1 THE USE OF THE FACET MODEL AND THE TOPOGRAPHIC PRIMAL SKETCH IN IMAGE ANALYSIS

LINDA G. SHAPIRO
ROBERT M. HARALICK
University of Washington

TING-CHUEN PONG
University of Minnesota

INTRODUCTION

The *facet model* states that all processing of digital-image data has its final authoritative interpretation relative to what the processing does to the underlying gray-tone intensity surface. The digital image's pixel values are noisy sampled observations of the underlying surface. Thus, in order to do any processing, we must estimate this underlying surface at each pixel position. This requires a model that describes what the general form of the surface would be in the neighborhood of any pixel if there were no noise. To estimate the surface from the neighborhood around a pixel, then, amounts to estimating the free parameters of the general form. The processing that takes place is then defined in terms of the estimated parameters.

The topographic primal sketch (Haralick, Watson, & Laffey, 1983) is one possible way of representing the fundamental structure of a digital image in a rich and robust way. The basis of the topographic primal sketch is the classification and grouping of the underlying image-intensity surface patches according to the categories defined by monotonic, gray-tone invariant functions of directional derivatives. Examples of such categories are peak, pit, ridge, ravine, saddle, flat, and hillside. From this initial classification, categories can be grouped to obtain a rich, hierarchical, and structurally complete representation of the fundamental image structure. By contrast, representations of the fundamental image structure only involving edges or the primal sketch as described by Marr (1976) are impoverished in the sense that they are insufficient for unambiguous matching. They also do not have the required invariance with respect to monotonically increasing gray-tone transformations.

The facet approach can also be used in classical gradient-based edge detection, in image segmentation, as well as in determining the topographic primal sketch of an image. The following sections discuss the facet model for image-data specialized to the sloped facet case and its direct application to gradient-edge detection; the facet model concepts as they can apply to image segmentation; the definition of the topographic primal sketch and how the information it requires can all come from the facet-model estimates; and three-dimensional object surface-shape estimation based on the patterns of the topographic primal sketch.

THE FACET MODEL FOR IMAGE DATA

The commonly used general forms for the facet model include piecewise constant (flat facet model), piecewise linear (sloped facet model), piecewise quadratic, and piecewise cubic. In the flat model, each ideal fitting neighborhood in the image is constant in gray tone. In the sloped model, each ideal fitting neighborhood has a gray tone surface that is a sloped plane. Similarly, in the quadratic and cubic models, regions have gray tone surfaces that are quadratic and cubic surfaces, respectively.

Given a noisy defocused image, and assuming one of these models, the problem is to estimate the parameters of the underlying surface for a given neighborhood and estimate the variance of the noise. These estimates can then be used in a variety of ways: edge detection, line detection, corner detection, and segmentation. In this section we review the parameter estimation problem for the sloped facet model and illustrate its use in the classic gradient edge detector application.

Sloped Facet Parameter and Error Estimation

In this discussion we employ a least-squares procedure to estimate the parameters of the sloped facet model for a given rectangular neighborhood whose row index set is R and whose column index set is C. The facet parameter estimates are obtained for the central neighborhood of each pixel on the image. We assume that for each $(r, c) \in R \times C$, the image function g is modeled by

$$g(r, c) = \alpha r + \beta c + \gamma + \eta(r, c)$$

where η is a random variable indexed on $R \times C$, which represents noise. We will assume that η is noise having mean 0 and variance σ^2 and that the noise for any two pixels is independent.

The least-squares procedure determines an $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, which minimize the sum of the squared differences between the fitted surface and the observed one:

$$\epsilon^2 = \sum_{r \in R} \sum_{c \in C} |\hat{\alpha} r + \beta c + \hat{\gamma} - g(r, c)|^2.$$

Taking the partial derivatives of ϵ^2 and setting them to zero results in

$$\begin{pmatrix} \frac{\partial \epsilon^2}{\partial \hat{\alpha}} \\ \frac{\partial \epsilon^2}{\partial \hat{\beta}} \\ \frac{\partial \epsilon^2}{\partial \hat{\gamma}} \end{pmatrix} = 2 \sum_{r \in R} \sum_{c \in C} (\hat{\alpha} r + \hat{\beta} c + \hat{\gamma} - g(r, c)) \begin{pmatrix} r \\ c \\ 1 \end{pmatrix} = 0. \quad (1)$$

Without loss of generality, we choose our coordinate system $R \times C$ so that the center of the neighborhood $R \times C$ has coordinates (0, 0). When the number of rows and columns is odd, the center pixel, therefore, has coordinates (0, 0). When the number of rows and columns is even, there is no pixel in the center but the point where the corners of the four central pixels meet has coordinates (0, 0). In this case, pixel centers will have coordinates of an integer plus a half.

The symmetry in the chosen coordinate system leads to

$$\sum_{r \in R} r = 0 \text{ and } \sum_{c \in C} c = 0$$

Hence,

$$\sum_r \sum_c \hat{\alpha} r^2 = \sum_r \sum_c r g(r, c),$$

$$\sum_r \sum_c \hat{\beta} c^2 = \sum_r \sum_c c g(r, c),$$

$$\sum_r \sum_c \hat{\gamma} = \sum_r \sum_c g(r, c).$$

Solving for $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ we obtain

$$\hat{\alpha} = \sum_r \sum_c r g(r, c) / \sum_r \sum_c r^2,$$

$$\hat{\beta} = \sum_r \sum_c c g(r, c) / \sum_r \sum_c c^2, \quad (2)$$

$$\hat{\gamma} = \sum_r \sum_c g(r, c) / \sum_r \sum_c 1.$$

Replacing $g(r, c)$ by $\alpha r + \beta c + \gamma + \eta(r, c)$ and simplifying the equations will allow us to explicitly see the dependence of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ on the noise. We obtain

$$\hat{\alpha} = \alpha + \left(\sum_r \sum_c r\eta(r, c) \Big/ \sum_r \sum_c r^2 \right),$$

$$\hat{\beta} = \beta + \left(\sum_r \sum_c c\eta(r, c) \Big/ \sum_r \sum_c c^2 \right),$$

$$\hat{\gamma} = \gamma + \left(\sum_r \sum_c \eta(r, c) \Big/ \sum_r \sum_c 1 \right).$$

From this it is apparent that $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\eta}$ are unbiased estimators for α, β, and γ, respectively, and have variances

$$V[\hat{\alpha}] = \sigma^2 / \sum_r \sum_c r^2,$$

$$V[\hat{\beta}] = \sigma^2 / \sum_r \sum_c c^2,$$

$$V[\hat{\gamma}] = \sigma^2 / \sum_r \sum_c 1.$$

Normally, distributed noise implies that $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are normally distributed. The independence of the noise implies that $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are independent since they are normal and that

$$E[(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)] = E[(\hat{\alpha} - \alpha)(\hat{\gamma} - \gamma)] = E[\hat{\beta} - \beta)(\hat{\gamma} - \gamma)] = 0$$

as a straightforward calculation shows.

Examining the squared error residual ϵ^2 we find that

$$\epsilon^2 = \sum_r \sum_c [(\hat{\alpha}r + \hat{\beta}c + \hat{\gamma}) - (\alpha r + \beta c + \gamma + \eta(r, c))]^2$$

$$= \sum_r \sum_c [(\hat{\alpha} - \alpha)^2 r^2 + (\hat{\beta} - \beta)^2 c^2 + (\hat{\gamma} - \gamma)^2 + \eta^2(r, c)$$

$$- 2(\hat{\alpha} - \alpha)r\eta(r, c) - 2(\hat{\beta} - \beta)c\eta(r, c) - 2(\hat{\gamma} - \gamma)\eta(r, c)].$$

Using the fact that

$$(\hat{\alpha} - \alpha) = \sum_r \sum_c r\eta(r, c) \Big/ \sum_r \sum_c r^2,$$

$$(\hat{\beta} - \beta) = \sum_r \sum_c c\eta(r, c) \Big/ \sum_r \sum_c c^2,$$

$$(\hat{\gamma} - \gamma) = \sum_r \epsilon R \sum_c \eta(r, c) \Big/ \sum_r \sum_c 1$$

we may substitute into the last three terms for ϵ^2 and obtain after simplification

$$\epsilon^2 = \sum_r \sum_c \eta^2(r, c) - (\hat{\alpha} - \alpha)^2 \sum_r \sum_c r^2 - (\hat{\beta} - \beta)^2 \sum_r \sum_c c^2$$

$$- (\hat{\gamma} - \gamma)^2 \sum_r \sum_c 1$$

Now notice that

$$\sum_r \sum_c \eta^2(r, c)$$

is the sum of the squares of

$$\sum_r \sum_c 1$$

independently distributed normal random variables. Hence,

$$\sum_r \sum_c \eta^2(r, c)/\sigma^2$$

is distributed as a chi-squared variate with

$$\sum_r \sum_c 1$$

degrees of freedom. Because $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are independent normals,

$$((\hat{\alpha} - \alpha)^2 \sum_r \sum_c r^2 + (\hat{\beta} - \beta)^2 \sum_r \sum_c c^2 + (\hat{\gamma} - \gamma)^2 \sum_r \sum_c 1)/\sigma^2$$

is distributed as a chi-squared variate with 3 degrees of freedom. Therefore, ϵ^2/σ^2 is distributed as a chi-squared variate with

$$\sum_r \sum_c 1 - 3$$

degrees of freedom. This means that $\epsilon^2/(\sum_r \sum_c 1 - 3)$ can be used as an unbiased estimator for σ^2.

Gradient-Based Facet-Edge Detection

Suppose that our model of the ideal image is one where each object part is imaged as a region that is homogeneous in gray tone. In this case the boundary between object parts will manifest itself as jumps in gray level between successive pixels on the image. A small neighborhood on the image that can be divided into two parts by a line passing through the middle of the neighborhood and in which all the pixels on one side of the line have one gray level is a neighborhood in which the dividing line is indeed an edge line. When such a neighborhood is fitted with the sloped

facet model, $\hat{\alpha}r + \hat{\beta}c + \hat{\gamma}$, a gradient magnitude of $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ will result. The gradient magnitude will be proportional to the gray-level jump. On the other hand if the region is entirely contained within a homogeneous area, then the true surface $\alpha r + \beta c + \gamma$ will have $\alpha = \beta = 0$ and the fitted, sloped facet model $\hat{\alpha}r + \hat{\beta}c + \hat{\gamma}$ will produce a value of $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ which is near zero. Hence, it is reasonable for edge detectors to use the estimated gradient magnitude $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ as the basis for edge detection. Such edge detectors are called "gradient based edge detectors." There are other kinds of edge detectors such as zero-crossing edge detectors. A discussion of how the facet model can be used to determine zero crossings of second directional derivatives as edges can be found in Haralick (1984).

The most interesting question in the use of the estimated gradient $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ as an edge detector is how large does the gradient have to be in order for it to be considered significantly different from 0. The discussion that answers this question begins by noting that $\hat{\alpha}$ has a normal distribution with mean α and variance $\sigma^2 / \sum_r \sum_c r^2$, that $\hat{\beta}$ has a normal distribution with mean β and variance $\sigma^2 / \sum_r \sum_c c^2$, and that $\hat{\alpha}$ and $\hat{\beta}$ are independent. Hence,

$$\frac{(\hat{\alpha} - \alpha)^2 \sum_r \sum_c r^2 + (\hat{\beta} - \beta)^2 \sum_r \sum_c c^2}{\sigma^2}$$

is distributed as a chi-square variate with 2 degrees of freedom. From this it follows that to test the hypothesis of no edge under the assumption that $\alpha = \beta = 0$, we use the statistic G

$$G = \frac{\hat{\alpha}^2 \sum_r \sum_c r^2 + \hat{\beta}^2 \sum_r \sum_c c^2}{\sigma^2}$$

which is distributed as a chi-squared variate with 2 degrees of freedom. If the statistic G has a high enough value, then we reject the hypothesis that there is no edge.

If the neighborhood used to estimate the facet is square, then $\sum_r \sum_c r^2 = \sum_r \sum_c c^2$ so that the test statistic is a multiple of the estimated squared gradient magnitude $\hat{\alpha}^2 + \hat{\beta}^2$. Such an edge operator is the well known Prewitt edge operator. However, by knowing the conditional distribution given no edge, it becomes easier to choose a threshold. For example, suppose we want the edge detector to work with a controlled

false-alarm rate. The false-alarm rate is the conditional probability that the edge detector classifies a pixel as an edge given that the pixel is not an edge. Suppose the false-alarm rate is to be held to 1%. Then since $P(X_2^2 > 9.21) = .01$, the threshold we must use must be at least 9.21.

But to use this technique, we must know the noise variance σ^2. Fortunately we can obtain a good estimate of σ^2. Each neighborhood's normalized squared residual error,

$$\epsilon^2 / \left(\sum_r \sum_c 1 - 3 \right),$$

can constitute an estimator for σ^2. This estimator is available for each neighborhood of the image. Because there are usually large numbers (thousands) of pixels in the image, the average of $\epsilon^2 / \left(\sum_r \sum_c 1 - 3 \right)$ taken over all the neighborhoods of the image is a very good and stable estimator of σ^2 if it can be assumed that the noise variance is the same in each neighborhood. If ϵ_n^2 represents the squared residential fitting error from the n^{th} neighborhood, then we may use

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} \epsilon_n^2 / \left(\sum_r \sum_c 1 - 3 \right)$$

in place of σ^2. Hence our test statistic G becomes

$$G = \frac{\hat{\alpha}^2 \sum_r \sum_c r^2 + \hat{\beta}^2 \sum_r \sum_c c^2}{\hat{\sigma}^2}$$

Under the hypothesis of no edge, G, being the ratio of two chi-squared statistics, would have an F distribution. But because the number of degrees of freedom of $\hat{\sigma}^2$ is so high, G has essentially a chi-squared distribution with two degrees of freedom. Thus if we wanted to detect edges and be assured that the false-alarm rate (the conditional probability of assigning a pixel as an edge given that it is not an edge) is less than ρ_0, we would use a threshold of θ_0 where $P(X_2^2 \geq \theta_0) = \rho_0$.

Figure 1.1a shows a controlled 100 × 100 image having a disk of diameter 63. The interior of this disk has gray level of 200. The background of the disk has gray level of 0. Independent Gaussian noise having mean zero and standard deviation of 40, 50, and 75 is added to the controlled image. The noisy images are shown in Figures 1.1b, 1.1c, and 1.1d, respectively.

A sloped facet model is fitted to each 5 × 5 neighborhood of each image and its $\hat{\alpha}$, $\hat{\beta}$, and ϵ^2 is computed. For the ideal image of Figure

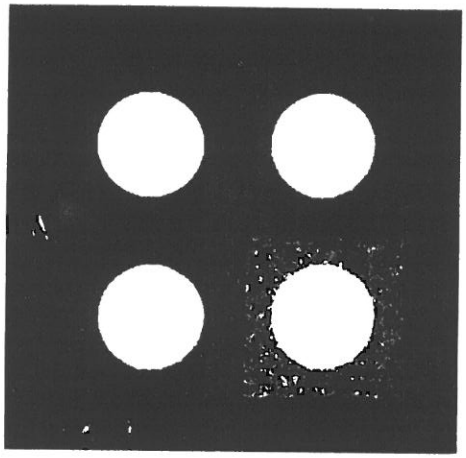


FIGURE 1.1. (a) (upper left) shows the controlled disk (background having value 0, disk having value 200) and noisy disks. (b) (the upper right) has noise standard deviation 40. (c) (the lower left) has noise standard deviation of 50. (d) (the lower right) has noise standard deviation of 75.

1.1a, the average squared residual fitting error ϵ^2, the average being taken over all neighborhoods, was 302.33. This corresponds to a standard deviation of about 17.4, which is 8.7% of the dynamic range of the image.

Obviously the fit will be perfect for all 5×5 neighborhoods that are entirely contained in the background or on the disk. The error must come from those neighborhoods that have some pixels from the background and some pixels from the disk. In these neighborhoods, the sloped fit is only an approximation. The neighborhood that has the worst fit is shown in Figure 1.2. The sloped fit there has an average

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	200
0	200	200	200	200

observed neighborhood

(a)

-64	-48	-32	-16	0
-28	-12	4	20	36
8	24	40	56	72
44	60	76	92	108
80	96	112	128	144

slope fitted neighborhood

(b)

64	48	32	16	0
28	12	-4	-20	-36
-8	-24	-40	-56	-72
-44	-60	-76	-92	92
-80	104	88	72	56

residual fitting errors

(c)

FIGURE 1.2. (a) illustrates a neighborhood for which the sloped fit is a relatively bad approximation. The fit produces an $\hat{\alpha} = 36$, $\hat{\beta} = 16$, and $\hat{\gamma} = 40$. The sloped fitted neighborhood is shown in (b), and the residual fitting errors are shown in (c). The total squared error from (c) is 82400. This divided by the degrees of freedom, $25 - 3 = 22$ yields an average squared error of 3746. The square root of 3746 is about 61.2, which represents the standard deviation of the residual fitting errors.

squared residual error of 3746. The standard deviation of fitting error is then 61.2, which represents 30.6% of the dynamic range.

For the noisy image of Figure 1.1b, the standard deviation of the fitting errors is $\sigma = 77.3$. This is just a little higher than the standard deviation of the noise because it takes into account the extent to which the data do not fit the model. In fact, assuming that the imperfectedness of the model and the noise are independent, we would expect to find a standard deviation of $\sqrt{17.4^2 + 75^2} = 77$, which is close to the 77.3 measured.

Figure 1.3a shows edges obtained when the statistic G computed on the ideal image of Figure 1.1a is thresholded at the value 120. Since $\hat{\sigma}^2 = 302.33$ and $\sum_r \sum_c r^2 = \sum_r \sum_c c^2 = 50$, this corresponds to selecting all neighborhoods having slopes greater than 26.94. Figure 1.3b, c, and d show the edges obtained when a 5×5 sloped facet model is employed and when the statistic G computed from each neighborhood of the noisy image of Figure 1.1d is thresholded at 4, 8, and 11. Since $\hat{\sigma}^2 = 5975.3$ for the noisy image, a threshold of 8 corresponds to selecting all neighborhoods having slopes greater than 30.92. These thresholds of 4, 8, and 11 guarantee (under the conditions of independent Gaussian noise) that the false alarm rates must be less than .1353, .01832, and .0041, respectively. The observed false-alarm rates for these thresholds are .1231, .0164, and .0042, respectively. The corresponding misdetection rates are .0205, .0820, and .1598.

As just mentioned, corresponding to each possible threshold is a false alarm rate and a misdetection rate. The misdetection rate is the conditional probability that a pixel is assigned "no edge" given that it is actually an "edge" pixel. One way to characterize the performance of an edge detector is to plot its false-alarm rate as a function of misdetection rate in a controlled experiment. Such a plot is called an "operating curve." Figure 1.4 shows two operating curves for the sloped facet edge detector.

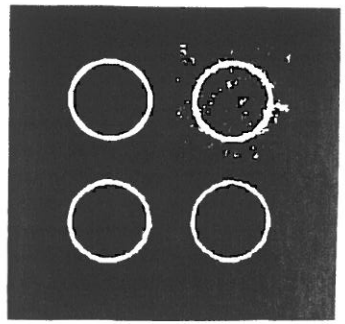


FIGURE 1.3. (a) (upper left) shows edges obtained when the statistic G computed using 5×5 neighborhoods on the ideal image of (a) is thresholded at the value 120. (b) (upper right) shows the edges obtained when the statistic G computed using 5×5 neighborhoods on the noisy image of (d) is thresholded at the value 4. (c) (lower left) and (d) (lower right) show thresholds of 8 and 11.

FIGURE 1.4. shows two operating curves for the 5 × 5 slope-facet gradient edge detector. The higher one corresponds to a noisy disk with noise standard deviation of 75 and the upper one corresponds to a noisy disk with noise standard deviation of 50.

The higher one corresponds to a noisy disk with noise standard deviation 75. The lower one corresponds to a noisy disk with noise standard deviation 50.

If it cannot be assumed that the noise variance is the same in each neighborhood, then the estimator using the average of the normalized squared residual errors for σ^2 is not proper. In this case, the local $\epsilon^2/(\sum_r \sum_c 1 - 3)$ can be used as an estimate of the variance in each neighborhood. However, this estimate is not as stable. It does have a higher variance than the estimate based on the average of the local variances, and it has a much lower number of degrees of freedom. Here, to test the hypothesis of no edge for the flat-world assumption, $\alpha = \beta = 0$, we use the ratio

$$F = \left(\left(\hat{\alpha}^2 \sum_r \sum_c r^2 + \hat{\beta}^2 \sum_r \sum_c c^2 \right) /2 \right) / \left(\epsilon^2 / \left(\sum_r \sum_c 1 - 3 \right) \right),$$

which has an F distribution with

$$\left(2, \sum_r \sum_c 1 - 3 \right)$$

degrees of freedom and reject the hypothesis for large values of F.

Again notice that F may be regarded as a significance or reliability measure associated with the existence of a nonzero sloped region in the domain $R \times C$. It is essentially proportional to the squared gradient of the region normalized by

$$\epsilon^2 / \left(\sum_r \sum_c 1 - 3 \right)$$

which is a random variable whose expected value is σ^2, the variance of the noise.

EXAMPLE: Consider the following 3 × 3 region:

3	5	9
4	7	7
0	3	7

Then $\hat{\alpha} = -1.17$, $\hat{\beta} = 2.67$, and $\hat{\gamma} = 5.00$. The estimated gray-tone surface is given by $\hat{\alpha}r + \hat{\beta}r + \hat{\gamma}$ and is

3.50	6.16	8.83
2.33	5.00	7.67
1.17	3.83	6.50

The difference between the estimated and the observed surface is the error and it is

0.50	1.17	−0.17
−1.67	−2.00	0.67
1.17	0.83	−0.50

From this we can compute the squared error $\epsilon^2 = 11.19$. The F statistic is then

$$\frac{[(-1.17)^2 \cdot 6 + (2.67)^2 \cdot 6]/2}{11.19/6} = 13.67.$$

If we were compelled to make a hard decision about the significance of the observed slope in the given 3 × 3 region, we would probably call it a nonzero sloped region, since the probability of a region with true zero slope giving an $F_{2,6}$ statistic of value less than 10.6 is 0.99. 13.67 is greater than 10.6, so we are assured that the probability of calling the region a nonzero sloped region when it is in fact a zero sloped region is

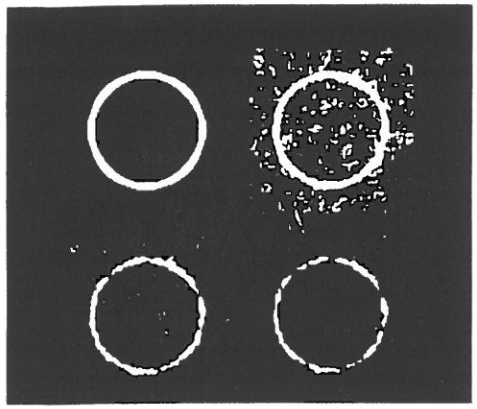


FIGURE 1.5. The edges obtained under a 5 × 5 sloped-facet model using the F statistic. (a) (upper left) shows the thresholded F statistic from the noiseless disk. (b) (upper right) shows the F statistic image of the noisy disk of (d) (noise standard deviation 75) thresholded at 2.32. (c) (lower left) and (d) (lower right) use thresholds of 5.04 and 7.06.

much less than 1%. The statistically oriented reader will recognize the test as a 1% significance level test.

Figure 1.5 shows the edges obtained when a 5 × 5 sloped facet model is employed and when the statistic F computed from each neighborhood of the noisy image of Figure 1.1d is thresholded at 2.32, 5.04, and 7.06. These thresholds should guarantee (under conditions of independent Gaussian noise) that the false-alarm rates are less than .1218, .0158, and .0042, respectively. These thresholds produce observed false-alarm rates of .1236, .0165, and .0042, indicating these were small but negligible departures from the independent Gaussian assumptions. Since these observed false-alarm rates are almost identical to the observed false-alarm rates from the Chi-square tests of Figure 1.3, we may compare the corresponding misidentification rates. The observed misidentification rates for the F test were .0792, .3224, and .5137, all of which are considerably higher than the observed misidentification of the corresponding Chi-square tests. It is obvious from a comparison of these images that the edge noise is worse in the F tests compared to the Chi-square tests. All this is to be expected because the noise meets the assumption of the Chi-square test, and the more one is able to correctly assume about reality, the better the results ought to be when the appropriate statistical test is used.

SEGMENTATION USING THE FACET MODEL

The facet model allows each neighborhood of an image to be thought of as a piecewise linear (or constant or quadratic or cubic) surface. One way in which this facet representation is useful is for image segmentation. There are a variety of approaches. We describe here one approach to determine an initial segmentation and a second approach to merge regions of an initial segmentation. The first approach was originally discussed in Haralick and Shapiro (1985) and the second approach in Pong, Shapiro, Watson, & Haralick (1984).

Determining an Initial Segmentation

To find regions of a segmentation we must look for connected sets of resolution cells that are surely on the same gray-tone intensity surface. To find edges we must look for pairs of adjacent regions having significantly different surfaces. To do segmentation, we must do both. This suggests the following hybrid linkage-combination technique. Select an appropriate-sized neighborhood. Run this neighborhood over the image. For each location where the neighborhood may be placed on the image, determine the parameters of a sloped facet-surface fit as well as the ϵ^2 error of the fit. Use this information to create an edge image.

Now perform a region-growing algorithm on the nonedge pixels. This means that no linking is performed across edge pixels, and edge pixels are not assigned to any region. Edges are barriers to the region-growing process. Such a region-growing technique is described in Haralick and Shapiro (1985). The image is scanned in usual raster-scan order: left to right and top to bottom. Each current pixel then neighbors four pixels (for 8-connectivity), which have been previously scanned. If a previously scanned pixel is not an edge pixel then it belongs to some already existing, but not necessarily completed, region segment. This not necessarily completed region segment has a mean and variance. If the value of the current pixel is not significantly different from the mean of such a neighboring region segment, the pixel is added to the segment and the mean and variance of the segment is updated. Here, significantly different means by a T test.

If there is more than one region that is close enough, then the test pixel is added to the closest region. If the means of any two competing neighboring regions are each close enough to the current pixel value and close enough to each other, then the two regions are merged and the pixel is added to the merged regions. If no neighboring region has its mean close enough, then a new segment is established containing the

current pixel and having a mean value that is the value of the current pixel.

We now define the T test precisely. Let R be a segment containing N pixels and whose pixels neighbor the current pixel. The mean $\bar{X}$ and the scatter S^2 of region R are defined by

$$\bar{X} = \frac{1}{N} \sum_{(r,c) \in R} I(r, c)$$

$$S^2 = \sum_{(r,c) \in R} (I(r, c) - \bar{X})^2$$

Let the current pixel have the value y.

Under the assumption that all the pixels in R and the test pixel y are independent and identically distributed normals, the statistic

$$T = \left| \frac{(N-1)N}{(N+1)} (y - \bar{X})^2 / S^2 \right|^{1/2}$$

has a T_{N-1} distribution. If T is small enough, y is added to region R, and the mean and scatter are updated using y. The new mean and scatter are given by

$$\bar{X}_{\text{new}} \leftarrow (N \bar{X}_{\text{old}} + y) / (N + 1)$$

and

$$S^2_{\text{new}} \leftarrow S^2_{\text{old}} + (y - \bar{X}) + N(\bar{X}_{\text{new}} - \bar{X}_{\text{old}})^2.$$

If T is too high, the value y is not likely to have arisen from the population of pixels in R. If y is different from all of its neighboring regions then it begins its own region. A slightly stricter linking criterion can require that not only must y be close enough to the mean of the neighboring regions, but that a neighboring pixel in that region must have a close enough value to y. This combines a centroid linkage and single-linkage criterion.

To give a precise meaning to the notion of too high a difference, we use an α level statistical significance test. The fraction α represents the probability that a T statistic with $N-1$ degrees of freedom will exceed the value $t_{N-1}(\alpha)$. If the observed T is larger than $t_{N-1}(\alpha)$, then we declare the difference to be significant. If the pixel and the segment really come from the same population, the probability that the test provides an incorrect answer is α.

The significance level α is a user-provided parameter. The value of $t_{N-1}(\alpha)$ is higher for small degrees of freedom and lower for larger degrees of freedom. Thus, region scatters considered to be equal, the

larger a region is, the closer the value of a pixel has to be to the mean of that region in order to be merged with that region.

Note that all regions initially begin as one pixel in size. To avoid the problem of division by 0 (for S^2 is necessarily 0 for 1-pixel regions and 0 for regions having identically valued pixels), a small positive constant can be added to S^2. One convenient way of determining the constant is to decide on a prior variance $V_0 > 0$ and an initial segment size N_0. The initial scatter for a new 1-pixel region is then given by $N_0 V_0$ and the new initial region size is given by N_0. This mechanism keeps the degrees of freedom of the T-statistic high enough so that a significant difference is not the huge difference required for a T-statistic with a small number of degrees of freedom.

Region Merging

The problem with using the initial segmentation just described as input to a higher-level algorithm attempting to recognize objects in the scene is that the regions are too small to be meaningful. This problem motivated us to develop a region-growing scheme that starts with an initial segmentation and produces a new segmentation having larger, hopefully more useful regions. Such a procedure could be repeated any number of times producing a sequence of rougher and rougher segmentations. The final result or the entire sequence of segmentations might prove useful to a higher-level process.

Once an initial segmentation has been produced, (either by the method previously described or some alternate method) properties of the initial regions are computed. These resulting property vectors are used in the region-merging process. Among the properties measured for each region, we have used the following in our experiments.

1. *Size* is simply the number of pixels in a region.
2. *Mean gray level* is the average gray-level intensity in a region.
3. *Elongation* is a measure of the shape of a figure. It is obtained by finding the covariance matrix M of the distribution of $(r - \bar{r}, c - \bar{c})$ where (r, c) represents the coordinates of a pixel in region R, and $(\bar{r}, \bar{c})$ is the center of mass of R.

The matrix M is defined by

$$M = \begin{pmatrix} \sum_{(r,c) \in R} (r - \bar{r})^2 & \sum_{(r,c) \in R} (r - \bar{r}) * (c - \bar{c}) \\ \sum_{(r,c) \in R} (c - \bar{c}) * (r - \bar{r}) & \sum_{(r,c) \in R} (c - \bar{c})^2 \end{pmatrix}$$

and $(\bar{r}, \bar{c})$ for a region R is given by

$$\bar{r} = 1/|R| \sum_{(r,c)\in R} r \text{ and } \bar{c} = 1/|R| \sum_{(r,c)\in R} c.$$

Two eigenvalues can be obtained from the matrix M. Elongation is defined as the ratio of the larger eigenvalue to the smaller.

Besides the property vector for each region, a region adjacency graph that gives topological information about the regions is also generated for a segmented image. Two regions $R1$ and $R2$ are said to be *adjacent* for a segmented image if there exists some pixel in $R1$ such that its 4(8)-neighborhood intersects $R2$. The region adjacency graph has nodes corresponding to regions and edges that connect together nodes representing adjacent regions.

Now we are ready to describe the merging process. Suppose an initial segmentation is given. We group regions using an iterative scheme. Each region is represented by a property vector. At each iteration, the property vector of a region can be replaced by some function of the property vectors of the regions constituting its best fitting neighborhood of regions, a concept to be soon defined. After convergence of the iterative procedure, connected sets of regions with similar revised property vectors become the new regions.

The merging algorithm has two phases. In phase 1 the properties of each region are updated based upon the properties of its region neighborhood. In phase 2 adjacent regions that have similar updated property values are merged together. We now describe the algorithm and its several variations in detail.

Phase 1

Suppose that the image-spatial domain has been divided into N nonoverlapping regions labeled $r(1), \ldots, r(N)$ with corresponding property vectors $p^k(1), \ldots, p^k(N)$ at the k^{th} iteration. Define the neighborhood of region r, $NBD(r)$, by

$$NBD(r) = \{r' \mid \text{region } r' \text{ is adjacent to region } r\}.$$

Suppose for some region r that $NBD(r) = \{r'(1), \ldots, r'(m)\}$. Then because neighboring is symmetric, r is also an element of $NBD(r'(j))$ for $j = 1, \ldots, m$. Thus region r participates in m different neighborhoods.

For a given neighborhood X, we define the variance of X, var(X), by

$$\text{var}(X) = \sum_{r(j)\in X} \| (p(j) - \bar{p}(X)) \|^2 / (|X| - 1)$$

where $\bar{p}(X)$ is the mean property vector of X, and $|X|$ is the cardinality of X. The best-fitting neighborhood of region r, $BF(r)$, is that one of the m neighborhoods it participates in that has lowest variance. Thus

$$BF(r) = X*$$

where $X* = NBD(r')$ for some $r' \in NBD(r)$ and

$$\text{var}(X) = \min_{r'\in NBD(r)} \text{var}(NBD(r')).$$

An iteration of the region growing algorithm starts with the set of regions $r(1), \ldots, r(N)$, with property vectors $p^k(1), \ldots, p^k(N)$, and replaces the property vector of each region by some function of the property vectors of its best-fitting neighborhood. That is,

$$p^{k+1}(n) = f(BF^k(r(n))), \; n = 1, \ldots, N$$

where, of course, $BF^k(r(n))$ depends on $p^k(n)$. The process is repeated until it reaches or approaches a fixed point. Then in phase 2, adjacent regions with identical or near identical property vectors are merged to form a new set of regions.

Phase 2

Suppose that we start with a segmented image whose regions are labeled $r^i(1), \ldots r^i(N_i)$. If the process for these regions reaches a fixed point at some iteration, we merge them to form a new set of regions $r^{i+1}(1), \ldots, r^{i+1}(N_{i+1})$ in the following way.

Construct a graph in which the nodes are the regions. Link together all pairs of regions (a) that are adjacent and (b) whose updated property vectors are close enough. Determine the connected components of the resulting graph. Each connected component corresponds to a subset of regions whose union constitutes one of the merged regions for the next cycle. For an image with T regions, let $n(i)$ be the number of neighbors for region i. Then the number of computations for an iteration of this algorithm is proportional to $n(1) + n(2) + \cdots + n(T)$, which gives a computational complexity $O(T \times \bar{n})$, where $\bar{n}$ is the average number of neighboring regions for all the regions. In most cases $\bar{n} << T$, which makes this an efficient algorithm.

Updating the Property Vectors

One of the most important steps in this region-growing algorithm is to update the property vector for each region. Three different alternatives have been tried. The first method uses the mean property vector of its

best-fitting neighborhood. At iteration k, the updated property vector of a region R is given by

$$p^{k+1}(R) = \overline{p}(R*)$$

where $R* = BF(R)$. Extensions of the theorem for the flat-facet model in (Haralick & Watson, 1981) guarantee the convergence of this method.

The second way of updating the property vector of a region is to make it take on the original property vector of the region that defines the best-fitting neighborhood. That is, instead of using $p^{k+1}(R) = \overline{p}(R*)$ as in method (1), the property vector is updated by

$$p^{k+1}(R) = p^k(S)$$

where $NBD(S) = BF(R)$. Results showed that the rate of merging by this method is faster than the first. But unfortunately, this method shows oscillatory behavior for some images; it does not always lead to a fixed point.

The third alternative is to calculate the mean by weighting the property vectors of a neighborhood by their region sizes. Our results show that the first approach, using the mean property vector, is the most reliable method.

Thresholded Flat-Facet Iteration

The updating schemes described in the previous section recompute the property vector of every region at each iteration. To prevent inaccurate segmentation due to property vectors changing too much, we need to inhibit the updating if the new property vector of a region is too different from the original one. To accomplish this, the idea of a thresholded flat-facet iteration is introduced. For a flat-facet iteration with threshold e, the updated property vector of a region R is given by

$$p^{k+1}(R) = p^k(R), \qquad \text{if } \|\overline{p}(R*) - p^k(R)\| > e$$
$$= \overline{p}(R*), \text{ otherwise.}$$

The convergence of the flat facet iteration is guaranteed (Haralick & Watson, 1981). The proof that the *thresholded* flat-facet iteration also converges was given in Pong, Shapiro, & Haralick (1985).

To illustrate the region-merging technique, we begin with a busy image (Figure 1.6a) and an equally busy initial segmentation (Figure 1.6b). Figure 1.7a shows the segmentation obtained after one iteration of region merging, using only the gray-tone property and employing the first approach to updating the property vectors. After only one iteration, some order begins to appear in the cluttered segmentation. Figure 1.7.b shows the segmentation obtained after two iterations: A few meaningful

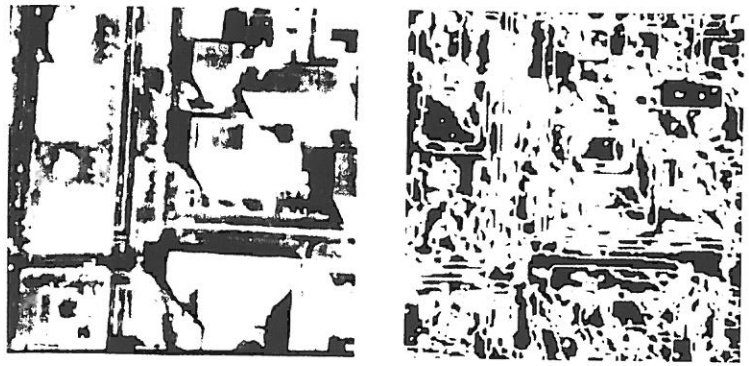


FIGURE 1.6. (a) shows an aerial image of buildings. (b) shows its initial segmentation using the techniques just described.

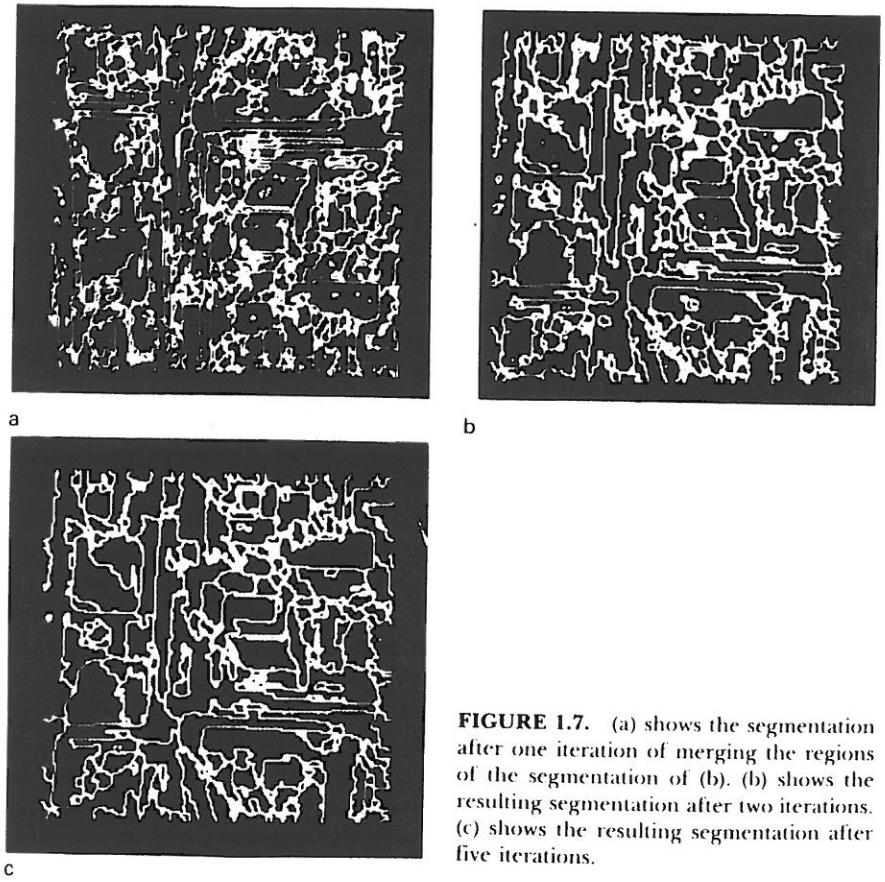


FIGURE 1.7. (a) shows the segmentation after one iteration of merging the regions of the segmentation of (b). (b) shows the resulting segmentation after two iterations. (c) shows the resulting segmentation after five iterations.

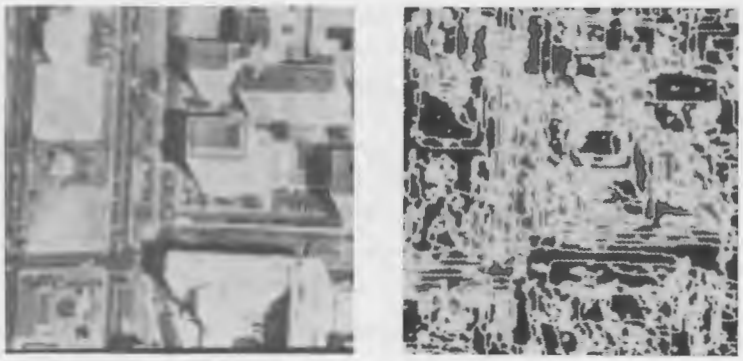
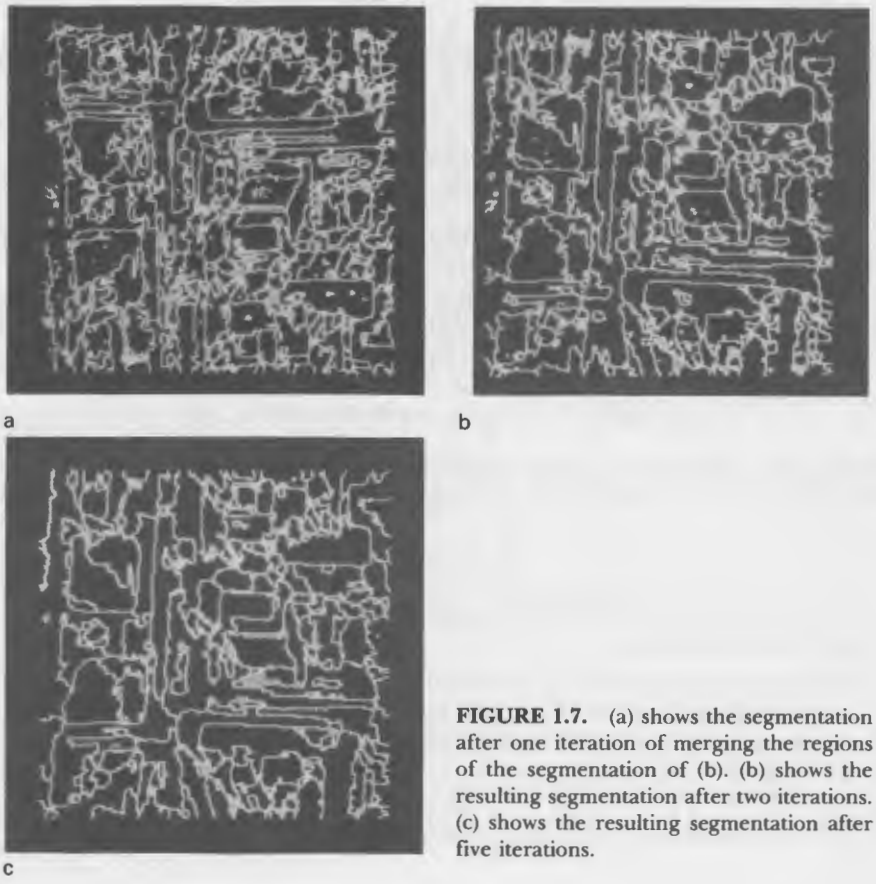


FIGURE 1.6. (a) shows an aerial image of buildings. (b) shows its initial segmentation using the techniques just described.



a

b

c

FIGURE 1.7. (a) shows the segmentation after one iteration of merging the regions of the segmentation of (b). (b) shows the resulting segmentation after two iterations. (c) shows the resulting segmentation after five iterations.

structures, such as streets and buildings, are now showing up. Figure 1.7c shows the segmentation obtained after five iterations. More merging has taken place and there are more meaningful regions to work with. Of course, we must remember that this is a blind, bottom-up segmentation technique. Unless it is used with a knowledge-based system that understands the semantics of some of the variations in gray tone of the original image, we cannot expect such a system to segment into meaningful objects.

THE TOPOGRAPHIC PRIMAL SKETCH

Our classification approach is based on the estimation of the first- and second-order directional derivatives. In two dimensions, the rate of change of a function f depends on direction. We denote the directional derivative of F at the point (r, c) in the direction of θ by $f'_\theta(r,c)$. It is defined as

$$f'_\theta(r, c) = \lim_{h \to 0} \frac{f(r + h^* \sin\theta, c + h^* \cos\theta) - f(r, c)}{h}$$

The direction angle θ is the clockwise angle from the column axis. It follows directly from this definition that

$$f'_\theta(r, c) = \frac{\partial f}{\partial r}(r, c) * \sin\theta + \frac{\partial f}{\partial c}(r, c) * \cos\theta.$$

We denote the second derivative of f at the point (r, c) in the direction θ by $f''_\theta(r, c)$, and it follows that

$$f''_\theta = \frac{\partial^2 f}{\partial r^2} * \sin^2\theta + 2 * \frac{\partial f}{\partial r\, \partial c} * \sin\theta * \cos\theta + \frac{\partial f}{\partial c^2} * \cos^2\theta.$$

The *gradient* of f is a vector whose magnitude,

$$\left(\left(\frac{\partial f}{\partial r} \right)^2 + \left(\frac{\partial f}{\partial c} \right)^2 \right)^{1/2}$$

at a given point (r, c) is the maximum rate of change of f at that point, and whose direction,

$$\tan^{-1} \begin{pmatrix} \partial f / \partial r \\ \partial f / \partial c \end{pmatrix}$$

is the direction in which the surface has the greatest rate of change.

The Mathematical Properties

We will use the following notation to describe the mathematical properties of our various topographic categories for continuous surfaces. Let

∇f = gradient vector of a function f;

$\|\nabla f\|$ = gradient magnitude;

$\omega^{(1)}$ = unit vector in direction in which second directional derivativehas greatest magnitude;

$\omega^{(2)}$ = unit vector orthogonal to $\omega^{(1)}$;

λ_1 = value of second directional derivative in the direction $\omega^{(1)}$;

λ_2 = value of second directional derivative in the direction of $\omega^{(2)}$;

$\nabla f \cdot \omega^{(1)}$ = value of first directional derivative in the direction of $\omega^{(1)}$; and

$\nabla f \cdot \omega^{(2)}$ = value of first directional derivative in the direction of $\omega^{(2)}$.

Without loss of generality, we assume $|\lambda_1| \geq |\lambda_2|$.

Each type of topographic structure in our classification scheme is defined in terms of the aforementioned quantities. In order to calculate these values, the first- and second-order partials with respect to r and c need to be approximated. These five partials are as follows:

$$\partial f/\partial r,\ \partial f/\partial c,\ \partial^2 f/\partial r^2,\ \partial^2 f/\partial c^2,\ \partial^2 f/\partial r\, \partial c.$$

The gradient vector is simply $(\partial f/\partial r,\ \partial f/\partial c)$. The second directional derivatives may be calculated by forming the *Hessian* where the Hessian is a 2×2 matrix defined as

$$H = \begin{vmatrix} \partial^2 f/\partial r^2 & \partial^2 f/\partial r\, \partial c \\ \partial^2 f/\partial c\, \partial r & \partial^2 f/\partial c^2 \end{vmatrix}$$

Hessian matrices are used extensively in nonlinear programming. Only three parameters are required to determine the Hessian matrix H, since the order of differentiation of the cross partials may be interchanged. That is,

$$\partial^2 f/\partial r\, \partial c = \partial^2 f/\partial c\, \partial r.$$

The eigenvalues of the Hessian are the values of the extrema of the second-directional derivative, and their associated eigenvectors are the directions in which the second-directional derivative is extremized. This can easily be seen by rewriting f''_θ as the quadratic form

$$f''_\theta = (\sin\theta \cos\theta) * H * \left| \frac{\sin\theta}{\cos\theta} \right| .$$

Thus

$$H\omega^{(1)} = \lambda_1\omega^{(1)}, \text{ and } H\omega^{(2)} = \lambda_2\omega^{(2)}.$$

Furthermore, the two directions represented by eigenvectors are orthogonal to one another. Since H is a 2×2 symmetric matrix, calculation of the eigenvalues and eigenvectors can be done efficiently and accurately using the method of Rutishauser. We may obtain the values of the first-directional derivative by simply taking the dot product of the gradient with the appropriate eigenvector:

$$\nabla f \cdot \omega^{(1)}$$

$$\nabla f \cdot \omega^{(2)}.$$

There is a direct relationship between the eigenvalues λ_1 and λ_2 and curvature in the directions $\omega^{(1)}$ and $\omega^{(2)}$: When the first-directional derivative $\nabla f \cdot \omega^{(1)} = 0$, then $\lambda_1/(1 + (\nabla f \cdot \nabla f))^{3/2}$ is the curvature in the direction $\omega^{(i)}$, $i = 1$ or 2.

Having the gradient magnitude and direction and the eigenvalues and eigenvectors of the Hessian, we can describe the topographic classification scheme. A peak occurs where there is a local maxima in all directions. In other words, we are on a peak if, no matter what direction we look in, we see no point that is as high as the one we are on. The curvature is downward in all directions. At a peak, the gradient is zero, and the second-directional derivative is negative in all directions. To test whether the second-directional derivative is negative in all directions, we just have to examine the value of the second-directional derivative in the directions that make it smallest and largest. A point is therefore classified as a peak if it satisfied the following conditions:

$$\|\nabla f\| = 0, \lambda_1 < 0, \lambda_2 < 0.$$

A pit is identical to a peak except that it is a local minima in all directions rather than a local maxima. At a pit the gradient is zero, and the second-directional derivative is positive in all directions. A point is classified as a pit if it satisfies the following conditions:

$$\|\nabla f\| = 0, \lambda_1 > 0, \lambda_2 > 0.$$

Ridge

A ridge occurs on a ridge-line, a curve consisting of a series of ridge points. As we walk along the ridge-line, the points to the right and left of us are lower than the ones we are on. Furthermore, the ridge-line may be flat, slope upward, slope downward, curve upward, or curve downward. A ridge occurs where there is a local maximum in one direction. Therefore, it must have negative second-directional derivative in the direction across the ridge and also a zero first-directional derivative in that same direction. The direction in which the local maximum occurs may correspond to either of the directions in which the curvature is "extremized," since the ridge itself may be curved. For nonflat ridges, this leads to the following first two cases for ridge characterization. If the ridge is flat, then the ridge-line is horizontal and the gradient along it is zero. This corresponds to the third case. The defining characteristic is that the second-directional derivative in the direction of the ridge-line is zero, while the second-directional derivative across the ridge-line is negative. A point is therefore classified as a ridge if it satisfies any one of the following three sets of conditions:

$$\|\nabla f\| \neq 0, \lambda_1 < 0, \nabla f \cdot \omega^{(1)} = 0$$

or

$$\|\nabla f\| \neq 0, \lambda_2 < 0, \nabla f \cdot \omega^{(2)} = 0$$

or

$$\|\nabla f\| = 0, \lambda_1 < 0, \lambda_2 = 0.$$

A geometric way of thinking about the definition for ridge is to realize that the condition $\nabla f \cdot \omega^{(1)} = 0$ means that the gradient direction (which is defined for nonzero gradients) is orthogonal to the direction $\omega^{(1)}$ of extremized curvature.

Ravine

A ravine (valley) is identical to a ridge except that it is a local minimum (rather than maximum) in one direction. As we walk along the ravine-line, the points to the right and left of us are higher than the one we are on. A point is classified as a ravine if it satisfies any one of the following three sets of conditions:

$$\|\nabla f\| \neq 0, \lambda_1 > 0, \nabla f \cdot \omega^{(1)} = 0$$

or

$$\|\nabla f\| \neq 0, \lambda_2 > 0, \nabla f \cdot \omega^{(2)} = 0$$

or

$$\|\nabla f\| = 0, \lambda_1 > 0, \lambda_2 = 0.$$

Saddle

A saddle occurs where there is a local maximum in one direction and a local minimum in a perpendicular direction. A saddle must therefore have positive curvature in one direction and negative curvature in a perpendicular direction. At a saddle, the gradient magnitude must be zero and the extrema of the second-directional derivative must have opposite signs. A point is classified as a saddle if it satisfies the following conditions:

$$\|\nabla f\| = 0, \lambda_1 * \lambda_2 < 0.$$

Flat

A flat is a simple, horizontal surface. It, therefore, must have zero gradient and no curvature. A point is classified as a flat if it satisfies the following conditions:

$$\|\nabla f\| = 0, \lambda_1 = 0, \lambda_2 = 0.$$

Given that the aforementioned conditions are true, a flat may be further classified as a *foot* or *shoulder*. A foot occurs at that point where the flat just begins to turn up into a hill. At this point, the third-directional derivative in the direction toward the hill will be nonzero, and the surface increases in this direction. The shoulder is an analogous case and occurs where the flat is ending and turning down into a hill. At this point, the maximum magnitude of the third-directional derivative is nonzero, and the surface decreases in the direction toward the hill. If the third-directional derivative is zero in all directions, then we are in a flat, not near a hill. Thus a flat may be further qualified as being a foot or shoulder, or not qualified at all.

Hillside

A hillside point is anything not covered by the previous categories. It has a nonzero gradient and no strict extrema in the directions of maximum and minimum second-directional derivative. If the hill is simply a tilted flat (i.e., has constant gradient), we call it a "slope." If its curvature is positive (upward), we call it a "convex hill." If its curvature is negative

(downward), we call it a "concave hill." If the curvature is up in one direction and down in a perpendicular direction, we call it a "saddle hill."

A point on a hillside is an "inflection point" if it has a zero-crossing of the second-directional derivative taken in the direction of the gradient. The inflection-point class is the same as the "step edge" defined by Haralick (1984) who classifies a pixel as a step edge if there is some point in the pixel's area having a negatively sloped zero-crossing of the second-directional derivative taken in the direction of the gradient.

To determine whether a point is a hillside, we just take the complement of the disjunction of the conditions given for all the previous classes. Thus, if there is no curvature, then the gradient must be nonzero. If there is curvature, then the point must not be a relative extremum. Therefore, a point is classified as a hillside if all three sets of the following conditions are true ('$\rightarrow$' represents the operation of logical implication):

$$\lambda_1 = \lambda_2 = 0 \rightarrow \|\nabla f\| \neq 0,$$

and

$$\lambda_1 \neq 0 \rightarrow \nabla f \cdot \omega^{(1)} \neq 0,$$

and

$$\lambda_2 \neq 0 \rightarrow \nabla f \cdot \omega^{(2)} \neq 0.$$

Rewritten as a disjunction of clauses rather than a conjunction of clauses, a point is classified as a hillside if any one of the following four sets of conditions are true:

$$\nabla f \cdot \omega^{(1)} \neq 0, \nabla f \cdot \omega^{(2)} \neq 0$$

or

$$\nabla f \cdot \omega^{(1)} \neq 0, \lambda_2 = 0$$

or

$$\nabla f \cdot \omega^{(2)} \neq 0, \lambda_1 = 0$$

or

$$\|\nabla f\| \neq 0, \lambda_1 = 0, \lambda_2 = 0.$$

We can differentiate between different classes of hillsides by the values of the second-directional derivative. The distinction can be made as follows:

Slope	if $\lambda_1 = \lambda_2 = 0$
Convex	if $\lambda_1 >= \lambda_2 >= 0, \lambda_1 \neq 0$
Concave	if $\lambda_1 <= \lambda_2 <= 0, \lambda_1 \neq 0$
Saddlehill	if $\lambda_1 * \lambda < 0$

A slope, convex, concave, or saddle hill is classified as an inflection point if there is a zero-crossing of the second-directional derivative in the direction of maximum first-directional derivative (i.e., the gradient).

Summary of the Topographic Categories

A summary of the mathematical properties of our topographic structures on continuous surfaces can be found in Table 1.1. The table exhaustively defines the topographic classes by their gradient magnitude, second-directional derivative extrema values, and the first-directional derivatives taken in the directions that extremize second-directional derivatives. Each entry in the table is either 0, +, −, or *. The 0 means not significantly different from zero on the positive side; − means significantly different from zero on the negative side, and * means it does not matter. The label "Cannot occur" means that it is impossible for the gradient to be nonzero and the first-directional derivative to be zero in two orthogonal directions.

From Table 1.1 one can see that our classification scheme is complete. All possible combinations of first and second-directional derivatives have a corresponding entry in the table. Each topographic category has a set of mathematical properties that uniquely determines it.

TABLE 1.1
Mathematical Properties of Topographic Structures

$\|\nabla f\|$	λ_1	λ_2	$\nabla f \cdot w^{(1)}$	$\nabla f \cdot w^{(2)}$	*Label*
0	−	−	0	0	Peak
0	−	0	0	0	Ridge
0	−	+	0	0	Saddle
0	0	0	0	0	Flat
0	+	−	0	0	Saddle
0	+	0	0	0	Ravine
0	+	+	0	0	Pit
+	−	−	−, +	−, +	Hillside
+	−	*	0	*	Ridge
+	*	−	*	0	Ridge
+	−	0	−, +	*	Hillside
+	−	+	−, +	−, +	Hillside
+	0	0	*	*	Hillside
+	+	−	−, +	−, +	Hillside
+	+	0	−, +	*	Hillside
+	+	*	0	*	Ravine
+	*	+	*	0	Ravine
+	+	+	−, +	−, +	Hillside
+	*	*	0	0	Cannot occur

(*Note:* Special attention is required for the degenerate case $\lambda_1 = \lambda_2 \neq 0$, which implies that $\omega^{(1)}$ and $\omega^{(2)}$ can be *any* two orthogonal directions. In this case, there *always* exists an extreme direction ω which is orthogonal to ∇f, and thus the first-directional derivative $\nabla f \cdot \omega$ is *always* zero in an extreme direction. To avoid spurious zero-directional derivatives, we choose $\omega^{(1)}$ and $\omega^{(2)}$ such that $\nabla f \cdot \omega^{(1)} \neq 0$ and $\nabla f \cdot \omega^{(2)} \neq 0$, unless the gradient is zero.)

Local Cubic Facet Model

In order to estimate the required partial derivatives, we perform a least-squares fit with a two-dimensional surface, f, to a neighborhood of each pixel. It is required that the function f be continuous and have continuous first- and second-order partial derivatives with respect to r and c in a neighborhood around each pixel in the rc plane.

We choose f to be a cubic polynomial in r, and c expressed as a combination of discrete orthogonal polynomials. The function f is the best discrete least-squares polynomial approximation to the image data in each pixel's neighborhood. More details can be found in Haralick and Watson (1981), in which each coefficient of the cubic polynomial is evaluated as a linear combination of the pixels in the fitting neighborhood.

To express the procedure precisely and without reference to a particular set of polynomials tied to neighborhood size, we will canonically write the fitted bicubic surface for each fitting neighborhood as

$$\begin{aligned} f(r, c) = {} & k_1 + k_2 r + k_3 c \\ & + k_4 r^2 + k_5 rc + k_6 c^2 \\ & + k_7 r^2 + k_8 r^2 c + k_9 r c^2 + k_{10} c^3, \end{aligned}$$

where the center of the fitting neighborhood is taken as the origin. It quickly follows that the needed partials evaluated at local coordinates (r, c) are

$$\begin{aligned} \partial f/\partial r &= k_2 + 2k_4 r + k_5 c + 3k_7 r^2 + 2k_8 rc + k_9 c^2 \\ \partial f/\partial c &= k_3 + k_5 r + 2k_6 c + k_8 r^2 + 2k_9 rc + 3k_{10} c^2 \\ \partial^2 f/\partial r^2 &= 2k_4 + 6k_7 r + 2k_8 c \\ \partial^2 f/\partial c^2 &= 2k_6 + 2k_9 r + 6k_{10} c \\ \partial^2 f/\partial r \partial c &= k_5 + 2k_8 r + 2k_9 c. \end{aligned}$$

It is easy to see that the above quantities are evaluated at the center of the pixel where local coordinates $(r, c) = (0, 0)$, only the constant terms will be of significance. If the partials need to be evaluated at an arbitrary point in a pixel's area, then a linear or quadratic-polynomial value must be computed.

THREE-DIMENSIONAL SHAPE ESTIMATION

Consider an image of a three-dimensional object illuminated by an arbitrary light source and viewed from an arbitrary position. Although ambiguities are possible, frequently the human viewer can estimate (a) the three-dimensional shape of the object, (b) the camera position, and (c) the location of the light source. The original "shape-from-shading" techniques (Horn, 1975) solve systems of differential equations to derive three-dimensional shape from gray-tone intensity variations and operate under a limiting set of restrictions. In addition to low level shading cues, we believe that the human viewer also recognizes patterns in the image that give cues leading to estimation of the shape of the object.

Extracting patterns from the original gray-tone image is, in most nontrivial cases, an impossible task. In fact, it is for this reason that syntactive pattern-recognition systems have had to first extract descriptions consisting of primitives, their properties, and their interrelationships from the image and then to parse these descriptions according to the rules of a grammar. Instead of trying to recognize patterns at the gray-tone intensity level, we propose to work at the topographic-labeling level. Our goal is to use patterns expressed in terms of ridges and valleys, peaks and pits, flats and hillsides to estimate three-dimensional shape.

Imaging Geometry

The relationship between scene coordinates and image coordinates is illustrated in Figure 1.8. We assume that the camera lens is at the origin and that the z-axis is directed toward the image plane which is in front of the lens. The image plane is placed at a distance f, the focal length of the lens, in front of the origin so that the image is oriented in the same way as the scene. As seen from Figure 1.11, the following relations hold for perspective projection:

$$u = \frac{fx}{z} \text{ and } v = \frac{fy}{z} .$$

In our discussion, the perspective projection is approximated by an orthographic projection. This approximation is good when the size of the objects being imaged is small compared to the viewing distance. In this case, appropriate coordinate systems can be chosen such that the following relations hold:

$$u = x \text{ and } v = y.$$

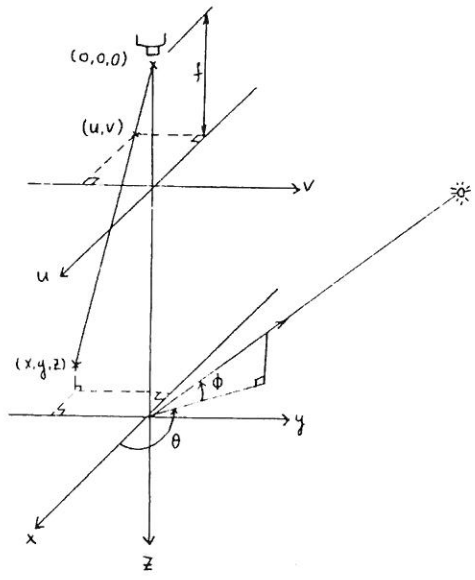


FIGURE 1.8. Relationship between scene coordinates and image coordinates.

Illumination Model

In the following discussion, we will use a simple illumination model that assumes a distant point-light source and a Lambertian reflectance model. A Lambertian surface scatters light equally in all direction. The brightness of a Lambertian surface illuminated by a distant point-light source is given by:

$$I = I_0 N \cdot L \tag{1}$$

where I_0 is a constant depending on the surface albedo and the intensity of the light source, N is the unit surface normal vector, and L is the unit vector of the illumination direction.

The unit vector which points in the direction of the light source can be specified by the two angles shown in Figure 1.11. The first is the azimuth (θ), which is the angle between the x-axis and the projection of the vector onto the x-y plane, while the second is the angle of elevation (ϕ) of the light source. If we represent this unit vector by $[a, b, c]$ then

$$a = \cos\theta \cos\phi,$$

$$b = \sin\theta \cos\phi, \text{ and,}$$

$$c = -\sin\phi.$$

In our discussion, we will consider only positive values of ϕ. Therefore, c is always less than zero.

If the height of the object surface above the x-y plane is expressed as a function of x and y,

$$z = S(x, y),$$

then the surface normal is given by the vector:

$$N = [S_x, S_y, -1]/(1 + S_x^2 + S_y^2)^{1/2}$$

where S_x and S_y denote first partials of S with respect to x and y, respectively. By carrying out the dot product in Equation 1, it follows that

$$I = I_0 \frac{aS_x + bS_y - c}{(1 + S_x^2 + S_y^2)^{1/2}} . \qquad (2)$$

Shape From Topographic Patterns

There are two possible methods for determining the pattern of topographic labels that will appear, given a particular three-dimensional shape category, a particular reflectance model, a particular light source, and a particular viewpoint. The first method is to work the problem analytically, obtaining exact equations for the illuminated surface. At each point the gradient, eigenvectors, and eigenvalues can be computed in order to determine precisely which sets of points have the various topographic labels. The second method is to work the problem experimentally, using software to generate digital images of illuminated three-dimensional surfaces, to fit these images with either polynomials, splines, or discrete cosines, and to assign topograpic labels to each pixel. The first method has the advantage of exactness and the disadvantage of becoming extremely difficult for all but the simplest surfaces. The second method has the advantage of being applicable to a wide variety of surfaces and illuminating conditions and the disadvantage of yielding some inaccurate results due to possible errors in fitting the gray-tone image. We have begun to experiment with both methods, starting with very simple surfaces, the Lambertian reflectance model, and point light sources. We have worked with four simple surfaces: (a) the top half of a cylinder, (b) the upper hemisphere of a sphere, (c) the top half of an ellipsoid, and (d) the upper half of a hyperboloid. Figures 1.9 and 1.10 illustrate the cylinder and the sphere, respectively.

Method 1: The Experimental Approach

The process for topographic classification can be done in one pass through the image. At each pixel of the image, the following four steps,

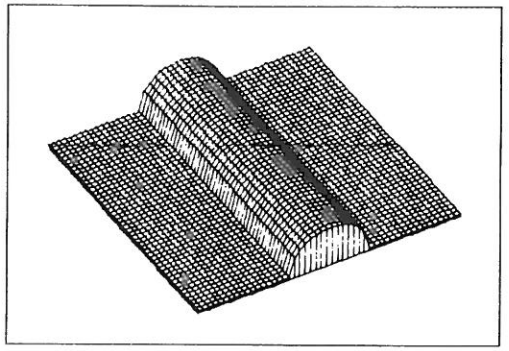


FIGURE 1.9. The cylindrical object used in our experiments.

which were discussed in detail in the previous section need to be performed.

1. Calculate the least-squares fitting coefficients of a two-dimension cubic polynomial in an $n \times n$ neighborhood around the pixel.
2. Use the coefficients calculated in step 1 to find the gradient, the gradient magnitude, and the eigenvalues and eigenvectors of the Hessian at the center of the pixel's neighborhood.
3. Search in the direction of the eigenvectors calculated in step 2 for a zero-crossing of the first-directional derivative within the pixel's area.
4. Recompute the gradient, gradient magnitude, and values of second directional derivative extrema at each zero crossing. Then classify the pixel based on Table 1.1.

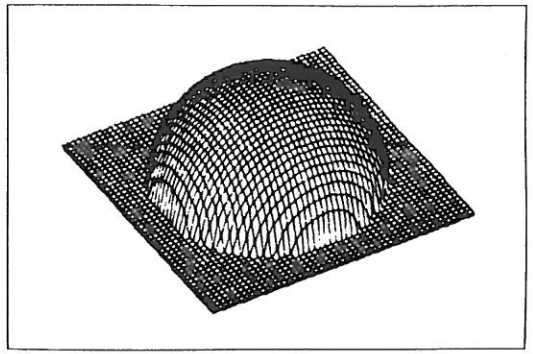


FIGURE 1.10. The spherical object used in our experiments.

Method 2: The Analytical Approach

Topographic Labels on the Cylinder

Consider a cylindrical surface given by:

$$s(x, y) = d - (r^2 - y^2)^{1/2} \text{ for } -r \leq y \leq r \tag{3}$$

where d is the distance of the x-y plane from the camera down the z-axis and r is the radius of the cylinder. This surface in which the axis of the cylinder lies along the x-axis was chosen to simplify calculations. Notice that since only the top half of the cylinder is considered, the sign of the square root in Equation 3 is taken as positive. By differentiating S with respect to x and y, we obtain

$$S_x = 0 \text{ and } S_y = y(r^2 - y^2)^{-1/2}.$$

It follows from Equation 2 that the intensity of the cylinder illuminated from direction (a, b, c) is given by:

$$I(x, y) = I_0(by - c(r^2 - y^2)^{1/2})r \tag{4}$$

After some simplifications, the first and second partials of I are found to be:

$$I_x = I_{xx} = I_{xy} = I_{yx} = 0,$$
$$I_y = I_0(b + cy(r^2 - y^2)^{-1/2})/r, \text{ and}$$
$$I_{yy} = I_0cr(r^2 - y^2)^{-3/2},$$

where the subscripted Is denote partial differentiation with respect to the subscript(s).

Since I_x is equal to zero, the gradient magnitude ($\|\nabla f\|$) is equal to the absolute value of I_y. Therefore, $\|\nabla f\| = 0$ when

$$I_y = I_0(b + cy(r^2 - y^2)^{-1/2})/r = 0$$

which implies

$$b(r^2y^2)^{1/2} = cy. \tag{5}$$

Upon solving Equation 5, we obtain

$$y^2 = r^2b^2/(b^2 + c^2).$$

Because c is always negative, the sign of y is taken to be the same as that of b in order for Equation 5 to be satisfied.

To determine the second-directional derivative extrema values and the first-directional derivatives taken in the directions which extremize second-directional derivatives, we form the Hessian:

$$H = \begin{bmatrix} 0 & 0 \\ 0 & I_0cr(r^2 - y^2)^{-3/2} \end{bmatrix}.$$

The eigenvalues of the Hessian are obtained as:

$$X_1 = I_0cr(r^2 - y^2)^{-3/2} \text{ and} \tag{6}$$
$$X_2 = 0; \tag{7}$$

their associated eigenvectors are:

$$w_1 = (0,1) \text{ and}$$
$$w_2 = (1,0).$$

Recall that c is always negative, therefore, X_1 is always negative for $-r < y < r$. By taking the dot product of the gradient with the eigenvectors, we obtain:

$$\nabla I \cdot w_1 = I_y = I_0(b + cy(r^2 - y^2)^{-1/2}/r \text{ and}$$

$$\nabla I \cdot w_2 = 0.$$

To determine the topographic labels, we need to consider two cases: (1) zero-gradient magnitude and (2) positive-gradient magnitude.

CASE 1: Zero-Gradient Magnitude. If we let $y_0 = rb(b^2 + c^2)^{-1/2}$, it follows from Equation 5 that $\|\nabla I\| = 0$ when $y = y_0$. By Equations 6 and 7, the second directional derivative extrema values at $y = y_0$ are

$$X_1 = I_0cr(r^2 - y_0^2)^{-3/2} \tag{8}$$
$$X_2 = 0. \tag{9}$$

Since X_1 is always less than zero, it follows directly from Table 1.1 that a ridge is located at $y = y_0$.

CASE 2: Positive Gradient Magnitude. If the gradient magnitude ($\|\nabla I\|$) is taken to be positive, then the value of the first-directional derivative in the direction of w_1 ($\nabla I.w_1$) is always non-zero because $\nabla I.w_1 = I_y$ and $\|\nabla f\| = |I_y|$. In this case, since X_1 is always negative and X_2 is always zero, it follows from row 11 of Table 1.1 that hillsides are located at those places where the gradient magnitudes are positive.

Topographic Labels on the Sphere

In the case of the sphere, the equation of a spherical surface with radius r is given by:

$$S(x, y) = d - (r^2 - x^2 - y^2)^{1/2} \quad \text{for } -r \leq x \leq r \text{ and } -r \leq y \leq r \tag{10}$$

Its intensity illuminated from direction $[a, b, c]$ is given by:

$$I(x, y) = I_0 * [ax + by - c(r^2 - x^2 - y^2)^{1/2}]/r \tag{11}$$

After some simplifications, the first and second partials of I are found to be:

$$I_x = I_0[a + cx(r^2 - x^2 - y^2)^{-1/2}]/r,$$
$$I_y = I_0[b + cy(r^2 - x^2 - y^2)^{-1/2}]/r,$$
$$I_{xx} = I_0c(r^2 - y^2)(r^2 - x^2 - y^2)^{-3/2}/r,$$
$$I_{xy} = I_{yx} = I_0cxy(r^2 - x^2 - y^2)^{-3/2}/r, \text{ and}$$
$$I_{yy} = I_0c(r^2 - x^2)(r^2 - x^2 - y^2)^{-3/2}/r.$$

The gradient magnitude ($\|\nabla I\|$) is given by:

$$\|\nabla I\| = (I_x^2 + I_y^2)^{1/2},$$

which is zero when

$$a(r^2 - x^2 - y^2)^{1/2} + cx = 0 \text{ and}$$
$$b(r^2 - x^2 - y^2)^{1/2} + cy = 0$$

are satisfied simultaneously. By squaring and invoking the constraint $a^2 + b^2 + c^2 = 0$ on the unit vector $[a, b, c]$, the solution to the simultaneous equations is found to be:

$$x = ra \text{ and } y = rb.$$

The Hessian for the intensity surface of the illuminated sphere is given by:

$$H = \frac{I_0c}{r(r^2 - x^2 - y^2)} * \begin{bmatrix} r^2 - y^2 & xy \\ xy & r^2 - x^2 \end{bmatrix}.$$

Its eigenvalues are found to be:

$$X_1 = I_0cr(r^2 - x^2 - y^2)^{-3/2} \text{ and}$$
$$X_2 = I_0c(r^2 - x^2 - y^2)^{-1/2}/r.$$

Notice that both eigenvalues are always less than zero since c is always less than zero. The eigenvector corresponding to X_1 is given by:

$$w_1 = [x(x^2 + y^2)^{-1/2}, y(x^2 + y^2)^{-1/2}]$$

and the eigenvector corresponding to X_2 is given by:

$$w_2 = [-y(x^2 + y^2)^{-1/2}, x(x^2 + y^2)^{-1/2}].$$

The dot product of the gradient with w_1 is

$$\nabla I \cdot w_1 = \frac{I_0}{(x^2 + y^2)^{1/2}} \left[x \left(a + \frac{cx}{(r^2 - x^2 - y^2)^{1/2}} \right) \right] \left[y \left(b + \frac{cy}{(r^2 - x^2 - y^2)^{1/2}} \right) \right] \tag{12}$$

and the dot product of the gradient with w_2 is

$$\nabla I \cdot w_2 = I_0(-ay + bx)(x^2 + y^2)^{-1/2}/r. \tag{13}$$

We determine the topographic labels by considering two cases.

CASE 1: Zero Gradient Magnitude. The gradient magnitude is equal to zero when $(x, y) = (ra, rb)$. Since both eigenvalues are less than zero on the illuminated sphere, it follows directly from Table 1.1 that a peak is located at $(x, y) = (ra, rb)$.

CASE 2: Positive Gradient Magnitude. In the case when the gradient magnitude is given to be positive, since both eigenvalues are known to be negative, it follows from Table 1.1 that there is a ridge at those locations where either $\nabla f \cdot w_1 = 0$ or $\nabla f \cdot w_2 = 0$ is satisfied. We obtain from Equations 12 and 13 that

$\nabla I.w_1 = 0$ when $(ax + by)(r^2 - x^2 - y^2)^{1/2} + c(x^2 + y^2) = 0$ and

$\nabla I.w_2 = 0$ when $-ay + bx = 0$.

Table 1.1 also says that hillsides appear at places where both $\nabla f \cdot w_1$ and $\nabla f \cdot w_2$ are non-zero.

Estimation of Surface Orientation

The topographic labels along with their quantitative measurements bear a strong relationship to the surface orientation of the three-dimensional object in the scene. Consider a spherical surface as previously described. The unnormalized surface orientation of such a surface can be represented in the gradient space by the vector $[p, q, -1]$, where

$$p = \frac{x}{(r^2 - x^2 - y^2)^{1/2}} \text{ and } q = \frac{y}{(r^2 - x^2 - y^2)^{1/2}}.$$

An alternative way of specifying surface orientation is the tilt and slant representation. Tilt specifies the orientation of the projection of the surface normal onto the image plane. Slant is the angle between the surface normal and viewing direction. The tilt and slant representation

and the gradient space representation can be related by the following formulas:

$$\text{Tan } \theta = q/p \text{ and}$$
$$\text{Tan } \theta = (p^2 + q^2)^{1/2} \text{ or}$$
$$\text{Cos } \theta = (1 + p^2 + q^2)^{1/2}.$$

In the case of a sphere,

$$\text{Tan } \theta = y/x \text{ and}$$
$$\text{Cos } \phi = (r^2 - x^2 - y^2)/r^2.$$

To see how the surface orientation of a spherical Lambertian surface can be derived from the topographic analysis of the image intensity surface, we need first to complete the analytical results of the previous section by considering the lower half of the sphere. The equation of the lower hemisphere of a sphere whose center is at $(0, 0, d)$ is given by:

$$S(x,y) = d + (r^2 - x^2 - y^2)^{1/2} \quad \text{for } -r \le x \le r \text{ and } -r \le y \le r.$$

Differentiating the aforementioned equation with respect to x and y, we obtain

$$p = \frac{-x}{(r^2 - x^2 - y^2)^{1/2}} \text{ and } q = \frac{-y}{(r^2 - x^2 - y^2)^{1/2}}.$$

Notice the sign difference between the surface orientations of the upper and lower hemispheres.

As for the upper hemisphere, after some simplification, we obtain for the lower hemisphere a similar set of expressions for I and its partials,

$$\begin{aligned}
I &= I_0[-ax - by - c(r^2 - x^2 - y^2)^{1/2}]/r,\\
I_x &= I_0[-a + cx(r^2 - x^2 - y^2)^{-1/2}]/r,\\
I_y &= I_0[-b + cy(r^2 - x^2 - y^2)^{-1/2}]/r,\\
I_{xx} &= I_0c(r^2 - y^2)(r^2 - x^2 - y^2)^{-3/2}/r,\\
I_{xy} &= I_{yx} = I_0cxy(r^2 - x^2 - y^2)^{-3/2}/r, \text{ and}\\
I_{yy} &= I_0c(r^2 - x^2)(r^2 - x^2 - y^2)^{-3/2}/r.
\end{aligned}$$

Notice that the second partials of I are the same for both halves of the sphere. Since the second partials make up the Hessian, it follows that the eigenvalues and eigenvectors for the two hemispheres are also identical. Recall that the eigenvalues and eigenvectors are given by

$$\begin{aligned}
X_1 &= I_0cr(r^2 - x^2 - y^2)^{-3/2},\\
X_2 &= I_0c(r^2 - x^2 - y^2)^{-1/2}/r,\\
w_1 &= [x(x^2 + y^2)^{-1/2}, y(x^2 + y^2)^{-1/2}], \text{ and}\\
w_2 &= [-y(x^2 + y^2)^{-1/2}, x(x^2 + y^2)^{-1/2}].
\end{aligned}$$

If we take the ratio of the smaller over the larger eigenvalue, we obtain

$$\frac{X^2}{X^1} = \frac{(r^2 - x^2 - y^2)}{r^2}.$$

This ratio is the square root of the cosine of the surface slant. Note that the signs of both X_1 and X_2 depend only on the sign of c, which is the negative of the sine of the angle of elevation of the light source. Therefore, the ratio is always positive and its square root is always justifiable. Furthermore, the ratio is always less than or equal to one, since X_2 is the smaller eigenvalue. Thus, we can obtain surface slant by taking the arccosine of the square root of X_2/X_1. The resulting angle is determined uniquely because the surface slant for a visible surface always lies between 0 and $\pi/2$.

The remaining component to be determined for the unit surface normal is the surface tilt. By considering w_1, the eigenvector corresponding to the larger eigenvalue, we can obtain the direction θ in which the second-directional derivative of I is extremized. That is,

$$\tan \theta = \frac{y}{(x^2 + y^2)^{1/2}} \cdot \frac{(x^2 + y^2)^{1/2}}{x} = \frac{y}{x},$$

which is identical to the tangent of the surface tilt. Thus we have $t = \theta$ or $t = \theta + \pi$. Unfortunately, there are two possible solutions. This is expected because each solution corresponds to one half of the sphere. This shows the ambiguity in local analysis of image shading.

Without any assumption about the location of the light source, we found from the aforementioned analytical results that the topographic labels on the underlying intensity surface resulting from a spherical Lambertian surface can only be peak, pit, ridge, valley, convex, or concave hillside. This is because any combination of I_{xx}, I_{yy}, and I_{xy} resulting from a spherical Lambertian surface can produce only either a semipositive or seminegative definite Hessian. Therefore, not all combinations of I_{xx}, I_{yy}, and I_{xy} are possible.

Furthermore, if we approximate three-dimensional surfaces locally by spherical surfaces, it is expected that the radii of the approximating spheres for points or a spherical surface are constant. Recall that the

radius, r, and the eigenvalues, X_1 and X_2, of a spherical Lambertian surface are related by the following expressions:

$$X_1 = \frac{I_0 c r}{(r^2 - x^2 - y^2)^{3/2}} \text{ and } X_2 = \frac{I_0 c}{r(r^2 - x^2 - y^2)^{1/2}} .$$

We obtain from the expression from X_1

$$(r^2 - x^2 - y^2)^{1/2} = \frac{I_0 c}{X_2 r} .$$

We then have from the expression for X_2

$$X_1 = I_0 c r \left(\frac{I_0 c}{X_2 r} \right)^{-3} ,$$

or

$$r^4 = \frac{X_1 I_0^2 c^2}{X_2^3} .$$

Since I_0 and c are fixed, we conclude that X_1/X_2^3 is constant for a spherical Lambertian surface. Therefore, an image point can be determined as resulting from a point on a spherical Lambertian surface only if it is labeled as a peak, pit, ridge, valley, convex hillside or concave hillside, and the radii of the approximating spheres at pixels within the neighborhood around that point are similar enough. What this suggests is that we should estimate surface tilt and surface slant locally from the eigenvalues and eigenvectors of the Hessian of the underlying intensity surface only if the underlying intensity surface is compatible with that of a spherical Lambertian surface.

It can be observed from the expressions of the eigenvalues that a pit, valley, and convex-hillside classification of the intensity surface of a spherical Lambertian surface corresponds to a positive c. This implies a light source below the object surface. Although this is physically possible, such illuminating condition can usually be ignored when solving practical problems. We thus further assume that a spherical Lambertian surface can only result in peak, ridge, and concave-hillside classifications.

Classification of Object Surfaces

We propose here a scheme for partial classification of three-dimensional object surfaces. The basic goal of this classification scheme is to group together pixels that are likely to come from the same surface patch. We limit our consideration to five types of object surfaces. They are planar, developable, spherical, elliptical, and hyperbolic surfaces.

Based on the previous discussions, it is evident that topographic labels together with the signs and magnitudes of their second directional derivatives bear a strong relationship to the nature of the three-dimensional object surface in the scene. This evidence leads to the assumption that maximally connected sets of pixels having the same topographic label belong to the same surface patches. A feature-extraction process is employed to extract these connected sets of topographic structures. The resulting structures are arcs, regions, and topographic labels. The desired topographic structures are then determined by applying a connected-components algorithm to the topographic labels within each region segment.

The assembled topographic structures may be divided into three categories: (1) areal structures, which consist of convex hillsides, concave hillsides, saddle hillsides, flat surfaces, and sloped surfaces; (2) arc structures, which include edges, ridges, and valleys; and (3) point structures, which include peaks, pits, and saddle points. In what follows, we suggest hypotheses that can be made about the three-dimensional objects based on the analytical results that we have derived and the results of the experiments that we have performed. We believe that three-dimensional object shape can be inferred by feeding this knowledge into a hypothesis-based reasoning system.

Areal Structures

Flat. A flat is a simple surface with zero gradient and no curvature. That is, the gray-level intensity is constant in a connected flat structure. Since the surface-normal vectors within a planar surface are constant, we can be almost certain that pixels belonging to a connected flat structure come from the same planar surface. Although this may not hold for shadow areas, we can usually separate shadow areas by identifying flat structures with relatively low intensity averages.

Hillsides. We first hypothesize that a concave/convex hillside assembly is part of a spherical, elliptical, or developable surface, and a saddle hillside assembly is part of a hyperbolic surface. Our first hypothesis is driven by the analytical and experimental results of the cylindrical, spherical, elliptical, and hyperbolic surfaces that we considered.

We further postulate that a concave/convex hillside assembly is part of a developable surface if it is adjacent to a straight and horizontal ridge. In particular, it is part of cylindrical surface if it is concave and the second-directional derivative of the hillside in the direction of the ridge is zero.

As a result of the previous section, a hillside assembly can be consid-

ered as part of a spherical surface if it is concave and the radii of the approximating spheres within the hillside assembly are similar enough. We have not been able to derive a complete classification scheme for all the areal structures. Nonetheless, since the assembled regions are likely to come from the same surface patches, they are good starting regions for shape-from-shading approaches.

Line and Point Structures

While edges are considered to be good indications of the discontinuities, peaks and ridges are found to be significant structures in the images of the conic surfaces that we have considered. The following observations are gathered from the topographic structures of the conic surface:

1. The ridge arcs obtained from the images of the sphere are found to be symmetrical around the peaks.
2. Ridges for the images of the ellipsoid and the hyperboloid are found to be symmetrical around the peaks only if the projection of the light vectors are parallel to one of the axes of these conic surfaces.
3. Straight ridge lines are found in the images of the cylinder. The gray-tone intensities along the ridge lines are found to be constant.
4. While the ridges around the peaks found in the image of the ellipsoid curve away from the light sources, those of the hyperboloid curve toward the light sources.
5. The peaks located in the images of the conic surfaces correspond to locations where the surface normals are pointing toward the light sources.

Results

We will show the analytical and experimental results of the topographic patterns on the cylinder and sphere. See Pong, Shapiro, and Haralick (1985) for results on other surfaces. Three illumination conditions are considered for each surface: (1) the light direction is $(0, 0, -1)$, which means directly above the center of the surface; (2) the light direction is $(0, \sqrt{3}/2, -1/2)$, which translates to azimuth 0° and elevation 30°; (3) the light direction is $(1/2, 1/2, -1/\sqrt{2})$, which translates to azimuth 45° and elevation 45°. The illuminated surfaces of the cylinder and the sphere are shown in Figure 1.11 and Figure 1.12 respectively.

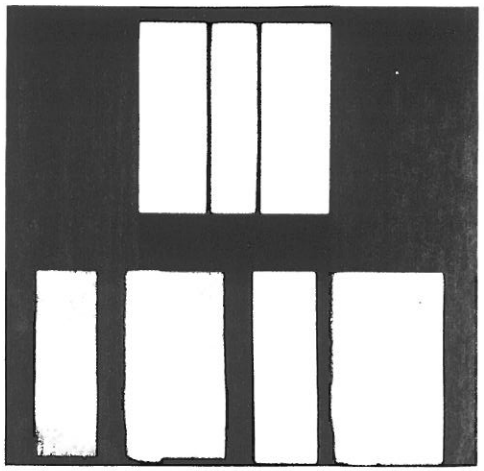


FIGURE 1.11. Shaded images of the cylinder of Figure 1.9.

Analytical Results for the Cylinder

When the light direction is from azimuth 0°, elevation 90°, analytical results indicate a ridge parallel to the axis of the cylinder and running along the center of the top half as shown in Figure 1.13. When the light direction is from azimuth 0° and elevation 30°, the ridge appears as in Figure 1.13. When the light direction is from azimuth 45° and elevation 45°, the ridge appears as in Figure 1.13. In all three cases, the remaining points of the cylinder are hillsides.

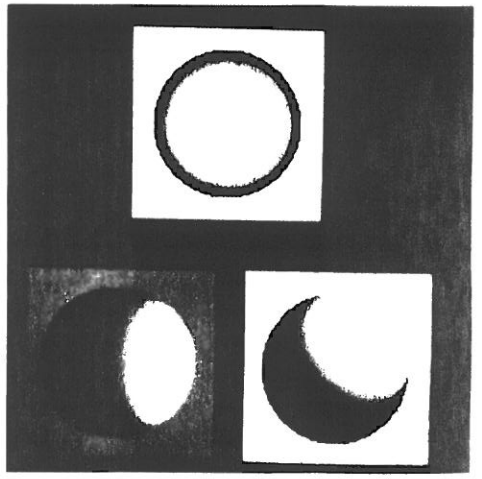


FIGURE 1.12. Shaded images of the sphere of Figure 1.10.

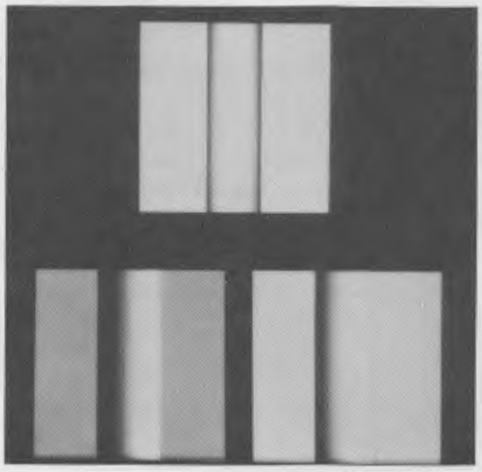


FIGURE 1.11. Shaded images of the cylinder of Figure 1.9.

Analytical Results for the Cylinder

When the light direction is from azimuth 0°, elevation 90°, analytical results indicate a ridge parallel to the axis of the cylinder and running along the center of the top half as shown in Figure 1.13. When the light direction is from azimuth 0° and elevation 30°, the ridge appears as in Figure 1.13. When the light direction is from azimuth 45° and elevation 45°, the ridge appears as in Figure 1.13. In all three cases, the remaining points of the cylinder are hillsides.

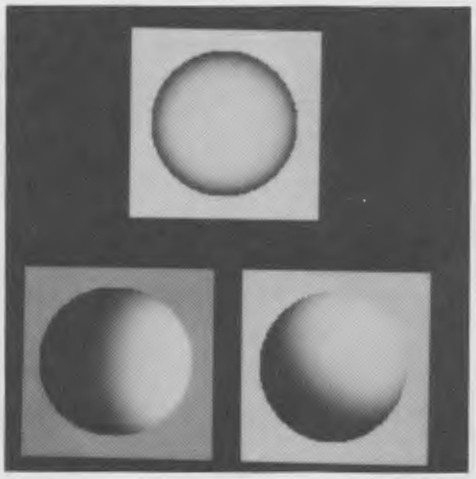


FIGURE 1.12. Shaded images of the sphere of Figure 1.10.

Analytical Results for the Sphere

When the light source is directly above the center of the sphere, the gradient magnitude is zero at (0, 0); therefore, a peak is located at the center of the sphere. The gradient magnitude is positive and the first-directional derivative in the direction w_2 is zero at the remaining points of the sphere. It follows from our analytical results that ridges locate at these points.

When the light direction is $(0, \sqrt{3}/2, -1/2)$, a peak is found at $(0, \sqrt{3}r/2)$. At the remaining points,

$$\nabla I.w_1 = 0 \text{ when } (x^2 + y^2) = \sqrt{3}/2y(r^2 - x^2 - y^2)^{1/2} \text{ and}$$

$$\nabla I.w_2 = 0 \text{ when } x = 0.$$

Therefore, there are ridges when either one of these two equations is satisfied and hillsides otherwise.

Similarly, a peak is found at $(r/2, r/2)$ when the light direction is $(1/2, 1/2, -1/\sqrt{2})$. Ridges can be located at places where either

$$\sqrt{2}(x^2 + y^2) = (x + y)(r^2 - x^2 - y^2)^{1/2} \text{ or}$$

$$x = y \text{ is satisfied.}$$

At the remaining points, hillsides are the correct categories. Figure 1.14 shows the topographic labels for the illuminated spheres.

Experimental Results

Experimentally, we are working in the GIPSY (General Image Processing System) environment. There currently exist GIPSY commands to

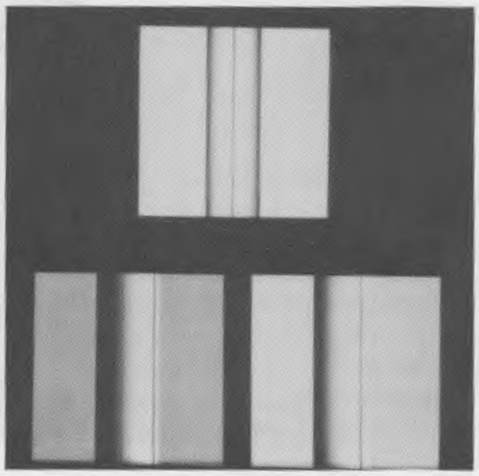


FIGURE 1.13. The analytically derived topographic labeling of the cylinder.

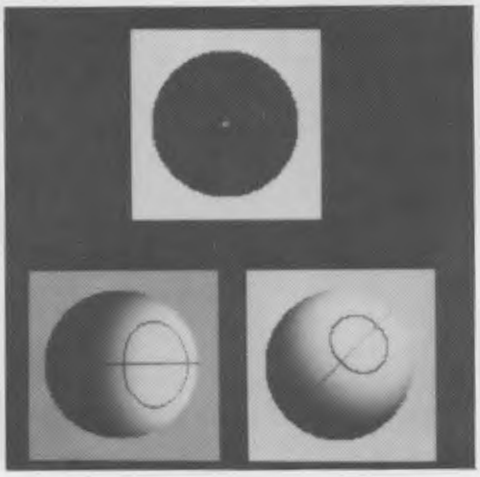


FIGURE 1.14. The analytically derived topographic labeling of the sphere.

construct three-dimensional surfaces, to produce images of these surfaces from various viewpoints and light directions, to fit these images with either cubic polynomials, splines, or discrete cosines, and to calculate the topographic labelings. Figures 1.15 and 1.16 show experimental results for the cylinder and sphere using cubic polynomial surface fitting. Experimental results show very good correspondence with the analytical results, except for the sphere when the light direction is $(0, 0, -1)$. In this case, when points are labeled *ridge* and have neighboring points in a direction orthogonal to the gradient that are also labeled *ridge*, our software reclassifies these ridge continuums as hillsides.

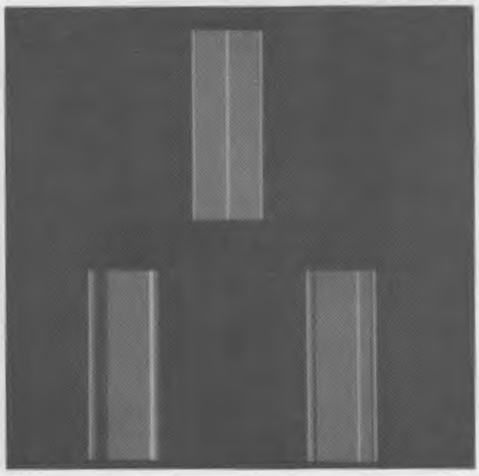


FIGURE 1.15. The experimental results for the cylinder.

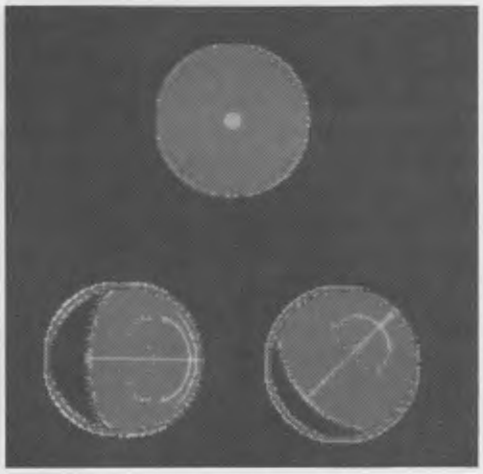


FIGURE 1.16. The experimental results for the sphere.

In addition to the images of the two simple surfaces, a synthetic image of a more complex surface was also used in testing. The surface is composed of cylindrical and spherical surface patches. Figure 1.17 shows the image of the surface when illuminated from azimuth 45° and elevation 45°. Figure 1.18 illustrates the topographic labels that resulted from the experimental method. Most of the resulting topographic labels are located at places where they are predicted by the analytical method.

Our results show that the most informative features found in the images of the cylinder and sphere are ridges and peaks. While the ridges found in the cylinder images are intuitive, the ellipse-like ridges found in the sphere images are unexpected. Although most ridge points found in the sphere images are weak ridges, experimental results show that these ridges are detectable. These ellipse-like ridges will be a definite clue to

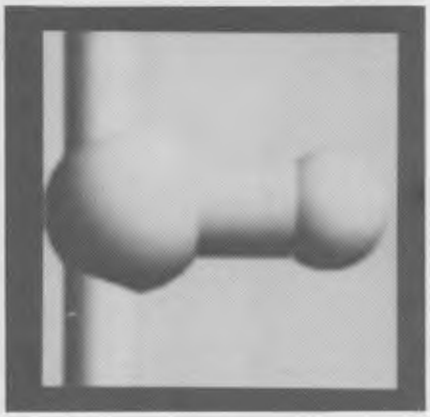


FIGURE 1.17. Gray-tone image of the synthetic object illuminated from azimuth 45° and elevation 45°.

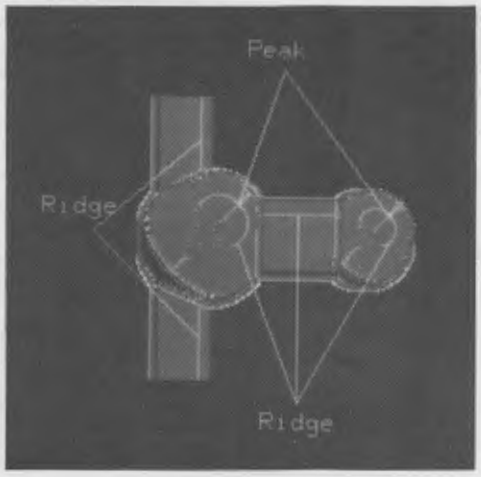


FIGURE 1.18. The topographic labeling of the synthetic object image.

three-dimensional surface identifications. Once the shape of the three-dimensional surface is hypothesized as cylindrical or spherical, information such as the direction of the light source and the cylinder/sphere radius may also be estimated by examining the topographic labels.

SUMMARY

The facet model for image processing estimates the underlying surface of a gray-tone intensity image and uses this estimate in processing the image. The facet model has been used in edge detection, several kinds of segmentation algorithms, and to construct the topographic primal sketch of the image. The topographic primal sketch, a pixel classification scheme, has been used to aid in the estimation of three-dimensional shape from two-dimensional views of shaded objects. In all of these different algorithms, the facet model has been shown to be a robust model for understanding and processing images.

REFERENCES

Haralick, R. M., "Digital step edges from zero crossing of second directional derivatives," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6* (1), Jan. 1984, pp. 58–68.

Haralick, R. M., & L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics, and Image Processing, 29*(1), Jan. 1985, pp. 100–132.

Haralick, R. M., & L. T. Watson, "A facet model for image data," *Computer Graphics and Image Processing, 15,* 1981, pp. 113–129.

Haralick, R. M., L. T. Watson, & T. J. Laffey, "The Topographic Primal Sketch," *International Journal of Robotics Research*, 2 (1), 1983, pp. 50–72.

Horn, B. K. P., "Obtaining shape from shading information," *The Psychology of Computer Vision*, P. H. Winston (Ed.), McGraw-Hill, New York, 1975, pp. 115–155.

Marr, D., "Early processing of visual information," *Philosophical Transactions Royal Society of London, B, 275*, 1976, pp. 483–534.

Pong, T. C., L. G. Shapiro, L. T. Watson, & R. M. Haralick, "Experiments in segmentation using a facet model region grower," *Computer Vision, Graphics, and Image Processing*, Jan. 1984, pp. 1–23.

Pong, T. C., L. G. Shapiro, & R. M. Haralick, "Shape Estimation from Topographic Primal Sketch," *Pattern Recognition, 18*, (5), 1985, pp. 333–348.