considers cameras as *radiometric* devices for measuring light energy, brightness, and color. Here, we focus instead on purely geometric camera characteristics. After introducing several models of image formation in Section 1.1—including a brief description of this process in the human eye in Section 1.1.4—we define the *intrinsic* and *extrinsic* geometric parameters characterizing a camera in Section 1.2, and finally show how to estimate these parameters from image data—a process known as *geometric camera calibration*—in Section 1.3.

1.1 IMAGE FORMATION

1.1.1 Pinhole Perspective

Imagine taking a box, using a pin to prick a small hole in the center of one of its sides, and then replacing the opposite side with a translucent plate. If you hold that box in front of you in a dimly lit room, with the pinhole facing some light source, say a candle, an inverted image of the candle will appear on the translucent plate (Figure 1.2). This image is formed by light rays issued from the scene facing the box. If the pinhole were really reduced to a point (which is physically impossible, of course), exactly one light ray would pass through each point in the plane of the plate (or *image plane*), the pinhole, and some scene point.

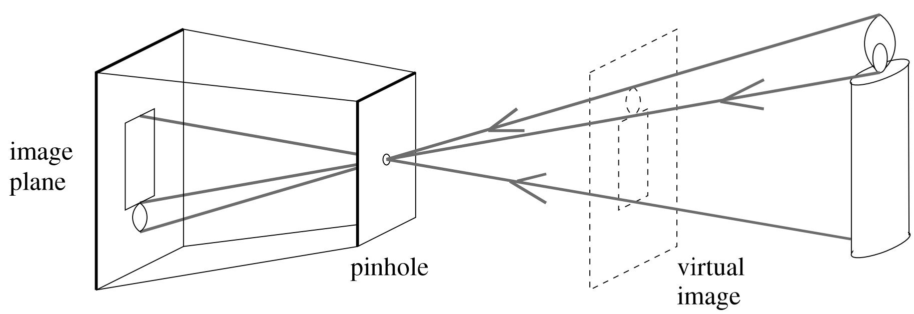

FIGURE 1.2: The pinhole imaging model.

In reality, the pinhole will have a finite (albeit small) size, and each point in the image plane will collect light from a cone of rays subtending a finite solid angle, so this idealized and extremely simple model of the imaging geometry will not strictly apply. In addition, real cameras are normally equipped with lenses, which further complicates things. Still, the *pinhole perspective* (also called *central perspective*) projection model, first proposed by Brunelleschi at the beginning of the fifteenth century, is mathematically convenient and, despite its simplicity, it often provides an acceptable approximation of the imaging process. Perspective projection creates inverted images, and it is sometimes convenient to consider instead a *virtual image* associated with a plane lying *in front* of the pinhole, at the same distance from it as the actual image plane (Figure 1.2). This virtual image is not inverted but is otherwise strictly equivalent to the actual one. Depending on the context, it may be more convenient to think about one or the other. Figure 1.3 (a) illustrates an obvious effect of perspective projection: the apparent size of objects depends on their distance. For example, the images b and c of the posts B and C have the same height, but A and C are really half the size of B. Figure 1.3 (b) illustrates another well-known effect: the projections of two parallel lines lying in some plane Φ appear to converge on a horizon line h formed by the intersection of the image plane Π with the plane parallel to Φ and passing through the pinhole. Note that the line L parallel to Π in Φ has no image at all.

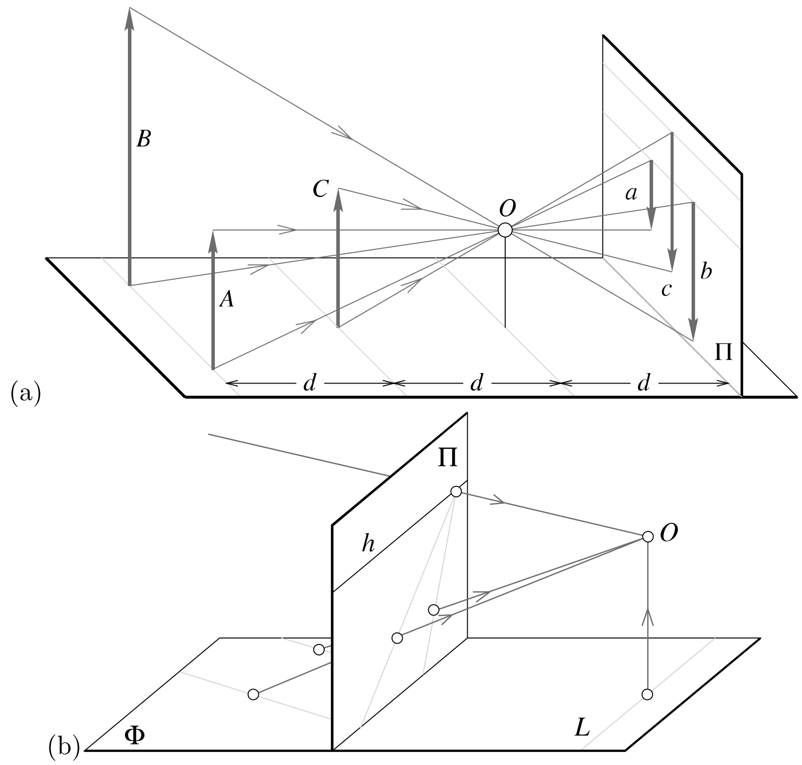

FIGURE 1.3: Perspective effects: (a) far objects appear smaller than close ones: The distance d from the pinhole O to the plane containing C is half the distance from O to the plane containing A and B; (b) the images of parallel lines intersect at the horizon (after Hilbert and Cohn-Vossen, 1952, Figure 127). Note that the image plane Π is *behind* the pinhole in (a) (physical retina), and *in front* of it in (b) (virtual image plane). Most of the diagrams in this chapter and in the rest of this book will feature the physical image plane, but a virtual one will also be used when appropriate, as in (b).

These properties are easy to prove in a purely geometric fashion. As usual, however, it is often convenient (if not quite as elegant) to reason in terms of reference frames, coordinates, and equations. Consider, for example, a coordinate system $(O, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ attached to a pinhole camera, whose origin O coincides with the pinhole, and vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ form a basis for a vector plane parallel to the image plane Π, itself located at a positive distance d from the pinhole along the vector $\boldsymbol{k}$ (Figure 1.4). The line perpendicular to Π and passing through the pinhole is called the optical axis, and the point c where it pierces Π is called the *image center*. This point can be used as the origin of an image plane coordinate frame, and it plays an important role in camera calibration procedures.

Let P denote a scene point with coordinates (X, Y, Z) and p denote its image

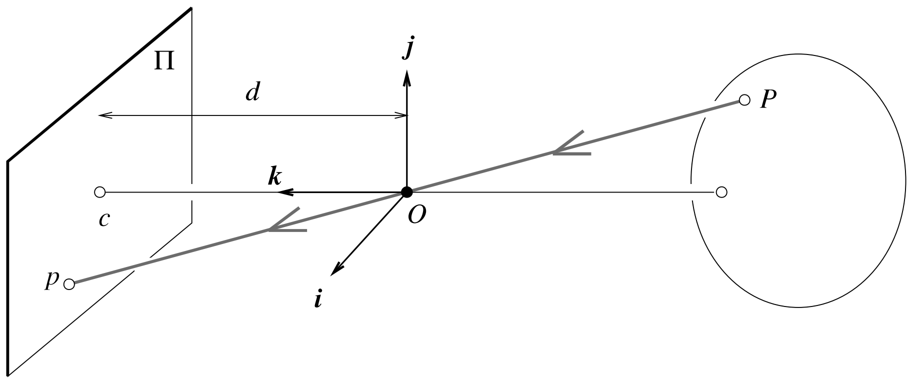

FIGURE 1.4: The perspective projection equations are derived in this section from the collinearity of the point P, its image p, and the pinhole O.

with coordinates (x, y, z). (Throughout this chapter, we will use uppercase letters to denotes points in space, and lowercase letters to denote their image projections.) Since p lies in the image plane, we have $z = d$. Since the three points P, O, and p are collinear, we have $\overrightarrow{Op} = \lambda \overrightarrow{OP}$ for some number λ, so

$$\begin{cases} x = \lambda X \\ y = \lambda Y \\ d = \lambda Z \end{cases} \iff \lambda = \frac{x}{X} = \frac{y}{Y} = \frac{d}{Z},$$

and therefore

$$\begin{cases} x = d\dfrac{X}{Z}, \\ y = d\dfrac{Y}{Z}. \end{cases} \tag{1.1}$$

1.1.2 Weak Perspective

As noted in the previous section, pinhole perspective is only an approximation of the geometry of the imaging process. This section discusses a coarser approximation, called *weak perspective*, which is also useful on occasion.

Consider the *fronto-parallel plane* Π_0 defined by $Z = Z_0$ (Figure 1.5). For any point P in Π_0 we can rewrite Eq. (1.1) as

$$\begin{cases} x = -mX, \\ y = -mY, \end{cases} \quad \text{where} \quad m = -\frac{d}{Z_0}. \tag{1.2}$$

Physical constraints impose that Z_0 be negative (the plane must be in front of the pinhole), so the *magnification* m associated with the plane Π_0 is positive. This name is justified by the following remark: consider two points P and Q in Π_0 and their images p and q (Figure 1.5); obviously, the vectors $\overrightarrow{PQ}$ and $\overrightarrow{pq}$ are parallel, and we have $||\overrightarrow{pq}|| = m||\overrightarrow{PQ}||$. This is the dependence of image size on object distance noted earlier.

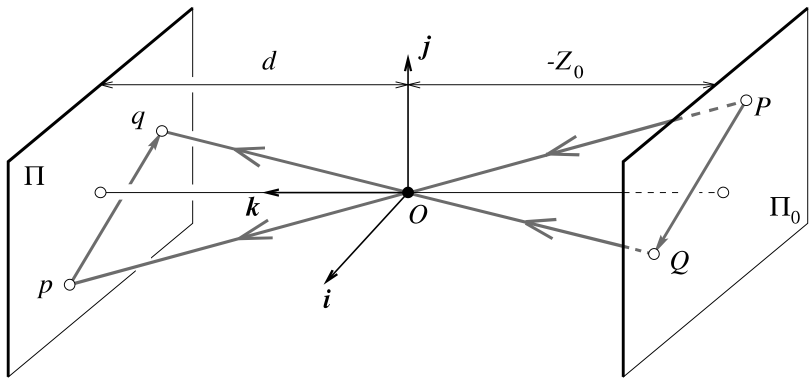

FIGURE 1.5: Weak-perspective projection. All line segments in the plane Π_0 are projected with the same magnification.

When a scene's relief is small relative to its average distance from the camera, the magnification can be taken to be constant. This projection model is called *weak perspective*, or *scaled orthography*.

When it is a priori known that the camera will always remain at a roughly constant distance from the scene, we can go further and normalize the image coordinates so that $m = -1$. This is *orthographic projection*, defined by

$$\begin{cases} x = X, \\ y = Y, \end{cases} \tag{1.3}$$

with all light rays parallel to the $\boldsymbol{k}$ axis and orthogonal to the image plane π (Figure 1.6). Although weak-perspective projection is an acceptable model for many imaging conditions, assuming pure orthographic projection is usually unrealistic.

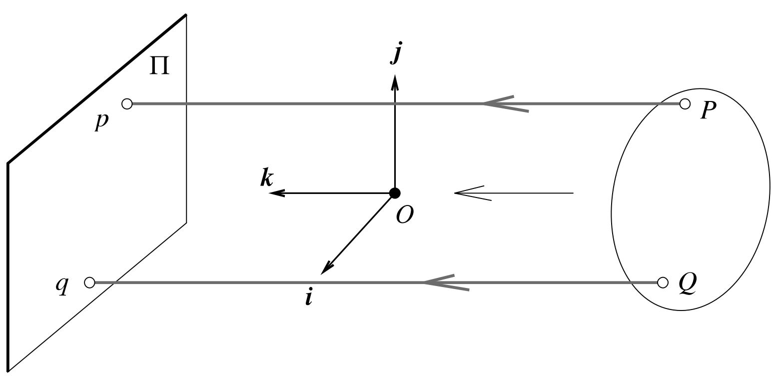

FIGURE 1.6: Orthographic projection. Unlike other geometric models of the image formation process, orthographic projection does not involve a reversal of image features. Accordingly, the magnification is taken to be negative, which is a bit unnatural but simplifies the projection equations.

1.1.3 Cameras with Lenses

Most real cameras are equipped with lenses. There are two main reasons for this: The first one is to gather light, because a single ray of light would otherwise reach each point in the image plane under ideal pinhole projection. Real pinholes have a finite size, of course, so each point in the image plane is illuminated by a cone of light rays subtending a finite solid angle. The larger the hole, the wider the cone and the brighter the image, but a large pinhole gives blurry pictures. Shrinking the pinhole produces sharper images but reduces the amount of light reaching the image plane, and may introduce *diffraction* effects. Keeping the picture in sharp focus while gathering light from a large area is the second main reason for using a lens.

Ignoring diffraction, interferences, and other physical optics phenomena, the behavior of lenses is dictated by the laws of geometric optics (Figure 1.7): (1) light travels in straight lines (*light rays*) in homogeneous media; (2) when a ray is reflected from a surface, this ray, its reflection, and the surface normal are coplanar, and the angles between the normal and the two rays are complementary; and (3) when a ray passes from one medium to another, it is *refracted*, i.e., its direction changes. According to Snell's law, if r_1 is the ray incident to the interface between two transparent materials with indices of refraction n_1 and n_2, and r_2 is the refracted ray, then r_1, r_2, and the normal to the interface are coplanar, and the angles α_1 and α_2 between the normal and the two rays are related by

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2. \tag{1.4}$$

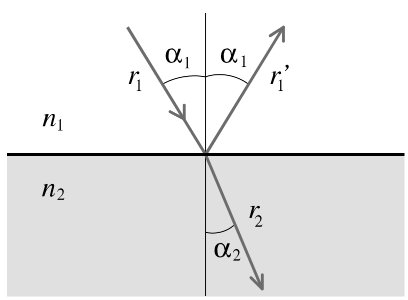

FIGURE 1.7: Reflection and refraction at the interface between two homogeneous media with indices of refraction n_1 and n_2.

In this chapter, we will only consider the effects of refraction and ignore those of reflection. In other words, we will concentrate on lenses as opposed to *catadioptric optical systems* (e.g., telescopes) that may include both reflective (mirrors) and refractive elements. Tracing light rays as they travel through a lens is simpler when the angles between these rays and the refracting surfaces of the lens are assumed to be small, which is the domain of *paraxial* (or *first-order*) geometric optics, and Snell's law becomes $n_1\alpha_1 \approx n_2\alpha_2$. Let us also assume that the lens is rotationally symmetric about a straight line, called its *optical axis*, and that all refractive surfaces are spherical. The symmetry of this setup allows us to determine

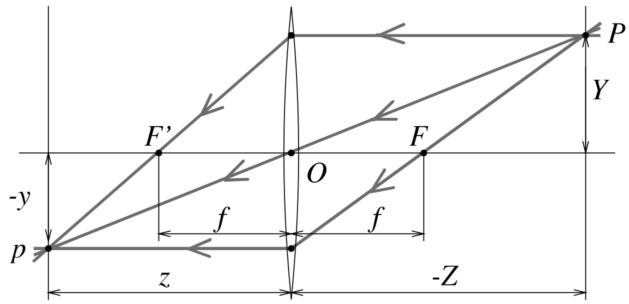

FIGURE 1.8: A thin lens. Rays passing through O are not refracted. Rays parallel to the optical axis are focused on the focal point F'.

the projection geometry by considering lenses with circular boundaries lying in a plane that contains the optical axis. In particular, consider a lens with two spherical surfaces of radius R and index of refraction n. We will assume that this lens is surrounded by vacuum (or, to an excellent approximation, by air), with an index of refraction equal to 1, and that it is *thin*, i.e., that a ray entering the lens and refracted at its right boundary is immediately refracted again at the left boundary.

Consider a point P located at (negative) depth Z off the optical axis, and denote by (PO) the ray passing through this point and the center O of the lens (Figure 1.8). It easily follows from the paraxial form of Snell's law that (PO) is not refracted, and that all the other rays passing through P are focused by the thin lens on the point p with depth z along (PO) such that

$$\frac{1}{z} - \frac{1}{Z} = \frac{1}{f}, \tag{1.5}$$

where $f = \frac{R}{2(n-1)}$ is the *focal length* of the lens.

Note that the equations relating the positions of P and p are exactly the same as under pinhole perspective projection if we take $d = z$ since P and p lie on a ray passing through the center of the lens, but that points located at a distance $-Z$ from O will be in sharp focus only when the image plane is located at a distance z from O on the other side of the lens that satisfies Eq. (1.5), the *thin lens equation.* Letting $Z \to -\infty$ shows that f is the distance between the center of the lens and the plane where objects such as stars (that are effectively located at $Z = -\infty$) focus. The two points F and F' located at distance f from the lens center on the optical axis are called the *focal points* of the lens. In practice, objects within some range of distances (called *depth of field* or *depth of focus*) will be in acceptable focus. As shown in the problems at the end of this chapter, the depth of field increases with the *f-number* of the lens, i.e., the ratio between the focal length of the lens and its diameter.

Note that the *field of view* of a camera, i.e., the portion of scene space that actually projects onto the retina of the camera, is not defined by the focal length alone but also depends on the effective area of the retina (e.g., the area of film that can be exposed in a photographic camera, or the area of the sensor in a digital camera; see Figure 1.9).

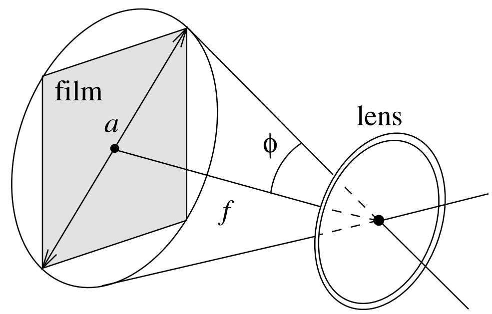

FIGURE 1.9: The field of view of a camera. It can be defined as 2ϕ, where $\phi \stackrel{\text{def}}{=} \arctan \frac{a}{2f}$, a is the diameter of the sensor (film, CCD, or CMOS chip), and f is the focal length of the camera.

A more realistic model of simple optical systems is the *thick lens*. The equations describing its behavior are easily derived from the paraxial refraction equation, and they are the same as the pinhole perspective and thin lens projection equations, except for an offset (Figure 1.10). If H and H' denote the *principal points* of the lens, then Eq. (1.5) holds when $-Z$ (resp. z) is the distance between P (resp. p) and the plane perpendicular to the optical axis and passing through H (resp. H'). In this case, the only undeflected ray is along the optical axis.

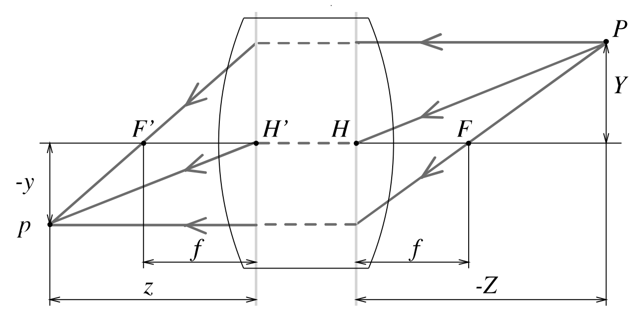

FIGURE 1.10: A simple thick lens with two spherical surfaces.

Simple lenses suffer from a number of *aberrations*. To understand why, let us remember that the paraxial refraction model is only an approximation, valid when the angle α between each ray along the optical path and the optical axis of the length is small and $\sin \alpha \approx \alpha$. This corresponds to a first-order Taylor expansion of the sine function. For larger angles, additional terms yield a better approximation, and it is easy to show that rays striking the interface farther from the optical axis are focused closer to the interface. The same phenomenon occurs for a lens, and it is the source of two types of *spherical aberrations* (Figure 1.11 [a]): Consider a point P on the optical axis and its paraxial image p. The distance between p and the intersection of the optical axis with a ray issued from P and refracted by the lens is called the longitudinal spherical aberration of that ray. Note that if an image plane Π were erected in P, the ray would intersect this plane at some distance from