Obtaining shape from shading information

Article · August 1989

CITATIONS
413

READS
779

1 author:

Berthold K. P. Horn
Massachusetts Institute of Technology
201 PUBLICATIONS **28,154** CITATIONS

SEE PROFILE

4 OBTAINING SHAPE FROM SHADING INFORMATION

Berthold Horn

4.1 INTRODUCTION

4.1.1 Shading as a Monocular Depth Cue

An image of a smooth object known to have a uniform surface will exhibit gradations of reflected light intensity which can be used to determine its shape. This is not obvious since at each point in the image we know only the reflectivity at the corresponding object point. For some points (called singular points here) the reflectivity does uniquely determine the local normal, but for almost all points it does not. Consequently, the shape of the surface cannot be found by local operations alone.

For many surfaces the fraction of the incident light which is scattered in a given direction is a smooth function of the angles involved. It is convenient to think of the situation as depending on the three angles shown in Fig. 4.1: the incident angle between the local normal and the incident ray, the emittance angle between the local normal and the emitted ray, and the phase angle between the incident and the emitted rays.

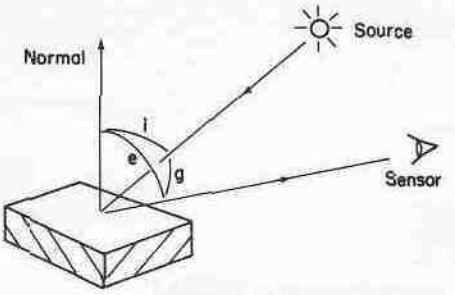

Fig. 4.1 Definition of the incident (i), emittance (e), and phase angle (g).

We will show that the shape can be obtained from the shading if we know the reflectivity function and the position of the light sources. The reflectivity and the gradient of the surface are related by a nonlinear first-order partial differential equation in two unknowns. The recipe for solving this equation involves setting up an equivalent set of five ordinary differential equations, three for the coordinates and two for the components of the gradient. These five equations can then be integrated numerically along certain curved paths. For while we cannot determine the gradient locally, we can, roughly speaking, determine its component in one special direction. Then taking a small step in this direction, we can repeat the process. The curve traced out on the object in this manner is called a characteristic. Figure 4.2 shows the characteristics determined for an experimental sphere. Their projections on the image plane will be referred to as the base characteristics. The shape of the visible surface of the object is thus given as a sequence of coordinates on characteristics along its surface.

Figures 4.3 and 4.4 show stereo pairs for three test cases. Figure 4.5 gives contour maps for the same three objects.

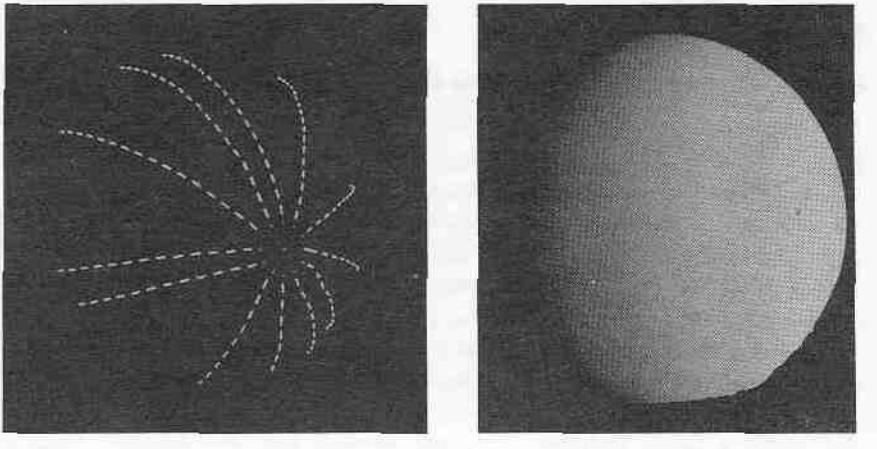

Fig. 4.2 Image of a sphere and a stereo pair of the characteristic curves obtained from the shading.

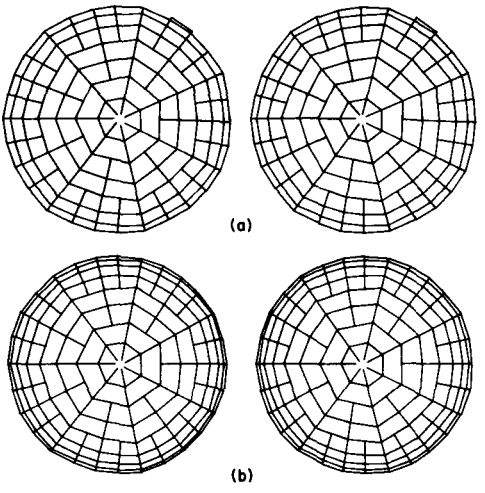

Fig. 4.3 Stereo pairs of solutions for dish-shaped, spherical, and bullet-shaped objects.

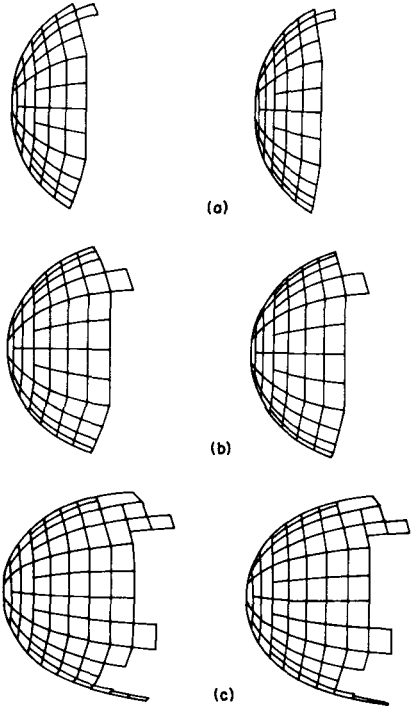

Fig. 4.4 Stereo pairs of same solutions as in previous figures (rotated 90°).

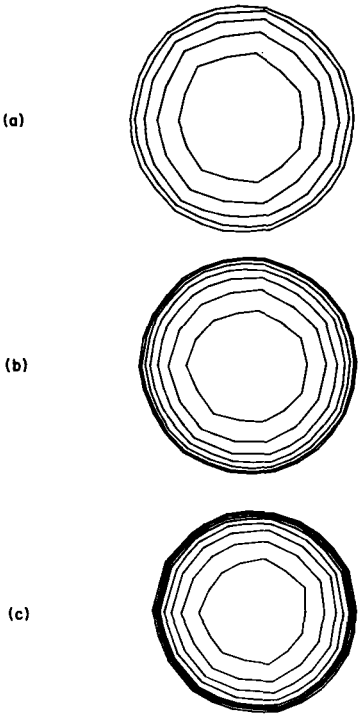

Fig. 4.5 Contour maps of same solutions as in previous figures.

An initial known curve on the object is needed to start the solution. Such a curve can usually be constructed near the singular points mentioned earlier using the known local normal. The only additional information needed is the distance to the singular point and whether the surface is convex or concave with respect to the observer at this point—such ambiguities arise in several other instances in the process of solution as will be seen.

To solve the equations, the reflectivity as a function of the three angles must be known as well as the geometry relating light source, object, and observer. Multiple or extended light sources increase the complexity of the solution algorithm presented. But all of this initially needed information can be deduced from the image if a calibration object of known shape is present in the same image. Furthermore, incorrect assumptions about the reflectivity function and the position of the light sources can lead to inconsistencies in the solution and it may be possible to utilize this information in the absence of a calibration object.

In practice it is found that if the object is at all complex, its image will be segmented by edges. Some of these represent the occlusion of one surface by another while others are angular edges (also called joints here) on a single object. Another kind of edge is the ambiguity edge. This is an edge which the characteristics cannot cross, indicating an ambiguity which cannot be resolved

locally. One can solve inside each region bounded by these various edges, but some global or external knowledge is needed to match up the regions. In the case of an angular edge on an object one can integrate up to the edge and then use the known location of the edge as an initial curve for another region.

A very similar situation obtains when one bridges a shadow. Since one edge of the shadow and the position of the light source is known, we can trace along the rays grazing the edge until the corresponding image points fall on an illuminated region. Since we know the path of each ray, we can calculate the coordinates of the point where it impinges on the object by triangulation. The edge of the shadow (which need not be on the same object) can then serve as an initial curve from which to continue the solution.

4.1.2 Applications

A number of interesting applications of this method can be mentioned. The first of these concerns the scanning electron microscope which produces images which are particularly easy to interpret because the intensity recorded is a function of the slope of the object at that point and is thus a form of shading. In optical and transmission electron microscopes, intensities depend instead on thickness and optical or electron density. The geometry of the scanning electron microscope allows several simplifications in the algorithm for determining shape from shading. Because of the random access capability of the microscope beam, it should be easy and useful to combine it with a small computer to obtain three-dimensional information directly.

Another interesting demonstration lies in the determination of lunar topography. Here the special reflectivity function of the material in the maria of the moon allows a very great simplification of the equations used in the shape-from-shading algorithm. The equations in fact reduce to one integral which must be evaluated along each of a family of predetermined straight lines in the image.

So far we have assumed that the surface is uniform in its photometric properties. Any nonuniformity will cause this algorithm to determine an incorrect shape. This is one of the uses of facial makeup, because by darkening certain slopes those slopes can be made to appear steeper. In some cases surface-markings can be detected if they lead to discontinuities of the calculated shape.

Judging by our wide use of monocular pictures (photographs or even paintings and woodcuts) of people and other smooth objects, humans are good at interpreting shading information. The shortcomings of our method which are related to the shading information available can be expected to be found in human visual perception too. It will of course be difficult to decide whether the visual system actually determines the shape quantitatively or whether it uses the shading information in a very qualitative way only.

4.1.3 History of the Problem

The literature on perception has only a few conjectures on the possibility of determining shape from the monocular depth-cue of shading. One relevant paper is on lunar topography[1] which gives complete details of a solution obtained in the form of an integral in the special case of the reflectivity function of the moon. It can be shown that the result is in fact a special case of the general solution derived here.

Various defects of image-dissector sensing devices affect the accuracy of shape measurements. Since very little was known about the characteristics of this device on other than theoretical grounds, a program was developed to measure properties such as resolution, signal-to-noise ratio, drift, settling time, scatter, and pinholes in the photocathode.[2] Software now compensates for some defects such as geometric distortion and nonuniform sensitivity using measurements from test patterns.

In parallel with the programming work, theoretical efforts were made to define and get around some of the difficulties of the method. Of particular interest were applications where the equations simplify greatly. Unfortunately the massive simplification found in the case of lunar topography seems unique.

4.2 THEORY

4.2.1 Reflectivity Functions

Consider a surface element of size dS inclined i with respect to the incident ray and e with respect to the emitted ray. The angles are measured with respect to the normal as was shown in Fig. 4.1. Let the incident light intensity be I_1 per unit area perpendicular to the incident ray. The amount of light falling on the surface element is then $I_1 \cos(i)\, dS$.

Let the emitted ray have intensity I_2 per unit solid angle per unit area perpendicular to the emitted ray. Therefore the amount of light intercepted by an area subtending a solid angle dw at the surface element will be $I_2 \cos(e)\, dS\, dw$. The reflectivity function $\phi(i, e, g)$ is then defined to be I_2/I_1.

If we want to be more precise about what units the intensity is measured in, we have to take into account the spectral distribution of the light emitted by the source, as well as the spectral sensitivity of the sensor. With this proviso we can speak of watts per unit area and watts per unit solid angle per unit area. We need not be too concerned with this if we either use white paint, or measure the reflectivity function with the same equipment later used in the shape-from-shading algorithm. It should be noted that for most surfaces the reflectivity function is dependent on the color of the light used. Typically the specular component of the reflected light, being reflected

before it has penetrated far into the surface, will be unchanged, while the mat component will be colored by pigments in the surface coating.

Several other definitions of the reflectivity function are in use which are multiples of the one defined here by π, 2, cos(e) and/or cos(i). The specific formulation chosen here makes the equation relating the incident light intensity to the image illumination very simple.

Surfaces where the three parameters i, e, and g are not sufficient to fully determine the reflectivity are unsuitable for this analysis. Examples are translucent objects and those with nonisotropic surface properties like hair and the mineral commonly called tigereye.

Perhaps the most important determinant of the reflectivity function is the microstructure of the surface. Different reflectivity functions may apply at different magnifications. At high magnification many objects become increasingly translucent. It is best therefore to determine the reflectivity function under conditions similar to those later used in the determination of the shape of the object.

One way to measure the reflectivity function is to employ a goniophotometer fitted with a small flat sample of the surface to be investigated. The device can be set for any combination of incident, emittance and phase angles.

To avoid having to move the source and the sensor into all possible positions with respect to a flat sample of the surface when measuring the reflectivity function, it is convenient to have a test-object which presents all possible values of i and e for a given g. The constraints are $i + e \leqslant g$, $e + g \leqslant i$ and $g + i \leqslant e$. Use of such an object is greatly simplified by using a telephoto lens and a distant source, giving almost constant g. It is convenient to tabulate the reflectivity versus i and e for each of a series of values of g. A sphere is the easiest test-object to use if one is willing to live with the decreasing accuracy in determining e as one approaches the edge.

One could also have an object of known shape in the same scene as the object to be analyzed. This solves the problem of having to know the source location and the transfer properties of the image forming system. In some cases objects of known shape and surface characteristics differing from those of the object under study are useful—for example a sphere with specular reflectivity can pinpoint the location of the light sources.

Alternately, one might hope to predict reflectivity functions on a theoretical basis starting with some assumed microstructure of the surface. White mat surfaces are usually finely divided grains of transparent material such as snow or crushed glass. White paint consists of transparent 'pigment' particles (e.g., SiO_2 or TiO_2) of high refractive index and small size (optimally about the wavelength of visible light) suspended in a transparent medium of low refractive index. If one chooses to model a regular arrangement of suspended particles of uniform size and makes restrictive assumptions, one can derive a reflectivity function and study its dependence on various parameters.

Another type of surface is that of a highly reflective material (such as a metal) where the light rays do not penetrate into the material. Choosing a particular type of surface depression and a statistical distribution of the size of these, one can again derive a reflectivity function.

Only a few such models have been studied and little hope exists for modelling real surfaces well enough without having to abandon closed expressions for the reflectivity function.

4.2.2 The Differential Equation of Image Illumination

This section contains the derivation of the image illumination equation and the analytical formulation of the shape-from-shading problem.

At a known point on the object we can calculate g. We should like to find the gradient (or at least its component in one direction) at this point so as to be able to continue the solution to a neighboring point. Measurement of the light reflected tells us something about i and e. Since only one measurement is involved, we cannot generally hope to determine both i and e locally, but only a relation between the two. There are exceptional points where the normal is locally fully determined and this is useful in finding initial conditions as explained later.

Suppose we collapse the two principal planes of the image-forming system together, forming the x-y plane as shown in Fig. 4.6. Let the z-axis coincide with the optical axis and extend toward the object. Let f be image-plane distance from the exit pupil and assume that the image and object space refractive indexes are equal.

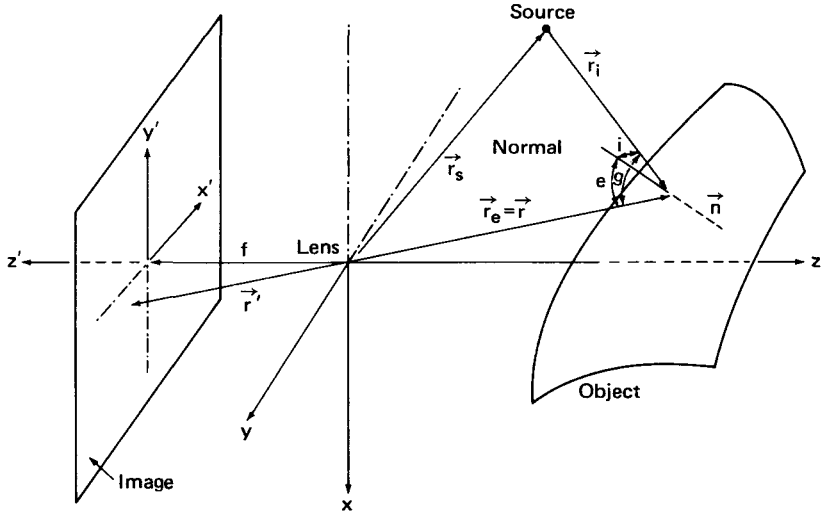

Fig. 4.6 Details of the geometry of image illumination and projection in the imaging system.

Let t be the ratio of image illumination to object luminescence.

Let a(x, y, z) be the incident light intensity (usually constant or obeys some inverse square law).

Let $A(x, y, z) = t \cdot a(x, y, z)$.

Let $\mathbf{r} = (x, y, z)$ be a point on the object and $\mathbf{r}' = (x', y', f)$ the corresponding point in the image.

Let $b(x', y')$ be the intensity measured at the image point (x', y').

Let p and q be the partial derivatives of z with respect to x and y.

Let $I = \cos(i)$, $E = \cos(e)$ and $G = \cos(g)$.

We have $A(\mathbf{r})\, \phi(I, E, G) = b(\mathbf{r}')$.

We would like to show that this equation must be a first-order partial differential equation. This will be true if it contains only x, y, z, and the first partial derivatives p and q. We emphasize that this image illumination equation is the main equation studied here.

When finding a solution we assume $A(\mathbf{r})$ and $\phi(I, E, G)$ are known and $b(\mathbf{r}')$ is obtained from the image. We want to show that the equation is a first-order, nonlinear partial differential equation in two independent variables of the form:

$$F(x, y, z, p, q) = 0$$

From the simple projection geometry we have

$$\mathbf{r}' = \left(\frac{f}{z}\right) \cdot \mathbf{r}$$

where f is the image plane distance from the exit pupil. We took care of image reversal by orienting the x' and y' axes appropriately. It remains to show that I, E and G are functions of x, y, z, p and q. An inward normal to the surface at the point $\mathbf{r}$ is $\mathbf{n} = (-p, -q, 1)$.

Let the light source be at $\mathbf{r}_s = (x_s, y_s, z_s)$. Then the incident ray will be $\mathbf{r}_i = \mathbf{r} - \mathbf{r}_s$, and the emergent ray $-\mathbf{r}_e = -\mathbf{r}$. Clearly then

$$I = \hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_i, \quad E = \hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_e \quad \text{and} \quad G = \hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_e$$

where the $\hat{}$'s denote unit vectors. All the terms thus involve only x, y, z, p and q. On substituting into the image illumination equation, it follows that we are dealing with a first-order nonlinear partial differential equation in the two unknowns x and y:

$$F(x, y, z, p, q) = A(\mathbf{r})\ \phi(I, E, G) - b(\mathbf{r}') = 0$$

4.2.3 Equivalent Ordinary Differential Equations

The usual method of dealing with a first-order nonlinear partial differential equation is to solve an equivalent set of five ordinary differential equations:

$$\dot{x} = F_p, \quad \dot{y} = F_q, \quad \dot{z} = pF_p + qF_q$$

$$\dot{p} = -F_x - pF_z \quad \text{and} \quad \dot{q} = -F_y - qF_z$$

The dot denotes differentiation with respect to s, a parameter which varies with distance along a so-called characteristic strip. The subscripts denote partial derivatives. These equations are solved along the characteristic strips. See, for example, Garabedian.[3] The characteristic strips are the characteristic curves described earlier (values of x, y, and z) plus the values of p and q on them.

It can be shown that these equations are equivalent to the image illumination equation. The demonstration is tedious and is omitted here. Interested readers may find the mathematics in books on partial differential equations.

Since we can multiply the equation $F = 0$ by any nonzero smooth function $\lambda(x, y, z, p, q)$ without altering the solution surface, we can obtain a different set of equations:

$$\dot{x} = \lambda F_p, \quad \dot{y} = \lambda F_q, \quad \dot{z} = \lambda(pF_p + qF_q)$$

$$\dot{p} = \lambda(-F_x - pF_z) \quad \text{and} \quad \dot{q} = \lambda(-F_q - qF_z)$$

The solution to this new set of equations will differ only in the values of the parameter s at any given point. For example if we let

$$\lambda = \frac{1}{\sqrt{F_p^2 + F_q^2 + (pF_p + qF_q)^2}}$$

the parameter s gives us arc-length along the characteristics. This is used in the programs to be described later. Of course we can only do this if the denominator is not zero.

At singular points and ambiguity edges the denominator will be zero since $F_p = F_q = 0$. A different choice for λ will be used later in the discussion of the scanning electron microscope.

4.2.4 Simplifying Situations

Since the general equations are fairly complex it is of great interest to find simplifying conditions. Some of these are presented in this section. But first we will need some partial derivatives, which it is convenient to introduce here.

The development of these partial derivatives requires some further simple notation. Let A be a vector (3-tuple) and $A = |A|$ be the magnitude of A. Also let $\hat{A} = A/|A|$ be the corresponding unit vector. Consider the dot-product $A \cdot B$ to be matrix multiplication of the 1 by 3 matrix A by the 3×1 matrix B^T (the transpose of B). Consider partial differentiation with respect to a vector to be the 3-tuple whose components are found by differentiating with respect to each component in turn. We denote this differentiation by a subscript. Then for example

$$A_A = \hat{A}$$

At times we will also need the partial derivatives of vectors with respect to vectors. These are defined as 3×3 matrices, the first row being the result of differentiating with respect to the first component and so on. Then for example

$$A_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We will also use partial derivatives of dot-produces of unit vectors with respect to vectors. For example

$$X = \hat{A} \cdot \hat{B} \quad \text{and we want } X_A$$

To avoid finding $\hat{A}_A$ we write $A X = A \cdot \hat{B}$ and then

$$A_A X + A X_A = A_A \cdot \hat{B}$$

Extending the definition of dot-product in the appropriate way we find

$$A_A \cdot \hat{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \hat{B}^T = \hat{B}$$

$$A X_A = \hat{B} - X \hat{A}$$

$$X_A = \left(\frac{1}{A}\right)(\hat{B} - X \hat{A})$$

Given these results, we get the important relations, i.e.,

$$I_r = I_{r_i} = \left(\frac{1}{r_i}\right)(\hat{n} - I \hat{r}_i)$$

$$I_n = \left(\frac{1}{n}\right)(\hat{r}_i - I \hat{n})$$

$$E_r = E_{r_e} = \left(\frac{1}{r_e}\right)(\hat{n} - E \hat{r}_e)$$

$$E_n = \left(\frac{1}{n}\right)(\hat{r}_e - E \hat{n})$$

$$G_r = G_{r_i} + G_{r_e} = \left(\frac{1}{r_e}\right)(\hat{r}_i - G \hat{r}_e) + \left(\frac{1}{r_i}\right)(\hat{r}_e - G \hat{r}_i)$$

$$G_n = 0$$

We will now proceed to list some simplifying situations:

1. *Distant source*: Collimated source or the object subtends a small angle at the source.

   $\mathbf{A_r} \cdot \mathbf{r_i} = 0$ and for a truly distant source:

   $$\mathbf{A_r} = 0$$

   Replace $\mathbf{r_i}$ by $k\mathbf{r_i}$ and let $k \to \infty$. Then

   $\mathbf{I_r} = 0$, $\mathbf{I_n}$ unchanged

   $\mathbf{E_r}$ and $\mathbf{E_n}$ unchanged

   $$\mathbf{G_r} = \left(\frac{1}{r_e}\right)(\hat{\mathbf{f}}_i - G\,\hat{\mathbf{f}}_e),\ \mathbf{G_n} = 0$$

   In addition choosing the z-axis along $\mathbf{r_i}$ removes further terms.

2. *Distant camera*: Telephoto lens or the object subtends a small angle at the camera.

   Replace $\mathbf{r_e}$ by $k\mathbf{r_e}$ and let $k \to \infty$. Then

   $\mathbf{I_r}$ and $\mathbf{I_n}$ unchanged

   $\mathbf{E_r} = 0$, $\mathbf{E_n}$ unchanged

   $$\mathbf{G_r} = \left(\frac{1}{r_i}\right)(\hat{\mathbf{f}}_e - G\,\hat{\mathbf{f}}_i),\ \mathbf{G_n} = 0$$

   In addition choosing the z-axis along $\mathbf{r_e}$ removes further terms.

3. *Distant source and distant camera*:

   $\mathbf{I_r} = 0$, $\mathbf{I_n}$ unchanged

   $\mathbf{E_r} = 0$, $\mathbf{E_n}$ unchanged

   $\mathbf{G_r} = 0$, $\mathbf{G_n} = 0$

   Most practical situations are an approximation of this case.

4. *Source at the camera*:

   $\mathbf{r_i} = \mathbf{r_e}$ $\quad I = E$ and $G = 1$

   $\mathbf{I_r} = \mathbf{E_r}$ unchanged

   $\mathbf{I_n} = \mathbf{E_n}$ unchanged

   $\mathbf{G_r} = 0$ and $\mathbf{G_n} = 0$

5. *Distant source at distant camera*:

   $\mathbf{I_r} = \mathbf{E_r} = \mathbf{G_r} = 0$

   $\mathbf{I_n} = \mathbf{E_n}$ unchanged, $\mathbf{G_n} = 0$

Choosing the object to be on the z-axis removes further terms. This is the simplest possible case.

6. *Uniform illumination*: Uniform illumination (or an approximation thereof) is fairly common and might at first sight appear not to fit into our framework. It can be shown however that uniform illumination is equivalent to a point source at the camera and an altered reflectivity function.

4.2.5 The Five Differential Equations of Shading

We will now make further use of the notation and results of the last section. Recall the image illumination equation:

$$F(x, y, z, p, q) = A(\mathbf{r})\ \phi(I, E, G) - b(\mathbf{r}') = 0$$

We know $A(\mathbf{r})$ and $\phi(I, E, G)$, and obtain $b(\mathbf{r}')$ from the image. We need F_x, F_y, F_z, F_p and F_q. Since $\mathbf{r} = (x, y, z)$ and $\mathbf{n} = (-p, -q, 1)$ we can get all of these derivatives from $F_\mathbf{r}$ and $F_\mathbf{n}$.

$$F_\mathbf{r} = A(\mathbf{r})\ \phi_\mathbf{r}(I, E, G) + A_\mathbf{r}(\mathbf{r})\ \phi(I, E, G) - b_\mathbf{r}(\mathbf{r}')$$

$$F_\mathbf{n} = A(\mathbf{r})\ \phi_\mathbf{n}(I, E, G)$$

Let $\mathbf{a} = (I, E, G)$. Then

$$\phi_\mathbf{r} = \phi_\mathbf{a}\, \mathbf{a}_\mathbf{r} \quad \text{and} \quad \phi_\mathbf{n} = \phi_\mathbf{a}\, \mathbf{a}_\mathbf{n}$$

Note that $\mathbf{a}_\mathbf{r}$ and $\mathbf{a}_\mathbf{n}$ are 3×3 matrices, the rows of which we computed in the previous section.

$$\mathbf{a}_\mathbf{r} = \begin{bmatrix} \left(\frac{1}{r_i}\right)(\hat{\mathbf{n}} - I\,\hat{\mathbf{r}}_i) \\ \left(\frac{1}{r_e}\right)(\hat{\mathbf{n}} - E\,\hat{\mathbf{r}}_e) \\ \left(\frac{1}{r_e}\right)(\hat{\mathbf{r}}_i - G\,\hat{\mathbf{r}}_e) + \left(\frac{1}{r_i}\right)(\hat{\mathbf{r}}_e - G\,\hat{\mathbf{r}}_i) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{r_i} & -\frac{I}{r_i} & 0 \\ \frac{1}{r_e} & 0 & -\frac{E}{r_e} \\ 0 & \left(\frac{1}{r_e} - \frac{G}{r_i}\right) & \left(\frac{1}{r_i} - \frac{G}{r_e}\right) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{n}} \\ \hat{\mathbf{r}}_i \\ \hat{\mathbf{r}}_e \end{bmatrix}$$

Note that this is the product of two 3×3 matrices. Similarly

$$a_n = \begin{bmatrix} \left(\frac{1}{n}\right)(\hat{r}_i - I\,\hat{n}) \\ \left(\frac{1}{n}\right)(\hat{r}_e - E\,\hat{n}) \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{I}{n} & \frac{1}{n} & 0 \\ -\frac{E}{n} & 0 & \frac{1}{n} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{n} \\ \hat{r}_i \\ \hat{r}_e \end{bmatrix}$$

To evaluate the derivative F_r we need $b_r(r')$.

$$b_r(r') = b_{r'} \cdot r'_r$$

Written out in full we have

$$(b_x, b_y, b_z) = \left(\frac{f}{z}\right)\left\{b_{x'}, b_{y'}, -\left[\left(\frac{x}{z}\right)b_{x'} + \left(\frac{y}{z}\right)b_{y'}\right]\right\}$$

where $b_{x'}$ and $b_{y'}$ are measured directly from the image.

Since the intensities measured from the image do not locally determine the normal, one might well ask what, roughly, such measurements do determine. The components of the gradient of the intensity are related to the second derivatives of the distance to the surface, while the intensity itself is related to the magnitude of the first derivatives. This relationship becomes exact for the case of a distant source at a distant camera.

It should be noted that the equation for F_r also involves A_r. Usually A is fairly constant over the area of the object recorded in the image, or at least satisfies a simple inverse-square equation.

If $A = (r_1/r_i)^2$, then $A_r = -2(r_1^2/r_i^4)r_i$

where r_i is the incident vector, and r_1 is the length of the incident vector to the singular point.

4.2.6 Initial Conditions and Singular Points

To select a particular solution surface among all possible solution surfaces one needs to specify an initial curve through which the solution surface must pass:

$$x = x(t), \quad y = y(t) \quad \text{and} \quad z = z(t)$$

Along this curve we must satisfy

$$z'(t) = p\,x'(t) + q\,y'(t)$$

$$F[x(t),\ y(t),\ z(t),\ p(t),\ q(t)] = 0$$

Here the dash represents differentiation with respect to t. This pair of nonlinear equations allows one to find p(t) and q(t) along the initial curve. There may be more than one solution, in which case there will be more than one solution surface. The characteristic strips sprout from the initial curve as for example in Fig. 4.7. The solution surface can be described parametrically:

$$x = x(s, t),\ y = y(s, t),\ z = z(s, t), \text{ and}$$

$$p = p(s, t),\ q = q(s, t)$$

Now it would be a great disadvantage if one always required an initial curve to start the solution from. Fortunately it is usually possible to calculate some initial curve if one makes some assumptions about the surface and uses the special singular points where the reflectivity uniquely determines the local normal.

The singular points are the brightest or the darkest points (depending on the reflectivity function). At all other points the normal cannot be locally determined. The singular points correspond to values of i and e for which the reflectivity is a unique global maximum or minimum. These may be either extrema or at the limiting values of the angles.

This method cannot be used if the surface does not contain a surface element oriented in this special direction. The points are found by looking for the brightest (or darkest) points in the image.

All we still need to know then is the distance of this point from the camera, but since one is usually only interested in relative distances this is not a serious restriction.

Unfortunately it will be found that the solution will not move from these singular points because $\dot{x} = \dot{y} = 0$. This is an indication that the algorithm needs to be informed about which way the surface is curved, convex or concave.

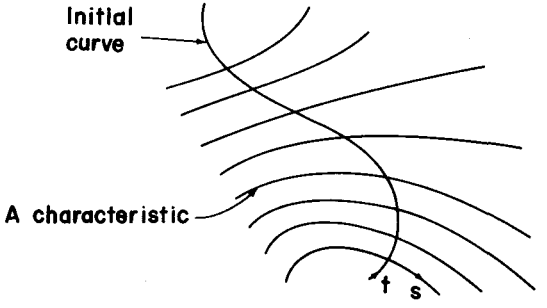

Fig. 4.7 Characteristic strips sprouting from an initial curve.

If the surface is convex (or concave) at the singular point and we have a guess at the radius of curvature (from the overall size of the object for example), we can get around the problem of singular points by constructing small spherical caps on them. Difficulties will be encountered if this point happens to be a saddle point, but the presence of a saddle point usually indicates that other singular points exist where the surface is either convex or concave.

Consider Fig. 4.8. Let **S** be the vector from the camera to the singular point (found from its known image coordinates and its distance from the camera). R is the estimated radius of curvature and ρ the distance we decide to step away from the singular point, determined in practice by considerations of uncertainty in the position of the singular point and the desired detail in the solution. The known normal at the singular point is $\hat{\mathbf{N}}_0$. We construct a spherical cap with center $\mathbf{S} - R\hat{\mathbf{N}}_0$.

Let

$$R_1^2 = R^2 - r^2$$

$$\mathbf{S}_1 = \mathbf{S} + (R_1 - R)\hat{\mathbf{N}}_0$$

$$\mathbf{X} = \hat{\mathbf{y}} \times \hat{\mathbf{N}}_0 \text{ where } \hat{\mathbf{y}} = (0, 1, 0)$$

$$\mathbf{Y} = \hat{\mathbf{N}}_0 \times \hat{\mathbf{X}}$$

$$\mathbf{T}(t) = \rho(\hat{\mathbf{X}} \cos(2\pi t) + \hat{\mathbf{Y}} \sin(2\pi t)) \qquad 0 \leqslant t < 1$$

Points on the initial circle are then given by

$$\mathbf{S}_1 + \mathbf{T}(t)$$

We also need an initial guess at p and q, so we construct $\mathbf{N}_1$, (an outward normal):

$$\mathbf{N}_1(t) = R_1\hat{\mathbf{N}}_0 + \mathbf{T}(t)$$

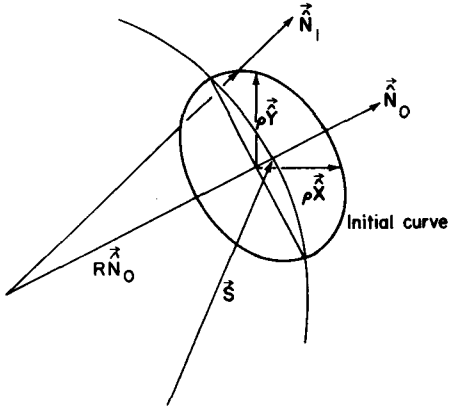

Fig. 4.8 Construction of the initial curve near a singular point.

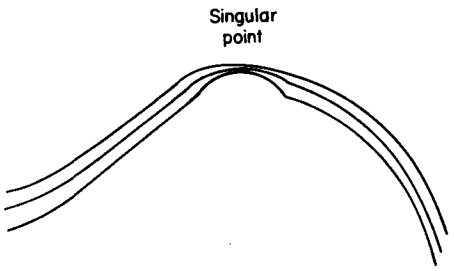

Fig. 4.9 Three solutions obtained for varying initial radius of curvature and the small effect which errors in the initial curve have on the solution.

The requirement for an initial guess at the radius of curvature is not as restrictive as it might seem, since the required accuracy is extremely low. This is because ρ is usually very much smaller than R, and hence a change in R affects the position of the initial curve very little. Even more importantly, the values derived for p and q need not be accurate since they are only used as a first guess in an iterative method of finding p and q on the initial curve before starting the solution. Figure 4.9 shows graphically how different radii might influence the solution in a typical case.

4.2.7 Nonpoint Sources

Uniform sources have already been dealt with. Perhaps the easiest other case is a circularly symmetric source at a distance large compared to the dimensions of the object.

Distant circularly symmetric sources can be replaced by a point source after modifying the reflectivity function. One merely convolves the reflectivity function with the spread function of the source (a bit of spherical trigonometry is involved). Strictly speaking, one should perform the same operation with the entrance pupil of the camera since it too subtends a finite angle at the object and accepts a bundle of light rays. Since ϕ is smooth (except at $I = 0$ and $I = 1$) it will be changed very little except at these points. The main change will be that ϕ does not tend to 0 as I tends to 0, but rather for some negative value of I. Also the specular component will be more smeared out.

Let the source intensity be I(a) per unit solid angle at the angle a from its center when viewed from the object. See Fig. 4.10. Then the new reflectivity function $\phi'(I, E, G)$ is

$$\phi'(I, E, G) = \frac{\int_0^{2\pi}\int_0^{a_0} I(a)\,\phi(I', E, G')\,a\,da\,dv}{\int_0^{2\pi}\int_0^{a_0} I(a)\,a\,da\,dv}$$

where a_0 = total angular diameter of the source

$\cos(A) = (\cos(g) - \cos(i)\cos(e))/(\sin(i)\sin(e))$

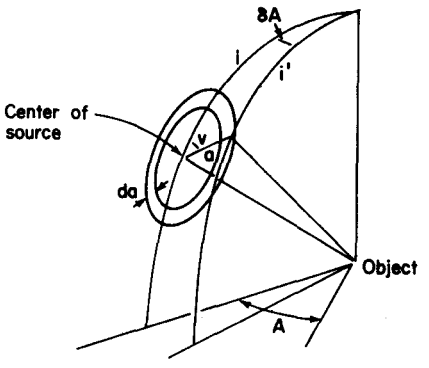

Fig. 4.10 Circularly symmetric source and quantities used in the convolution.

$$\cos(i') = \cos(i)\cos(a) + \sin(i)\sin(a)\cos(v)$$
$$\sin(\delta A) = \sin(i')\sin(a)/\sin(v)$$
$$\cos(g') = \cos(A + \delta A)\sin(i')\sin(e) + \cos(i')\cos(e)$$

When the source distribution is not easily treated as above one can introduce a different A_k for each source and replace the main equation by

$$\sum_k A_k(\mathbf{r})\ \phi(I_k, E, G_k) = b(\mathbf{r}')$$

Difficulties in finding initial conditions will be encountered with multiple sources unless they are of special kinds (e.g., a point source plus a uniform source).

4.2.8 Shadows and Other Edges

Several kinds of edges appear in an image, each with its own properties and problems for our algorithm:

1. Overlap—special case of occlusion of one object by another in which the line of occlusion corresponds to an angular edge on the occluding object. There is a discontinuity in z. The program must detect this or it will erroneously continue a solution across such an edge.
2. View edges—special case of occlusion where no angular edge is involved. The surface is smooth and E tends to 0 as we approach it. This is easily detected by the program during the calculation of the solution.
3. Joints—angular edges on an object. There are discontinuities in the derivatives of z. One cannot continue p and q across such an edge. It is possible however to use the position of the edge as a new initial curve. This and the previous condition can be detected as a step in the intensity distribution or from a highlight on the edge.

4. Shadow edges–here I tends to 0 as we approach the edge and again the program can easily detect this.
5. Projected shadow edges–if the shadow is bridged this edge may serve as a new initial curve as described below.
6. Ambiguity edges–some are lines of aggregation of singular points (on which $\lambda \to \infty$). The characteristics will not cross an ambiguity edge.

If the single source is not at the camera, shadows will appear. Solutions can be carried across shadows since the position of the source is known and one can construct a ray through the last illuminated point and trace it until it meets another illuminated region. The place where a glancing ray first strikes the surface on the other side of the shadow can be determined by triangulation on the source-surface ray and the surface-eye ray. Only the coordinates and not the local gradient of this new point will be known. It is necessary to carry this operation out for all characteristics entering the shadow, producing a new initial curve at the other edge of the shadow where we can restart the solution. In practice care has to be taken because of noise.

4.3 SPECIAL APPLICATIONS

4.3.1 The Scanning Electron Microscope

This section deals with two applications in which the equations simplify considerably. The first is scanning electron microscopy.

The scanning electron microscope device uses an electron beam which is focused and deflected much like the beam of a cathode ray tube and impinges on a specimen in an evacuated chamber. As shown in Fig. 4.11 the narrow ray penetrates into the specimen for some small distance, creating secondary

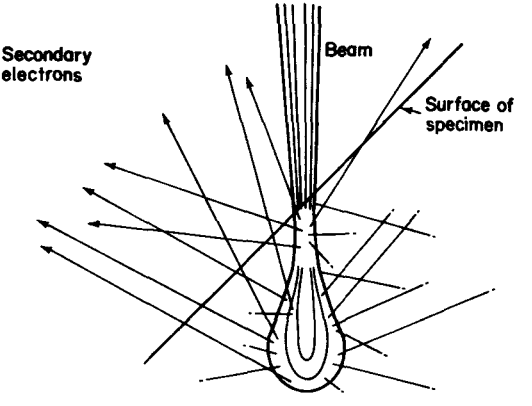

Fig. 4.11 Detail of electron beam impinging on specimen.

electrons along its path. The depth of penetration, the spread, and the number of secondary electrons are all functions of the material of that portion of the specimen. The number of secondary electrons which emerge from the specimen back into the vacuum through the surface will depend strongly on the inclination of the surface with respect to the beam.

These relatively slow secondary electrons are then attracted by a positively charged grid and impinge on a phosphor-coated photomultiplier. (See Fig. 4.12.) In this way a current is generated proportional to the number of secondary electrons escaping the specimen. The emerging electrons are analogous to photons reflected at a surface but in direct contrast to optical surfaces the intensity is least when the incident beam is perpendicular. Thus the steep edges are outlined more brightly. Strangely, people find this effect appealing.

The output is used to modulate the intensity of the beam in a cathode ray tube while both beams are scanned synchronously in a television-like raster. The image created exhibits shading and is remarkably easy to interpret topographically. This is quite unlike the normal use of optical or transmission electron microscopes which portray density and thickness.

The magnification is easily increased by decreasing the deflection in the microscope. The resolution is poor compared to the transmission electron microscope because of the spread of the beam as it enters the specimen, but the depth of field is much better than that of an optical microscope because of the very narrow beam (extremely high f-number).

There are important cases where the shape must be determined and stereoscopic methods are not applicable. This may be because at the magnification used the specimen appears smooth without significant surface detail or because it is difficult to line up the second image. Since the

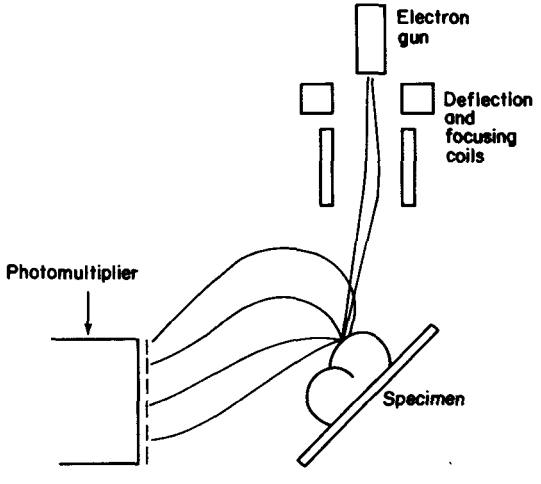

Fig. 4.12 A scanning electron microscope.

equations for this case turn out to be so simple, it should be rewarding to tie a scanning electron microscope directly into a small computer.

A little thought shows that this electron microscope situation is analogous to the case where a source is at the camera (or equivalently, the case where we have uniform illumination). Note that no shadows appear. Moreover the projection is ordinarily near-orthogonal. Because of these two effects the five ordinary differential equations simplify considerably:

$$\dot{x} = F_p,\ \dot{y} = F_q,\ \dot{z} = pF_p + qF_q$$

$$\dot{p} = -F_x - pF_z \quad \text{and} \quad \dot{q} = -F_y - qF_z$$

Now

$$F_n = A\,\phi_I I_n \quad \text{and} \quad F_r = -b_r$$

$$I = \frac{n \cdot \hat{z}}{n} = \frac{1}{n} \quad \text{where } n = (-p, -q, 1)$$

$$I_n = \left(\frac{1}{n}\right)(\hat{z} - I\,\hat{n}) = \left(\frac{\hat{z}}{n}\right) - \left(\frac{1}{n^3}\right)\hat{n}$$

$$I_p = \left(\frac{1}{n^3}\right)p \quad \text{and} \quad I_q = \left(\frac{1}{n^3}\right)q$$

Hence

$$\dot{x} = F_p = \left(\frac{A\,\phi_I}{n^3}\right)p,\ \dot{y} = F_q = \left(\frac{A\,\phi_I}{n^3}\right)q$$

$$\dot{z} = \left(\frac{A\,\phi_I}{n^3}\right)(p^2 + q^2)$$

$$\dot{p} = -b_x \quad \text{and} \quad \dot{q} = -b_y$$

If $\phi_I \neq 0$ everywhere, we can change to a new measure s along the characteristic by multiplying all equations by $\lambda = n^3/(A\,\phi_I)$ and we get

$$\dot{x} = p,\ \dot{y} = q,\ \dot{z} = p^2 + q^2$$

$$\dot{p} = b_x \frac{n^3}{A\,\phi_I},\ q = b_y \frac{n^3}{A\,\phi_I}$$

Notice that the changes in x and y along the characteristics are given in this case by the partial derivatives of z with respect to x and y. This constrains the characteristics therefore to grow in the direction of the surface's gradient at each point. Thus this extremely simple case has characteristics which are curves of steepest descent. Also note that the

equation for z does not couple back into the system of equations because of the orthogonal projection. This increases accuracy. The equations happen to be very similar to the Eikonel equations for the paths of light rays in refractive media. It may be possible to find ready-made solutions to some special cases by using this analogy.

We assumed that $\phi_I \neq 0$; this is equivalent to assuming that an inverse exists which allows us to find I from a measurement of the image intensity:

$$\psi[\phi(I, I, 1)] = I$$

Let

$$\xi(x) = \frac{1 - \psi^2(x)}{2\psi(x)}$$

Then

$$\xi[\phi(I, I, 1)] = \left(\frac{1}{2}\right)(p^2 + q^2)$$

So we can find at each point the magnitude, but not the direction of the local gradient.

Let us turn to the question of ambiguities since the subject is easy enough in the electron microscope case. Assume the camera and light source are at the same position and consider the two surfaces:

$$z = z + x^3, \quad z = z + |x|^3$$

Clearly they cannot be distinguished in monocular views since they produce identical intensity distributions in the image. This manifests itself in a slowing down of the characteristics as they approach the line $x = 0$ (alternatively $\lambda \to \infty$). They cannot cross this line aggregation of singular points. Note that the characteristics approach this line at right angles and that the edge is determined locally, since in general each point on an ambiguity edge is a singular point.

A second kind of ambiguity edge can occur parallel to characteristics, separating those which can be reached from one singular point from those

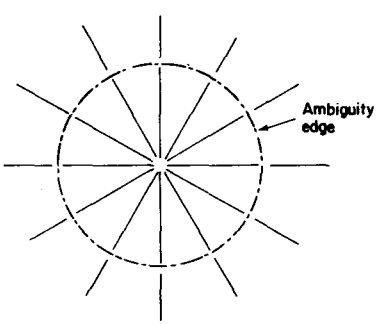

Fig. 4.13 A locally determined ambiguity edge: $f = 1/(x^2 + y^2 - 1)^2$.

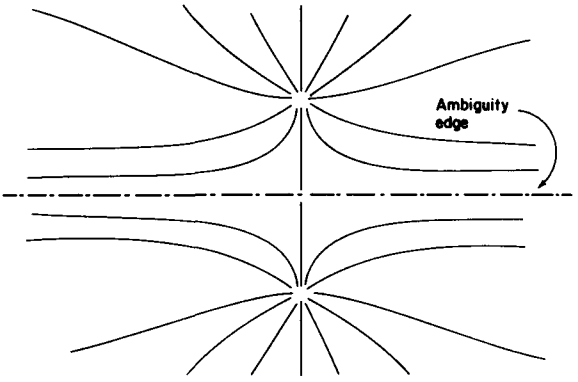

Fig. 4.14 A globally determined ambiguity edge: $f = 1/[1 + x^2 + (y-1)^2] + 1/[1 + x^2 + (y+1)^2]$.

reachable only from another. This kind of edge is not locally determined, since a change in the surface is possible which removes one of the singular points and makes all the characteristics accessible from the other. This can be done without altering an area near the two given points previously separated by an ambiguity edge. Figure 4.13 shows a locally defined ambiguity edge and Fig. 4.14 shows a globally defined one.

Both types of ambiguity edge occur in the general case but are not so easily studied there. They divide the image into regions within each of which a solution can be obtained. Typically most such regions will have one singular point from which one may obtain initial conditions (provided one makes a decision about whether the surface is concave or convex and knows the distance to the singular point).

4.3.2 Lunar Topography

The other very interesting simplification to the general shape from shading equations occurs when we introduce the special reflectivity function which applies to the material in the maria of the moon. This in fact was the first shape from shading problem solved both theoretically and in an operating algorithm.[1] Using the special reflectivity function and the fact that the sun is a distant source, it is possible (but very tedious) to show that the equations simplify so that the base characteristics (i.e., the projection of the characteristics on the image plane) become straight lines radiating from the zero-phase point, the point corresponding to $g = 0$. The camera lies directly between this point and the sun. Actually this is true only when the sun is located at negative z; for positive z (i.e., in front of the camera), the relevant point is the π-phase point, directly in the sun.

The variation of light reflected from the surface of the moon with phase and inclination of the surface has been studied for a long time. At a given lunar phase g, all possible combinations of incident angle i and emittance

angle e are represented by some portion of the surface. A fairly good approximation is the Lommel-Seeliger formula:[4]

$$\phi(I, E, G) = \frac{\Gamma_0(I/E)}{(I/E) + \lambda(G)}$$

Where Γ_0 is a constant and the function $\lambda(G)$ is defined numerically by a table. This formula can also be derived from a simplified model of the lunar surface. A slight gain in accuracy is possible if Γ_0 is allowed to vary with G as well. In particular Fesenkov finds the more accurate formula

$$\phi(I, E, G) = \frac{\Gamma_0(I/E)[1 + \cos^2(\alpha/2)]}{(I/E) + \lambda_0[1 + \tan^2(\alpha/2)]}$$

where

$$\tan(\alpha) = \frac{G - (I/E)}{\sqrt{1 - G^2}}$$

A recent theoretical model is that of Hapke[5] which corresponds fairly closely to the measured reflectivity function. In most of these formulas we find that for a given G, ϕ is a constant for constant I/E. The lines of constant I/E are meridians.

At full moon, when $G = 1$ we find that the whole face has constant luminosity. This is quite unlike the effect on a sphere coated with a typical mat paint where the image intensity would vary as

$$\sqrt{1 - \left(\frac{r}{R}\right)^2}$$

Where R is the radius of the image and r the distance from the center of the image. The full moon thus has the same appearance as a flat disc if one is used to objects with normal mat surfaces. This may explain the flat appearance of the full moon.

In the case of pictures taken of the lunar surface from nearby (e.g., from orbit) we have the following:

1. Distant source (the moon subtends an angle of about .03 milliradians at the sun).
2. Near point source (the sun subtends an angle of about 10 milliradians at the moon).
3. Camera at the origin.
4. The reflectivity function is constant for constant I/E.

It can be demonstrated that the solution to the lunar topography problem is a special case of the more general formulation given in this paper. But since the details are tedious, we only note that the ordinary differential equation that constrains r has the simple solution:

$$\frac{r(s)}{r(0)} = e^{-P\int_0^s [\tan(\alpha)/Q^2]\, ds}$$

where

$$L = \frac{x_0}{z_0}\cos(t) + \frac{y_0}{z_0}\sin(t)$$

and

$$Q = \sqrt{s^2 + 2sL + \left(\frac{r_0}{z_0}\right)^2} \qquad P = \mathrm{sgn}(z_0)\sqrt{\left(\frac{r_0}{z_0}\right)^2 - L^2}$$

Note that r(0) is the distance to the point from where the integration was started, t is a parameter which varies from characteristic to characteristic, s is a parameter that varies along a given characteristic and $\hat{f}_0 = (x_0, y_0, z_0)$ is a unit vector parallel to the direction from the sun to the moon.

To sum up, as one advances along each characteristic in turn, one calculates G, measures b/A and uses ψ to obtain $\tan(\alpha)$, which is then used in the evaluation of the above integral. Here $\psi(b/A, G) = I/E$, that is to say, ψ is a kind of inverse function for $\phi(I, E, G)$. The process is much simpler than the general shape from shading algorithm.

Let us list some of the major points of interest.

1. The base characteristics are predetermined straight lines (independent of the image). This makes for high accuracy and ease in planning a picture-taking mission.
2. Only a single integral needs to be evaluated, not five differential equations.
3. The primary input is the intensity, not its gradients, again making for high accuracy.
4. Although, as usual, the reflected light-intensity does not give a unique normal, it does determine the slope component in the direction of the characteristic. J. van Diggelen[6] first noted a special case of this when he solved the lunar topography problem for the special case of an area near the terminator (line separating sunlit from dark areas). The characteristics are such that the slope along them can be determined locally. The slope at right angles to the characteristics cannot be determined locally.
5. Although T. Rindfleisch did not mention it in his paper[1] it is very easy to bridge shadows since each light ray lies in a sun-camera-characteristic plane. Its image can thus be traced on the base characteristic until we again meet a lighted area. One need not even make special provisions for this, but just use $\tan(\alpha)$ for grazing incidence (intensity = 0) in the shaded section.

4.4 IMPLEMENTATION

4.4.1 Measuring the Reflectivity Function

The reflectivity functions of some paints were measured using large rubber spheres as calibration objects. Both camera and source were moved as far away as possible to achieve almost constant phase angle g. The image of a convex object is especially useful because it contains two points for all possible combinations of the incident and emittance angles (i and e) for a given phase (angle g). The position of the light source is measured, as well as the distance from the front of the sphere to the entrance pupil. The image dissector is focused on the edge of the sphere.

With the sphere temporarily illuminated from several sources, a program finds its exact position and size, as well as the difference in horizontal and vertical deflection sensitivity of the image dissector. It is now possible to calculate the points in the image which correspond to given incident and emittance angles. For a number of choices of both of these angles one then reads the intensity at a small raster of points near these positions and averages them to reduce noise and the effect of pinholes in the photocathode. Since there are usually two places in the image with the same incident and emittance angle, a check on the data is possible. The resultant table of values (usually normalized with respect to the brightest intensity) can be printed and the whole process repeated after moving the light source to a new position for a new phase angle. The program accounts for such things as change in incident light intensity as the light source gets moved around.

Clearly the points for given incident angle lie on a circle on the surface of the sphere. Similarly for points with a given emittance angle. These two circles may intersect in two, one, or no points. One can find this intersection by first finding the line along which the planes containing these circles intersect. Examine Fig. 4.15. Applying the sine and cosine laws,

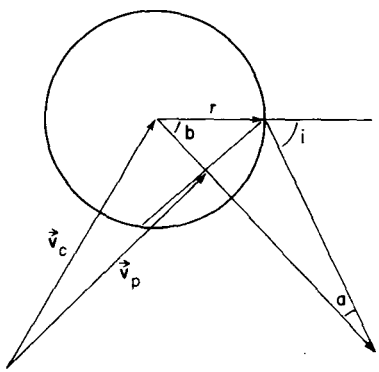

Fig. 4.15 Finding points for given incident and emittance angles.

$$I = \cos(i) \quad \text{as usual and} \quad D = |v_s|$$

$$\frac{D}{\sin(\pi - i)} = \frac{r}{\sin(a)} \qquad b = i - a$$

$$r\cos(b) = r\,[\cos(i)\cos(a) + \sin(i)\sin(a)]$$

$$= \left(\frac{r}{D}\right)[I\sqrt{D^2 - r^2(1 - I^2)} + r(1 - I^2)]$$

$$d = r\cos(b)$$

$$v_p = v_c + d\,\hat{v}_s$$

The equation of the plane in which the circle of points with given incident angle i lies is

$$v \cdot \hat{v}_s = v_p \cdot \hat{v}_s \qquad \text{where } v = (x, y, z)$$

One can find a similar equation for the plane in which the circle of points with given emittance angle e lies. The introduction of an arbitrary third plane allows us to find one point v on the intersection of the first two. The line of intersection of the first two planes must be parallel to the cross product of their normals (let them be v_{s1} and v_{s2}). So the equation of the line we are looking for is

$$(v - v_a) = k\,v_1 \qquad \text{where } v_1 = v_{s1} \times v_{s2}$$

The points we are trying to find must also lie on the sphere, that is,

$$(v - v_c)^2 = r^2$$

$$(v_a + k\,v_1 - v_c) = r^2$$

$$k^2 v_1 \cdot v_1 + 2\,k\,v_1 \cdot (v_a - v_c) + (v_a - v_c)^2 - r^2 = 0$$

The above equation may have no solution for k, in which case no point exists for the given incident and emittance angle. Otherwise we can use the two solutions and substitute back to obtain the desired coordinates which are then transformed into image coordinates. Figure 4.16 shows the sort of table that results for a given phase angle g.

The first paint investigated was a mat white paint consisting of particles of SiO_2 and TiO_2 suspended in a transparent base. Very roughly one finds that the reflectivity function behaves like cos(i) for a given g. After playing with polynomial fits for a while, the following fairly accurate formula was found by a process of little interest here:

		I →									
		1.00	0.97	0.93	0.87	0.78	0.68	0.56	0.43	0.29	0.15
	1.00					0.77					
	0.97				0.87	0.78	0.66				
	0.93			0.93	0.88	0.78	0.67	0.57			
	0.87		0.97	9.94	0.89	0.79	0.67	0.57	0.45		
E	0.78	0.99	0.98	0.95	0.90	0.81	0.68	0.59	0.46	0.32	
↓	0.68		0.98	0.95	0.91	0.82	0.71	0.59	0.47	0.33	0.18
	0.56			0.94	0.90	0.83	0.74	0.61	0.48	0.34	0.17
	0.43				0.88	0.79	0.74	0.62	0.50	0.34	0.18
	0.29					0.79	0.70	0.58	0.42	0.30	0.15
	0.15						0.65	0.50	0.38	0.26	0.13

Fig. 4.16 Table of reflectivity (for a white mat paint) vs. $I = \cos i$ and $E = \cos e$ for $G = \cos g = 0.81$. The intervals chosen correspond to constant size steps in the angles. Note the blank areas for combinations of angles which cannot form a spherical triangle.

$$\phi(I, E, G) = \frac{(1 + G)(2 + G)}{6}\left[I + \frac{1 + 2IEG - (I^2 + E^2 + G^2)}{16(1 - G)}\right]$$

For reasonable angles the above formula is about 5 percent accurate, becoming worse for extreme angles. The repeatability of this measurement was disappointingly low, depending on the depth of the paint coat and the details of its application. Much of the investigation of the behavior of the image dissector was the result of efforts to trace the remaining causes of inaccuracy.

Some other paints and an eggshell showed a mat component similar to the above, plus a very strong specular component (which is small except near the point for which $i = e$ and $i + e = g$). This component is very sensitive to small changes in the surface properties such as can be brought about by handling the object.

The image of a convex object with such a surface will usually have two local maxima in intensity. One of these will be broad (corresponding to the mat component), the other narrow and bright (corresponding to the specular component). These may be distinguished by a computer program on the basis of just these properties. It would then be possible to start a solution from the mat maximum (which is not a global maximum) rather than the specular maximum. This might be a good idea because of the increased accuracy (for one thing the normalization of image intensities would be more accurate).

In an attempt to track down poor results in the first try at finding reflectivity functions accurately, the image dissector was investigated in some detail[2]. Among problems found were:

1. Unequal deflection sensitivity in horizontal and vertical directions (differed by 12 percent).

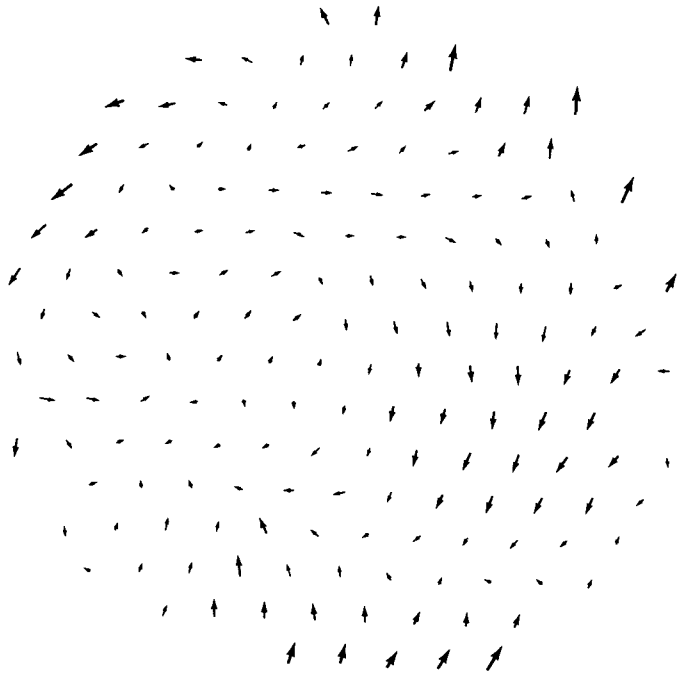

Fig. 4.17 Geometric distortion in image dissector for a triangular raster of points covering the photo-cathode. (The arrows are exaggerated three times.)

2. Twist of image varying with distance from center of field of view. (See Fig. 4.17.)
3. Poor resolution (3 line-pairs/mm — radius of tube 38 mm).
4. Pinholes in the photocathode (about 20 of up to 0.5 mm in size).
5. Nonuniform sensitivity of the photocathode (varies more than 30 percent).
6. Fairly long settling time of the deflection coils (on the order of 300 microseconds).
7. A large amount of scatter, which reduces the contrast by almost one-third and causes intensities measured on the image of a uniform square on a dark background to vary by 20 percent, depending on how close to the edge the measurement is taken.

4.4.2 Numerical Methods for Solving the Equations

The five ordinary differential equations were at first solved using a well-known Runge-Kutta method. The idea is to average together several estimates of the derivatives of the five variables (x, y, z, p, and q) with respect to the

parameter s. The first estimates are the actual derivatives at s. These are used to take a half step forward and calculate new values for x , y , z, p , and q as though derivatives higher than the first were zero. We then calculate the derivatives at this new point to get a second set. Next we start over from s, probing out again by a half step but now using the second set of derivatives. We then get a third set of derivatives at $s + h/2$. The third set is used in the final probe from s which now extends fully to $s + h$ where we get a fourth set of derivatives. The official full step is taken using a weighted average of the four sets of derivatives found in this way. Written out in symbols this becomes:

Let h be the step-size (for the parameter s).
Let $\mathbf{Y} = (x, y, z, p, q)$.
And let the equations for the derivatives be

$$\mathbf{Y}' = \mathrm{F}(s, \mathbf{Y})$$

(In our case, F is actually independent of s.)
Denote $\mathbf{Y}(s_n)$ by $\mathbf{Y}_n$ then at the n^{th} step

$$\mathbf{K}_1 = h\,\mathrm{F}(s_n, \mathbf{Y}_n)$$

$$\mathbf{K}_2 = h\,\mathrm{F}\left(s_n + \frac{h}{2}, \mathbf{Y}_n + \frac{\mathbf{K}_1}{2}\right)$$

$$\mathbf{K}_3 = h\,\mathrm{F}\left(s_n + \frac{h}{2}, \mathbf{Y}_n + \frac{\mathbf{K}_2}{2}\right)$$

$$\mathbf{K}_4 = h\,\mathrm{F}(s_n + h, \mathbf{Y}_n + \mathbf{K}_3)$$

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \left(\frac{1}{6}\right)(\mathbf{K}_1 + 2\mathbf{K}_2 + 2\mathbf{K}_3 + \mathbf{K}_4)$$

This method is easy to start (requires no previous values of **Y**) and stable, but requires four time-consuming evaluations of the derivatives per step. For this reason various predictor-modifier-corrector methods were tried and the simplest was found to give adequate accuracy

$$\mathbf{P}_{n+1} = \mathbf{Y}_n + 2\,h\,\mathrm{F}(s_n, \mathbf{Y}_n)$$

$$\mathbf{M}_{n+1} = \mathbf{P}_{n+1} - \left(\frac{4}{5}\right)(\mathbf{P}_n - \mathbf{C}_n)$$

$$\mathbf{C}_{n+1} = \mathbf{Y}_n + \left(\frac{h}{2}\right)[\mathrm{F}(s_n, \mathbf{M}_{n+1}) + \mathrm{F}(s_n, \mathbf{Y}_n)]$$

$$\mathbf{Y}_{n+1} = \mathbf{C}_{n+1} + \left(\frac{1}{5}\right)(\mathbf{P}_{n+1} - \mathbf{C}_{n+1})$$

P, **M**, and **C** are the predictor, modifier and corrector respectively. This method is stable and requires only two derivative evaluations per step, but is

not self-starting. The Runge-Kutta method was retained for the first step in the integration. Stability and accuracy were not serious concerns since the noise in the data input contributes far more to errors in the solution.

Under optimal conditions (using methods to cancel out most of the distortion and nonuniformity of photocathode sensitivity) the program was allowed to scan a sphere of 100 mm radius. A sphere was then fitted by an iterative least-square method to the data points found. The data points nowhere deviated from the fitted sphere by more than 10 mm, and by less than 5 mm except near the very edge of the image. Such accuracy will not usually be obtained because of nonuniformity in the paint, shortcomings of the sensing device, etc. For many purposes, however, less accuracy is quite acceptable and for object recognition in particular a more important criterion is that similar objects are distorted in similar ways.

4.4.3 A Program Solving the Characteristics in Parallel

It soon became apparent that integrating along each characteristic independently has many disadvantages in the general case, even though it works well for lunar topography. The first reason is that characteristics spreading out from the singular point begin to separate and leave large portions of the image unexplored. One obtains only a very uneven sampling of the surface of the object. With a more parallel approach new characteristics can be created as one goes along and some others can be deleted in areas where characteristics approach each other too closely.

Next we find that the base characteristics (projections of the characteristics onto the image) may sometimes cross! This would not be possible if the solution were exact, since it indicates that the surface is double-valued or at least that its gradient is double-valued. Solution of this problem is easy if the integrations are carried along in parallel, but involves lengthy comparison tests otherwise.

Once it had been demonstrated that the equations were correct and a numerical solution possible, it was decided to write a program which would explore the surface of the object by moving along all the characteristics in parallel and by interpolating new characteristics when needed. Accuracy in the solution was traded for more noise immunity. The solution is achieved by taking all characteristics one step forward at the same time. An effort to find a convenient coordinate system for this approach produced the notation and resultant equations given here. The solution was previously worked out in a different coordinate system requiring manipulation of extreme complexity.

The values stored for each point (x, y, z, intensity, p, q and pointers to the previous point on the same characteristic) are here arranged not by characteristic but by "ring." A ring is a curve of constant arc-distance from the singular point. That is, the n^{th} points on all the characteristics form one

ring. The complete data structure is made up of a number of rings, the first of which is the initial curve. As before, individual characteristics may stop for a variety of reasons and this causes breaks to appear in the current ring. Some rings thus represent closed curves and others more distant from the singular point are broken into sections, the final ring having no active point on it.

As we have seen, one of the main inducements for using the parallel solution method is to allow interpolation of new characteristics. This is one of the reasons why the number of points in a ring may change from one to the next and why each point has to have a pointer into the previous ring indicating which element is its predecessor in the same characteristic.

It should be noted that the use of constant size steps along the characteristics may produce difficulties on complex objects. For even with smooth surfaces the curves of constant arc-distance from the singular point may have cusps. An alternative, which would circumvent this problem, would be the use of steps traversing a constant increment in intensity. This would turn the rings into contours of constant intensity.

We have already described how one can obtain p(t) and q(t) on the initial curve by solving the set of nonlinear equations:

$$p(t)\, x_t(t) + q(t)\, y_t(t) - z_t(t) = 0$$

$$A(r)\, \phi(I, E, G) - b(r') = 0$$

When solving a difference equation approximation from noisy data we can expect the solution for p and q to become progressively more inaccurate. Yet the above pair of equations must hold on any path along the surface of the object. In particular one can use them on the curve defined by one ring to determine values of p and q.

For the initial curve we had the additional difficulty that the two equations might have more than one solution and we selected one on the basis of some external knowledge (e.g., that the object is convex near the singular point). We have assumed that the object is smooth and therefore we will have fairly good values for p and q and cannot get into this difficulty at nonsingular points. Even a simple Newton-Raphsen method will suffice to get us more accurate values of p and q.

Let

$$g(p, q) = p\, x_t + q\, y_t - z_t$$

$$h(p, q) = \phi(I, E, G) - \frac{b}{A}$$

and suppose: $g(p + \delta p, q + \delta q) = h(p + \delta p, q + \delta q) = 0$.

Then ignoring other than first-order terms we have

$$\begin{pmatrix} g_p & g_q \\ h_p & h_q \end{pmatrix} \begin{pmatrix} \delta p \\ \delta q \end{pmatrix} = \begin{pmatrix} g(p, q) \\ h(p, q) \end{pmatrix}$$

That is,

$$\begin{pmatrix} x_t & y_t \\ p & q \end{pmatrix} \begin{pmatrix} \delta p \\ \delta q \end{pmatrix} = \begin{pmatrix} g(p, q) \\ h(p, q) \end{pmatrix}$$

Here x_t, y_t and z_t have to be estimated from difference approximations. One may not want to apply the full correction $(\delta p, \delta q)$. More than one iteration will not be required since after the first iteration p and q are very close to the correct values. We will call this process the sharpening of p and q.

When the separation between two neighboring points in a ring becomes greater than 1.5 times the step size along the characteristic, a new characteristic is interpolated. Its x, y, z, p, and q values are set to the average of its neighbors. A more complicated interpolation method can also be used which constructs the line of intersection of the tangent planes at the two neighboring points. It then finds the point on this line closest to the two neighbors and finally uses a point half-way between the point determined previously by the simpler method and this new point. This method does not, however, add significantly to the accuracy of the solution.

If two neighboring points in a section of a ring come closer than 0.7 times the step-size, one is deleted (it is important that this factor be less than 0.75, that is, one half of the factor used in the interpolation decision, or succesive rings on a flat region will have points interpolated on one step, only to be removed on the next, with consequent loss of accuracy).

Finally one wants to stop neighboring characteristics from crossing over each other. Consider the two points a and b on one ring and their successors c and d on the next as in Fig. 4.18. The test consists of checking whether c is to the left of the directed line through bd and whether d is to the right of the directed line through ac. Both tests are needed. If either fails, the corresponding characteristic is terminated, causing a break to appear in the ring at that point. The test is equivalent to checking whether the line segment cd falls in front of the line segment ab (and does not cross it). This test is applied across short breaks in rings as well to stop neighboring sections of the ring from crossing over each other.

Care has to be taken if the remaining sections of a ring all fall on one side of the singular point, since the break then actually encompasses an arc of

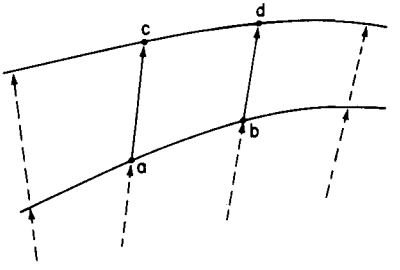

Fig. 4.18 The four points used in the crossing test.

more than π and crossing tests applied across it will invariable terminate more characteristics on either side of it. This can be avoided if the crossing test is not applied to points whose images fall too far apart in terms of the projection of the current step size.

Rather than use the intensities at a small raster of points to estimate the local gradient, it was decided to use a difference approximation from intensities measured at neighboring points. Using as many as possible of the intensities of the point itself and its five immediate neighbors, we can apply a simple least-squares method to estimate the gradient. (See Fig. 4.19.) Some of the points may not exist as explained previously and the characteristic is terminated if less than three points are available or only three which are nearly collinear. Suppose the coordinates of the points are (x'_k, y'_k) (image coordinate system) and the intensities are b_k. We wish to find b_0, $b_{x'}$ and $b_{y'}$, to minimize the following expression:

$$\sum_k (b_{x'} x'_k + b_{y'} y'_k + b_0 - b_k)^2$$

This happens when

$$\begin{bmatrix} \sum x'^2_k & \sum x'_k y'_k & \sum x'_k \\ \sum x'_k y'_k & \sum y'^2_k & \sum y'_k \\ \sum x'_k & \sum y'_k & \sum 1 \end{bmatrix} \begin{bmatrix} b_{x'} \\ b_{y'} \\ b_0 \end{bmatrix} = \begin{bmatrix} \sum b_k x'_k \\ \sum b_k y'_k \\ \sum b_k \end{bmatrix}$$

From $b_{x'}$ and $b_{y'}$ we can find b_x, b_y, and b_z by using the camera projection equations of an earlier section.

For good noise immunity and some ability to detect surface detail indicating that the solution is invalid, the intensity for each solution point is not read from only one image point. Small tilted rectangular rasters of points are established around each point of the solution as shown in Fig. 4.20. One axis of the rectangle is parallel to the base characteristic at that point, and the size is adjusted to correspond to the projection on the image of a square on the object of side-length equal to the step size. The intensity recorded for a solution point is the average of the intensities read for the points in this raster and the rms/average is used to make the edge-crossing decision. The rasters of all the points in the data structure almost but not quite touch and taken together almost cover the total area of the image explored. This insures that the data is not much affected by pinholes in the photocathode of the image

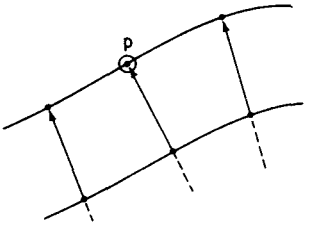

Fig. 4.19 The five neighbors used in determining the intensity gradient at p.

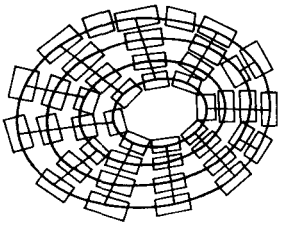

Fig. 4.20 Covering the image with the rasters of points read for each solution point.

dissector and that edge crossing can easily be detected without reducing the resolution.

This program spends more than half its time accessing the image dissector. Between 20 and 100 intensities are read for each point in the solution, and each access takes about 0.2 to 1.0 milliseconds. A complete solution requires from 1 to 5 minutes of real time.

4.4.4 Operation of the Program

First the program needs to be given such parameters as the position of the light source, the distance to the object, focal length of the lens and the step size to be used in the integration. It then proceeds to find a point of maximum intensity (for some reflectivity functions one needs to search for a minimum). This search can be directed to allow a choice of one of several possible maxima. The program then assumes that this point of maximum intensity is a singular point and that the object is convex at this point (in some cases we would like to assume it to be concave). After constructing an initial curve (a small circle) around the singular point, it proceeds to read the intensities at the corresponding image points. The nonlinear equations for p and q on this curve are than solved iteratively.

All intensities are normalized with respect to the intensity at the singular point unless the surface has a specular component. In the latter case, the intensities on the initial curve are used to establish a normalization value (the specular reflectivity is too variable for use in normalization). It is assumed that the initial curve has been chosen large enough to fall outside the region of strong specular reflection.

For each step in the parameter s, the following procedure is then carried out:

1. For each point calculate the normal (**n**), the incident vector ($\mathbf{r}_i$) and the emittance vector ($\mathbf{r}_e$). From these obtain the derivatives $I_{\mathbf{n}}$, $E_{\mathbf{n}}$ and $G_{\mathbf{n}}$.

2. Calculate ϕ_I, ϕ_E, ϕ_G and hence ϕ_n.
3. Then obtain F_p, F_q and λ.
4. Add $(\delta x, \delta y, \delta z)$ to (x, y, z) to get the point on the next ring for each characteristic. Here $(\delta x, \delta y, \delta z) = \lambda(F_p, F_q, pF_p + qF_q)$.
5. Interpolate new points where the points in the new ring are too far apart and delete points where they are too close together. Produce breaks where characteristics have crossed over adjacent characteristics.
6. Now read the intensities for all the points. Terminate those characteristics with points of very low intensity or high rms/average.
7. Calculate $b_{x'}$, $b_{y'}$ for all those points for which enough neighbors exist. From these values obtain b_x, b_y and b_z by the projection equations.
8. Now use n, r_i and r_e to calculate I_r, E_r and G_r.
9. Next use ϕ_I, ϕ_E and ϕ_G to calculate ϕ_r.
10. Then obtain F_x, F_y and F_z.
11. Add $(\delta p, \delta q)$ to (p, q) to obtain p and q for the uninterpolated points on the new ring. Here $(\delta p, \delta q) = [\lambda(-F_x - pF_z), \lambda(-F_y - qF_z)]$.
12. Interpolate p and q for the new points.
13. Sharpen up the values for p and q on all points in the new ring.
14. Garbage-collect various items.

The simpler Euler method for solving the differential equations could be replaced by a Runge-Kutta method with increases in running time of a factor of two, but little improvement in accuracy. Distortions in the imaging device produce distortions in x and y, while nonuniformities in the sensitivity will affect p and q and hence z. The only effect of low resolution will be that some edges will not be noticed and the solution erroneously continued across them.

It should be apparent where the various tests for terminating the characteristics fit into the above schema. The program terminates characteristics in the following situations:

1. The characteristic has moved out of the field of view of the image dissector.
2. The rms/average for the intensities read in the raster has become too great, indicating overlap of two objects or an angular joint on one object or some surface detail that is being missed.
3. The intensity has become too low, indicating a shadow region.
4. λ is too large, indicating approach to either another singular point or an ambiguity edge.
5. There are too few neighbors to construct a good estimate of the local intensity gradient.
6. I too small—indicating approach to a shadow edge.

7. E too small—indicating approach to an edge of the object.
8. The characteristic crossed over a neighboring one.
9. The intensity is equal to or greater than that measured at the singular point, indicating another singular point or ambiguity edge.

Note that several of these conditions are redundant to ensure that even with an inexact solution at least one will fail at the right place.

4.4.5 Summary and Conclusions

After defining the reflectivity function, an equation was found relating the intensity measured in the image of a smooth opaque object to the shape of the object. This equation was then shown to be a first-order nonlinear partial differential equation in two unknowns and the equivalent set of five ordinary differential equations was derived. Two especially simple cases were discussed, namely applications to lunar topography and the scanning electron microscope. Methods were described for obtaining the auxiliary information required (e.g., the reflectivity function) and how to avoid the need for an initial known curve on the object. Of importance too is the method demonstrated for continuously updating p and q (sharpening) as the solution progresses.

The analytical approach to the problem of determining shape from shading was developed to demonstrate that an exact solution is possible and to determine just what the limitations of this approach are. This is not to say that a more heuristic, approximate approach does not have its merits too for certain types of objects. It was decided to produce a program to allow experimentation with the solution method because many ideas in the field of artificial intelligence and visual perception are of little value until they can be tried on real data. Fortunately an image dissector was available to provide input of image intensities to the computer.

4.5 RELATED QUESTIONS

4.5.1 Likely Source Distributions

Since the complexity of the algorithms presented here increase with the complexity of the light source distribution and since we only know how to bridge shadows cast by one source, it is important to know which light source distributions occur in practice. First one notes that the situations found difficult by humans are almost certainly going to give difficulties to our algorithm. For example, when two sources cast shadows (such as on a road lighted by widely spaced streetlamps) the shape of unfamiliar objects becomes difficult to ascertain because of the crossed shadows. If the incident intensity varies greatly from one image area to another (such as in a lightly wooded forest) the tangle of lighted and dark areas makes perception more difficult.

On the other hand one would expect natural conditions to be particularly easy, as in the case of one point source somewhat above the observer (the sun) combined with a very diffuse (almost uniform) source (the sky). The diffuse source will not throw sharp shadows of its own. The absence of either of the two sources makes vision only slightly more difficult.

One would expect photographers to have something to contribute to this subject and introductory booklets on artificial light photography confirm the above conclusions. The beginner is advised to use a number of lights with different characteristics as follows:

1. The main light—The ideal main light is a large spot light approximating the effect of the sun. It is usually placed 45 degrees above and 45 degrees to the side of the subject. Its purpose is to establish the 'form of the subject' and fix the ratio of lighted to dark areas. The exact ratio is not important but the position of the source should result in good shading (which increases as the source is moved further from the camera) without too much shadow area (in which detail is more difficult to perceive).
2. The fill-in light (or axial light)—Its purpose is to lighten slightly the shadows cast by the main light and approximates the effect of the sky. This light is placed near the camera to prevent it from casting new shadows of its own and to simulate the effect of uniform lighting. The appearance of shadows within shadows is considered extremely "ugly" and should be avoided since it makes the picture more difficult to interpret. The ratio of fill-in light intensity to main light intensity is usually chosen to be about 1 to 3.
3. The accent light—Its purpose is to enliven the rendering by adding highlights and 'sparkle'. It should be a small collimated source which can be directed to illuminate small sections of the subject. It is placed behind and to the side of the subject so that it cannot cast shadows of its own. This light can add catchlights (specular reflections such as on eyes or metal objects) and bright outlines (particularly on hair).
4. The background light—Its purpose is to 'separate' the subject from the background. It illuminates the background only, such that the intensity reflected by the subject will nowhere match that of the background. This ensures that the two can be easily 'separated' because the edge between them will be visible.

Other hints are that too many lights spoil the effect, having the main light at the camera creates a "flat" image, shadows crossing edges on the subject are to be avoided and that light parts of the image draw the attention of the viewer. It is interesting to note how much of what is vaguely formulated in these introductions to photography can be understood from the point of view of shading.

4.5.2 Human Performance and the Science of Cosmetics

Judging by the popularity of monocular pictures of people and other smooth objects, humans are good at interpreting shading information. Since they use the same basic information as our shape-from-shading algorithm we expect to find similar shortcomings. Supposing the human visual system does not use the shading information in simple heuristic ways only, one might expect that the perception system 'solves' the equations or a much simplified form of them. Since this cannot be done locally (the way some portions of an edge-finding process might work) it is difficult to suggest an elegant and simple physiological mechanism and a place to look for it.

When a surface whose photometric properties are taken to be uniform is treated so as to change these properties in some areas, the apparent shape is changed. This of course is one of the uses of makeup. The shape of a face for example can be made to conform more closely to what a person thinks is currently considered ideal. This is achieved by making some areas darker (causing them to appear steeper) and others lighter. Areas lightened usually include singular points and cause a change in the apparent skin darkness (a normalization effect) and will change the apparent shape in areas other than the singular points.

These modifications can change the shape perceived when viewed under the right lighting conditions. The effect will change somewhat with orientation and may at times disappear when no reasonable shape would give rise to the shading observed. Because of a number of surface oils the skin has a specular component in its reflectivity. It is also fairly translucent. Both of these effects are sometimes controlled with talcum powder. The removal of the specular components makes the surface appear more rounded and soft.

4.5.3 Generating Shaded Images

The inverse problem of producing images of a specified scene with shading and shadows is vastly different from the method of shape-from-shading. Most programs written for this purpose can be used for objects bounded by planes only. The main issues of optimization of the calculation of which surfaces are visible to the source and camera respectively have been dealt with in some detail. Although the two problems are inverses of one another, the methods used are quite different.

An interesting problem of a mathematical nature (and incidentally with application to cutting woodcuts) is that of producing curved lines in a plane such that the density of lines is proportional to the shading in the image of some real or imagined object. Preferrably one would like as small a number of 'unnecessary' breaks in the lines as possible, i.e., the lines should either close on themselves or leave the image. Another restriction

one might apply is that the lines should not cross (when producing woodcuts one would most likely also reflect some of the surface texture in the choice of lines).

For a special case, a solution is immediately at hand. This is the case where we have a distant camera at a distant source and a reflectivity function ϕ such that

$$\phi(I, I, 1) = I = \frac{1}{\sqrt{1 + p^2 + q^2}}$$

Here the contour lines give a solution, with no crossing lines and no "unnecessary" breaks. One of the most attractive features of contour maps is perhaps just this fact that they provide some shading information.

4.5.4 Determining Shape from Texture Gradients

A problem related to that of determining shape using shading is that of determining shape from the depth-cue of texture gradients. A textured surface will produce an image in which the texture is distorted in a way reflecting both the direction and the amount of the inclination of the surface. An image of a tilted surface with a random dot pattern for example will be compressed in one direction (the average distance between dots is decreased) by an amount proportional to the inclination of the surface. Both direction and magnitude of the gradient can thus be determined—except for a two-way ambiguity.

In practice it may not always be easy to determine such texture gradients reliably because of low resolution of the imaging device and scatter, causing a reduction in constrast. Some simple textures may be handled by simple counting or distance measurements as suggested above, while more complicated textures like a plastered wall will need more sophisticated techniques, such as two-dimensional correlation. Some experimentation with this technique showed promise, but did not supply very reliable gradients and the method was slow.

The next problem is how to obtain the shape from the texture gradients. Starting at some point (whose distance from the camera we assume known), we use some external knowledge to resolve the two-way ambiguity. We can now take a small step in any direction and find the gradient at this new point. Continuing in this way we trace out some curve on the surface of the object (somewhat analogous to the characteristics in the shape-from-shading method, except that here the curve is quite arbitrary).

Let s be the arc-distance along the curve, z the distance to the initial point, and p and q the components of the gradient. Then

$$z(s) = z_0 + \int_0^s (p, q) \cdot ds$$

If one takes small enough steps, one can continue to resolve the ambiguity at each step by using the assumption of smoothness. This can be done until we meet a point where the gradient is zero. To continue past such a point would require some external knowledge to again resolve the two-way ambiguity. An aggregation of points with zero inclination can form an ambiguity edge which cannot be crossed.

Clearly we can reach a given point through many paths from the initial point. This allows us some error checking, but there certainly are better ways of making use of the excess information. For that is what we have, since we know from the solution to the shape-from-shading problem that only one value is required at each point for the determination of the shape, while we here have two (the components of the gradient). Most commonly when faced with such an excess of information one can make use of some least-squares technique to improve the accuracy. Perhaps a relaxation method on a grid would be useful.

REFERENCES

1. Rindfleisch, T.: Photometric Method for Lunar Topography, *Photometric Eng.*, 32(2):262–276 (1966).
2. Horn, Berthold K. P.: The Image Dissector Eyes, *M.I.T. Artificial Intelligence Laboratory Memo* 178, 1969.
3. Garabedian, D. R.: "Partial Differential Equations," John Wiley & Sons, New York, 1964.
4. Fesenkov, V. P.: Photometric Investigations of the Lunar Surface, *Astronomochheskii Zh.*, 5:219–234 (1929).
5. Willingham, D. E.: The Lunar Reflectivity Model for Ranger Block III Analysis, *Jet Prop. Lab. Tech. Rept.* 32-664, Pasadena, Calif., November, 1964.
6. Van Diggelen, J.: A Photometric Investigation of the Slopes and Heights of the Ranges of Hills in the Maria of the Moon, *Bull. Astron. Inst. Netherlands*, 11(423):283–289 (1951).

PATRICK HENRY WINSTON, Editor

The Psychology of Computer Vision