Contents

CHAPTER 1

Light and Shading

The brightness of a pixel in the image is a function of the brightness of the surface patch in the scene that projects to the pixel. In turn, the brightness of the patch depends on the fraction of the incident light that gets reflected, and on how much incident light arrives at the patch.

Camera response: Modern cameras respond linearly to middling intensities of light, but have pronounced non-linearities for darker and brighter illumination. This allows the camera to reproduce the very wide dynamic range of natural light without saturating. For most purposes, because very bright and very dark patches are not significant cues, it is enough to assume that the camera response is linearly related to the intensity of the surface patch. Write $\boldsymbol{X}$ for a point in space that projects to $\boldsymbol{x}$ in the image, $I_{patch}(\boldsymbol{X})$ for the intensity of the surface patch at $\boldsymbol{X}$, and $I_{camera}(\boldsymbol{x})$ for the camera response at $\boldsymbol{x}$. Then our model is:

$$I_{camera}(\boldsymbol{x}) = kI_{patch}(\boldsymbol{x})$$

where k is some constant to be determined by calibration. Generally, we assume that this model applies and that k is known if needed. Under some circumstances, a more complex model is appropriate: we discuss how to recover such models in section 1.10.

Surface reflection: Different points on a surface may reflect more or less of the light that is arriving. Darker surfaces reflect less light, and lighter surfaces reflect more. While there are a rich set of possible physical effects, most can be ignored. Section 1.1.1 describes the relatively simple model that is sufficient for almost all purposes in computer vision.

Illumination: The amount of light a patch receives depends on the overall intensity of the light, and on the geometry. The overall intensity may change because some **luminaires** (the formal term for light sources) may be shadowed, or may have strong directional components. Geometry affects the amount of light arriving at a patch because surface patches facing the light collect more radiation and so are brighter than surface patches tilted away from the light, an effect known as **shading**. Section 1.1.2 describes the most important model used in computer vision; section **??** describes a much more complex model that is necessary to explain some important practical difficulties in shading inference.

The brightness of a pixel is profoundly ambiguous. Surprisingly, people can disentangle these effects quite effectively. Often, but not always, people can tell whether objects are in bright light or in shadow, and do not perceive objects in shadow as having dark surfaces. People can usually tell whether changes of brightness are caused by changes in reflection or by shading (cinemas wouldn't work if we got it right all the time, however). Typically, people can tell that shading comes from the geometry of the object, but sometimes get shading and markings mixed up. For example, a streak of dark makeup under a cheekbone will often look like a

shading effect, making the face look thinner. Section **??** describes various methods to infer properties of the world from the shading signal.

1.1 MODELLING PIXEL BRIGHTNESS

1.1.1 Reflection at Surfaces

Most surfaces reflect light by a process of **diffuse reflection**. Diffuse reflection scatters light evenly across the directions leaving a surface, so the brightness of a diffuse surface doesn't depend on the viewing direction. Examples are easy to identify with this test: most cloth has this property, as do most paints, rough wooden surfaces, most vegetation, and rough stone or concrete. The only parameter required to describe a surface of this type is its **albedo**, the fraction of the light arriving at the surface that is reflected. This does not depend on the direction in which the light arrives or the direction in which the light leaves. Surfaces with very high or very low albedo are difficult to make. For practical surfaces, albedo lies in the range 0.05-0.90 (see Brelstaff and Blake (), who argue the dynamic range is closer to 10 than the 18 implied by these numbers; also ()`BBrefs`
).

Mirrors are not diffuse, because what you see depends on the direction in which you look at the mirror. The behavior of a perfect mirror is known as **specular reflection**. For an ideal mirror, light arriving along a particular direction can leave only along the **specular direction**, obtained by reflecting the direction of incoming radiation about the surface normal (figure 1.1). Usually some fraction of incoming radiation is absorbed; on an ideal specular surface, this fraction does not depend on the incident direction.

If a surface behaves like an ideal specular reflector, you could use it as a mirror, and, from this test, relatively few surfaces actually behave like ideal specular reflectors. Imagine a near perfect mirror made of polished metal; if this surface is suffers slight damage at a small scale, then around each point there will be a set of small facets, pointing in a range of directions. In turn, this means that light arriving in one direction will leave in several different directions, because it strikes several facets, and so the specular reflections will be blurred. As the surface becomes less flat, these distortions will become more pronounced; eventually, the only specular reflection that is bright enough to see will come from the light source. This mechanism means that, in most shiny paint, plastic, wet or brushed metal surfaces, one sees a bright blob — often called a **specularity** — along the specular direction from light sources, but few other specular effects. Specularities are easy to identify, because they are small and very bright (Figure 1.2; ()).

1.1.2 Sources and their Effects

The main source of illumination outside is the sun, whose rays all travel parallel to one another in a known direction because it is so far away. We model this behavior with a **distant point light source**. This is the most important model of lighting, and is quite effective for indoor scenes as well as outdoor scenes. Because the rays are parallel to one another, a surface that faces the source cuts more rays (and so collects more light) than one oriented along the direction the rays travel. The

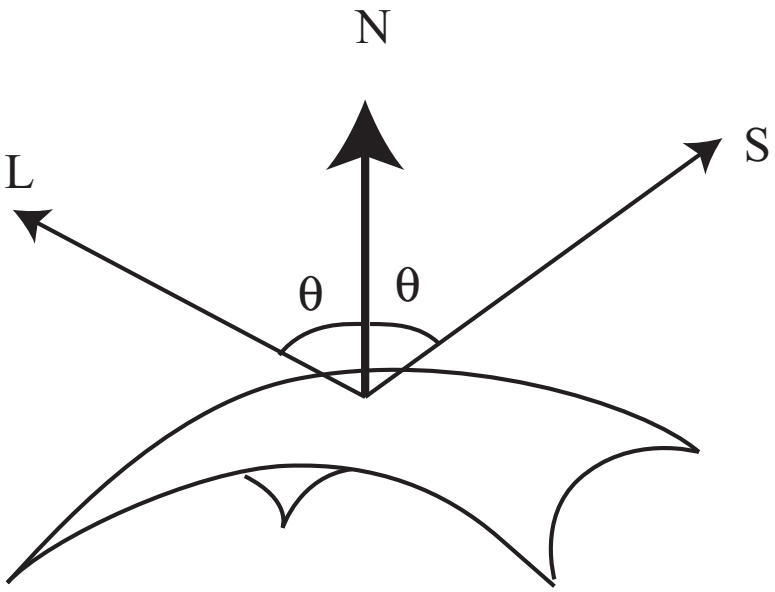


FIGURE 1.1: A specular surface patch, showing the surface normal ($\boldsymbol{N}$), and the direction to the light source ($\boldsymbol{L}$). The specular direction $\boldsymbol{S}$ is coplanar with the normal and the source direction, and has the same angle to the normal that the source direction does.

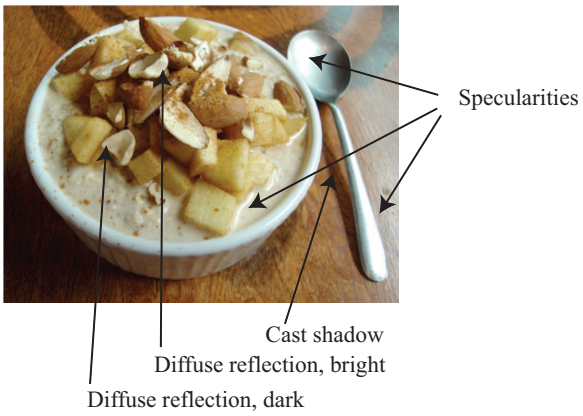


FIGURE 1.2: This photograph, published on flickr by mlinksva, illustrates a variety of illumination effects. There are specularities on the metal spoon and on the milk. The bright diffuse surface is bright because it faces the light direction. The dark diffuse surface is dark because it is tangential to the illumination direction. The shadows appear at surface points that cannot see the light source.

amount of light collected by a surface patch in this model is proportional to the cosine of the angle θ between the illumination direction and the normal (Figure 1.3). The figure yields **Lambert's cosine law**, which states the brightness of a diffuse

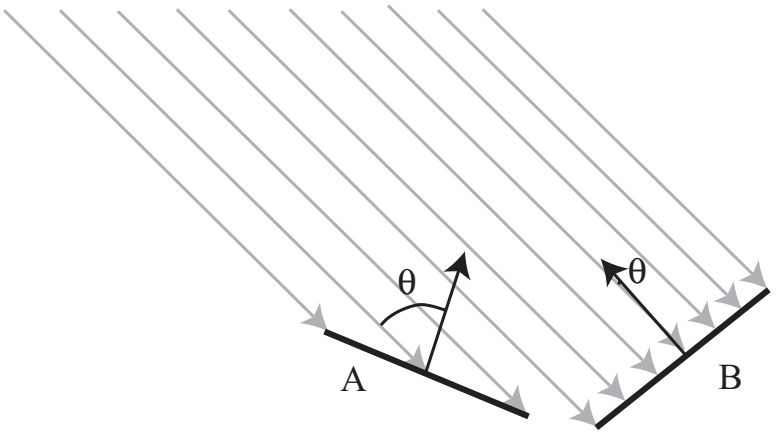


FIGURE 1.3: Two surface patches are illuminated by a distant point source, whose rays are shown as light arrowheads. Patch A is tilted away from the source (θ is close to 90^0) and collects less energy, because it cuts fewer light rays per unit surface area. Patch B, facing the source (θ is close to 0^0), collects more energy.

patch illuminated by a distant point light source is given by

$$I = \rho I_0 \cos\theta,$$

where I_0 is the intensity of the light source, θ is the angle between the light source direction and the surface normal, and ρ is the diffuse albedo. This law predicts bright image pixels come from surface patches that face the light directly and dark pixels come from patches that see the light only tangentially, so that the shading on on a surface provides some shape information. We explore this cue in 1.3.6.

If the surface cannot see the source, then it is in **shadow**. Since we assume that light arrives at our patch only from the distant point light source, our model suggests that shadows are deep black; in practice, they very seldom are, because the shadowed surface usually receives light from other sources. Outdoors, the most important such source is the sky, which is quite bright. Indoors, light reflected from other surfaces illuminates shadowed patches. These **interreflections** can have a significant effect on the brightness of other surfaces, too. These effects are sometimes modelled by adding a constant **ambient illumination** term to the predicted intensity.

1.1.3 The Lambertian + Specular model

For almost all purposes, it is enough to model all surfaces as being diffuse with specularities. This is the **lambertian+specular model**. Specularities are relatively seldom used in inference (section 1.3.2 sketches two methods), and so there is no need to have a formal model of their structure. Because specularities are small and bright, they are relatively easy to identify and remove with straightforward methods (find small bright spots, and replace them by smoothing the local pixel values). More sophisticated specularity finders use color information (section **??**).

Thus, to apply the lambertian+specular model, we find and remove specularities, and then use Lambert's law (section 1.1.2) to model image intensity. By far the most usual case involves an infinitely distant source. For this case, write $\boldsymbol{N}(x)$ for the unit surface normal at $\boldsymbol{x}$, $\boldsymbol{S}$ for a vector pointing from $\boldsymbol{x}$ toward the source with length I_o (the source intensity), $\rho(\boldsymbol{x})$ for the albedo at $\boldsymbol{x}$, and $Vis(\boldsymbol{S}, \boldsymbol{x})$ for a function that is one when $\boldsymbol{x}$ can see the source and zero otherwise. Then the intensity at $\boldsymbol{x}$ is

$$I(\boldsymbol{x}) = \rho(\boldsymbol{x})\,(\boldsymbol{N}\cdot\boldsymbol{S})\,\mathrm{Vis}(\boldsymbol{S},\,\boldsymbol{x}) + \rho(\boldsymbol{x})A + S$$

$$\begin{array}{ccccccc} \text{Image} & = & \text{Diffuse} & + & \text{Ambient} & + & \text{Specular} \\ \text{intensity} & & \text{term} & & \text{term} & & \text{term} \end{array}$$

This model can still be used for a more complex source (for example, an area source), but in that case it is more difficult to determine an appropriate $\boldsymbol{S}(\boldsymbol{x})$.

1.2 AREA SOURCES AND INTERREFLECTIONS

An **area source** is an area that radiates light. Area sources occur quite commonly in natural scenes — an overcast sky is a good example — and in synthetic environments — for example, the fluorescent light boxes found in many industrial ceilings. Shadows from area sources are very different from shadows cast by point sources, and interreflections are explained by an area source model, too.

1.2.1 Area Source Shadows

One seldom sees dark shadows with crisp boundaries indoors. Instead, one could see no visible shadows, or shadows that are rather fuzzy diffuse blobs, or sometimes fuzzy blobs with a dark core (figure 1.4). These effects occur because rooms tend to have light walls and diffuse ceiling fixtures which act as area sources. As a result, the shadows one sees are area source shadows.

To compute the intensity at a surface patch illuminated by an area source, we can break the source up into infinitesimal source elements, then sum effects from each element. If there is an occluder, then some surface patches may see none of the source elements. Such patches will be dark, and lie in the **umbra** (a Latin word meaning "shadow"). Other surface patches may see some, but not all, of the source elements. Such patches may be quite bright (if they see most of the elements), or relatively dark (if they see few elements), and lie in the **penumbra** (a compound of Latin words meaning "almost shadow"). One way to build intuition is to think of a tiny observer looking up from the surface patch. At umbral points, this observer will not see the area source at all whereas at penumbral points, the observer will see some, but not all, of the area source. An observer moving from outside the shadow, through the penumbra and into the umbra will see something that looks like an eclipse of the moon (figure 1.5). The penumbra can be large, and can change quite slowly from light to dark. There may even be no umbral points at all, and, if the occluder is sufficiently far away from the surface, the penumbra could be very large and almost indistinguishable in brightness from the unshadowed patches. This is why many objects in rooms appear to cast no shadow at all (figure 1.6).

FIGURE 1.4: This photograph, published on Flickr by Reiner Kraft, shows characteristic area source shadow effects. Notice that A is much darker than B; there must be some shadowing effect here, but there is no clear shadow boundary. Instead, there is a fairly smooth gradient. The pillow casts a complex shadow, with two distinct regions: there is a core of darkness (the *umbra*) surrounded by a partial shadow (*penumbra*).

1.2.2 Modelling Interreflection

In an interreflection model, each surface patch receives power from all the radiating surfaces it can see. These surfaces might radiate power that they generate internally because they are luminaires, or they might simply reflect power. The general form of the model will be:

$$\begin{pmatrix}\text{Power leaving}\\ \text{a patch}\end{pmatrix} = \begin{pmatrix}\text{Power generated}\\ \text{by that patch}\end{pmatrix} + \begin{pmatrix}\text{Power received}\\ \text{from other patches and reflected}\end{pmatrix}$$

This means we need to be able to model the power received from other patches and reflected. We will develop a model assuming that all surfaces are diffuse.e model, we will assume that all surfaces are diffuse. This leads to a somewhat simpler model, and describes all effects that are currently of interest to vision (it is complicated, but not difficult, to build more elaborate models). We will also need some radiometric terminology.

1.2.2.1 The Illumination at a Patch due to an Area Source

Consider a surface patch relatively close to the area source. The source will subtend a large solid angle at the patch, and if the patch moves a little further from the source, this solid angle will hardly change (if the source was infinite, it would not change at all). As a result, the amount of light the patch collects from the source will hardly change. This explains the widespread use of area sources in illumination engineering — they generally yield fairly uniform illumination. To build a more detailed model, we need units in which to measure illumination.

The appropriate unit is **radiance**, defined as

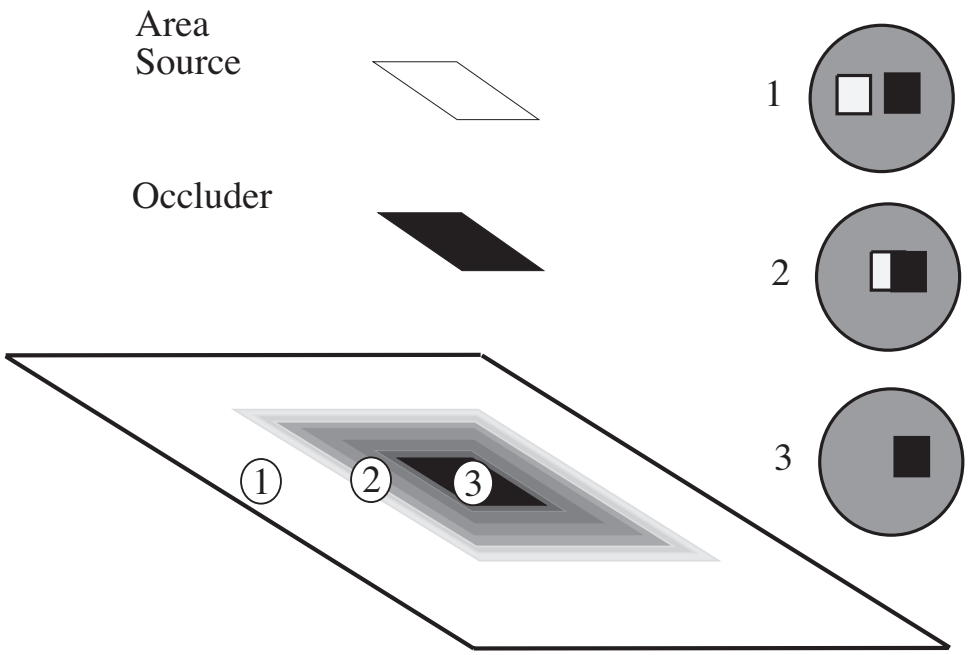


FIGURE 1.5: Area sources generate complex shadows with smooth boundaries, because from the point of view of a surface patch, the source disappears slowly behind the occluder. Regions where the source cannot be seen at all are known as the *umbra*; regions where some portion of the source is visible are 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.

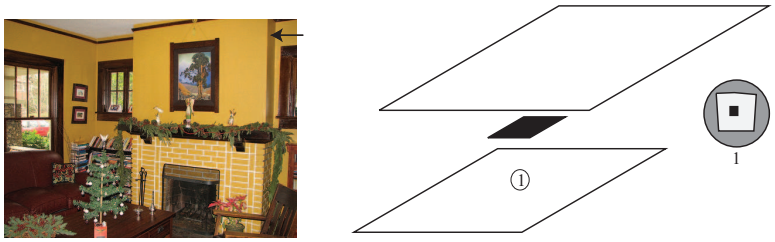


FIGURE 1.6: The photograph on the **left**, published on Flickr by MGShelton, shows an interior decorated for a holiday. Notice the lighting has some directional component (the vertical stripe indicated by the arrow is dark, because it does not face the main direction of lighting), but there are few visible shadows. On the **right**, a drawing to show why; here there is a small occluder and a large area source. The occluder is some way away from the shaded surface. Generally, at points on the shaded surface the incoming hemisphere looks like that at point 1. The occluder blocks out some small percentage of the area source, but the amount of light lost is too small to notice (compare figure 1.5).

the power (amount of energy per unit time) traveling at some point in a specified direction, per unit area *perpendicular to the direction of travel*, per unit solid angle

The units of radiance are watts per square meter per steradian ($Wm^{-2}sr^{-1}$). The definition of radiance may look strange, but it is consistent with the most basic phenomenon in radiometry: A small surface patch viewing a source frontally collects more energy than the same patch viewing a source radiance along a nearly tangent direction — the amount of energy a patch collects from a source depends both on how large the source looks from the patch *and* on how large the patch looks from the source.

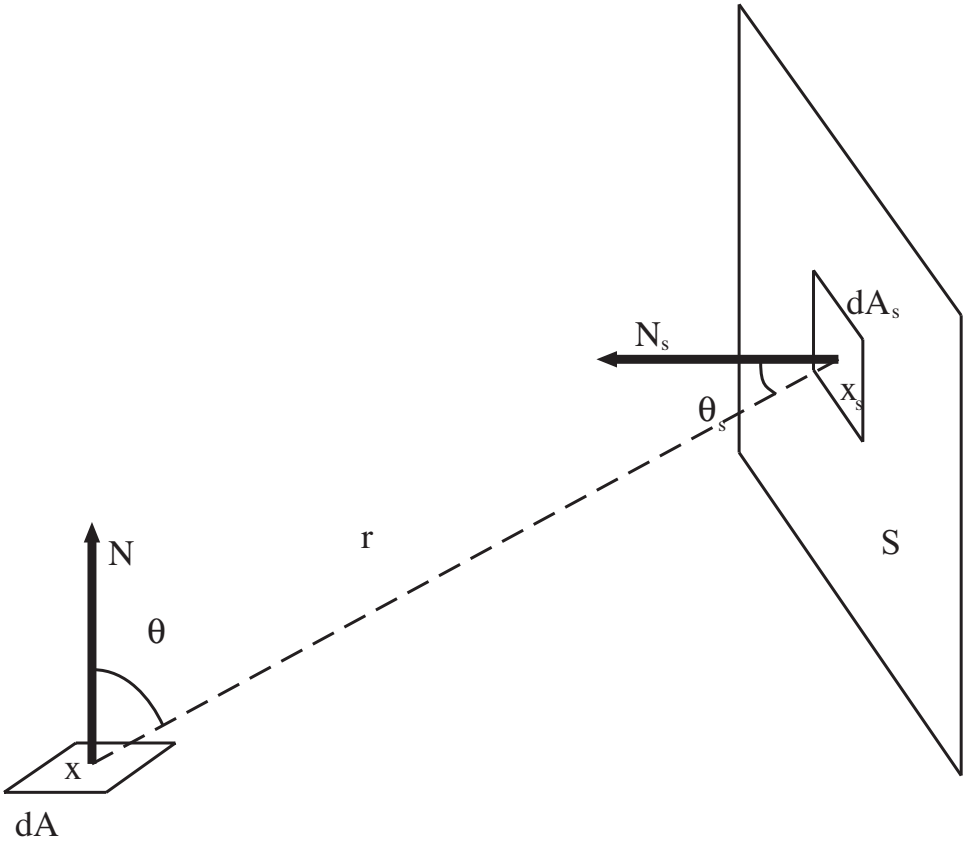


FIGURE 1.7: Terminology for the power transferred from an element to an element, and from an area source to an element. One can use this power to establish that radiance does not go down along straight line paths in a vacuum (or, for reasonable distances, in clear air), using an energy conservation argument sketched in the text.

It is important to remember that the square meters in the units for radiance are *foreshortened* (i.e., perpendicular to the direction of travel) to account for this phenomenon. Assume we have two elements, one at $\boldsymbol{x}$ with area dA and the other

at $\boldsymbol{x}_s$ with area dA_s. Write the angular direction from $\boldsymbol{x}$ to $\boldsymbol{x}_s$ as $\boldsymbol{x} \to \boldsymbol{x}_s$, and define the angles θ and θ_s as in Figure 1.7. Then the the solid angle subtended by element 2 at element 1 is

$$d\omega_{2(1)} = \frac{\cos\theta_s dA_s}{r^2},$$

so the power leaving $\boldsymbol{x}$ toward $\boldsymbol{x}_s$ is

$$\begin{aligned} d^2P_{1\to 2} &= (\text{radiance})(\text{foreshortened area})(\text{solid angle}) \\ &= L(\boldsymbol{x}, \boldsymbol{x} \to \boldsymbol{x}_s)(\cos\theta dA)(d\omega_{2(1)}) \\ &= L(\boldsymbol{x}, \boldsymbol{x} \to \boldsymbol{x}_s)\left(\frac{\cos\theta\cos\theta_s}{r^2}\right) dA_s dA. \end{aligned}$$

By a similar argument, the same expression yields the power arriving at $\boldsymbol{x}$ from $\boldsymbol{x}_2$; this means that, in a vacuum, *radiance is constant along (unoccluded) straight lines*.

FIGURE 1.8: A patch with area dA views an area source S. We compute the power received by the patch by summing the contributions of each element on S.

We can now compute the power that an element dA collects from an area source, by summing the contributions of elements over that source. Using the notation of figure 1.8, we get

$$dP_{S\to dA} = \left(\int_S L(\boldsymbol{x}_s, \boldsymbol{x}_s \to \boldsymbol{x})\left(\frac{\cos\theta_s\cos\theta}{r^2}\right) dA_s\right) dA$$

To get a more useful area source model, we need further units.

1.2.2.2 Radiosity and Exitance

We are dealing with diffuse surfaces, and our definition of a diffuse surface is that the intensity (formally, the radiance) leaving the surface is independent of the direction in which it leaves. There is no point in describing the intensity of such a surface with radiance (which explicitly depends on direction). The appropriate unit is **radiosity**, defined as

the total power leaving a point on a surface per unit area on the surface

Radiosity, which is usually written as $B(\boldsymbol{x})$, has units watts per square meter (Wm^{-2}). To obtain the radiosity of a surface at a point, we can sum the radiance leaving the surface at that point over the whole exit hemisphere. Thus, if $\boldsymbol{x}$ is a point on a surface emitting radiance $L(\boldsymbol{x}, \theta, \phi)$, the radiosity at that point is

$$B(\boldsymbol{x}) = \int_\Omega L(\boldsymbol{x}, \theta, \phi)\cos\theta d\omega,$$

where Ω is the exit hemisphere, $d\omega$ is unit solid angle, and the term $\cos\theta$ turns foreshortened area into area (look at the definitions again). We could substitute $d\omega = \sin\theta d\theta d\phi$, using the units of figure 1.9.

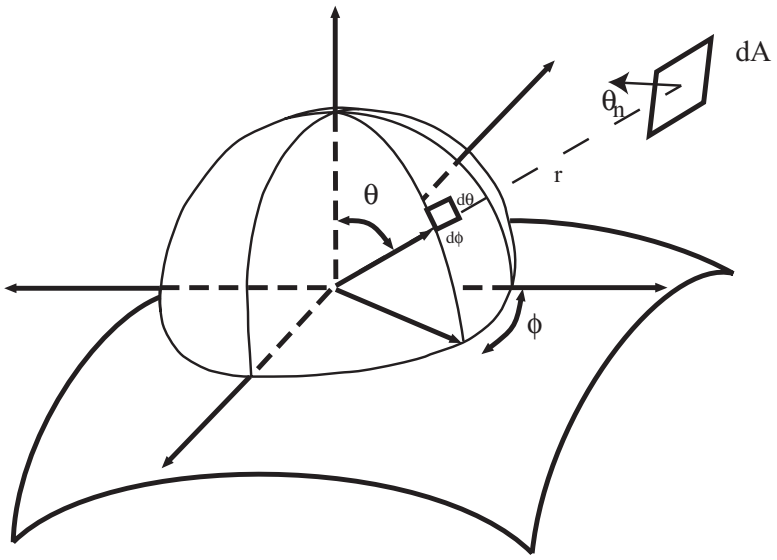


FIGURE 1.9: Angular coordinates on a sphere. An infinitesimal patch of area dA at distance r from the center subtends solid angle $d\omega = (1/r^2)\cos\theta_n dA$, using the notation of the figure. This solid angle $d\omega$ is cut out by a step of $d\phi$ in ϕ and a step of $d\theta$ in θ, so we can write $d\omega = \sin\theta d\theta d\phi$.

Consider a surface element as in figure 1.8. We have computed how much power it receives from the source as a function of the source's radiance. The surface element is diffuse, and its albedo is $\rho(\boldsymbol{x})$. The albedo is the fraction of incoming power that the surface radiates, so the radiosity due to power received from the area source is

$$B(\boldsymbol{x}) = \frac{dP_{S\to dA}}{dA} = \left(\int_S L(\boldsymbol{x}_s, \boldsymbol{x}_s \to \boldsymbol{x}) \left(\frac{\cos\theta_s \cos\theta}{r^2}\right) dA_s\right)$$

Now if a point $\boldsymbol{u}$ on a surface has radiosity $B(\boldsymbol{u})$, what is the radiance leaving the surface in some direction? We write L for the radiance, which is independent of angle, and we must have

$$B(\boldsymbol{u}) = L(\boldsymbol{u}) \int_\Omega L(\boldsymbol{x}, \theta, \phi) \cos\theta d\omega = L(\boldsymbol{u})\pi$$

This means that if the area source has radiosity $B(\boldsymbol{x}_s)$, then the radiosity at the element *due to the power received from the area source* is

$$B(\boldsymbol{x}) = \frac{\rho}{\pi}\left(\int_S L(\boldsymbol{x}_s, \boldsymbol{x}_s \to \boldsymbol{x}) \left(\frac{\cos\theta_s \cos\theta}{r^2}\right) dA_s\right)$$

Our final step is to model illumination that is generated internally in a surface — light generated by a luminaire, rather than reflected from a surface. We assume there are no directional effects in the luminaire, and that power is uniformly distributed across outgoing directions (this is the least plausible component of the model, but is usually tolerable). We use the unit **exitance**, which is defined as

the total power internally generated power leaving a point on a surface per unit area on the surface.

1.2.2.3 An Interreflection Model

We can now write a formal model of interreflections for diffuse surfaces by substituting terms into the original expression. Recall that radiosity is power per unit area, write $E(\boldsymbol{x})$ for exitance at the point $\boldsymbol{x}$, write $\boldsymbol{x}_s$ for a coordinate that runs over all surface patches, $\mathcal{S}$ for the set of all surfaces, dA for the element of area at $\boldsymbol{x}$, $\text{Vis}(\boldsymbol{x}, \boldsymbol{x}_s)$ for a function that is one if the two points can see each other and zero otherwise, and $\cos\theta$, $\cos\theta_s$, r as in figure 1.8. We obtain

$$\begin{array}{ccccc} \text{Power leaving a patch} & = & \text{Power generated by that patch} & + & \text{Power received from other patches and reflected} \\ B(\boldsymbol{x})dA & = & E(\boldsymbol{x})dA & + & \frac{\rho(\boldsymbol{x})}{\pi}\int_{\mathcal{S}}\left[\frac{\cos\theta\cos\theta_s}{r^2}\text{Vis}(\boldsymbol{x},\boldsymbol{x}_s)\right]B(\boldsymbol{x_s})dA_s dA \end{array}$$

and so, dividing by area, we have

$$B(\boldsymbol{x}) = E(\boldsymbol{x}) + \frac{\rho(\boldsymbol{x})}{\pi}\int_{\mathcal{S}}\left[\frac{\cos\theta\cos\theta_s}{r^2}\text{Vis}(\boldsymbol{x},\boldsymbol{x}_s)\right]B(\boldsymbol{x_s})dA_s$$

It is usual to write

$$K(\boldsymbol{x},\boldsymbol{x}_s) = \frac{\cos\theta\cos\theta_s}{\pi r^2}$$

and refer to K as the interreflection kernel. Substituting gives

$$B(\boldsymbol{x}) = E(\boldsymbol{x}) + \rho(\boldsymbol{x})\int_{\mathcal{S}} K(\boldsymbol{x},\boldsymbol{x}_s)\text{Vis}(\boldsymbol{x},\boldsymbol{x}_s)B(\boldsymbol{x}_s)dA_{\boldsymbol{x}_s}$$

an equation where 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 places to start for this topic are **?**) or **?**)). The model produces good predictions of observed effects (Figure 1.18).

1.2.2.4 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 $\boldsymbol{B}$, which contains the value of the radiosity for each patch. In particular, the ith component of $\boldsymbol{B}$ is the radiosity of the ith patch.

We write the incoming radiosity at the ith patch due to radiosity on the jth patch as

$$B_{j\to i}(\boldsymbol{x}) = \rho(\boldsymbol{x})\int_{\text{patch } j}\text{visible}(\boldsymbol{x},\boldsymbol{v})K(\boldsymbol{x},\boldsymbol{v})dA_{\boldsymbol{v}}B_j,$$

where $\boldsymbol{x}$ is a coordinate on the ith patch and $\boldsymbol{v}$ is a coordinate on the jth patch. Now this expression is not a constant, and so we must average it over the ith patch to get

$$\overline{B}_{j\to i} = \frac{1}{A_i}\int_{\text{patch } i} \rho_d(\boldsymbol{x}) \int_{\text{patch } j} \text{visible}(\boldsymbol{x},\boldsymbol{v})K(\boldsymbol{x},\boldsymbol{v})dA_{\boldsymbol{v}}dA_{\boldsymbol{x}}B_j,$$

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

$$\begin{aligned} B_i &= E_i + \sum_{\text{all } j} \overline{B}_{j\to i} \\ &= E_i + \sum_{\text{all } j} K_{ij}B_j, \end{aligned}$$

where

$$K_{ij} = \frac{1}{A_i}\int_{\text{patch } i} \rho_d(\boldsymbol{x}) \int_{\text{patch } j} \text{visible}(\boldsymbol{x},\boldsymbol{v})K(\boldsymbol{x},\boldsymbol{v})dA_{\boldsymbol{v}}dA_{\boldsymbol{x}}.$$

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; **?**) is an excellent account, as is the book of **?**).

1.3 INFERENCE FROM SHADING

1.3.1 Radiometric Calibration and High Dynamic Range Images

Real scenes often display a range of intensities that is much larger than cameras can cope with. Film and CCD's respond to energy. A property called **reciprocity** means that, if a scene patch casts intensity E onto the film, and if the shutter is open for time Δt, the response is a function of $E\Delta t$ alone. In particular, we will get the same outcome if we image one patch of intensity E for time Δt and another patch of intensity E/k for time $k\Delta t$. The actual response that the film produces is a function of $E\Delta t$; this function may depend on the imaging system, but is typically somewhat linear over some range, and sharply non-linear near the top and bottom of this range, so that the image can capture very dark and very light patches without saturation. It is usually monotonically increasing.

There are a variety of applications where it would be useful to know the actual radiance (equivalently, intensity) arriving at the imaging device. For example, we might want to compare renderings of a scene with pictures of the scene, and to do that we need to work in real radiometric units. We might want to use pictures of a scene to estimate the lighting in that scene so we can postrender new objects into the scene; these would need to be lit correctly. To infer radiance, we must determine the film response, a procedure known as **radiometric calibration**. As we shall see, to do this will require more than one image of a scene, obtained at different exposure settings. Imagine we are looking at a scene of a stained glass

window lit from behind in a church. At one exposure setting, we would be able to resolve detail in the dark corners, but not on the stained glass, which would be saturated. At another setting, we would be able to resolve detail on the glass, but the interior would be too dark. If we have both settings, we may as well try to recover radiance with a very large dynamic range — producing a **high dynamic range image**.

Now assume we have multiple registered images, each obtained using a different exposure time. At the i, j'th pixel, we know the image intensity value $I_{ij}^{(k)}$ for the k'th exposure time, we know the value of the k'th exposure time Δt_k, and we know that the intensity of the corresponding surface patch E_{ij} is the same for each exposure, but we do not know the value of E_{ij}. Write the camera response function f, so that

$$I_{ij}^{(k)} = f(E_{ij}\Delta t_k).$$

There are now several possible approaches to solve for f. We could assume a parametric form — say, polynomial — then solve using least squares. Notice that we must solve not only for the parameters of f, but also for E_{ij}. For a color camera, we solve for calibration of each channel separately. Mitsunaga and Nayar have studied the polynomial case in detail (). Though the solution is not unique, ambiguous solutions are strongly different from one another, and most cases are easily ruled out. Furthermore, one does not need to know exposure times with exact accuracy to estimate a solution, as long as there are sufficient pixel values; instead, one estimates f from a fixed set of exposure times, then estimates the exposure times from f, then re-estimates. This procedure is stable.

Alternatively, because the camera response is monotonic, we can work with its inverse $g = f^{-1}$, take logs, and write

$$\log g(I_{ij}^{(k)}) = \log E_{ij} + \log \Delta t_k$$

We can now estimate the values that g takes at each point and the E_{ij} by placing a smoothness penalty on g. In particular, we minimize

$$\sum_{i,j,k} (\log g(I_{ij}^{(k)}) - \log E_{ij} + \log \Delta t_k)^2 + \text{smoothness penalty on } g$$

by choice of g. Debevec and Malik penalize the second derivative of g (). Once we have a radiometrically calibrated camera, estimating a high dynamic range image is relatively straightforward. We have a set of registered images. At each pixel location, we seek the estimate of radiance that predicts the registered image values best. In particular, we assume we know f. We seek an E_{ij} such that

$$\sum_k w(I_{ij})(I_{ij}^{(k)} - f(E_{ij}\Delta t_k))^2$$

is minimized. Notice the weights, because our estimate of f is more reliable when I_{ij} is in the middle of the available range of values than at larger or at smaller values.

Registered images are not essential for radiometric calibration. For example, it is sufficient to have two images where we believe the histogram of E_{ij} values is

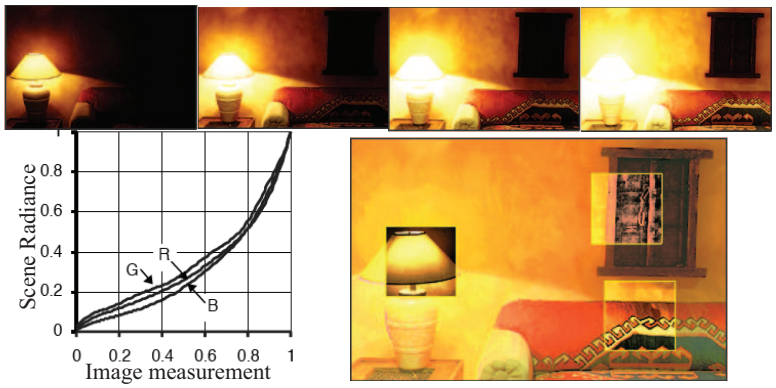


FIGURE 1.10: It is possible to calibrate the radiometric response of a camera from multiple images obtained at different exposures. The **top** row four different exposures of the same scene, ranging from darker (shorter shutter time) to lighter (longer shutter time). Note how, in the dark frames, the lighter part of the image shows detail, and in the light frames, the darker part of the image shows detail; this is the result of non-linearities in the camera response. On the **bottom left**, we show the inferred calibration curves for each of the R, G and B camera channels. On the **bottom right**, a composite image, illustrating the results. The dynamic range of this image is far too large to print; instead, the main image is normalized to the print range. Overlaid on this image are a set of boxes where the radiances in the box have also been normalized to the print range; these show how much information is packed into the high dynamic range image. *Figure 7 a, b, c from "Radiometric Self Calibration", Mitsunaga and Nayar, in the fervent hope that permission will be granted*

the same (). This occurs, for example, when the images are of the same scene, but are not precisely registered. Another option is to modify the imaging device; if each 2x2 block of pixels (say) is modified so that three of the pixels have known neutral density filters in front of them, it is possible to pay camera resolution to obtain increased dynamic range ().

1.3.2 Curvature and Specularities

`Either refer to or use the writeup in geofeats here`

1.3.3 Inferring Lightness and Illumination

Lightness constancy is the skill that allows humans to report whether a surface is white, gray, or black (the **lightness** of the surface) despite changes in the intensity of illumination (the **brightness**). There is a lot of evidence that human lightness constancy involves two processes: One compares the brightness of various image patches and uses this comparison to determine which patches are lighter and which darker; the second establishes some form of absolute standard to which these comparisons can be referred (e.g. **?**)).

1.3.3.1 A Simple Model of Image Brightness

The radiance arriving at a pixel depends on the illumination of the surface being viewed, its BRDF, its configuration with respect to the source, and the camera responses. We assume that the scene is plane and frontal, that surfaces are Lambertian (or that specularities have been removed), and that the camera responds linearly to radiance.

This yields a model of the camera response C at a point $\boldsymbol{x}$ as the product of an illumination term, an albedo term, and a constant that comes from the camera gain:

$$C(\boldsymbol{x}) = k_c I(\boldsymbol{x})\rho(\boldsymbol{x}).$$

If we take logarithms, we get

$$\log C(\boldsymbol{x}) = \log k_c + \log I(\boldsymbol{x}) + \log \rho(\boldsymbol{x}).$$

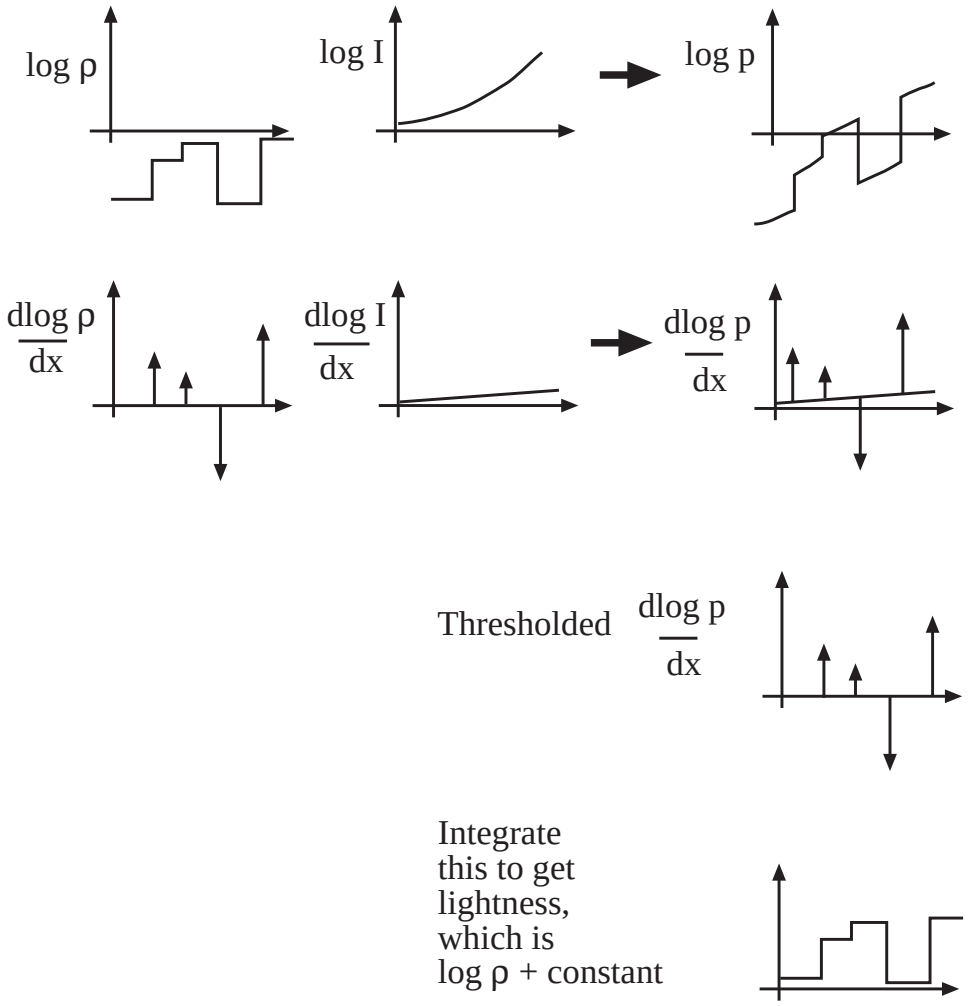
We now make a second set of assumptions:

- First, we assume that albedoes change only quickly over space — this means that a typical set of albedoes will look like a collage of papers of different grays. This assumption is quite easily justified: There are relatively few continuous changes of albedo in the world (the best example occurs in ripening fruit), and changes of albedo often occur when one object occludes another (so we would expect the change to be fast). This means that spatial derivatives of the term $\log \rho(\boldsymbol{x})$ are either zero (where the albedo is constant) or large (at a change of albedo).
- Second, illumination changes only slowly over space. This assumption is somewhat realistic. For example, the illumination due to a point source will change relatively slowly unless the source is very close — so the sun is a source that is particularly good for this example. As another example, illumination inside rooms tends to change very slowly because the white walls of the room act as area sources. This assumption fails dramatically at shadow boundaries, however. We have to see these as a special case and assume that either there are no shadow boundaries or that we know where they are.

1.3.3.2 Recovering Lightness from the Model

It is relatively easy to build algorithms that use our model. The earliest algorithm (the Retinex algorithm of **?**)) has fallen into disuse. The key insight of Retinex is that small gradients are changes in illumination, and large gradients are changes in lightness. We can use this by differentiating the log transform, throwing away small gradients, and integrating the results (**?**). There is a constant of integration missing, so lightness ratios are available, but absolute lightness measurements are not. Figure 1.11 illustrates the process for a one-dimensional example, where differentiation and integration are easy.

This approach can be extended to two dimensions as well. Differentiating and thresholding is easy: At each point, we estimate the magnitude of the gradient; if the magnitude is less than some threshold, we set the gradient vector to zero or



FIGUREE 1.11: The lightness algorithm is easiest to illustrate for a 1D image. In the **top row**, the graph on the **left** shows $\log \rho(x)$, that on the **center** $\log I(x)$, and that on the **right** their sum, which is $\log C$. The log of image intensity has large derivatives at changes in surface reflectance and small derivatives when the only change is due to illumination gradients. Lightness is recovered by differentiating the log intensity, thresholding to dispose of small derivatives, and integrating at the cost of a missing constant of integration.

else we leave it alone. The difficulty is in integrating these gradients to get the log albedo map. The thresholded gradients may not be the gradients of an image because the mixed second partials may not be equal (integrability again; compare with Section 1.3.4.4).

The problem can be rephrased as a minimization problem: choose the log albedo map whose gradient is most like the thresholded gradient. This is a relatively

Algorithm 1.1: Determining the Lightness of Image Patches

Form the gradient of the log of the image
At each pixel, if the gradient magnitude is below
  a threshold, replace that gradient with zero
Reconstruct the log-albedo by solving the minimization
  problem described in the text
Obtain a constant of integration
Add the constant to the log-albedo, and exponentiate

simple problem because computing the gradient of an image is a linear operation. The x-component of the thresholded gradient is scanned into a vector $\boldsymbol{p}$ and the y-component is scanned into a vector $\boldsymbol{q}$. We write the vector representing log-albedo as $\boldsymbol{l}$. Now the process of forming the x derivative is linear, and so there is some matrix $\mathcal{M}_x$, such that $\mathcal{M}_x\boldsymbol{l}$ is the x derivative; for the y derivative, we write the corresponding matrix $\mathcal{M}_y$.

The problem becomes to find the vector $\boldsymbol{l}$ that minimizes

$$| \mathcal{M}_x\boldsymbol{l} - \boldsymbol{p} |^2 + | \mathcal{M}_y\boldsymbol{l} - \boldsymbol{q} |^2 .$$

This is a quadratic minimization problem, and the answer can be found by a linear process. Some special tricks are required because adding a constant vector to $\boldsymbol{l}$ cannot change the derivatives, so the problem does not have a unique solution. We explore the minimization problem in the exercises.

The constant of integration needs to be obtained from some other assumption. There are two obvious possibilities:

- we can assume that the *brightest patch is white*;
- we can assume that the *average lightness is constant*.

We explore the consequences of these models in the exercises.

More sophisticated algorithms are now available, but there were no quantitative studies of performance until recently. Grosse *et al.* built a dataset for evaluating lightness algorithms, and show that a version of the procedure we describe performs extremely well compared to more sophisticated algorithms (). The major difficulty all these approaches find is with shadow boundaries, which we discuss in section **??**.

1.3.4 Photometric Stereo: Shape from multiple shaded images

It is possible to reconstruct a patch of surface from a series of pictures of that surface taken under different illuminants. First, we need a camera model. For simplicity, we choose a camera situated so that the point (x, y, z) in space is imaged to the

point (x, y) in the camera (the method we describe works for the other camera models described in chapter **??**).

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 1.12). 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.

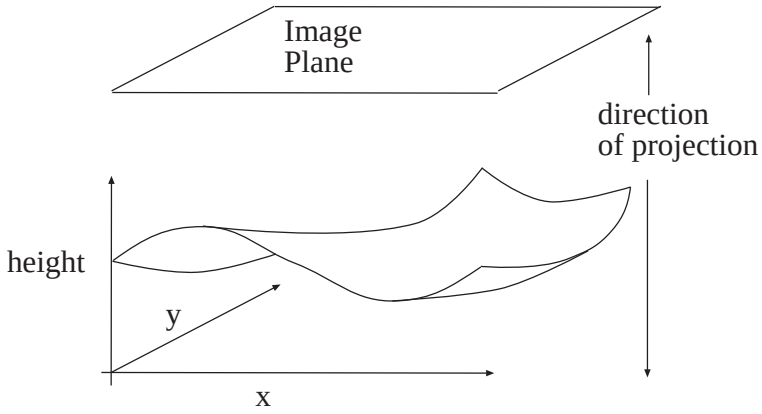


FIGURE 1.12: A Monge patch is a representation of a piece of surface as a height function. For the photometric stereo example, we assume that an orthographic camera — one that maps (x, y, z) in space to (x, y) in the camera — is viewing a Monge patch. This means that the shape of the surface can be represented as a function of position in the image.

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 different images of a surface in a fixed view illuminated by different sources. This method recovers 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 $\boldsymbol{x}$ on the surface is

$$B(\boldsymbol{x}) = \rho(\boldsymbol{x})\boldsymbol{N}(\boldsymbol{x}) \cdot \boldsymbol{S}_1,$$

where $\boldsymbol{N}$ is the unit surface normal and $\boldsymbol{S}_1$ is the source vector. We can write $B(x, y)$ for the radiosity of a point on the surface because there is only one point on the surface corresponding to the point (x, y) in the camera. Now we assume that the response of the camera is linear in the surface radiosity, and so have that the

value of a pixel at (x, y) is

$$\begin{aligned} I(x,y) &= kB(\boldsymbol{x}) \\ &= kB(x,y) \\ &= k\rho(x,y)\boldsymbol{N}(x,y)\cdot\boldsymbol{S}_1 \\ &= \boldsymbol{g}(x,y)\cdot\boldsymbol{V}_1, \end{aligned}$$

where $\boldsymbol{g}(x,y) = \rho(x,y)\boldsymbol{N}(x,y)$ and $\boldsymbol{V}_1 = k\boldsymbol{S}_1$, where k is the constant connecting the camera response to the input radiance.

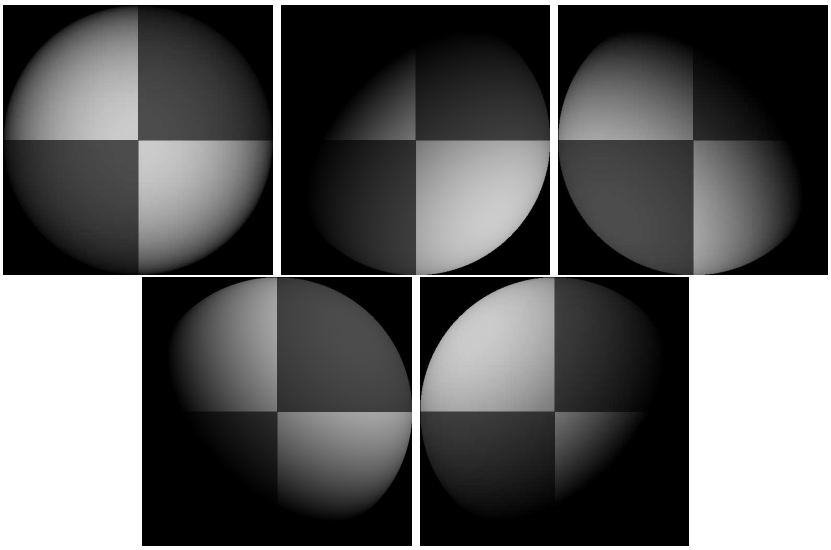


FIGURE 1.13: Five synthetic images of a sphere, all obtained in an orthographic view from the same viewing position. These images are shaded using a local shading model and a distant point source. This is a convex object, so the only view where there is no visible shadow occurs when the source direction is parallel to the viewing direction. The variations in brightness occuring under different sources code the shape of the surface.

In these equations, $\boldsymbol{g}(x,y)$ describes the surface and $\boldsymbol{V}_1$ is a property of the illumination and of the camera. We have a dot product between a vector field $\boldsymbol{g}(x,y)$ and a vector $\boldsymbol{V}_1$, which could be measured; with enough of these dot products, we could reconstruct $\boldsymbol{g}$ and so the surface.

1.3.4.1 Normal and Albedo from Many Views

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

$$\mathcal{V} = \begin{pmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \\ \cdots \\ \boldsymbol{V}_n^T \end{pmatrix}.$$

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

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

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

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

and $\boldsymbol{g}$ 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.

Substantial regions of the surface may be in shadow for one or the other light (see Figure 1.13). We assume that all shadowed regions are known, and deal only with points that are not in shadow for any illuminant. More sophisticated strategies can infer shadowing because shadowed points are darker than the local geometry predicts.

1.3.4.2 Measuring Albedo

We can extract the albedo from a measurement of $\boldsymbol{g}$ because $\boldsymbol{N}$ 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 1.14 shows albedo recovered using this method for the images of Figure 1.13.

1.3.4.3 Recovering Normals

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

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

Figure 1.15 shows normal values recovered for the images of Figure 1.13.

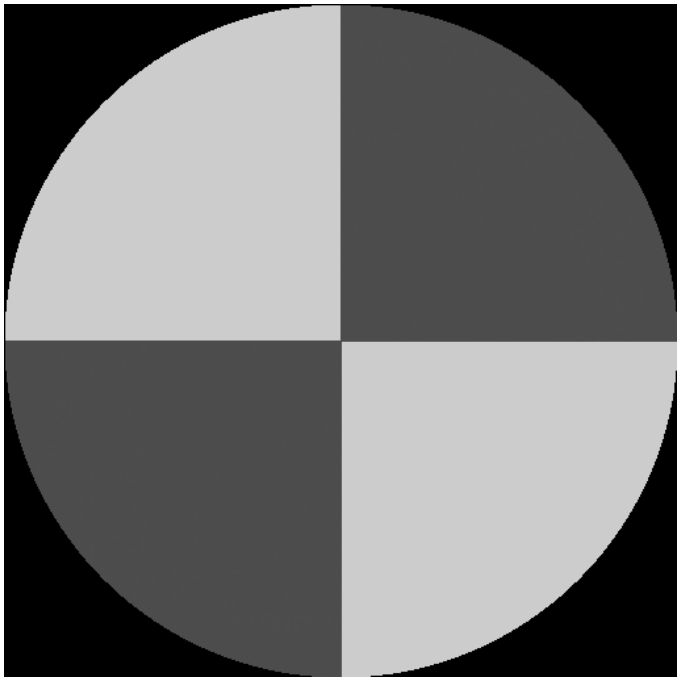


FIGURE 1.14: The magnitude of the vector field $\boldsymbol{g}(x, y)$ recovered from the input data of Figure 1.13 represented as an image — this is the reflectance of the surface.

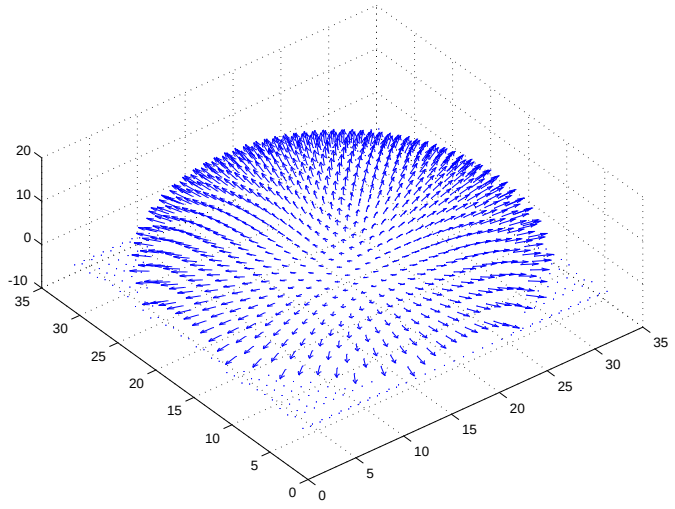


FIGURE 1.15: The normal field recovered from the surface of Figure 1.13.

1.3.4.4 Shape from Normals

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

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

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

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

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.

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 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). 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 1.3.3.2. Figure 1.16 shows the reconstruction obtained for the data of Figure 1.13.

1.3.5 Qualitative Properties of Interreflections

Photometric stereo as described uses the model that light at a surface patch comes only from a distant light source. One can refine the method to take into account nearby light sources, but it is much more difficult to deal with interreflections. Once

Algorithm 1.2: Photometric Stereo

Obtain many images in a fixed view under different illuminants
Determine the matrix $\mathcal{V}$ from source and camera information
Inferring albedo and normal:
For each point in the image array that is not shadowed
  Stack image values into a vector $\boldsymbol{i}$
  Solve $\mathcal{V}\boldsymbol{g} = \boldsymbol{i}$ 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?

Integration:
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

one accounts for interreflections, the brightness of each surface patch could be affected by the configuration of every other surface patch, making a very nasty global inference problem. While there have been attempts to build methods that can infer shape in the presence of interreflections (), the problem is extremely difficult. One source of difficulties is that one may need to account for every radiating surface in the solution, even distant surfaces one cannot see.

An alternative strategy to straightforward physical inference is to understand the qualitative properties of interreflected shading. By doing so, we may be able to identify cases that are easy to handle, the main types of effect, and so on. The effects can be quite large. For example, Figure 1.17 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

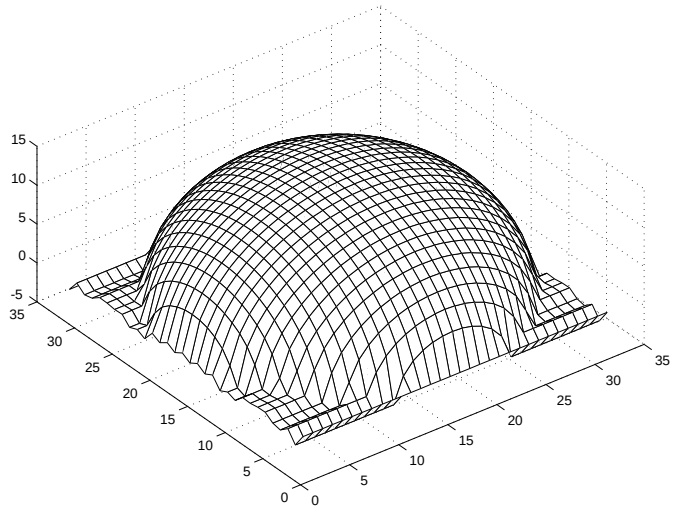


FIGURE 1.16: The height field obtained by integrating the normal field of Figure 1.15 using the method described in the text.

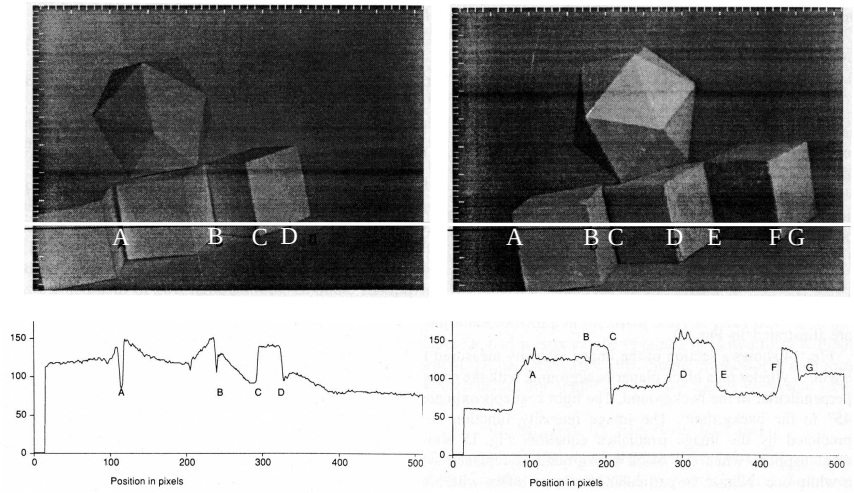


FIGURE 1.17: The column on the **left** shows data from a room with matte black walls and containing a collection of matte black polyhedral objects; that on the **right** shows data from a white room containing white objects. The images are qualitatively different, with darker shadows and crisper boundaries 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*

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.

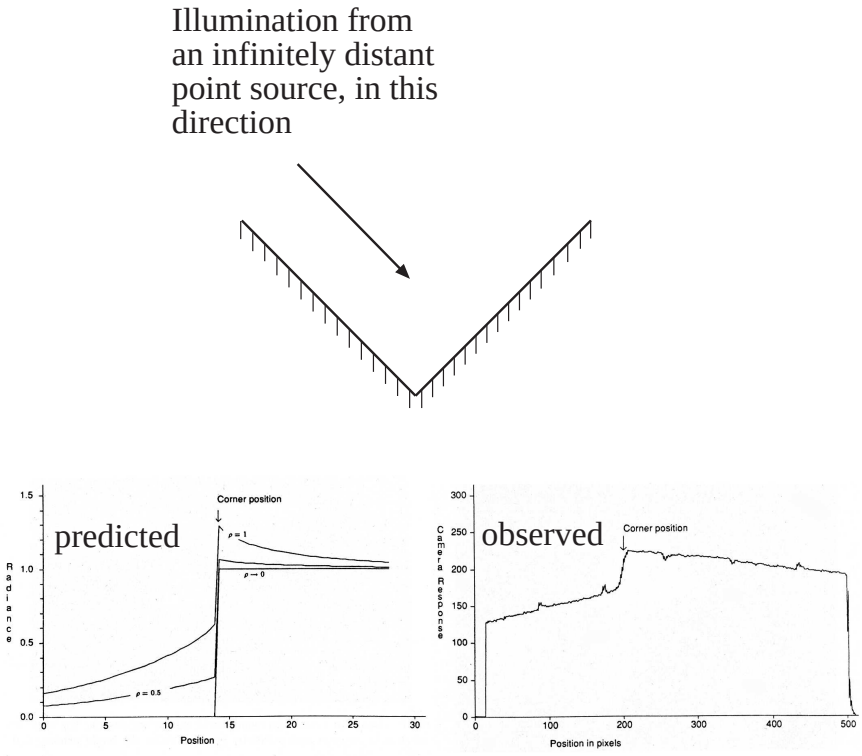


FIGURE 1.18: The model described in the text produces quite accurate qualitative predictions for interreflections. The **top** figure shows a concave right-angled groove illuminated by a point source at infinity where the source direction is parallel to the one face. On the **left** of the bottom row 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*

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 1.19. 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

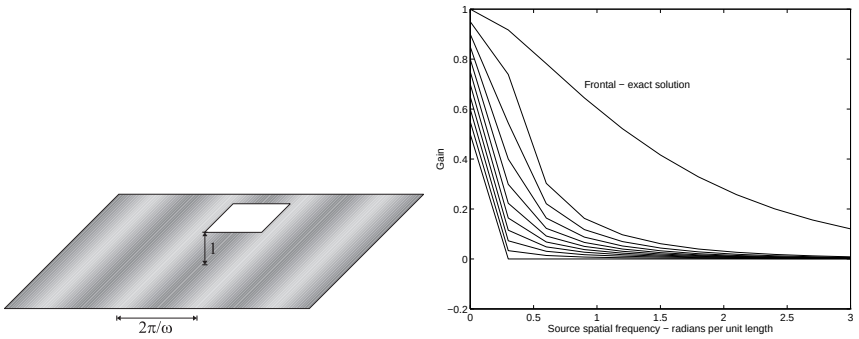


FIGURE 1.19: A small patch views a plane with sinusoidal radiosity of unit amplitude. This patch has a (roughly) sinusoidal radiosity due to the effects of the plane. We refer to the amplitude of this component as the *gain of the patch*. The graph shows numerical estimates of the gain for patches at 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. *Figure from "Shading Primitives: Finding Folds and Shallow Grooves," J. Haddon and D.A. Forsyth, Proc. Int. Conf. Computer Vision, 1998* © *1998 IEEE*

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 2ω. The patch is small enough that its contribution to the plane's radiosity can be ignored. If the patch is slanted by σ 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 1.19. 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 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 1.3.3, 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

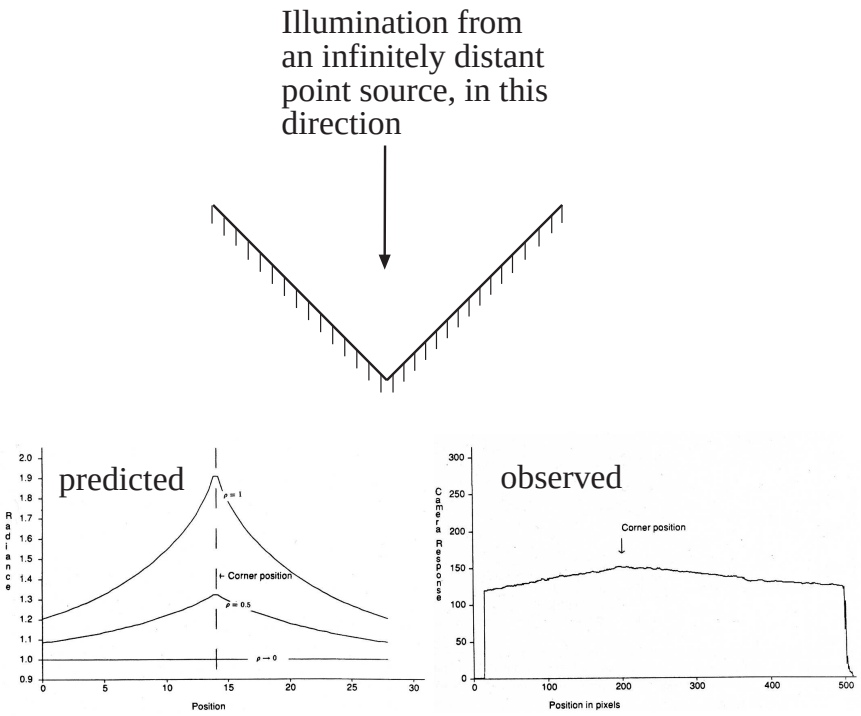


FIGURE 1.20: Reflexes at concave edges are a common qualitative result of interreflections. The figure on the **top** shows the situation here; a concave right-angled groove illuminated by a point light source at infinity, whose source vector is along the angle bisector. The graph on the **left** shows the intensity predictions of an interreflection model for this configuration; the case $\rho \rightarrow 0$ 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*

patches that appear mainly in concave regions (illustrated in Figure 1.20 and Figure 1.21). 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.

1.3.6 Shape from one shaded image

There is quite good evidence that people get some perception of shape from the shading pattern in a single image, though the details are uncertain and quite complicated (see the notes for a brief summary). You can see this evidence in practice: whenever you display a reconstruction of a surface obtained from images, it is a good idea to shade that reconstruction using image pixels, because it always looks more accurate. In fact, quite bad reconstructions can be made to look good like

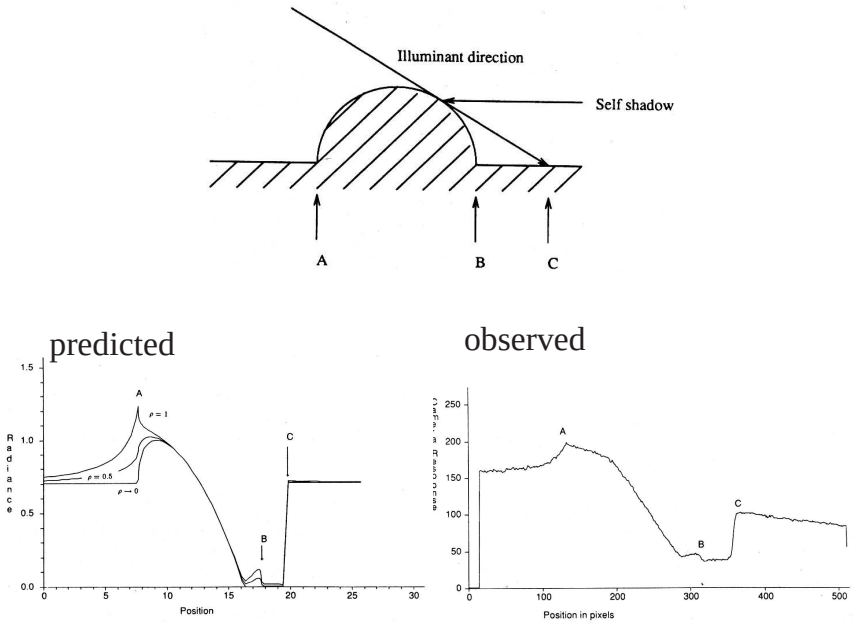


FIGURE 1.21: Reflexes occur quite widely; they are usually caused by a favorable view of a large reflecting surface. In the geometry shown on the **top**, the shadowed region of the cylindrical bump sees the plane background at a fairly favorable angle — if the background is large enough, near half the 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*

this. White *et al.* use this trick to replace surface albedos in movies; for example, they can change the pattern on a t-shirt in a movie (). Their method builds and tracks very coarse geometric reconstructions, and uses a form of regression to recover the original shading pattern of the object, then shade the coarse geometric reconstruction using the original shading pattern (figure 1.22).

The cue to shape must come from the fact that a surface patch that faces the light source is brighter than one that faces away from the source. But going from this observation to a working algorithm remains an open question. The key seems to be an appropriate use of the **image irradiance equation**. Assume we have a surface in the form $(x, y, f(x, y))$ viewed orthographically along the z axis. Assume that the surface albedo is uniform and known. Assume also that the model of section **??** applies, so that the shading at a point with normal $\boldsymbol{N}$ is given by some function $R(\boldsymbol{N})$ (the function of our model is $R(\boldsymbol{N}) = \boldsymbol{N} \cdot \boldsymbol{S}$, but others could be

FIGURE 1.22: On the **left**, an original frame from a movie sequence of a deforming plastic bag. On the **right**, two frames where the original texture has been replaced by another. The method used is a form of regression; its crucial property is that it has a very weak geometric model, but is capable of preserving the original shading field of the image. If you look closely at the *albedo* (i.e. the black pattern) of the bag, you may notice that it is inconsistent with the wrinkles on the bag; but because the shading has been preserved, the figures look quite good. This is indirect evidence that shading is a valuable cue to humans. Little is known about how this cue is to be exploited, however.

used). Now the normal of our surface is a function of the two first partial derivatives

$$p = \frac{\partial f}{\partial x},\ q = \frac{\partial f}{\partial y}$$

so we can write $R(p, q)$. Assume that the camera is radiometrically calibrated, so we can proceed from image values to irradiance values. Write the irradiance at x, y as $I(x, y)$. Then we have

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

This is a first order partial differential equation, because p and q are partial derivatives of f. In principle, we could set up some boundary conditions and solve this equation. Doing so reliably and accurately for general images remains outside our competence, forty years since the problem was originally posed by Horn ().

There are a variety of difficulties here. The physical model is a poor model of what actually happens at surfaces, because any particular patch is illuminated by other surface patches, as well as the source. We expect to see a rich variety of geometric constraints on the surface we reconstruct, and it is quite difficult to formulate shape from shading in a way that accomodates these constraints and still has a solution. Shading is a worthwhile cue to exploit, because we can observe shading at extremely high spatial resolutions, but this means we must work with very high dimensional models to reconstruct. Some schemes for shading reconstruction can be unstable, but there appears to be no theory to guide us to stable schemes. We very seldom actually see isolated surfaces of known albedo, and there are no methods that are competent to infer both shading and albedo, though there is some reason to hope that such methods can be built. We have no theory that is capable of predicting the errors in shading based reconstructions from first principles. All this makes shape inference from shading in a single image one of the most frustrating open questions in computer vision.

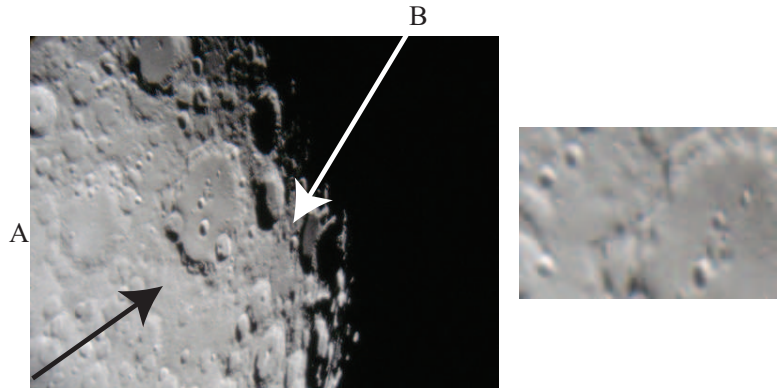


FIGURE 1.23: A picture of the moon, published on Flickr by makelessnoise, shows two important mechanisms by which it might be possible to infer surface shape from single images. First, patches that face the light (like A, on the **left**) are brighter than those that do not (B). The thick black arrow shows the approximate illumination direction. Second, shadows pick out relief — for example, small dents in a surface (the moon's craters, in the detail patch on the **right**), have a bright face facing the light and a dark face which is in shadow.

1.4 NOTES

Index