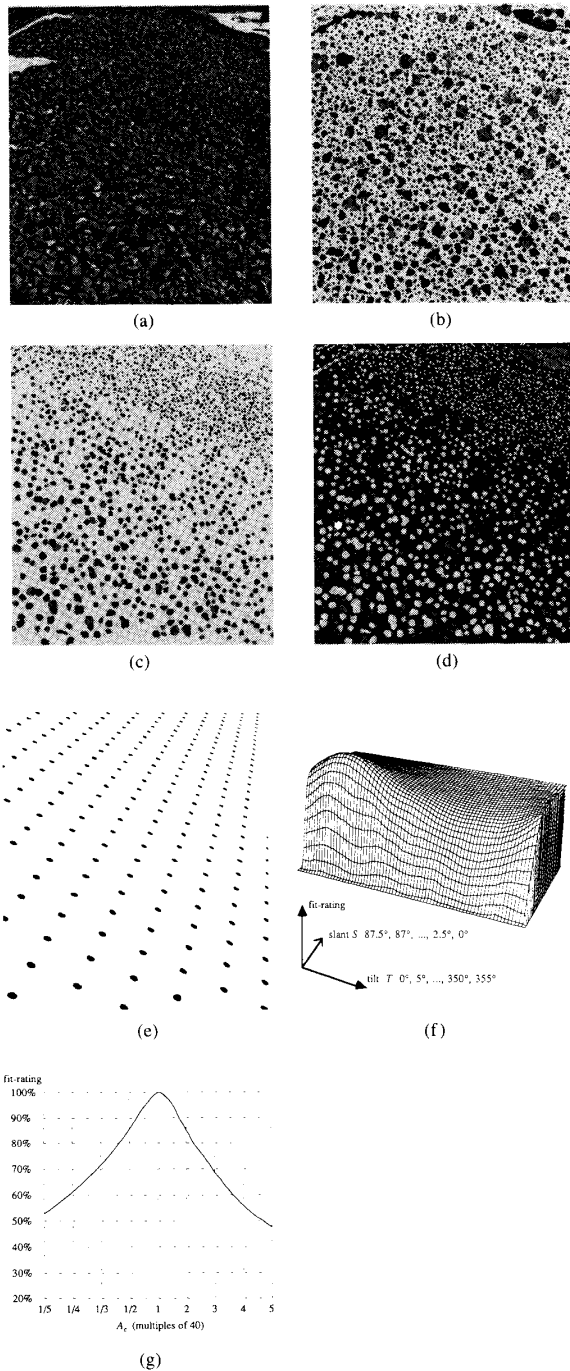


Fig. 3. (a) A rock pile. (b) Disks corresponding to positive-contrast regions of relatively uniform gray level. Disks are shown with a darkness proportional to the contrast of the region. (c) Extracted texels. These are all regions (sets of overlapping disks) having area within a factor of two of the area expected by the best planar fit ( $A_c = 40$ , slant  $62.5^\circ$ , tilt  $65^\circ$ ). The texels that fit the plane most closely are printed darkest. (d) The texels superimposed on a dark reproduction of the original. (e) Synthetic image to illustrate the planar fit  $A_c = 40$ , slant  $62.5^\circ$ , tilt  $65^\circ$ . (f) and (g) Rating of various possible planar fits. In (f) slant and tilt are varied while  $A_c$  is constant at 40. In (g)  $A_c$  is varied while slant and tilt are constant at  $62.5^\circ$  and  $65^\circ$ , respectively.

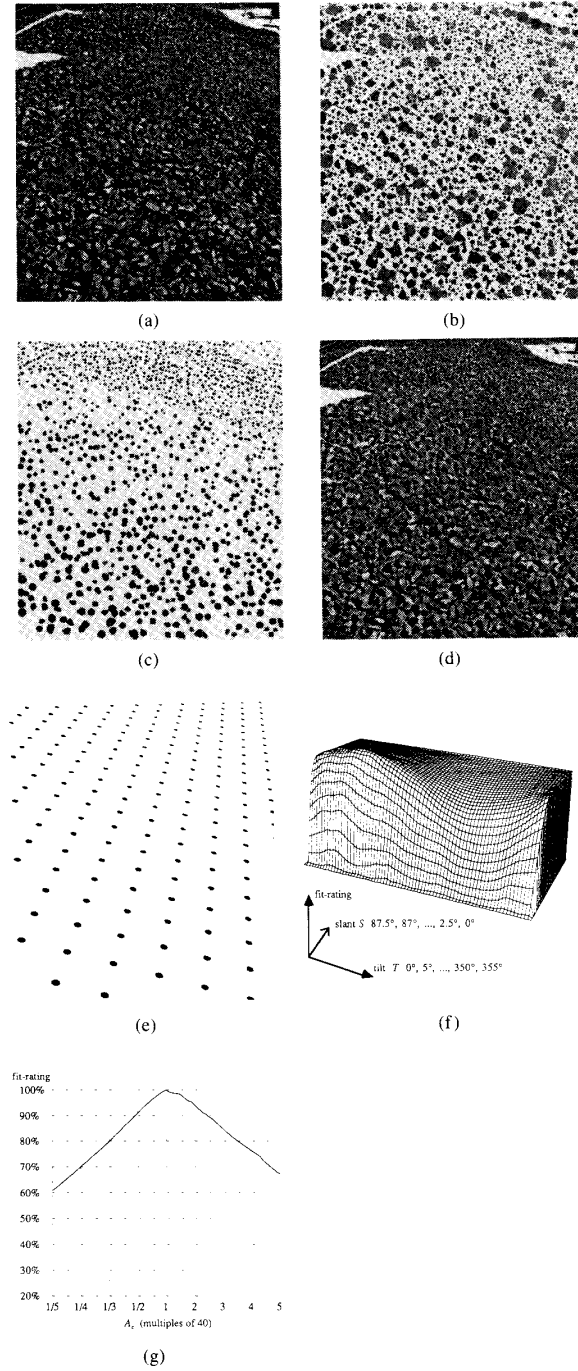


Fig. 4. (a) A rock pile. (b) Disks corresponding to negative-contrast regions of relatively uniform gray level. Disks are shown with a darkness proportional to the contrast of the region. (c) Extracted texels. These are all regions (sets of overlapping disks) having area within a factor of two of the area expected by the best planar fit ( $A_c = 40$ , slant  $60^\circ$ , tilt  $75^\circ$ ). The texels that fit the plane most closely are printed darkest. (d) The texels superimposed on a bright reproduction of the original. (e) Synthetic image to illustrate the planar fit  $A_c = 40$ , slant  $60^\circ$ , tilt  $75^\circ$ . (f) and (g) Rating of various possible planar fits. In (f) slant and tilt are varied while  $A_c$  is constant at 40. In (g)  $A_c$  is varied while slant and tilt are constant at  $60^\circ$  and  $75^\circ$ , respectively.



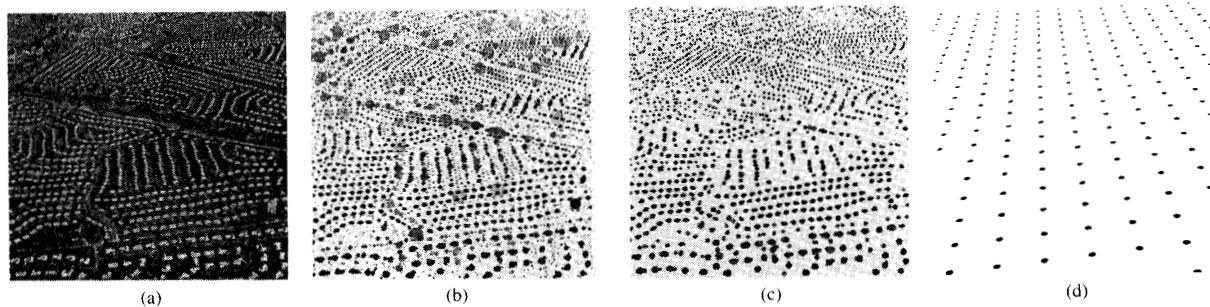


Fig. 5. Aerial view of Levittown, PA: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

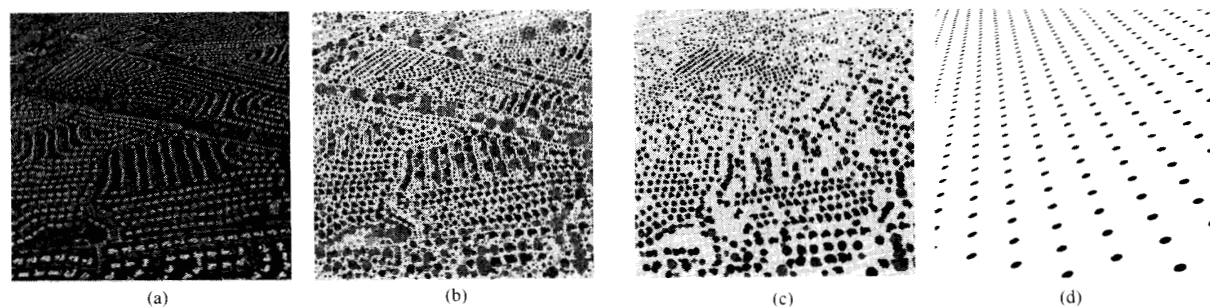


Fig. 6. Aerial view of Levittown, PA: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

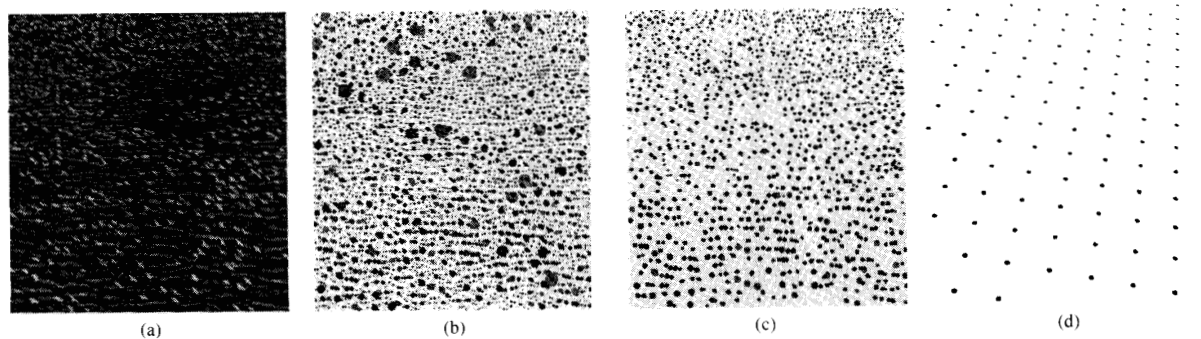


Fig. 7. Snow geese over Back Bay: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

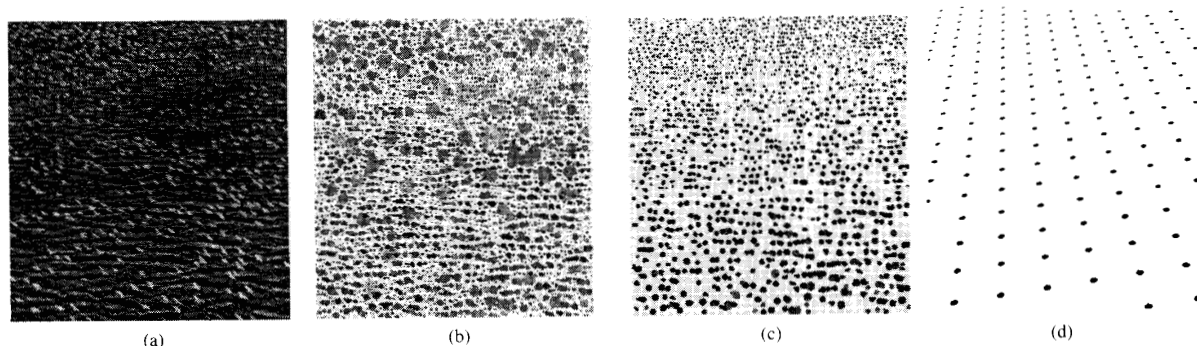


Fig. 8. Snow geese over Back Bay: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.



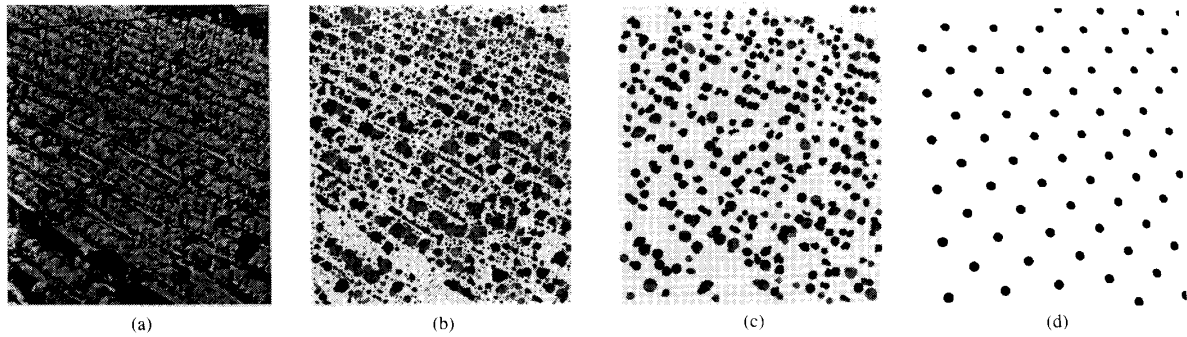


Fig. 9. Prayer at a mosque: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

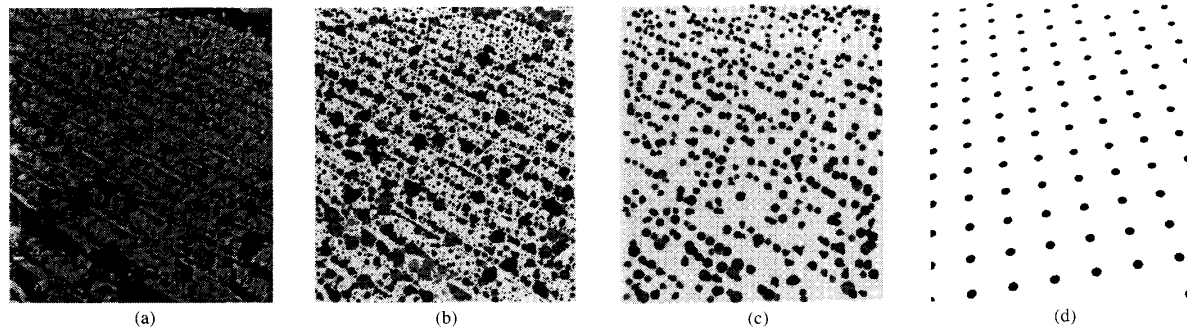


Fig. 10. Prayer at a mosque: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

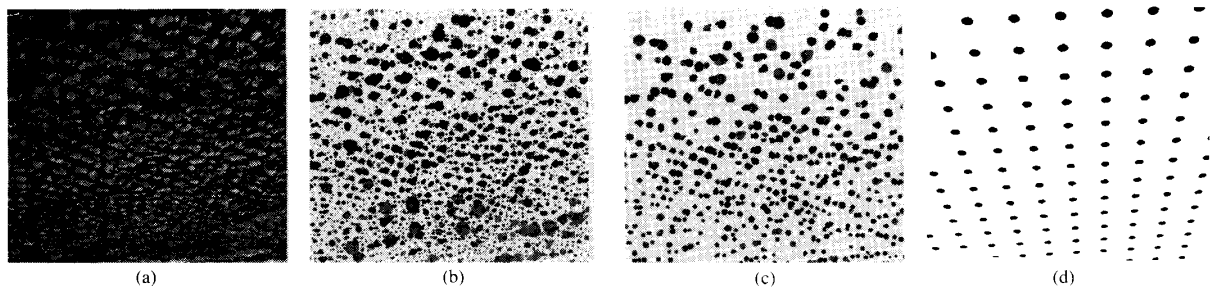


Fig. 11. Fleecy clouds: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

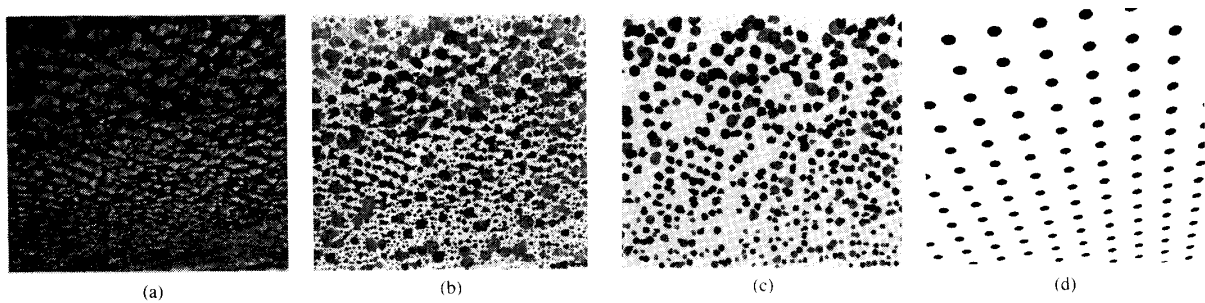


Fig. 12. Fleecy clouds: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

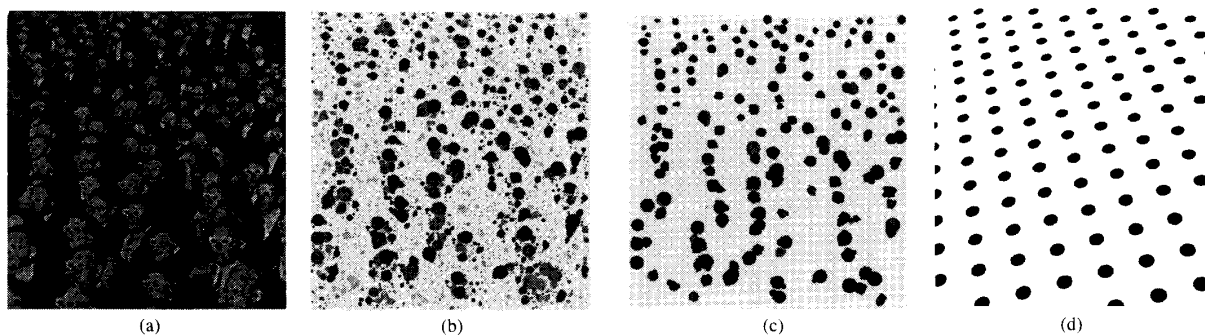


Fig. 13. Audience at a 3-D movie: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

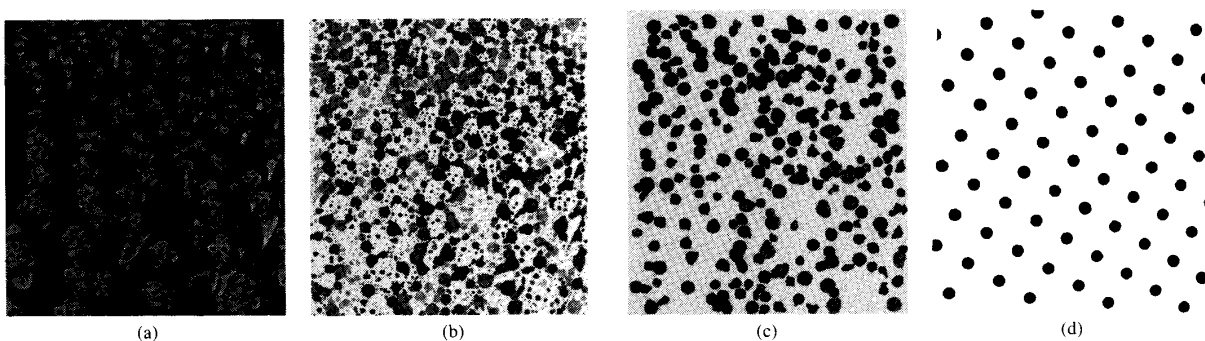


Fig. 14. Audience at a 3-D movie: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

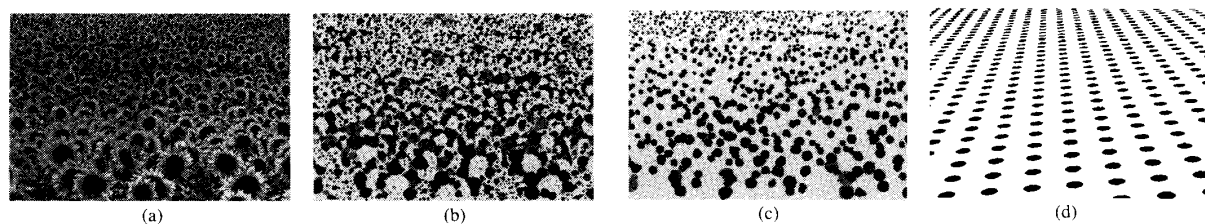


Fig. 15. Sunflowers: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

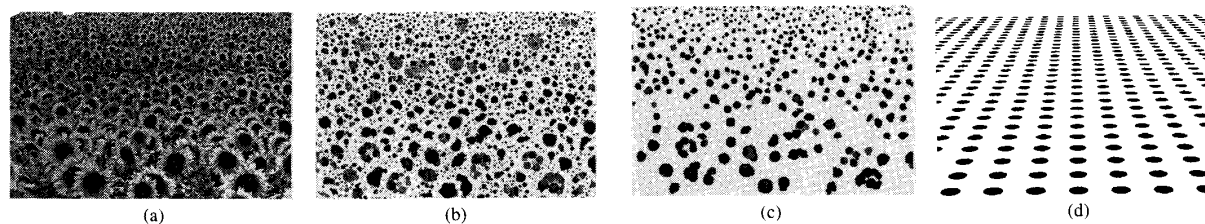


Fig. 16. Sunflowers: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

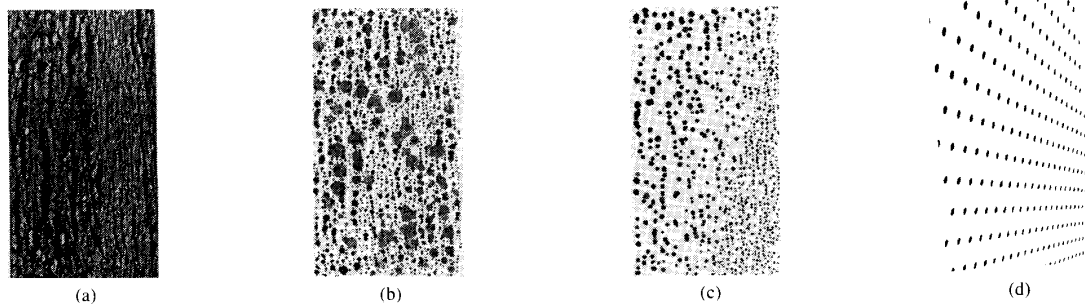


Fig. 17. Tree trunk: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

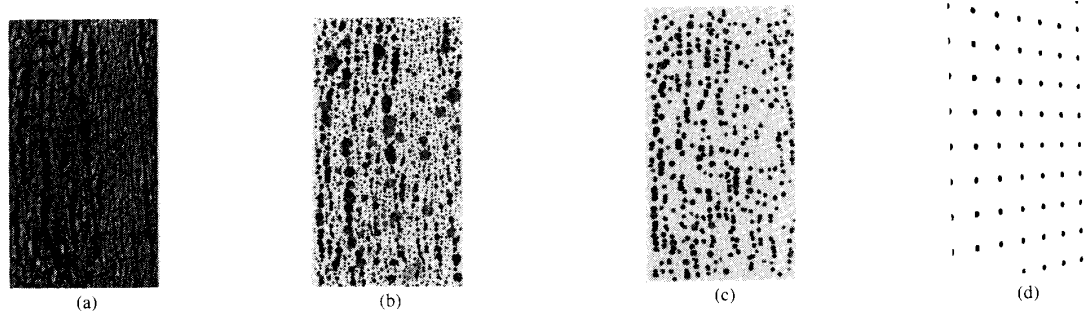


Fig. 18. Tree trunk: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

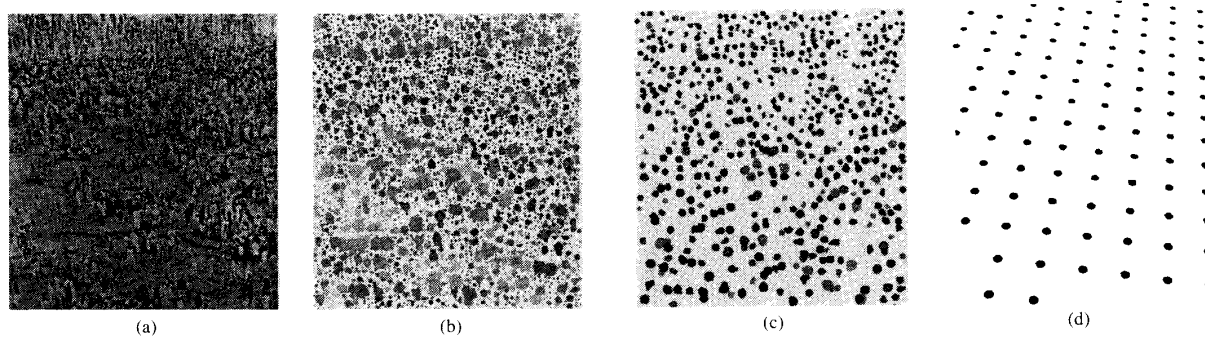


Fig. 19. Bathers on the Ganges: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

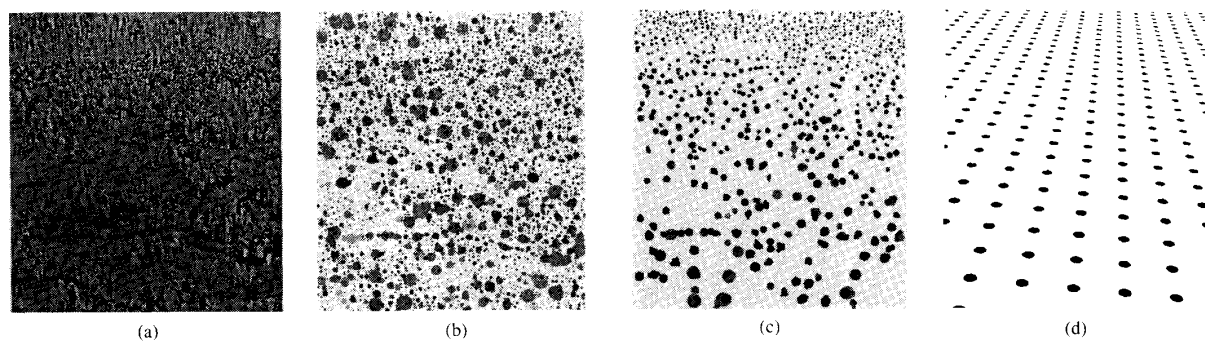


Fig. 20. Bathers on the Ganges: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

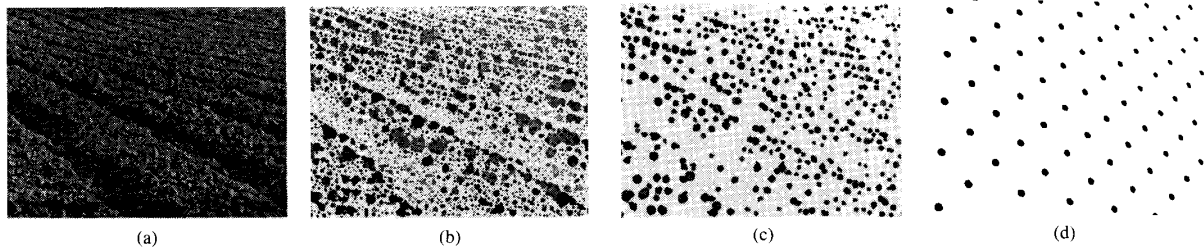


Fig. 21. A plowed field: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

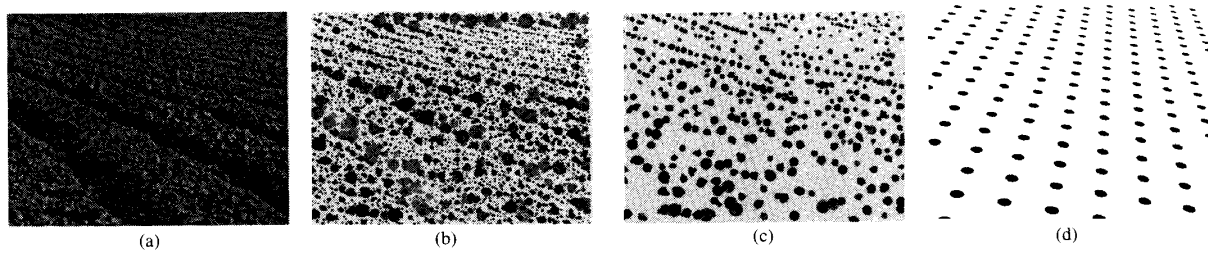


Fig. 22. A plowed field: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

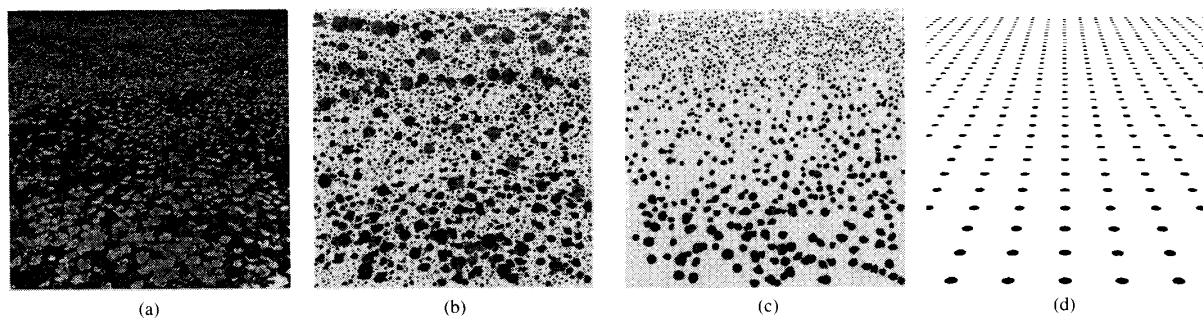


Fig. 23. A field of flowers: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

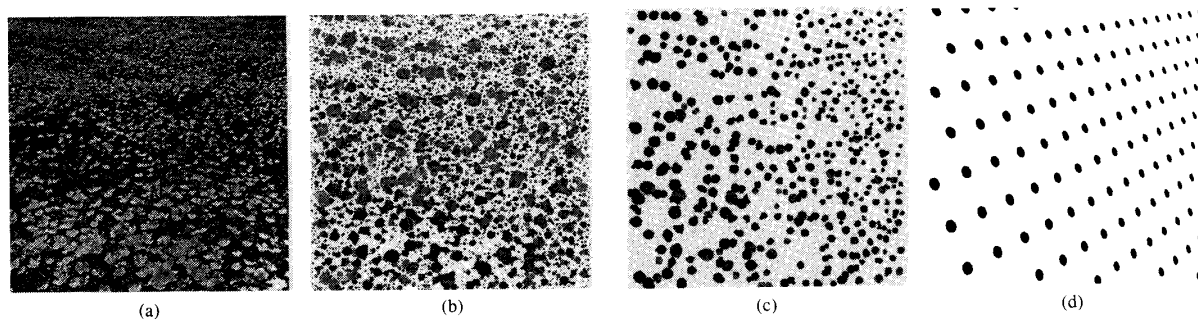


Fig. 24. A field of flowers: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

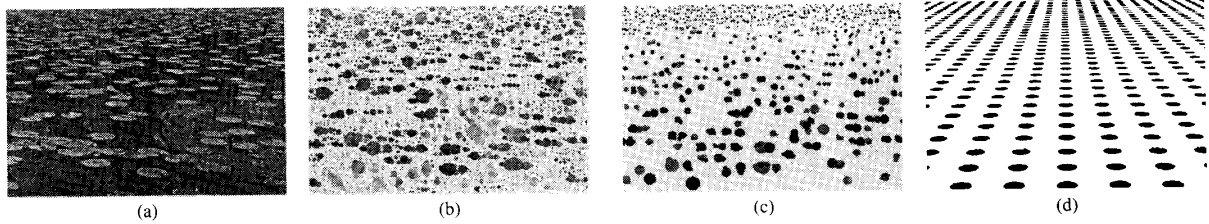


Fig. 25. Water lilies: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

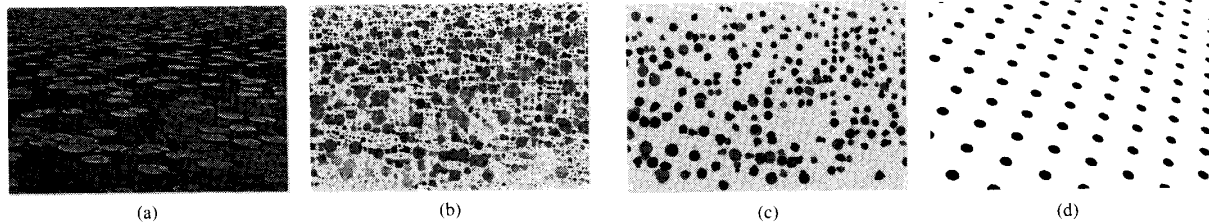


Fig. 26. Water lilies: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

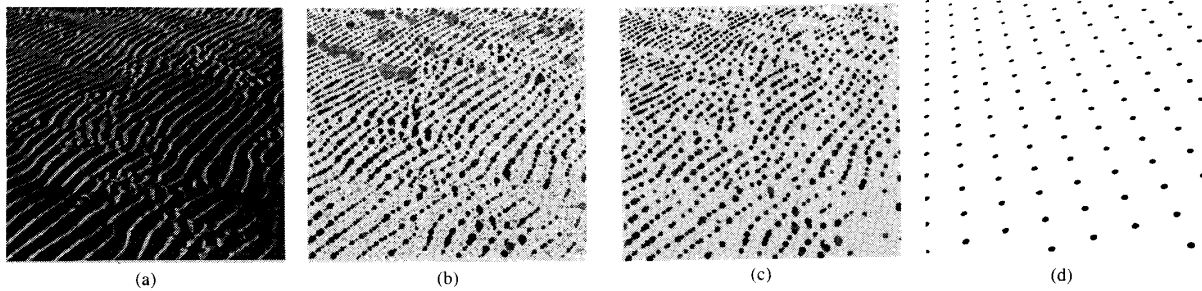


Fig. 27. Ripple marks in a shallow sea: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

the range  $50^\circ$ – $90^\circ$ , use the disks to form three candidate texels<sup>8</sup>: one large region consisting of all the disks, and two smaller regions resulting from splitting the large region at the concavity.<sup>9</sup> Mark mutual exclusion between candidate texels that share a disk: at most one of them can contribute support to a planar fit and be chosen as a true texture element.

<sup>8</sup>The particular values  $50^\circ$  and  $90^\circ$  are not critical; we have found that the range  $50^\circ$ – $90^\circ$  is large enough to capture the regions of interest and yet small enough to prevent a combinatorial explosion in the number of candidate texels generated.

<sup>9</sup>Region splitting is implemented as follows. We begin with a set  $P$  of overlapping disks, which together cover an image region  $R$ . The largest concavity in  $R$  is found by computing the angles formed by every pair of neighboring disks on the border of  $R$ . Suppose that  $X$  and  $Y$  are two neighboring disks on the border of  $R$ , and that they form a concavity that causes a split into smaller, more convex regions. The concavity is split by 1) removing  $X$  from  $P$  and repeating the above process, and then 2) removing  $Y$  from  $P$  and repeating the above process.

#### IV. SURFACE ESTIMATION AND TEXEL IDENTIFICATION

Our goal in analyzing image texture is to find a spatial layout of homogeneously-textured surfaces that could result in the given image texture. We do this by testing many spatial layouts and choosing the one that is the most consistent with a maximal subset of the candidate texels. The surface parameters are determined at the same time that the true texels are chosen from among the candidates.

##### A. The Expected Distribution of Texel Areas for a Planar Surface

The current implementation is restricted to fitting a single planar surface to the image, based on the observed areas of the candidate texels. In order to find a planar fit to the candidate texels, we need to know the distribution of texel areas that occurs in an image of an idealized textured plane. To derive this relationship, we assume a

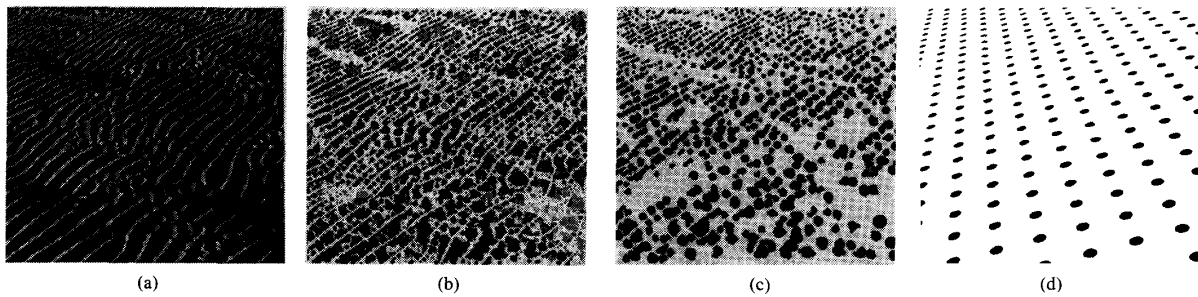


Fig. 28. Ripple marks in a shallow sea: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

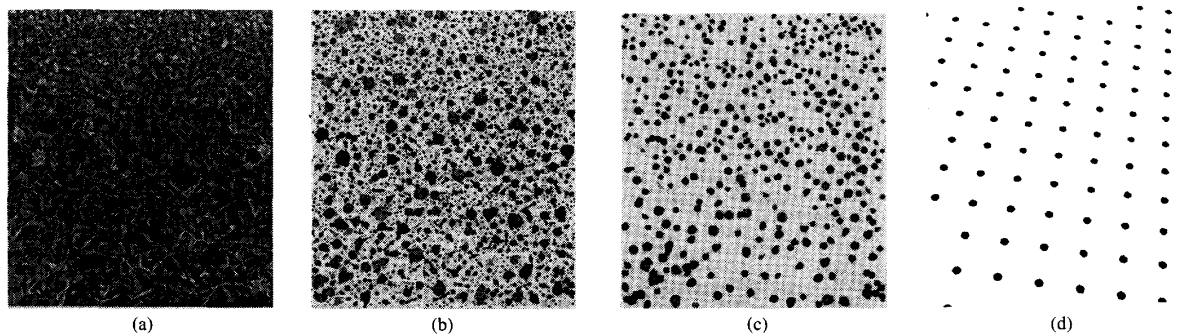


Fig. 29. Water hyacinths: positive contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

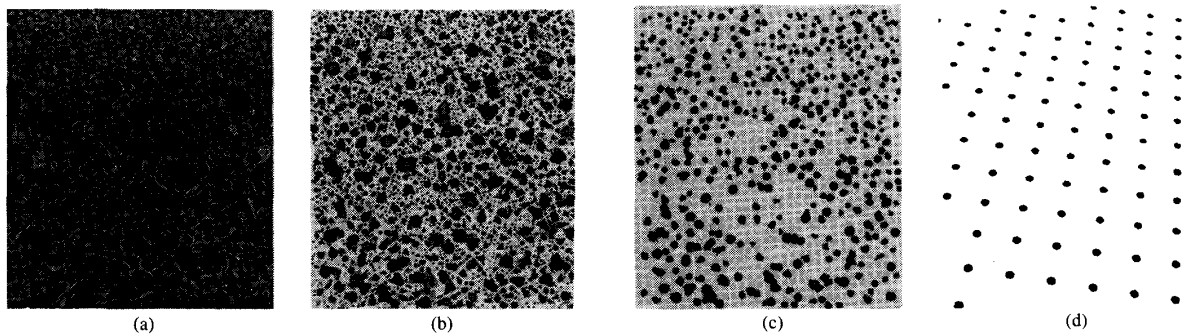


Fig. 30. Water hyacinths: negative contrast texture. (a) Original image, (b) regions detected, (c) texels extracted, and (d) synthetic display of recovered orientation.

planar textured surface covered with identical texels, where the texels show no three-dimensional relief (the texels are "painted" on the surface). Natural textures are typically more complicated: they are composed of highly variable texels that show three-dimensional relief. Our experiments show that the equations derived from consideration of idealized textures are useful for analyzing a variety of natural textures as well (Section V).

We derive two expressions to describe the size of image texels. The first expression characterizes the foreshortened image-texel dimension  $F_i$ ; this is the texel dimension

parallel to the tilt (Fig. 31). The second expression characterizes the unforeshortened image-texel dimension  $U_i$ ; this is the texel dimension perpendicular to the tilt. Combining these we obtain an expression for  $A_i$ , the expected image-texel area.

As illustrated in Fig. 31, an image location is specified by  $(X, Y)$ , in a normalized coordinate system which does not depend on the number of pixels in the image:  $X$  and  $Y$  are zero at the image center, and are  $-1$  or  $1$  at the image border. (For notational simplicity, we are assuming square images.)



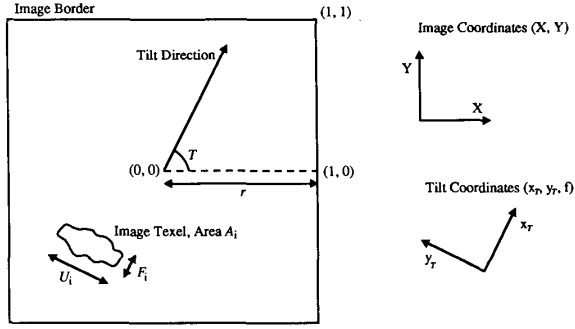


Fig. 31. Image and tilt coordinate systems.

The camera parameters we use are the focal length  $f$  and the physical width of the film  $r$ . The final expressions use only the ratio  $r/f$ , which is a measure of the field-of-view of the camera lens. Telephoto lenses have a low value of  $r/f$ , whereas fish-eye lenses have a high value of  $r/f$ .

The slant and tilt of the textured plane are denoted by  $S$  and  $T$ , respectively. Simple derivations may be obtained by defining a *tilt coordinate system*,  $(x_T, y_T, z_T)$ , with origin at the focal point, the  $z_T$  axis perpendicular to the image plane, and the  $x_T$  axis in the tilt direction. The view direction is along the positive  $z$  axis; thus this is a left-handed coordinate system. A point  $(X, Y)$  in image coordinates is transformed to tilt coordinates by

$$x_T = X \cos T + Y \sin T$$

$$y_T = X \sin T + Y \cos T$$

$$z_T = f.$$

Shown in Fig. 32 is the  $x_T$ - $z_T$  plane, which is perpendicular to both the image plane and the textured plane. The angle  $\theta = \tan^{-1}(x_T(r/f))$  for an image location with tilt coordinates  $(x_T, y_T, f)$ ; given raw image coordinates  $(X, Y)$ ,

$$\theta = \tan^{-1}((X \cos T + Y \sin T)(r/f)). \quad (6)$$

From the geometry in Fig. 32 we derive that [4]

$$F_i = F_p \frac{f}{g} \cos S (1 - \tan \theta \tan S)^2.$$

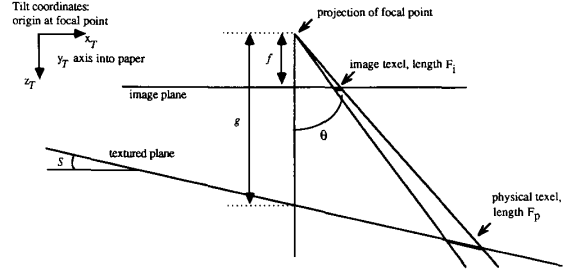
The only approximation made is that  $\theta$  does not change significantly across the texel. To eliminate the dependence on the surface-depth  $g$ , we calculate  $F_i$  in terms of  $F_c$ , the foreshortened dimension of a texel at the image center:

$$F_i = F_c (1 - \tan \theta \tan S)^2. \quad (7)$$

Similarly, we derive that  $U_i$ , the unforeshortened image-texel dimension, is related to  $U_c$ , the unforeshortened dimension of a texel at the image center by

$$U_i = U_c (1 - \tan \theta \tan S). \quad (8)$$

If the image texel has a compact shape, the area  $A_i$  of the image texel is proportional to the product of  $F_i$  and  $U_i$ .

Fig. 32. The  $x_T$ - $z_T$  plane. Since  $y_T$  is constant, both the image plane and the textured plane are perpendicular to the drawing.

(For example, if the physical texel is a circle, the image texel is an ellipse with  $A_i = (\pi/4) F_i U_i$ ; if the physical texel is a rectangle, the image texel is a parallelogram with  $A_i = F_i U_i$ .) Thus,  $A_i = k F_i U_i$ , where  $k$  is a constant of proportionality which depends upon the texel shape. Therefore  $A_i$ , the area of a texel at location  $(X, Y)$  in the image<sup>10</sup> is related to  $A_c$ , the area of a texel at the image center by

$$A_i = A_c (1 - \tan \theta \tan S)^3. \quad (9)$$

Substituting  $\theta$  from (6) into (9), we see that the following values are needed to predict the texel area anywhere in the image:

- $A_c$ , the area that would be measured for a texel located at the center of the image.
- $S$  and  $T$ , the slant and tilt of the textured plane.
- The ratio  $r/f$ , which is related to the field-of-view of the camera lens.

In our work we assume that the field-of-view is known. The other three quantities ( $A_c$ ,  $S$ ,  $T$ ) form the parameter space we search to find the best planar fit for a given texture image.

### B. Fitting a Planar Surface to the Candidate Texels

Having extracted candidate texels from an image of a textured surface, we find the orientation of the textured plane that best agrees with the observed areas of the candidate texels. A planar surface is characterized by the triple  $(A_c, S, T)$ , where  $A_c$  is the texel area expected in the image center,  $S$  is the slant, and  $T$  is the tilt. In order to find the best planar fit for the image texture, we discretize the possible values of  $A_c$ ,  $S$ , and  $T$ , and evaluate each possible planar fit. For each choice of  $(A_c, S, T)$ , (9) gives the expected texel area at each image location. These expected areas are compared to the region areas actually occurring in the image, and a fit-rating is computed for the plane. The plane that receives the highest fit-rating is selected as the estimate of the textured surface. The candidate texels that support the best planar fit are interpreted as true image texels.

<sup>10</sup>As pointed out by a reviewer, this result can also be derived from equations (2-1) and (3-2) in [18]; they also use the approximation that  $\theta$  is constant across the texel.



The rating of a planar fit is computed by summing contributions from all the candidate texels. If a texel at location  $(X, Y)$  has an area which is close to the ideal area in (9), then the texel provides a large amount of support for the planar fit. As the actual texel area deviates from the ideal area, the support for the planar fit decreases; we use an exponentially-decreasing weighting function. The rating of a planar fit is computed as

$$\begin{aligned} \text{fit rating} &= \sum_{\text{all regions}} (\text{region area}) \\ &\quad \cdot |\text{region contrast}| e^{-(\text{region-fit})^2/4} \\ \text{where region-fit} &= \frac{\max(\text{expected area, actual area})}{\min(\text{expected area, actual area})} \end{aligned} \quad (10)$$

The region-fit is 2.0 for a candidate texel that is either half as big or twice as big as the size predicted by the planar fit.

We begin with a coarse fit, in which the  $(A_c, S, T)$  space is searched at sparse locations:  $A_c$  (in units of pixels) takes on the values  $\{10, 20, 40, 80, 160, 320, 640\}$ ,  $S$  takes on the values  $\{0^\circ, 5^\circ, 10^\circ, \dots, 70^\circ, 75^\circ, 80^\circ\}$ , and  $T$  takes on the values  $\{0^\circ, 20^\circ, 40^\circ, \dots, 300^\circ, 320^\circ, 340^\circ\}$ . To refine the planar fit, a more detailed search of the  $(A_c, S, T)$  space is done in the neighborhood of the best plane from the coarse fit:  $S$  is changed in increments of  $2.5^\circ$ ,  $T$  in increments of  $5^\circ$ , and  $A_c$  in increments of less than 25 percent. As illustrated in parts (e) of Figs. 3 and 4 the fit-rating values change smoothly as a function of  $A_c$ , slant, and tilt. The plane that receives the highest fit-rating is selected as the best estimate of the textured surface. True image texels are those regions that have an area close to the area expected by the best planar fit.

#### V. APPLICATION OF THE ALGORITHM TO REAL IMAGES

We have conducted experiments with a variety of images of natural textures, having different mixes of texel shapes, number of fields, types of gradients, tilt directions, and three-dimensional texel effects. The results of the performance of the algorithm on a large variety of textures should help in judging the strengths, weaknesses, and generality of the algorithm and its current implementation. Part (a) of Figs. 3–30 show 14 of the images we have used in our experiments. A few of the images are photographs of outdoor scenes taken by one of the authors in Urbana, Illinois. The rest are illustrations in books (see [4] for references), which we have rephotographed. All of these images are digitized off of the photographic negatives using a drum scanner. The images are 512 by 512 pixels; the image sizes in the figures vary because image borders have been trimmed. All of the images are processed the same way; the method has no parameters that need to be tuned to particular images. As was described in Sections III and IV, the processing of an image is divided into three main phases: fit disks to the uniform im-

age regions, construct candidate texels from the disks, and fit a planar surface to the candidate texels.

The results of each phase are illustrated for one texture, a rock pile, in Figs. 3 and 4. Fig. 3 shows the results obtained for the positive contrast (bright) texture over dark background, and Fig. 4 shows the results for the negative contrast (dark) texture over bright background. The original image is shown in part (a) of each figure. Part (b) of each figure shows the extracted disks that model the regions of uniform gray level in the original image. Overlapping sets of these disks are used to make the list of candidate texels. It is impossible to display all the disks in a single image, since many disks are spatially contained in larger disks. This spatial containment typically means that either 1) the large disk is part of a texture element and the small disks are subtexture, or 2) the small disks are texels and the large disk is supertexture. In case 1) the large disk usually has higher contrast than the smaller disks, whereas in case 2) the smaller disks usually have higher contrast than the large disk. Wherever disks overlap, our figures show the disk of higher contrast. Therefore most subtexture disks in part (b) are not visible: they are covered by a larger, higher-contrast disk corresponding to part of a texture element.

The parameters of the best planar fit are illustrated by the synthetic texture images in part (e) of the figures. The detected texels are shown in parts (c) and (d): these are all candidate texels having area within a factor of two of the area expected by the best planar fit.

Parts (f) and (g) illustrate the change of fit-rating as a function of  $A_c$ , slant, and tilt. The height fields in part (f) of each figure show fit-rating as a function of slant and tilt, with  $A_c$  fixed at the value that produces the best planar fit for the texture in question.<sup>11</sup> The height fields flatten out near the back because tilt becomes less important as slant decreases; tilt is irrelevant when the slant is zero. The graphs in part (g) of each figure show fit-rating as a function of  $A_c$ , with slant and tilt fixed at the values that produce the best planar fit for the texture in question.

Figs. 5–30 illustrate selected results for 13 additional images of natural textures. The results obtained for each image are illustrated in two successive figures. The first figure shows the results for the positive-contrast texture, and the second figure shows the results for the negative-contrast texture. Parts (a), (b), and (c) of Figs. 5–30 are analogous to the corresponding parts of Figs. 3 and 4, whereas part (d) of Figs. 5–30 is analogous to part (e) of Figs. 3 and 4. For brevity, the details shown in parts (d), (f), and (g) of Figs. 3 and 4 are not repeated for the textures in Figs. 5–30.

The shape of the fit-rating peak is related to the properties of the image texture. A sharp fit-rating peak indicates that the texels have small size variance. This observation is supported by the fit-rating plots for the aerial view of houses (Figs. 5 and 6) and by the field of sun-

<sup>11</sup>In these height fields, the fit-rating values have been squared for display purposes.

flowers (Figs. 15 and 16), although, for brevity, these plots are not shown in the figures. If the texel sizes have larger variance, as for the clouds (Figs. 11 and 12) and the rock pile (Figs. 3 and 4), then the peak is much broader. (In the rock-pile image, the nonplanarity of the original textured surface also contributes to the broadness of the fit-rating peak.) The texels shown in part (c) of the figures are those candidate texels having area within a factor of two of the area expected by the planar fit. Using this same factor of two for all images causes incomplete extraction of texels in images where texel size is highly variable. More complete texel extraction can be achieved by adjusting the criteria for choosing texels from the set of candidate texels: the criteria should vary as a function of the broadness of the fit-rating peak in  $(A_c, S, T)$  space.

The accuracy of the results may be illustrated in two ways. First, the reader can compare his perception of the textured surfaces (part (a) of Figs. 3–30) with the planar surface fitted by the program. Agreement with human perception is quite good for many of the images. Second, since the processing of the positive-contrast and negative-contrast regions is performed totally independently, the agreement between the slants and tilts obtained by the two analyses strengthens the confidence in the results. (Note that the  $A_c$  parameters are not expected to be similar for the positive-contrast and negative-contrast regions—the positive-contrast and negative-contrast regions may be of very different sizes.) However, the two analyses may not always lead to the same estimates of slant and tilt, because a texture may not be homogeneous in both texel size and texel separation. Thus, an agreement among multiple analyses (such as the two discussed here) should not be required; instead, a method of automatically assessing the accuracies of the results obtained by different analyses, and selecting and integrating the pertinent analyses must be devised. Work is underway to address this problem.

Table I summarizes the planar fits obtained for all images. These fits use slants that are multiples of  $2.5^\circ$  and tilts that are multiples of  $5^\circ$ . The slant and tilt values computed from the positive-contrast and negative-contrast regions are frequently within  $15^\circ$  of each other. For reference, a  $30^\circ$  difference in tilt is equal to the angular distance between adjacent numbers on a clock face. A  $30^\circ$  difference in slant, on the other hand, is a more serious error. In many of those images that have a large discrepancy between the two planar fits, attributes of the original texture lead us to expect the fits to differ in accuracy. We have identified four reasons for the observed discrepancies. In the field of flowers (Fig. 23) and the water lilies (Fig. 25), the spaces between the texels are less regular than are the areas of the texels; therefore the fit to the negative-contrast regions is not as accurate as the fit to the positive-contrast regions. A second reason the background regions produce inaccurate results is because the properties of the physical texels are more important than the properties of background regions. In images where the physical texels are separated by gaps, the intertexel spacing carries more information than does the shape or area

TABLE I

| Description             | Figures | Fit to positive-contrast regions |              |             | Fit to negative-contrast regions |              |             | Difference   |            |
|-------------------------|---------|----------------------------------|--------------|-------------|----------------------------------|--------------|-------------|--------------|------------|
|                         |         | $A_c$                            | slant        | tilt        | $A_c$                            | slant        | tilt        | slant        | tilt       |
| A rock pile             | 3, 4    | 40                               | $62.5^\circ$ | $65^\circ$  | 40                               | $60^\circ$   | $75^\circ$  | $2.5^\circ$  | $10^\circ$ |
| Aerial view of houses   | 5, 6    | 35                               | $62.5^\circ$ | $95^\circ$  | 60                               | $67.5^\circ$ | $110^\circ$ | $5^\circ$    | $15^\circ$ |
| Birds flying over water | 7, 8    | 35                               | $45^\circ$   | $80^\circ$  | 40                               | $57.5^\circ$ | $100^\circ$ | $12.5^\circ$ | $20^\circ$ |
| Prayer at a mosque      | 9, 10   | 160                              | $27.5^\circ$ | $50^\circ$  | 120                              | $42.5^\circ$ | $100^\circ$ | $15^\circ$   | $50^\circ$ |
| Fleecy clouds           | 11, 12  | 100                              | $55^\circ$   | $275^\circ$ | 160                              | $55^\circ$   | $280^\circ$ | $0^\circ$    | $5^\circ$  |
| 3D movie audience       | 13, 14  | 280                              | $45^\circ$   | $105^\circ$ | 320                              | $7.5^\circ$  | $330^\circ$ | large        |            |
| Sunflowers              | 15, 16  | 160                              | $70^\circ$   | $95^\circ$  | 200                              | $70^\circ$   | $90^\circ$  | $0^\circ$    | $5^\circ$  |
| A tree trunk            | 17, 18  | 70                               | $65^\circ$   | $345^\circ$ | 80                               | $42.5^\circ$ | $0^\circ$   | $25.5^\circ$ | $15^\circ$ |
| Bathers on the Ganges   | 19, 20  | 100                              | $45^\circ$   | $80^\circ$  | 80                               | $65^\circ$   | $85^\circ$  | $20^\circ$   | $5^\circ$  |
| A plowed field          | 21, 22  | 80                               | $42.5^\circ$ | $40^\circ$  | 100                              | $65^\circ$   | $80^\circ$  | $22.5^\circ$ | $40^\circ$ |
| A field of flowers      | 23, 24  | 50                               | $70^\circ$   | $90^\circ$  | 140                              | $52.5^\circ$ | $20^\circ$  | large        |            |
| Water lilies            | 25, 26  | 120                              | $75^\circ$   | $90^\circ$  | 160                              | $52.5^\circ$ | $70^\circ$  | $22.5^\circ$ | $20^\circ$ |
| Ripples                 | 27, 28  | 50                               | $52.5^\circ$ | $105^\circ$ | 120                              | $62.5^\circ$ | $105^\circ$ | $10^\circ$   | $0^\circ$  |
| Water Hyacinths         | 29, 30  | 100                              | $37.5^\circ$ | $80^\circ$  | 100                              | $40^\circ$   | $80^\circ$  | $2.5^\circ$  | $0^\circ$  |

of the background regions. Thus, the results for the negative-contrast regions of the movie image (Fig. 14) and the lily pad image (Fig. 26) are inaccurate because the area of the background regions poorly reflects the intertexel spacing. A third reason for discrepancies between the two slant and tilt estimates is a large variability in texel area (as occurs in Fig. 9, the image of prayer at a mosque). This causes a broad peak in the planar fit space; hence the exact peak location is not as accurate for these images as for others. A fourth reason for inaccurate results is that the current extraction of uniform regions fragments noncompact regions in an arbitrary way, increasing the variabilities of the measured areas. This effect can be seen in the background of the movie image (Fig. 14). For nearly all of the images, at least one of the two analyses produces results that are in good agreement with human perception.

## VI. SUMMARY

We have presented a general discussion of the problem of recovering scene-layout information from the texture cues present in an image. We argue that extraction of texels is useful and perhaps even necessary for correct interpretation of texture gradients in the face of subtexture and supertexture. In order to separate texture elements from other regions (such as subtexture) it is necessary to perform texel identification and surface fitting simultaneously.

We have presented an implementation that is based on these ideas; the implementation is restricted to the detection of gradients of texel area. A multiscale region detector is developed from the response of an ideal disk to convolution with a Laplacian-of-Gaussian ( $\nabla^2 G$ ) over a range of scales. The output of the region detector is used to form a list of candidate texels. These candidate texels then provide the evidence needed to choose a good planar fit to the image texture; at the same time, the best planar fit determines which of the candidate texels are true texels. Results are shown for a variety of natural textures.

One consequence of the integration approach presented in this paper is that all regions whose properties are not unified by the gradient of a given property are treated as noise. For any given property, the noise regions do not contribute significantly to the fit-rating quality by virtue

of the exponential function in (10). Such regions could be the result of noise in the original image or in the region detection process. However, these noise regions could be valid texels if the gradient of a different property is considered, and they could quite possibly support the same surface orientation as the nonnoise regions. As long as there is some property whose gradient is supported by regions occupying a sufficiently large image area, the corresponding regions must be treated as texels. This is why the use of multiple texture gradients is necessary. A goal of our ongoing research is to estimate surface orientation from an integrated analysis of several relevant texture gradients, including area gradients, aspect-ratio gradients, and density gradients.

Because of the significant variability which is characteristic of natural textures, texture gradient as a cue of surface orientation appears to be more useful to obtain a coarse judgment of surface orientation and scene layout than as a source of obtaining accurate estimates. Stereo and other sources of scene information may be more appropriate for obtaining greater accuracy, e.g., for extracting shapes of curved, complex surfaces. In this sense, the analysis based on planar surfaces may suffice for most natural scenes containing textured surfaces, although mathematically (or for use with synthetic textures, where texture variability could be controlled), the approach presented in this paper could be extended to apply to curved surfaces. The extension required would be only in the surface fitting process. A much more important use of texture cues in real scenes is for segmentation of a scene into different textured surfaces [25]. With such segmentation available, it would be possible to identify image parts to which the approach of this paper could be applied meaningfully. As we stated earlier, we have not addressed the problem of texture segmentation in this paper.

#### REFERENCES

- [1] J. Aloimonos and M. Swain, "Shape from texture," in *Proc 9th Int. Joint Conf. AI*, 1985, pp. 926-931.
- [2] J. Aloimonos, "Detection of surface orientation from texture I: The case of planes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1986, pp. 584-593.
- [3] R. Bajcsy and L. Lieberman, "Texture gradient as a depth cue," *Comput. Graphics Image Processing*, vol. 5, pp. 52-67, 1976.
- [4] D. Blostein, "Recovering the orientation of textured surfaces in natural scenes," Ph.D. dissertation, Univ. Illinois, Coordinated Science Lab. Rep. UILU-ENG-87-2219, Apr. 1987.
- [5] D. Blostein and N. Ahuja, "Representation and three-dimensional interpretation of image texture: An integrated approach," in *Proc. IEEE First Int. Conf. Computer Vision*, June 1987, pp. 444-449.
- [6] —, "A multi-scale region detector," *Comput. Vision, Graphics, Image Processing*, vol. 45, no. 1, pp. 22-41, Jan. 1989.
- [7] M. L. Braunstein and J. W. Payne, "Perspective and form ratio as determinants of relative slant judgments," *J. Exp. Psychol.*, vol. 81, no. 3, pp. 584-590, 1969.
- [8] J. Crowley and A. Parker, "A representation for shape based on peaks and ridges in the difference of low pass transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 2, pp. 156-170, Mar. 1984.
- [9] J. E. Cutting and R. T. Millard, "Three gradients and the perception of flat and curved surfaces," *J. Exp. Psychol.: General*, vol. 113, no. 2, pp. 198-216, 1984.
- [10] L. Davis, L. Janos, and S. Dunn, "Efficient recovery of shape from texture," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, no. 5, pp. 485-492, Sept. 1983.
- [11] J. Gibson, *The Perception of the Visual World*. Boston, MA: Houghton Mifflin, 1950.
- [12] K. Ikeuchi, "Shape from regular patterns (an example of constraint propagation in vision)," MIT A.I. Memo 567, Mar. 1980.
- [13] K. Kanatani, "Detection of surface orientation and motion from texture by a stereological technique," *Artificial Intell.*, vol. 23, pp. 213-237, 1984.
- [14] K. Kanatani and T. Chou, "Shape from texture: General principle," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition 86*, Miami, FL, June 1986, pp. 578-583.
- [15] J. Kender, "Shape from texture," Ph.D. dissertation, Carnegie-Mellon Univ., Rep. CMU-CS-81-102, Nov. 1980.
- [16] H. Nakatani, S. Kimura, O. Saito, and T. Kitahashi, "Extraction of vanishing point and its application to scene analysis based on image sequence," in *Proc. Int. Conf. Pattern Recognition*, 1980, pp. 370-372.
- [17] R. Nevatia and K. R. Babu, "Linear feature extraction and description," *Comput. Graphics Image Processing*, vol. 13, pp. 257-269, 1980.
- [18] Y. Ohta, K. Maenobu, and T. Sakai, "Obtaining surface orientation from texels under perspective projection," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1981, pp. 746-751.
- [19] R. J. Phillips, "Stationary visual texture and the estimation of slant angle," *Quart. J. Psychol.*, vol. 22, pp. 389-397, 1970.
- [20] A. Rosenfeld, "A note on automatic detection of texture gradients," *IEEE Trans. Comput.*, vol. C-24, pp. 988-991, Oct. 1975.
- [21] R. R. Rosinski, "On the ambiguity of visual stimulation: A reply to Eriksson," *Perception Psychophys.*, vol. 16, no. 2, pp. 259-263, 1974.
- [22] R. Rosinski and N. Levine, "Texture gradient effectiveness in the perception of surface slant," *J. Exp. Child Psychol.*, vol. 22, pp. 261-271, 1976.
- [23] K. A. Stevens, "Slant-tilt: The visual encoding of surface orientation," *Biol. Cybern.*, vol. 46, pp. 183-195, 1983.
- [24] D. Vickers, "Perceptual economy and the impression of visual depth," *Perception and Psychophys.*, vol. 10, no. 1, pp. 23-27, 1971.
- [25] H. Voorhees and T. Poggio, "Detecting textons and texture boundaries in natural textures," in *Proc. IEEE First Int. Conf. Computer Vision*, June 1987, pp. 25-258.
- [26] A. P. Witkin, "Recovering surface shape and orientation from texture," *Artificial Intell.*, vol. 17, pp. 17-45, 1981.
- [27] —, "Scale space filtering," in *Proc. Eighth Int. Joint Conf. Artificial Intelligence*, Karlsruhe, West Germany, Aug. 1983, pp. 1019-1022.



**Dorothea Blostein** (S'87-M'88) received the B.S. degree in mathematics and computer science from the University of Illinois, Urbana-Champaign, in 1978, the M.S. degree in computer science from Carnegie-Mellon University, Pittsburgh, PA, in 1980, and the Ph.D. degree in computer science from the University of Illinois in 1987.

From 1980 to 1982, she worked at Intel Corporation. During 1987-1988, she worked at the University of Illinois Computer-Based Education Research Laboratory. She is currently an Assistant Professor in the Department of Computing and Information Science at Queen's University in Kingston, Ontario. Her research interests include computer vision, pattern recognition, computer music, and user-interface design.



**Narendra Ahuja** (S'79-M'79-SM'85) received the B.E. degree with honors in electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1972, the M.E. degree with distinction in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1974, and the Ph.D. degree in computer science from the University of Maryland, College Park, in 1979.

From 1974 to 1975 he was Scientific Officer in the Department of Electronics, Government of India, New Delhi. From 1975 to 1979 he was at the Computer Vision Laboratory, University of Maryland, College Park. Since 1979 he has been with the University of Illinois at Urbana-Champaign where he is currently (1988-) a Professor in the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute. His interests are in computer vision, robotics, image processing, and par-

allel algorithms. He has been involved in teaching, research, consulting, and organizing conferences in these areas. His current research emphasizes integrated use of multiple image sources of scene information to construct three-dimensional descriptions of scenes, the use of the acquired three-dimensional information for object manipulation and navigation, and multiprocessor architectures for computer vision.

Dr. Ahuja received the University Scholar Award (1985), Presidential Young Investigator Award (1984), National Scholarship (1967-1972), and President's Merit Award (1966). He has coauthored the books *Pattern Models* (Wiley, 1983) with Bruce Schachter, and *Motion and Structure from Image Sequences* (Springer-Verlag, to appear) with Juyang Weng and Thomas Huang. He is Associate Editor of the journals *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, and *Computer Vision, Graphics, and Image Processing*. He is a member of the American Association for Artificial Intelligence, the Society of Photo-Optical Instrumentation Engineers, and the Association for Computing Machinery.