# 5

# *Sources, Shadows, and Shading*

Surfaces are bright or dark for two main reasons: their albedo and the amount of light they are receiving. A model of how the brightness of a surface is obtained is usually called a *shading model*. Shading models are important because with an appropriate shading model we can interpret pixel values. If the right shading model applies, it is possible to reconstruct objects and their albedos using just a few images. Furthermore, we can interpret shadows and explain their puzzling and seldom-noticed absence in most indoor scenes.

## 5.1 QUALITATIVE RADIOMETRY

We should like to know how "bright" surfaces are going to be under various lighting conditions, and how this "brightness" depends on local surface properties, on surface shape, and on illumination. As we saw in chapter 4, foreshortening means that different sources can have the same effect on a surface. The most powerful tool for analyzing this problem is to think about *what a source looks like from the surface*. This *qualitative radiometry* is one of these tricks that looks unsophisticated—no hard math—but is extremely powerful. In some cases, this technique gives qualitative descriptions of "brightness" without even knowing what the term means.

Recall from Section 4.1.1 and Figure 4.1 that a surface patch sees the world through a hemisphere of directions at that patch. The radiation arriving at the surface along a particular direction passes through a point on the hemisphere. If two surface patches have equivalent incoming hemispheres, they must have the same incoming radiation, *whatever the outside world looks like*. This means that any difference in "brightness" between patches with the same incoming hemisphere is a result of different surface properties. In particular, if two surface patches with the same BRDF see the same incoming hemisphere, then the radiation they output must be the same.
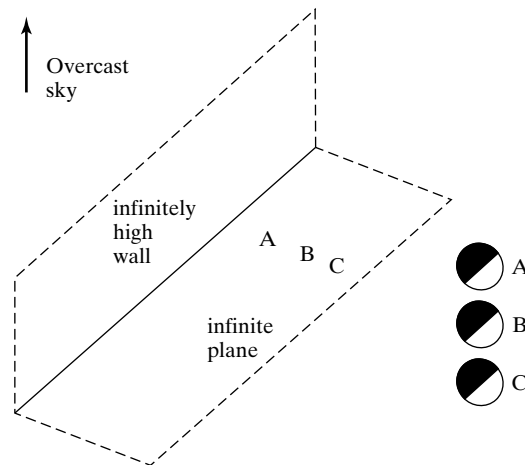
**Figure 5.1**   A geometry in which a qualitative radiometric solution can be obtained by thinking about what the world looks like from the point of view of a patch. We wish to know what the brightness looks like at the base of two different infinitely high walls. In this geometry, an infinitely high matte black wall cuts off the view of the overcast sky—which is a hemisphere of infinite radius and uniform "brightness". On the right, we show a representation of the directions that see or do not see the source at the corresponding points, obtained by flattening the hemisphere to a circle of directions (or, equivalently, by viewing it from above). Since each point has the same input hemisphere, the brightness must be uniform.

Lambert determined the distribution of "brightness" on a uniform plane at the base of an infinitely high black wall illuminated by an overcast sky (see Figure 5.1). In this case, every point on the plane must see the same hemisphere—half of its viewing sphere is cut off by the wall, and the other half contains the sky, which is uniform—and the plane is uniform, so every point must have the same "brightness".

A second example is somewhat trickier. We now have an infinitely thin black wall that is infinitely long in only one direction and on an infinite plane (Figure 5.2). A qualitative description would be to find what the curves of equal "brightness" look like. It is fairly easy to see that all points on any line passing through the point $p$ in Figure 5.2 see the same input hemisphere and so must have the same "brightness". Furthermore, the distribution of "brightness" on the plane must have a symmetry about the line of the wall—we expect the brightest points to be along the extension of the line of the wall and the darkest to be at the base of the wall.

## 5.2 SOURCES AND THEIR EFFECTS

### 5.2.1 Radiometric Properties of Light Sources

We define a *light source* to be anything that emits light *that is internally generated* (i.e., not just reflected). To describe a source, we need a description of the radiance it emits in each direction. Typically, internally generated radiance is dealt with separately from reflected radiance. This is because, although a source may reflect light, the light it reflects depends on the environment, whereas the light it generates internally usually does not.
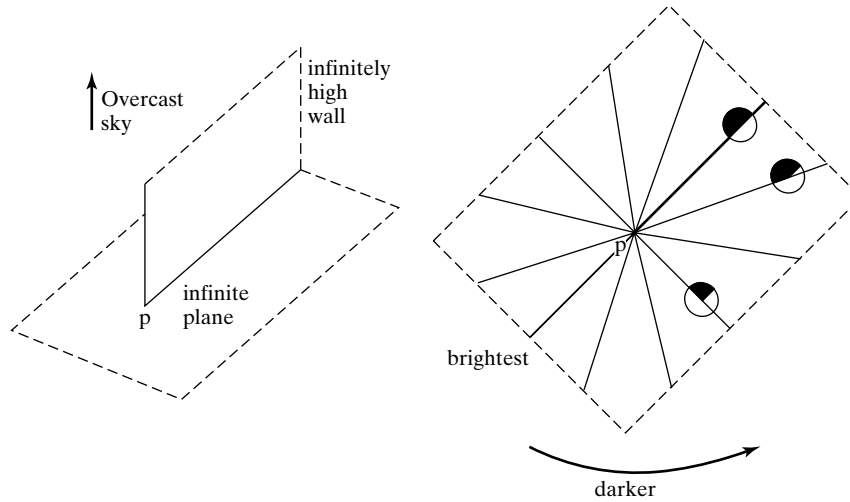
**Figure 5.2**   We now have a matte black, infinitely thin, half-infinite wall on an infinite white plane (shown on the **left**). This geometry also sees an overcast sky of infinite radius and uniform "brightness". In the text, we show how to determine the curves of similar "brightness" on the plane. These curves are shown on the **right**, depicted on an overhead view of the plane; the thick line represents the wall. Superimposed on these curves is a representation of the input hemisphere for some of these isophotes. Along these curves, the hemisphere is fixed (by a geometrical argument), but they change as one moves from curve to curve.

We seldom need a complete description of the radiance a source emits in each direction. It is more usual to model sources as emitting a constant radiance in each direction (possibly with a family of directions zeroed, like a spotlight). The proper quantity in this case is the *exitance*, defined as

the internally generated energy radiated per unit time and per unit area on the radiating surface.

Exitance is similar to radiosity, and can be computed as

$$E(P) = \int_{\Omega} L_e(P, \theta_o, \phi_o) \cos \theta_o \, d\omega,$$

Together with a description of the exitance, we need a description of the geometry of the source, which has profound effects on the spatial variation of light around the source and on the shadows cast by objects near the source. Sources are usually modeled with quite simple geometries for two reasons: first, many synthetic sources can be modeled as point sources, line sources, or area sources fairly effectively; second, sources with simple geometries can still yield surprisingly complex effects.

### 5.2.2  Point Sources

A common approximation is to assume that the light source is an extremely small sphere, in fact, a point; such a source is known as a *point source*. It is a natural model to use because many sources are physically small compared with the environment in which they stand. We can obtain
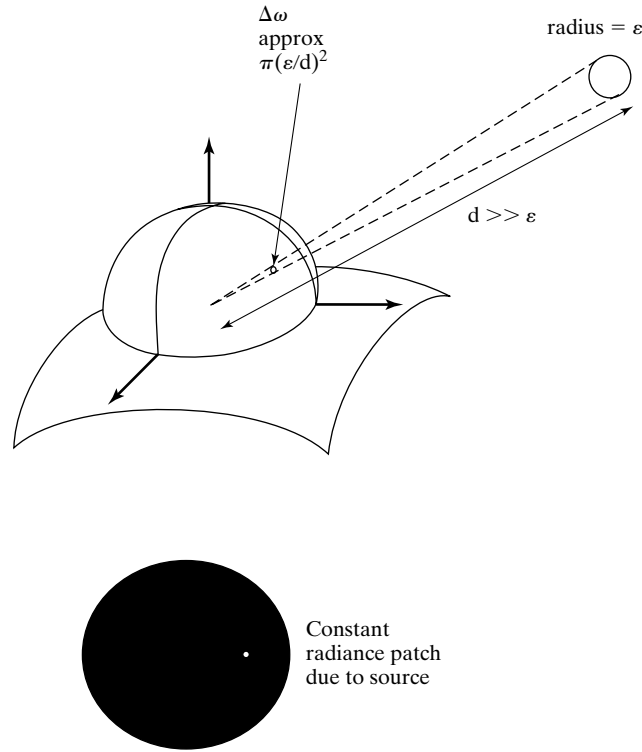
**Figure 5.3**   A surface patch sees a distant sphere of small radius; the sphere produces a small illuminated patch on the input hemisphere of the sphere. In the text, by reasoning about the scaling behavior of this patch as the distant sphere moves farther away or gets bigger, we obtain an expression for the behavior of the point source.

a model for the effects of a point source by modeling the source as a very small sphere that emits light at each point on the sphere, with an exitance constant over the sphere.

Assume that a surface patch is viewing a sphere of radius $\epsilon$, at a distance $r$ away, and that $\epsilon \ll r$ (Figure 5.3). The assumption that the sphere is far away from the patch relative to its radius almost always applies for real sources. Now the solid angle that the source subtends is $\Omega_s$. This behaves approximately proportional to

$$\frac{\epsilon^2}{r^2}.$$

The pattern of illumination that the source creates on the hemisphere (roughly) scales, too. As the sphere moves away, the rays leaving the surface patch and striking the sphere move closer together (roughly) evenly, and the collection changes only slightly (a small set of new rays is added at the rim—the contribution from these rays must be small because they come from directions tangent to the sphere). In the limit as $\epsilon$ tends to zero, no new rays are added.

The radiosity due to the source is obtained by integrating the pattern generated by the source, times $\cos\theta_i$, over the patch of solid angle. As $\epsilon$ tends to zero, the patch shrinks and the $\cos\theta_i$ is close to constant. If $\rho$ is the surface albedo, all this means the expression for radiosity due to the point source is

$$\rho \left(\frac{\epsilon}{r}\right)^2 E \cos\theta,$$

where $E$ is a term in the exitance of the source integrated over the small patch. We do not need a more detailed expression for $E$ (to determine one, we would need to actually do the integral we have shirked).

**A Nearby Point Source**    The angle term can be written in terms of $N(P)$ (the unit normal to the surface) and $S(P)$ (a vector from $P$ to the source whose length is $\epsilon^2 E$) to yield the standard model of a **nearby point source:**

$$\rho_d(P)\frac{N(P)\cdot S(P)}{r(P)^2}.$$

 This is an extremely convenient model, because it gives an explicit relationship between radiosity and shape (the normal term). In this model, $S$ is usually called the *source vector*.  It is common (and incorrect!) to omit the dependency on distance to the source from this model.

**A Point Source at Infinity**    The sun is far away; as a result, the terms $1/r(P)^2$ and $S(P)$ are essentially constant. In this case, the point source is referred to as being a *point source at infinity*. If all the surface patches we are interested in are close together with respect to the distance to the source, $r(P) = r_0 + \Delta r(P)$ where $r_0 \gg \Delta r(P)$. Furthermore, $S(P) = S_0 + \Delta S(P)$, where $\mid S_0 \mid \gg \mid \Delta S(P) \mid$. We now have

$$\frac{N\cdot S(P)}{r(P)^2} = \frac{N\cdot(S_0+\Delta S(P))}{(r_0+\Delta r(P))^2} \approx \frac{N\cdot S_0}{r_0^2}.$$

Now both $S_0$ and $r_0$ are constants, and there is no particular point in keeping them explicitly in the expression. Instead, we can write $S = (1/r_0^2)S_0$, and our model for the radiosity due to a point source at infinity becomes

$$B(P) = \rho_d(P)(N\cdot S).$$

The term $S$ is again known as the *source vector*. Typically, this model is used by inferring a source vector that is appropriate, rather than by computing its value from the exitance of the source and the geometry.

**Choosing a Point Source Model**    A point source at infinity is a good model for the sun, for example, because the solid angle that the sun subtends is small and essentially constant wherever it appears in the field of view (this test means that our approximation step is valid). If we use linear sensors with an unknown gain, for example, we can roll the source intensity and the unknown gain into the source vector term—this is quite often done without comment.

As you should expect from the derivation, a point source at infinity is a poor model *when the distance between objects is similar in magnitude to the distance to the source*. In this case, we cannot use the series approximation to pretend that the radiosity due to the source does not go down with distance to the source.

The heart of the problem is easy to see if we consider what the source looks like from different surface patches. It must look bigger to nearer surface patches (however small its radius); this means that the radiosity due to the source must go up. If the source is sufficiently distant—for example, the sun—we can ignore this effect because the source does not change in apparent size for any plausible motion.

However, for configurations like a light bulb in the center of a room, the solid angle subtended by the source goes up as the inverse square of the distance, meaning that the radiosity

due to the source will do so too. The correct model to use in this case is the point source of Section 5.2.2. The difficulty with this model is that radiosity changes sharply over space, in a way that is inconsistent with experience. For example, if a point source is placed at the center of a cube, then the the model predicts that radiosity in the corners is roughly one third that at the center of each face —but the corners of real rooms are nowhere near as dark as that. It is quite common in practice to suppress the distance term in the nearby point source model to account for this—an activity that is radiometrically incorrect but that tends to yield a better model. The explanation of this apparent contradiction must wait until we have discussed shading models.

### 5.2.3  Line Sources

A *line source* has the geometry of a line—a good example is a single fluorescent light bulb. Line sources are not terribly common in natural scenes or in synthetic environments, and we discuss them only briefly. Their main interest is as an example for radiometric problems; in particular, the radiosity of patches reasonably close to a line source changes as the reciprocal of distance to the source (rather than the square of the distance). The reasoning is more interesting than the effect. We model a line source as a thin cylinder with diameter $\epsilon$. Assume for the moment that the line source is infinitely long and that we are considering a patch that views the source frontally, as in Figure 5.4.

Figure 5.4 sketches the appearance of the source from the point of view of patch 1; now move the patch closer and consider patch 2—the width of the region on the hemisphere corresponding to the source changes, but not the length (because the source is infinitely long). In turn, because the width is approximately $\epsilon/r$, the radiosity due to the source must go down with the reciprocal of distance. It is easy to see that with a source that is not infinitely long, this applies as long as the patch is reasonably close.

### 5.2.4  Area Sources

An *area source* is an area that radiates light. Area sources are important for two reasons. First, they occur quite commonly in natural scenes—an overcast sky is a good example—and in syn-
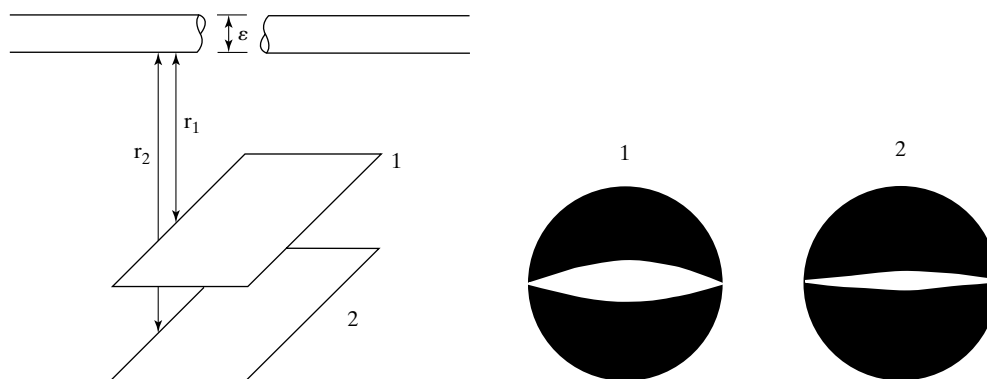


**Figure 5.4**   The radiosity due to a line source goes down as the reciprocal of distance for points that are reasonably close to the source. On the left, two patches viewing an infinitely long, narrow cylinder with constant exitance along its surface and diameter $\epsilon$. On the right, the view of the source *from each patch*, drawn as the underside of the input hemisphere seen from below. Notice that the length of the source on this hemisphere does not change, but the width does (as $\epsilon/r$). This yields the result.

thetic environments—for example, the fluorescent light boxes found in many industrial ceilings. Second, a study of area sources allows us to explain various shadowing and interreflection effects. Area sources are normally modeled as surface patches whose emitted radiance is independent of position and of direction—they can be described by their exitance.

An argument similar to that used for line sources shows that, for points not too distant from the source, the radiosity due to an area source does not change with distance to the source. This is because, if the area is large enough with respect to the distance to the source, the area subtended by the source on some input hemisphere is about the same as we move toward and away from the source. This explains the widespread use of area sources in illumination engineering—they generally yield fairly uniform illumination. For our applications, we need a more exact description of the radiosity due to an area source, so we need to write out the integral.

**The Exact Radiosity due to an Area Source**     Assume we have a diffuse surface patch that is illuminated by an area source with exitance $E(Q)$ at the source point $Q$. Instead of writing angles in coordinates, we write $\overrightarrow{QP}$ for the direction from $Q$ to $P$ (more notation
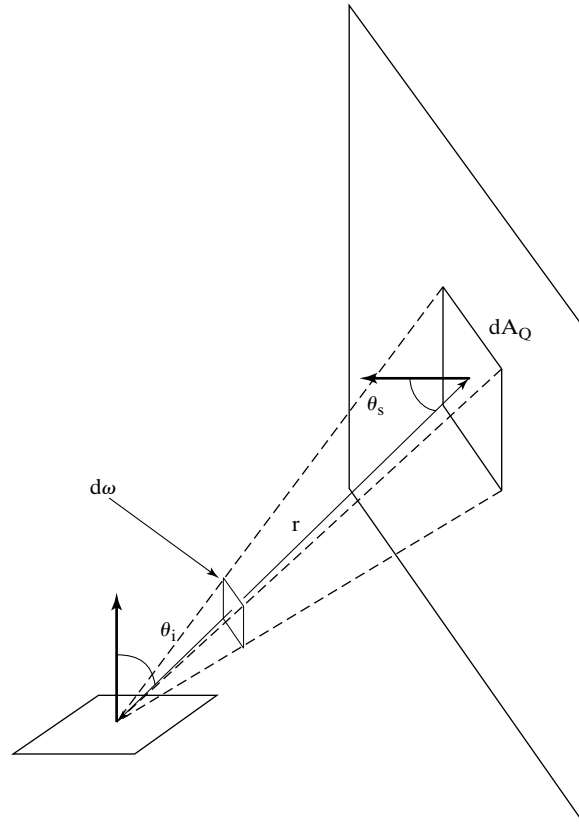


**Figure 5.5**   A diffuse source illuminates a diffuse surface. The source has exitance $E(Q)$, and we wish to compute the radiosity on the patch due to the source. We do this by transforming the integral of incoming radiance at the surface into an integral over the source area. This transformation is convenient because it avoids us having to use different angular domains for different surfaces. However, it still leads to an integral that is usually impossible in closed form.

is illustrated in Figure 5.5). The radiosity on the surface is obtained by summing the incoming radiance over all incoming directions. This integral can be transformed into an integral over the source as follows:

$$B(P) = \rho_d(P) \int_\Omega L_i(P, \overrightarrow{QP}) \cos \theta_i \, d\omega$$

$$= \rho_d(P) \int_\Omega L_e(Q, \overrightarrow{QP}) \cos \theta_i \, d\omega$$

$$= \rho_d(P) \int_\Omega \left( \frac{1}{\pi} E(Q) \right) \cos \theta_i \, d\omega$$

$$= \rho_d(P) \int_{\text{Source}} \left( \frac{1}{\pi} E(Q) \right) \cos \theta_i \left( \cos \theta_s \frac{dA_Q}{r^2} \right)$$

$$= \rho_d(P) \int_{\text{Source}} E(Q) \frac{\cos \theta_i \cos \theta_s}{\pi r^2} \, dA_Q.$$

The transformation works because radiance is constant along straight lines and because $E(Q) = (1/\pi)L_e(Q)$. It is useful because it means we do not have to worry about consistent angular coordinate systems. However we transform them, integrals describing the effect of area sources are generally difficult or impossible to do in closed form.

## 5.3 LOCAL SHADING MODELS

We have studied the physics of light because we want to know how bright things will be, and why, in the hope of extracting object information from these models. Currently, we know the radiosity at a patch *due to a source* but this is not a shading model. Radiance could arrive at surface patches in other ways (e.g., it could be reflected from other surface patches); we need to know which components to account for.

The easiest model to manipulate is a *local shading model*, which models the radiosity at a surface patch as the sum of the radiosity due only to light internally generated at sources. This means that we assume that light is not reflected from surface to surface, but instead leaves a source, arrives at some surface, and proceeds directly to the camera. This model is palpably unphysical, but is easy to analyze. The model supports a variety of algorithms and theories (see Section 5.4). Unfortunately, this model often produces wildly inaccurate predictions. Even worse, there is little reliable information about when it is safe to use this model.

An alternate model is to account for all radiation (Section 5.5). This takes into account radiance arriving from sources and that arriving from radiating surfaces. This model is physically accurate, but usually hard to manipulate.

### 5.3.1 Local Shading Models for Point Sources

The local shading model for a set of point sources is obtained by writing out the radiosity due to light internally generated at sources. This gives

$$B(P) = \sum_{s \in \text{sources visible from } P} B_s(P),$$

where $B_s(P)$ is the radiosity due to source $s$. This expression is fairly innocuous, but notice that if all the sources are point sources at infinity, the expression becomes

$$B(P) = \sum_{s\in\text{sources visible from }P} \rho_d(P)N(P) \cdot S_s,$$

so that if we confine our attention to a region where all points can see the same sources, we could add all the source vectors to obtain a single virtual source that had the same effects. The relationship between shape and shading is pretty direct here —the radiosity is a measurement of one component of the surface normal.

For point sources that are not at infinity, the model becomes

$$B(P) = \sum_{s\in\text{sources visible from }P} \rho_d(P)\frac{N(P) \cdot S(P)}{r_s(P)^2},$$

where $r_s(P)$ is the distance from the source to $P$; the presence of this term means that the relationship between shape and shading is somewhat more obscure.

**The Appearance of Shadows**    In a local shading model, shadows occur when the patch can not see one or more sources. In this model, point sources produce a series of shadows with crisp boundaries; shadow regions where no source can be seen are particularly dark. Shadows cast with a single source can be crisp and black depending on the size of the source and the albedo of other nearby surfaces (which could reflect light into the shadow and soften its boundary). It was a popular 19th Century pastime to cast such shadows onto paper and then draw them, yielding the silhouettes still occasionally found in antiques shops.

The geometry of the shadow cast by a point source on a plane is analogous to the geometry of viewing in a perspective camera (Figure 5.6). Any patch on the plane is in shadow if a ray from the patch to the source passes through an object. This means that there are two kinds of shadow boundary. At *self shadow boundaries*, the surface is turning away from the light, and a ray from the patch to the source is tangent to the surface. At *cast shadow boundaries*, from the perspective of the patch, the source suddenly disappears behind an occluding object. Shadows cast onto curved surfaces can have extremely complex geometries, however.
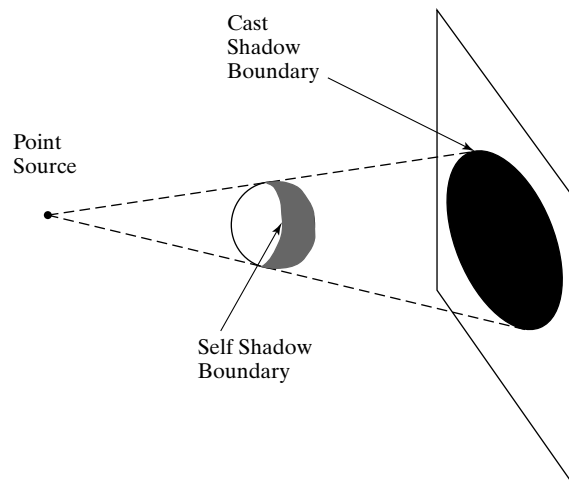


**Figure 5.6**    Shadows cast by point sources on planes are relatively simple. Self shadow boundaries occur when the surface turns away from the light and cast shadow boundaries occur when a distant surface occludes the view of the source.

If there are many sources, the shadows are less dark (except at points where no source is visible), and there can be many qualitatively distinct shadow regions (each source casts its own shadow—some points may not see more than one source). One example of this effect occurs in televised soccer matches. Because the stadium has multiple bright distant point-like illuminants spaced evenly around the perimeter of the stadium, there is a set of shadows radiating evenly around each player's feet. These shadows typically become brighter or darker as the player moves around usually because the illumination due to other sources and to interreflections in the region of the shadow increases or decreases.

### 5.3.2  Area Sources and Their Shadows

The local shading model for a set of area sources is significantly more complex because it is possible for patches to see only a portion of a given source. The model becomes

$$B(P) = \sum_{s \in \text{ all sources}} \left\{ \int_{\text{visible component of source } s} \text{Radiosity due to source} \right\}$$

$$= \sum_{s \in \text{ all sources}} \int_{\text{visible component of source } s} \left\{ E(Q) \frac{\cos \theta_Q \cos \theta_s}{\pi r^2} \, dA_Q \right\}$$

using the terminology of Figure 5.5; usually, we assume that $E$ is constant over the source.

Area sources do not produce dark shadows with crisp boundaries. This is because, from the perspective of a viewing patch, the source appears slowly from behind the occluding object (think of an eclipse of the moon—it is an exact analogy). It is common to distinguish between points in the *umbra* (a Latin word meaning "shadow")—which cannot see the source at all—and points in the *penumbra* (a compound of Latin words meaning "almost shadow")—which see part of the source. The vast majority of indoor sources are area sources of one form or another, so the effects are quite easy to see; hold an arm quite close to the wall and look at the shadow it casts. There is a dark core, which gets larger as the arm gets closer to the wall (this is the umbra), surrounded by a lighter region with a fuzzier boundary (the penumbra). Figure 5.7 illustrates the geometry.

### 5.3.3  Ambient Illumination

One problem with local shading models should be apparent immediately; they predict that some shadow regions are arbitrarily dark because they cannot see the source. This prediction is inaccurate in almost every case because shadows are illuminated by light from other diffuse surfaces. This effect can be significant. In rooms with light walls and area sources, it is possible to see shadows only by holding objects close to the wall or close to the source. This is because a patch on the wall sees all the other walls in the room; until an object is close to the wall, it blocks out only a small fraction of the visual hemisphere of each patch.

For some environments, the total irradiance a patch obtains from other patches is roughly constant and roughly uniformly distributed across the input hemisphere. This must be true for the interior of a sphere with a constant distribution of radiosity (by symmetry) and (by accepting a model of a cube as a sphere) is roughly true for the interior of a room with white walls. In such an environment, it is sometimes possible to model the effect of other patches by adding an *ambient illumination* term to each patch's radiosity. There are two strategies for determining this term. First, if each patch sees the same proportion of the world (e.g., the interior of a sphere), we can add the same constant term to the radiosity of each patch. The magnitude of this term is usually guessed.
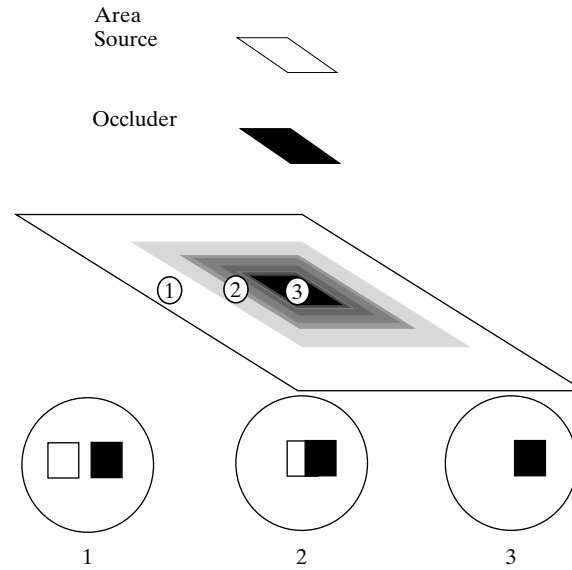
**Figure 5.7**   Area sources generate complex shadows with smooth boundaries, because from the point of view of a surface patch, the source disappears slowly behind the occluder. Regions where the source cannot be seen at all are known as the *umbra*; regions where some portion of the source is visible are known as the *penumbra*. A good model is to imagine lying with your back to the surface looking at the world above. At point 1, you can see all of the source; at point 2, you can see some of it; and at point 3, you can see none of it.

Second, if some patches see more or less of the world than others (this happens if regions of the world occlude a patch's view, e.g., a patch at the bottom of a groove), this can be taken into account. To do so, we need a model of the world *from the perspective of the patch under consideration*. A natural strategy is to model the world as a large, distant polygon of constant radiosity, where the view of this polygon is occluded at some patches (see Figure 5.8). The result is that the ambient term is smaller for patches that see less of the world. This model is often more accurate than adding a constant ambient term. Unfortunately, it is much more difficult to extract information from this model, possibly as difficult as for a global shading model.

## 5.4  APPLICATION: PHOTOMETRIC STEREO

We reconstruct a patch of surface from a series of pictures of the surface taken under different illuminants. For simplicity, we use an orthographic camera and choose a coordinate system such that the point $(x, y, z)$ in space projects onto the point $(x, y)$ in the image (the method we describe works for the other camera models described in chapter 1).

In this case, to measure the shape of the surface, we need to obtain the depth to the surface. This suggests representing the surface as $(x, y, f(x, y))$—a representation known as a *Monge patch* after a French military engineer who first used it (Figure 5.9). This representation is attractive because we can determine a unique point on the surface by giving the image coordinates. Notice that to obtain a measurement of a solid object, we would need to reconstruct more than one patch because we need to observe the back of the object.

*Photometric stereo* is a method for recovering a representation of the Monge patch from image data. The method involves reasoning about the image intensity values for several
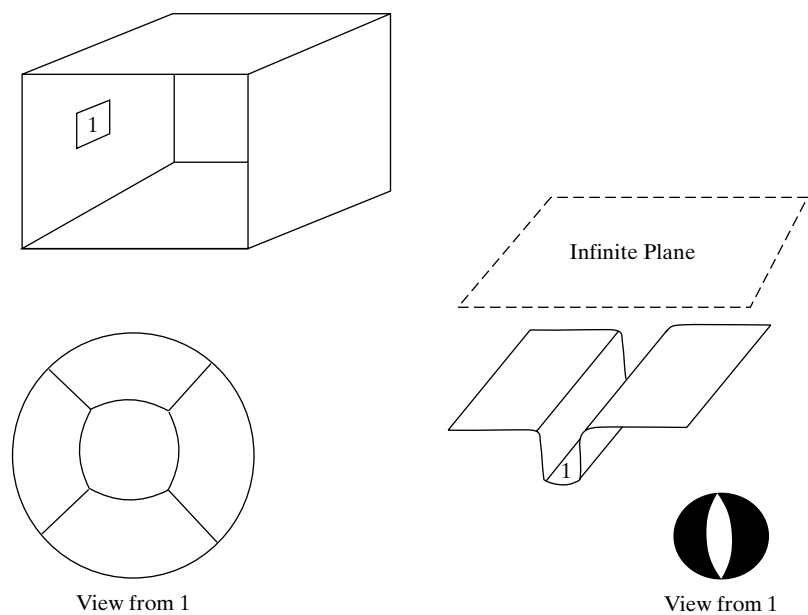
**Figure 5.8**    Ambient illumination is a term added to the radiosity predictions of local shading models to model the effects of radiosity from distant, reflecting surfaces. In a world like the interior of a sphere or of a cube (the case on the left), where a patch sees roughly the same thing from each point, a constant ambient illumination term is often acceptable. In more complex worlds, some surface patches see much less of the surrounding world than others. For example, the patch at the base of the groove on the right sees relatively little of the outside world, which we model as an infinite polygon of constant exitance; its input hemisphere is shown below.
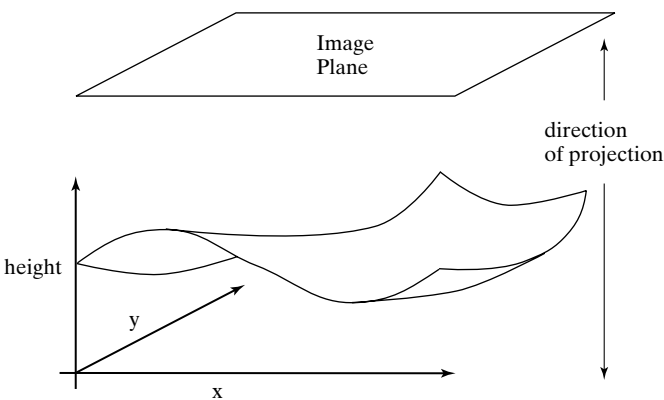


**Figure 5.9**    A Monge patch is a representation of a piece of surface as a height function. For the photometric stereo example, we assume that an orthographic camera—one that maps $(x, y, z)$ in space to $(x, y)$ in the camera—is viewing a Monge patch. This means that the shape of the surface can be represented as a function of position in the image.

different images of a surface in a fixed view illuminated by different sources. This method re-
covers the height of the surface at points corresponding to each pixel; in computer vision circles,
the resulting representation is often known as a *height map*, *depth map*, or *dense depth map*.

  Fix the camera and the surface in position and illuminate the surface using a point source
that is far away compared with the size of the surface. We adopt a local shading model and
assume that there is no ambient illumination (more about this later) so that the radiosity at a
point $P$ on the surface is

$$B(P) = \rho(P)N(P) \cdot S_1,$$

where $N$ is the unit surface normal and $S_1$ is the source vector. With our camera model, there is
only one point $P$ on the surface for each point $(x, y)$ in the image, and we can write $B(x, y)$ for
$B(P)$. Now we assume that the response of the camera is linear in the surface radiosity, so the
value of a pixel at $(x, y)$ is

$$I(x, y) = kB(x, y)$$
$$= k\rho(x, y)N(x, y) \cdot S_1$$
$$= g(x, y) \cdot V_1,$$

where $k$ is the constant connecting the camera response to the input radiance, $g(x, y) =
\rho(x, y)N(x, y)$, and $V_1 = kS_1$.

  In these equations, $g(x, y)$ describes the surface and $V_1$ is a property of the illumination
and of the camera. We have a dot product between a vector field $g(x, y)$ and a vector $V_1$, which
could be measured; with enough of these dot products, we could reconstruct $g$ and so the surface.

### 5.4.1 Normal and Albedo from Many Views

Now if we have $n$ sources, for each of which $V_i$ is known, we stack each of these $V_i$ into a known
matrix $\mathcal{V}$, where

$$\mathcal{V} = \begin{pmatrix} V_1^T \\ V_2^T \\ \dots \\ V_n^T \end{pmatrix}.$$

For each image point, we stack the measurements into a vector

$$i(x, y) = \{I_1(x, y), I_2(x, y), \dots, I_n(x, y)\}^T.$$

Notice that we have one vector per image point; each vector contains all the image brightnesses
observed at that point for different sources. Now we have

$$i(x, y) = \mathcal{V}g(x, y),$$

 and $g$ is obtained by solving this linear system—or rather, one linear system per point in the
image. Typically, $n > 3$ so that a least squares solution is appropriate. This has the advantage
that the residual error in the solution provides a check on our measurements.

  The difficulty with this approach is that substantial regions of the surface may be in shadow
for one or the other light (see Figure 5.10). There is a simple trick that deals with shadows. If
there really is no ambient illumination, then we can form a matrix from the image vector and
multiply both sides by this matrix; this zeroes out any equations from points that are in shadow.
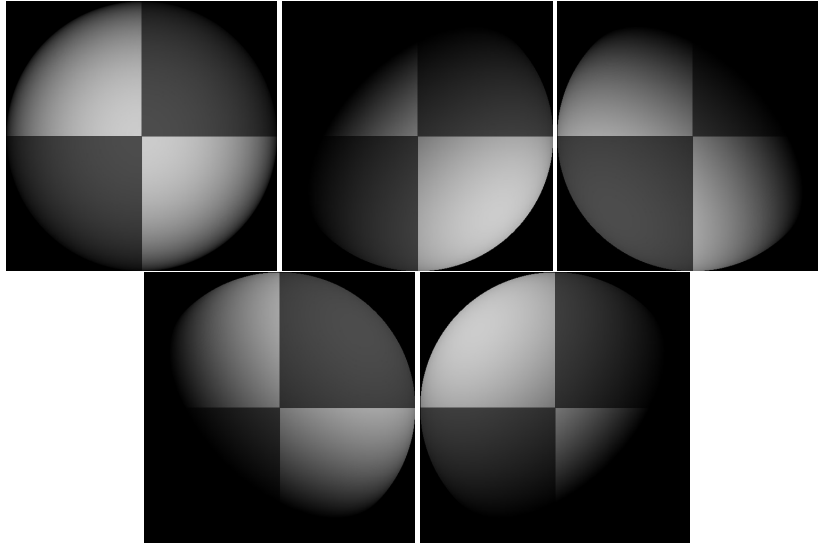
**Figure 5.10**   Five synthetic images of a sphere, all obtained in an orthographic view from the same viewing position. These images are shaded using a local shading model and a distant point source. This is a convex object, so the only view where there is no visible shadow occurs when the source direction is parallel to the viewing direction. The variations in brightness occuring under different sources code the shape of the surface.

We form

$$\mathcal{I}(x, y) = \begin{pmatrix} I_1(x, y) & \cdots & 0 & 0 \\ 0 & I_2(x, y) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & I_n(x, y) \end{pmatrix}$$

and

$$\mathcal{I}\boldsymbol{i} = \mathcal{I}\mathcal{V}\boldsymbol{g}(x, y),$$

and $\mathcal{I}$ has the effect of zeroing the contributions from shadowed regions, because the relevant elements of the matrix are zero at points that are in shadow. Again, there is one linear system per point in the image; at each point, we solve this linear system to recover the $\boldsymbol{g}$ vector at that point.

**Measuring Albedo**    We can extract the albedo from a measurement of $\boldsymbol{g}$ because $N$ is the unit normal. This means that $|\boldsymbol{g}(x, y)| = \rho(x, y)$. This provides a check on our measurements as well. Because the albedo is in the range zero to one, any pixels where $|\boldsymbol{g}|$ is greater than one are suspect—either the pixel is not working or $\mathcal{V}$ is incorrect. Figure 5.11 shows albedo recovered using this method for the images of Figure 5.10.

**Recovering Normals**    We can extract the surface normal from $\boldsymbol{g}$ because the normal is a unit vector

$$N(x, y) = \frac{1}{|\boldsymbol{g}(x, y)|}\boldsymbol{g}(x, y).$$

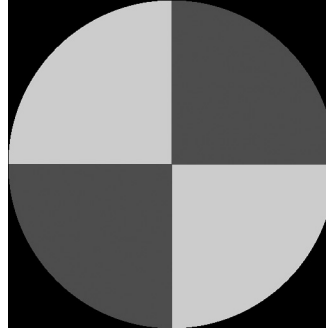Figure 5.12 shows normal values recovered for the images of Figure 5.10.

**Figure 5.11**   The magnitude of the vector field $\boldsymbol{g}(x, y)$ recovered from the input data of Figure 5.10 represented as an image—this is the reflectance of the surface.

### 5.4.2 Shape from Normals

The surface is $(x, y, f(x, y))$, so the normal as a function of $(x, y)$ is

$$
N(x, y) = \frac{1}{\sqrt{1 + \frac{\partial f}{\partial x}^2 + \frac{\partial f}{\partial y}^2}} \left\{ -\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1 \right\}^T
$$

To recover the depth map, we need to determine $f(x, y)$ from measured values of the unit normal.

Assume that the measured value of the unit normal at some point $(x, y)$ is $(a(x, y), b(x, y), c(x, y))$. Then

$$
\frac{\partial f}{\partial x} = \frac{a(x, y)}{c(x, y)} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{b(x, y)}{c(x, y)}.
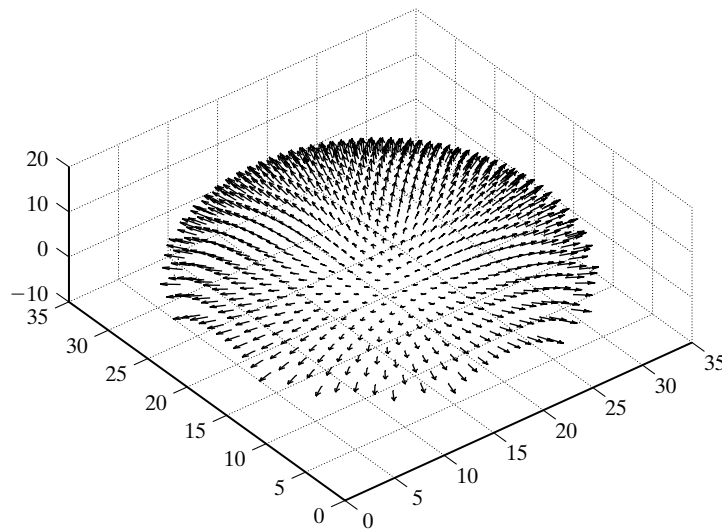$$



**Figure 5.12**   The normal field recovered from the input data of Figure 5.10.

We have another check on our data set, because

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

so we expect that

$$\frac{\partial \left( \frac{a(x,y)}{c(x,y)} \right)}{\partial y} - \frac{\partial \left( \frac{b(x,y)}{c(x,y)} \right)}{\partial x}$$

should be small at each point. In principle it should be zero, but we would have to estimate these partial derivatives numerically and so should be willing to accept small values. This test is known as a test of *integrability*, which in vision applications always boils down to checking that mixed second partials are equal.

---

**Algorithm 5.1:** Photometric Stereo

Obtain many images in a fixed view under different illuminants
Determine the matrix $\mathcal{V}$ from source and camera information
Create arrays for albedo, normal (3 components),
   p (measured value of $\frac{\partial f}{\partial x}$) and
   q (measured value of $\frac{\partial f}{\partial y}$)
For each point in the image array
   Stack image values into a vector $\boldsymbol{i}$
   Construct the diagonal matrix $\mathcal{I}$
   Solve $\mathcal{I}\mathcal{V}\boldsymbol{g} = \mathcal{I}\boldsymbol{i}$ to obtain $\boldsymbol{g}$ for this point
   Albedo at this point is $|\boldsymbol{g}|$
   Normal at this point is $\frac{\boldsymbol{g}}{|\boldsymbol{g}|}$
   p at this point is $\frac{N_1}{N_3}$
   q at this point is $\frac{N_2}{N_3}$
end
Check: is $(\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x})^2$ small everywhere?

Top left corner of height map is zero
For each pixel in the left column of height map
   height value = previous height value + corresponding q value
end
For each row
   For each element of the row except for leftmost
      height value = previous height value + corresponding p value
   end
end

---

**Shape by Integration**     Assuming that the partial derivatives pass this sanity test, we can reconstruct the surface up to some constant depth error. The partial derivative gives the change in surface height with a small step in either the *x* or the *y* direction. This means we can
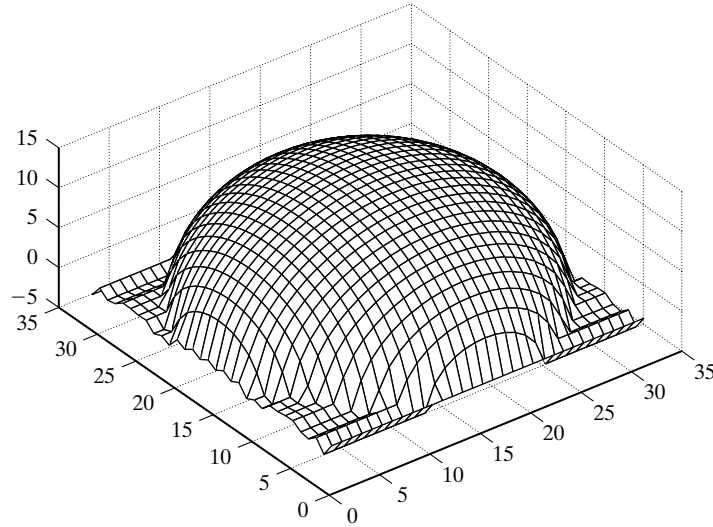
**Figure 5.13**    The height field obtained by integrating the normal field of Figure 5.12 using the method described in the text.

get the surface by summing these changes in height along some path. In particular, we have

$$f(x, y) = \oint_C \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \cdot d\boldsymbol{l} + c,$$

where $C$ is a curve starting at some fixed point and ending at $(x, y)$ and $c$ is a constant of integration, which represents the (unknown) height of the surface at the start point. The recovered surface does not depend on the choice of curve (exercises).

For example, we can reconstruct the surface at $(u, v)$ by starting at $(0, 0)$, summing the $y$-derivative along the line $x = 0$ to the point $(0, v)$, and then summing the $x$-derivative along the line $y = v$ to the point $(u, v)$:

$$f(u, v) = \int_0^v \frac{\partial f}{\partial y}(0, y) \, dy + \int_0^u \frac{\partial f}{\partial x}(x, v) \, dx + c.$$

This is the integration path given in Algorithm 5.1. Any other set of paths would work as well although it is probably best to use many different paths and average so as to spread around the error in the derivative estimates. Figure 5.13 shows the reconstruction obtained for the data of Figure 5.10.

Another approach to recovering shape is to choose the function $f(x, y)$ whose partial derivatives most look like the measured partial derivatives. We explore this approach for a similar problem in Section 6.5.2.

## 5.5 INTERREFLECTIONS: GLOBAL SHADING MODELS

Local shading models can be quite misleading. In the real world, each surface patch is illuminated not only by sources, but also by light reflected off other surface patches (a phenomenon known as *interreflection*). A model that incorporates interreflection effects is known as a *global shading model*. Interreflections lead to a variety of complex shading effects, which are still quite poorly
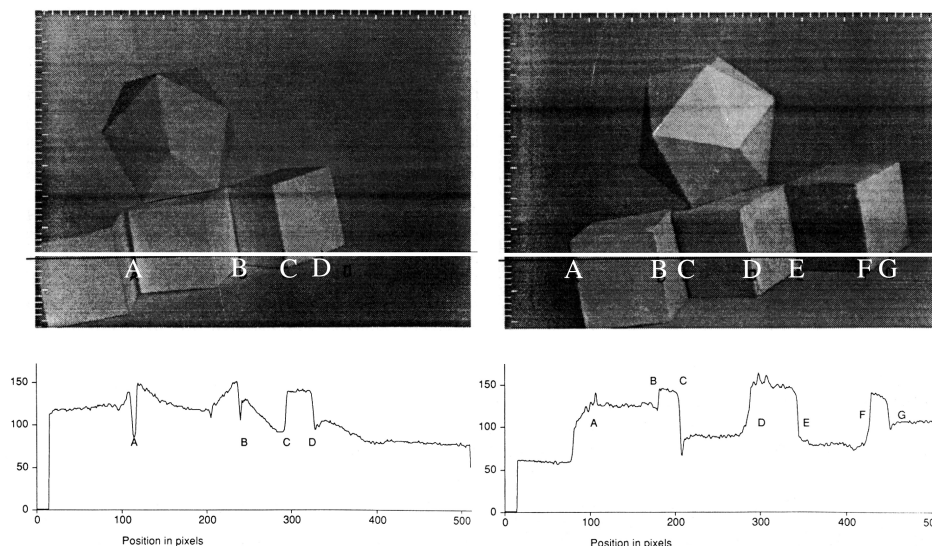
**Figure 5.14**   The column on the **left** shows data from a room with matte black walls and containing a collection of matte black polyhedral objects; that on the **right** shows data from a white room containing white objects. The images are qualitatively different, with darker shadows and crisper boundaries in the black room and bright reflexes in the concave corners in the white room. The graphs show sections of the image intensity along the corresponding lines in the images. *Figure from "Mutual Illumination," by D.A. Forsyth and A.P. Zisserman, Proc. CVPR, 1989,* © *1989 IEEE*

understood. Unfortunately, these effects occur widely, and it is still not yet known how to simplify global shading models without losing essential qualitative properties.

For example, Figure 5.14 shows views of the interior of two rooms. One room has black walls and contains black objects. The other has white walls and contains white objects. Each is illuminated (approximately!) by a distant point source. Given that the intensity of the source is adjusted appropriately, the local shading model predicts that these pictures would be indistinguishable. In fact, the black room has much darker shadows and crisper boundaries at the creases of the polyhedra than the white room. This is because surfaces in the black room reflect less light onto other surfaces (they are darker), whereas in the white room other surfaces are significant sources of radiation. The sections of the camera response to the radiosity (these are proportional to radiosity for diffuse surfaces) shown in the figure are hugely different qualitatively. In the black room, the radiosity is constant in patches as a local shading model would predict, whereas in the white room slow image gradients are quite common—these occur in concave corners, where object faces reflect light onto one another.

This effect also explains why a room illuminated by a point light source does not show the sharp illumination gradients that a local shading model predicts (recall Section 5.2.2). The walls and floor of the room reflect illumination back, and this tends to light up the corners, which would otherwise be dark.

### 5.5.1  An Interreflection Model

It is well understood how to predict the radiosity on a set of diffuse surface patches. The total radiosity leaving a patch is its exitance —which is zero for all but sources—*plus* all the radiosity

that is reflected from the patch:

$$B(P) = E(P) + B_{\text{refl}}(P).$$

From the point of view of our patch, there is no distinction between energy leaving another patch due to exitance and that due to reflection. This means we can take the expression for an area source and use it to obtain an expression for $B_{\text{refl}}(Q)$. In particular, from the perspective of our patch, the patch at $R$ in the world is equivalent to an area source with exitance $B(R)$. This means

$$B_{\text{refl}}(P) = \rho_d(P) \int_{\text{world}} \text{visible}(P, Q) B(Q) \frac{\cos \theta_P \cos \theta_Q}{\pi d_{PQ}^2} \, dA_Q$$

$$= \rho_d(P) \int_{\text{world}} \text{visible}(P, Q) K(P, Q) B(Q) \, dA_Q,$$

where the terminology is that of Figure 5.15 and

$$\text{visible}(P, Q) = \left\{ \begin{array}{l} 1 \text{ if } P \text{ can see } Q, \\ 0 \text{ if } P \text{ cannot see } Q. \end{array} \right.$$

The term $\text{visible}(P, Q) K(P, Q)$ is usually referred to as the *interreflection kernel*. Substituting the expression for $B_{\text{refl}}(P)$ gives

$$B(P) = E(P) + \rho_d(P) \int_{\text{world}} \text{visible}(P, Q) K(P, Q) B(Q) \, dA_Q.$$

In particular, the solution appears inside the integral. Equations of this form are known as Fredholm integral equations of the second kind. This particular equation is a fairly nasty sample of the type because the interreflection kernel is generally not continuous and may have singularities. Solutions of this equation can yield quite good models of the appearance of diffuse surfaces, and the topic supports a substantial industry in the computer graphics community (good
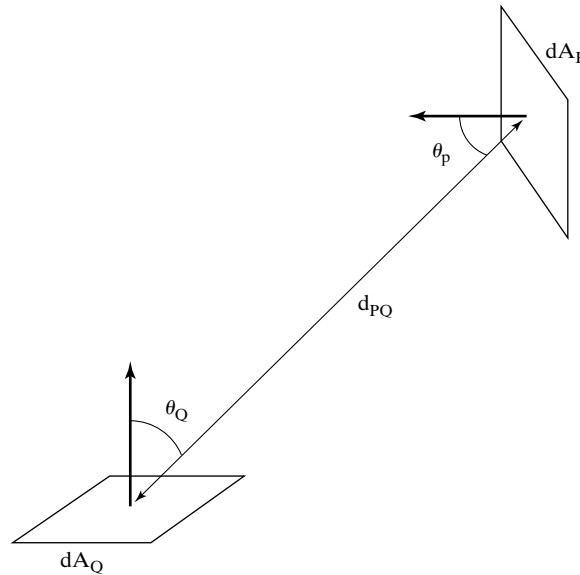


**Figure 5.15**   Terminology for expression derived in the text for the interreflection kernel.
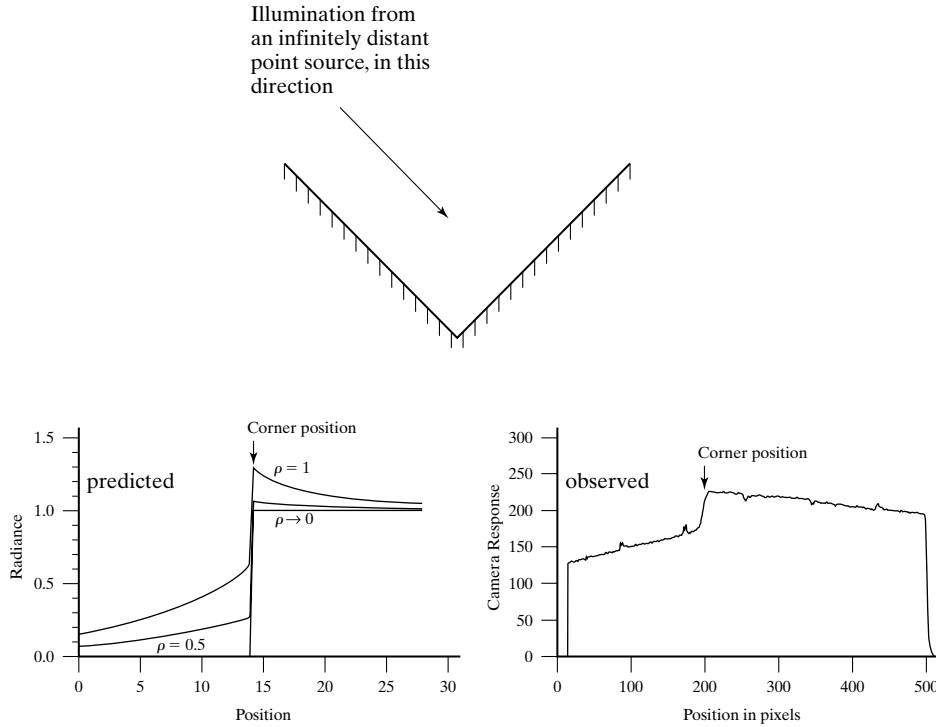
**Figure 5.16**   The model described in the text produces quite accurate qualitative predictions for interreflections. The **top** figure shows a concave right-angled groove illuminated by a point source at infinity where the source direction is parallel to the one face. On the **left** of the bottom row is a series of predictions of the radiosity for this configuration. These predictions have been scaled to lie on top of one another; the case $\rho \to 0$ corresponds to the local shading model. On the **right**, an observed image intensity for an image of this form for a corner made of white paper, showing the roof-like gradient in radiosity associated with the edge. A local shading model predicts a step. *Figure from "Mutual Illumination," by D.A. Forsyth and A.P. Zisserman, Proc. CVPR, 1989,* © *1989 IEEE*

places to start for this topic are Cohen and Wallace (1993) or Sillion (1994)). The model produces good predictions of observed effects (Figure 5.16).

### 5.5.2  Solving for Radiosity

We sketch one approach to solving the global shading model to illustrate the methods. Subdivide the world into small, flat patches and approximate the radiosity as being constant over each patch. This approximation is reasonable because we could obtain an accurate representation by working with small patches. Now we construct a vector $B$, which contains the value of the radiosity for each patch. In particular, the $i$th component of $B$ is the radiosity of the $i$th patch.

   We write the incoming radiosity at the $i$th patch due to radiosity on the $j$th patch as

$$B_{j \to i}(P) = \rho_d(P) \int_{\text{patch } j} \text{visible}(P, Q) K(P, Q) \, dA_Q B_j,$$

where $P$ is a coordinate on the $i$th patch and $R$ is a coordinate on the $j$th patch. Now this expression is not a constant, and so we must average it over the $i$th patch to get

$$\overline{B}_{j\to i} = \frac{1}{A_i} \int_{\text{patch } i} \rho_d(P) \int_{\text{patch } j} \text{visible}(P, Q) K(P, Q) \, dA_P \, dA_Q B_j,$$

where $A_i$ is the area of the $i$th patch. If we insist that the exitance on each patch is constant, too, we obtain the model

$$B_i = E_i + \sum_{\text{all } j} \overline{B}_{j\to i}$$

$$= E_i + \sum_{\text{all } j} K_{ij} B_j,$$

where

$$K_{ij} = \frac{1}{A_i} \int_{\text{patch } i} \rho_d(P) \int_{\text{patch } j} \text{visible}(P, Q) K(P, Q) \, dA_P \, dA_Q.$$

The elements of this matrix are sometimes known as *form factors*.

This is a system of linear equations in $B_i$ (although an awfully big one—$K_{ij}$ could be a million by a million matrix) and, as such, can in principle be solved. The tricks that are necessary to solve the system efficiently, quickly, and accurately are well beyond our scope; Sillion (1994) is an excellent account, as is the book of Cohen and Wallace (1993).

### 5.5.3 The Qualitative Effects of Interreflections

We should like to extract shape information from radiosity. This is relatively easy to do with a local model (see Section 5.4 for some details), but the model describes the world poorly, and little is known about how severely this affects the resulting shape information. Extracting shape information from a global shading model is difficult for two reasons. First, the relationship between shape and radiosity is complicated because it is governed by the interreflection kernel. Second, there are almost always surfaces that are not visible, but radiate to the objects in view. These so-called "distant surfaces" mean it is hard to account for all radiation in the scene using an interreflection model because some radiators are invisible and we may know little or nothing about them.

All this suggests that understanding qualitative, local effects of interreflection is important; armed with this understanding, we can either discount the effects of interreflection or exploit them. This topic remains largely an open research topic, but there are some things we can say.

**Smoothing and Regional Properties**    First, interreflections have a characteristic smoothing effect. This is most obviously seen if one tries to interpret a stained glass window by looking at the pattern it casts on the floor; this pattern is almost always a set of indistinct colored blobs. The effect is seen most easily with the crude model of Figure 5.17. The geometry consists of a patch with a frontal view of an infinite plane, which is a unit distance away and carries a radiosity $\sin \omega x$. There is no reason to vary the distance of the patch from the plane because interreflection problems have scale invariant solutions—this means that the solution for a patch two units away can be obtained by reading our graph at $2\omega$. The patch is small enough that its contribution to the plane's radiosity can be ignored. If the patch is slanted by $\sigma$ with respect to the plane, it carries radiosity that is nearly periodic, with spatial frequency $\omega \cos \sigma$. We refer to the amplitude of the component at this frequency as the gain of the patch and plot the gain in Figure 5.17. The important property of this graph is that high spatial frequencies have a difficult time jumping the gap from the plane to the patch. This means that shading effects with high
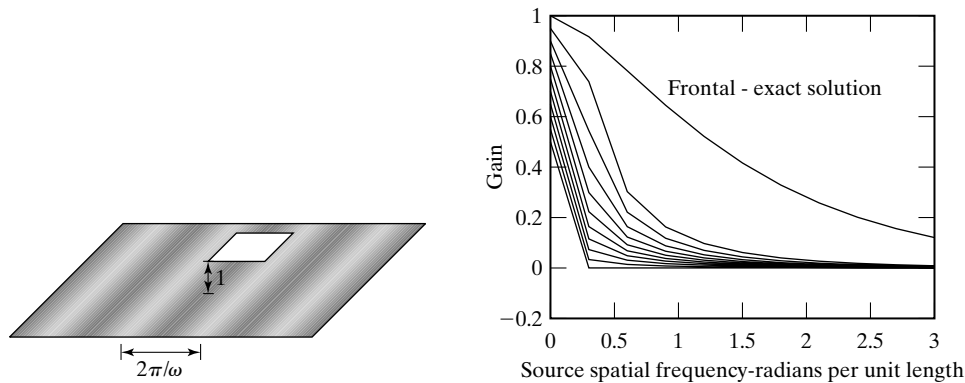
**Figure 5.17**    A small patch views a plane with sinusoidal radiosity of unit am-
plitude. This patch has a (roughly) sinusoidal radiosity due to the effects of the
plane. We refer to the amplitude of this component as the *gain of the patch*. The
graph shows numerical estimates of the gain for patches at 10 equal steps in slant
angle, from 0 to $\pi/2$, as a function of spatial frequency *on the plane*. The gain
falls extremely fast, meaning that large terms at high spatial frequencies must
be regional effects, rather than the result of distant radiators. This is why it is
hard to determine the pattern in a stained glass window by looking at the floor
at foot of the window. *Reprinted from "Shading Primitives: Finding Folds and
Shallow Grooves," by J. Haddon and D.A. Forsyth, Proc. Int. Conf. Computer
Vision, 1998 © 1998 IEEE*

spatial frequency and high amplitude generally cannot come from distant surfaces (unless they
are abnormally bright).

The extremely fast fall-off in amplitude with spatial frequency of terms due to distant
surfaces means that, if one observes a high amplitude term at a high spatial frequency, *it is very
unlikely to have resulted from the effects of distant, passive radiators* (because these effects die
away quickly). There is a convention, which we see in Section 6.5.2, that classifies effects in
shading as due to reflectance if they are fast ("edges") and the dynamic range is relatively low
and due to illumination otherwise. We can expand this convention. There is a mid range of spatial
frequencies that are largely unaffected by mutual illumination from distant surfaces because the
gain is small. Spatial frequencies in this range cannot be transmitted by distant passive radiators
unless these radiators have improbably high radiosity. As a result, spatial frequencies in this
range can be thought of as *regional properties*, which can result only from interreflection effects
within a region.

The most notable regional properties are probably *reflexes*— small bright patches that ap-
pear mainly in concave regions (illustrated in Figure 5.18 and Figure 5.19). A second important
effect is *color bleeding*, where a colored surface reflects light onto another colored surface. This
is a common effect that people tend not to notice unless they are consciously looking for it. It is
quite often reproduced by painters.

## 5.6  NOTES

Shading models are handled in a quite unsystematic way in the vision literature. The point source
approximation is widely abused; you should use it with care and inspect others' use of it with
suspicion. We believe we are the first to draw the distinction between (a) the physical effects of
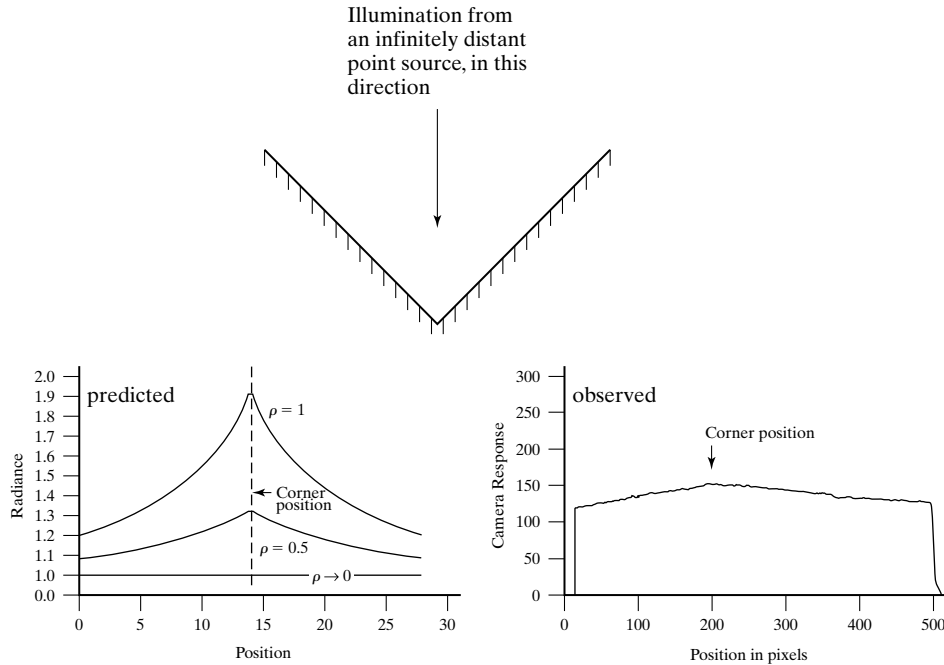sources, and (b) the shading model.

Illumination from
an infinitely distant
point source, in this
direction



**Figure 5.18**   Reflexes at concave edges are a common qualitative result of interreflections. The figure on the **top** shows the situation here; a concave right-angled groove illuminated by a point light source at infinity, whose source vector is along the angle bisector. The graph on the **left** shows the intensity predictions of an interreflection model for this configuration; the case $\rho \to 0$ is a local shading model. The graphs have been lined up for easy comparison. As the surface's albedo goes up, a roof-like structure appears. The graph on the **right** shows an observation of this effect in an image of a real scene. *Figure from "Mutual Illumination," by D.A. Forsyth and A.P. Zisserman, Proc. CVPR, 1989,* © *1989 IEEE*

## Local Shading Models

The great virtue of local shading models is that the analysis is simple. The primary characteristic of a local shading model is that, on a surface of constant albedo, the radiosity of a surface patch is a function of the normal alone. This means that one can avoid the abstraction of reflectance and sources and instead simply code the properties of surface *and source* as a *reflectance map*. The reflectance map is a function that takes a representation of the normal, and returns the radiosity to be expected at a point with that normal.

Horn started the systematic study of shading in computer vision, with important papers on recovering shape from a local shading model using a point source (Horn, 1970, 1975), with a more recent account in Horn (1990). The methods discussed have largely fallen into disuse (at least partially because they appear unable to cope with the difficulties created by a global shading model), so we do not survey the vast literature here. A comprehensive summary is in Horn and Brooks (1989). Shape and albedo are ambiguous; with appropriate changes in albedo, surfaces of different shapes can generate the same image (Belhumeur, Kriegman and Yuille, 1999, Kriegman and Belhumeur, 1998). Because the surface normal is the key in local shading models, such models typically yield elegant links between surface shading and curvature (Koenderink and van Doorn, 1980).
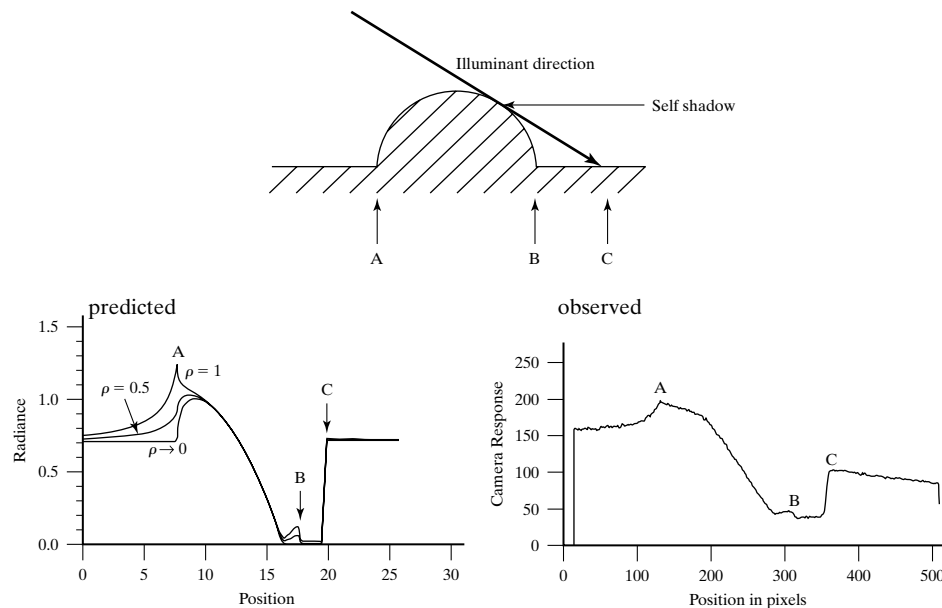
**Figure 5.19**    Reflexes occur quite widely; they are usually caused by a favor-
able view of a large reflecting surface. In the geometry shown on the **top**, the
shadowed region of the cylindrical bump sees the plane background at a fairly
favorable angle —if the background is large enough, near half the hemisphere
of the patch at the base of the bump is a view of the plane. This means there
will be a reflex with a large value attached to the edge of the bump and inside
the cast shadow region (which a local model predicts as black). There is another
reflex on the other side, too, as the series of solutions (again normalized for easy
comparison) on the **left** show. On the **right**, an observation of this effect in a real
scene. *Figure from "Mutual Illumination," by D.A. Forsyth and A.P. Zisserman,
Proc. CVPR, 1989, © 1989 IEEE*

## Interreflections

The effects of global shading are often ignored in the shading literature, which causes a reflex
response of hostility in one of the authors. The reason to ignore interreflections is that they are
extremely hard to analyze, particularly from the perspective of inferring object properties given
the output of a global shading model. If interreflection effects do not change the output of a
method much, then it is probably all right to ignore them. Unfortunately, this line of reasoning is
seldom pursued because it is quite difficult to show that a method is stable under interreflections.
The discussion of spatial frequency issues follows Haddon and Forsyth (1998*a*), after an idea
of Koenderink and van Doorn (1983). Apart from this, there is not much knowledge about the
overall properties of interreflected shading, which is an important gap in our knowledge. An
alternative strategy is to iteratively reestimate shape using a rendering model (Nayar, Ikeuchi and
Kanade, 1991*a*).

      Horn is also the first author to indicate the significance of global shading effects (Horn,
1977). Koenderink and van Doorn (1983) noted that the radiosity under a global model is ob-
tained by taking the radiosity under a local model, and applying a linear operator. One then
studies that operator; in some cases, its eigenfunctions (often called *geometrical modes*) are
informative. Forsyth and Zisserman (1989, 1990, 1991) then demonstrated a variety of the qual-
itative effects due to interreflections.

### Photometric Stereo

In its original form, photometric stereo is due to Woodham. There are a number of variants of this useful idea (Horn, Woodham and Silver, 1978, Woodham, 1979, 1980, 1989, 1994). There are a variety of variations on photometric stereo. One interesting idea is to illuminate the surface with three lights of different colors (and in different positions) and use a color image. For an appropriate choice of colors, this is equivalent to obtaining three images, so the measurement process is simplified.

Generally, photometric stereo is used under circumstances where the illumination is quite easily controlled, so that it is possible to ensure that no ambient illumination is in the image. It is relatively simple to insert ambient illumination into the formulation given; we extend the matrix $\mathcal{V}$ by attaching a column of ones. In this case, $g(x, y)$ becomes a four-dimensional vector, and the fourth component is the ambient term. However, this approach does not guarantee that the ambient term is constant over space. Instead, we would have to check that this term was constant and adjust the model if it were not.

Photometric stereo depends only on adopting a local shading model. This model need not be a Lambertian surface illuminated by a distant point source. If the radiosity of the surface is a known function of the surface normal satisfying a small number of constraints, photometric stereo is still possible. This is because the intensity of a pixel in a single view determines the normal up to a one-parameter family. This means that two views determine the normal. The simplest example of this case occurs for a surface of known albedo illuminated by a distant point source.

In fact, if the radiosity of the surface is a $k$-parameter function of the surface normal, photometric stereo is still possible. The intensity of the pixel in a single view determines the normal up to a $k + 1$ parameter family, and $k + 1$ views give the normal. For this approach to work, the radiosity needs to be given by a function for which our arithmetic works (e.g., if the radiosity of the surface is a constant function of the surface normal, it is not possible to infer any constraint on the normal from the radiosity). One can then recover shape and reflectance maps simultaneously (Garcia-Bermejo, Diaz Pernas and Coronado, 1996, Mukawa, 1990, Nayar, Ikeuchi and Kanade, 1990, and Tagare and de Figueiredo, 1992, 1993).

### Alternative Shading Representations

Instead of trying to extract shape information from the shading signal, one might try to match it to a collection of different possible examples. This suggests studying what kinds of shaded view a surface can generate. The collection of available shadings is notably limited (Belhumeur and Kriegman, 1998). A knowledge of this collection's structure is valuable because it makes it possible to understand how to compare shaded images without being confused by illumination changes. Illumination changes are a particular problem in face finding and recognition applications (Adini, Moses and Ullman, 1997, Phillips and Vardi, 1996). Knowing the possible variations in illumination seems to help (Georghiades, Kriegman and Belhumeur, 1998, 2000, Jacobs, Belhumeur and Basri, 1998).

Another possibility is to extend the notion of qualitative analysis of interreflections to obtain a *shading primitive*—a shading pattern that is characteristic and stably linked to a shape pattern. For example, narrow grooves and deep holes in surfaces are dark, and cylinders have a characteristic extended pattern of shading. Few such primitives are known but some appear to be useful (Haddon and Forsyth, 1998*a*, 1998*b*).

## PROBLEMS

**5.1.** What shapes can the shadow of a sphere take if it is cast on a plane and the source is a point source?

**5.2.** We have a square area source and a square occluder, both parallel to a plane. The source is the same size as the occluder, and they are vertically above one another with their centers aligned.
   **(a)** What is the shape of the umbra?
   **(b)** What is the shape of the outside boundary of the penumbra?

**5.3.** We have a square area source and a square occluder, both parallel to a plane. The edge length of the source is now twice that of the occluder, and they are vertically above one another with their centers aligned.
   **(a)** What is the shape of the umbra?
   **(b)** What is the shape of the outside boundary of the penumbra?

**5.4.** We have a square area source and a square occluder, both parallel to a plane. The edge length of the source is now half that of the occluder, and they are vertically above one another with their centers aligned.
   **(a)** What is the shape of the umbra?
   **(b)** What is the shape of the outside boundary of the penumbra?

**5.5.** A small sphere casts a shadow on a larger sphere. Describe the possible shadow boundaries that occur.

**5.6.** Explain why it is difficult to use shadow boundaries to infer shape, particularly if the shadow is cast onto a curved surface.

**5.7.** An infinitesimal patch views a circular area source of constant exitance frontally along the axis of symmetry of the source. Compute the radiosity of the patch due to the source exitance $E(\boldsymbol{u})$ as a function of the area of the source and the distance between the center of the source and the patch. You may have to look the integral up in tables—if you don't, you're entitled to feel pleased with yourself— but this is one of few cases that can be done in closed form. It is easier to look up if you transform it to get rid of the cosine terms.

**5.8.** As in Figure 5.17, a small patch views an infinite plane at unit distance. The patch is sufficiently small that it reflects a trivial quantity of light onto the plane. The plane has radiosity $B(x, y) = 1 + \sin ax$. The patch and the plane are parallel to one another. We move the patch around parallel to the plane, and consider its radiosity at various points.
   **(a)** Show that if one translates the patch, its radiosity varies periodically with its position in $x$.
   **(b)** Fix the patch's center at $(0, 0)$; determine a *closed form* expression for the radiosity of the patch at this point as a function of $a$. You'll need a table of integrals for this (if you don't, you're entitled to feel very pleased with yourself).

**5.9.** If one looks across a large bay in the daytime, it is often hard to distinguish the mountains on the opposite side; near sunset, they are clearly visible. This phenomenon has to do with scattering of light by air—a large volume of air is actually a source. Explain what is happening. We have modeled air as a vacuum and asserted that no energy is lost along a straight line in a vacuum. Use your explanation to give an estimate of the kind of scales over which that model is acceptable.

**5.10.** Read the book *Colour and Light in Nature*, by Lynch and Livingstone, published by Cambridge University Press, 1995.

### Programming Assignments

**5.11.** An area source can be approximated as a grid of point sources. The weakness of this approximation is that the penumbra contains quantization errors, which can be quite offensive to the eye.
   **(a)** Explain.
   **(b)** Render this effect for a square source and a single occluder casting a shadow onto an infinite plane. For a fixed geometry, you should find that as the number of point sources goes up, the quantization error goes down.

**(c)** This approximation has the unpleasant property that it is possible to produce arbitrarily large quantization errors with any finite grid by changing the geometry. This is because there are configurations of source and occluder that produce large penumbrae. Use a square source and a single occluder, casting a shadow onto an infinite plane, to explain this effect.

**5.12.** Make a world of black objects and another of white objects (paper, glue and spraypaint are useful here) and observe the effects of interreflections. Can you come up with a criterion that reliably tells, *from an image*, which is which? (If you can, publish it; the problem looks easy, but isn't).

**5.13.** (This exercise requires some knowledge of numerical analysis.) Do the numerical integrals required to reproduce Figure 5.17. These integrals aren't particularly easy: If one uses coordinates on the infinite plane, the size of the domain is a nuisance; if one converts to coordinates on the view hemisphere of the patch, the frequency of the radiance becomes infinite at the boundary of the hemisphere. The best way to estimate these integrals is using a Monte Carlo method on the hemisphere. You should use importance sampling because the boundary contributes rather less to the integral than the top.

**5.14.** Set up and solve the linear equations for an interreflection solution for the interior of a cube with a small square source in the center of the ceiling.

**5.15.** Implement a photometric stereo system.
  **(a)** How accurate are its measurements (i.e., how well do they compare with known shape information)? Do interreflections affect the accuracy?
  **(b)** How repeatable are its measurements (i.e., if you obtain another set of images, perhaps under different illuminants, and recover shape from those, how does the new shape compare with the old)?
  **(c)** Compare the minimization approach to reconstruction with the integration approach; which is more accurate or more repeatable and why? Does this difference appear in experiment?
  **(d)** One possible way to improve the integration approach is to obtain depths by integrating over many different paths and then average these depths (you need to be a little careful about constants here). Does this improve the accuracy or repeatability of the method?