Computer Vision Handout 2

Chapter 2

Image Formation

In Chapter 1, we discussed some salient aspects of the human visual system—in particular, of the human eye. As illustrated in Figure 2.1, the human eye forms an inverted image of the scene on its retina. The retina, in turn, as we saw in Chapter 1, senses the image, encodes it, and then transmits the encoded image to the brain. The role of a camera in a

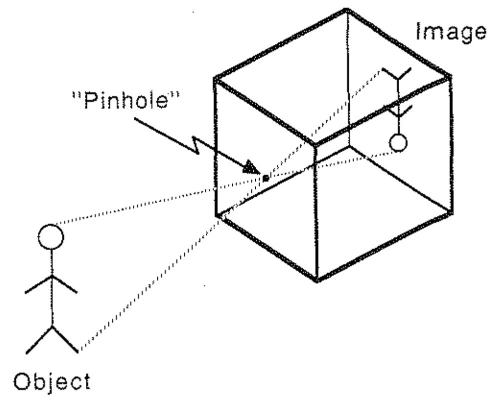


Figure 2.2 Pinhole camera. A pinhole camera is the simplest device to form an image of a three-dimensional scene on a two-dimensional surface: Straight rays of light pass through a "pinhole" and form an inverted image on a surface.

computer-vision system is analogous to the role of the eye in the human visual system. Let us now explore the relationship between the three-dimensional world and its two-dimensional image produced by a camera. This relationship is at the heart of every nonheuristic attempt to recover the properties of a scene from its one or multiple images. Hence, it is important that we understand this relationship well.

Our discussion here will proceed along three lines: geometry, radiometry, and sensing. First, we shall study the geometry of image formation. Then, we shall examine the relationship between the amount of light radiating from a surface and the amount of light impinging on the image of the surface. Finally, we shall turn our attention to the sensing of the image—that is, to the conversion of the image into a representation that is amenable to storage, processing, and analysis by an electronic computer. A word of caution: All the models presented in this chapter are just first-order approximations that will certainly be in need of refinement as computer vision advances.

2.1 Geometry

The simplest imaging device is a **pinhole camera** of the type illustrated in Figure 2.2. Ideally, a pinhole camera has an infinitesimally small aperture—a "pinhole"—through which light enters the camera and forms an image on the camera surface facing the aperture. Geometrically, the image is formed by straight rays of light that travel from the object through the aperture to

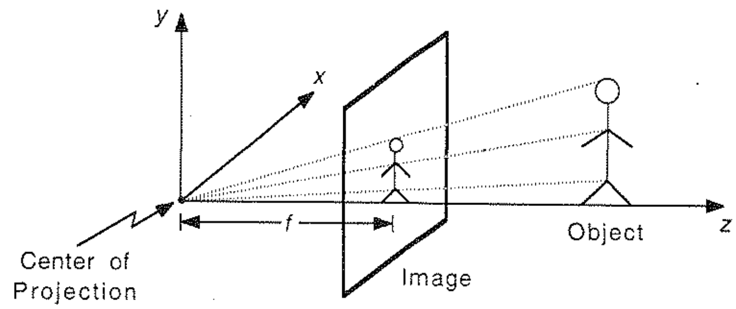


Figure 2.3 Perspective projection. In *perspective projection*, each object point is projected onto a surface along a straight line through a fixed point called the *center of projection*. (Throughout this book, we shall use the terms *object* and *scene* interchangeably.) The projection surface here is a plane. Perspective projection closely models the geometry of image formation in a pinhole camera (see Figure 2.2), except, in perspective projection, we are free to choose the location of the projection surface such that the image is not inverted.

the image plane; here, as elsewhere, we use the terms *object* and *scene* interchangeably. Such a mapping from three dimensions onto two dimensions is called *perspective projection*. Let us first examine perspective projection, and then let us examine two linear approximations to perspective projection. Subsequently, we shall consider the role of a lens in image formation.

2.1.1 Perspective Projection

Perspective projection, also known as **central projection**, is the projection of a three-dimensional entity onto a two-dimensional surface by straight lines that pass through a single point, called the **center of projection**. Perspective projection closely models the geometry of image formation in a pinhole camera.

As illustrated in Figure 2.2, the image in a pinhole camera is inverted. As far as analysis goes, this inversion of the image is mildly inconvenient. Hence, it is customary instead to consider the geometrically equivalent configuration of Figure 2.3 in which the image is on the same side of the center of projection as the scene, and, as a result, the image is not inverted. Now, if we denote the distance of the image plane from the center of projection by f, and we denote the object coordinates and the image coordinates by the subscripts o and i, respectively, then it is clear from similar triangles that

$$x_i = \frac{f}{z_o}\, x_o \qquad \text{and} \qquad y_i = \frac{f}{z_o}\, y_o\,.$$

These equations are the fundamental equations for perspective projection onto a plane.

Homogeneous Coordinates

The preceding equations for perspective projection onto a plane are nonlinear. It is often convenient to linearize these equations by mapping each point (x, y, z) in three-dimensional space onto the following line that passes through the origin in four-dimensional space: $(X, Y, Z, W) = (wx, wy, wz, w)$, where w is a dummy parameter that sweeps a straight line in four-dimensional space. The new coordinates, (X, Y, Z, W), are called **homogeneous coordinates**. (Historically, the use of homogeneous coordinates in computer vision goes back to Roberts in 1965 [Roberts 1965].) Although the homogeneous coordinates of a point in three-dimensional space are not unique as every point in three-dimensional space is represented by a whole family of points in four-dimensional space, the homogeneous coordinates are unambiguous if we exclude the origin in four-dimensional space. That is, barring the origin, every point in four-dimensional space represents no more than a single point in three-dimensional space. Specifically, (X, Y, Z, W), $W \neq 0$, in four-dimensional space represents the single point $(x, y, z) = (X/W, Y/W, Z/W)$ in three-dimensional space. Despite their redundancy, homogeneous coordinates are extremely useful as they allow us to express several otherwise nonlinear transformations linearly; see, for instance, [Ballard and Brown 1982] and [Wolberg 1990]. In homogeneous coordinates, perspective projection onto a plane may be expressed as follows:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ W_i \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_o \\ Y_o \\ Z_o \\ W_o \end{bmatrix},$$

where, once again, the subscripts o and i denote the object coordinates and the image coordinates, respectively. You can easily verify the equivalence of this linear expression to the preceding nonlinear equations for perspective projection by making the following two substitutions: $(X_i, Y_i, Z_i, W_i) = (\alpha x_i, \alpha y_i, \alpha z_i, \alpha)$ and $(X_o, Y_o, Z_o, W_o) = (\beta x_o, \beta y_o, \beta z_o, \beta)$. Although the exact definition of a linear transform must await Section 2.3.2, it is sufficient to note here that a transform is linear if and only if it can be expressed as a matrix multiplication, as in the preceding expression for perspective projection in homogeneous coordinates.

Figure 2.4 Photograph illustrating a vanishing point. Parallel straight lines converge at a single point under perspective projection. This point is called the *vanishing point* of the straight lines. (Photograph by Herbert Gehr, from the magazine *Life*, July 1947, © Time Warner, Inc.)

Vanishing Point and Vanishing Line

An important concept in the context of perspective projection is that of a vanishing point. The **vanishing point** of a straight line under perspective projection is that point in the image beyond which the projection of the straight line cannot extend. That is, if the straight line were infinitely long in space, the line would appear to "vanish" at its vanishing point in the image. As the vanishing point of a straight line depends only on the orientation of the line, and not on the position of the line, the notion of a vanishing point is frequently explained in the context of parallel lines. Barring the degenerate case where parallel lines are all parallel to the image plane, parallel straight lines in space project perspectively onto straight lines that on extension intersect at a single point in the image plane. The common intersection of the straight lines in the image, which is the vanishing point, corresponds to "points at infinity" in the receding direction of the parallel straight lines in space. The photograph in Figure 2.4 illustrates a vanishing point beautifully.

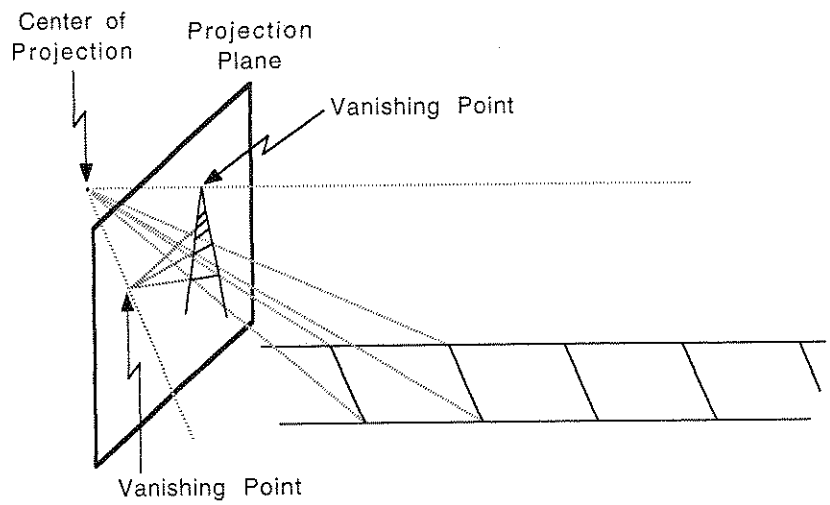


Figure 2.5 The vanishing point. The *vanishing point* of a straight line under perspective projection is that point on the projection surface at which the line would appear to "vanish" if the line were infinitely long in space. The location of the vanishing point of a straight line depends only on the orientation of the straight line in space, and not on the line's position: For any given spatial orientation, the vanishing point is located at that point on the projection surface where a straight line passing through the center of projection with the given orientation would intersect the projection surface.

As illustrated in Figure 2.5, the vanishing point of any given straight line in space is located at that point in the image where a parallel line through the center of projection intersects the image plane. It follows easily that the vanishing point of every straight line that is confined to some plane—actually, for every straight line that is parallel to this plane—lies somewhere along a particular straight line in the image plane. This line, called the **vanishing line** of the plane, is located where a parallel plane through the center of projection intersects the image plane.

Planar Versus Spherical Perspective Projection

Although it is to the geometry of Figure 2.3 that most people refer when they speak of perspective projection, this geometry is not always the most convenient to analyze. In Figure 2.3, the image is formed at the intersection of a cone of projection rays with a plane. Let us call projection along a cone of projection rays onto a plane **planar perspective projection**. The image in planar perspective projection depends on more than just the position of the center of projection: It also depends on the orientation and position of the

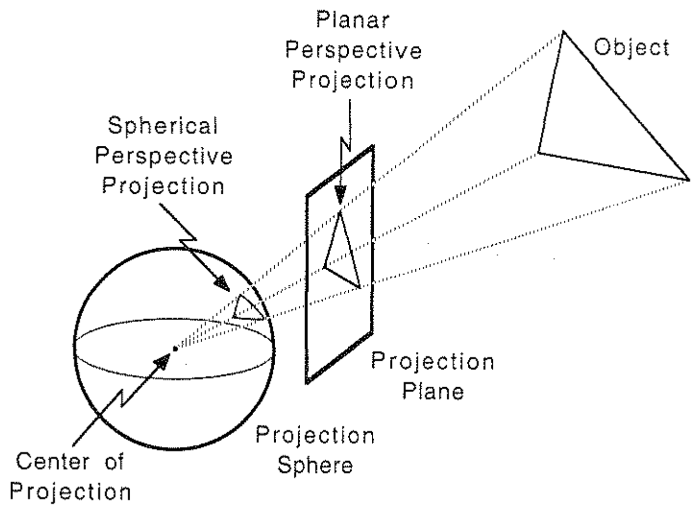


Figure 2.6 Planar and spherical perspective projection. In perspective projection, each object point is projected along a straight line through the center of projection: In *planar perspective projection*, the projection surface is a plane, and in *spherical perspective projection*, the projection surface is a sphere that is centered at the center of projection.

imaging surface, which is a plane. We can remove such a dependence by instead projecting the scene onto a unit sphere that is centered at the center of projection, as illustrated in Figure 2.6; the unit sphere in this context serves as a convenient device to represent the two-dimensional manifold of projection rays.[1] Let us call projection along a cone of projection rays onto a

1. It is noteworthy that the earliest extant geometrical investigation of vision, the treatise *Optics* by Euclid [Euclid c. 300 B.C.], confined itself to the study of the relative orientations of projection rays. Although the premise of the *Optics* that object points are visible when rays emitted by the eye are incident on them is now known to be false, the geometrical analysis therein remains accurate, and is, in fact, remarkable. For instance, the proof of the following proposition of the *Optics* contains the seeds of the notion of a vanishing point: "Parallel lines, when seen from a distance, appear not to be equally distant from each other" (p. 358, [Euclid c. 300 B.C.], English translation). We shall have further occasion to appreciate the present-day relevance of the *Optics* to computer vision.

Euclid, the author of the *Optics*, is best known for his *Elements*, a treatise without equal in the history of mathematics. The *Elements* lay the foundations of axiomatic geometry, and has been used as a text virtually unchanged for over 2000 years. Despite the preeminence of the *Elements*, all that is known with certainty of Euclid's life is that he founded a school at Alexandria circa 300 B.C. and taught mathematics there. An anecdote relates that, on being asked by a student what he would gain by learning geometry, Euclid called his slave and said, "Give him three obols [ancient Greek coins] since he must needs make gain by what he learns" (see [Gillispie 1971]).

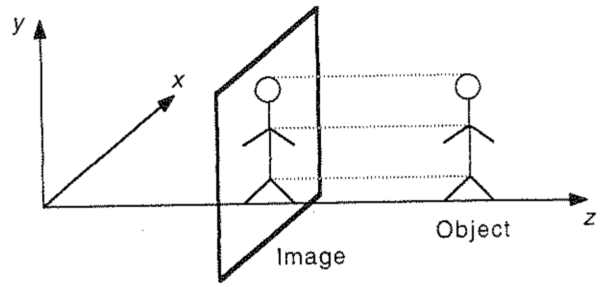


Figure 2.7 Orthographic projection. In *orthographic projection*, each object point is projected onto a plane along a straight line orthogonal to the plane.

sphere, which is not necessarily of unit radius but is centered at the center of projection, **spherical perspective projection**. Under spherical perspective projection, straight lines map onto great-circle arcs—that is, they map onto arcs of circles that are centered at the center of the sphere. Once again, the vanishing point of a straight line is located at that point in the image where a parallel line through the center of projection intersects the imaging surface; hence, we now get two vanishing points for *every* orientation of a straight line. It is not difficult to see that an image formed under planar perspective projection along with its center of projection defines the corresponding image under spherical perspective projection. Of course, we can go from spherical perspective projection to planar perspective projection too.

2.1.2 Orthographic Projection

Orthographic projection, as illustrated in Figure 2.7, is the projection of a three-dimensional entity onto a plane by a set of parallel rays orthogonal to this plane. In the figure, we have $x_i = x_o$ and $y_i = y_o$, where, once again, the subscripts o and i denote the object coordinates and the image coordinates, respectively. Under conditions that we shall examine in this section, orthographic projection closely approximates perspective projection up to a uniform scale factor. When valid, such an approximation is extremely convenient as orthographic projection, unlike perspective projection, is a linear transformation in three-dimensional space.

Consider the perspective-projection geometry of Figure 2.3 once again. As the object is moved away from the center of projection along the z-axis, the image size clearly decreases. More important, the magnification factor (f/z_o) in the perspective-projection equations becomes less sensitive to z_o. That is, $f/(z_o + \Delta z_o)$ tends to be more closely approximated by (f/z_o) as

$\frac{f}{z_0+\Delta z_0} = \frac{f/z_0}{1+\Delta z_0/z_0} \approx f/z_0$ if $\frac{\Delta z_0}{z_0} \to 0$ or $\frac{z_0}{\Delta z_0} \to \infty$

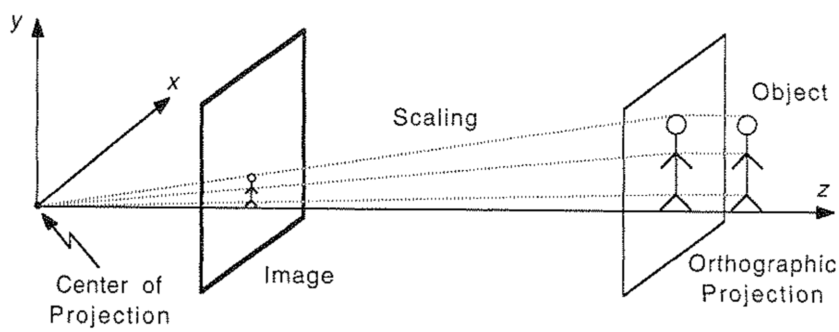


Figure 2.8 Approximation of perspective projection by orthographic projection. Perspective projection onto a plane can be approximated by orthographic projection, followed by scaling, when (1) the object dimensions are small compared to the distance of the object from the center of projection, and (2) compared to this distance, the object is close to the straight line that passes through the center of projection and is orthogonal to the image plane (this line is the z-axis here).

$(z_0/\Delta z_0)$ tends to become large. This increasingly close approximation of $f/(z_0 + \Delta z_0)$ by (f/z_0) might lead you to believe that, whenever the average depth of an object is large compared to the object's range of depths, perspective projection can be approximated by orthographic projection up to the scale factor (f/z_0). Such a hypothesis, however, is incorrect.

Two conditions are necessary and sufficient for perspective projection to be approximated closely by orthographic projection up to a uniform scale factor:

1. The object must lie close to the optical axis; in consistency with the terminology for the imaging geometry of a lens in Section 2.1.4, the **optical axis** here is defined as the line through the center of projection that is orthogonal to the image plane.
2. The object's dimensions must be small.

Both *close* and *small* here are with respect to the distance of the object from the center of projection. Figure 2.8 graphically illustrates the approximation of perspective projection as a two-step process: orthographic projection onto a nearby plane parallel to the image plane, and then, perspective projection onto the image plane. The latter is equivalent to uniform scaling. It is not difficult to see that the two projections in tandem approximate direct perspective projection closely only when both the conditions specified here are satisfied. To verify this assertion, you need simply to consider projections of various wire-frame cuboids, each cuboid with one face parallel to the image plane.

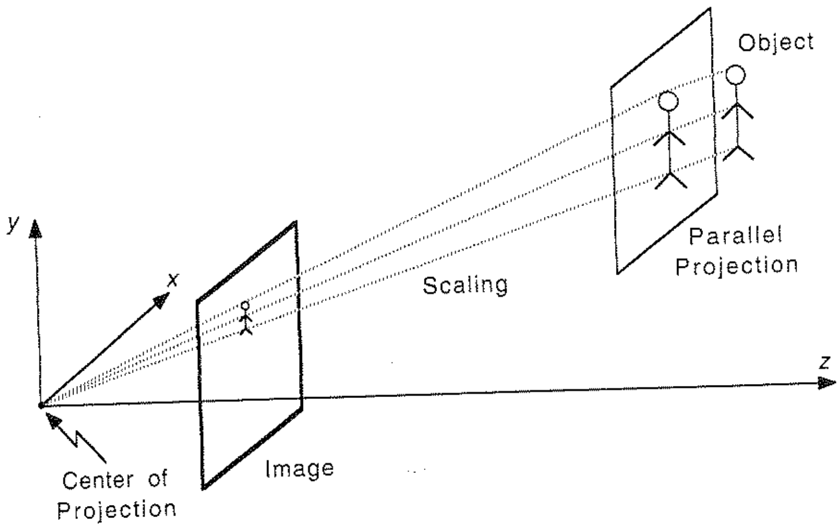


Figure 2.9 Approximation of perspective projection by parallel projection. *Parallel projection* onto a plane is a generalization of orthographic projection in which all the object points are projected along a set of parallel straight lines that may or may not be orthogonal to the projection plane. Perspective projection onto a plane can be approximated by parallel projection, followed by scaling, whenever the object dimensions are small compared to the distance of the object from the center of projection. The direction of parallel projection in such an approximation is along the "average direction" of perspective projection.

2.1.3 Parallel Projection

Parallel projection is a generalization of orthographic projection in which the object is projected onto the image plane by a set of parallel rays that are not necessarily orthogonal to this plane. Parallel projection, like orthographic projection, is a linear transformation in three-dimensional space. Under conditions that we shall examine in this section, parallel projection too provides a convenient approximation to perspective projection up to a uniform scale factor.

As illustrated in Figure 2.9, perspective projection can be approximated by parallel projection up to a uniform scale factor whenever the object's dimensions are small compared to the average distance of the object from the center of projection. The direction of parallel projection is along the "average direction" of perspective projection. When the object, in addition to being *small*, is *close* to the optical axis, the parallel-projection direction can be taken to lie along the optical axis, and we get orthographic projection.

Even when the dimensions of the scene are not small compared to the average distance of the scene from the center of projection, it may be possible

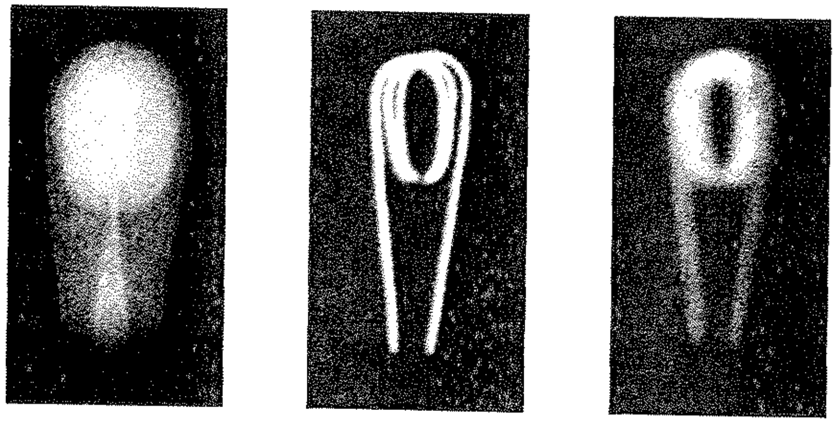


Figure 2.10 Three pinhole-camera photographs (enlarged) of an incandescent filament, each photograph acquired with a circular aperture of a different size. From left to right, the diameter of the aperture is 0.06 inch, 0.015 inch, and 0.0025 inch, respectively; in each case, the distance between the aperture and the image plane is 4 inches. Simple ray tracing would lead us to believe that, as the aperture size is decreased, the image will become sharper. However, when the aperture is reduced below a certain size, rectilinear ray tracing is inadequate for purposes of analysis, and we need to consider *diffraction*, a term used to describe the bending of light rays around the edges of opaque objects. Diffraction, whose extent is inversely related to the ratio of the width of the aperture to the wavelength of the incident light, increasingly blurs the image as the aperture is reduced beyond a certain point. (From [Ruechardt 1958] with permission.)

to partition the scene into smaller subscenes, each of whose dimensions are small compared to its average distance from the center of projection. Under such circumstances, we could approximate perspective projection of the whole scene by a set of parallel projections, each parallel projection applying to a different subscene and having its own projection direction and scale factor. Such an approximation seems to have been proposed first by Ohta, Maenobu, and Sakai [Ohta, Maenobu, and Sakai 1981]; this approximation has subsequently been termed **paraperspective projection** by Aloimonos and Swain [Aloimonos and Swain 1988].

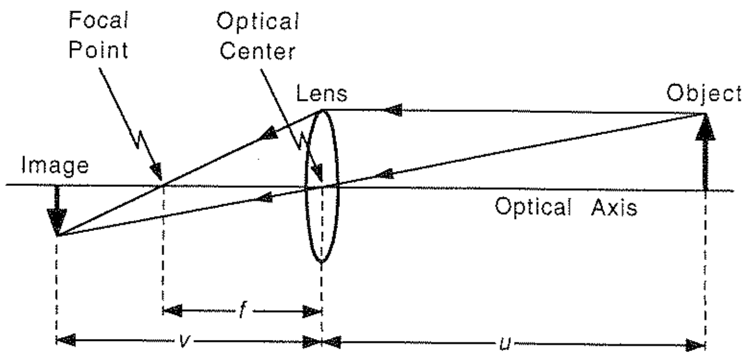
2.1.4 Imaging with a Lens

Thus far, in this chapter, we have considered a pinhole camera, its imaging geometry, and approximations to this geometry. Let us now turn to imaging with a lens. As the size of the aperture of a pinhole camera is reduced, simple ray tracing would lead us to believe that the image will become progressively sharper. However, as demonstrated by the photographs in Figure 2.10, the image will become sharper only up to a point. Below a certain aperture size, rectilinear ray tracing is inadequate for purposes of analysis, and we need to consider **diffraction**, a term used to describe the bending of light rays around the edges of opaque objects. In general, the smaller the width of the aperture relative to the wavelength of the incident light, the more pronounced is the diffraction; see [Ruechardt 1958] for an elementary and readable account of diffraction and related optical phenomena. As the aperture size is reduced, in addition to an increase in the image blurring due to diffraction, there is a decrease in the image intensity (i.e., in the image brightness), the image intensity being directly proportional to the area of the aperture. These considerations lead us to use lenses, the aim of using a lens being to duplicate the pinhole geometry without resorting to undesirably small apertures.

Under ideal circumstances, a lens gathers all the light radiating from an object point toward the lens's finite aperture, and brings this light into focus at a single distinct image point. However, lenses have limitations of their own. The principal limitation of lenses is that, strictly speaking, a lens can bring into focus only those object points that lie within one particular plane parallel to the image plane. We are assuming here, as elsewhere, that the lens is *thin* and that its **optical axis**—that is, its axis of rotation—is perpendicular to the imaging surface, which is a plane. As illustrated in Figure 2.11, a thin lens ideally obeys the **thin-lens equation**, which is also known as the **lens law**:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f},$$

where u is the distance of an object point from the plane of the lens, v is the distance of the focused image of the object point from this plane, and f is the focal length of the lens. It is clear from this equation that, as the object distance u becomes increasingly large with respect to the focal length f, the image distance v approaches f. In fact, the **focal length** of a lens can be defined as the distance from the plane of the lens at which any object point that is located at infinity is brought into focus. Axial object points located at infinity are brought into focus at the **focal point**, which is located on the optical axis at a distance f from the lens; the plane perpendicular to the optical axis at the focal point is called the **focal plane**. Even when objects are not quite at infinity, it is often reasonable to assume that they are brought into focus approximately on the focal plane. Unless stated otherwise, we shall assume throughout this book that the objects are brought into focus on the focal plane, and that the image plane is coincident with the focal plane. As we discussed earlier, the purpose of using a lens is to duplicate the pinhole geometry while maintaining a reasonable aperture size. The aperture



Thin-Lens Equation: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

Figure 2.11 Thin-lens equation: $1/u + 1/v = 1/f$, where u is the distance of an object point from the plane of the lens, v is the distance of the image of the object point from this plane, and f is the focal length of the lens. The *thin-lens equation,* also known as the *lens law,* governs the relationship between the distances of an object and its image from the lens, both distances measured along the lens's *optical axis,* which is the axis of rotation of the lens. As illustrated, we can geometrically determine the position of the image of an off-axis object point by invoking the following two rules: (1) all rays parallel to the optical axis of the lens are deflected by the lens to pass through the lens's focal point, which is at a distance f from the lens along its optical axis; and (2) all rays through the lens's *optical center,* which is a central point in the lens along its optical axis, pass through the lens undeflected. Thus, the optical center of the lens plays the role of a "pinhole" in a pinhole camera (see Figure 2.2), the purpose of using a lens being to duplicate the pinhole geometry while gathering light over a much larger aperture than is possible with a pinhole camera.

here, as elsewhere, is assumed to be circular, and, for all practical purposes, to be in the plane of the lens. As illustrated in Figure 2.11, the effective center of perspective projection for a thin lens is at the lens's **optical center,** which is a central point in the lens along its optical axis through which light rays may be assumed to pass undeflected.

The **field of view** of an imaging device describes the cone of viewing directions of the device. This cone, which comprises all the directions from which rays of light strike the image plane after passing through the effective center of projection of the lens, is almost always chosen to be symmetrical about the optical axis of the lens. For any given size of an image, the field of view of an imaging device is inversely related to the magnification of the lens, and, hence, to its focal length (see Figure 2.11). **Wide-angle lenses** have

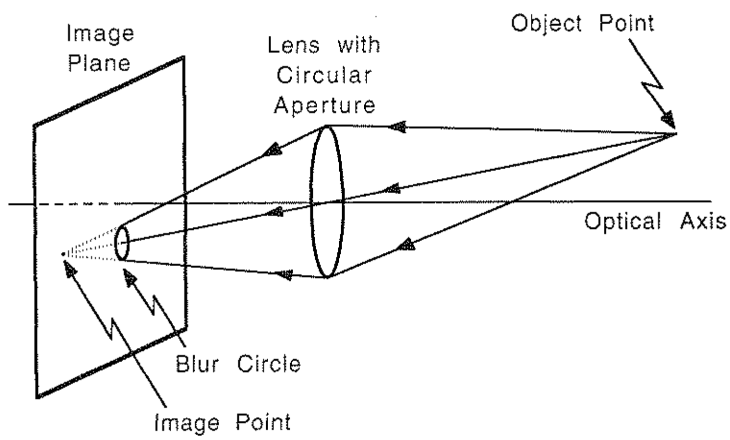


Figure 2.12 Misfocus blur. When a lens brings the image of an object point into focus either in front of or behind the image plane, rather than on the image plane, what appears on the image plane is a blur. If the in-focus image is farther from the lens than the image plane, then this blur has the same shape as the aperture through which light crosses the lens; if the in-focus image is closer to the lens than the image plane, then this blur has the inverted shape of the aperture. As is geometrically evident from the figure, the size of the blur is proportional to the size of the aperture—if we assume that the aperture lies in the plane of the lens, then the factor of proportionality, which may be used as an index of the misfocus, is the ratio of the distance of the in-focus image from the image plane to the distance of the in-focus image from the plane of the lens.

small focal lengths, and, as a result, they have large fields of view. **Telephoto lenses**, on the other hand, have large focal lengths, and, as a result, they have small fields of view. As a practical matter, the perspective imaging geometry of an imaging device is approximated closely by orthographic projection (up to a uniform scale factor) whenever a telephoto lens is used to view a distant scene that has a relatively small range of depth. Clearly, such an approximation is inappropriate when a wide-angle lens is used.

Now, as is clear from the thin-lens equation, for any given position of the image plane, only points within one particular object plane are brought into focus on the image plane by an ideal thin lens. As illustrated in Figure 2.12, points that do not lie within the particular plane brought into focus are imaged as **blur circles**—also known as **circles of confusion**—each blur circle being formed by the intersection of the corresponding cone of light rays with the image plane. As is clear from the figure, the diameter of a blur circle is proportional to the diameter of the aperture. Hence, as the aperture size is decreased, the range of depths over which the world is approximately in focus, better known as the **depth of field**, increases, and errors in focusing become less important. This increase in depth of field, of course, is accompanied by a reduction in the image intensity, to compensate for which it might be necessary to use longer exposure times for image sensing. Clearly, we have a fundamental tradeoff here: loss of resolution (i.e., discriminability) in space, versus that in time—equivalently, image blur due to misfocus, versus image blur due to motion during image capture. This tradeoff, of course, is precisely what aperture-adjustment mechanisms in cameras allow us to control.

Aberrations and Diffraction

Although lenses allow us to overcome some of the limitations of a pinhole camera, they are not without problems. Every lens typically exhibits several imperfections or aberrations. An **aberration** of a lens is any failure of the lens to bring together at the following specific point all the light radiating toward the lens from a single object point: the point that lies along the straight line through the object point and the optical center of the lens, at a distance governed by the lens law (see Figure 2.11).

There are several types of aberrations that a lens might exhibit; for an extensive discussion of aberrations, see the *Manual of Photogrammetry* [Slama 1980] and [Hecht and Zajac 1974]. To begin with, not only does an ideal lens bring into focus just one plane, but also this plane depends on the wavelength of the incident light; this dependence is a consequence of the dependence of the refractive index of the lens on the wavelength of the incident light. As a result, we have **chromatic aberrations** that are caused by radiation at different wavelengths from a single point being brought into focus at different points, which has the effect of blurring the image. Even with monochromatic radiation—that is, with radiation at a single wavelength—the image may be blurred owing to the inability of the lens to bring into focus at a single point all the light rays radiating toward the lens from a single object point. **Spherical aberration** describes the failure of a lens to bring into focus at a single point monochromatic light rays originating at a single point on the optical axis of the lens; **coma** and **astigmatism** describe the failure of the lens to do the same for monochromatic rays from an off-axis point. The three-dimensional blurring of individual image points is not the only possible aberration. Even when a lens is capable of bringing into focus at a single point all the radiation of every wavelength impinging on the lens from a single object point, we are not guaranteed a perfect image. That is, we are not guaranteed an image that is perfect in the sense of its being a planar perspective projection of the scene when this scene is planar

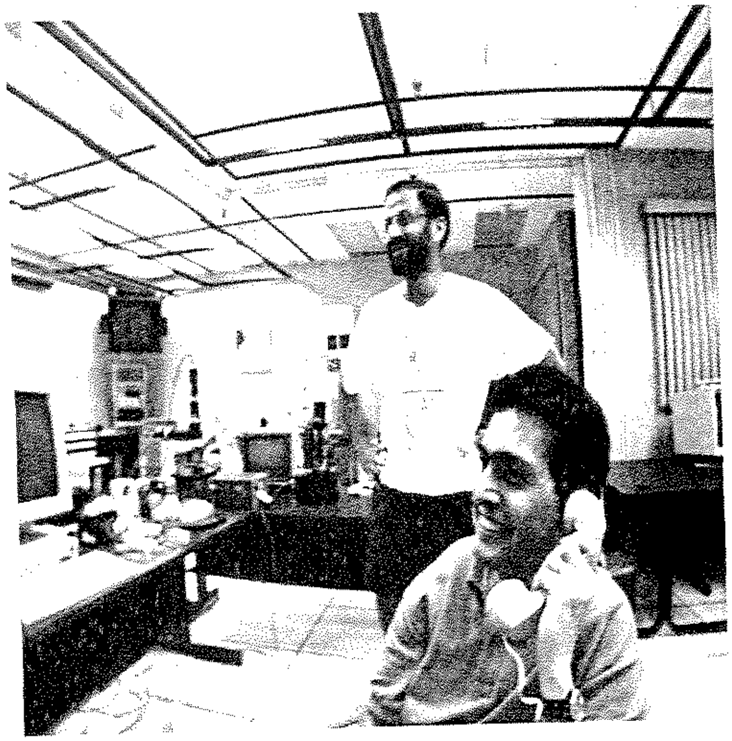


Figure 2.13 Image distortion. Even when a lens brings the image of each object point into focus at a single point on the image plane, we are not guaranteed a perfect image: The geometry of the image may not conform to a planar perspective projection of the scene. Such an aberration, termed *image distortion*, causes straight lines in the scene to appear bowed in the image, as in the ceiling of the room in the photograph shown.

and orthogonal to the optical axis of the lens. The image of such a scene may be brought into focus not on a plane, but instead on a curved surface—such an aberration is termed **curvature of field**. Further, even when a planar scene that is orthogonal to the optical axis of the lens is brought into focus on a single plane that is orthogonal to the optical axis of the lens, the image may be distorted—such an aberration is simply called **distortion**. Image distortion, which is illustrated for a general scene in Figure 2.13, is of particular concern when a wide-angle lens is used. Manufacturers of optical equipment seek to minimize the net effect of the various aberrations on the overall image quality by designing complex lens systems that are composed of several carefully selected and aligned individual lens elements. A reduction in the aperture size is also helpful in reducing the effect of aberrations on the image, but such a reduction could lead to an unacceptable reduction in the intensity of the image.

Even if a lens were perfectly free of aberrations, the physical nature of light would preclude a perfect image. We would still need to take into account the effects of **diffraction**, which, as we saw earlier, describes the deviation of light from a rectilinear path at the edges of opaque objects. A lens whose image quality is limited by diffraction—rather than by aberrations—is said to be **diffraction limited**. As a result of diffraction, the image of a point object formed by an aberration-free lens obeying the lens law is not a point on the image plane even when this plane is at the distance v dictated by the lens law. In particular, if the aperture of the lens is circular, then the image of a point object is a circular disc surrounded by progressively fainter rings; such a diffraction pattern is called the **Airy pattern**, after the astronomer who first derived its equation in the early nineteenth century. The radius of the central disc of the Airy pattern is $1.22\,\lambda v/d$, where λ is the wavelength of the incident light, v is the distance of the image plane from the lens, and d is the diameter of the circular aperture of the lens; see, for instance, [Hecht and Zajac 1974] and [Goodman 1968]. Thus, in a diffraction-limited imaging system, we can improve the image quality in two ways: (1) by increasing the size of the aperture, and (2) when feasible, by reducing the wavelength of the light forming the image. The former strategy, which also increases the brightness of the image, is adopted in the design of telescopes, and the latter strategy is adopted in the design of microscopes.

Camera Calibration

Despite all the approximations and problems with lenses, it must be emphasized that perspective projection is an extremely useful and convenient model for the geometry of image formation by a lens. We must, however, always bear in mind that that's just what perspective projection is: It is a model.

To derive three-dimensional geometric information from an image, it is necessary to determine the parameters that relate the position of a scene point to the position of its image. This determination is known as **camera calibration**, or, more accurately, as **geometric camera calibration**. Let us assume that the perspective-projection model is valid. Let us further assume a global coordinate frame for the scene, and an independent two-dimensional frame for the image. We need to relate the spatial positions and orientations of these two frames, and to determine the position of the center of projection. In addition, to account for the transformation undergone by an image between its capture on the image plane and its display, we need to determine two independent scale factors, one for each image coordinate axis.

As perspective projection and image scaling along any direction in the image are both linear transformations in homogeneous coordinates, each of these operations, and, therefore, the complete mapping from a scene position to its image position, can be expressed as a multiplicative matrix in homogeneous coordinates. Given the image positions and scene coordinates of six points, it is straightforward to derive a closed-form solution to this matrix (see [Ballard and Brown 1982]); more points offer greater robustness. Ganapathy [Ganapathy 1984] has shown that this matrix, in turn, provides closed-form solutions to the six **extrinsic camera parameters** and to the four **intrinsic camera parameters**. Of the six extrinsic camera parameters, three are for the position of the center of projection, and three are for the orientation of the image-plane coordinate frame. Of the four intrinsic camera parameters, two are for the position of the origin of the image coordinate frame, and two are for the scale factors of the axes of this frame. Although the distance of the image plane from the center of projection cannot be modeled independently of the scale factors of the axes of the image, as indicated in our discussion of lenses, this distance is often well approximated by the focal length of the lens. On the other hand, if the scale factors of the image axes are known a priori, this distance too may be calibrated.

Typically, camera calibration is pursued using a known calibration object whose images exhibit a large number of distinct points that can be identified easily and located accurately in the image. Clearly, it is desirable that the calibration object be easy to generate and to measure accurately, and that the shape of the object be conducive to simplifying the calibration computations. One object that meets these criteria comprises either one or multiple planar rectilinear grids [Tsai 1986].

Tsai [Tsai 1986] argues that, in practice, it is necessary to model and calibrate image distortion in addition to the ideal-case parameters we have discussed. He reviews previous calibration techniques, and then describes a now widely used calibration procedure designed for accuracy, robustness, and efficiency. Also of interest here is the work by Fischler and Bolles [Fischler and Bolles 1981], who investigate the determination of the extrinsic camera parameters under knowledge of the intrinsic camera parameters; this problem is called the **exterior camera-orientation problem**. In particular, Fischler and Bolles show that, given the image positions and scene locations of three points, the extrinsic camera parameters have at most four solutions, each solution expressible in closed form.