11

Reflectance Map: Shape from Shading

In the previous chapter we introduced the reflectance map and the image irradiance equation, and we used them to recover surface orientation from registered images taken under different lighting conditions. In this chapter we concentrate on the recovery of surface shape from a single image. This is a more difficult problem that will require the development of more advanced tools. We first examine the case of a linear reflectance map. It turns out that, under point-source illumination, the reflectance maps of the surface material in the maria of the moon and on rocky planets such as Mercury are functions of linear combinations of the components of the gradient. Next, we consider the shape-from-shading problem when the reflectance map is rotationally symmetric. This applies, for example, to images taken with the scanning electron microscope. We then solve the general case.

The image irradiance equation can be viewed as a nonlinear first-order partial differential equation. The traditional methods for solving such equations depend on growing characteristic strips. This is a sequential process. We are more interested in methods that ultimately lead to parallel algorithms. Consequently, we formulate a minimization problem that leads to a relaxation algorithm on a grid. We choose to minimize the integral of the difference between the observed brightness and that predicted for the estimated shape.

It is, of course, very important to know whether a solution to these problems exists and whether there is more than one solution. Unfortunately, these existence and uniqueness questions are difficult to decide without detailed assumptions about the reflectance map. We briefly explore what is known in this regard and then finish the chapter by showing how the ideas developed here can be applied to improve the results obtained by means of the photometric stereo method discussed in the previous chapter.

11.1 Recovering Shape from Shading

How can we recover the shape of a surface from a single image? Different parts of the surface are oriented differently and thus will appear with different brightnesses. We can take advantage of this spatial variation of brightness, referred to as *shading*, in estimating the orientation of surface patches. Measurement of brightness at a single point in the image, however, only provides one constraint, while surface orientation has two degrees of freedom. Without additional information, we cannot recover the orientation of a surface patch from the image irradiance equation

$$E(x, y) = R(p, q).$$

We have already discussed one method for introducing another constraint: the use of additional images taken under different lighting conditions.

11.1.1 Growing a Solution

But what if we have only one image? People can estimate the shapes of facial features using a single picture reproduced in a magazine. This suggests that there is enough information or that we implicitly introduce additional assumptions. Many surfaces are smooth, lacking discontinuities in depth. Also, there are often no discontinuities in the partial derivatives. An even wider class of objects have piecewise-smooth surfaces, with departures from smoothness concentrated along edges.

The assumption of smoothness provides a strong constraint. Neighboring patches of the surface cannot assume arbitrary orientations. They have to fit together to make a continuous, smooth surface. Thus a global method exploiting a smoothness constraint can be envisioned.

11.1.2 Linear Reflectance Maps

To begin with, we consider some special cases. Suppose that

$$R(p, q) = f(ap + bq),$$

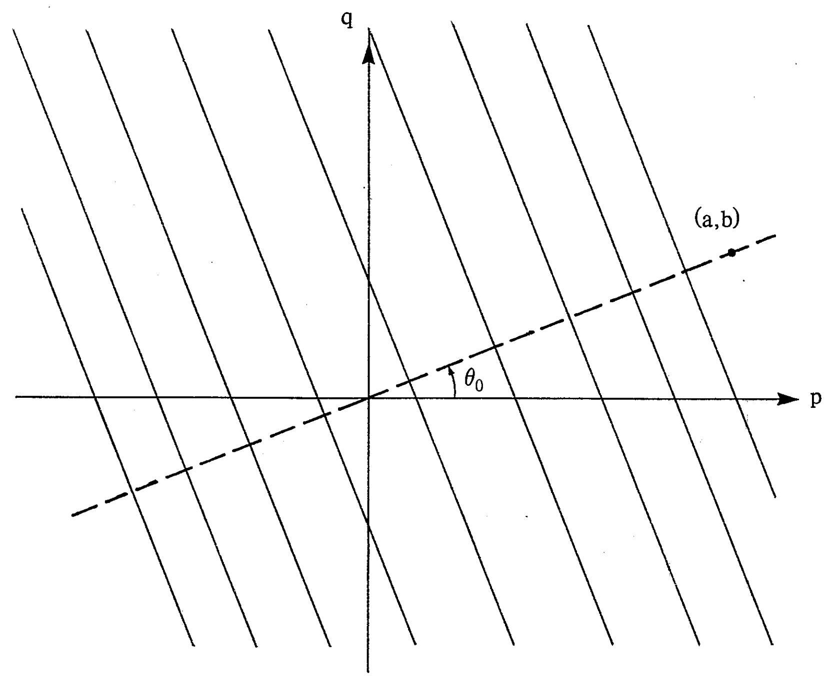

Figure 11-1. A reflectance map that is a function of a linear combination of the components of the gradient is particularly simple. The contours of constant brightness are parallel straight lines in gradient space.

where a and b are constants (figure 11-1).

Here f is a strictly monotonic function that has an inverse, f^{-1} (figure 11-2). From the image irradiance equation we then have

$$ap + bq = f^{-1}(E(x, y)).$$

We cannot determine the gradient (p, q) at a particular image point from a measurement of image brightness alone, but we do have one equation that constrains its possible values.

The slope of the surface, in a direction that makes an angle θ with the x-axis, is

$$m(\theta) = p \cos\theta + q \sin\theta.$$

This is the directional derivative. Now choose a particular direction θ_0 (figure 11-1), where $\tan\theta_0 = b/a$, that is,

$$\cos\theta_0 = a/\sqrt{a^2 + b^2} \quad \text{and} \quad \sin\theta_0 = b/\sqrt{a^2 + b^2}.$$

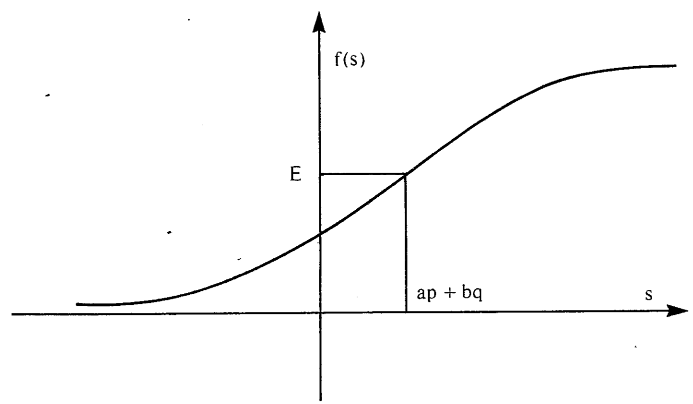

Figure 11-2. If the function f is continuous and monotonic, an inverse can be found and $s = ap+bq$ can be recovered from the brightness measurement $E(x, y)$.

The slope in this direction is

$$m(\theta_0) = \frac{ap+bq}{\sqrt{a^2+b^2}} = \frac{1}{\sqrt{a^2+b^2}} f^{-1}\big(E(x,y)\big).$$

Thus we can determine the slope in a particular direction. Note that we know nothing about the slope in the direction at right angles to this, however.

Starting at a particular image point we can take a small step of length $\delta\xi$, producing a change in z of $\delta z = m\,\delta\xi$. Thus

$$\frac{dz}{d\xi} = \frac{1}{\sqrt{a^2+b^2}} f^{-1}\big(E(x,y)\big),$$

where

$$x(\xi) = x_0 + \xi\cos\theta \qquad \text{and} \qquad y(\xi) = y_0 + \xi\sin\theta.$$

Suppose that we start the solution at the point $(x_0, y_0, z_0)^T$ on the surface. Integrating the differential equation for z derived above, we obtain

$$z(\xi) = z_0 + \frac{1}{\sqrt{a^2+b^2}} \int_0^{\xi} f^{-1}\big(E(x,y)\big)\, d\xi,$$

where x and y in the integrand are the linear functions of ξ given above. In this fashion we obtain a profile of the surface along a line in the special direction defined above (one of the straight lines in figure 11-3). The profile is called a *characteristic curve*. In practice, of course, the integrand will not be given as a formula, so that numerical integration is called for.

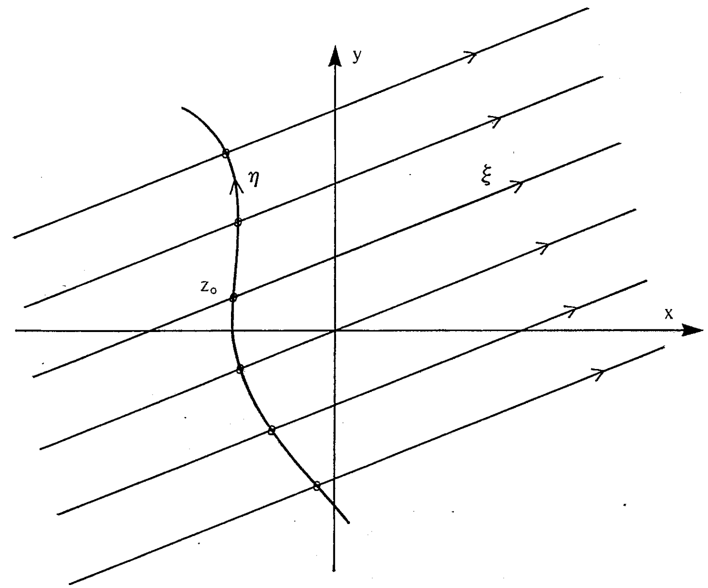

Figure 11-3. The base characteristics are parallel straight lines when the reflectance map is a function of a linear combination of the components of the gradient. The surface can be recovered by integration along these lines, provided the height $z_0(\eta)$ along some initial curve is given.

We cannot determine the absolute distance to the surface—the constant of integration—since the absolute distance does not influence the shading, only variations in depth do. If we require information about absolute distance, we shall need to know the value z_0 at one point. The shape can be recovered without this additional information, however.

Now suppose that we are given initial information not just at a point, but as a profile $z(\eta)$ along some curve that is nowhere parallel to the special direction (a, b) (figure 11-3). Then we can integrate along lines starting at points of this initial curve. The whole surface can be explored in this way if the initial curve extends far enough. The general case, to be explored later, is similar in that the surface is determined by integration along special curves in the image. The general case differs, however, in that these curves are not predetermined straight lines.

The special case discussed here is of practical importance because the material in the maria of the moon has reflectance properties that can be closely approximated by some function of $\cos\theta_i/\cos\theta_e$, as already mentioned. The reflectance map, in this case, is a function of a linear combination of p and q. This was the version of the shape-from-shading problem that first received attention. We use orthographic projection for simplicity here, but the method can be extended to the case of perspective projection.

11.1.3 Rotationally Symmetric Reflectance Maps

If the light source is distributed in a rotationally symmetric fashion about the viewer, then the reflectance map is rotationally symmetric, too. That is, we can write

$$R(p,q) = f(p^2 + q^2)$$

for some f. One situation leading to a rotationally symmetric reflectance map is provided by a hemispherical sky, if we assume that the viewer is looking straight down from above. Another example is that of a point source at essentially the same place as the viewer.

Now suppose that the function f is strictly monotonic and differentiable, with inverse f^{-1}. From the image irradiance equation we obtain

$$p^2 + q^2 = f^{-1}\bigl(E(x,y)\bigr).$$

The direction of steepest ascent makes an angle θ_s with the x-axis, where $\tan\theta_s = q/p$, so that

$$\cos\theta_s = p/\sqrt{p^2+q^2} \qquad \text{and} \qquad \sin\theta_s = q/\sqrt{p^2+q^2}.$$

The slope in the direction of steepest ascent is

$$m(\theta_s) = \sqrt{p^2+q^2} = \sqrt{f^{-1}\bigl(E(x,y)\bigr)}.$$

Thus in this case we can find the slope of the surface, given its brightness, but we cannot find the direction of steepest ascent.

Suppose we did know the direction of steepest ascent, given by (p,q). Then we could take a small step of length $\delta\xi$ in the direction of steepest ascent. The changes in x and y would be given by

$$\delta x = \frac{p}{\sqrt{p^2+q^2}}\delta\xi \qquad \text{and} \qquad \delta y = \frac{q}{\sqrt{p^2+q^2}}\delta\xi.$$

The change in z would be

$$\delta z = m\,\delta\xi = \sqrt{p^2+q^2}\,\delta\xi = \sqrt{f^{-1}\bigl(E(x,y)\bigr)}\,\delta\xi.$$

To simplify these equations, we could take a step of length $\sqrt{p^2+q^2}\,\delta\xi$ rather than $\delta\xi$. Then

$$\delta x = p\,\delta\xi, \quad \delta y = q\,\delta\xi, \quad \delta z = (p^2+q^2)\,\delta\xi = f^{-1}\bigl(E(x,y)\bigr)\,\delta\xi.$$

The problem with this approach is that we need to determine the values of p and q at the new point in order to continue the solution. We need to develop equations for the changes δp and δq in p and q, respectively.

Before we address this issue, let us look at the image brightness gradient $(E_x, E_y)^T$. We know that a planar surface patch gives rise to a region of uniform brightness in the image. Thus a nonzero brightness gradient can occur only where the surface is curved. To find the brightness gradient, we differentiate the image irradiance equation

$$E(x,y) = f(p^2+q^2)$$

with respect to x and y. Let r, s, and t be the second partial derivatives of z with respect to x and y as defined by

$$r = \frac{\partial^2 z}{\partial x^2}, \quad \frac{\partial^2 z}{\partial x \partial y} = s = \frac{\partial^2 z}{\partial y \partial x}, \quad t = \frac{\partial^2 z}{\partial y^2}.$$

Then, using the chain rule for differentiation, we obtain

$$E_x = 2(p\,r + q\,s)f' \quad \text{and} \quad E_y = 2(p\,s + q\,t)f',$$

where $f'(s)$ is the derivative of $f(s)$ with respect to its single argument s.

Now we return to the problem of determining the changes δp and δq occasioned by the step $(\delta x, \delta y)$ in the image plane. We find

$$\delta p = r\,\delta x + s\,\delta y \quad \text{and} \quad \delta q = s\,\delta x + t\,\delta y$$

by simple differentiation. In our case $\delta x = p\,\delta\xi$ and $\delta y = q\,\delta\xi$, so that

$$\delta p = (p\,r + q\,s)\,\delta\xi \quad \text{and} \quad \delta q = (p\,s + q\,t)\,\delta\xi,$$

or

$$\delta p = \frac{E_x}{2f'}\delta\xi \qquad \text{and} \qquad \delta q = \frac{E_y}{2f'}\delta\xi.$$

In the limit as $\delta\xi \to 0$, we obtain the differential equations

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

$$\dot{p} = \frac{E_x}{2f'}, \quad \dot{q} = \frac{E_y}{2f'},$$

where the dots denote differentiation with respect to ξ. Given starting values, this set of five ordinary differential equations can be solved numerically to produce a curve on the surface of the object. Curves generated in this fashion are called *characteristic curves*, and in this particular case they happen to be the curves of steepest ascent. These curves are everywhere perpendicular to the contours of constant height. In the case treated previously, in which the reflectance map was a linear function of p and q, the characteristic curves were parallel planar sections of the surface.

By differentiating $\dot{x} = p$ and $\dot{y} = q$ one more time with respect to ξ, we obtain the alternate formulation

$$\ddot{x} = \frac{E_x}{2f'}, \quad \ddot{y} = \frac{E_y}{2f'}, \quad \dot{z} = f^{-1}\big(E(x,y)\big).$$

Naturally, these equations can only be solved numerically, since E_x and E_y are image brightness measurements, not functions of x and y given in closed form.

The special case discussed above is of practical importance, since scanning electron microscopes produce images analogous to those produced in an optical system with a light source disposed around the viewer in a rotationally symmetric fashion. In such a device a focused beam of electrons strikes a surface in a position determined by two orthogonal deflection coils. Secondary electrons are generated as a result of collisions between the incident primary electrons and the atoms in the material. Some of these escape and are collected by an electrode. Secondary electrons generated deep inside the material have less of a chance to escape than those generated near the surface. The secondary electron flux is thus lowest when the beam strikes the surface at right angles and is highest at grazing incidence. The probing beam scans out a raster, while the brightness of a cathode ray tube scanned in the same fashion is modulated in proportion to the secondary electron current. The result is a (highly magnified) picture of the surface. People find such pictures easy to interpret, because they exhibit shading due to the dependence of brightness on surface orientation. The only strange thing about these images is that surface patches perpendicular to the viewer appear darkest, not brightest, in a scanning electron microscope picture.

11.1.4 The General Case

Suppose that we have the coordinates of a particular point on the surface and that we wish to extend the solution from this point. Taking a small step $(\delta x, \delta y)$, we note once more that the change in depth is given by

$$\delta z = p\,\delta x + q\,\delta y,$$

where p and q are the first partial derivatives of z with respect to x and y (figure 11-4). We cannot proceed unless p and q are also known. Unfortunately, the image irradiance equation provides only one constraint; this is not enough information to allow a solution for both p and q.

Suppose for the moment that we did know p and q at the given point. Then we could extend the solution from (x, y) to $(x + \delta x, y + \delta y)$. But to

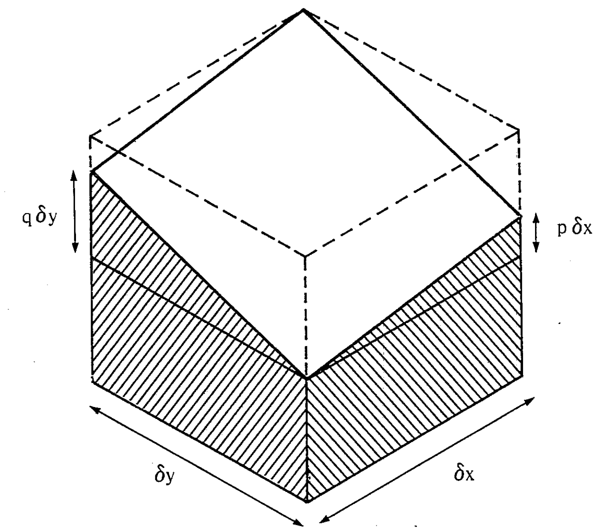

Figure 11-4. The change in height, δz, is the sum of $p\,\delta x$, the change in height due to a small step in the x-direction, and $q\,\delta y$, the change in height due to a small step in the y-direction.

continue from there we would need the new values of p and q at that point (figure 11-5). Now the changes in p and q can be computed using

$$\delta p = r\,\delta x + s\,\delta y \quad \text{and} \quad \delta q = s\,\delta x + t\,\delta y,$$

where r, s, and t are the second partial derivatives of z with respect to x and y. This can be written in a more compact form as

$$\begin{pmatrix} \delta p \\ \delta q \end{pmatrix} = \mathbf{H} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix},$$

where $\mathbf{H}$ is the *Hessian matrix* of second partial derivatives:

$$\mathbf{H} = \begin{pmatrix} r & s \\ s & t \end{pmatrix}.$$

The Hessian provides information on the curvature of the surface. For small surface inclinations, its determinant is the Gaussian curvature, to be introduced later. Also, the *trace* of the Hessian (the sum of its diagonal elements) is the Laplacian of depth, which for small surface inclinations is twice the so-called *mean curvature*. We shall explore surface curvature in chapter 16, where we discuss extended Gaussian images.

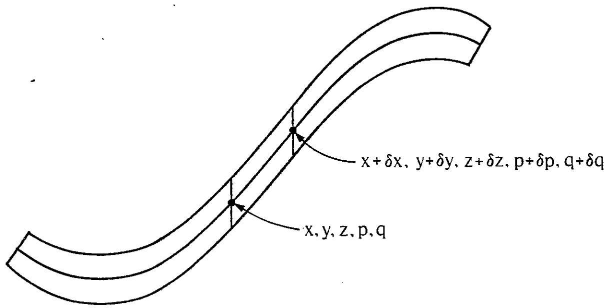

Figure 11-5. The solution of the shape-from-shading problem is determined by solving five differential equations for x, y, z, p, and q. The result is a characteristic strip, a curve in space along which surface orientation is known.

To use the Hessian matrix for computing the changes in p and q, we need to know its components, the second partial derivatives of z. To keep track of them we would need still higher derivatives. We could go on differentiating ad infinitum. Note, however, that we have not yet used the image irradiance equation! Differentiating it with respect to x and y, and using the chain rule, we obtain

$$E_x = r\,R_p + s\,R_q \qquad \text{and} \qquad E_y = s\,R_p + t\,R_q,$$

or

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \mathbf{H} \begin{pmatrix} R_p \\ R_q \end{pmatrix},$$

where the Hessian $\mathbf{H}$ once again makes an appearance. This is a relationship between the gradient $(E_x, E_y)^T$ in the image and the gradient $(R_p, R_q)^T$ in the reflectance map. We cannot solve for $\mathbf{H}$, since we have only two equations and three unknowns r, s, and t, but fortunately we do not need the individual elements of $\mathbf{H}$. While we cannot continue the solution in an arbitrary direction, we can do so in a specially chosen direction. This is the key idea. Let

$$\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{pmatrix} R_p \\ R_q \end{pmatrix} \delta\xi,$$

where $\delta\xi$ is a small quantity. Then

$$\begin{pmatrix} \delta p \\ \delta q \end{pmatrix} = \mathbf{H} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \mathbf{H} \begin{pmatrix} R_p \\ R_q \end{pmatrix} \delta\xi,$$

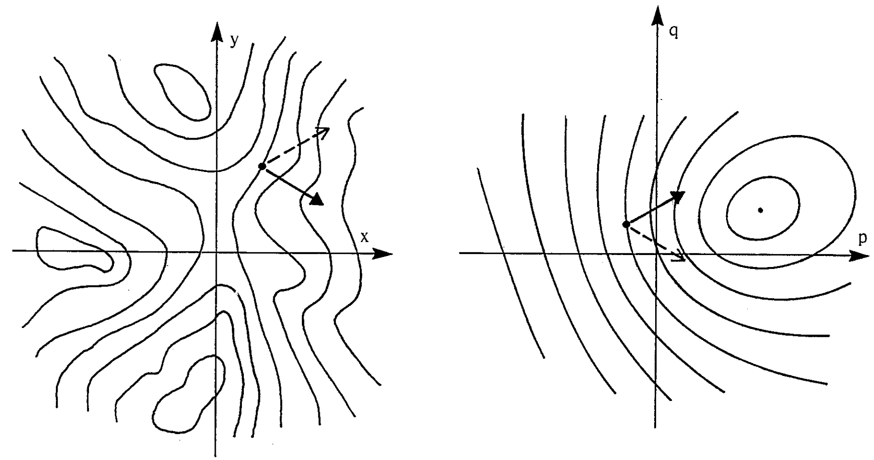

Figure 11-6. Curiously, the step taken in pq-space is parallel to the gradient of $E(x, y)$, while the step taken in xy-space is parallel to the gradient of $R(p, q)$.

or

$$\begin{pmatrix} \delta p \\ \delta q \end{pmatrix} = \begin{pmatrix} E_x \\ E_y \end{pmatrix} \delta\xi.$$

Thus, if the direction of the change in the image plane is parallel to the gradient of the reflectance map, then the change in (p, q) can be computed. The direction of the change in gradient space is parallel, in turn, to the gradient in the image (figure 11-6). We can summarize all this in five ordinary differential equations:

$$\dot{x} = R_p, \qquad \dot{y} = R_q, \qquad \dot{z} = p\,R_p + q\,R_q,$$
$$\dot{p} = E_x, \qquad \dot{q} = E_y,$$

where the dots denote differentiation with respect to ξ. A solution of these differential equations is a curve on the surface. The parameter ξ will vary along this curve. By rescaling the equations, we can easily arrange for ξ to be any function of length along the curve.

11.2 Characteristic Curves and Initial Curves

The curves traced out by the solutions of the five ordinary differential equations are called *characteristic curves*, and their projections in the image are called *base characteristics*. The solutions for x, y, z, p, and q actually form a *characteristic strip*, since they define not only a curve in space but surface orientation along this curve as well (figure 11-5).

To obtain the whole surface we must patch together characteristic strips. Each requires a point where initial values are given in order to start the solution. If we are given an initial curve on the surface, a solution for the surface can be obtained as long as this curve is nowhere parallel to any of the characteristics. On this curve, starting values of p and q can be obtained using the image irradiance equation,

$$E(x, y) = R(p, q),$$

and the known derivatives of z along the curve. Suppose, for example, that the initial curve is given in terms of a parameter η, as $x(\eta)$, $y(\eta)$, and $z(\eta)$. Then, along this curve,

$$\frac{\partial z}{\partial \eta} = p\frac{\partial x}{\partial \eta} + q\frac{\partial y}{\partial \eta}.$$

We have just derived the method of characteristic strip expansion for solving first-order partial differential equations. In our case the relevant equation is the image irradiance equation, a (possibly very nonlinear) first-order partial differential equation.

Figure 11-7 shows a digitized picture of a face, the face with base characteristics superimposed, and the face with a contour map of the recovered shape.

11.3 Singular Points

We are normally not given an initial curve along with the image of an object. How much can we tell about shape in the absence of such auxiliary information? Are there any points where surface orientation can be determined directly? Suppose that $R(p, q)$ has a unique isolated maximum at (p_0, q_0); that is,

$$R(p, q) < R(p_0, q_0) \qquad \text{for all } (p, q) \neq (p_0, q_0).$$

Also assume that at some point (x_0, y_0) in the image, $E(x_0, y_0) = R(p_0, q_0)$. Then it is clear that at this point the gradient (p, q) is uniquely determined to be (p_0, q_0). It would seem, then, that we could start the solution at such a singular point. Unfortunately, at a maximum of $R(p, q)$ the partial derivatives R_p and R_q are zero. Thus the solution will not move from such a point because $\dot{x}$ and $\dot{y}$ are zero. One way to bypass this apparent impasse is to construct a small "cap" at this point and start the solution at the edge of this cap, as we shall show in the next section.

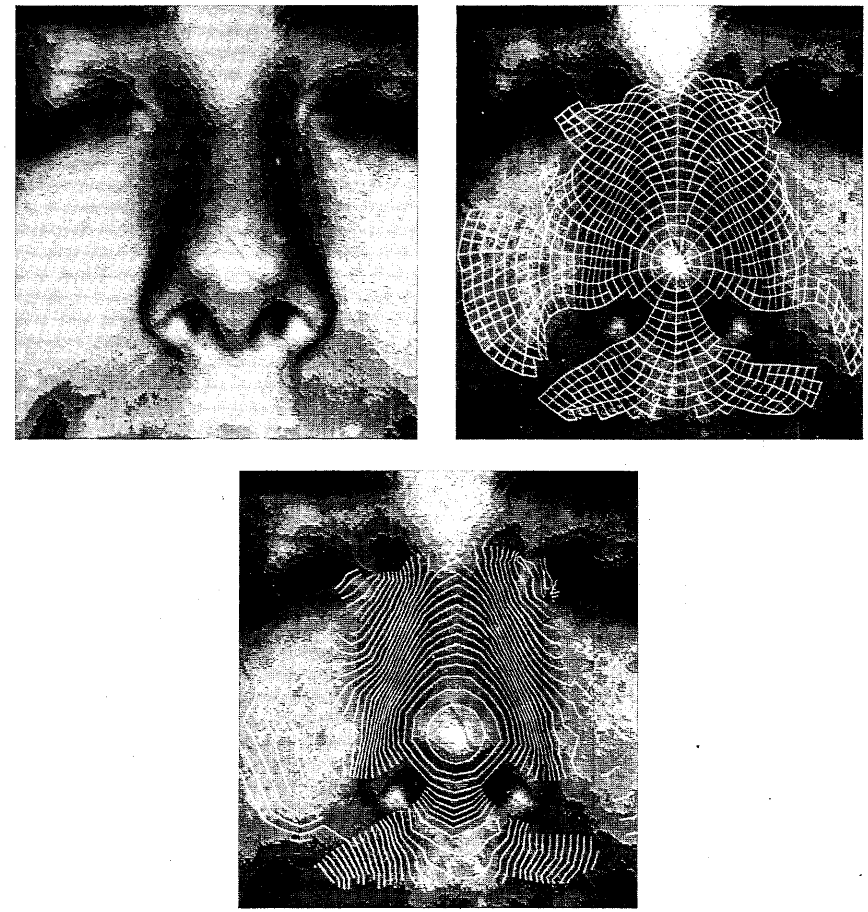

Figure 11-7. The shape-from-shading method is applied here to the recovery of the shape of a nose. The first picture shows the (crudely quantized) gray-level image available to the program. The second picture shows the base characteristics superimposed, while the third shows a contour map computed from the elevations found along the characteristic curves.

11.4 Power Series near a Singular Point

To observe what happens near a singular point, consider the reflectance map

$$R(p, q) = \frac{1}{2}(p^2 + q^2).$$

In this case we have a unique isolated minimum. Let there be a singular point at the origin such that $E(0,0) = 0$. We conclude that $(p,q) = (0,0)$ at this point. If the surface is smooth enough, we can expand z as a power series in x and y with the first-order terms missing. If we also ignore higher-order terms near the origin, we can write

$$z = z_0 + \frac{1}{2}(ax^2 + 2bxy + cy^2).$$

Thus

$$p = ax + by \quad \text{and} \quad q = bx + cy.$$

Substituting these values in the formula for the reflectance map, we obtain

$$E(x,y) = \frac{1}{2}(a^2 + b^2)x^2 + (a+c)bxy + \frac{1}{2}(b^2 + c^2)y^2.$$

Our task is to determine the coefficients a, b, and c, given the image brightness and its derivatives near the origin. Before we go on, observe that the surface

$$z = z_0 - \frac{1}{2}\left(ax^2 + bxy + \frac{1}{2}cy^2\right)$$

gives rise to exactly the same shading pattern, so we already know that there will be at least two solutions.

The brightness gradient is given by

$$E_x = (a^2 + b^2)x + (a+c)b\,y,$$
$$E_y = (a+c)b\,x + (b^2 + c^2)y.$$

Thus $(E_x, E_y)^T = (0,0)^T$ at $(x,y) = (0,0)$, as it should. We cannot use the brightness gradient to recover the shape. Differentiating again, we obtain the three equations

$$E_{xx} = a^2 + b^2, \quad E_{xy} = (a+c)b, \quad E_{yy} = b^2 + c^2,$$

in the three unknowns a, b, and c. Three second-order polynomials in three unknowns can have up to eight solutions. The three equations found here have a rather special form, however, and there are only four solutions, as shown in exercise 11-10.

In any case, given one of these local solutions, we can construct a small cap. The edge of this region then constitutes an initial strip for the method of characteristic strip expansion, since p and q as well as z are known on the edge. Note also that the solution will move away from the edge, since R_p and R_q are nonzero there.

The above analysis can be generalized to singular points away from the origin and to other rotationally symmetric reflectance maps. It provides a means for starting a solution a small distance away from the singular point. A possible problem is that more than one shape might give rise to the same shading, since the nonlinear equations containing the coefficients of the power series near the singular point can have more than one solution, as they did here.

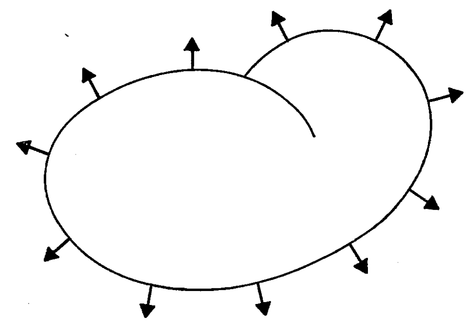

Figure 11-8. The occluding boundary provides an important constraint on the solutions of the shape-from-shading problem. The surface orientation of any solution has to match the known surface orientation along the silhouette.

11.5 Occluding Boundaries

At what other point is the surface orientation known? If the object has a smooth surface, then the silhouette also provides valuable information (figure 11-8). The occluding boundary is the curve on the surface that projects to the silhouette. The orientation there is known, since the tangent plane includes the direction to the observer and also the tangent at the corresponding point on the silhouette. In other words, the surface normal on the occluding boundary lies in a plane parallel to the image plane and is perpendicular to the silhouette.

The only problem with this kind of information is that the slope of the surface is infinite on the occluding boundary. It is thus difficult to incorporate this information as an "initial curve." Nevertheless, it is possible to show that if the reflectance map is a strictly monotonic function of a quadratic function of p and q, then there is a unique surface corresponding to a particular shaded image that exhibits a simple closed silhouette. Conversely, if the reflectance map is a linear function in p and q, then an infinite number of surfaces gives rise to the same shading. In many cases shading and auxiliary information determine a surface uniquely. In some cases they do not, unfortunately: The shading on a small patch of the

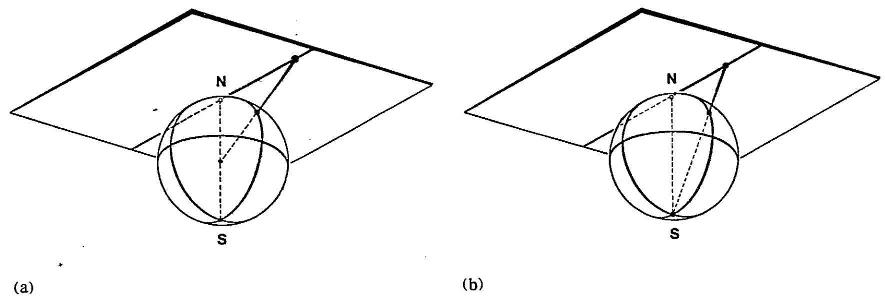

Figure 11-9. Because points on the Gaussian sphere specify directions in space, the reflectance map can be plotted on the Gaussian sphere. (a) More commonly, we project the upper hemisphere onto an infinite plane, called the gradient space. If we want to deal with the occluding boundary of an object, another projection is more useful. (b) Here, the whole sphere, except for one point, is projected onto a plane, called the stereographic plane. (Figures reproduced with permission from the chapter by Woodham in *Image Understanding 1984*, edited by S. Ullman & W. Richards, Ablex Publishing Corp., Norwood, New Jersey, 1984.)

surface of an object, for example, without any other information, does not determine the local shape of the surface.

11.6 Stereographic Projection

Orientation has two degrees of freedom. We can specify the orientation of a patch by giving its gradient (p, q). Alternatively we can erect a unit normal $\hat{\mathbf{n}}$. As noted in the previous chapter, we can use the Gaussian sphere to represent the direction in which the surface normal is pointing. The Gaussian sphere itself is often inconvenient to use because of its curved surface. This is why we usually project it onto a plane to obtain the *gradient space* (figure 11-9a).

Consider an axis through the sphere parallel to the z-axis. We can project points on the "northern" hemisphere onto a plane tangent at the "north" pole, using the center of the sphere as the center of projection. This is called the *gnomonic* projection. It is easy to show that position in this plane equals $(-p, -q)$. One disadvantage of gradient space (the plane so defined) is that we can only project one hemisphere onto the plane if we want to avoid ambiguity.

Often we are only concerned with surface elements facing the viewer. These correspond to points on the northern hemisphere. But at times directions in the other hemisphere are needed also. In a scene lit from behind, for example, the direction to the light source can be specified by a point in the southern hemisphere. We just came across another difficulty with

gradient space. Orientations of surface patches on the occluding boundary correspond to points on the equator of the Gaussian sphere, which project to infinity in gradient space.

One way out of these difficulties is provided by the *stereographic projection*. Here again we project onto a plane tangent at the north pole, but this time the center of projection is the south pole (figure 11-9b). All points on the sphere, except for the south pole, can be mapped. The equator projects to a circle of radius two. Let us call the coordinates in stereographic space f and g. We show in exercise 11-13 that

$$f = \frac{2p}{1 + \sqrt{1 + p^2 + q^2}} \quad \text{and} \quad g = \frac{2q}{1 + \sqrt{1 + p^2 + q^2}}.$$

Conversely,

$$p = \frac{4f}{4 - f^2 - g^2} \quad \text{and} \quad q = \frac{4g}{4 - f^2 - g^2}.$$

An added advantage of stereographic space is that it is a conformal projection of the Gaussian sphere. That is, angles on the surface of the sphere are projected faithfully into equal angles in the plane. One disadvantage, however, is that some formulae become more complicated when expressed in stereographic coordinates.

11.7 Relaxation Methods

The method of characteristic strip expansion suffers from a number of practical problems, including sensitivity to measurement noise. Special means must be employed to prevent adjacent characteristics from crossing over each other as a result of small errors accumulating in the numerical integration of the differential equations. This method also makes it hard to utilize the information on surface orientation available on the occluding boundary. Finally, it suggests neither biological nor parallel machine implementations.

More desirable would be an iterative scheme similar to one of the finite-difference methods used for solving elliptic second-order partial differential equations. This would immediately suggest ways to incorporate boundary and other auxiliary information.

11.7.1 Minimization in the Continuous Case

Our objective is to find two functions, $f(x, y)$ and $g(x, y)$, that ensure that the image irradiance equation,

$$E(x, y) = R_s(f, g),$$

is satisfied, where $R_s(f, g)$ is the reflectance map expressed in stereographic coordinates. We also want $f(x, y)$ and $g(x, y)$ to correspond to a smooth surface. Of the many ways to measure smoothness, we choose one that penalizes rapid changes in f and g. We try to minimize the integral

$$e_s = \iint_I \left((f_x^2 + f_y^2) + (g_x^2 + g_y^2)\right) dx\, dy,$$

where f_x, f_y, g_x, and g_y are the first partial derivatives of f and g with respect to x and y. Other measures of departure from "smoothness" could also be used. These would lead to somewhat different algorithms.

So far, the plan is to minimize e_s subject to the constraint that f and g must satisfy the image irradiance equation. In practice, there are errors in both the measurements of the irradiance and the determination of the reflectance map. Instead of insisting on equality of $E(x, y)$ and $R_s(f, g)$, we could try to minimize the error

$$e_i = \iint_I \left(E(x, y) - R_s(f, g)\right)^2 dx\, dy.$$

Overall, then, we are to minimize $e_s + \lambda e_i$, where λ is a parameter that weights the errors in the image irradiance equation relative to the departure from smoothness. This parameter should be made large if brightness measurements are very accurate, and small if they are very noisy.

The minimization of an integral of the form

$$\iint F(f, g, f_x, f_y, g_x, g_y)\, dx\, dy$$

is a problem in the calculus of variations (a topic covered in the appendix). The corresponding Euler equations are

$$F_f - \frac{\partial}{\partial x} F_{f_x} - \frac{\partial}{\partial y} F_{f_y} = 0,$$

$$F_g - \frac{\partial}{\partial x} F_{g_x} - \frac{\partial}{\partial y} F_{g_y} = 0,$$

where F_f is the partial derivative of F with respect to f. In the present case,

$$F = (f_x^2 + f_y^2) + (g_x^2 + g_y^2) + \lambda\left(E(x, y) - R_s(f, g)\right)^2.$$

The aim is to minimize the integral of F. The Euler equations for this problem yield

$$\nabla^2 f = -\lambda\left(E(x, y) - R_s(f, g)\right) \frac{\partial R_s}{\partial f},$$

$$\nabla^2 g = -\lambda\left(E(x, y) - R_s(f, g)\right) \frac{\partial R_s}{\partial g},$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

is the Laplacian operator. The result is a coupled pair of elliptic second-order partial differential equations. These can be solved by iterative methods once the values of f and g on the silhouette are introduced.

11.7.2 Minimization in the Discrete Case

We can develop a numerical method either by approximating the continuous solution found in the previous section or by directly minimizing a discrete version of the integral. Readers uncomfortable with the calculus of variations may be more satisfied with the latter approach. We can measure the departure from smoothness at the point (i, j) by

$$s_{i,j} = \frac{1}{4}\left((f_{i+1,j} - f_{i,j})^2 + (f_{i,j+1} - f_{i,j})^2 + (g_{i+1,j} - g_{i,j})^2 + (g_{i,j+1} - g_{i,j})^2\right),$$

while the error in the image irradiance equation is given by

$$r_{ij} = \left(E_{ij} - R_s(f_{ij}, g_{ij})\right)^2,$$

where E_{ij} is the observed image irradiance at the grid point (i, j). We seek a set of values $\{f_{ij}\}$ and $\{g_{ij}\}$ that minimize

$$e = \sum_i \sum_j (s_{ij} + \lambda r_{ij}).$$

Differentiating e with respect to f_{kl} and g_{kl}, we obtain

$$\frac{\partial e}{\partial f_{kl}} = 2(f_{kl} - \overline{f}_{kl}) - 2\lambda\left(E_{kl} - R_s(f_{kl}, g_{kl})\right) \frac{\partial R_s}{\partial f},$$

$$\frac{\partial e}{\partial g_{kl}} = 2(g_{kl} - \overline{g}_{kl}) - 2\lambda\left(E_{kl} - R_s(f_{kl}, g_{kl})\right) \frac{\partial R_s}{\partial g},$$

where $\overline{f}$ and $\overline{g}$ are local averages of f and g:

$$\overline{f}_{i,j} = \frac{1}{4}\left(f_{i+1,j} + f_{i,j+1} + f_{i-1,j} + f_{i,j-1}\right),$$

$$\overline{g}_{i,j} = \frac{1}{4}\left(g_{i+1,j} + g_{i,j+1} + g_{i-1,j} + g_{i,j-1}\right).$$

We have to be careful when performing this differentiation, since f_{kl} and g_{kl} occur in four terms of the sum; the new subscripts k and l are introduced to avoid confusion with the subscripts i and j that occur in the sum.

The extremum is to be found where the above derivatives of e are equal to zero. If we rearrange the resulting equations by solving for f_{kl} and g_{kl}, an iterative solution method suggests itself:

$$f_{kl}^{n+1} = \overline{f}_{kl}^{n} + \lambda\big(E_{kl} - R_s(f_{kl}^n, g_{kl}^n)\big)\frac{\partial R_s}{\partial f},$$
$$g_{kl}^{n+1} = \overline{g}_{kl}^{n} + \lambda\big(E_{kl} - R_s(f_{kl}^n, g_{kl}^n)\big)\frac{\partial R_s}{\partial g},$$

where the new values for f and g at each grid point are obtained using the old values of f and g in evaluating $R_s(f,g)$, $\partial R_s/\partial f$, and $\partial R_s/\partial g$. It can be shown that a stable method is obtained if we use the local averages $\overline{f}$ and $\overline{g}$ in evaluating $R_s(f,g)$ and the two partial derivatives, provided suitable boundary conditions are introduced and λ is small enough.

The simple iterative scheme described above can be improved in various ways. The estimates of the Laplacians of f and g, proportional to $(\overline{f}_{kl} - f_{kl})$ and $(\overline{g}_{kl} - g_{kl})$, can be replaced by more accurate formulae, for example. Then the local average is computed as

$$\overline{f}_{i,j} = \frac{1}{5}\left(f_{i+1,j} + f_{i,j+1} + f_{i-1,j} + f_{i,j-1}\right) + \frac{1}{20}\left(f_{i+1,j+1} + f_{i+1,j-1} + f_{i-1,j-1} + f_{i-1,j+1}\right),$$

and similarly for $\overline{g}_{i,j}$. The computation of the average can be represented by the stencil

$$\frac{1}{20}\begin{array}{|c|c|c|}\hline 1 & 4 & 1 \\ \hline 4 & & 4 \\ \hline 1 & 4 & 1 \\ \hline\end{array}$$

which is derived directly from one we used earlier to approximate the Laplacian operator.

Figure 11-10 shows a picture made from the image of a small resin droplet obtained by means of a scanning electron microscope. To apply the iterative shape-from-shading scheme, we have to know the reflectance map. A commonly used model of secondary electron emission from a surface suggests that brightness should vary as the secant of the incident angle, $\sec\theta_i$. Using this model, we obtained the shape shown in figure 11-11.

11.7.3 Application to Photometric Stereo

The gradient values computed at adjacent image points using the photometric stereo method are not necessarily consistent. Even in the case of a

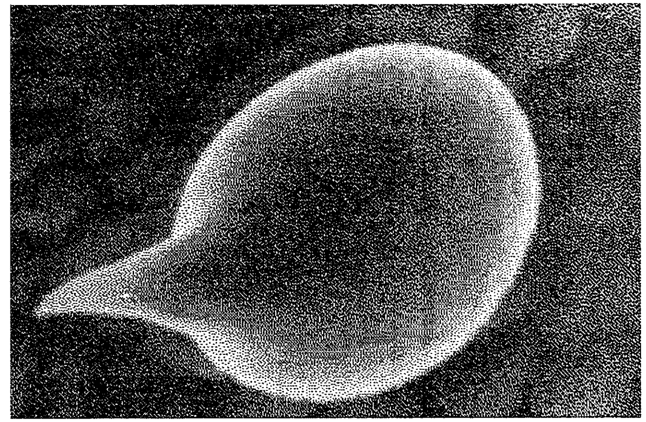

Figure 11-10. Display of the image of a small resin droplet on a flower of a *Cannabis sativa* plant. (Reproduced by permission from the book *Magnifications—Photography with the Scanning Electron Microscope* by David Scharf, Schocken Books, New York, 1977.)

planar surface, there can be fluctuations in estimated surface orientation due to measurement errors. If we know that the surface is smooth, we can use the method presented in this chapter to improve the results of the photometric stereo method.

If there are n images, we can formulate this problem in terms of the minimization of

$$e = \iint_I \left((f_x^2 + f_y^2) + (g_x^2 + g_y^2)\right)\,dx\,dy + \sum_{i=1}^{n}\lambda_i \iint_I \left(E_i(x,y) - R_i(f,g)\right)^2\,dx\,dy,$$

where E_i is the brightness measured in the i^{th} image and R_i is the corresponding reflectance map. The constant multipliers λ_i are parameters that weight the errors in the image irradiance equations relative to the departure from smoothness. They are unequal if the information provided by the cameras is not all equally reliable.

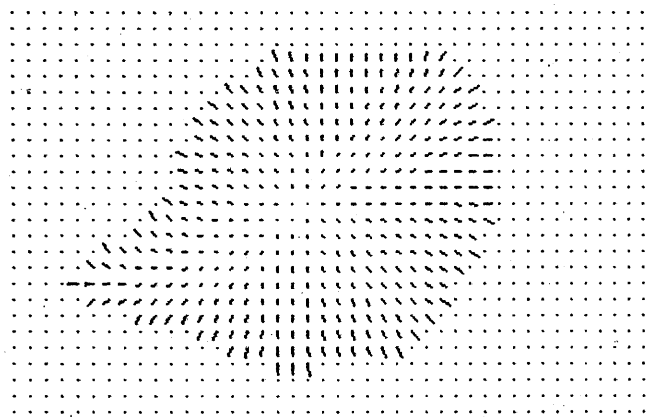

Figure 11-11. Needle diagram calculated by the iterative scheme under the assumption that the reflectance map is $\sec\theta_i$. (The surface orientation data are actually available on a finer grid; they are sampled coarsely here for display purposes.) The needle diagram is the estimate of the shape of the surface of the resin droplet shown in the previous figure. (Figure kindly provided by Katsushi Ikeuchi.)

The Euler equations in this case yield

$$\nabla^2 f = -\sum_{i=1}^{n} \lambda_i \big(E_i(x,y) - R_i(f,g)\big) \frac{\partial R_i}{\partial f},$$

$$\nabla^2 g = -\sum_{i=1}^{n} \lambda_i \big(E_i(x,y) - R_i(f,g)\big) \frac{\partial R_i}{\partial g}.$$

The corresponding discrete equations suggest an iterative scheme:

$$f_{kl}^{n+1} = \overline{f}_{kl}^{n} + \sum_{i=1}^{n} \lambda_i \big(E_{i;kl} - R_i(f_{kl}, g_{kl})\big) \frac{\partial R_i}{\partial f},$$

$$g_{kl}^{n+1} = \overline{g}_{kl}^{n} + \sum_{i=1}^{n} \lambda_i \big(E_{i;kl} - R_i(f_{kl}, g_{kl})\big) \frac{\partial R_i}{\partial g}.$$

The simple photometric stereo method discussed in the previous chapter can be used to obtain good initial values for $\{f_{kl}\}$ and $\{g_{kl}\}$. This will ensure rapid convergence to the solution.

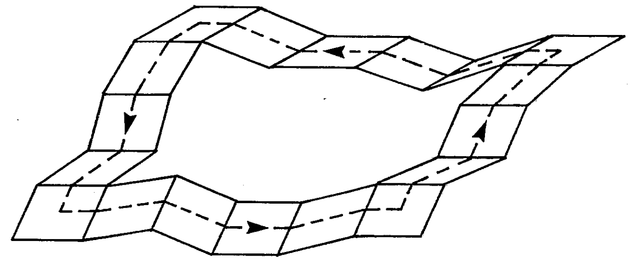

Figure 11-12. If surface orientation is known, elevation of the surface above some reference plane can be determined by integration along curves in the image. To be consistent, the integral of the surface gradient along a closed curve should be zero, since the overall change in elevation when one walks in a closed loop on a single-valued surface is zero.

11.8 Recovering Depth from a Needle Diagram

Several machine vision methods, including photometric stereo, produce surface shape information in the form of a *needle diagram*, in which surface orientation is given for every picture cell. In some cases we may want to represent surface shape in a different way. Often a *depth map*, giving height above some reference plane, is a desirable objective.

Given p and q, the partial derivatives of $z(x,y)$ with respect to x and y, we can recover $z(x,y)$ by integrating along arbitrary curves in the plane

$$z(x,y) = z(x_0,y_0) + \int_{(x_0,y_0)}^{(x,y)} (p\,dx + q\,dy).$$

In practice, p and q are recovered from noisy image data by imperfect methods. Thus the above integral might depend on the path chosen. Indeed, an integral along a closed path, as shown in figure 11-12, can be different from zero. Now since both p and q are available, we actually have more information than we really need. This suggests that we use a least-squares method to find the surface that best fits the imperfect estimate of the surface gradient.

We can, for example, choose $z(x,y)$ so as to minimize the error

$$\iint_I \big((z_x - p)^2 + (z_y - q)^2\big)\, dx\, dy,$$

where p and q are the given estimates of the components of the gradient, while z_x and z_y are the partial derivatives of the best-fit surface. Again, this is a problem in the calculus of variations. We have to minimize an integral of the form

$$\iint F(z, z_x, z_y)\, dx\, dy.$$

The Euler equation is

$$F_z - \frac{\partial}{\partial x} F_{z_x} - \frac{\partial}{\partial y} F_{z_y} = 0,$$

so that from

$$F = (z_x - p)^2 + (z_y - q)^2$$

we obtain

$$\frac{\partial}{\partial x}(z_x - p) + \frac{\partial}{\partial y}(z_y - q) = 0,$$

or just

$$\nabla^2 z = p_x + q_y.$$

This equation accords with intuition, since it states that the Laplacian of the desired surface must equal $p_x + q_y$, which is an estimate of the Laplacian based on the given data.

As usual, we also need to know what to do about the boundary of the region over which this equation is to be solved. The *natural boundary condition* (see appendix) for an integral of the form

$$\iint F(z, z_x, z_y)\, dx\, dy$$

is

$$F_{z_x} \frac{dy}{ds} - F_{z_y} \frac{dx}{ds} = 0.$$

Here s is arclength along the boundary. We note at this point that $(dx/ds, dy/ds)^T$ is a tangent vector. In our case we obtain

$$(z_x - p)\frac{dy}{ds} = (z_y - q)\frac{dx}{ds},$$

or

$$(z_x, z_y)^T \cdot \left(\frac{dy}{ds}, -\frac{dx}{ds}\right)^T = (p, q)^T \cdot \left(\frac{dy}{ds}, -\frac{dx}{ds}\right)^T,$$

where the vector

$$\left(\frac{dy}{ds}, -\frac{dx}{ds}\right)^T$$

is a normal to the boundary curve at the point s. This result is eminently reasonable, since it states that the normal derivative of the desired surface must equal the estimate of the normal derivative obtained from the data.

An iterative method can be used to solve this equation. It can be based on a discrete approximation of the equation or a least-squares analysis applied directly to a discrete approximation of the original error integral, thus sidestepping the need for application of the calculus of variations.

Initial values can be generated by some simple scheme, such as integrating $p(x, 0)$ along the x-axis to obtain one profile, then integrating $q(x_0, y)$ along y starting at each point x_0 on the x-axis. (This crude method by itself, of course, does not produce a particularly good surface.)

11.9 References

The scanning electron microscope is described in *Scanning Electron Microscopy* by Wells [1974]. Many interesting pictures made using such instruments are shown in *Tissues and Organs: a Text-Atlas of Scanning Electron Microscopy* by Kessel & Kardon [1979] and in *Magnifications—Photography with the Scanning Electron Microscope* by Scharf [1977].

There are numerous books discussing partial differential equations, among them *Partial Differential Equations: Theory and Technique* by Carrier & Pearson [1976] and volume II of *Methods of Mathematical Physics* by Courant & Hilbert [1962]. But perhaps the most relevant for the first-order equations explored in this chapter is Garabedian's *Partial Differential Equations* [1964].

The calculus of variations is also the topic of many books, including *Calculus of Variations: With Applications to Physics & Engineering* by Weinstock [1974] and volume I of *Methods of Mathematical Physics* by Courant & Hilbert [1953]. In some of the exercises we relate the methods used in this chapter to *regularization* techniques for producing well-posed problems from ill-posed ones. Regularization is discussed by Tikhonov & Arsenin in *Solutions of Ill-Posed Problems* [1977].

There was a lot of interest in determining the shape of the surface features of the moon from telescopic images taken from the earth, at least until we could send probes, and finally people, to the vicinity of our rocky satellite. Since the libation of the moon, as well as the ratio of the radius of the earth to the distance between the two bodies, is small, we always see the moon from essentially the same direction. Thus binocular stereo can be ruled out as a viable method for recovering surface shape in this instance. Astronomers used shadows to estimate the relief of crater edges above the surrounding terrain. Van Digellen [1951] was the first to suggest the possibility of using shading, but he was only able to make some heuristic estimates of surface slope in the direction of the light source. Rindfleisch [1966] used photometric models developed in Russia by Fesenkov [1962] and others to derive a complex integration method for recovering the shape along profiles that we now know to be characteristic lines.

Horn [1970, 1975a] found the general solution of the shape-from-shading problem and later [1977] reworked the solution to make use of the reflectance map. Woodham [1979, 1981, 1984] provides excellent discussions of this topic, using the Hessian matrix as a tool.

The existence and uniqueness of solutions to the nonlinear first-order partial differential equation were explored by Bruss [1981, 1982, 1983], Brooks [1982], and Deift & Sylvester [1981].

Horn [1970] searched for a way to reformulate the problem so that the solution would take the form of a parallel iterative algorithm on a grid, much like the one he later used for the computation of lightness [1974]. Strat [1979] developed the first such algorithm. This algorithm was not, however, able to deal with the occluding boundary, since it used the gradient to parameterize surface orientation. Ikeuchi & Horn [1981] rectified this problem by introducing stereographic coordinates and a "lack-of-smoothness" term, now recognized as a regularization term. (For other ideas on parallel computation in vision, see Ballard, Hinton, & Sejnowshi [1983]. For a discussion of the use of regularization in dealing with ill-posed early vision problems, see Poggio & Torre [1984].)

Unfortunately the method of Ikeuchi and Horn, in turn, did not guarantee the integrability of the resulting needle diagram. Horn & Brooks [1985] have remedied this deficiency by using the surface normal to parameterize surface orientation. They avoid the use of a regularizing term in their work.

Shape can also be calculated from texture gradients or regular patterns; see Horn [1970], Bajcsy & Lieberman [1976], Witkin [1981], and Ikeuchi [1984]. These methods are inherently simpler, however, since more information is available at each image point than when shading is used.

Most shape-from-shading methods require that the reflectance map be given. There have been attempts to reduce dependence on such detailed knowledge. Pentland [1984], for example, tries to extract information locally. This inevitably requires strong assumptions, such as that the surface is spherical. Local methods cannot lead to unique results, since it is known from the work of Bruss [1981, 1982, 1983] and Brooks [1982] that singular points and occluding boundaries provide strong constraints, which are not available to a method that only considers shading in a small region of the image.

A compromise between exact knowledge of the reflectance map and not knowing anything at all is the use of a parameterized reflectance map. It is possible, for example, to recover the position of the light source from the image when certain assumptions are made. See Pentland [1982], Lee [1983], and Brooks & Horn [1985].

One can say something about the relationship between shading and surface shape without detailed solution of the image irradiance equation. For an example, see Koenderink & van Doorn [1980]. The silhouette of the image of an object provides a great deal of information. Marr [1977] was concerned with the recovery of surface shape information from the occluding contour alone. Stevens [1981] considered the recovery of surface shape from special contours on the surface.

Much work has been devoted to understanding the interaction of light with the surface layer of an object. Typically a surface layer model is either so complex that only numerical results can be obtained or so simple as to be unrealistic. The photometric model of the lunar surface was refined by Minnaert [1961], whose reflectance function is used in exercise 11-1. Useful models for glossy reflection were developed by Torrance & Sparrow [1967] and Trowbridge & Reitz [1975].

11.10 Exercises

11-1 Consider a sphere of Lambertian material with center on the optical axis. Assume that the light source is a point source of unit intensity in the direction $(-p_s, -q_s, 1)^T$.

(a) What is the irradiance $E(x, y)$? Assume that the radius of the sphere is R and that the image is obtained using orthographic projection.

(b) Show that contours of constant brightness in the image are nested ellipses of equal eccentricity. Hint: What are the contours of constant brightness on the sphere?

(c) What do the contours of constant brightness in the image look like if we assume instead that the surface has the reflectance properties of the material in the maria of the moon? Hint: What are the contours of constant brightness on the sphere in this case?

11-2 Show that the slope of a surface in the direction that makes an angle θ with the x-axis is

$$m(\theta) = p\cos\theta + q\sin\theta.$$

Find the direction of steepest ascent. Conclude that the slope in the direction of steepest ascent is

$$m(\theta_s) = \sqrt{p^2 + q^2}.$$

11-3 Show that the two surfaces

$$z_1 = 2(x^2 + xy + y^2) \qquad \text{and} \qquad z_2 = (x^2 + 4xy + y^2)$$

give rise to the same shading near the origin if a rotationally symmetric reflectance map applies.

11-4 When are the base characteristics parallel straight lines in the image plane? That is, what class of reflectance maps lead to solutions that have this property, independent of the shape of the surface? Hint: For the base characteristics to be straight lines there must be a proportionality between $\dot{x}$ and $\dot{y}$ in the equations for the characteristic strip.

11-5 Suppose that the reflectance map is linear in p and q, so that

$$R(p,q) = ap + bq + c.$$

We have an image, including the silhouette of a simple convex object of shape $z = f(x,y)$. Show that the surface

$$\overline{z} = f(x,y) + g(bx - ay),$$

for an arbitrary differentiable function $g(s)$, will give rise to the same image. Does the surface $\overline{z}$ have the same silhouette? Assume that the derivative of g is bounded.

11-6 Scanning electron microscope images are unusual in that surface patches inclined relative to the viewer are brighter than a patch that is orthogonal to the viewing direction. One might imagine that simply printing such images in negative form would improve their interpretability. This is not the case. Explain why. Hint: "Shadows" in scanning electron microscope images are dark.

11-7 Here we explore the importance of singular points in reducing the ambiguity in the shape-from-shading problem. Suppose that we have a paraboloid defined by the equation

$$z(x,y) = z_0 + \frac{1}{2}(x^2 + y^2)$$

and a reflectance map

$$R(p,q) = p^2 + q^2.$$

(a) Show that the image can be written

$$E(x,y) = x^2 + y^2$$

and that there is a singular point at the origin.

(b) Now demonstrate that the image irradiance equation that applies here can be expressed in polar coordinates as

$$z_r^2 + \frac{1}{r^2} z_\theta^2 = r^2,$$

where z_r and z_θ are the partial derivatives of z with respect to r and θ, respectively.

(c) Show that the solution of this equation is the paraboloid we started with, provided that $z_\theta = 0$. (We cannot, of course, distinguish a convex paraboloid from a concave paraboloid, or recover its absolute distance from the viewer.)

(d) Now suppose instead that $z_\theta = k$. Show that the solution in this case is

$$z(r) = z_0 \pm \int_{r_0}^{r} \sqrt{t^2 - \frac{k^2}{t^2}}\, dt.$$

Over what range of r is this expression valid?

(e) Perform the indicated integration and conclude that

$$z(r) = z_0 \pm \frac{1}{2}\left(\sqrt{r^4 - k^2} + k \sin^{-1}\frac{k}{r^2}\right).$$

(f) The solution is unique if the image includes the singular point. Is the solution unique if we are only given the part of the image in an annular ring around the singular point? What if we are given the image in a simply connected region that does not include the singular point?

11-8 For a surface of low inclination (that is, small z_x and z_y), the Gaussian curvature and the mean curvature are given by

$$\kappa_1\kappa_2 \approx z_{xx}z_{yy} - z_{xy}z_{yx} \quad \text{and} \quad \frac{1}{2}(\kappa_1 + \kappa_2) \approx \frac{1}{2}(z_{xx} + z_{yy}).$$

Express these results in terms of the Hessian matrix $\mathbf{H}$.

11-9 The Gaussian curvature of a surface $z(x,y)$ can be written in the form

$$\kappa = \frac{z_{xx}z_{yy} - z_{xy}z_{yx}}{\left(1 + z_x^2 + z_y^2\right)^2}.$$

Rewrite this in terms of p and q as well as r, s, and t as defined in this chapter.

A *ruled surface* is one that can be generated by sweeping a straight line, called the *generator*, through space. At each point on a ruled surface, we can find a tangent that lies in the surface (the generator for that part of the surface). A hyperboloid is an example of a ruled surface. *Developable surfaces* constitute a subclass of ruled surfaces. Intuitively, a developable surface is one that can be cut open and flattened out. In the case of a developable surface, all points on a tangent that lies in the surface have a common normal direction. Cylindrical and conical surfaces are examples of developable surfaces (figure 11-13). The Gaussian curvature on a developable surface is everywhere zero.

Suppose we are told that a surface we are viewing happens to be developable. Show that we can perform a local shading analysis to recover the second partial derivatives of the surface. Specifically, show that

$$r = \frac{E_x^2}{R_p E_x + R_q E_y}, \quad s = \frac{E_x E_y}{R_p E_x + R_q E_y}, \quad t = \frac{E_y^2}{R_p E_x + R_q E_y}.$$

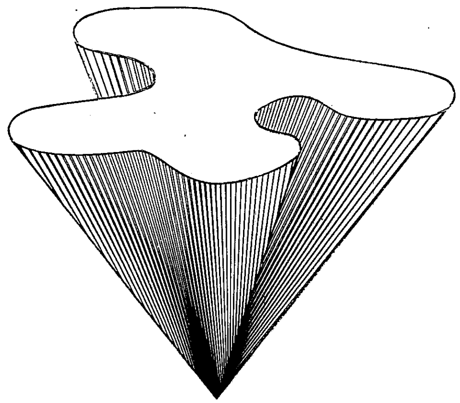

Figure 11-13. A developable surface, such as a cone, has zero Gaussian curvature everywhere.

11-10 Here we consider in detail what happens near a singular point in the case of a simple reflectance map. Suppose that the surface shape near the origin is given by

$$z = z_0 + \frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Let the reflectance map be

$$R(p, q) = \frac{1}{2}(p^2 + q^2).$$

We wish to determine exactly how many solutions there are locally for surface shape, and how they are related to one another.

(a) Show that

$$E(x, y) = \frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} a & b \\ b & c \end{pmatrix}^2\begin{pmatrix} x \\ y \end{pmatrix}.$$

(b) Show that

$$\begin{pmatrix} E_{xx} & E_{xy} \\ E_{yx} & E_{yy} \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}^2 = \begin{pmatrix} a^2+b^2 & (a+c)b \\ (a+c)b & b^2+c^2 \end{pmatrix}.$$

Next, we evaluate the second partial derivatives in a rotated coordinate system. Suppose that

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

For convenience, we call the rotation matrix in this equation $\mathbf{R}(\theta)$. We shall also use the shorthand notation $c = \cos\theta$ and $s = \sin\theta$.

(c) Show that

$$\begin{pmatrix} E_{x'x'} & E_{x'y'} \\ E_{y'x'} & E_{y'y'} \end{pmatrix} = \mathbf{R}(\theta)\begin{pmatrix} E_{xx} & E_{xy} \\ E_{yx} & E_{yy} \end{pmatrix}\mathbf{R}^T(\theta)$$
$$= \begin{pmatrix} c^2E_{xx} + 2sc\,E_{xy} + s^2E_{yy} & (c^2-s^2)E_{xy} - sc(E_{xx}-E_{yy}) \\ (c^2-s^2)E_{xy} - sc(E_{xx}-E_{yy}) & s^2E_{xx} - 2sc\,E_{xy} + c^2E_{yy} \end{pmatrix}.$$

Hint: Use the chain rule for differentiation.

(d) Show that the off-diagonal elements, $E_{x'y'}$ and $E_{y'x'}$, are zero when

$$\tan 2\theta = \frac{2E_{xy}}{E_{xx} - E_{yy}}.$$

Find expressions for $\sin 2\theta$ and $\cos 2\theta$.

(e) Show that

$$\cos^2\theta = \frac{1}{2}\frac{\sqrt{(E_{xx}-E_{yy})^2 + 4E_{xy}^2} + (E_{xx}-E_{yy})}{\sqrt{(E_{xx}-E_{yy})^2 + 4E_{xy}^2}},$$

$$\sin^2\theta = \frac{1}{2}\frac{\sqrt{(E_{xx}-E_{yy})^2 + 4E_{xy}^2} - (E_{xx}-E_{yy})}{\sqrt{(E_{xx}-E_{yy})^2 + 4E_{xy}^2}}.$$

Conclude that

$$E_{x'x'} = \frac{1}{2}\left(1 + \sqrt{(E_{xx}-E_{yy})^2 + 4E_{xy}^2}\right),$$

$$E_{y'y'} = \frac{1}{2}\left(1 - \sqrt{(E_{xx}-E_{yy})^2 + 4E_{xy}^2}\right).$$

Hint: The eigenvalues and eigenvectors of a symmetric 2×2 matrix are developed in exercise 3-5.

(f) Show that the second-order polynomial for z does not contain a cross-term in the rotated coordinate system. That is, we can write

$$z = a'(x')^2 + c'(y')^2.$$

Find a' and c' in terms of a, b, and c.

(g) Show that there are exactly four surfaces that have the observed second-order partial derivatives of image brightness. How are they related? Hint: What are the Gaussian curvatures of the solutions?

11-11 A surface $z(x, y)$ with continuous second partial derivatives has to satisfy an *integrability constraint*, that is,

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x},$$

or $p_y = q_x$. The iterative shape-from-shading scheme presented in this chapter does not guarantee this. Suppose you wish to minimize the brightness error,

$$\iint_I \big(E(x, y) - R(p, q)\big)^2 \, dx \, dy,$$

by suitable choice of the two functions $p(x, y)$ and $q(x, y)$, subject to the integrability constraint $p_y = q_x$.

(a) Show that the appropriate Euler equations are

$$\big(E(x, y) - R(p, q)\big) R_q = -\lambda_x \quad \text{and} \quad \big(E(x, y) - R(p, q)\big) R_p = +\lambda_y,$$

where $\lambda(x, y)$ is a Lagrange function and

$$R_p = \frac{\partial R}{\partial p}, \qquad R_q = \frac{\partial R}{\partial q}, \qquad \lambda_x = \frac{\partial \lambda}{\partial x}, \qquad \lambda_y = \frac{\partial \lambda}{\partial y}.$$

(b) Conclude that the desired functions $p(x, y)$ and $q(x, y)$ must satisfy the equation

$$\big((E-R)R_{pp} - R_p^2\big) p_x + \big((E-R)R_{pq} - R_p R_q\big)(p_y + q_x) + \big((E-R)R_{qq} - R_q^2\big) q_y \\ = (E_x R_p + E_y R_q),$$

as well as the constraint $p_y = q_x$. Hint: Total derivatives with respect to x and y are required in order to eliminate the Lagrange multiplier.

(c) Show that you end up with the same Euler equation if you try to minimize

$$\iint_I (E(x, y) - R(z_x, z_y))^2 \, dx \, dy$$

by suitable choice of $z(x, y)$.

11-12 Suppose instead that you wish to minimize the sum of the brightness error and the deviation from integrability,

$$\iint_I \Big(\big(E(x, y) - R(p, q)\big)^2 + \lambda(p_y - q_x)^2\Big) \, dx \, dy,$$

by suitable choice of the two functions $p(x, y)$ and $q(x, y)$.

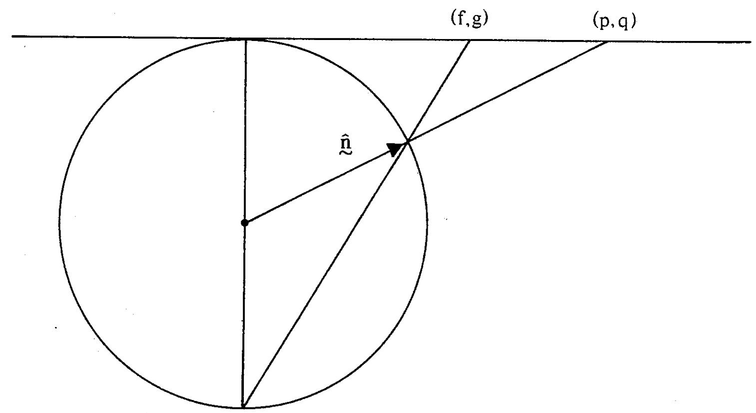

Figure 11-14. Cross section of the Gaussian sphere and a plane tangent at the north pole. The gnomonic projection is obtained by connecting points on the surface of the sphere to its center. The stereographic projection is obtained by connecting points on the surface of the sphere to its south pole.

(a) How does this approach differ from that of the previous problem? Hint: Note that λ here is a constant.

(b) Show that the appropriate Euler equations are

$$\big(E(x, y) - R(p, q)\big) R_p = \lambda(p_{yy} - q_{xy}),$$

$$\big(E(x, y) - R(p, q)\big) R_q = \lambda(q_{xx} - p_{xy}).$$

(c) Suggest an iterative scheme based on isolation of the central values in discrete approximations of the second-order derivatives p_{yy} and q_{xx}.

11-13 Here we explore the stereographic projection of the Gaussian sphere. A cross section through the Gaussian sphere useful for understanding the gnomonic and stereographic projections is shown in figure 11-14.

(a) Show that the relationship between the stereographic projection and the gnomonic projection can be expressed in the form

$$f = \frac{2p}{1 + \sqrt{1 + p^2 + q^2}}, \qquad g = \frac{2q}{1 + \sqrt{1 + p^2 + q^2}}.$$

(b) Show that

$$p = \frac{4f}{4 - f^2 - g^2} \quad \text{and} \quad q = \frac{4g}{4 - f^2 - g^2}.$$

(c) Further, show that the integrability condition $p_y = q_x$ can be expressed in terms of stereographic coordinates as

$$f_y(4 + f^2 - g^2) - g_x(4 - f^2 + g^2) + 2fg(g_y - f_x) = 0.$$

11-14 Show that

$$\nabla \times (-p, -q, 1)^T = 0,$$

where $\nabla \times \mathbf{n}$ is the *curl* of the vector $\mathbf{n}$. Hint: First show that

$$(-p, -q, 1)^T = \nabla\big(z_0 - z(x, y)\big),$$

where $\nabla f(x, y, z)$ is the gradient

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)^T$$

of the function $f(x, y, z)$.

11-15 Suppose that we have calculated a discrete needle map $\{(p_{ij}, q_{ij})\}$ by means of some shape-from-shading method. We now wish to recover the surface $\{z_{ij}\}$. To do this, we minimize

$$\sum_{i=1}^{n} \sum_{j=1}^{m} (z_x - p_{ij})^2 + (z_y - q_{ij})^2$$

by suitable choice of $\{z_{ij}\}$. Use the estimates

$$z_x \approx \frac{1}{2h}\big(z_{i,j+1} - z_{i,j} + z_{i+1,j+1} - z_{i+1,j}\big),$$

$$z_y \approx \frac{1}{2h}\big(z_{i+1,j} - z_{i,j} + z_{i+1,j+1} - z_{i,j+1}\big),$$

for the derivatives. Note that these estimates are unbiased for the corners where four picture cells meet, not for picture-cell centers. Show that a necessary condition for a minimum is that

$$\frac{1}{h}\left(4z_{k,l} - (z_{k+1,l-1} + z_{k-1,l+1} + z_{k+1,l+1} + z_{k-1,l-1})\right)$$

$$= \big(p_{k,l-1} - p_{k,l} + p_{k-1,l-1} - p_{k-1,l}\big) + \big(q_{k-1,l} - q_{k,l} + q_{k-1,l-1} - q_{k,l-1}\big).$$

Compare this result to that obtained in the continuous case in this chapter using the calculus of variations.

11-16 So far we have assumed that we know where the light sources are and that we do not know the shape of the object. At times we may in fact know the shape of some object in the scene, but not where the light is coming from. Assume that the surface is Lambertian and that it is illuminated by a distant point source. How can we recover the direction $\mathbf{s}$ to that source?

(a) Minimize the integral of the square of the difference between the observed and the predicted brightness,

$$\iint_I \big(E(x, y) - \mathbf{n}(x, y) \cdot \mathbf{s}\big)^2 \, dx\, dy.$$

(b) Show that

$$\iint_I E\mathbf{n}\, dx\, dy = \left[\iint_I \mathbf{n}\mathbf{n}^T\, dx\, dy\right] \mathbf{s},$$

where $\mathbf{n}\mathbf{n}^T$ is the dyadic product of $\mathbf{n}$ with itself, which is a 3×3 matrix.

(c) Conclude that the unknown source position can be computed as follows:

$$\mathbf{s} = \left[\iint_I \mathbf{n}\mathbf{n}^T\, dx\, dy\right]^{-1} \iint_I E\mathbf{n}\, dx\, dy.$$

(d) Develop a suitable discrete scheme based on this analysis. Replace the integrals by sums.

(e) How many points on the surface of the object must be measured to ensure that the sum of the dyadic products is a nonsingular matrix? Warning: This part is harder than the rest of the problem.

(f) Can the same sort of thing be done for a surface whose brightness is proportional to $\sqrt{\cos\theta_i / \cos\theta_e}$? Hint: In this case, it may be inconvenient to minimize the difference between the observed and the predicted brightness. Instead, minimize a related quantity in order to make sure that the resulting equations are tractable.