Technical Report 968

Finding Texture Boundaries in Images

Harry Voorhees

MIT Artificial Intelligence Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

Technical Report No. 968 June 1987

Finding Texture Boundaries in Images

by

Harry Voorhees

A thesis submitted to the Department of Electrical Engineering and Computer Science on May 8, 1987 in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering and Computer Science.

Abstract. Texture provides one cue for identifying the physical cause of an intensity edge, such as occlusion, shadow, surface orientation or reflectance change. Marr, Julesz, and others have proposed that texture is represented by small lines or blobs, called "textons" by Julesz [1981*a*], together with their attributes, such as orientation, elongation, and intensity. Psychophysical studies suggest that texture boundaries are perceived where distributions of attributes over neighborhoods of textons differ significantly.

However, these studies, which deal with synthetic images, neglect to consider two important questions: How can these textons be extracted from images of natural scenes? And how, exactly, are texture boundaries then found? This thesis proposes answers to these questions by presenting an algorithm for computing blobs from natural images and a statistic for measuring the difference between two sample distributions of blob attributes. As part of the blob detection algorithm, methods for estimating image noise are presented, which are applicable to edge detection as well.

Thesis supervisor: Dr. Tomaso Poggio
Professor of Psychology

Acknowledgments

Many members of the Artificial Intelligence Laboratory contributed to this research. I'd especially like to thank Jim Mahoney and Jim Little for many thoughtful discussions; Alan Yuille and Michael Gennert, for their mathematical expertise; Ann Knowles, Dave Braunegg, Walter Gillett, Jim Mahoney, and Jim Little, for carefully reviewing drafts of this paper; and Bror Saxberg and Anselm Spoerri, for recommending useful statistics books. Most of all, I'd like to thank Tomaso Poggio for years of great ideas and enthusiastic support.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the U.S. Department of Defense under Office of Naval Research contract N00014–85–K–0124 and Army contract DACA76–85–C–0010. The financial support of Hughes Aircraft Company is also gratefully acknowledged.

Contents

Chapter 1 **Introduction**

Vision has been characterized as a process which converts a large array of intensity data into a symbolic description of objects in the scene. A critical step in the computation of symbols from signals is the identification of physical discontinuities in the scene—sharp changes in surface depth and orientation, reflectance, and illumination. Of these events, depth discontinuities (occluding boundaries) are undoubtedly the most important, since segmenting the image into distinct surfaces and, ultimately, objects, seems crucial for interpretation.

Edge detection—the detection of intensity changes in images—has proven useful for locating physical discontinuities. But since all types of discontinuities cause intensity changes, other processes are needed to interpret the types of physical discontinuities present. Texture, along with stereo, color, and motion, provides a cue.

As noted by Riley [1981], most drastic changes in texture can only be explained by occlusion, while projection-type changes may be due to a change in surface orientation as well. Stereopsis combined with surface interpolation can locate depth and orientation discontinuities (e.g., Terzopoulos [1985]). Color is useful for distinguishing material changes from other types of discontinuities (e.g., Rubin & Richards [1982]). Structure-from-motion can find occluding boundaries (e.g., Hildreth [1984]). Multiscale intensity edge analysis may

provide some additional cues for interpretation (e.g., Richards et al. [1982], Beattie [1984]).

Table 1.1 summarizes the kinds of discontinuities which various vision modules can identify, ignoring specularities and light sources in the scene. One cannot expect texture, or any other visual channel, to completely and correctly interpret all physical discontinuities present in a scene. However, each module can provide cues regarding the types of changes present. The vision group at the M.I.T. Artificial Intelligence Laboratory is currently building a "Vision Machine" which will attempt to integrate coarse, symbolic descriptions of intensity, stereo, motion, color, and texture discontinuities. Together the data-driven modules, combined with higher-level knowledge, should provide a powerful method for segmenting and interpreting surfaces in a scene.

It is with this goal in mind—the interpretation of physical discontinuities—that I investigate the visual processing of texture information.

Table 1.1 Cues for Interpreting Image Intensity Changes[1]

Module	*Discontinuity Detected*	*Physical Interpretations*
Stereo	Depth discontinuity	Occlusion
	Orientation discontinuity	Surface Orientation
Motion	Optical flow discontinuity	Occlusion
Color	Hue discontinuity	Material (often occlusion)
Texture	Projection-type texture changes	Orientation (or occlusion)
	Intensity change only	Shadow
	Other large texture changes	Material (often occlusion)

[1]Charniak and McDermott [1985, Figure 3.57] tabulate the inverse mapping between columns two and three, namely, the kinds of discontinuities which can occur at each type of physical boundary.

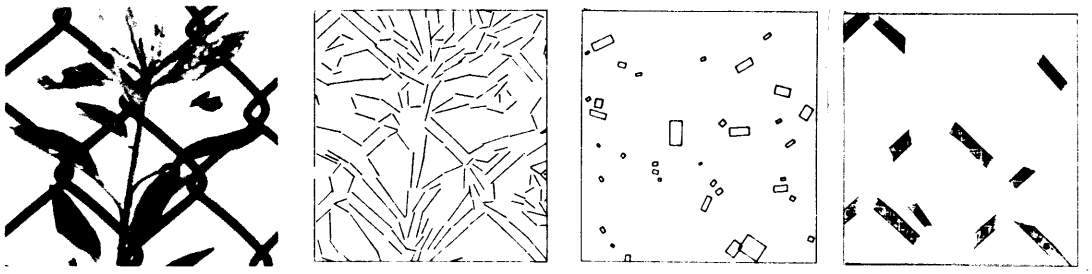


Figure 1.1 The raw primal sketch of the image shown in a is a set assertions of intensity changes and their attributes, including (b) edge segments, (c) blobs, and (d) bars [Marr 1982, p. 72].

1.1 Theory of human texture vision

The currently accepted computational theory of human texture vision is a result of research in artificial intelligence, psychophysics, and neurophysiology. Below we discuss developments most relevant to the research described in this thesis.

1.1.1 The primal sketch theory

David Marr pioneered the development of a computational framework for human vision in his theory of the primal sketch. Its main thesis was that a "symbolic" representation of visual information is constructed early in the interpretation process, without the use of higher level knowledge, and that even early visual processing, such as texture discrimination, operates at a symbolic level [Marr 1976*b*].

The first of these representations is the raw primal sketch, a representation of intensity changes in the image. Sharp edges, extended edges, thin lines or bars, and compact blobs are represented along with their attributes, such as orientation, contrast, and fuzziness (spatial scale) (see Figure 1.1). Edge assertions are constructed by parsing the output of different receptive fields

which detect intensity changes. The important point is that the raw primal sketch is not simply an array of detector outputs, but a symbolic description of the intensity changes present in the image.

As discussed earlier, it is useful to represent intensity changes since they usually correspond to physical events in the scene. The existence of retinal receptive fields suggest that such a computation is biologically plausible. As discussed in the next section, psychophysical experiments suggest a token-based representation of texture. The raw primal sketch is therefore supported on physical, biological, and psychophysical grounds.

In the full primal sketch, the intensity primitives, along with termination points (endpoints) of edges and thin lines, act as place tokens, the elements on which texture discrimination and other visual processes are based. Much of the power of the primal sketch theory stems from the fact that these place tokens are grouped, recursively, on the basis of proximity, curvilinearity, similarity, and local configuration into new place tokens [Marr 1976*b*].

According to Marr, texture processing involves two types of operations: local discriminations and grouping of place tokens. The discriminations are based on first-order statistics[2] of token attributes computed over relatively small neighborhoods. These measures include total density of blobs and contours, densities at different orientations, and mean and variance of size, lightness, and nearest-neighbor distance [Marr 1976*a*].

The same grouping operations which construct small aggregations of place tokens serve to identify textured regions. Figure 1.2 shows the result of aggre-

[2]First-order statistics describe the probablility distribution of a single measurement. Second-order statistics describe the probability distribution for pairs of points according to their configuration. In general, nth-order statistics describe the probability distribution of configurations of n points.

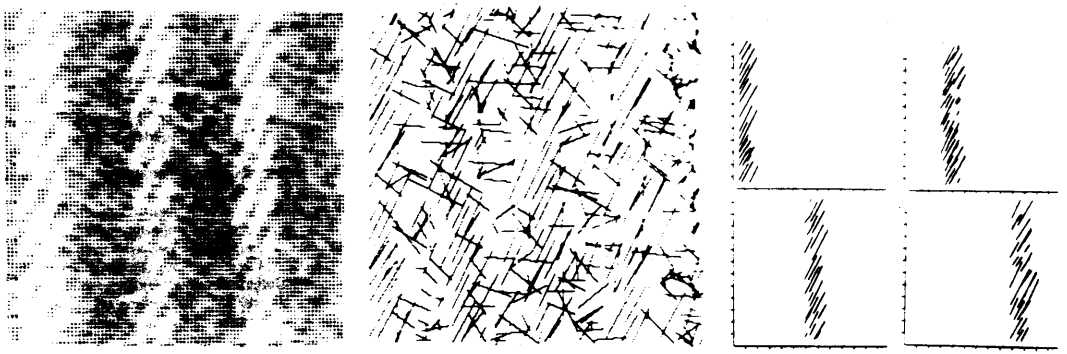


Figure 1.2 (a) Image of herringbone, (b) edges, and (c) edges at the predominant orientation (60°) grouped on the basis of orientation and proximity [Marr 1976a, pp. 658–659].

gating similarly oriented edges from an image of herringbone fabric. To some extent, the grouping is driven by global statistics; in the herringbone example, the predominant orientation is selected for grouping on the basis of a global histogram of edge orientations. In this data-driven mode, texture vision is responsible for figure and ground separation and the extraction of form, with little use of higher-level knowledge.

Despite its plausibility, the feasibility of Marr's proposal has not yet been proven. Many details of the computation need to be defined before the theory can be fully implemented.

1.1.2 Psychophysical studies

Recent psychophysical studies of texture perception support the computation of local, first-order statistics of image feature attributes and densities. These studies have shown that texture discrimination can occur immediately (within 160 milliseconds), before the eye has time to move. Presumably, this parallel, extra-foveal "preattentive" process directs the fovea to quickly scrutinize areas of interest within the scene (the so-called "attentive" mode of vision) [Julesz

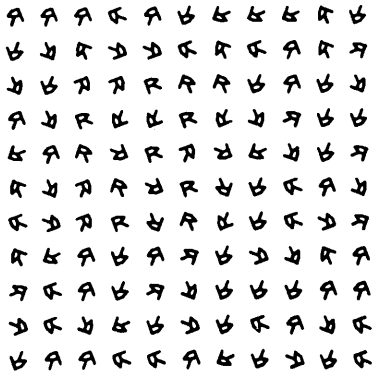


Figure 1.3 An example satisfying Julesz's conjectures: The central region of R's is indistinguishable from the background of reflected R's [Julesz and Bergen 1983, p. 1623].

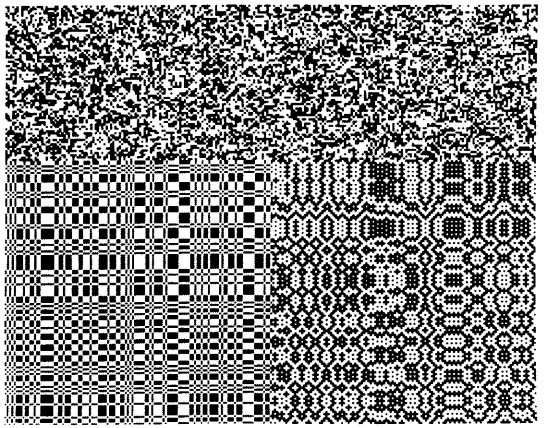


Figure 1.4 A counterexample to Julesz's original conjecture: The regions have identical third-order statistics yet are discriminable [Julesz and Bergen, p. 1630].

and Bergen 1983].

Julesz has studied human texture perception using synthetic images, usually composed of dots or symbols. Originally, he proposed that textures having identical second-order, global statistics of image points were preattentively indistinguishable, regardless of whether their higher order statistics differed. With few exceptions, textures having different second-order statistics were discriminable, Julesz [1975] conjectured. In Figure 1.3, for example, the central region of R's is not immediately distinguishable from the background of reflected R's; the perception supports his conjecture, since the probability of a dipole landing on a particular pair of intensity values is the same in both regions. Since second-order statistics of points reflect characteristics of textures such as directionality and coarseness, they have been used extensively as texture measures in the pattern recognition field (Haralick [1979] and Van Gool et al. [1985] review work in this area).

Counterexamples to Julesz's original conjecture were found, however. Caelli, Julesz, and Gilbert [1978; Julesz 1981*b*] and Gagalowitz [1981] constructed examples like the one shown in Figure 1.4 having identical third-order statistics which were discriminable. Julesz eventually abandoned second-order statistics and proposed a "texton theory" based on first-order measures of local image features [Julesz 1981*b*; Julesz and Bergen 1983]. He now explained the perception of Figure 1.3 as equal distributions of line segment orientations. In addition to attributes of line segments or elongated blobs, textons included their termination points (endpoints) and crossings (intersections). Figure 1.5 shows a purported demonstration of crossings as textons. More recently, Julesz [1986] has measured the maximum spacing between line segments perceived as a single texton and between textons perceived as a texture.

Phenomenologically, texton theory resembles the primal sketch theory, but with the minor addition of crossings as tokens, and the major omission of filtering and grouping operations. Julesz's results, based on synthetic images, say nothing about the process of detecting textons, and little about the process of finding texture boundaries.

Another psychophysicist, Beck, has proposed a theory based on first-order statistics of attributes of texture elements. Differences in attributes between adjacent neighborhoods are measured, after which "decision units" somehow segment the display [Beck et al. 1983]. Beck and his colleagues also observe that the perceived strength of texture segmentation increases with neighborhood size and concentration of strong difference locations, but is inhibited by the presence of difference signals of comparable strength everywhere (texture gradients). Beck's theory differs from Marr's by omitting segmentation based explicitly on element similarity. His proposed stages are essentially an abstract of a subset of the primal sketch theory.

More recently, Muller [1986] has presented psychophysical evidence supporting a difference-based theory of texture discrimination. He shows that "boundary-based cues"—differences obtained by local comparisons of texture elements—are perceived more strongly than are "field-based cues"—boundaries between homogeneous texture fields (see Figure 1.6). He also presents an O'Brian-Craik-Cornsweet illusion in which the values of texton attributes along the boundary affect the perception of textons inside the region. Muller concludes that texture discrimination may be analogous to intensity edge detection, being based on texture difference boundaries, as well as similarity-based grouping.

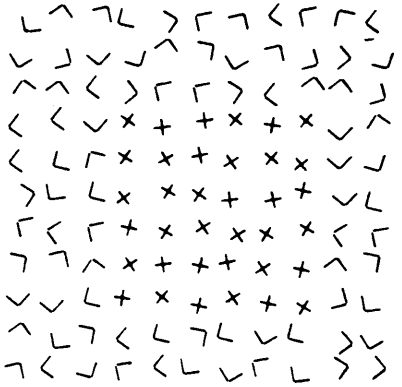


Figure 1.5 This example purportedly demonstrates that crossings are textons [Julesz and Bergen 1983, p. 1627].

Figure 1.6 These examples of Muller [1986] suggest that detection of local differences at boundaries (a) play a stronger role in texture discrimination than does detection of similarity within regions (b).

1.1.3 Using texture to interpret edges

While much attention has been devoted to the study of texture discrimination, relatively few studies investigate texture for the purpose of recovering information about the physical surfaces present in a scene. Methods of recovering shape from texture are described by Stevens [1980] and Witkin [1981]. Less effort has been focused on the problem of interpreting texture discontinuities.

One approach to the problem is to locate texture boundaries independently of other cues. Riley [1981] shows that boundaries due only to texture (and not intensity) can be used in motion correspondence and shape recognition, implying that the human visual system indeed represents texture boundaries explicitly. Riley also bases his argument on the physical significance of texture boundaries: Surface orientation changes cause only "geometric changes" of texture elements (changes in orientation, density, and length), while structural (material) boundaries may cause in changes in color, contrast, size and shape of elements as well.

Since humans can detect detect boundaries due only to texture changes, a complete theory of human vision must include an autonomous texture boundary detection process. But consider an alternative approach to the problem: Instead of detecting texture boundaries independently, one could use texture to label intensity or other types of edges. The usefulness of such a hypothesize-and-test approach might be justified by the observation that, in natural scenes, most texture boundaries correspond to intensity changes. Of course, the human visual system might also use such an approach attentively, e.g., to disambiguate conflicting preattentive cues.

Witkin [1982] labeled intensity edges in this manner by computing the correlations of intensity values between lines parallel to the selected intensity

edge. If the correlations are high on either side of the edge, but drop across it, the edge is due to occlusion. If the correlations remain high everywhere, but the regression parameters are offset across the edge, the intensity edge is due to shadow. While this computation is implausible as a model of human vision, it is interesting as a method for interpreting the physical cause of intensity changes.

1.2 Questions to be answered

The psychophysical studies, based on synthetic images of line segments or other symbols, neglect two important issues: First, how can textons be reliably extracted from images of natural scenes? And second, how, exactly, are texture boundaries then detected? Even Marr's proposal was sketchy and was never actually implemented. In this thesis, I offer answers to these questions by describing an implemented algorithm for computing textons and texture boundaries from images.

I conjecture that texture in natural images can be represented using blobs and their attributes. Chapter 3 describes an implemented algorithm for detecting such blobs in natural images using Laplacian of Gaussian ($\nabla^2 G$) filtering. As part of the algorithm, methods for estimating image noise are presented in Chapter 2, which are applicable to edge detection as well.

Chapter 4 presents some psychophysical experiments that investigate which attributes of blobs are compared by the human visual system and how small a difference in attribute values is perceivable. This leads to the development of a texture difference statistic in Chapter 5, which answers a critical question not addressed by Marr or the psychophysicists, namely, How are two sample dis-

tributions of texton attributes compared? The boundaries detected using the statistic accord with perception, as demonstrated by both real and synthetic examples.

Chapter 6 questions the significance of terminators and crossings, proposed by Julesz as textons. It shows that purported demonstrations of their salience can be explained by other attributes of blobs obtained by filtering the image. Chapter 7 concludes with some speculation on the computational aspects of preattentive and attentive vision, and on the third major question of texture vision: How can texture boundaries be interpreted according to their physical causes?

Chapter 2 **Noise Estimation**

When dealing with images of natural scenes, it is necessary to estimate the amount of noise in an image. In edge or blob detection, the noise estimate determines a threshold above which an intensity change is considered "significant." In this section, we show how noise can be estimated from a histogram of the image convolved with a Gaussian or a Laplacian of Gaussian filter, depending on which is used for edge or blob detection.

2.1 Gaussian filtering

If one assumes that the smoothed image, $f(x,y) = I * G$, consists of white Gaussian noise,

$$\Pr_f(z) = G(\sigma_n; z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}(\frac{z}{\sigma_n})^2\}$$

then, using the probability law for a function of random variables, the partial derivatives f_x and f_y have p.d.f.'s

$$\Pr_{f_x}(z) = \Pr_{f_y}(z) = \int_{-\infty}^{\infty} \Pr_f(z) \Pr_f(z-s)ds = G(\varsigma; z),$$

where $\varsigma = \sqrt{2}\,\sigma_n$.

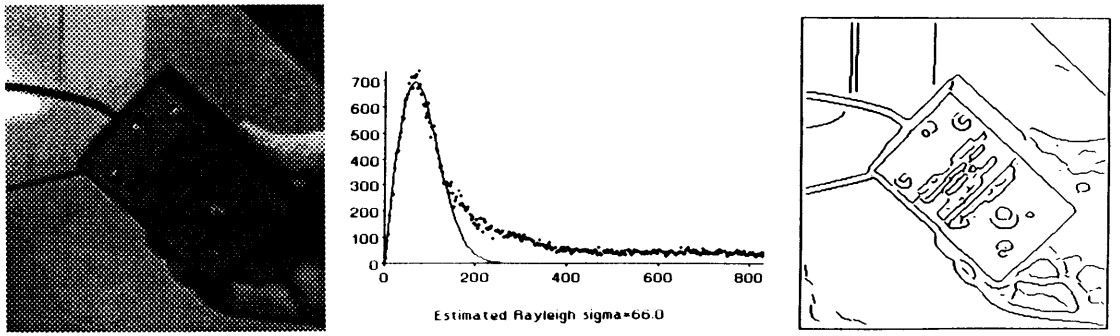
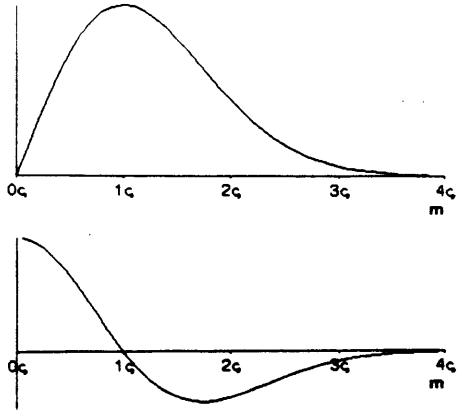


Figure 2.1 Simple noise estimation example. (a) Image of mouse, containing smooth intensity surfaces. (b) Magnitude of gradient histogram with Rayleigh distribution fit to peak at $m = \varsigma$. (c) Canny [1983] edges detected using hysteresis with magnitude of gradient threshold $\tau_1 = 2\varsigma$, $\tau_2 = 4\varsigma$.



Rayleigh distribution and its first derivative

Again using the probability law, the p.d.f. of the magnitude of the gradient $\|\nabla f\| = \sqrt{f_x^2 + f_y^2}$ is

$$\Pr_{\|\nabla f\|}(m) = m \int_0^{2\pi} \Pr_{f_x}(m \cos\theta) \Pr_{f_y}(m \sin\theta) d\theta = mG(\varsigma; m), \quad \text{for } m > 0$$

(see, e.g., Parzen [1960], pp. 320–321). The magnitude of gradient distribution is a Rayleigh distribution, graphed at left, with a maximum value at $m = \varsigma$. The addition of real edges, whose gradients are stronger than those of the noise distribution, affect mainly the tail of the distribution, and do not significantly affect the location of the peak. Bracho and Sanderson [1985], who applied this idea to region growing, noted that the assumption holds as long as a large portion of the image consists of roughly uniform intensity which satisfies the white noise model. Often the background of the image is such a region. In this case, the noise can be estimated by constructing a histogram $h(m)$ of $\|\nabla f\|$, and simply measuring the location of the peak as ς. Figure 2.1 shows an example for which this simple method works well.

While this strategy works well for images containing some uniform regions,

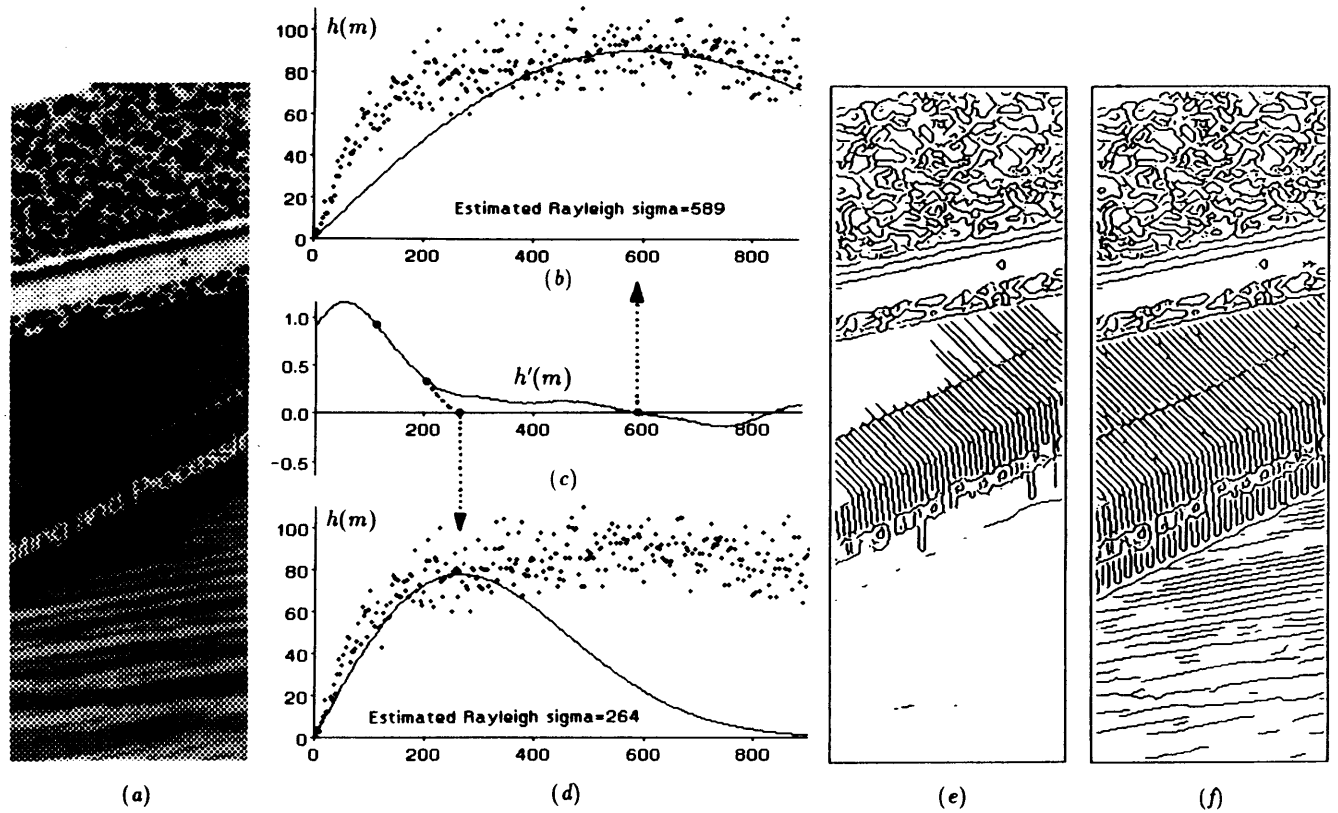


Figure 2.2 More complicated noise estimation example. Since this image (a) contains no smooth intensity surfaces, fitting the Rayleigh distribution to the global maximum of the gradient histogram, $h(m)$ (b) results in too high a edge threshold (e). A better fit is obtained (d) by extrapolating the first derivative of the smoothed histogram (c), resulting in a better edge threshold (f). Edge images e and f were thresholded with hysteresis with $\tau_1 = 2\varsigma$, $\tau_2 = 4\varsigma$.

it can fail when the image contains many strong texture edges, as shown in Figure 2.2*a*. In this case, the many edges of strong gradient obliterate the the local maximum of $h(m)$ due to the noise distribution. It is therefore preferable to fit the Rayleigh distribution to the steep, rising portion of the histogram.

The fitting can be done easily by using the first derivative of the smoothed histogram, $h'(m)$. Either the location where h' crosses zero or where it "would have crossed" zero, whichever quantity is less, estimates ς. The latter position is computed by linearly extrapolating the downward sloping portion of h' after

its maximum, as shown in Figure 2.2*c*. The next points on the curve at 80% and 20% of the maximum define the extrapolation line. The actual zero crossing, of course, corresponds to the first local maximum of $h(m)$, which is used as the noise estimate for images satisfying the uniform region assumption. The estimate of ς could be refined using a least-squares approach, but the simple method described here seems adequate in practice.

When constructing the histogram, it is desirable to use as large a number of buckets as possible, in order to estimate ς as accurately as possible. On the other hand, a large number of buckets will tend to make the histogram noisy, necessitating smoothing in order to locate the correct local maximum. Only the minimal amount of smoothing necessary should be performed in order to accurately estimate ς. As it turns out, the amount of smoothing depends on how spread out the histogram is, and the location of the global maximum before smoothing provides a good measure of this.[1]

Once ς is estimated, one can choose a threshold to exclude most edges due to noise. To remove noise with a confidence of c, requires a threshold τ such that

$$c = \frac{\int_0^\tau zG(\varsigma; z)dz}{\int_0^\infty zG(\varsigma; z)dz},$$

yielding $\tau/\varsigma = \sqrt{-2\ln c}$. Hence, removing noise with a confidence of 99% requires $\tau/\varsigma \approx 3$.

Figures 2.2*e* and 2.1*c* demonstrate that this improved method works for images consisting entirely of highly textured images, as well as for images

[1]**Implementation note:** I used a histogram of 500 buckets which spanned the lower 20% of the range of the magnitude of the gradient, not including gradients near the array boundaries. Convolving it with a Gaussian $G(\sigma = i/8; m)$—where i is the index of the global maximum of the $h(m)$—insured that $h'(m)$ was sufficiently smooth.

containing untextured areas. In these examples, I used the edge detector of Canny [1983], thresholding the edges with hysteresis, using thresholds $\tau_1 = 2\varsigma$ and $\tau_2 = 4\varsigma$. In many applications, it is desirable to remove weak "real" edges as well, in which case higher threshold factors are used.

The noise estimator presented here is an improvement over Canny's original scheme, which uses the 30th percentile of the magnitude of gradient histogram [Canny, Lisp code]. Simply using a fixed percentile of the histogram does not provide a robust estimate of noise, since the percentile varies with the density of real edges in the image.

2.2 Laplacian of Gaussian filtering and blob detection

For Laplacian of Gaussian edge or blob detection, one wishes to choose a non-zero threshold τ, to exclude most responses due to noise. If, as before, we assume the smoothed image $f = I * G$ consists of white Gaussian noise with standard deviation σ, then the p.d.f of $\nabla^2 f \equiv f_{xx} + f_{yy}$ is $\Pr_{\nabla^2 f}(z) = G(\hat{\varsigma}; z)$, where $\hat{\varsigma} = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 4^2}\,\sigma_n = \sqrt{20}\,\sigma_n$. In this case the histogram of $\nabla^2 f$ is a Gaussian distribution. Again, real edges mainly affect the tails of the distribution so the Gaussian distribution is fit to only the central portion of the histogram. In my implementation, the inflection points on either side of the global maximum of the smoothed histogram serve as bounds for fitting the Gaussian distribution, and an iterative least-squares technique is used to determine the parameters of the best-fitting Gaussian.[2] Figure 2.3 shows the

[2]**Implementation Note:** I used a histogram of 200 buckets which spanned the range of $I * \nabla^2 G$ (excluding values affected by boundary conditions) and smoothed the histogram with a Gaussian filter having standard deviation 4.0 to identify the inflection points. The corresponding points of the unsmoothed histogram were used for fitting the ideal Gaussian.

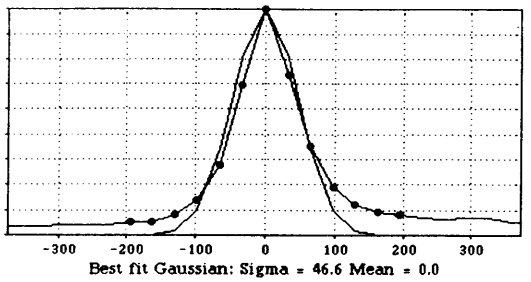


Figure 2.3 Laplacian of Gaussian noise estimation example. The histogram of $I * \nabla^2 G(\sigma = 1.5)$ for the image of Figure 2.2a, with the Gaussian distribution fit to its core. The dots indicate points of the histogram used for fitting.

histogram of $I * \nabla^2 G$ and the approximating Gaussian for the image of books is shown in Figure 2.2*a*.

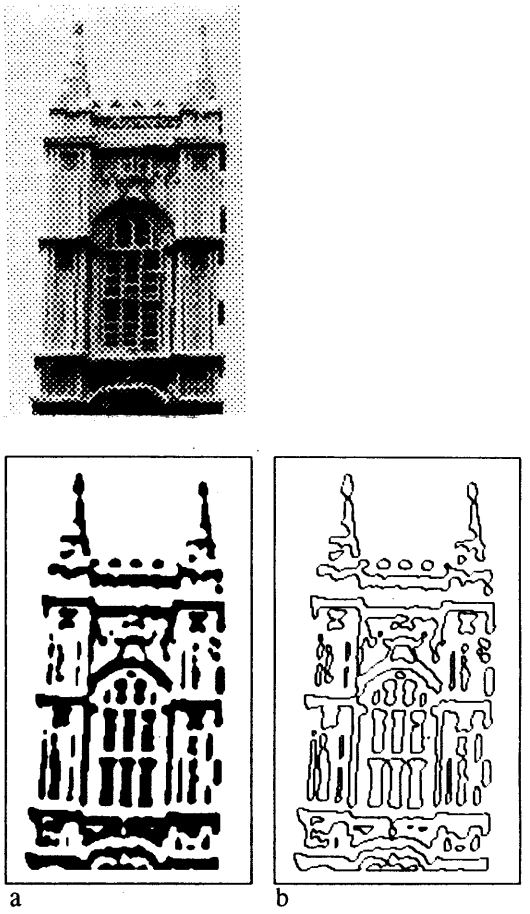
The estimate of $\hat{\varsigma}$ is used in the next chapter to choose an initial level for blob detection. The relationship between threshold τ and confidence level c is given by the standard normal distribution. Avoiding responses due to noise with $c = 99\%$, for example, requires $\tau/\hat{\varsigma} \approx 2.3$.

Chapter 3 **Blob Detection**

How can textons be computed from images of natural scenes? This chapter presents a method for detecting blobs in images—small compact or thin elongated regions which are lighter or darker than their surround—which often characterize the texture of the underlying surface.

Unlike the primal sketch algorithm for blob detection [Marr 1976*a*], the computation described here is not based on linking together nearby edge fragments which enclose or nearly enclose a small region. Rather, blobs are detected directly by thresholding the image convolved with the Laplacian of Gaussian filter. In this sense blobs may be regarded as duals of edges, which can be detected using the zero crossings of $I * \nabla^2 G$ (see illustration at right). Instead of thresholding at zero, however, a small threshold is used which removes some connections between blobs due to noise. The threshold is positive for dark blobs (and negative for light blobs), and its magnitude is proportional to the noise estimate ς, whose computation is described in Section 2.2. In the examples here, dark blobs are identified where $I * \nabla^2 G > \ell_0$, where $\ell_0 \approx 0.8\varsigma$, and where $\nabla^2 G$ is approximated with a difference of Gaussians mask, described by Marr and Hildreth [1980].

A similar computation for removing noise can be achieved by biasing one of the Gaussians by a slight factor—so that the difference of Gaussians mask



Blobs (a) as the duals of level crossings (b).

has non-zero area—and thresholding the convolved image at zero [Richards et al. 1982].

3.1 Discounting effects of illumination on blob size

For the purpose of relating computed texture to the properties of physical surfaces, it is desirable to discount the effects of illumination. This can be partially achieved by employing the logarithm of intensity values.

Image brightness is a multiple of surface reflectance and illumination:

$$b(x, y) = r(x, y, ...)\, i(x, y).$$

Here, the "reflectance" term $r(x, y, ...)$ is a function of surface orientation with respect to the viewer and the light source, as well as albedo, so it can include specular as well as lambertian components [Horn 1986].

The presence of a shadow decreases the brightness difference between a surface feature and its surround, and therefore decreases the response of $I * \nabla^2 G$ over the shaded region. Hence, if pixel values are linear with brightness, a shadow can reduce the size of blobs, as shown in Figure 3.1*a,b*. Making I proportional to the the logarithm of brightness, however, makes $I * \nabla^2 G$ proportional to ratios of b, which are independent of local illumination. As a result, shadowed areas do not affect the size of blobs within them, as shown in Figure 3.1*c,d*.[1]

Implementing such a scheme, of course, requires that cameras be calibrated. Most cameras have neither purely linear nor logarithmic response, suggesting the use of a look-up table for transforming their output to $I = \log b$.

[1]This assumes that illumination changes occur at a larger scale than reflectance changes do. Thin shadows can cause blobs too.

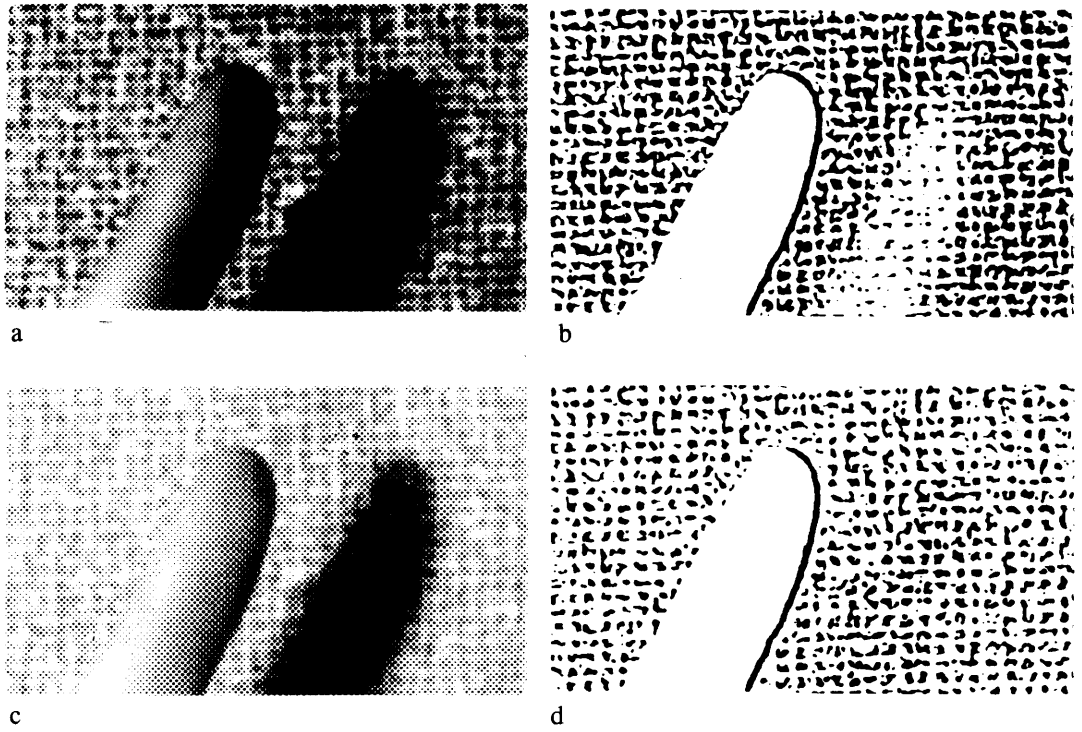


Figure 3.1 Discounting effects of illumination on blob size. (a) Image I of finger casting a shadow on burlap. Intensity values are roughly linear with brightness. (b) Shadow reduces sizes of blobs using $I * \nabla^2 G$. (c) Logarithm of image, $\log I$. (d) Density remains the same using $(\log I) * \nabla^2 G$.

3.2 Blob segmentation

Even though some noise has been removed, one cannot simply treat connected components of the thresholded $I * \nabla^2 G$ as textons, because these regions can span large distances in the image. In order to compute meaningful attributes of blobs, such regions must be segmented into localized blobs and bars. Thin connections between otherwise compact bars must be eliminated. Also, elongated regions must be separated at bends and intersections into roughly linear bars. Simple procedures for performing these two types of segmentation are described next.

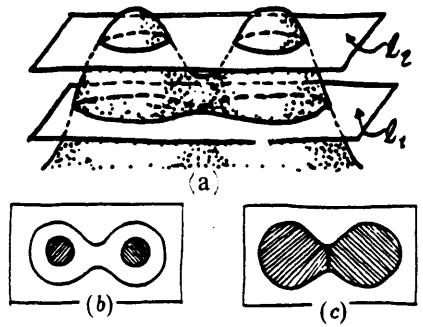


Figure 3.2 Segmenting regions into compact blobs using multiple levels. (a) The surface $I * \nabla^2 G$ sliced at levels ℓ_1 and ℓ_2. (b) The blob at ℓ_1 is split into two blobs at ℓ_2. (c) The blobs at ℓ_2 are labeled and grown over the original blob to segment it.

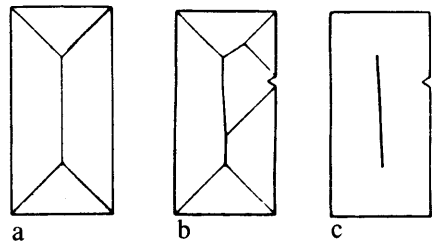


Figure 3.3 A more robust medial axis transform. The medial axis transform of a rectangle (a) is undesirably sensitive to slight changes in the bounding contour (b) [Marr 1982, p. 304]. Fitting straight line segments to the skeleton and removing those which are short relative to their distance from the perimeter yields a more robust representation (c).

3.2.1 Compact blob segmentation

Often, for a particular level, two nearby extrema of $I * \nabla^2 G$ may be joined into a single region, containing a thin neck. These connections are sensitive to the level chosen and may result in connected regions which span large portions of the image. Such regions are segmented using multiple levels. Whenever a higher level would result in a region being split into two or more regions, I use the new regions provided that they are relatively stable. The stability criterion (similar to the stability criterion for fingerprints proposed by Witkin [1983]) requires that the new regions exist over at least as large a range of levels as the underlying one. The splitting operation is applied recursively at several levels. Finally, the new, smaller regions are numbered and then grown back over the original regions in order to segment them. A stage of the process is illustrated in Figure 3.2.

3.2.2 Bar segmentation

Elongated regions are segmented at junctions and bends by using a refinement of the medial axis transform. The medial axis, obtained by thinning the region down to its skeleton, has been criticized for being sensitive to slight perturbations in the region contour (see Figure 3.3). However, a stable representation can be obtained by keeping only those segments of the skeleton which represent elongated sections of the region.

The pruning is accomplished as follows: Straight line segments are fit to the skeleton, with knot points placed at intersections as well as bends.[2] Each

[2]**Implementation Note:** The thinning algorithm and straight line fitting algorithm are described in Pavlidis [1982], on pp. 203–206 and pp. 283–292, respectively.

line segment corresponds to a piece of the skeleton. For each such skeleton segment, the average distance to the perimeter of the region (computed while thinning) is computed, and only segments whose lengths are longer than α_{min} times the average blob diameter are retained. At this stage, a maximum aspect ratio, α_{max}, can be imposed on bars, if desired, by splitting very long segments. In my implementation $\alpha_{min} = 1.5$ and $\alpha_{max} = 3.0$. The algorithm removes insignificant skeleton segments, retaining only those which represent elongated, bar-like parts of the region. The remaining segments are numbered and are grown back over the original region, segmenting it into approximately linear segments. Compact regions, which have no significant skeleton segments, are not segmented. Figure 3.4 shows a simple example.

The blob detection and segmentation operations are illustrated, step-by-step, in Figure 3.5.

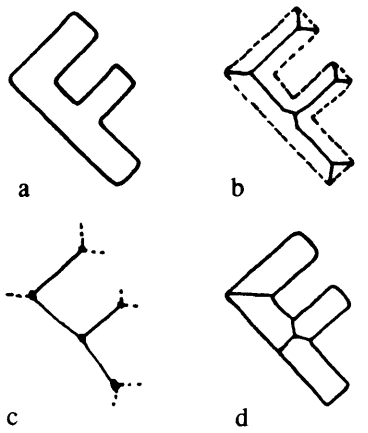


Figure 3.4 Segmenting regions into bars at junctions and intersections. (a) A bar-like figure to be segmented. (b) Skeleton (medial axis transform) obtained by thinning. (c) Straight line skeleton with segments having small aspect ratios removed. (d) Segmentation obtained by growing labeled segments back over original figure.

3.3 Attribute computation and removal of spurious blobs

At this point the blob regions have been segmented into localized blobs, whose attributes can be computed. The length, width, and orientation of each blob are computed from its second moments (see, for example, [Horn, 1986]). In the examples here, the geometric attributes are portrayed by rendering the blobs as rectangles or ellipses.

For each blob, average intensity and contrast are also computed. Contrast is computed by averaging the quantity $c(s) = \nabla I * G \cdot \vec{n}(s)$ along the perimeter, where $\vec{n}(s)$ is the outward-pointing unit normal to the blob's perimeter at point s;

$$\vec{n}(s) = \pm \frac{\nabla I * G}{\| \nabla I * \nabla^2 G \|}$$

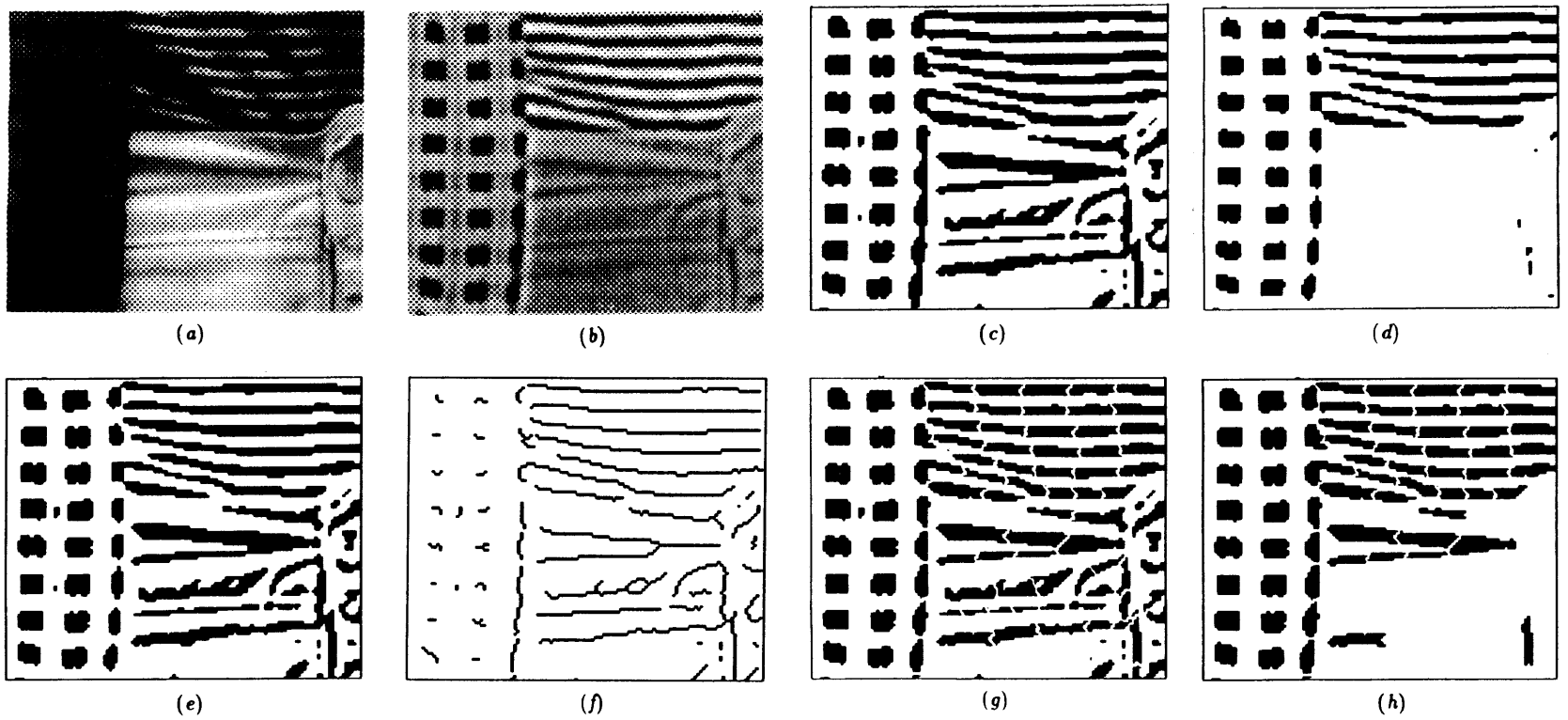


Figure 3.5 Steps of blob detection and segmentation. (a) Image I. (b) $I * \nabla^2 G$. (c) $I * \nabla^2 G > \ell_1$. (d) $I * \nabla^2 G > \ell_2$, $\ell_2 > \ell_1$. (e) Segmentation obtained by growing a over c. (f) Significant skeleton segments of e. (g) Segmentation obtained by growing f over e. (h) Result of removing spurious blobs from g. (i) Geometric blob attributes displayed as rectangles.

for dark and light blobs, respectively. Assuming a step edge of amplitude a, the contrast at a point on the perimeter is $c(s) = \pm a/2\pi\sigma$, where σ is the standard deviation of the Gaussian smoothing filter G. The contrast is negative for dark blobs and positive for light ones.

An intensity edge of a uniform region may result in a thin bar running alongside the edge. From the viewpoint of edge detection, the the artificial edge is known as a "spurious" crossing, and its strength (slope) can be higher than that of zero crossings which correspond to actual intensity edges [Berzins 1984]. Along a spurious portion of a contour, $(I * G)'(I * G)''' < 0$ in the direction perpendicular to the contour; hence $c(s)$ is positive for dark blobs and negative for light blobs at a spurious contour, and of opposite sign otherwise.

Only dark blobs having $c(s) > \tau = 2.5\varsigma$ along significantly more than half (60%) of their perimeter are maintained. (The computation of the gradient noise estimate, ς, is described in Section 2.1.) This removes any remaining blobs due to noise as well as any spurious blobs caused by strong edges along only one side of them. Analogously, significant light blobs must have $c(s) < -\tau$ along at least 60% of their contour. Tiny blobs, whose areas are less than $3\sigma^2$, where σ is the scale of blob detection, are also discarded as noise. Figures 3.6 and 3.7 show more examples of blobs detected from natural scenes. Figure 3.8 shows dark blobs detected in a random-dot counterexample to Julesz's conjecture.

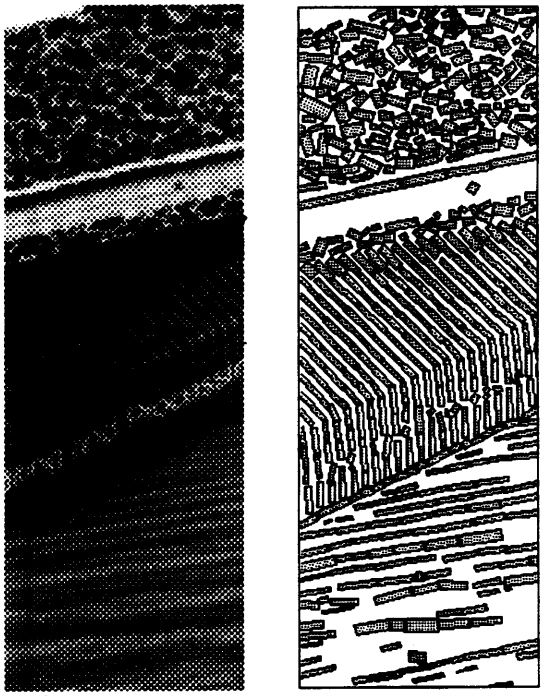


Figure 3.6 Blobs due to surface markings. (a) 160 × 454-pixel close-up of books on a wooden table top. (b) Strong, dark blobs found at scale $\sigma = 1.5$ pixels.

3.4 Analysis of blob detection algorithm

The proposed blob segmentation scheme is less biologically plausible than one which employs oriented feature detectors of various sizes and orientations (as

Figure 3.7 Blobs due to surface roughness. (a) A 352 × 389-pixel image of a pant leg and sock against a pitted ceiling tile. Color does not provide a cue for locating discontinuities in this scene, since all surfaces are white. (b) All dark blobs detected at scale $\sigma = 2$ pixels. (c) Strong blobs found.

implemented by Mahoney [1987] for line drawings). My method, however, postpones solving the problem of feature integration and is quite efficient, employing local operations. As currently implemented on a Symbolics 3600 Lisp Machine, the entire blob detection process takes about 6 minutes on the 160×454-pixel books image shown in Figure 3.6. Of this time, about 2 minutes are used for convolution, differentiation, and noise estimation; 3 minutes for blob segmentation (using 5 levels); and 1 minute for attribute computation and thresholding. On the Thinking Machines Connection Machine the algorithm is expected to take a few seconds.

3.5 Why blobs?

I conjecture that texture in natural scenes can be represented using blobs (as detected by my algorithm) and their attributes. This proposal contrasts with most implemented feature-based computations of texture, which employ edges. I claim that blobs are more useful than edges for representing tex-

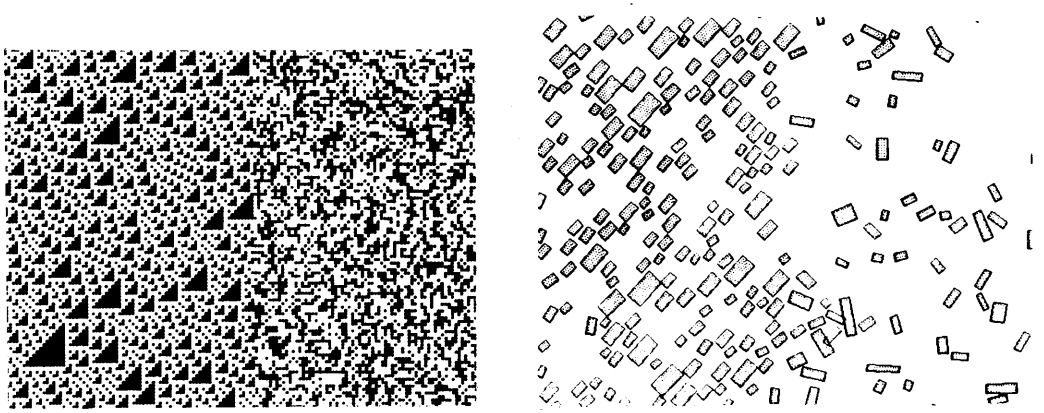


Figure 3.8 (a) In this counterexample to Julesz's conjecture, the two regions have identical second-order statistics but are discriminable [Julesz and Bergen 1983, p. 1630]. (b) The computed blobs, with geometric attributes shown as rectangles, provide a better explanation of texture discrimination.

ture for two reasons. First, attributes of blobs capture more information than attributes of edges. First-order statistics of blob attributes account for higher-order statistics of edge attributes. In particular, the average perpendicular distance between edges, a second-order measure used by some (e.g., Kjell and Dyer [1985]) are captured by the average width of the blobs between them. Second, blobs provide a more appropriate resolution for representing the physical events which cause intensity changes. A blob can describe the approximate location, size, and orientation of an impression in a surface, even when its bounding contour (i.e., set of edges) is not well defined.

Interestingly, in the images which I have examined, dark blobs are useful than light ones as textons. Small impressions in the surface, such as spaces between threads in a fabric, or gaps between leaves on a tree manifest themselves as small dark regions in the image. Less light is reflected from such surfaces because of shadows cast by objects nearer the light source, or because of the oblique angle of these surfaces with respect to the viewer. Even surface markings, such as type on a page, are more often than not dark on light, resulting

in isolated dark blobs.

How to automatically choose the appropriate scale for blob detection (σ) is not addressed here. One approach is to compute blobs at multiple scales, and integrate them into a single description, as Canny [1983] suggested for edges. Another possibility is to first determine the "natural scale" for blob detection. In each of my examples, a single scale sufficed. In general, though, different scales may be required over different parts of the image.

If desired, endpoints and crossings of bars can be identified as terminators and crossings. However, it is not clear from natural images that these features are as relevant to the interpretation of texture as some psychophysical experiments might suggest. As discussed in Chapter 6, some of Julesz's examples can be explained in terms of geometric and luminance attributes of blobs. The next chapter investigates which of these attributes are significant.

Chapter 4 **Perceivable Differences in Blob Attributes**

What attributes of textons are compared by the human visual system? Feature-based theories of texture vision mention density, size, orientation, and intensity of blobs. They do not, however, state how large a change in each of these attributes is necessary to cause a texture boundary. Without such data, it is hard to justify which attributes among a set of interrelated attributes are compared. If lengths and widths of blobs are compared, for example, is it also necessary to compare their areas? What about their elongations?

I am aware of little other work which measures the minimum differences in geometric attributes of blobs which cause texture boundaries. Julesz mentions blobs in his texton theory, but the psychophysical experiments he presents deal mainly with terminators and crossings as textons [Julesz and Bergen 1983; Julesz 1984, 1986].

I performed some informal psychophysical experiments to investigate exactly what kinds of changes in size, orientation, and density of blobs are perceived. One purpose of the experiments was to measure, for each attribute, the minimum magnitude of attribute change perceived as significant. Another goal was to infer from these minimum perceivable differences which, among a set of interrelated attributes, are compared by the human visual system.

The experiments employed synthetic images composed of filled-in rectangles or ellipses of varying dimensions, orientations and spacings. Each image consisted of two distributions of textons, one distribution per region. Usually, the attributes of textons were normally distributed in each region according to specified mean and standard deviation. Orientation was also modeled as a uniform distribution. I investigated under what circumstances the two distributions were discriminable.

Within each region, textons are nominally placed at a grid of locations, with specified horizontal and vertical spacing, but are offset by a random amount. The offset components are normally distributed according to a specified standard deviation, which typically equals 1/10 the mean spacing. This model of texton placement facilitates the creation of regions of uniform density and avoids overlapping textons, while providing just enough randomness to prevent the perception of colinear arrangements of textons.

In order to measure only preattentive capabilities, each image was exposed for 150 milliseconds followed by a masking pattern. This brief stimulus allowed little time for eye movement or, presumably, for scrutiny of individual elements. After each image was flashed, the subject was required to locate any perceived texture boundary. The boundary did not have to appear sharp, but had to be located approximately correctly. The actual boundaries were always circular arcs, with positions chosen randomly, but so that neither region was too small.

For each type of attribute change investigated, a series of images was used, with different magnitudes of change in each. Each such set of images investigating a particular kind of attribute change constitutes an experiment. During any one session, the subject was shown images from several experiments, in random order, so he would not become tuned to the type of changes being

investigated.

Of course, the minimum perceivable difference in an attribute is probably not a precisely defined quantity, and the discriminability of a texture boundary can vary from one observation to another. Hence, each experiment was performed a number of times, and the minimum difference boundary correctly located over half the time is reported. Each result is illustrated with a figure showing the approximate minimum difference perceived. While these figures, viewed attentively, do not verify the results, they do illustrate magnitude of the changes involved.

4.1 Orientation differences

Orientation is the easiest geometric attribute to study, since it is the only one which is not confounded with others. Apparently this is why orientation examples are often used by psychologists to demonstrate a perceptual capability, e.g., the ability to discriminate two sharp modes from one.

Experiment 1. The first orientation experiment measured the minimum perceivable difference in orientation, assuming no deviation. Orientations which might result in colinear grouping (multiples of 45 degrees, given the nominal square grid) were carefully avoided. Images having different orientation differences were shown, and the minimum difference preattentively perceived was about 15°. Figure 4.1 shows this critical difference, along with a difference of 10° which is not preattentively perceivable, and a difference of 20° which certainly is. Rectangles were used in the orientation experiments since their orientations seemed more salient than those of ellipses of the same dimensions.

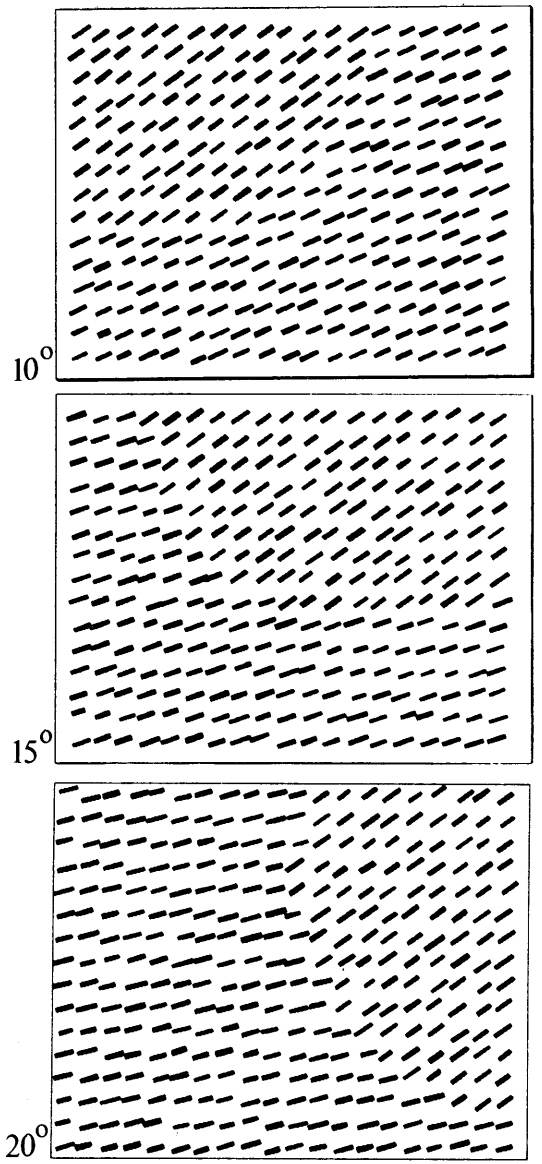


Figure 4.1 Orientation differences of 15° *or more are preattentively perceivable.*

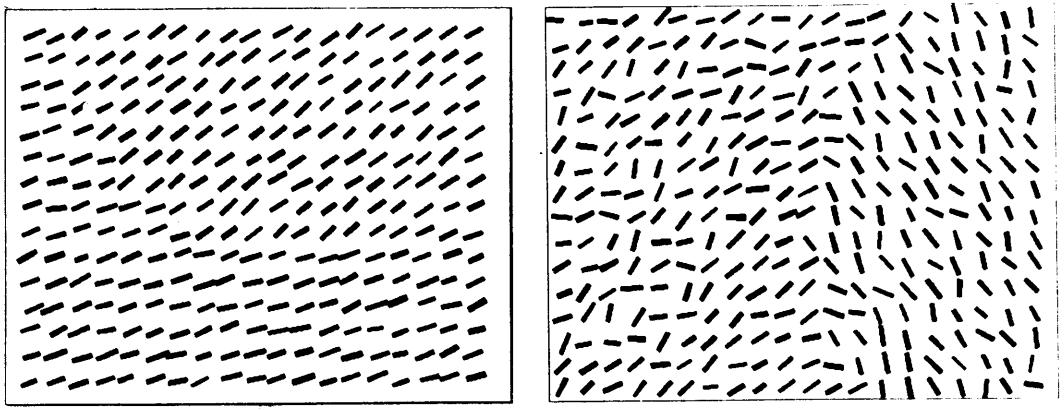


Figure 4.2 (a) At a difference in means of 15°, standard deviations up to 5° are perceivable. (b) At a 90° difference, the maximum standard deviation is about 20°.

Psychologists have noted that the visual system has a special sensitivity for vertical and horizontal orientations. Treisman, for example, found that bars which differ slightly in orientation from a homogeneous field of surrounding bars are spotted faster if their orientation is vertical than if the surrounding bars are vertical [Treisman 1985]. It would not be surprising if humans exhibited increased orientation discriminability at vertical and horizontal orientations. In the above experiment, the average of the two orientations was always about 25°. It would be interesting to repeat the experiment, using different pairs of orientations, to determine whether the minimum perceivable difference is orientation dependent.

Experiment 2. The next experiment investigated the effect of non-zero variance on orientation discriminability. In each image, the standard deviations of orientation in both regions were equal. Orientations which differ by 15° remain discriminable up to standard deviations of about 5°, as shown in Figure 4.2*a*.

One would expect larger standard deviations to be tolerated as the difference in means increases. This experiment measured the maximum standard deviation at the maximum possible difference in means, 90°. Indeed, standard deviations up to about 20° are discriminable, as shown in Figure 4.2*b*.

Experiment 3. The maximum standard deviation of a normal orientation distribution preattentively distinguishable from a uniform orientation distribution was investigated. The result, between 12° and 15°, is pictured in Figure 4.3.

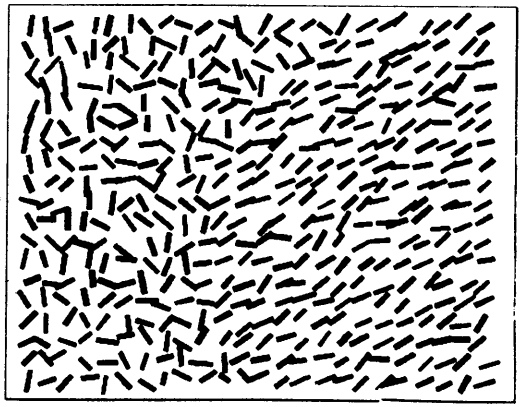


Figure 4.3 The maximum standard deviation of a unimodal orientation distribution preattentively discriminable from a uniform distribution is about 15°.

Experiment 4. This experiment measured whether two sharp modes were discriminable from a uniform distribution. Avoiding horizontal or vertical orientations, I found that two modes were not usually preattentively discriminable from a uniform distribution. An example is shown in Figure 4.4. The result agrees with Riley's [1981] conclusion which is based on preattentive shape recognition from texture boundaries.

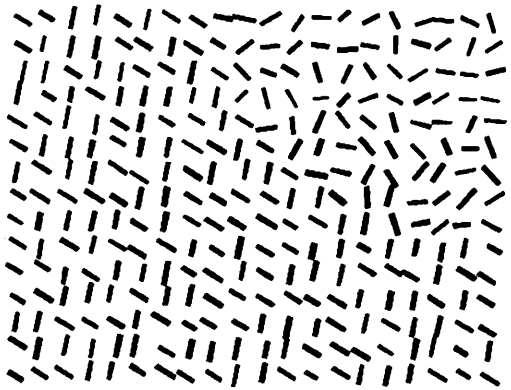


Figure 4.4 A bimodal orientation distribution is not preattentively discriminable from a uniform distribution.

4.2 Size and density differences

The individual effects of blob density and blob size attributes are harder to measure because they are interrelated. It seems reasonable to compare any of the following attributes: length, width, elongation (length divided by width), area (length times width), texton area density (fraction of area occupied by textons), and texton density (number of textons per unit area). Since it is impossible to vary one of these attributes without affecting others, several experiments are needed to measure the minimum perceivable difference in any single attribute, and to determine whether the attribute need be compared.

I assumed that size and density comparisons were scale-independent, and hence, that critical differences should be expressed as ratios rather than as differences of the attributes. For consistency, I express such ratios as values greater than one. It is often more convenient to use the logarithms of the attributes, in which case the absolute difference of values is the appropriate measure of comparison.

In the synthetic images the attributes most directly specified are width, length, and number density. (The number density equals one divided by the square of the specified mean spacing.) I denote the logarithms of width, length, and number density by w, l, and n, respectively. These three attributes are independent, and provide a convenient basis for evaluating the other attributes listed above. The logarithm of area, a, equals $l + w$, while the logarithm of elongation, e, equals $l - w$. The logarithm of texton area density, α, equals $l + w + n$. The remaining experiments are designed to determine which of these attributes the preattentive visual system compares, and the minimum perceivable difference for each.

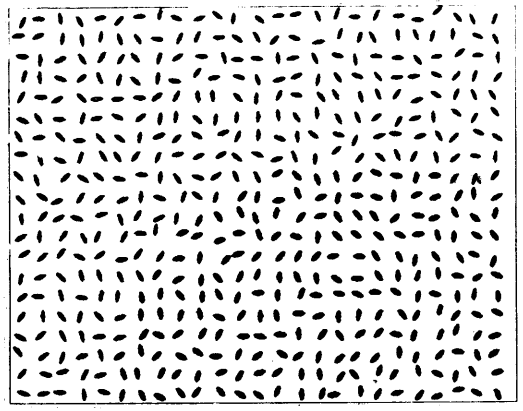


Figure 4.5 With lengths held equal, the minimum perceivable width ratio is about 1.25.

4.2.1 Size differences

The first set of size and density experiments measured perceivable differences in texton size attributes, keeping texton number densities fixed.

Experiment 5. The minimum perceptible difference in width was first measured, keeping lengths and number density equal between the two regions. Testing the width ratios above 1 at intervals of 0.05, a minimum perceivable width ratio of 1.25 was required for preattentive discrimination. This difference is pictured in Figure 4.5.

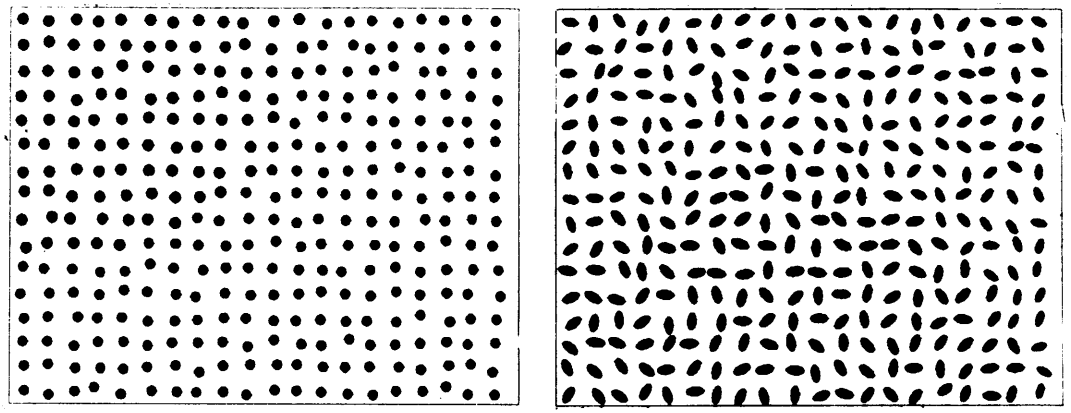


Figure 4.7 When widths and lengths varied by the same amount, a minimum area ratio of 1.25 is required for discrimination. This difference is shown for (a) compact and (b) elliptical blobs.

As expected, increasing the variance of the width distributions required larger differences in means for discrimination. All the size and density results reported here were obtained using zero standard deviations (delta functions) for the attributes under investigation.

Experiment 6. A similar experiment was performed for length, holding widths and number densities equal. The minimum perceivable length ratio was around 1.25 to 1.30, as shown in Figure 4.6.

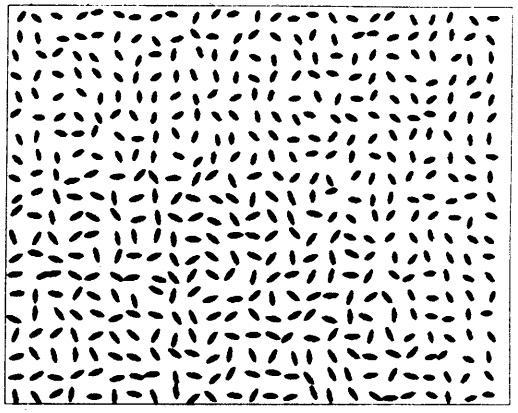


Figure 4.6 With widths held equal, the minimum perceivable length ratio is about 1.25 also.

Experiment 7. The minimum perceivable difference in texton area was measured by keeping length proportional to width. For both elongated (elliptical) and compact (circular) blobs, a minimum perceivable area ratio of 1.25 (to the nearest 0.05) was obtained; this difference is shown in Figure 4.7. It seems, then, that the width and length results obtained in Experiments 5 and 6 might alternatively be explained by significant differences in texton area.

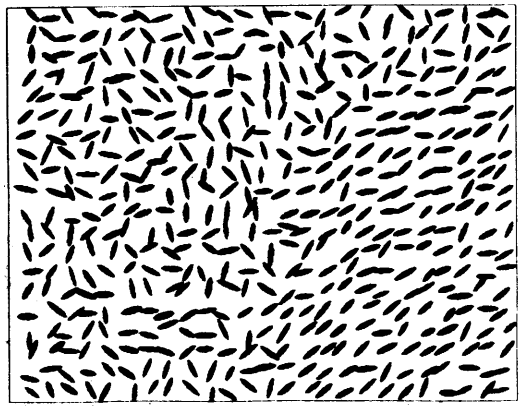


Figure 4.8 When blob area is held constant, the minimum perceivable ratio in length or width is still about 1.25.

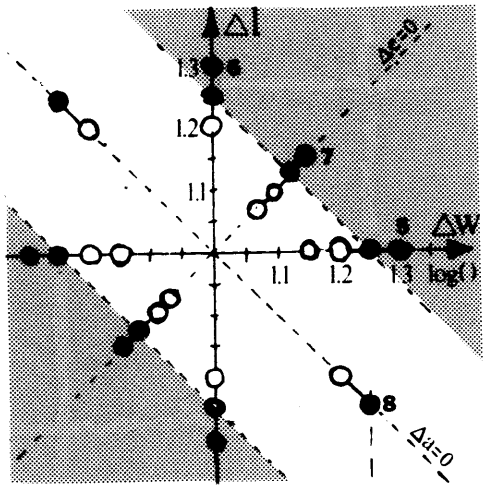


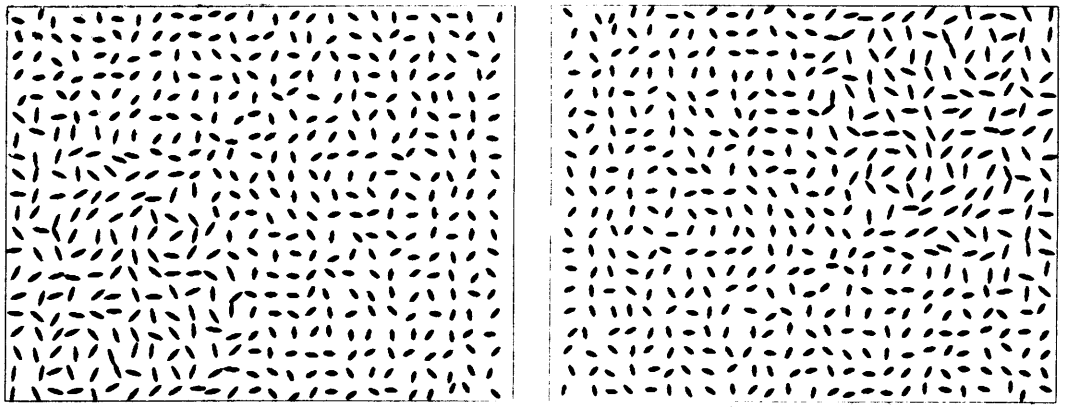
Figure 4.9 The results of Experiments 5–8, plotted on the plane $\Delta n = 0$. The region $|\Delta a| > \log 1.25$ is shaded.

Experiment 8. In this experiment, texton elongation was investigated by varying texton length inversely with width, so the areas remained equal, as in Figure 4.8. The minimum perceivable difference occurred at width and length ratios of about 1.25, at which point the elongation difference was about 1.55, as shown in the figure. Unlike area, elongation could not account for the results of Experiments 5 and 6, where the elongation ratios at perceivable differences were only 1.25 to 1.30.

The results are conveniently analyzed by plotting them as shown in Figure 4.9. The horizontal axis represents change in the logarithm of width, Δw, while the vertical axis represents change in the logarithm of length, Δl. The data points examined near the critical differences are shown, as well as the minimum perceivable differences themselves. A filled in circle represents a perceivable difference (i.e., the boundary was correctly located more than half the time), while an open circle represents an imperceivable one (in which the boundary was correctly located less than half the time). The results of Experiment 5 are plotted on the line $\Delta l = 0$, Experiment 7 on the line $\Delta l = \Delta w$, etc. Each experiment generates a symmetric set of points, since the direction in which the difference is computed is arbitrary.

The hypothesis that independent comparisons of these attributes account for texture discrimination is expressed graphically as follows: For each logarithmic attribute compared, a pair of lines is drawn normal to the axis of the attribute, intersecting the axis at plus and minus the minimum perceivable difference. If the hypothesis is true, then a set of these lines will delimit imperceivable from perceivable texture differences, i.e., the open circles from solid ones. The task is to find a set of such lines, if any.

Using this technique, it is clear that texton area must be computed, since,



a b

Figure 4.10 (a) At $\Delta w = -\log 1.15$, $\min \Delta l = 1.30$. *(b) At* $\Delta l = -\log 1.15$, $\min \Delta w = \log 1.25$.

among the attributes tested, only the constraint $\Delta l + \Delta w > \log 1.25$ accounts for the result of Experiment 7. As shown by the shading in Figure 4.9, this constraint might also account for the width and length results, given the data so far.

It is not clear which of the attributes length, width, and elongation are compared. The remaining set of data points, from Experiment 8, does not provide enough information to say which constraint or constraints enclose the region of perceptible size differences. To determine this, two more experiments were performed.

Experiment 9. With a log width difference of $\Delta w = -\log 1.15$, the minimum positive difference in log length was measured. The critical difference, shown in Figure 4.10*a.* is approximately $\Delta l = \log 1.30$. Distributions with differences $\Delta w = -\log 1.10$ and $\Delta l = 1.30$ were also found to be discriminable.

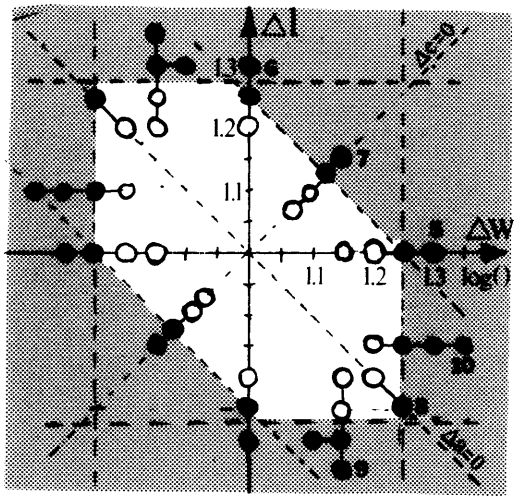


Figure 4.11 Experiments 5–10 suggest that width, length, and texton area are computed, as indicated by the shaded regions.

Experiment 10. The converse experiment was performed, holding the log length difference at $\Delta l = -\log 1.15$ while width differences were varied. The minimum log width difference, shown in Figure 4.10 *b* , is about $\Delta w = \log 1.25$.

These results are added to the graph in Figure 4.11. Together, the data imply that length and width are compared, and perceivable differences are approximately defined by the constraints, $\Delta w > \log 1.25$ or $\Delta l > \log 1.30$. The region defined by the length, width, and area constraints is shaded. It is apparent that comparing elongations is not necessary, since doing so would not further constrain the space of imperceivable differences.

The size experiments above suggest that texton widths, lengths, and areas are compared. So far, however, we have only considered data in one plane of the three-dimensional space parameterized by Δw, Δh, and Δn, namely, the plane $\Delta n = 0$, since texton number densities were held equal in each image. Each attribute comparison by the human visual system, if independent, determines a plane in this space. Instead of comparing w, l, or a it is theoretically possible that $w \pm kn$, $l \pm kn$, or $a \pm kn$, respectively, are compared, where k is a constant, given the experimental data. Of these possibilities, only $\alpha = a + n$, the texton area density, has an obviously physically significant meaning. In Section 4.3 I argue, with some psychophysical evidence, that $w \pm kn$ and $l \pm kn$ are not computed. The data therefore suggests that w, l, and a or α are compared. Determining whether a or α is compared requires experiments involving differences in texton number density. These experiments are described next.

4.2.2 Density differences

Texton number and area density is investigated in this set of experiments. The purpose of these experiments is to determine the minimum perceivable differences in these attributes, and to suggest which of the density and area attributes n, α, and a are compared.

In this case, the experimental results are to be plotted in a plane indexed by Δa and Δn. The plane corresponds to a slice of the Δw-Δl-Δn space at $\Delta w = \Delta l$. Therefore, in each image used in these experiments, texton width is varied by the same ratio as texton length between the two regions. Both compact or elongated texton are used.

We already have one set of data points on this plane from Experiment 7, plotted along the line $\Delta n = 0$. The remaining experiments seek to determine whether the discriminable region is bounded by vertical lines, suggesting that area is compared; horizontal lines, suggesting that number density is compared; or diagonal lines, suggesting that area density is compared.

Experiment 11. In Experiment 7, texton area density was varied by varying texton area while keeping texton number density constant. This experiment tested the opposite case: Texton number density was varied (by varying the mean spacing), while texton dimensions were held constant. The minimum perceivable number density ratio was about 1.30 for both compact and elongated textons; the latter case is shown in Figure 4.12.

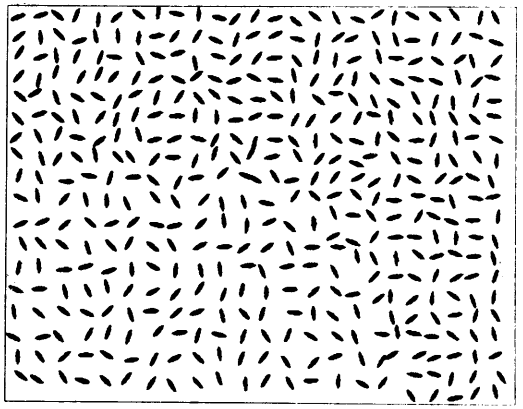


Figure 4.12 The minimum perceivable difference in texton number density, with areas held equal, is about 1.30.

Experiment 12. In this experiment, the texton area density was varied by varying both texton number density and texton area in the same direction, i.e., $\Delta n = \Delta a = \Delta\alpha/2$. The minimum texton area density ratio perceiv-

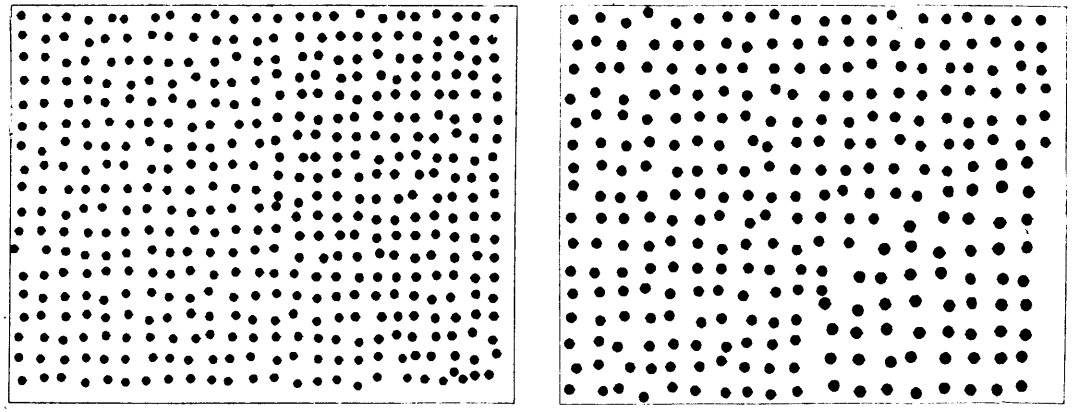


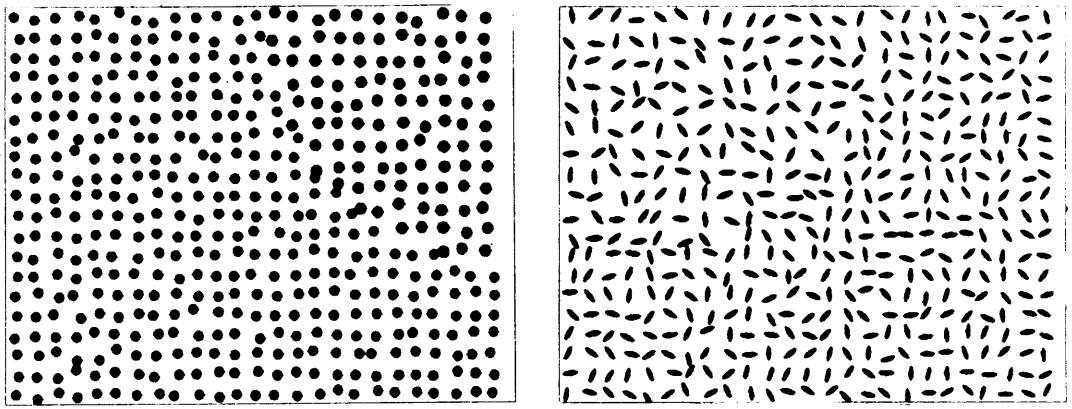
Figure 4.13 (a) When both texton number density (spacing) and texton area are varied together, a texton area density of at least 1.30 is required for discrimination. (b) When these quantities are varied inversely, so that texton area density remains constant, a minimum number density ratio of about 1.30 is required for discrimination.

able was about 1.30, at which point the number density and area ratios were approximately 1.15. The critical difference is shown in Figure 4.13*a*.

Experiment 13. This time, the texton number density was varied inversely with the texton area, i.e., $\Delta n = -\Delta a$, so $\Delta \alpha = 0$. The minimum perceivable number density ratio, shown in Figure 4.13*b*, was about 1.30.

Experiment 14. The texton area ratio was held at 1.15, while number density was varied in the inverse direction. Under this constraint, the minimum number density ratio perceived was approximately 1.25, as shown in Figure 4.14*a*.

Experiment 15. The texton number density was held at 1.15, while the area ratio was varied in the opposite direction. The minimum perceivable texton area ratio was about 1.30, as shown in Figure 4.14*b*.



a b

Figure 4.14 (a) With $\Delta n = -1.15$, and Δa positive, $\Delta a > \log 1.25$ yields discriminable textures. (b) With $\Delta a = -1.15$, and Δn positive, $\Delta n > \log 1.30$ yields discriminable textures.

The data points tested in the experiments are plotted in Figure 4.15. Experiments 7, 11, and 12 demonstrate that texton area density must be computed, and that $\Delta\alpha > \log 1.25$ is perceivable. Experiments 11, 13, and 14 define a horizontal boundary of the imperceptible region of the Δa-Δn plane, since $\Delta n > \log 1.25$ is perceivable. Hence, texton number density (or the square root of its inverse, average spacing) appear to be compared.

The results of Experiments 15 and 13 suggest that texton area is also compared, and that the critical ratio is around 1.30. Additional experiments should be done at other points in the neighborhood $\Delta a = -1.25$ to -1.35 and $\Delta n = 1.05$ to 1.25 to check this critical value, which is based only on Experiment 15. If it is correct, then it is unlikely that $\Delta a - \Delta n$, is compared, since it would not further constrain the region of indiscriminable texture pairs. This quantity, which represents the product of texton area and surrounding empty space, is implausible as an attribute anyway.

From this set of experiments, it appears that texture boundaries are perceived when the ratios of texton number density or texton area density exceed

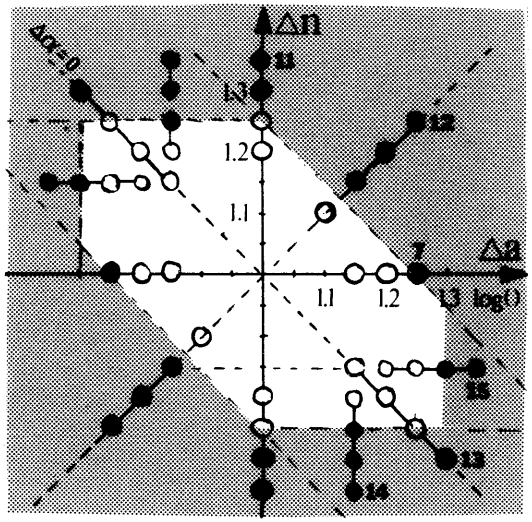


Figure 4.15 Experiments 7 and 11–15 suggest that texton number density, texton area density, and possibly texton area are compared, as indicated by the shaded regions.

about 1.25, or when the ratios of texton area exceed about 1.30. These conclusions are based on the assumption that density comparisons are not biased with respect to changes in texton width or length. Geometrically, I assumed that the bounding planes in Δw-Δl-Δn-space are perpendicular to the plane $\Delta w = \Delta l$ in which the data was sampled. The next section presents some evidence that these assumptions are reasonable.

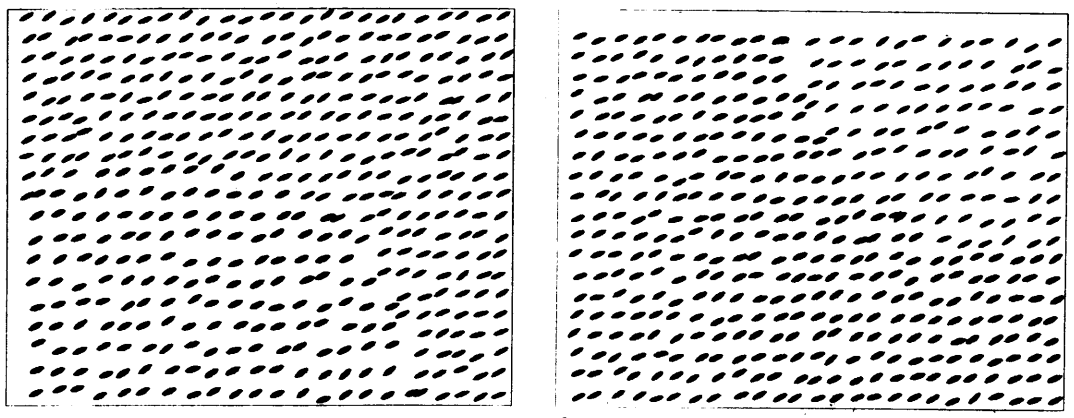
4.3 Combinations of width and number density differences

The preceeding two sets of experiments assumed that number density n is combined with width and length homogeneously, i.e., that n, $n + l + w$, or $n-l-w$ are compared, but that $n \pm w$ and $n \pm l$ are not. This assumption seems reasonable, since it reflects attributes we commonly refer to. Notwithstanding, the significance of the less intuitive combinations can be tested in the same fashion.

Two experiments were performed to investigate the possibility that $w - n$ or $w + n$ (number density normalized or scaled by width) are compared.

Experiment 16. The minimum difference in width was measured, holding lengths and texton area density constant. That is, $\Delta l = 0$, while $\Delta n = -\Delta w$. The minimum width ratio was still approximately 1.25 in this case, as shown in Figure 4.16*a*.

Experiment 17. The minimum difference in width was measured varying texton number density with width, i.e., $\Delta n = \Delta w$, as $\Delta l = 0$. Again, width ratios of 1.25 or more were perceivable, as shown in Figure 4.16*b*.



a b

Figure 4.16 (a) When texton width is varied inversely with texton number density, so that texton area density remains constant, the width minimum width ratio perceivable is approximately 1.25. (b) When width is varied with number density, the minimum width ratio is also around 1.25.

As shown in Figure 4.17, the experimental data implies that w and n, rather than $w - n$ or $w + n$, are computed. The data from this plane, combined with the results from the other two planes explored, provides persuasive evidence for independent computations and comparisons of n and w. Presumably, an experiment involving n and l would yield analogous results. The result suggests that Experiments 5–15 indeed identified the appropriate measures for size and density discrimination.

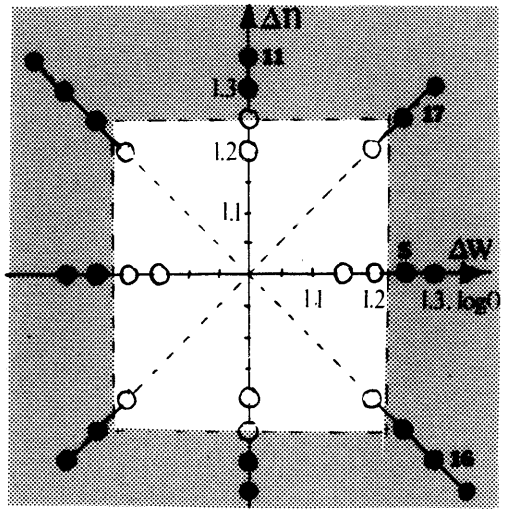
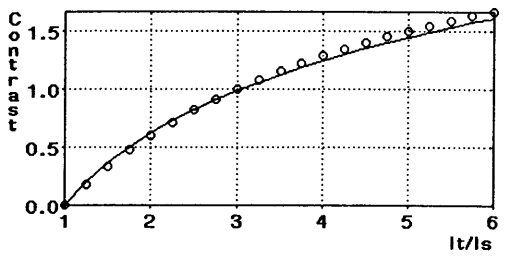


Figure 4.17 The results of Experiments 16 and 17 suggest that Δw and Δn, rather than $\Delta w \pm \Delta n$, are compared.

4.4 Luminance differences

From the standpoint of recovering information about the surfaces in a scene, blob contrast provides a more useful attribute than average blob intensity. As argued in Section 3.1, the difference in logarithms of blob to background intensity provides an illumination-independent measure of blob reflectance.

Interestingly, this corresponds nicely with Beck's measurements of human contrast discrimination [Beck et al. 1987]. Using synthetic images of squares,



Plot of Rayleigh contrast (circles) and $\beta \log(I_t/I_s)$ *(curve) show that these two contrast measures are almost identical over normal range of* I_t/I_s*. Here,* $\alpha = 0.25$ *and* $\beta = 0.9$.

he and his associates found that the Rayleigh contrast function closely describes the perceptual strength of relative blob intensities. Rayleigh contrast is defined as $(I_t - I_m)/I_t$, where I_t is the intensity of the texton and I_m is the "local mean intensity", i.e., $I_m = \alpha I_t + (1 - \alpha)I_s$, where I_s is the intensity of the surround. This function corresponds closely to the proposed difference of logarithms measure over reasonable ranges of intensity ratios I_t/I_s and ratios of center to surround sizes α, as shown at right.

Beck et al. [1987] determined that, when comparing fields of squares having contrast of the same sign, the perceptual strength of segregation varied approximately linearly with the ratio of the blobs' Rayleigh contrast measures. This suggests that, as with the geometric attributes, the logarithm of contrast is the appropriate measure to compare. They also discovered that differences in the sign of contrast, even small ones, yielded strong segregation. This finding is consistent with physiological evidence for on-center and off-center cells, they noted. Computationally, Beck's finding supports the separate detection of dark and light blobs, and the comparison of densities of dark and light blobs without regard to contrast magnitude.

Finally, Beck et al. [1987] discovered that the size and contrast of the squares were not independent attributes, but that increases in contrast could compensate for decreases in size. They note that the result is consistent with the computation performed by spatial frequency channels. With respect this proposal, their result suggests that texton maps from different scales are not integrated into a single map before detecting texture boundaries. Hence, the use of textons at a single scale, as implemented here, may not be unrealistic.

4.5 Summary and discussion

The psychophysical experiments presented in this chapter offer answers to two questions:

- Which attributes of blobs are compared?
- And what magnitude of changes are perceived?

Interrelated blob size and density attributes were sorted out by using texton width, length, and number density as a basis. It was determined that two homogeneous textures are discriminable if any of the following conditions are met:

- Orientation: $|\Delta\theta| > 15°$;
- Width: $|\Delta w| > \log 1.25$;
- Length: $|\Delta l| > \log 1.25$;
- Area (possibly): $|\Delta a| > \log 1.30$;
- Number density: $|\Delta n| > \log 1.30$; or
- Area density: $|\Delta \alpha| > \log 1.25$.

Each of these attributes appears to be compared independently.

The experiments were performed on one subject, a master's degree candidate, age 25, with corrected-to-normal vision. The subject was not naive as to the purpose of the experiments. (However, for any particular test image, the type and magnitude of texture boundary, as well as its location, was unknown to him.) More extensive and more controlled experiments should be performed to obtain more precise estimates of "average" performance.

Nevertheless, the minimum perceivable differences are compatible with previously reported results. Schatz [1977] reports that differences in dot density of 20% are perceivable, close to my figure for texton area density (20 or 25 percent). The existence of orientation channels which span roughly 15° is also documented [Schatz 1977]. The results obtained here for isolated blob textures are somewhat larger than those measured by Julesz and Bergen [1983] for periodic bars. They determined that orientation is judged with accuracy of 6° and that width is estimated to within 4–6% error, suggesting that no more than twice these figures are necessary for discrimination.

Despite experimental uncertainty in the values of the minimum difference figures, small changes in these values are not likely to affect the conclusions concerning which geometric attributes of textons are compared (with the possible exception of texton area). It is unlikely that unusual combinations of attributes (such as $n - l$) are compared for compact blobs.

From the standpoint of locating physical discontinuities in the scene, the minimum perceptible difference values make sense. Small changes in the orientation or contrast of textons are usually due to noise or gradual changes in the surface, while sharp changes are more likely the result of physical discontinuities.

4.5.1 Role of filtering

One might object that the experiments above ignore the role of spatial filtering. In particular, the results are based on the assumption that the blobs detected by the visual system match the textons of the synthetic images. Because the textons are elliptical and fairly well isolated, at small scales of filtering (relative to the blob size) this assumption is valid. Figure 4.18 shows elliptical textons

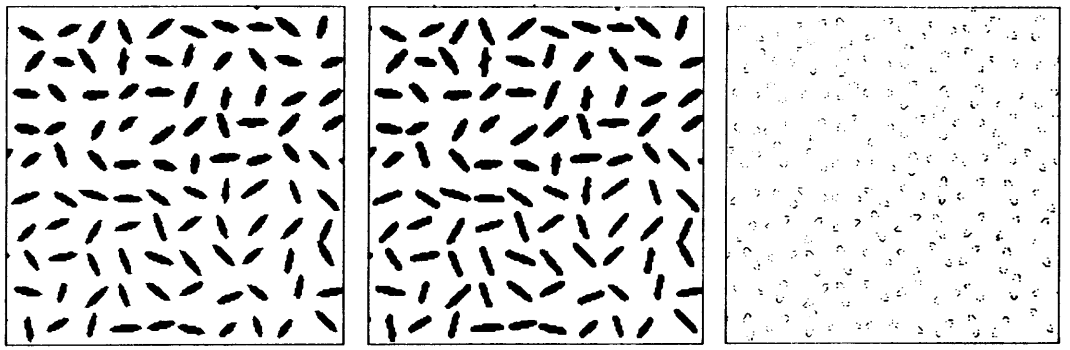


Figure 4.18 (a) Image of ellipses of different widths. (b) Blobs detected at scale $\sigma = 1.5$ pixels. (c) "Exclusive or" of a and b show their similarity.

differing in width by 25%, along with blobs detected at a single scale. As seen by the "exclusive or" of their bit maps, the blobs correspond closely to the original image. It seems unlikely that contrast differences due to differences in blob size are high enough in any of the experiments to account for differences in geometric attributes.

The use of logarithms of geometric attributes was based on the assumption that comparisons of size and density are scale-independent. Experiments involving textons of different sizes should probably be performed to check the validity of this hypothesis.

4.5.2 Local computation of density

A blob or edge is not perceived as a texture element unless it lies in a field of elements which is sufficiently dense. Julesz [1986] determined that for compact elements to be perceived as textons, the spacing between them should not exceed twice their diameter. Local computations of density, then, may serve two purposes: First, to detect density boundaries, and second, to deter-

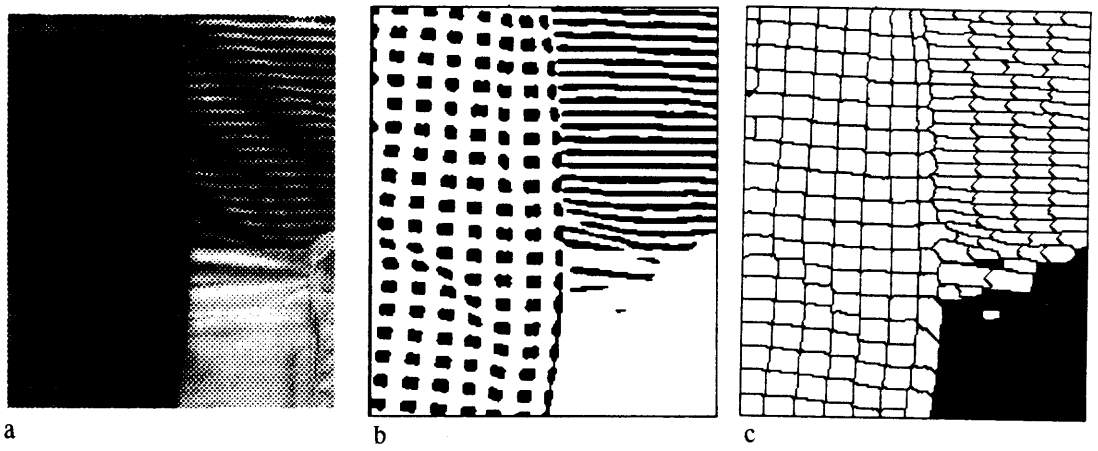
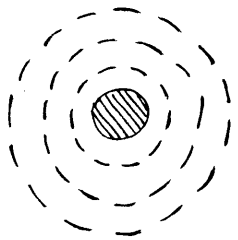


Figure 4.19 (a) Image of mismatched clothing and (b) dark blobs detected. (c) Partial discrete Voronoi tesselation obtained by growing in parallel each labeled blob by one pixel in all four directions, by up to 1.5 times its width.



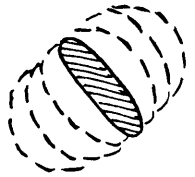
Proposed nearest-neighbor distance detector for compact blobs

mine whether the density is high enough in a neighborhood to compare other attributes.

Marr [1976*a*] suggested that the distance to the nearest one or two neighbors be computed for each texton. On a computer, this attribute can be efficiently computed by growing the blobs of a numbered blob map in parallel, resulting in a discrete Voronoi tesselation of the plane, as shown in Figure 4.19. Biologically, one might imagine that each feature detector is surrounded by a series of concentric detectors of larger sizes which measure the distance to the nearest texton. Such detectors could also determine whether the blob was sufficiently close to other blobs to constitute a texture element. Perhaps this inter-blob spacing is is the human visual system's local measurement of texton number density.

For a bar which is part of a longer line, the relevant nearest neighbor distances are in the directions perpendicular to the bar's major axis. Otherwise, its two nearest neighbors are probably its colinear, adjacent neighbors, and

clearly, a set of bars constituting a single line does not constitute a texture. For this reason, I suggest that distance detectors surrounding a bar detector extend only in the direction across the bar, as shown at right.



Proposed nearest-neighbor distance detector for bars

Furthermore, I believe that Julesz's minimum spacing criterion can be generalized to elongated blobs as follows: A blob is perceived as part of a texture if the distance between it and its nearest neighbor (or neighbors, measured as indicated above) is less than two or three times the blob *width*. The proposed criterion subsumes that of compact blobs, and seems plausible for parallel bars. Whether it accurately describes the perception of unconnected, randomly oriented bars is not as obvious. Further psychophysical investigation of this question would be useful.

Some of the density experiments should also be repeated using closely spaced, very long, parallel bar regions. Since the lengths of bars segments are not well-defined, it would not be surprising if density attributes were asymmetric with respect to width and length in this case.

4.5.3 Other evidence of independent comparisons of attributes

Treisman [1985] presents persuasive evidence that preattentive texture discrimination involves the comparison of individual attributes, but not conjunctions of them. While it is easy, for example, to detect a boundary between elongated blobs and compact blobs, and between dark and light blobs, it much harder to discriminate a region of dark elongated and light compact blobs from a region of light elongated and dark compact blobs. Her experiments indicate that while the former task can be done preattentively, the latter involves serial search.

The discovery that minimum perceivable differences depend on single at-

tributes, and not combinations of them (other than those attributes so defined) are consistent with Treisman's results. They too suggest that preattentive texture discrimination is based on the independent comparison of texton attributes (except for the possible interaction of contrast and width, mentioned previously). It will be interesting to see whether physiologists discover pathways for the various attributes, just as they have for different visual processes, including intensity, color, stereo, and motion [Treisman 1985].

Chapter 5 **Comparing Attribute Distributions**

Chapter 3 showed how textons can be extracted from natural images. Chapter 4 investigated what kinds of changes in the attributes of these textons cause texture boundaries. This chapter proposes how such texture boundaries can be located.

One approach is to test the existence of a texture boundary at an edge detected by some other means, e.g., an intensity edge. This hypothesize-and-test strategy is discussed in Chapter 7. Humans, however, have the ability to identify texture boundaries in the absence of other cues, as demonstrated by many of the synthetic images in the previous chapter. The independent computation of texture boundaries is investigated here.

Both locating and testing the existence of a texture boundary involve the comparison of attributes of textons lying in neighborhoods on either side of a line segment. The last chapter presented evidence that each attribute can be treated independently. The problem of identifying texture boundaries can therefore be posed as the question, How are two sample distributions of texton attribute values compared?

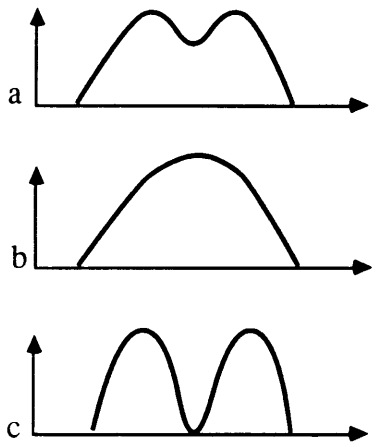
5.1 Statistical approaches

The comparison of two sets of samples to determine whether they are from the same distribution is a major subject of statistics. However, the standard statistical tests are not appropriate for detecting texture boundaries.

Using a two-sided test, one tests the null hypothesis, $H_o : f_1 = f_2$ (there is no texture boundary), against the alternative hypothesis $H_A : f_1 \neq f_2$ (there is a texture boundary). A statistic is computed, from which the probability that the differences are due to chance is determined.

Simple parametric tests, such as the Student's T test of means and the Snedecor F test of variances, are not accurate for this application, because attributes of textons are generally not normally distributed.[1] A distribution can have more than one mode. Even unimodal distributions are not necessarily normal, especially since natural distributions of textons often contain outliers resulting from other "events", such as wrinkles in the surface or shadows.

The existence of multimodal distributions and outliers suggests that a parametric approach could be employed if the histogram were first parsed into individual modes. Rules, together with parametric tests, would then determine whether the two distributions are equal. To be robust though, multiple scales of analysis, and rather complex matching rules would be required (e.g., two overlapping modes should sometimes be interpreted as one, as shown at left). This symbolic approach might be applicable to attentive texture discrimination, where the subject can say how the two distributions differ. But as a model of preattentive vision—which apparently tells the subject "where" but



The complication with the parsing approach: Distribution a is similar to both distributions b and c.

[1] de Souza [1983] has applied these tests to scan lines of intensity pixels to locate edges in x-ray images.

not "what" texture boundaries occur [Julesz 1986, p. 245]—this strategy seems unnecessarily complex. A non-parametric approach is more plausible.

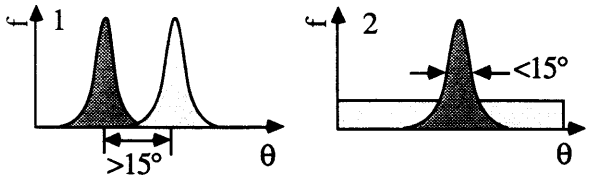
However, most non-parametric and parametric tests suffer from another problem: sensitivity to texture gradients. A systematic change in attribute values between two regions would yield high confidence that a texture boundary exists, even when the difference is imperceivable. Making such tests insensitive to imperceivable differences suggests that histograms, rather than exact values, of the samples are compared.

The idea, then, is to see whether a non-parametric statistic, computed on histograms of the sample distributions, can identify texture boundaries. The statistic and its threshold are based on the capabilities of preattentive perception, rather than on a theoretical probability that the underlying distributions are equal.

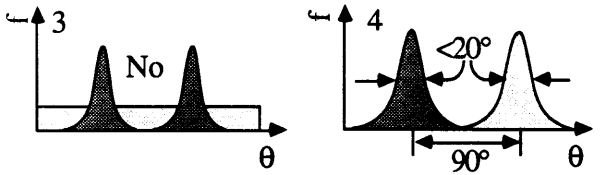
5.2 A proposed statistic for preattentive texture discrimination

Let us start by considering orientation. The capabilities of human texture vision hint at how humans compare orientation distributions. In the last chapter, the following capabilities, illustrated at right, were measured:

- *Capability 1:* Two single-valued distributions are discriminable if they differ by at least 15°.
- *Capability 2:* A single-valued distribution is distinguishable from a uniform distribution. More generally, a normal distribution is distinguishable from a uniform distribution if its standard deviation is less than about 15°.



Capabilities of preattentive orientation discrimination



Capabilities of preattentive orientation discrimination (continued)

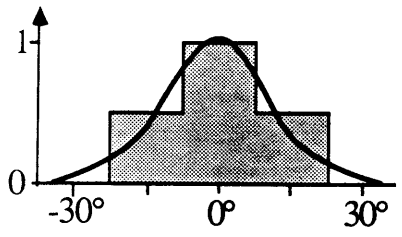
- *Capability 3:* A two-valued distribution is not usually distinguishable from a uniform distribution.
- *Capability 4:* Two normal distributions with means which differ by 90°, and equal standard deviations, σ, are discriminable if σ is less than about 20°.

What computation is consistent with the capabilities of human texture perception listed above? The fact that orientations which differ by less than 15° are not generally discriminable, regardless of the size of each sample, suggest that histograms—rather than exact values—of the samples are compared. I propose that the spacing between the histogram buckets equals the minimum perceivable difference. Biologically, this means that outputs from a relatively small number of receptor pools are compared, each pool being sensitive to a range of attribute values. The hypothesis is consistent with neurophysiological evidence of orientation-selective channels in the human visual system [Schatz 1977, p. 55].

To reduce the effects of discretization, the histogram buckets can overlap. It is also reasonable to assume that the receptor fields are differentially stimulated according to how well the texton's orientation matches that of the field. For simplicity, I approximated response curves with the function

$$R(\theta) = \begin{cases} 1 & \text{if } |\theta| < 7.5° \\ 0.5 & \text{if } 7.5° \leq |\theta| < 22.5° \\ 0 & \text{otherwise,} \end{cases}$$

by smoothing the histogram with the mask $(0.5, 1.0, 0.5)$. The exact shape of the mask is not critical, although, as we shall see, satisfying all capabilities does place constraints on it.

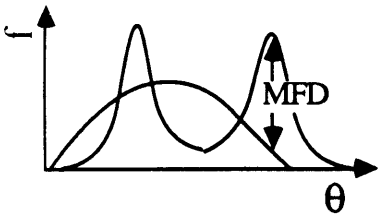


$R(\theta)$

The question of texture boundary detection now becomes, What statistic is used to compare two histograms of texton attributes? I propose using the statistic,

$$\max_i |f_{1i} - f_{2i}|,$$

where f_{1i} and f_{2i} are the frequencies of samples in corresponding buckets i of the two histograms ($f_i = n_i / \sum_i n_i$, where n_i equals the number of samples in bucket i). The maximum frequency difference (MFD) ranges between 0 (for two identical histograms) and 1 (when one bucket contains all the samples of one histogram, while the corresponding bucket in the other histogram is empty). If the MFD statistic is sufficiently high—above 0.5—the two textures are perceived as different.



The MFD statistic

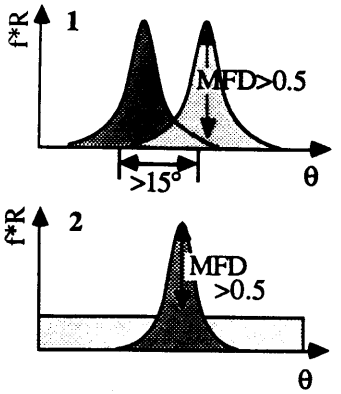
5.2.1 Accounts for preattentive capabilities

This computation is consistent with all of the perceptual capabilities listed above, as illustrated at right and on the next page:

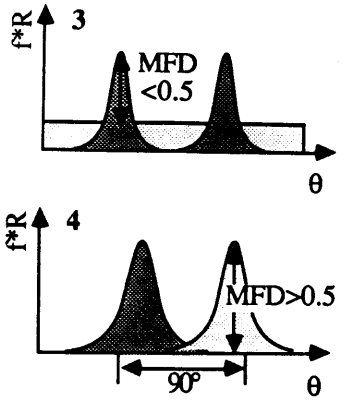
(1) The discrimination of two sharp, unimodal distributions. If the locations of two sharp modes differ by much more than 15°, the MFD is around 1, and the distributions are distinguishable. At imperceivable differences, the MFD is near 0. At the minimum perceivable difference, the MFD is near its threshold value, 0.5.

(2) The discrimination of a normal distribution from a uniform distribution. The less sharp the distribution, the lower its maximum frequency, and the less distinguishable it is from a uniform distribution. At a standard deviation of 15°, the MFD equals $\hat{f}_N - \hat{f}_U$, where

$$\hat{f}_N = \int R(\theta) G(\theta; \sigma = 15°) d\theta \approx 0.62, \text{and}$$



MFD and capabilities of preattentive orientation discrimination



MFD and capabilities of preattentive orientation discrimination (continued)

$$\hat{f}_U = \int \frac{15°}{180°} R(\theta) \approx 0.17.$$

The result, 0.45, is consistent with the threshold, considering the precision of the experiment.

(3) The nondiscrimination of a two-valued distribution from a uniform distribution. The MFD statistic equals $0.5 - \hat{f}_U \approx 0.33$, which is below the threshold.

(4) The discrimination of two non-sharp, normal distributions. At a 90° difference in means, two normal distributions are discriminable if their standard deviations are less than about 20°. In the borderline case, the MFD equals $\int R(\theta)G(\theta; \sigma = 20°)d\theta \approx 0.51$, consistent with the threshold.

The fact that the same threshold is implied by a number of different experiments investigating the limits of human texture perception indicates that computation of the MFD statistic is plausible and probably robust.

5.2.2 Constraints on response curves of receptive pools

The relations above show that my choices of response curve and MFD threshold are consistent with the perceptual capabilities identified. The constraints imposed by the measured capabilities on these choices can be formulated more generally. Letting

ρ = the range of the attribute (180° for orientation),
$\Delta\theta$ = the minimum perceivable difference (15° for orientation),
δ = the perceptibility threshold
$R(\theta)$ = the response curve, as before

and assuming $R(0) = 1$, Capabilities 1–4, imply the first four of the following constraints:

$$
\begin{aligned}
&(1)\quad \delta \approx \text{ the average value of } |R(\theta) - R(\theta - \Delta\theta)| \\
&(2)\quad \delta \approx \int G(\theta; \sigma = \Delta\theta) - \frac{1}{\rho}\int R(\theta)d\theta \\
&(3)\quad \delta > \frac{1}{2} - \frac{1}{\rho}\int R(\theta)d\theta \\
&(4)\quad \delta \approx \int G(\theta; \sigma = \tfrac{4}{3}\Delta\theta)R(\theta)d\theta \\
&(5)\quad \delta \approx \frac{1}{4\,\Delta\theta}\int_{-2\,\Delta\theta}^{2\,\Delta\theta} R(\theta)d\theta
\end{aligned}
$$

The fifth constraint results from a capability, which was discovered when comparing the MFD with other non-parametric statistics, discussed next.

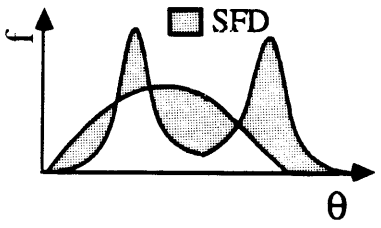
5.2.3 Comparison with other statistics

While the proposed MFD statistic seems plausible, one might wonder whether other common choices would suffice. Here, the MFD statistic is compared with some other statistics.

Sum of frequency differences

Schatz [1977] proposed a computation of preattentive texture discrimination for synthetic images based on the orientations of real and virtual line segments. He compared histograms having overlapping buckets, using the sum of frequency differences,

$$\sum_i |f_{1i} - f_{2i}|,$$



The SFD statistic

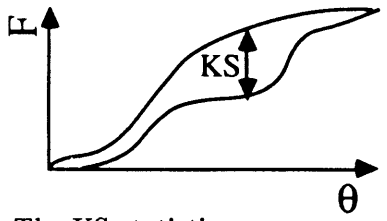
Empirically, he determined a threshold for the statistic. However, the sum of frequency differences (SFD) statistic is not consistent with the psychophysical results obtained here. In particular, it does not account for Capability 4—the fact that sharper modes at different locations are more discriminable than wider ones.

Kologmorov-Smirnov statistic

A well-known, non-parametric test for comparing two attribute distributions is the Kologmorov-Smirnov (KS) test, which employs the statistic

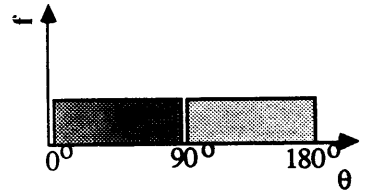
$$\max_x |F_1(x) - F_2(x)|,$$

where F_1 and F_2 are the cumulative sample distribution functions (i.e., $F(x) = \int_{-\infty}^{x} f(\xi)d\xi$). The statistic ranges between 0 and 1. If a histogram of the sample distributions is used, the KS statistic is identical to the MFD statistic, except that the cumulative histograms are compared.



The KS statistic

Computationally, the MFD statistic is simpler than the KS statistic, since it requires fewer summations of outputs of receptor pools. Furthermore, there is no natural starting point for a cumulative orientation distribution. At additional computational expense, the maximum KS statistic, KS_{max}, can be computed over all possible starting points. If both statistics accorded with perception, the KS statistic would be preferable.



Since the cumulative distribution can start at the location of the maximum frequency difference, $KS_{max} \geq$ MFD always. For comparisons involving sharp peaks (Capabilities 1 and 3), $KS_{max} =$ MFD.

To compare their relative performance, let us consider a case where the two statistics are dramatically different: two uniform distributions which do

not overlap, shown at left. Here, $KS_{max} = 1$, its maximum possible value, implying that the distributions are discriminable (regardless of the threshold value). By contrast, the MFD statistic is only $\frac{90°}{15°} \int R \approx 0.33$, predicting that they are not. In fact, psychophysical experiments determined that the distributions are not preattentively discriminable. As shown in Figure 5.1, it is even hard to discriminate them attentively!

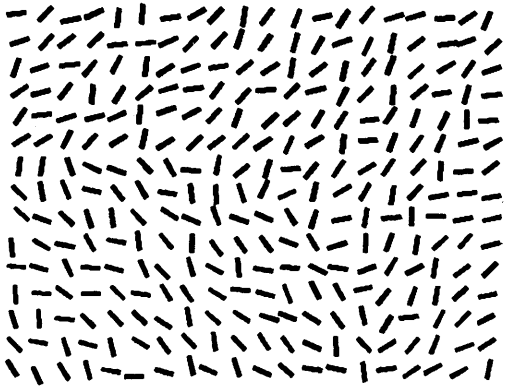
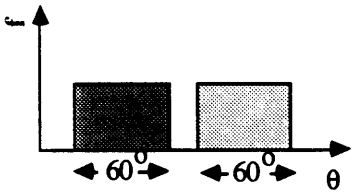


Figure 5.1 Two non-overlapping uniform distributions of range 90° are not discriminable.

Taking this investigation a step further, I determined the minimum range of each non-overlapping uniform distribution which did not yield a preattentive boundary. The result:

- *Capability 5:* Two non-overlapping uniform distributions are discriminable if the range of each is about 60° or less.

The 60° case is shown in Figure 5.2. Here, the MFD is $\frac{60°}{15°} \int R = 0.5$, providing further evidence supporting the threshold for this statistic. This example also demonstrates that the sum of frequency differences, which equals its maximum value, does not provide an adequate explanation of texture perception.



Capability 5

The MFD statistic also predicts that two versus two sharp modes are barely discriminable, if sufficiently spaced, but that three versus three modes of equal frequency are not. It would be interesting to test this capability.

Silvey [1975] feels that two-sided non-parametric tests are not practical for many applications, because the alternative hypothesis, $f_1 \neq f_2$, is too broad. This seems to be the case with texture, where only certain differences in the attribute distributions are perceived as boundaries. The MFD statistic proposed here, which is less sensitive than the KS, SFD, and other non-parametric statistics, and is insensitive to outliers, more accurately identifies significant texture changes.

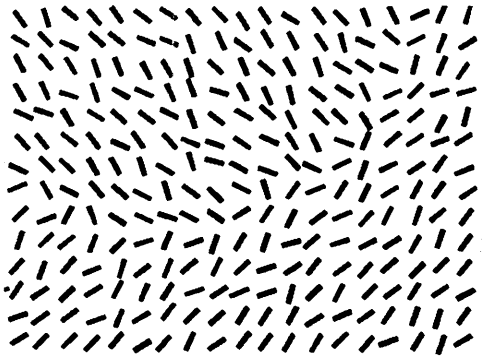


Figure 5.2 Two non-overlapping uniform distributions of range 60° (or less) are preattentively discriminable.

Minimum number difference

Biologically, the computation of the maximum frequency difference suggests that the outputs of selectively tuned receptors over a region are summed and compared with those from adjacent regions, and that the maximum response is attenuated (divided) by the sum of all the responses. Wondering whether the division could be eliminated, I considered using the maximum number difference (MND), computed by comparing the number of samples in each histogram:

$$\max_i |n_{1i} - n_{2i}|.$$

A hypothesis was that this statistic would obviate the need for minimum density checking, since it would be low in regions of low density, and would implicitly detect number density boundaries. However, results obtained using fixed neighborhood sizes were considerably noisier than those obtained using the MFD statistic. The result suggests that the division is indeed necessary if fixed neighborhood sizes are used.

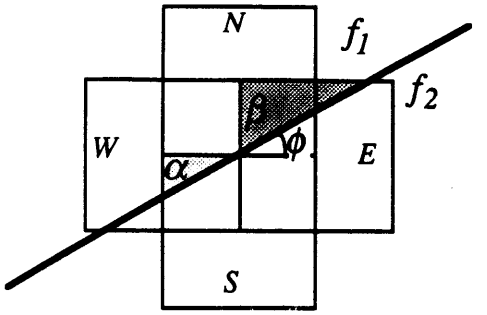
5.3 Detecting texture boundaries

To detect texture boundaries, the MFD is computed by comparing adjacent neighborhoods of textons at a grid of points which span the entire image array. At each point, the maximum difference over configurations of neighborhood pairs at all orientations is desired. This value can be computed from the MFD statistic measured in two directions (x and y), by assuming that the texture edge is locally linear. The next section describes the exact computation and an approximation to it.

5.3.1 Computing the maximum MFD at a point

Assume that two attribute distributions, f_1 and f_2, are separated by a texture edge which is locally linear, passing through the test point at angle ϕ, as shown at right. Assuming (without loss of generality) that $0° \leq \phi < 45°$, and denoting the four neighborhoods by their compass directions, N, S, E, and W, the distribution over each neighborhood is:

$$\begin{aligned} f_N &= (1-\alpha)f_1 + \alpha f_2 \\ f_S &= (1-\alpha)f_2 + \alpha f_1 \\ f_W &= (1-\beta)f_1 + \beta f_2 \\ f_E &= (1-\beta)f_2 + \beta f_1 \end{aligned}$$



Local model of texture edge separating two distributions

where

$$\alpha = \tfrac{1}{8}\tan\phi \quad \text{and} \quad \beta = \begin{cases} \frac{1}{2}(1-\tan\phi) & \text{if } 0 \leq \tan\phi < \frac{1}{2} \\ \frac{1}{8}\cot\phi & \text{if } \frac{1}{2} \leq \tan\phi < 1. \end{cases}$$

Then

$$\begin{aligned} d_x &\equiv \mathrm{MFD}(W,E) = (1-2\beta)\max|f_1 - f_2| \quad \text{and} \\ d_y &\equiv \mathrm{MFD}(N,S) = (1-2\alpha)\max|f_1 - f_2|, \end{aligned}$$

so the desired statistic is

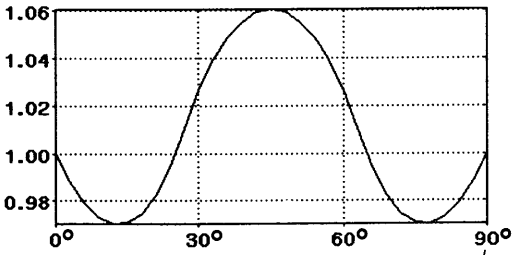
$$\max|f_1 - f_2| = \begin{cases} d_x/4 + d_y & \text{if } 0 \leq \tan\phi < \frac{1}{2} \\ \frac{8(d_x+d_y)+4\sqrt{4(d_x+d_y)^2-15d_xd_y}}{15} & \text{if } \frac{1}{2} \leq \tan\phi < 1. \end{cases}$$

The interval containing $\tan\phi$ can be determined from either of the expressions

$$\tan\phi = \begin{cases} \frac{1}{d_y/d_x+1/4} & (\text{if } 0 \leq \tan\phi < \frac{1}{2}) \quad \text{or} \\ 2(1-\frac{d_y}{d_x}) + \sqrt{4(\frac{d_y}{d_x})^2 - 7\frac{d_y}{d_x} + 4} & (\text{if } \frac{1}{2} \leq \tan\phi < 1), \end{cases}$$

since both functions are monotonic in d_x/d_y, crossing only at $\tan\phi = \frac{1}{2}$.

It turns out that the simpler function $\sqrt{d_x^2 + d_y^2}$ approximates $\max|f_1 - f_2|$ to within about 6% error, as shown at left. (It is only an approximation because d_x and d_y are not derivatives of an underlying surface.) Thus, if the edge passes through the point tested, the root-mean-square of the MFD's in the x and y directions estimates the MFD across the edge. Provided that the sampling grid is much finer than the neighborhood size, the spatial maxima of the statistic will be estimated with sufficient accuracy. At points away from the maxima, the approximation might not hold, but the values will clearly decrease with distance from the edge, and will not affect the ridge points.



Plot of $\frac{\max|f_1 - f_2|}{\sqrt{d_x^2 + d_y^2}}$ *vs.* ϕ

5.3.2 Finding ridges

At each scale, the MFD statistic is computed on a grid of points whose spacing is 1/4 the neighborhood width (i.e., neighborhoods overlapped by a factor of four, and attributes of textons from adjacent, non-overlapping neighborhoods are compared at each point[2]). The resulting MFD array is smoothed[3] and ridges are identified using non-maximum suppression. Texture boundaries are marked at ridges which lie above the MFD threshold and are sufficiently long—at least 1 to 2 times the neighborhood width. Thresholding with hysteresis and/or gap filling could further improve the results. The algorithm is therefore analogous to Canny's intensity-based edge detector [Canny 1983], with the MFD statistic replacing $|\nabla I * G|$.

[2]**Implementation Note:** In my implementation, a texton was included in the sample if the texton's center fell within the neighborhood. This approach is valid provided that blobs and bars are segmented so that they are not longer than the neighborhood size.

[3]**Implementation Note:** I used a Gaussian of standard deviation $\sigma \approx 1$.

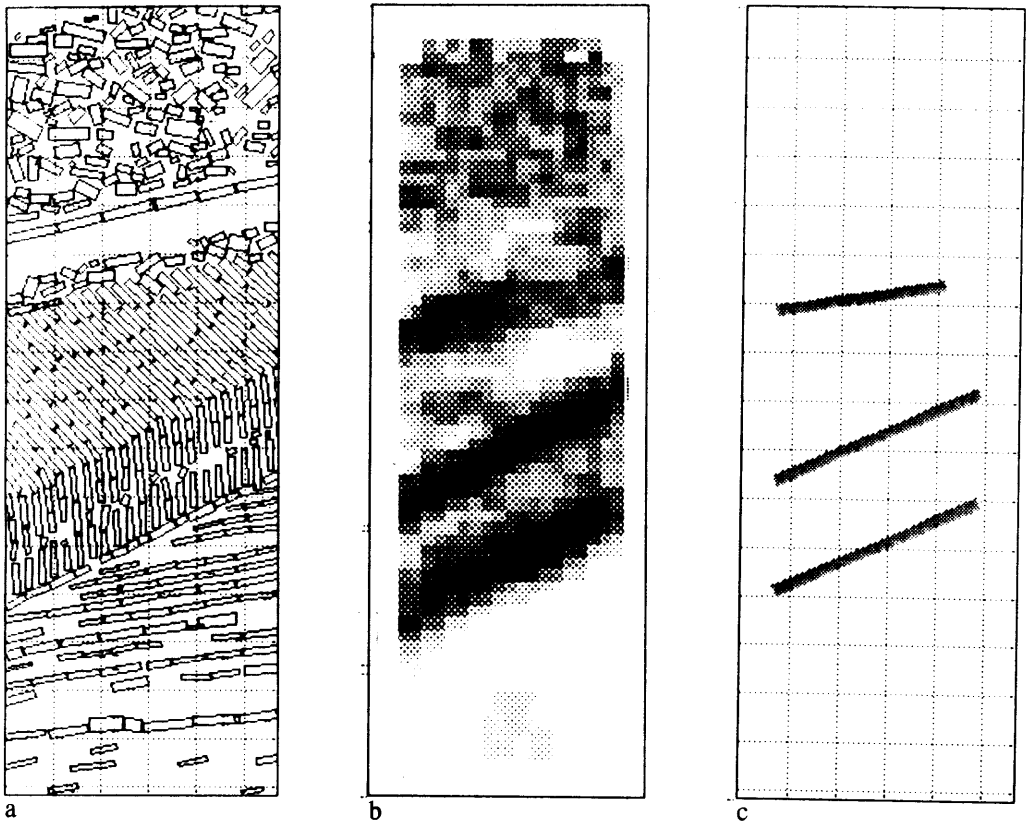


Figure 5.3 (a) Only elongated blobs are used to detect orientation boundaries. Blobs detected from an image of books (Figure 3.6) having elongation ≥ 1.5 are shown with some of the 28^2-pixel neighborhoods used to compute (b) the orientation MFD array. (c) Significantly high ridges in b locate the boundaries.

Figure 5.3 shows the orientation boundaries detected in the image of books and the textons and neighborhood size used to detect them. The choice of neighborhood size is discussed next.

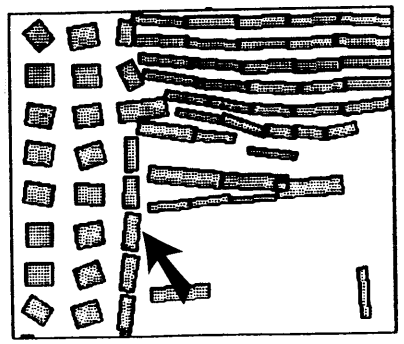
5.3.3 Neighborhood size

In real images texture boundaries occur at different scales. This is due to two factors: First, the spacing of textons varies. Second, the width of a transition

between two textured areas varies.

Julesz [1986] measured the upper limit on texton spacing. He found that compact textons were perceived as a texture provided that they were separated by no more than twice their diameter. At the other extreme, textons extracted from real images can be separated by less than one times their diameter. A neighborhood which covers several closely spaced textons may not cover enough sparsely spaced textons to adequately characterize the underlying distribution.

Multiple neighborhood sizes are also needed because transitions can exist between two homogeneous textured regions. This observation is overlooked in most psychophysical experiments (including those in Chapter 4), where texture boundaries are abrupt and can be detected using small neighborhoods. In real images, however, textons adjacent to a physical edge may not be representative of the texture of their surface. This is especially probable at occluding edges, where texton "events" themselves may be occluded or sheared. A thin shadow cast by the surface nearer the light source can also create non-representative textons at the boundary, as shown at left. Guaranteed identification of the boundary suggests that the largest neighborhood should be at least 4 textons wide, at which point textons in a two texton wide transition region constitute a minority.[4] On the other hand, the perception of narrow (two texton wide) textured regions suggest that small neighborhood sizes are also needed. At all scales, it would probably be beneficial to separate the neighborhoods compared by a small amount.



Textons near a boundary may be (a) corrupted by occlusion (see Figure 3.5a) or (b) created by a thin shadow (see Figure 3.7a.

[4]One might argue that two nearby, parallel small-scale boundaries on either side of the corrupted textons could be identified as a single large-scale boundary by some later stage of processing. But if the transition between the two homogeneous regions is smooth, there is no guarantee that either small-scale boundary will be perceivable.

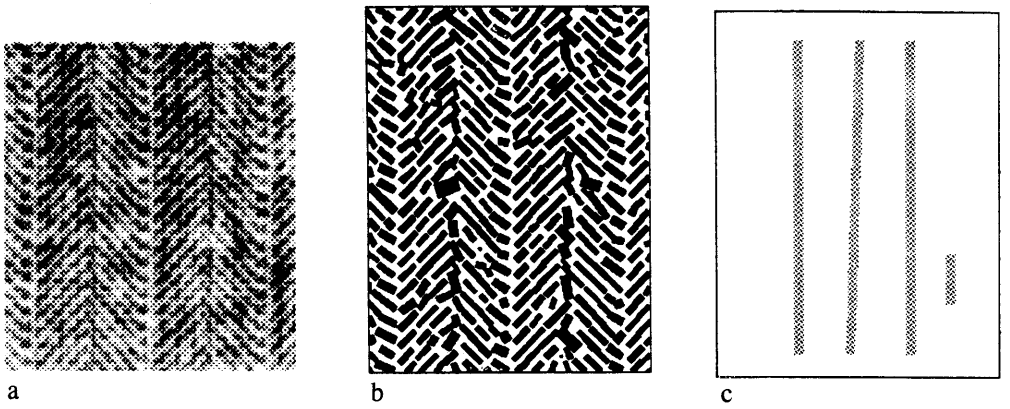


Figure 5.4 (a) Marr's herringbone example is revisited here (photo from Brodatz [1966]). (b) Blobs computed and (c) orientation boundaries.

It would be interesting to perform some psychophysical experiments to determine the range of neighborhood sizes used during preattentive vision. Images like Muller's [1986] example of "field-based cues" (Figure 1.6*b*) could be used for this purpose. Specifically, the largest size could be estimated by measuring the maximum transition width at which a single boundary is perceived between two discriminable regions.

In my implementation, I use neighborhoods of sizes 5, $5\sqrt{2}$, and 10 times the average texton width, squared. If the average spacing between textons is about one to two times their width (see Section 4.5.2), the smallest neighborhood covers about four textons on the average, while the largest neighborhood size covers an average of 4^2 textons. In images I examined, at least one of these scales, usually the middle one, yielded satisfactory results. Marr's example is revisited in Figure 5.4, showing the orientation boundaries detected at this scale.

Figure 5.5 shows orientation boundaries detected at the smallest and largest scales for the image of a sock. The width of the spline equals the grid spacing, and is therefore proportional to the scale. Figure 5.6 shows the same for an-

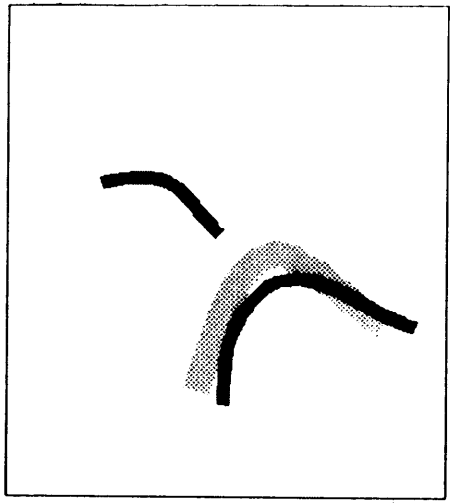


Figure 5.5 Orientation boundaries detected in image of sock using neighborhood sizes 44^2 and 88^2 pixels. The image and blobs are shown in Figure 3.7.

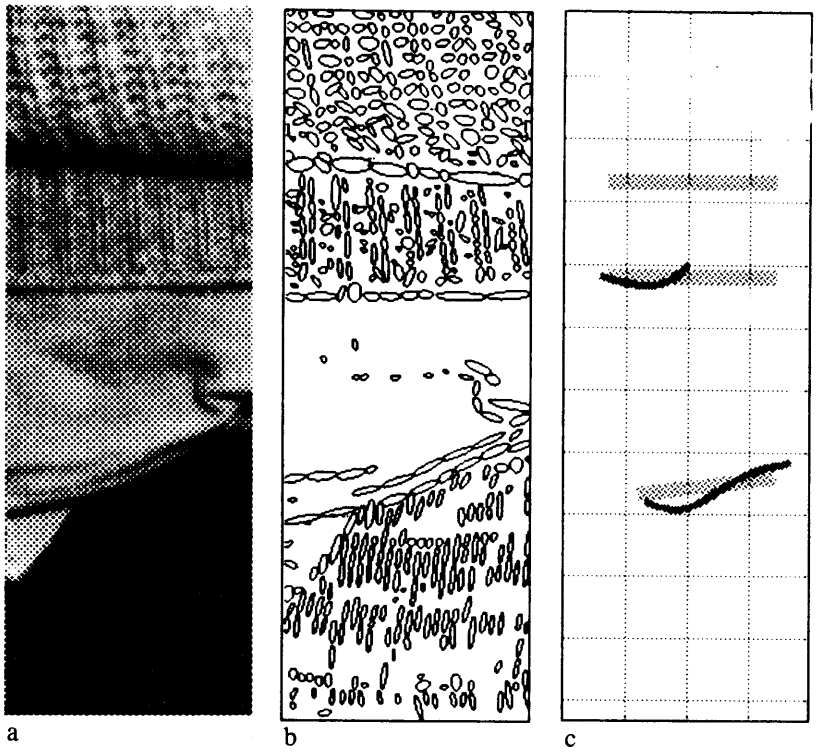


Figure 5.6 (a) A 160 × 454-pixel close-up of a sweater, shirttail, and corduroy pants. (b) Dark blobs detected at $\sigma = 1.5$ pixels. (c) Orientation boundaries detected using neighborhood sizes 20^2 and 40^2 pixels (larger size shown by grid).

other image of clothing. Exactly how texture boundaries from different scales are integrated, and how the "natural scale" for texture boundary detection is chosen, are open questions.

To implement the minimum density requirement, an MFD computed from a frequency distribution having fewer than 4 textons is reduced by 1/4 for each texton lacking from the smaller neighborhood. It would probably be better, especially for larger neighborhoods, to include in each sample distribution only those textons having nearby neighbors, computed using the discrete Voronoi diagram discussed in Section 4.5.2.

5.3.4 Applicability of MFD statistic to other attributes

The statistic above and its threshold are based on orientation discriminability experiments. It is not clear that the same results would be obtained for other attributes, which, unlike orientation, span a linear range.

Ranges used for geometric attributes are based on the scale of blob detection. The upper bound can be generous, since extra buckets do not affect the MFD. A notable exception is length. Long unsegmented bars, whose lengths are arbitrary, should simply be treated as "very long" by lumping all their lengths into the uppermost length bucket.

Not all orientation capabilities are necessarily applicable to the other attributes. In particular, it would be quite surprising if two non-overlapping uniform distributions of width, length, area, density, or contrast were not discriminable. Experiments like this are hard to perform for attributes other than orientation because, as noted in Chapter 4, such attributes are interrelated. In any case, it might not matter. Uniform distributions of these attributes (over the histogram range) are much less likely to occur in natural images than are uniform distributions of orientation. It would be useful to measure the maximum standard deviation of two discriminable, normal distributions at a given difference in means (as was done to determine orientation Capability 4). For these attributes, the only experiments reported involve the minimum difference in means.

In my implementation, I use the same statistic, threshold,. and smoothing envelope derived for orientation, with the minimum perceivable difference applicable to the attribute. I suspect that psychophysical experiments will demonstrate that this approach is valid for these attributes as well, but that the appropriate threshold and smoothing envelope may be different. Figure 5.7

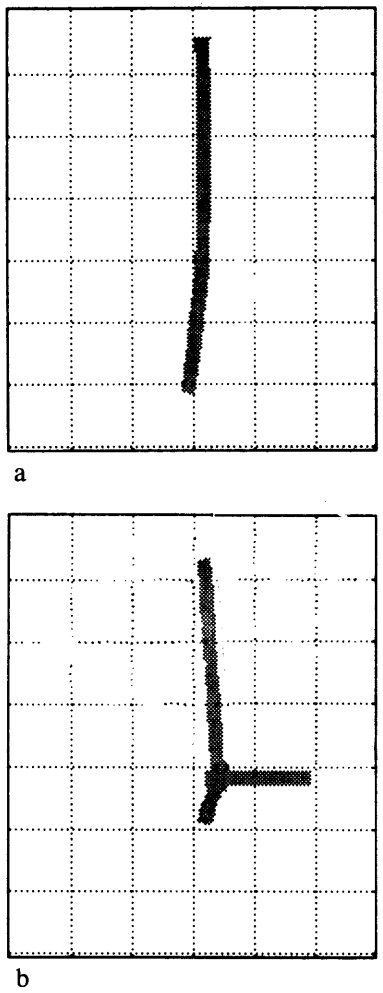


Figure 5.7 (a) Length boundaries and (b) contrast boundaries detected in the image of mismatched clothing (Figure 4.19a).

shows width and contrast boundaries detected in the image of mismatched clothing, where blobs are due to surface markings.

5.3.5 Combining attribute boundaries

I have shown how boundaries might be detected for any particular attribute. How to combine boundaries of different attributes, like the problem of combining boundaries from different scales, is an open question.

Instead of integrating boundaries from each of the attributes, it is possible that the MFD's from each attribute are combined into a single texture difference statistic, from which a single set of texture boundaries are detected.

There is evidence that, at least attentively, the strength of a texture boundary increases if more attributes change there [Beck et al. 1983]. From this observation one might infer that the composite statistic should be defined as the sum or norm of MFD's of individual attributes. However, the minimum difference experiments from Chapter 4 imply that combinations of attributes (aside from the few attributes so defined) do not yield texture boundaries. Furthermore, if the component differences were summed, a significant maximum in one attribute could be canceled by a minimum in another, resulting in a missed boundary. Using the maximum of the MFD's as the composite statistic is more consistent with these results. It can alternatively be computed directly as the maximum difference of buckets from all histograms between two neighborhoods.

If individual attribute boundaries are not detected explicitly, this would imply that subjects are not aware of the types of texture boundaries from preattentive viewing. It would be interesting to perform some psychophysical experiments, in order to possibly reject this hypothesis.

a b c

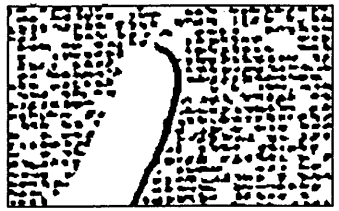
Figure 5.8 (a) Image of leopard (from Time magazine, Feb. 23, 1987, p. 40), (b) Blobs detected, and (c) area density boundaries detected.

5.3.6 Area density boundaries

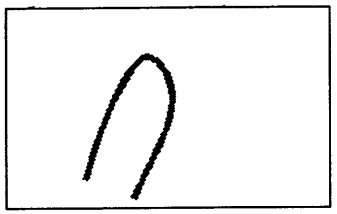
It should be noted that for finding occluding boundaries, area density of textons is probably the most useful attribute. Two different surfaces are usually at different depths, so it is unlikely that both yield textons at a single scale of analysis. Texture boundaries based on blob attributes are more likely to occur at orientation discontinuities or in images where the depth of field is small.

Texton area density, like the other attributes, can be defined on a very local basis as the ratio of a blob's area to that of its surround. Biologically, the "surround" may be computed by the proposed concentric detectors which measure spacing, and geometrically, as a texton's Voronoi polygon.

Alternatively, the average texton area be computed over neighborhoods directly from the blob pixel map, and an edge detector applied to the result. This method of texton area boundary detection does not require the symbolic representation of the primal sketch or the algorithm for attribute comparison which explain the perception of other attribute boundaries. Figures 5.8 and 5.9 show boundaries detected using this method.



a



b

Figure 5.9 (a) Blobs detected from image of finger (Figure 3.1a) and (b) area density boundaries detected.

It is possible that area density boundaries are computed instead of contrast boundaries. Since the response of $I * \nabla^2 G$ is proportional to the contrast of the blob, some contrast boundaries will manifest themselves as density boundaries if the proper level is chosen.

Implementing this computation, however, is not trivial, since the proper threshold levels must be chosen. To avoid locating boundaries which are artifacts of the particular level chosen, perhaps a boundary stability criterion could be used; this would require a boundary to be detected at a range of thresholds. The symbolic approach to finding contrast boundaries, presented here, avoids this issue by using a single level for detecting blobs, which is a simple function of noise in the image.

5.4 Summary

Summarizing, I have proposed and implemented the following algorithm for detecting texture boundaries:

1. Blobs are extracted from the intensity image using $I * \nabla^2 G$ filtering, thresholding, and segmentation operations. Attributes of the blobs—their orientation, width, length, area, nearest-neighbor spacing, and area density—are computed.

2. For each attribute, histograms of blob attribute values over local neighborhoods are constructed. The spacing of the histogram buckets, which overlap, approximately equals the minimum perceptible difference for the attribute. The neighborhoods overlap, and are computed over a variety of scales. A histogram is not computed where the textons are too sparse.

3. At each scale, histograms from adjacent neighborhoods are compared. The maximum difference in frequencies over all buckets is computed. The MFD statistic, which ranges from 0 to 1, measures the texture difference between the two neighborhoods.

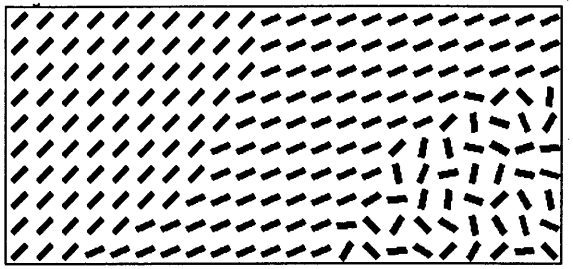
4. Texture boundaries occur at local maxima (ridges) of the difference array which lie above a threshold. The threshold accords with human preattentive capabilities. Overlapping neighborhoods serve to localize the texture boundaries, with uncertainty proportional to the width of the neighborhood.

The plausibility of the computation was demonstrated using blobs extracted from several natural images. Of course, many details remain to be filled in. How the natural scales for blob detection and texture boundary detection are chosen, or how the results from different scales are integrated, are among the open questions.

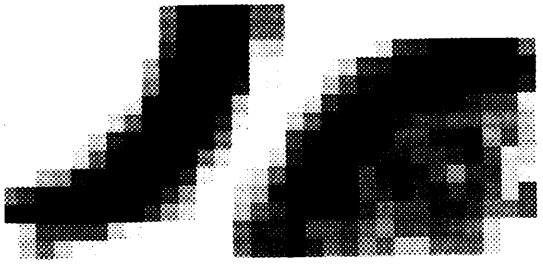
5.5 Comparison with other theories

The proposed computation of texture boundary detection, fills in some of the details of the theories offered by Marr, Beck, and Julesz—details which are critical for implementation, and which make the theories relevant to natural images. Although each of their theories is based on the premise that first order differences in texton attributes are compared, each offers some unique, if sketchy, ideas about how this is done. The next chapter shows how some observations of Julesz can be reinterpreted in terms of the theory proposed here. Below, my proposal is related to particular ideas of Marr and Beck.

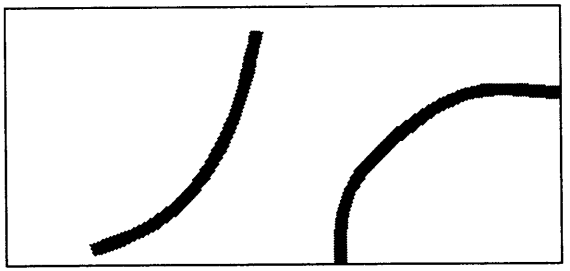
5.5.1 Similarity grouping



a



b



c

Figure 5.10 The MFD statistic accounts for apparent similarity grouping in a, even when textons on only one side of a boundary are similar. (b) The orientation MFD array and (c) computed boundaries.

Marr [1976*a*,*b*] and others have proposed that nearby texture elements are grouped together, on the basis of similarity, into textured regions. This operation, together with difference detection, he claimed, identified texture boundaries. Just how these two complementary operations interacted, however, was not explained.

The non-parametric MFD statistic combines the notions of similarity grouping and difference detection. If many textons in a neighborhood are similar and would therefore group together, they form a peak in the sample histogram. This results in a high MFD if an adjacent neighborhood does not have a peak at the same location—either because the textons cluster around another value or because they do not group together at all. The sensitivity of the MFD to both phenomena is illustrated explicitly in Figure 5.10.

The use of texture boundaries, rather than textured regions, as a basis for preattentive texture discrimination is consistent with the Muller's results, reviewed in Section 1.1.2, which show that texture boundary cues are more salient than textured region cues [Muller 1986]. Moreover, I conjecture that what Muller calls "boundary-based cues", computed at different scales, account for preattentive texture vision, without the need to identify large regions of similar textons.

Beck hypothesized that "Dissimilarity [is] the basic factor underlying similarity grouping" [Beck 1982, p. 301]. Now, the MFD statistic provides persuasive evidence.

5.5.2 Recursive grouping

According to the primal sketch theory, image tokens are not only created by simple intensity features like blobs but also by sets of tokens grouped together on the basis of proximity or colinearity, possibly combined with a similarity constraint. These "place tokens," defined recursively, are the primitives on which texture measures are applied [Marr 1976*a*].

Although smoothing can sometimes account for proximity grouping, Riley [1981] has shown that symbolic grouping operations are also needed. He constructed synthetic images of texture elements formed by triplets of colinear checkerboards, as shown in Figure 5.11. The elements were embedded in a field of grey, so that the line segments formed by the triples were not detectable by simple $I * \nabla^2 G$ filtering. Riley found that differences in the orientation of these virtual lines were instantly perceivable, implying that local, explicit grouping indeed occurs. The details of this computation remain to be implemented.

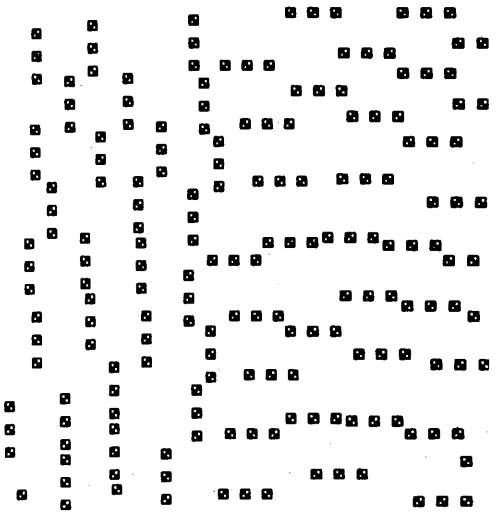


Figure 5.11 The perception of an orientation boundary here suggests that textons indeed can be formed by local grouping operations [Riley 1981].

5.5.3 Dense texture boundary suppression

Even though changes in a texton attribute can result in a texture boundary, boundaries do not occur if texture changes everywhere over the image. In on region of Figure 5.10*a*, almost all textons differ significantly from their neighbors, but a dense grid of boundaries is not detected. To account for this phenomenon, Beck et al. [1983] propose that texture boundary detection is suppressed if the density of strong difference signals is high. Their explanation would therefore require a mechanism whereby difference signals are attenuated by the presence of strong, non-colinear signals, or by suppressing texture boundary detection in regions where the density of signals was high.

Beck's proposal apparently rests on the assumption that simple differences between texton attributes (or perhaps means of distributions of texton attributes) are computed. Our theory accounts for this phenomenon as part of the computation of the MFD statistic over neighborhoods of reasonable size. If texton variance is high in a region, then the frequency of any particular attribute bucket will be small, and hence differences in frequency between neighborhoods in the region will be small. As a result, no boundaries would be detected.

Chapter 6 Terminators and Crossings as Textons

In addition to blobs, terminators (line endpoints) and crossings (line intersections) have been proposed as textons. In the representation proposed here, terminations could be marked at endpoints of bars (which do not touch other bars at that point) and crossings could be marked where three or more bars meet.[1] The significance of these structural features as textons is controversial, though, because some purported demonstrations of their significance can be explained by geometric and luminance attributes of blobs.

6.1 Another explanation of Julesz's experiments

In their texton theory of vision, Julesz and Bergen [1983; Julesz 1984] present a number of experiments which they claim demonstrates that terminators and crossings are textons. They present synthetic images which yield preattentive texture boundaries where density of terminators and crossings change. Below, I show that changes in blob attributes obtained by $\nabla^2 G$ filtering may provide a more accurate account of texture perception in these examples.

[1] Rearick [1985] has implemented a similar computation, detecting terminators and crossings in natural images using endpoints and junctions of the skeleton obtained by contrast-enhancing, thresholding, and thinning the image array.

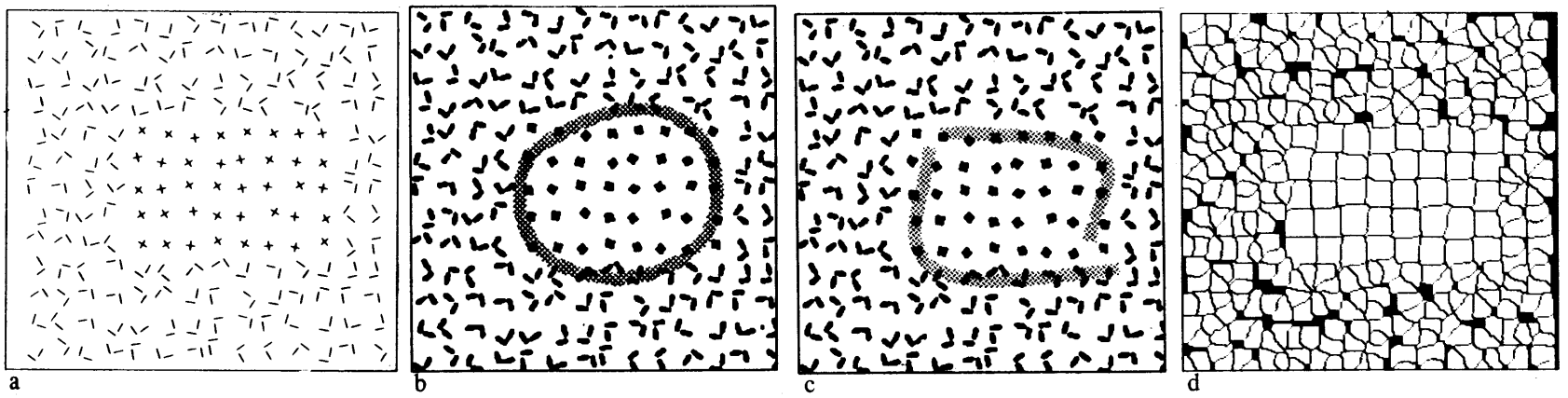
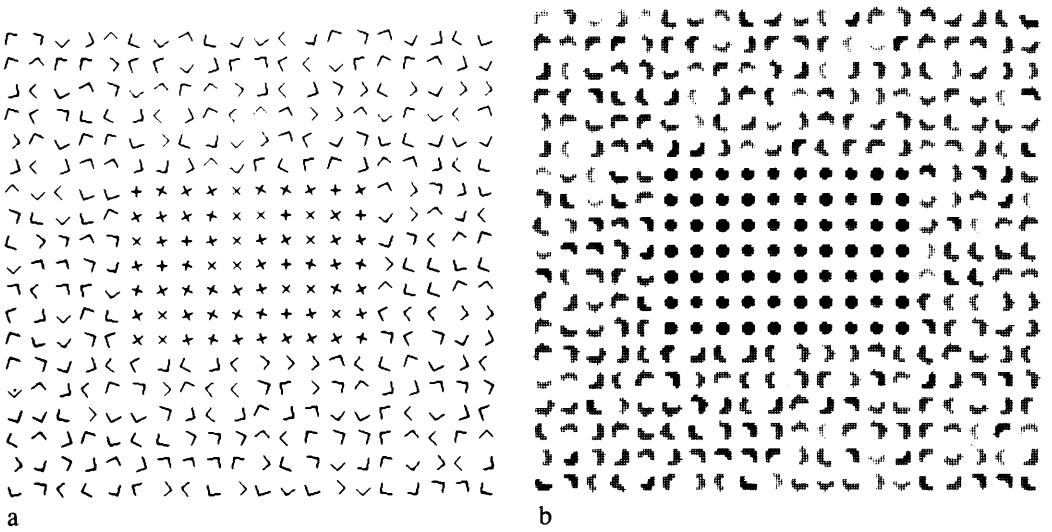


Figure 6.1 (a) A purported demonstration of crossings as textons [Julesz and Bergen 1983]. (b) However, the +'s yield wider blobs than the L segments after $\nabla^2 G$ filtering, shown here with detected width boundaries. (c) Nearest-neighbor (center-to-center) spacing boundaries also account for perception. (d) The partial Voronoi tesselation (whose computation is described in Section 4.5.2) defines the neighbor relations.

Figure 6.1*a*, modeled after Julesz [1984], contains a region of +'s on a background of disconnected L shapes. Julesz attributes the boundary to the difference in number of crossings (one versus two per texton). However, the boundary is just as easily explained by a difference in the width or spacing of blobs obtained by $\nabla^2 G$ filtering and segmentation, as shown in Figures 6.1b and c.

When the segments forming the L's are joined together, as in Figure 6.2, the density of terminators as well as crossings differ between the two regions. Perceptually, though, the +'s appear somewhat darker than the L's. The contrasts of the blobs obtained by filtering the image are indeed higher, as shown in Figure 6.2*b*. Computationally, the perception is explained by a contrast boundary (Figure 6.2*c*). Again, crossings are not required as an explanation.

It would interesting to repeat the psychophysical experiments involving crossings, lowering the contrast of the +'s so that no contrast boundary would be detected. If the regions were not preattentively discriminable, this would disprove the conjecture that crossings are textons.



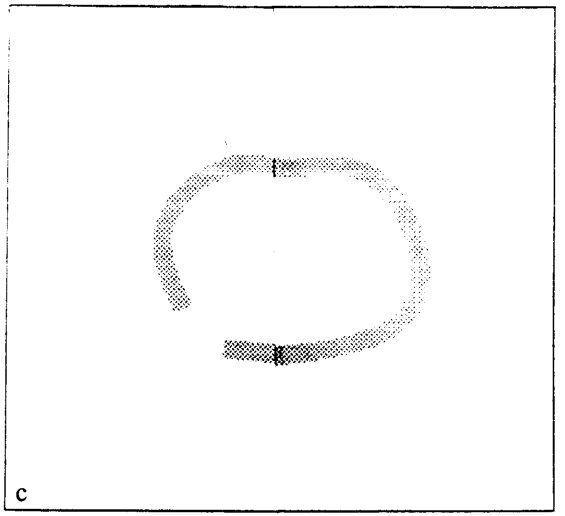
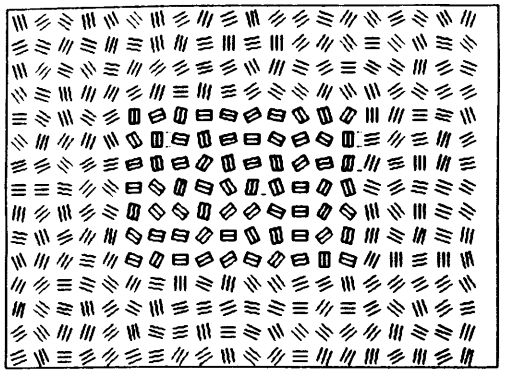


Figure 6.2 (a) Crossings do not necessarily account for the perception of this image (adapted from Julesz [1984]) either. (b) $\nabla^2 G$ filtering yields blobs of higher contrast for the inner region, as shown here by grey-level coding of blob contrast. (The variation within each region is due to discretization effects; squarely aligned +'s contain $\sqrt{2}$ times as many pixels as diagonally aligned ones, and hence are darker.) (c) Contrast boundaries account for discrimination.

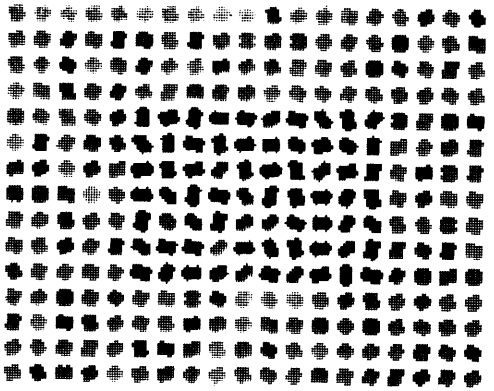
An example involving terminators is shown in Figure 6.4*a*, after Julesz [1984]. In this example, Julesz maintains that discrimination is due to the difference in number of terminators—6 for the triple bars, versus 0 for the figure 8's—as well as to a small difference in the density of line segments. But again, the inner figures appear darker, and the resulting blobs yield a slight contrast boundary. Here too a blob attribute may provide an alternative explanation of perception.

A similar image, Figure 6.3*a*, provides a counterexample to Julesz's conjecture. In this case, his theory predicts that the E's are distinguishable from the 6's based on a terminator density difference of 3 versus 1 per texton (a higher ratio than 5 to 2, which is significant, according to Julesz and Bergen [1983], Figure 5). However, a texture boundary is not preattentively perceivable. In this case blobs computed at the same scale as the previous example have similar contrasts and geometric attributes, as shown in Figure 6.3b. In this example, blob attributes better account for texture perception than do terminators.

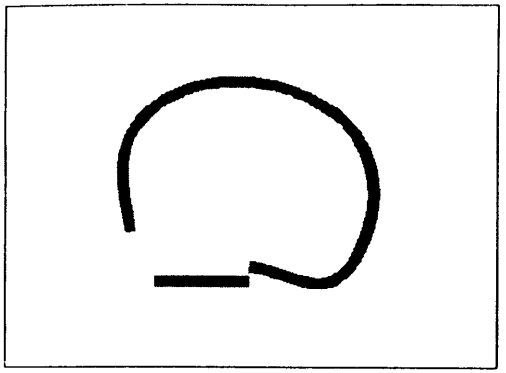
The examples above show that Julesz and Bergen's demonstrations of ter-



a



b



c

Figure 6.4 The discrimination of the 8's and triple bars (a) is also explained by a difference in blob contrast. The extracted blobs with grey-level proportional to contrast (b), and resulting contrast boundary (c).

minators and crossings as textons are not compelling. In fact, their proposition regarding terminators incorrectly predicts the perception of an example (Figure 6.3) very similar to ones used as evidence for it.

The erroneous conclusions of texton theory are due to its omission of filtering operations. Beck would agree: In discussing orientation discrimination, he notes, "What is important is not the orientation of lines per se but whether the change in the orientation causes feature detectors to be differentially stimulated" [Beck et. al 1983, p. 9].

It should also be noted that crossings and textons are not nearly as common in natural images as they are in synthetic ones. Very seldom in a natural scene would connected groups of blobs form character-like shapes, each having a distinctive number of terminators or crossings. It is not surprising then, that the simple blob attributes, whose relevance to natural images is demonstrated in Section 5.3, better account for perception in the artificial examples as well. Julesz and Bergen recognized that a precise definition of terminators for blobs having significant width was problematic, and concluded, "we are not certain

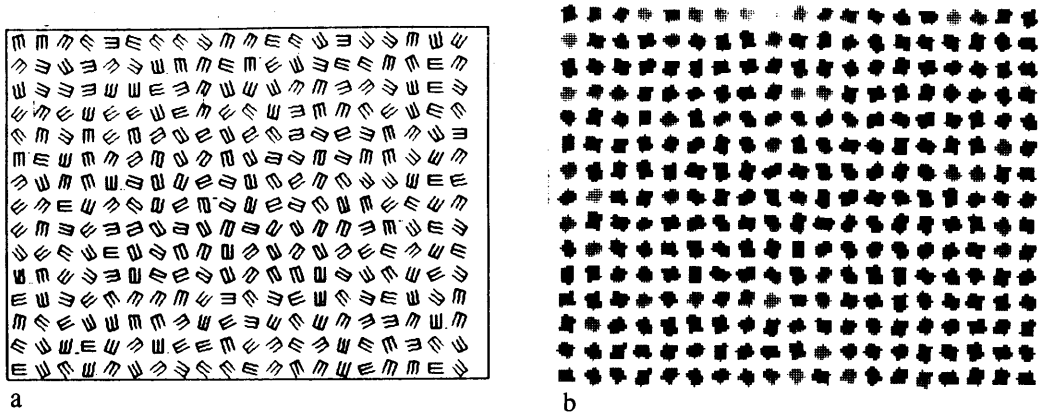


Figure 6.3 (a) No preattentive texture boundary is perceived between the E's and 6's, even they differ in number of terminators. (b) The extracted blobs, shown with blackness proportional to contast, yield no significant difference in contrast or geometric attributes.

whether terminators are independent textons" [Julesz and Bergen 1983, p. 1628].

6.2 Evidence of terminators as textons

If the examples of Julesz and Bergen are not convincing, then is there any evidence that terminators or crossings are textons? There is for terminators. Treisman [1985] showed that a circle with a small break in its curve is preattentively distinguishable from a field of closed circles. The difference appears to be due to the number of terminators, since the resulting blobs are similar. (She also argues that the difference is not due to closure.) Although her experiment concerned "pop out" phenomenon—the immediate recognition of textons which differ from their neighbors—it has implications for preattentive boundary detection as well. For texture vision, it suggests that a field of textons having two terminators each is preattentively distinguishable from a field of textons having none.

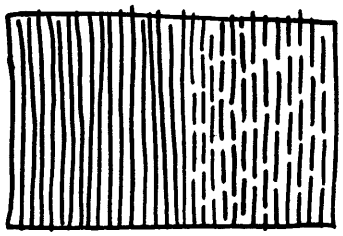


Figure 6.5 A texture boundary which is explainable by a difference in terminator density.

An example directly concerning texture boundaries, and perhaps more relevant to natural images, is shown in Figure 6.5. Here, the perceived texture boundary can probably only be accounted for by a difference in terminators. The widths, orientations, and density of bars on each side is the same. The lengths of dark bars cannot explain the difference because the specific lengths of very long bars are arbitrary. Here, the systematic presence of terminators on one side of the boundary versus none on the other seems to account for the difference.

Both of these examples involved the presence versus absence of terminators. I conjecture that only very sharp differences in the density of terminators result in texture boundaries. Figure 6.3 suggests that minimum significant ratio is greater than 3/1. Psychophysical experiments using images like Figure 6.5 might help answer this question. If the conjecture is true, then the discrimination of Figure 6.4 should not be affected by lowering the contrast of the figure 8's.

There appears to be some neurophysiological findings consistent with the computation of termination points [Beck 1982, pp. 310–311]. It would be interesting, however, to discover whether perceivable differences in terminator density occur often enough in natural images to justify their use in texture discrimination.

There is less reason to believe that crossings are textons, since no accurate psychophysical evidence has been presented to suggest it. Treisman's pop out tests, in fact, indicate that crossings are not preattentively perceivable [Treisman 1985]. It is also worth noting that Marr [1976*a,b*] mentions terminations, but not crossings, as assertions of the raw primal sketch.

In summary, the geometric and luminance attributes of blobs seem more

relevant to natural images than do crossings or terminators. Psychophysics indicate that differences in these attributes which cause texture boundaries (around 25%) are small compared to terminators, which require much larger ratios.

Chapter 7 **Conclusion**

The previous chapters proposed answers to two main questions of texture vision: How are textons extracted from images of natural scenes? and, how are texture boundaries then located? I conclude with some speculation on the third main question: How are texture boundaries interpreted according to their physical cause?

In order to interpret texture boundaries according to their physical cause, it is necessary to know which attributes change at a texture boundary. Large changes in density, for example, can only be due to occlusion, while changes in orientation may be due to a surface orientation discontinuity. Whether the preattentive system keeps track of the type of texture boundaries, or just creates a single, unlabeled texture boundary map, is not known.

It is already known that preattentive vision is unaware of "what ... textons are" [Julesz 1986, p. 245]. It would be interesting to perform some psychophysical experiments in which the subject is asked to identify the attribute type of texture boundary, to suggest whether any preattentive interpretation is possible.

If the preattentive visual system actually computes a non-parametric statistic like the one proposed, then not all interpretation can be performed preattentively. This is because the MFD statistic does not tell how two distributions

differ.

Consider an orientation boundary. If each of the orientation distributions on either side of the edge consists of a single sharp peak located at different means, the boundary may be due to a change in surface orientation. But if the distributions differ in the number of modes (one versus two, say) the boundary must be caused by occlusion or a change in surface material. Interpreting the physical cause of an edge can therefore require information about the *how* the attribute distributions differ, information which is not available from the non-parametric statistic proposed. If preattentive vision indeed computes a statistic similar to the MFD, this suggests that certain texture boundaries require attentive vision for interpretation.

These observations suggest that attentive vision is needed to interpret the physical cause of some, if not all, texture boundaries. One approach to achieve this, described in the next section, is to symbolically describe the attribute distributions on either side of a texture boundary. Rules are then applied to interpret possible physical causes.

7.1 Interpreting texture boundaries

Preattentively detected texture boundaries may be examined attentively to interpret the physical cause of the boundary. One possible computational model of this process is sketched here.

At each boundary of interest, the textons on either side of it are identified by constructing a window alongside the boundary. The windows are drawn as strips parallel to the boundary on either side of it. Choosing the appropriate window sizes is important. Each window must be sufficiently large

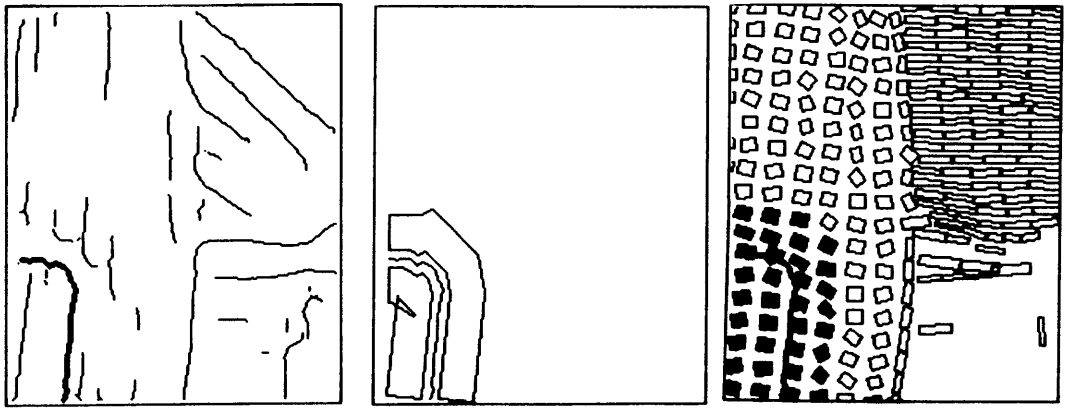


Figure 7.1 (a) *Large scale intensity edges of image of mismatched clothing (Figure 5.7a) with selected edge highlighted,* (b) *regions drawn alongside selected edge, and* (c) *included blobs.*

to accurately characterize the local texton distribution, but small enough to avoid covering more than one texton distribution. The implemented algorithm initially chooses a very narrow strip as the window, followed by successively wider ones, until a sufficient number of textons lie within the window (either partially or entirely). In these examples, at least 15 textons were required in each window, provided that the window did not exceed a maximum width. The windows are positioned a small distance away from the boundary, due to uncertainty in the boundary location, and to avoid textons associated with the boundary itself, such as thin shadows caused by occlusion. Histograms are constructed of the attributes of the textons lying in each window, defining the sample distribution functions to be analyzed. Figure 7.1 shows an example.

The distributions of texton attributes are described symbolically using an algorithm written by Carlotto [1984]. The procedure identifies Gaussian-shaped modes, which may overlap, in a histogram. To accomplish this, the

histogram is smoothed and differentiated twice. Pairs of zero crossings of the second derivative determine the initial estimates of the locations and widths of each mode. These estimates are refined using a maximum likelihood technique. Each mode is described by its position (mean), width (standard deviation), and the fraction of data it covers (relative frequency). Modes which cover less than a specified minimum percentage of the data are discarded.

The symbolic descriptions of each pair of attribute distributions—one from each side of the edge—are compared to suggest the physical cause of the boundary. A set of heuristics can be constructed by elaborating Riley's observations regarding the kinds of attribute changes caused by different types of physical discontinuities. Riley [1981] noted that at a surface orientation discontinuity, only geometric attributes such as orientation, density, and size of texture features change, while at a structural discontinuity (i.e., a occluding or material boundary) any attributes, including intensity, shape, and contrast can change.

In particular, the following conditions provide cues regarding the type of boundary present:

1. Large changes in texton area density or size: Occlusion.

2. Clear differences in the number of modes of any attribute: Occlusion.

3. Change in orientations of elongated textons or number density, small changes in width and length consistent with laws of projection, similar area density: Surface orientation discontinuity.

4. Similar in all attributes, including intensity: Surface marking (applicable if intensity edges are tested).

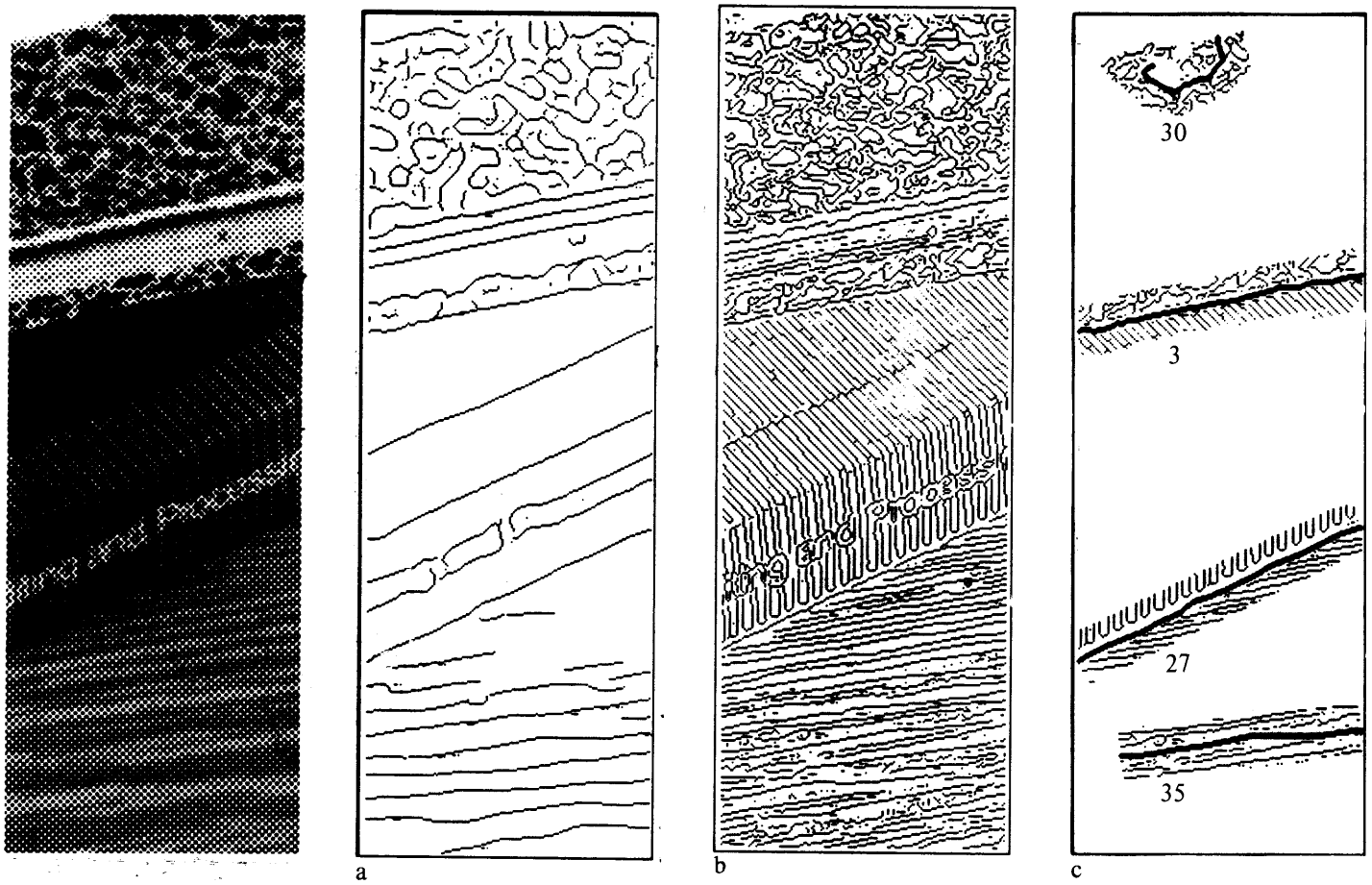
5. Difference in intensity, but similar in all other attributes, including contrast: Shadow.

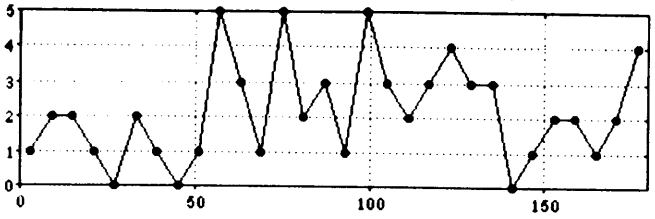
Figure 7.2 demonstrates an partial implementation of the approach, which uses the orientations of edge pixels as textons. In this example, only the number and positions of orientation modes are compared to suggest the physical cause of selected intensity edges.

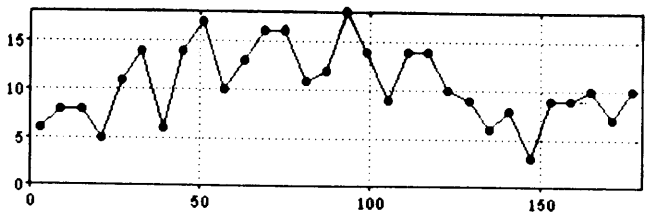
The conditions of the rules need to be formulated more precisely, and multiple scales of analysis performed, to avoid errors in ambiguous cases. For example, the widths of modes must be considered when comparing their number in condition number 2. As noted in Section 5.1, a distribution of two overlapping modes, for example, could be represented symbolically as either a single or pair of modes at a single scale of representation.

Besides preattentive texture boundaries, the proposed computation could be applied to other types of boundaries as well. Intensity, motion, and depth boundaries could all be tested as texture boundaries. For intensity boundaries, such an approach may be motivated by the observation that in the real world, texture boundaries often coincide with intensity changes. In this context, texture can be regarded as a cue for labeling an intensity edge directly according to its physical origin—a discontinuity in surface depth, orientation, reflectance or illumination. Such a hypothesize-and-test strategy resembles Witkin's correlation approach to intensity edge labeling [Witkin 1982], described in Section 1.1.3.

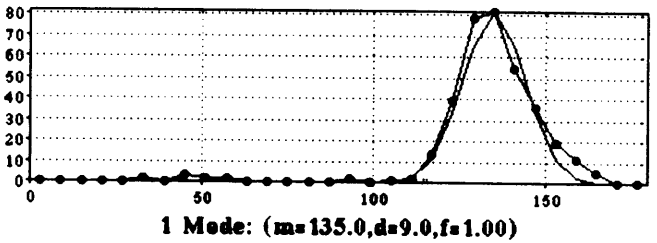
Figure 7.2 Interpreting intensity edges. (a) Large scale edges of image books to be interpreted. (b) Here, fine scale ($\sigma = 1.5$) intensity edge pixels are employed as textons. (c) Selected long, large scale ($\sigma = 4.0$) intensity edges (shown in bold) are tested as intensity boundaries. Facing page: Histograms of the orientations of texton edge pixels on either side of each edge (shown in **c**) *are constructed, analyzed, and compared to interpret the physical cause of the edge.*

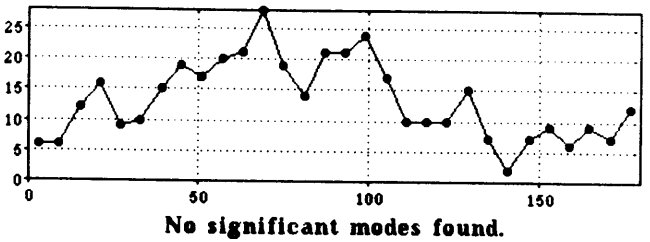




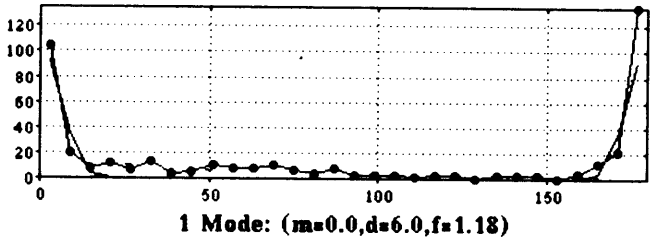


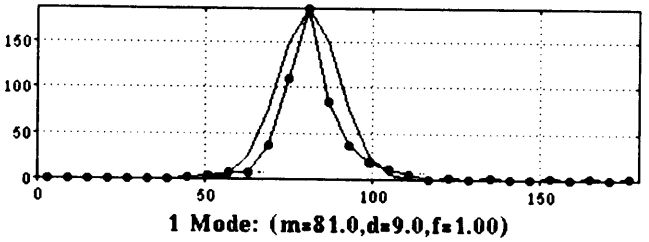
Edge 30: No modes for either histogram found. Interpretation: Surface marking or shadow.



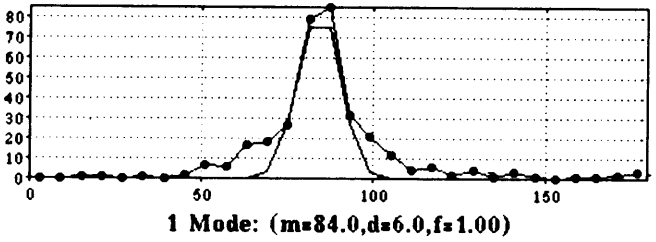


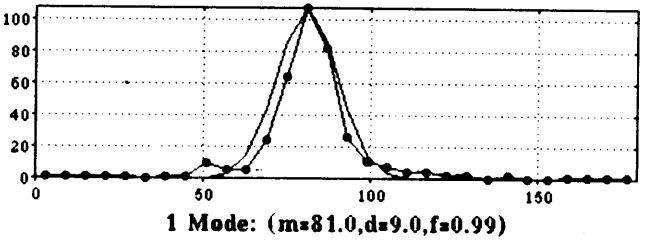
Edge 3: Different numbers of modes: 1 0 Interpretation: Occlusion.



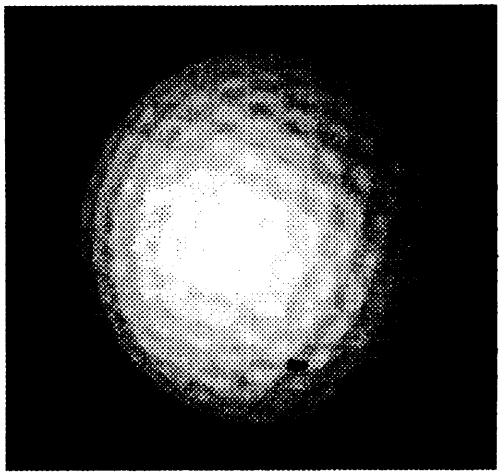


Edge 27: Modes at different angles: 0.0, 81.0. Interpretation: Surface Orientation or occlusion.

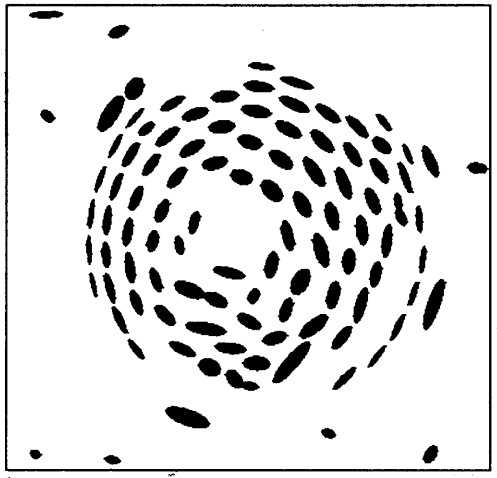




Edge 35: Modes at similar angles: 84.0, 81.0. Interpretation: Surface marking or shadow.



a



b

Figure 7.3 Image of golf ball (a) shows that even blobs due to surface impressions (b) can provide a cue to the object's shape.

7.2 Shape from texture and texture boundaries

The components of condition number 3, above, which express surface orientation discontinuities in terms of attributes of projected textons must be rather complex. It may be more natural to detect such boundaries using a shape-from-texture algorithm. Stevens [1980], for example, proposes how shape can be determined from attributes of structural features. As illustrated in Figure 7.3, blobs may be useful for this purpose.

Shape-from-texture algorithms have not been very practical, to date, in part because they assume that texture on three-dimensional surfaces is homogeneous, and need to know where this assumption might be violated. Two-dimensional texture boundaries, preattentively detected, can provide such information. A shape-from-texture algorithm can therefore find surface orientation discontinuities (or possibly provide some more general description of shape), while occluding and shadow boundaries can be detected by simple rules, maybe preattentively. In this way the texture boundary and shape-from-texture computations can complement each other.

7.3 Neighborhood sizes of preattentive and attentive vision

The trade-off between neighborhood size and localization is a well-known problem in texture boundary detection. Humans are sensitive to changes in the number of modes and to large differences in variance of attribute distributions, statistics which require large neighborhoods to compute. Yet we are somehow able to locate texture boundaries quite precisely. The "uncertainty principle" of intensity edge detection [Canny 1983] thus becomes more acute in the context of texture, where the data (the textons) are sparse.

This problem may be resolved by separating the task into two stages, as the human visual system apparently does. Preattentively, a non-parametric statistic like the one proposed can be computed in order to locate texture boundaries approximately. Since it employs small neighborhoods, it can be computed relatively efficiently, but as a result, it cannot specify how two distributions differ. Hence it does not always provide enough information to infer the physical cause of a texture boundary.

To obtain precise information about the kind of change in the attribute distribution, an attentive mode is used. Since the boundary is located fairly accurately, relatively large areas can be examined for the purpose of characterizing the distributions. This allows the detection of more subtle changes than is possible preattentively and permits physical interpretation of the boundary. Once the distributions on either side of the boundary are characterized attentively, the position of the texture boundary can be refined by examining textons in the immediate vicinity of the transition.

For some time researchers have supposed that dichotomy of preattentive and attentive vision is a result of making optimal use of limited computational resources to achieve both fast, parallel processing over the entire image, and detailed scrutiny over areas of interest. The detection and interpretation of texture boundaries provides a good example of the need for these two modes of vision.

7.4 Summary of results

For over a decade various scientists have proposed that texture vision operates by detecting first-order differences in the attributes of texture tokens.

This research, for the first time, has explained some of the details of such a computation—details which are crucial to demonstrate the applicability of the theory to images of natural scenes.

Below, I summarize the main results of this research and describe their applicability to other vision modules as well:

- *Chapter 2:* A technique for estimating noise in natural images was described. The method, which does not depend on a smooth surface assumption, is useful for $I * \nabla^2 G$-based blob detection, and improves the performance of Canny's $\nabla I * G$-based edge detector.

- *Chapter 3:* A method for computing and segmenting blobs from natural images was presented. This research demonstrated the usefulness of such blobs as textons for finding texture boundaries, proving the relevance of feature-based theories of texture to images of natural scenes. As features of the raw primal sketch, the blobs should also be useful for shape-from-texture and image registration for motion and stereo correspondence problems.

- *Chapter 4:* From psychophysical experiments the minimum perceivable differences in geometric blob attributes were measured. These experiments also suggested which attributes are compared by the human visual system.

- *Chapter 5:* A minimum difference statistic was proposed which accounts for the large variety of preattentively perceivable texture boundaries. The statistic, which is biologically plausible, successfully locates texture boundaries in natural scenes. A similar computation might serve to locate color and motion discontinuities (see Spoerri and Ullman [1987]).

- *Chapter 6:* The role of terminators and crossings as textons was challenged, by showing that purported demonstrations of their salience are better accounted for by attributes of blobs. It was conjectured that only changes in terminator density higher than previously suggested could cause texture boundaries.

- *Chapter 7:* Some possible roles of preattentive and attentive vision in interpreting texture boundaries and in determining shape from texture were discussed. An attentive check for texture boundaries can be performed on intensity, color, depth, or motion boundaries as well.

As in any challenging area of research, a new theory often raises as many questions as it answers. In this case, open questions include:

- How is the "natural scale" of textons determined, or, how can output of multiple feature detectors at a location be integrated?

- How are nearby blobs grouped into higher-level place tokens?

- How, exactly, should density be measured, and the minimum density requirement computed?

- How are the various attribute boundaries integrated?

- Is the physical interpretation of texture boundaries ever performed preattentively?

- What, exactly, are the rules for interpreting the physical cause of texture boundaries?

- How is texture information integrated with other vision modules such as color, motion, and intensity?

Future research in artificial intelligence, neurophysiology, and psychophysics should further decipher the computational mystery of human vision.

References

Beattie, R. J. "A Semantically-based Multi-Level Edge Detection Scheme," *Proc., 7th Int'l Conf. on Pattern Recognition*, pp. 653–655, 1984.

Beck, Jacob. "Textural Segmentation," in *Organization and Representation in Perception* (Jacob Beck, ed.), Chapter 15, Erlbaum, Hillsdale, NJ, 1982.

Beck, Jacob, K. Prazdny, and Azriel Rosenfeld. "A Theory of Textural Segmentation," in *Human and Machine Vision.* (Jacob Beck, Barbara Hope, Azriel Rosenfeld, eds.) Academic Press, New York, pp. 1-38, 1983.

Beck, Jacob, Anne Sutter, and Richard Ivry. "Spatial Frequency Channels and Perceptual Grouping in Texture Segmentation," *Computer Vision, Graphics, and Image Processing*, Vol. 37, No. 2, pp. 299–325, Feb. 1987.

Berzins, Valdis. "Accuracy of Laplacian Edge Detectors," *Computer Vision, Graphics, and Image Processing*, Vol. 27, No. 2, pp. 195–210, Aug. 1984.

Bracho, Rafael, and Arthur C. Sanderson. "Segmentation of Images Based on Intensity Gradient Information," *Proc., Computer Vision and Pattern Recognition Conf.*, pp. 341–347, 1985.

Brodatz, Phil. *Textures: A Photographic Album for Artists and Designers.* Dover Publications Inc., New York, 1966.

Caelli, Terry, Bela Julesz, and Edgar Gilbert. "On Perceptual Analyzers Underlying Visual Texture Discrimination: Part II," *Biological Cybernetics*, Vol. 29, pp. 201–214, 1978.

Canny, John Francis. "Finding Edges and Lines in Images," M.I.T. Artificial Intelligence Lab, Technical Report No. 720, June 1983.

Carlotto, Mark J. "Texture Classification based on Hypothesis Testing Approach," *Proc., 7th Int'l Conf. on Pattern Recognition*, pp. 93–96, 1984.

Charniak, Eugene, and Drew McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley Publishing Co., Reading, MA, pp. 126–129, 1985.

Gagalowicz, André. "A New Method of Texture Fields Synthesis: Some Applications to the Study of Human Texture Vision," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-3, No. 5, pp. 520–533, Sept. 1981.

Haralick, Robert M. "Statistical and Structural Approaches to Texture," *Proc. of the IEEE*, Vol. 67, No. 5, pp. 786–804, May 1979.

Hildreth, Ellen C. "Computations Underlying the Measurement of Visual Motion," *Artificial Intelligence*, Vol. 23, No. 3, pp. 309–354, Aug. 1984.

Horn, Berthold Klaus Paul. "Determining Lightness from an Image," *Computer Graphics and Image Processing*, Vol. 3, No. 1, pp. 277–299, Dec. 1974.

———. "Understanding Image Intensities," *Artificial Intelligence*, Vol. 8., No. 2, pp. 201–231, April 1977.

———. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

Julesz, Bela. "Experiments in the Visual Perception of Texture," *Scientific American*, Vol. 232, No. 4, pp. 34–43, April 1975.

———. "Textons, the Elements of Texture Perception, and their Interactions," *Nature*, Vol. 290, pp. 91–97, 12 March 1981*a*.

———. "A Theory of Preattentive Texture Discrimination Based on First-Order Statistics of Textons," *Biological Cybernetics*, Vol. 41, pp. 131–138, 1981*b*.

———. "A Brief Outline of the Texton Theory of Human Vision," *Trends in Neuroscience*, Vol. 7, pp. 41–45, 1984.

———. "Texton Gradients: The Texton Theory Revisited," *Biological Cybernetics*, Vol. 54, pp. 245–251, 1986.

Julesz, B. and J. R. Bergen. "Textons, The Fundamental Elements in Preattentive Vision and Perception of Textures," *The Bell System Technical Journal*, Vol. 62, No. 6, pp. 1619–1645, July–Aug. 1983.

Kjell, Bradley P. and Charles R. Dyer. "Edge Separation and Orientation Texture Measures," *Proc., Computer Vision and Pattern Recognition Conf.*, pp. 306–311, 1985.

Mahoney, James V. "Image Chunking: Defining Spatial Building Blocks for Scene Analysis," M.S. Thesis, M.I.T. Dept. of E.E.C.S., Jan. 1987.

Marr, David. "Analyzing Natural Images: A Computational Theory of Texture Vision," *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. XL, pp. 647–662, 1976*a*.

———. "Early Processing of Visual Information," *Philosophical Trans. of the Royal Society of London, Biological Sciences*, Vol. B 275, No. 942, pp. 483–524, 19 Oct. 1976*b*.

———. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman and Co., San Francisco, 1982.

Marr, David and Ellen Hildreth. "Theory of Edge Detection," *Proc. of the Royal Society of London*, Vol. B 207, pp. 187–217, 1980.

Muller, Michael J. "Texture Boundaries: Important Cues for Human Texture Discrimination," *Proc., Computer Vision and Pattern Recognition Conf.*, pp. 464–468, 1986.

Parzen, Emanuel. *Modern Probability Theory and its Applications*, John Wiley & Sons, New York, 1960.

Pavlidis, Theo. *Algorithms for Graphics and Image Processing.* Computer Science Press, Rockville, MD, 1982.

Rearick, T. C. "A Texture Analysis Algorithm Inspired by a Theory of Preattentive Vision," *Proc., Computer Vision and Pattern Recognition Conf.*, pp. 312–317, 1985.

Richards, W., H. K. Nishihara, and B. Dawson. "CARTOON: A Biologically Motivated Edge Detection Algorithm," M.I.T. Artificial Intelligence Lab, A.I. Memo No. 668, June 1982.

Riley, Michael Dennis. "The Representation of Image Texture," M.I.T. Artificial Intelligence Lab, Technical Report No. 649, Sept. 1981.

Rubin, J. M. and W. A. Richards. "Color Vision and Image Intensities: When are Changes Material?" *Biological Cybernetics*, Vol. 45, pp. 215–226, 1982.

Schatz, Bruce R. "The Computation of Immediate Texture Discrimination," M.I.T. Artificial Intelligence Lab, A.I. Memo No. 426, Aug. 1977. (Summarized in Carnegie Mellon University, Dept. of Computer Science, Report No. CMU-CS-78-152, 1978).

Silvey, S. D. *Statistical Inference*, Monographs in Applied Probability and Statistics, Chapman and Hall, London, 1975.

de Souza, Peter. "Edge Detection Using Sliding Statistical Tests," *Computer Vision, Graphics, and Image Processing*, Vol. 23, No. 1, pp. 1–14, July 1983.

Spoerri, Anselm and Shimon Ullman. "The Early Detection of Motion Boundaries," M.I.T. Artificial Intelligence Lab, A.I. Memo No. 935, 1987.

Stevens, Kent. "Surface Shape from Local Analysis of Texture and Contour," M.I.T. Artificial Intelligence Lab, Technical Report No. 512, Feb. 1980.

Terzopolous, Demetri. "Computing Visible-Surface Representations," M.I.T. Artificial Intelligence Lab, A.I. Memo No. 800, March 1985.

Treisman, Anne. "Preattentive Processing in Vision," *Computer Vision, Graphics, and Image Processing*, Vol. 31, No. 2, pp. 156–177, Aug. 1985.

Van Gool, L., P. Dewaele, and A. Oosterlinck. "Texture Analysis Anno 1983," *Computer Vision, Graphics, and Image Processing*, Vol. 29, No. 3, pp. 336–357, March 1985.

Witkin, Andrew P. "Recovering Surface Shape and Orientation from Texture," *Artificial Intelligence*, Vol. 17, Nos. 1–3, pp. 17–45, Aug. 1981.

———. "Intensity-Based Edge Classification," *Proc., Nat'l Conf. on Artificial Intelligence (AAAI-82)*, pp. 36–41, 1982.

———. "Scale-Space Filtering," *Proc., Int'l Joint Conf. on Artificial Intelligence (IJCAI-83)*, pp. 1019–1022, 1983.