

CHAPTER 2

Cameras, Light and Shading

2.1 CAMERAS

2.1.1 The Pinhole Camera

A *pinhole camera* is a light-tight box with a very small hole in the front (Figure 2.1). Think about a point on the back of the box. The only light that arrives at that point must come through the hole, because the box is light-tight. If the hole is very small, then the light that arrives at the point comes from only one direction. This means that an inverted image of a scene appears at the back of the box (Figure 2.1). An appropriate sensor (CMOS sensor; CCD sensor; light sensitive film) at the back of the box will capture this image.

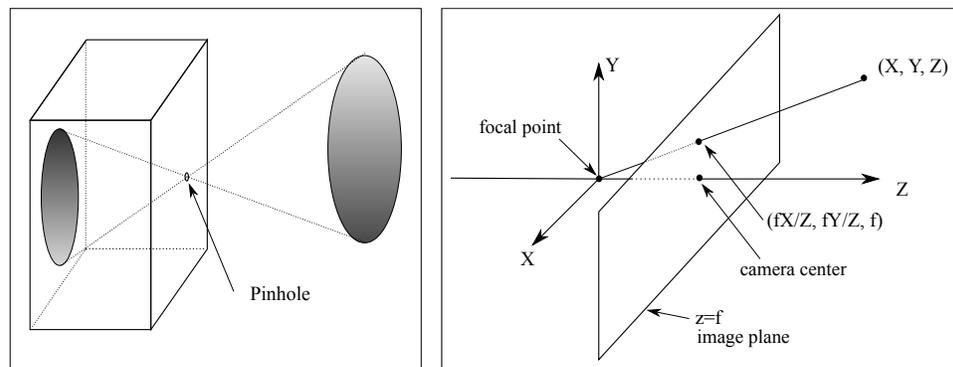


FIGURE 2.1: *The pinhole imaging model. On the left, a light-tight box with a pinhole in it views an object. The only light that a point on the back of the box sees comes through the very small pinhole, so that an inverted image is formed on the back face of the box. On the right, the usual geometric abstraction. The box doesn't affect the geometry, and is omitted. The pinhole has been moved to the back of the box, so that the image is no longer inverted. The image is formed on the plane $z = f$, by convention. Notice the coordinate system is left-handed, because the camera looks down the z -axis. This is because most people's intuition is that z increases as one moves into the image. The text provides some more detail on this point.*

Pinhole camera models produce an upside-down image. This is easily dealt with in practice (turn the image the right way up). An easy way to account for this is to assume the sensor is *in front* of the hole, so that the image is not upside-down. One could not build a camera like this (the sensor blocks light from the hole) but it is a convenient abstraction. There is a standard model of this camera, in a standard coordinate system. The coordinate system is left-handed even though coordinate systems in 3D are usually right-handed coordinate systems. This is because most

people's intuition is that z *increases* as one moves into the image. The pinhole – usually called the *focal point* – is at the origin, and the sensor is on the plane $z = f$. This plane is the *image plane*, and f is the *focal length*. We ignore any camera body and regard the image plane as infinite.

Under this highly abstracted camera model, almost any point in 3D will map to a point in the image plane. We *image* a point in 3D by constructing a ray through the 3D point and the focal point, and intersecting that ray with the image plane. The focal point has an important, distinctive, property: It cannot be imaged, and it is the only point that cannot be imaged.

Similar triangles yields that the point (X, Y, Z) in 3D is imaged to

$$(fX/Z, fY/Z, f)$$

on the sensor (Figure 2.1). Notice that the z -coordinate is the same for each point on the image plane, so it is quite usual to ignore it and use the model

$$(X, Y, Z) \rightarrow (fX/Z, fY/Z).$$

The focal length just scales the image. In standard camera models, other scaling effects occur as well, and we write projection as if $f = 1$, yielding

$$(X, Y, Z) \rightarrow (X/Z, Y/Z).$$

The projection process is known as *perspective projection*. The point where the z -axis intersects the image plane (equivalently, where the ray through the focal point perpendicular to the image plane intersects the image plane) is the *camera center*. Remarkably, in almost every publication in computer vision the camera is expressed in left-handed coordinates and everything else works in right-handed coordinates. The exercises demonstrate that there is no real difficulty here.

Remember this: *Most practical cameras can be modelled as a pinhole camera. The standard model of the pinhole camera maps*

$$(X, Y, Z) \rightarrow (X/Z, Y/Z).$$

Figure 2.1 shows important terminology (focal point; image plane; camera center).

2.1.2 Perspective Effects

Perspective projection has a number of important properties, summarized as:

- lines project to lines;
- more distant objects are smaller;
- lines that are parallel in 3D meet in the image;

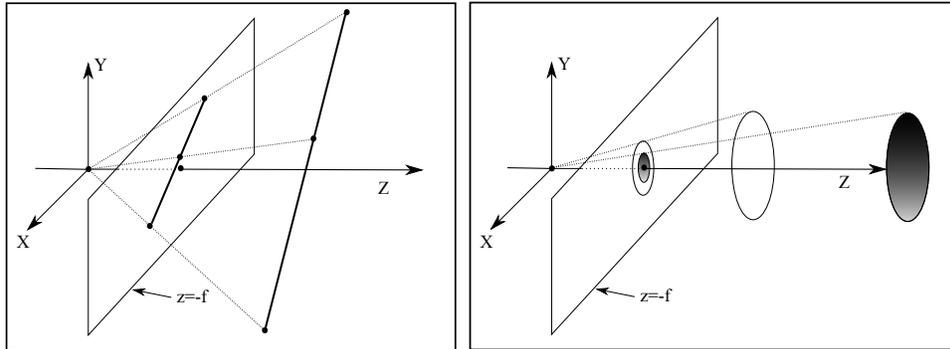


FIGURE 2.2: *Perspective projection maps almost any 3D line to a line in the image plane (left). Some rays from the focal point to points on the line are shown as dotted lines. The family of all such rays is a plane, and that plane must intersect the image plane in a line as long as the 3D line does not pass through the focal point. On the right, two 3D objects viewed in perspective projection; the more distant object appears smaller in the image.*

- planes have horizons;
- planes image as half-planes.

Lines project to lines: Almost every line in 3D maps to a line in the image. You can see this by noticing that the image of the 3D line is formed by intersecting rays from the focal point to each point on the 3D line with the image plane. But these rays form a plane, so we are intersecting a plane with the image plane, and so obtain a line (Figure 2.2). The exceptions are the 3D lines through the focal point – these project to points.

More distant objects are smaller: The further away an object is in 3D, the smaller the image of that object, because of the division by Z (Figure 2.2).

Lines that are parallel in 3D meet in the image: Now think about a set of infinitely long parallel railroad tracks. The sleepers supporting the tracks are all the same size. Distant sleepers are smaller than nearby sleepers, and arbitrarily distant sleepers are arbitrarily small. This means that parallel lines will meet in the image. The point at which the lines in a collection of parallel lines meet is known as the *vanishing point* for those lines (Figure 2.3). The vanishing point for a set of parallel lines can be obtained by intersecting the ray from the focal point and parallel to those lines with the image plane (Figure 2.3).

Planes have horizons: Now think about the image of a plane. As Figure 2.5 shows, the plane through the focal point and parallel to that plane produce a line in the image, known as the *horizon* of the plane.

Planes image as half-planes: For an abstract perspective camera, any point on the plane can be imaged to a point on the image plane. In practical cameras, we cannot image points that lie behind the camera in 3D. Now cast a ray through the focal point and some point \mathbf{x} in the image plane. If \mathbf{x} is on one side of the horizon, the ray will hit the plane in the $z > 0$ half space and so we can see the

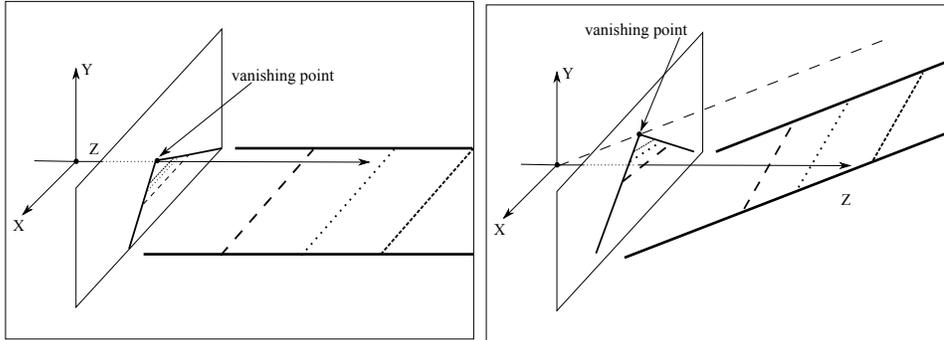


FIGURE 2.3: Perspective projection maps a set of parallel lines to a set of lines that meet in a point. On the **left**, a set of lines parallel to the z -axis, with “railway sleepers” shown. As these sleepers get further away, they get smaller in the image, meaning the projected lines must meet. The vanishing point (the point where they meet) is obtained by intersecting the ray parallel to the lines and through the focal point with the image plane. On the **right**, a different pair of parallel lines with a different vanishing point. The figure establishes that, if there are more than two lines in the set of parallel lines, all will meet at the vanishing point.

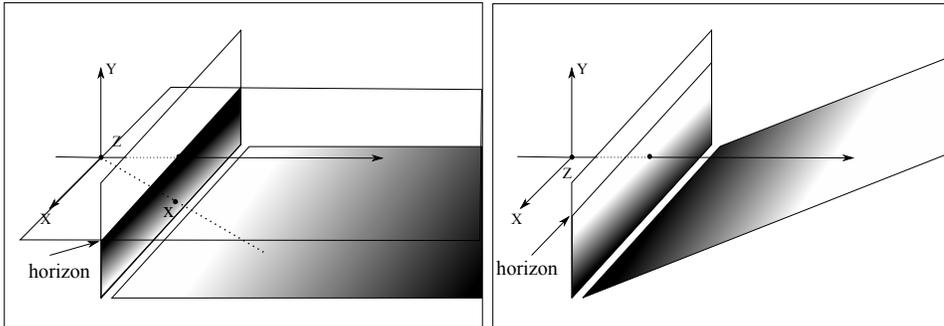


FIGURE 2.4: **Left** shows a plane in 3D (in this case, $y = -1$). The intersection of the plane through the focal point parallel to the 3D plane (in this case, $y = 0$) and the image plane, forms an image line called the horizon. This line cuts the image plane into two parts. Construct the ray through the focal point and a point \mathbf{x} in the image plane. For \mathbf{x} on one side of the horizon, this ray will intersect the 3D plane in the half space $z > 0$ (and so in front of the camera, shown here). If \mathbf{x} is on the other side of the horizon, the intersection will be in the half space $z < 0$ (and so behind the camera, where it cannot be seen). **Right** shows a different 3D plane with a different horizon. The gradients on the planes indicate roughly where points on the 3D plane appear in the image plane (light points map to light, dark to dark).

plane. If it is on the other side, it will hit the plane in the $z < 0$ half space, so we cannot see the plane.

Remember this: *Under perspective projection:*

- *points project to points;*
- *lines project to lines;*
- *more distant objects are smaller;*
- *lines that are parallel in 3D meet in the image;*
- *planes have horizons;*
- *planes image as half-planes.*

2.1.3 Scaled Orthographic Projection and Orthographic Projection

Under some circumstances, perspective projection can be simplified. Assume the camera views a set of points which are close to one another compared with the distance to the camera. Write $\mathbf{X}_i = (X_i, Y_i, Z_i)$ for the i 'th point, and assume that $Z_i = Z(1 + \epsilon_i)$, where ϵ_i is quite small. In this case, the distance to the set of points is much larger than the *relief* of the points, which is the distance from nearest to furthest point. The i 'th point projects to $(fX_i/Z_i, fY_i/Z_i)$, which is approximately $(f(X_i/Z)(1 - \epsilon_i), f(Y_i/Z)(1 - \epsilon_i))$. Ignoring ϵ_i because it is small, we have the projection model

$$(X, Y, Z) \rightarrow (f/Z)(X, Y) = s(X, Y).$$

This model is usually known as *scaled orthographic projection*. The model applies quite often. One important example is pictures of people. Very often, all body parts are roughly the same distance from the camera — think of a side view of a pedestrian seen from a motor car. Scaled orthographic projection applies in such cases. It is not always an appropriate model. For example, when a person is holding up a hand to block the camera's view, perspective effects can be significant (Figure ??).

Occasionally, it is useful to rescale the camera (or assume that $f/Z = 1$), yielding $(X, Y, Z) \rightarrow (X, Y)$. This is known as *orthographic projection*.

Remember this: *Scaled orthographic projection maps*

$$(X, Y, Z) \rightarrow s(X, Y)$$

where s is some scale. The model applies when the distance to the points being viewed is much greater than their relief. Many views of people have this property.

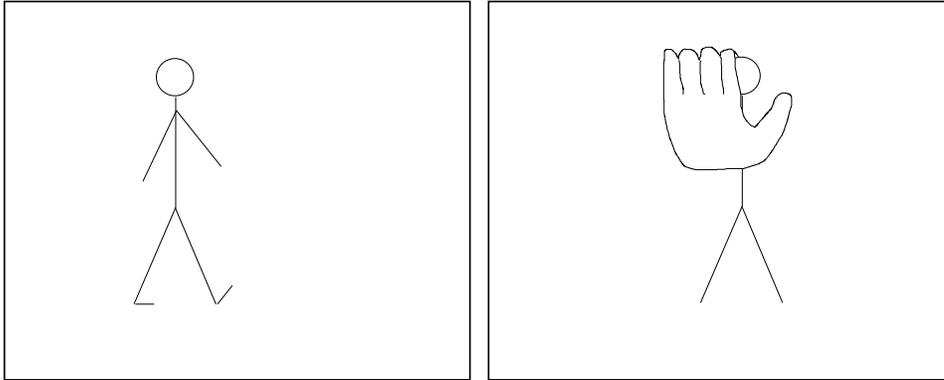


FIGURE 2.5: The pedestrian on the **left** is viewed from some way away, so the distance to the pedestrian is much larger than the change in depth over the pedestrian. In this case, which is quite common for views of people, scaled orthography will apply. The celebrity on the **right** is holding a hand up to prevent the camera viewing their face; the hand is quite close to the camera, and the body is an arm's length away. In this case, perspective effects are strong. The hand looks big because it is close, and the head looks small because it is far.

2.1.4 Lenses

One practical version of a pinhole camera is a *camera obscura* – the box is built as a room, and you can stand in the room and see the view on the back wall (some examples are at <https://www.atlasobscura.com/lists/camera-obscura-places>; the internet yields amusing disputes about the correct plural form of the term). You can also build a simple pinhole camera with a matchbox, some tape, a pin, and some light sensitive film do the trick. Getting good images takes trouble, however.

A large hole in front of the camera will cause the image at the back to be brighter, but blurrier, because each point on the sensor will average light over all directions that happen to go through the hole. If the hole is smaller, the image will get sharper, but darker. In practical cameras, achieving an image that is both bright and focused is the job of the lens system. There may be one or several lenses that light passes through before reaching the sensor at the back of the camera. Each of these lenses is built from refracting materials. The shape and position of the lenses, together with the refractive index of the materials they are built of, determine the path that light follows through the lens system. Generally, the lens system is designed to collect as much light as possible at the input and produce a focused image on the image plane. Remarkably, the many or most lens systems result in an imaging geometry that can be modelled with a pinhole camera model, and lens system effects are ignored in all but quite specialized applications of computer vision.

Lens systems are designed and modelled using geometric optics, but lens designs always involve compromises. The result is that cameras with lenses differ from pinhole cameras in some ways that are worth knowing about, although they are not always important. First, in an abstract pinhole camera, all objects at what-

ever distance are in focus. Geometric optics means that a lens with this property admits very little light, so it is common to work with cameras that have a limited *depth of field* – the range of distances to the camera over which objects are in focus on the image plane. Second, manufacturing difficulties and cost considerations mean that lenses will have various *aberrations*. The net effect of most aberrations is a tendency to defocus some objects under some circumstances, but *chromatic aberrations* can cause colors to be less crisp and objects to have “halos” of color. Chromatic aberration occurs because light of different wavelengths takes slightly different paths through a refracting object. Various lens coatings can correct chromatic aberration, but the resulting lens system will be more expensive. Third, in most lens systems, the periphery of the image tends to be brighter than it would be in a pure pinhole camera. For more complex lens systems, an effect in the lens known as *vignetting* can darken the periphery somewhat. Finally, lenses may cause *geometric distortions* of the image. The most noticeable effect of these distortions is that straight lines in the world may project to curves in the image. Most common is *barrel distortion*, where a square is imaged as a bulging barrel; *pincushion distortion*, where the square bulges in rather than out, can occur (Figure ??).

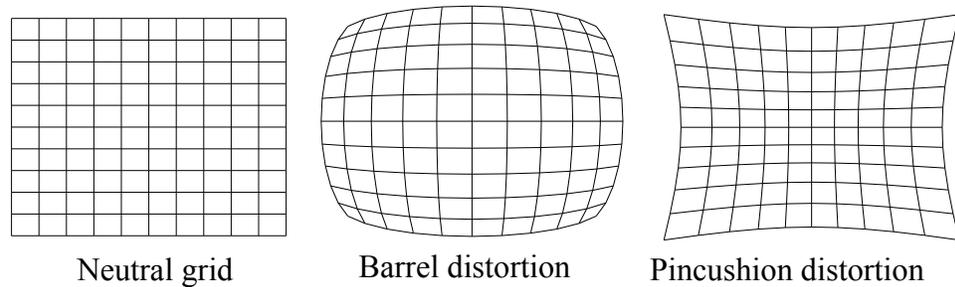


FIGURE 2.6: On the **left** a neutral grid observed in a non-distorting lens (and viewed frontally to prevent any perspective distortion). **Center** shows the same grid, viewed in a lens that produces barrel distortion. **Right**, the same grid, now viewed in a lens that produces pincushion distortion.

2.2 DEPTH CAMERAS

TODO: depth cams

2.3 LIDAR

TODO: LIDAR

2.4 LIGHT AND SURFACES

Three major phenomena determine the brightness of a pixel: the response of the camera to light, the fraction of light reflected from the surface to the camera, and the amount of light falling on the surface. Each can be dealt with quite straightforwardly.

Camera response: Modern camera sensors respond linearly to light. This linear response is adjusted in software, because humans find linear images confusing (such images tend to be too dark in most places, and too light in others). The *camera response function* or *CRF* determines what value is reported at each location. Typical CRF's are close to linear in mid-ranges, but have pronounced nonlinearities for darker and brighter illumination. This allows the camera to reproduce the very wide dynamic range of natural light without saturating.

Write \mathbf{X} for a point in space that projects to \mathbf{x} in the image, $I_{patch}(\mathbf{X})$ for the intensity of the surface patch at \mathbf{X} , $C(\cdot)$ for the camera response function, and $I_{camera}(\mathbf{x})$ for the camera response at \mathbf{x} . Then our model is:

$$I_{camera}(\mathbf{x}) = C(I_{patch}(\mathbf{x})).$$

It is quite usual to assume that the camera response is linearly related to the intensity of the surface patch. In this case, $C(I_{patch}(\mathbf{x})) = kI_{patch}(\mathbf{x})$, and it is common to assume that k is known if needed. A CRF can be recovered from enough image data, if required (Section 4.1.1).

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. There is a rich set of possible physical effects, but most can be ignored. Section 2.4.1 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 could change because some *luminaires* (the formal term for light sources) might be shadowed, or might 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 2.4.2 describes the most important model used in computer vision; Section 2.4.4 describes a much more complex model that is necessary to explain some important practical difficulties in shading inference.

2.4.1 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 ?, who argue the dynamic range is closer to 10 than the 18 implied by these numbers). 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

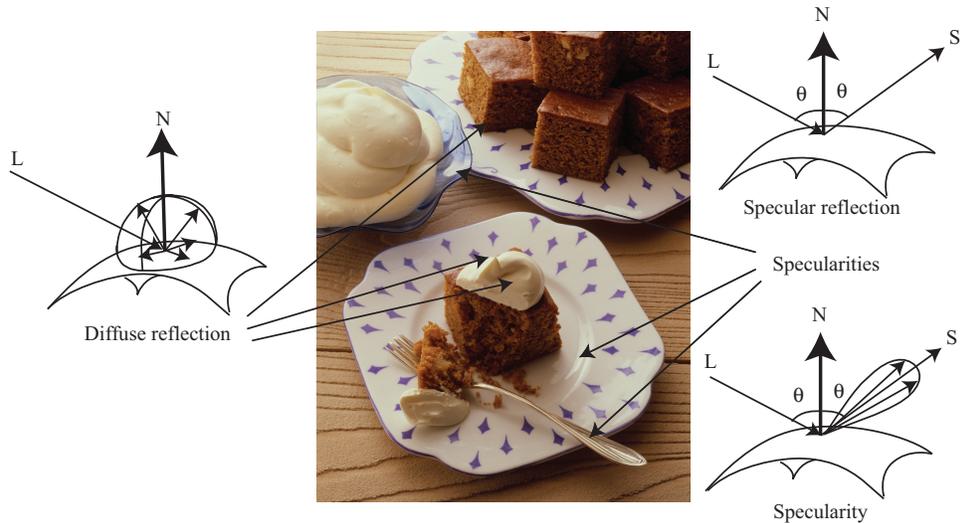


FIGURE 2.7: The two most important reflection modes for computer vision are diffuse reflection (**left**), where incident light is spread evenly over the whole hemisphere of outgoing directions, and specular reflection (**right**), where reflected light is concentrated in a single direction. The specular direction \mathbf{S} is coplanar with the normal and the source direction (\mathbf{L}), and has the same angle to the normal that the source direction does. Most surfaces display both diffuse and specular reflection components. In most cases, the specular component is not precisely mirror like, but is concentrated around a range of directions close to the specular direction (**lower right**). This causes specularities, where one sees a mirror like reflection of the light source. Specularities, when they occur, tend to be small and bright. In the photograph, they appear on the metal spoon and on the plate. Large specularities can appear on flat metal surfaces (arrows). Most curved surfaces (such as the plate) show smaller specularities. Most of the reflection here is diffuse; some cases are indicated by arrows. Martin Brigdale © Dorling Kindersley, used with permission.

incoming radiation about the surface normal (Figure 2.7). 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 based on this test, relatively few surfaces actually behave like ideal specular reflectors. Imagine a near perfect mirror made of polished metal; if this surface 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 2.7; ?). Most surfaces reflect only some of the incoming light in a specular component, and we can represent the percentage of light that is specularly reflected with a *specular albedo*. Although the diffuse albedo is an important material property that we will try to estimate from images, the specular albedo is largely seen as a nuisance and usually is not estimated.

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, and so there is no need for 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 []. Thus, to apply the lambertian+specular model, we find and remove specularities, and then use Lambert’s law (Section 2.4.2) to model image intensity.

2.4.2 Sources and Their Effects

The main source of illumination outdoors 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 (because it is like the sun and because it is easy to use), and can be 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 in which the rays travel. The 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 2.8). The figure yields *Lambert’s cosine law*, which states the brightness of a diffuse 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 that 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 a surface provides some shape information. We explore this cue in Section ??.

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. This means that, for example, we tend to see few shadows in rooms with white walls, because any shadowed patch receives a lot of light from the walls. These effects are sometimes modelled by adding a constant *ambient illumination* term to the predicted intensity. The ambient term

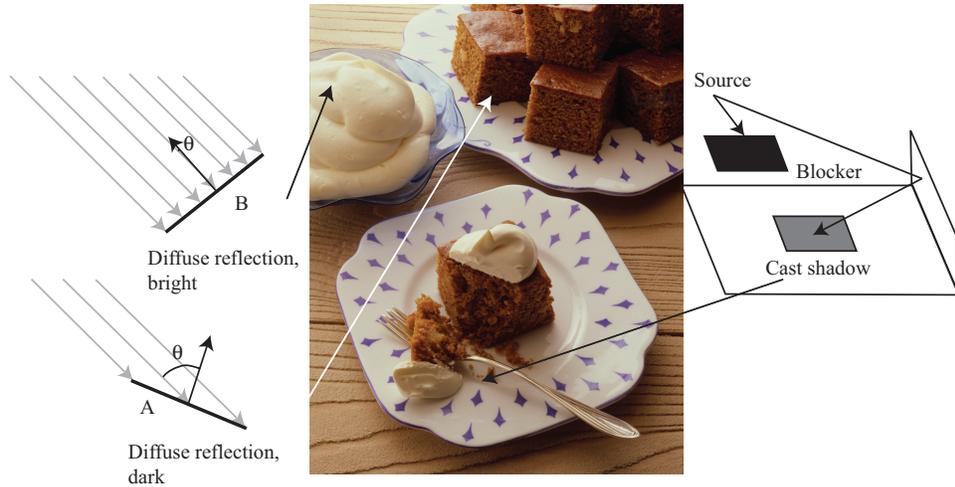


FIGURE 2.8: The orientation of a surface patch with respect to the light affects how much light the patch gathers. We model surface patches as illuminated by a distant point source, whose rays are shown as light arrowheads. Patch A is tilted away from the source (θ is close to 90°) and collects less energy, because it cuts fewer light rays per unit surface area. Patch B, facing the source (θ is close to 0°), collects more energy, and so is brighter. Shadows occur when a patch cannot see a source. The shadows are not dead black, because the surface can see interreflected light from other surfaces. These effects are shown in the photograph. The darker surfaces are turned away from the illumination direction. Martin Brigdale © Dorling Kindersley, used with permission.

ensures that shadows are not too dark, but this is not a particularly good model of the spatial properties of interreflections. We have sketched the effects to be aware of in Section 2.4.4.

2.4.3 The Local Shading Model for Distant Luminaires

Surfaces reflect light onto one another (*interreflections*), meaning that the light arriving at a surface could have come directly from a luminaire, but it could also have been reflected from some other surface. Really accurate physical models of how light is distributed on scenes are now very well known [?] and are extremely useful in computer graphics. These models are very hard to use for inference, because every variable affects every other variable. For example, changes in the orientation of one surface element affect how much light it reflects onto every other surface element.

This means we must simplify the model, and so we must be using a model that isn't exact, meaning we need to keep track of what that model will do well and what it will do badly. The usual simplification is a *local shading model*, where we assume that shading is caused only by light that comes from the luminaire (i.e., that there are no interreflections).

Now assume that the luminaire is an infinitely distant source. For this case, write $\mathbf{N}(x)$ for the unit surface normal at \mathbf{x} , \mathbf{S} for a vector pointing from \mathbf{x} toward the source with length I_o (the source intensity), $\rho(\mathbf{x})$ for the albedo at \mathbf{x} , and $Vis(\mathbf{S}, \mathbf{x})$ for a function that is 1 when \mathbf{x} can see the source and zero otherwise. Then, the intensity at \mathbf{x} is

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

Image	=	Diffuse	+	Ambient	+	Specular (mirror-like)
intensity		term		term		term

2.4.4 Shading Effects from Area Sources

The local shading model is a good rough and ready model, but it isn't right. It predicts dark shadows with sharp boundaries. These are quite common outdoors where the sun is the most important light source, but are uncommon indoors. To understand why, we must look at area sources.

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. Area sources are common in illumination engineering, because they tend not to cast strong shadows and because the illumination due to the source does not fall off significantly as a function of the distance to the source. Detailed models of area sources are complex, but a simple model is useful to understand shadows. Shadows from area sources are very different from shadows cast by point sources. 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 2.9). These effects occur indoors 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 2.9). The penumbra can be large, and can change quite slowly from light to dark. There might 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 2.10).

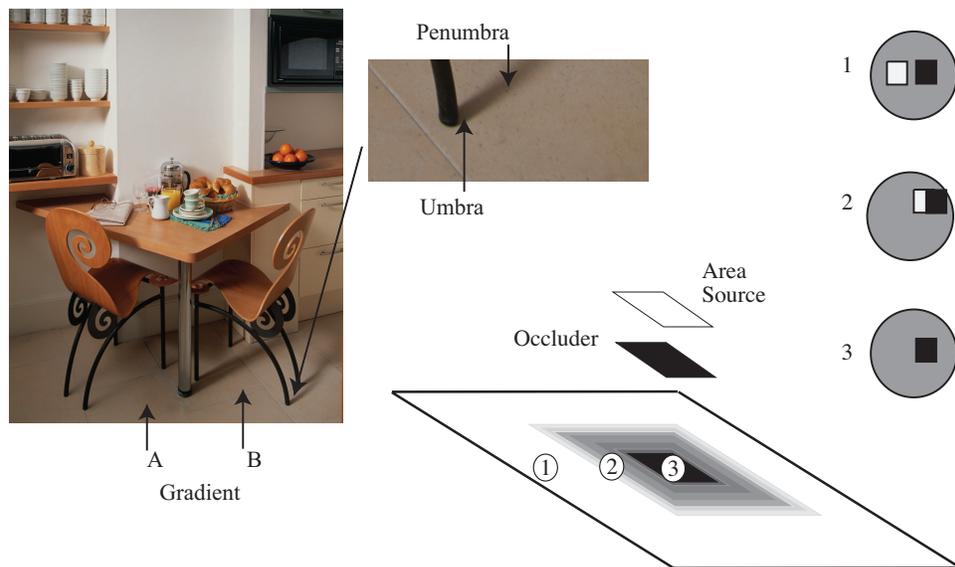


FIGURE 2.9: Area sources generate complex shadows with smooth boundaries, because from the point of view of a surface patch, the source disappears slowly behind the occluder. **Left:** a photograph, showing characteristic area source shadow effects. Notice that A is much darker than B; there must be some shadowing effect here, but there is no clear shadow boundary. Instead, there is a fairly smooth gradient. The chair leg casts a complex shadow, with two distinct regions. There is a core of darkness (the umbra—where the source cannot be seen at all) surrounded by a partial shadow (penumbra—where the source can be seen partially). A good model of the geometry, illustrated **right**, 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. Peter Anderson © Dorling Kindersley, used with permission.



FIGURE 2.10: The photograph on the **left** shows a room interior. Notice the lighting has some directional component (the vertical face indicated by the arrow is dark, because it does not face the main direction of lighting), but there are few visible shadows (for example, the chairs do not cast a shadow on the floor). 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 2.9). Jake Fitzjones © Dorling Kindersley, used with permission.