CHAPTER 1

Geometric Camera Models

There are many types of imaging devices, from animal eyes to video cameras and radio telescopes. They may or may not be equipped with lenses: For example the first models of the *camera obscura* (literally, dark chamber) invented in the 16th century did not have lenses, but instead used a *pinhole* to focus light rays onto a wall or translucent plate and demonstrate the laws of perspective discovered a century earlier by Brunelleschi. Pinholes were replaced by more and more sophisticated lenses as early as 1550, and the modern photographic or digital camera is essentially a camera obscura capable of recording the amount of light striking every small area of its backplane (Figure 1.1).

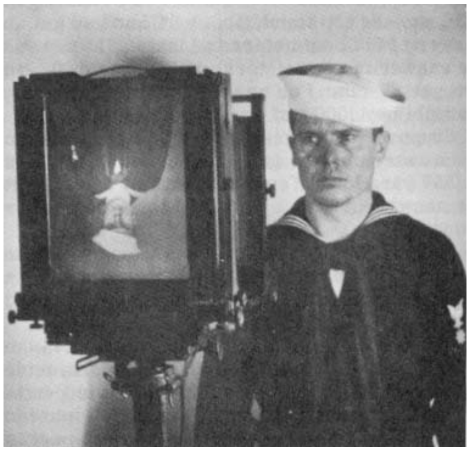

FIGURE 1.1: Image formation on the backplate of a photographic camera. *Figure from US NAVY MANUAL OF BASIC OPTICS AND OPTICAL INSTRUMENTS, prepared by the Bureau of Naval Personnel, reprinted by Dover Publications, Inc., (1969).*

The imaging surface of a camera is in general a rectangle, but the shape of the human retina is much closer to a spherical surface, and panoramic cameras may be equipped with cylindrical retinas. Imaging sensors have other characteristics. They may record a spatially discrete picture (like our eyes with their rods and cones, 35mm cameras with their grain, and digital cameras with their rectangular picture elements or pixels), or a continuous one (in the case of old-fashioned TV tubes, for example). The signal that an imaging sensor records at a point on its retina may itself be discrete or continuous, and it may consist of a single number (black-and-white camera), a few values (e.g., the RGB intensities for a color camera, or the responses of the three types of cones for the human eye), many numbers (e.g., the responses of hyperspectral sensors) or even a continuous function of wavelength (which is essentially the case for spectrometers). Chapter **??** 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, you will see an inverted image of the candle appearing 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 of course physically impossible), exactly one light ray would pass through each point in the plane of the plate (or *image plane*), the pinhole, and some scene point.

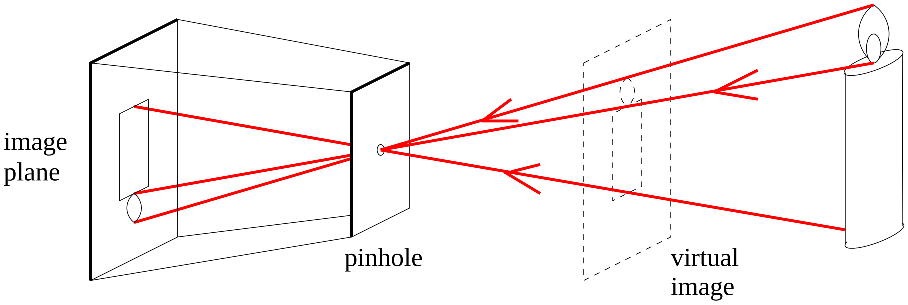

FIGURE 1.2: The pinhole imaging model.

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.

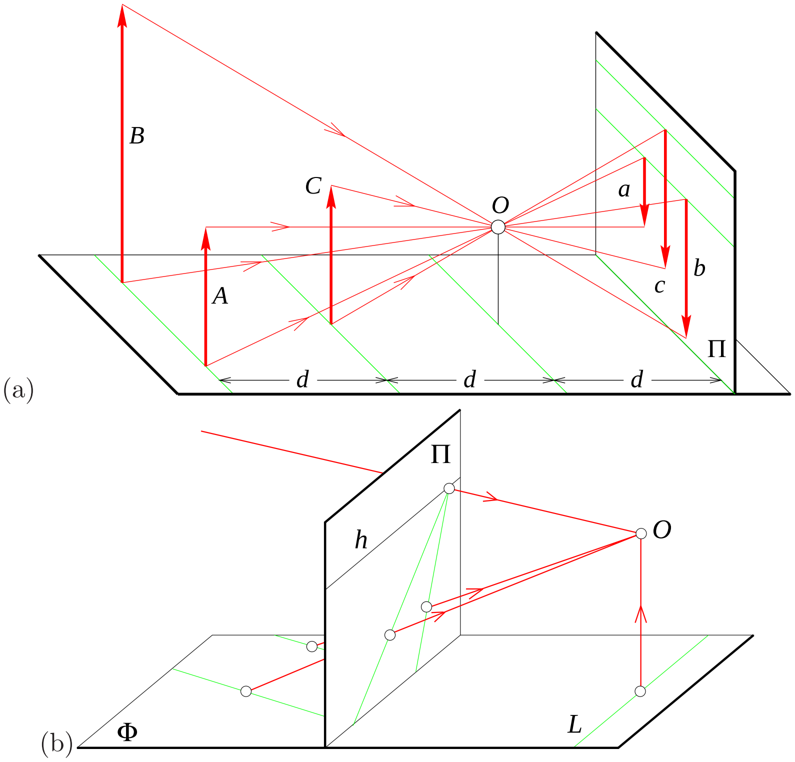

FIGURE 1.3: Perspective effects: (a) far objects appear smaller than close ones: The distance d from the pinhole O to the plane containing C is half the distance from O to the plane containing A and B; (b) the images of parallel lines intersect at the horizon (after Hilbert and Cohn-Vossen, 1952, Figure 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 of course 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

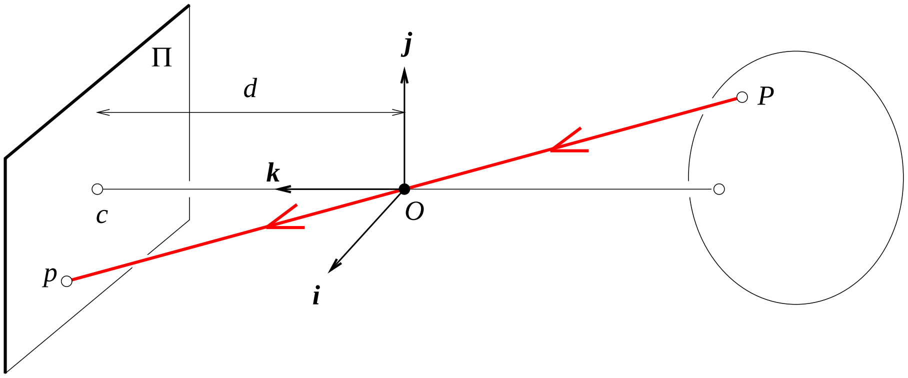

FIGURE 1.4: The perspective projection equations are derived in this section from the collinearity of the point P, its image p and the pinhole O.

with coordinates (x, y, z) (we will use throughout 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

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

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.

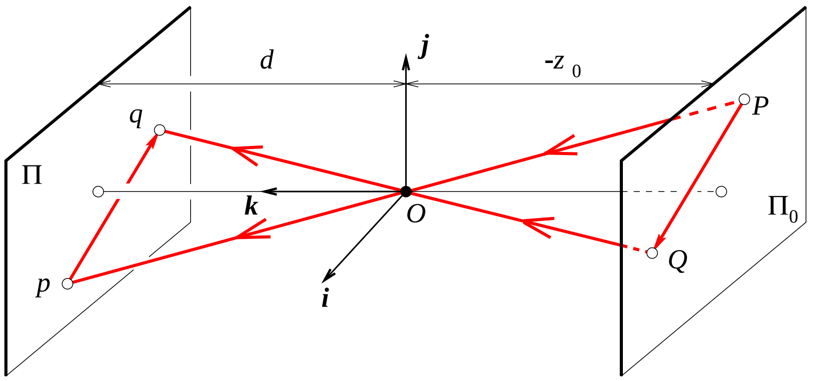

FIGURE 1.5: Weak-perspective projection: All line segments in the plane Π_0 are projected with the same magnification.

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.

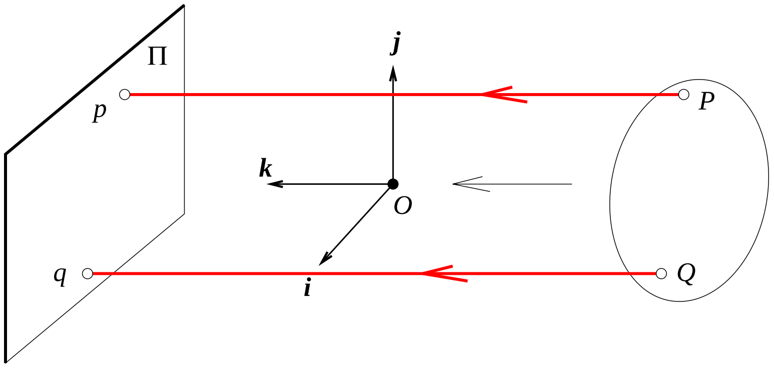

FIGURE 1.6: Orthographic projection. Unlike other geometric models of the image formation process, orthographic projection does not involve a reversal of image features. Accordingly, the magnification is taken to be negative, which is a bit unnatural but simplifies the projection equations.

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, since 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 Descartes' 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

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

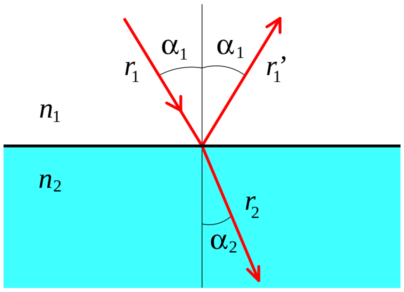

FIGURE 1.7: Reflection and refraction at the interface between two homogeneous media with indices of refraction n_1 and n_2.

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

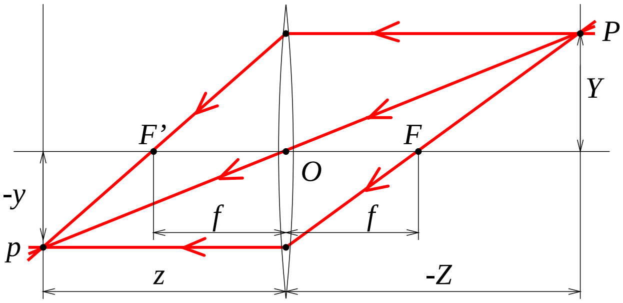

FIGURE 1.8: A thin lens. Rays passing through O are not refracted. Rays parallel to the optical axis are focused on the focal point F'.

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 only be in sharp focus 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 exercises, 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).

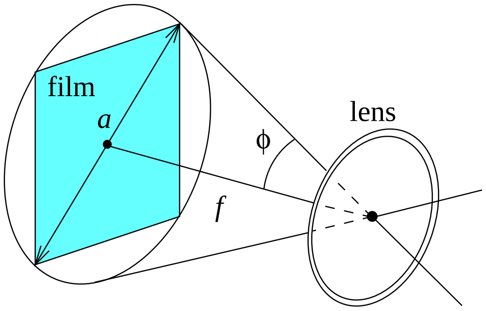

FIGURE 1.9: The field of view of a camera. It can be defined as 2ϕ, where $\phi \overset{\text{def}}{=} \arctan \frac{a}{2f}$, a is the diameter of the sensor (film, CCD, or CMOS chip) and f is the focal length of the camera.

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.

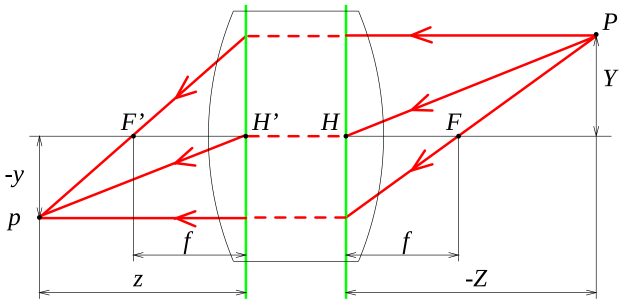

FIGURE 1.10: A simple thick lens with two spherical surfaces.

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 the axis, called the transverse spherical aberration of that ray. Together, all rays passing through P and refracted by the lens form a circle of confusion centered in P as they intersect Π. The size of that circle will change if we move Π along the optical axis. The circle with minimum diameter is called the *circle of least confusion*, and its center does not coincide (in general) with p.

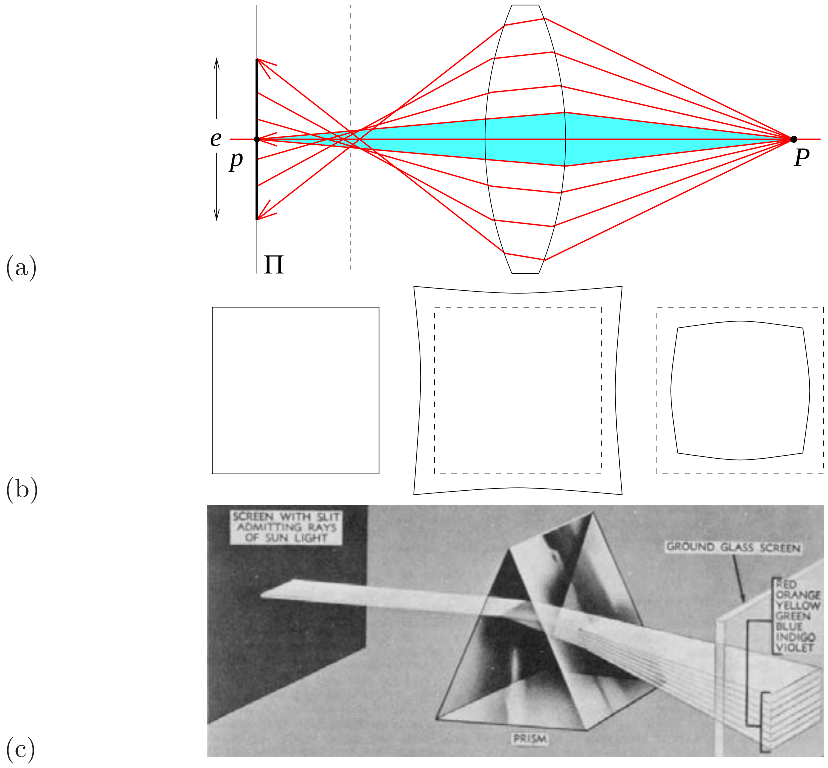

FIGURE 1.11: Aberrations. (a) Spherical aberration: The grey region is the paraxial zone where the rays issued from P intersect at its paraxial image p. If an image plane π were erected in p, the image of p in that plane would form a circle of confusion of diameter e. The focus plane yielding the circle of least confusion is indicated by a dashed line. (b) Distortion: From left to right, the nominal image of a fronto-parallel square, pincushion distortion, and barrel distortion. (c) Chromatic aberration: The index of refraction of a transparent medium depends on the wavelength (or colour) of the incident light rays. Here, a prism decomposes white light into a palette of colours. *Figure from US NAVY MANUAL OF BASIC OPTICS AND OPTICAL INSTRUMENTS, prepared by the Bureau of Naval Personnel, reprinted by Dover Publications, Inc., (1969).*

Besides spherical aberration, there are four other types of *primary aberrations* caused by the differences between first- and third-order optics, namely *coma*, *astigmatism*, *field curvature* and *distortion*. A precise definition of these aberrations is beyond the scope of this book. Suffice to say that, like spherical aberration, the first three degrade the image by blurring the picture of every object point. Distortion, on the other hand, plays a different role and changes the shape of the image

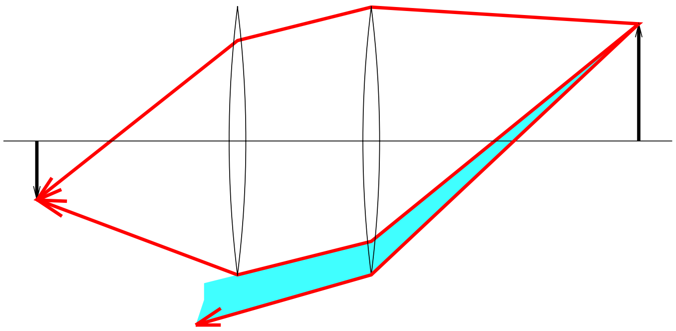

FIGURE 1.12: Vignetting effect in a two-lens system. The shaded part of the beam never reaches the second lens. Additional apertures and stops in a lens further contribute to vignetting.

as a whole (Figure 1.11 [b]). This effect is due to the fact that different areas of a lens have slightly different focal lengths. The aberrations mentioned so far are monochromatic, i.e., they are independent of the response of the lens to various wavelengths. However, the index of refraction of a transparent medium depends on wavelength (Figure 1.11 [c]), and it follows from the thin lens equation (Eq. [1.5]) that the focal length depends of wavelength as well. This causes the phenomenon of *chromatic aberration*: Refracted rays corresponding to different wavelengths will intersect the optical axis at different points (longitudinal chromatic aberration) and form different circles of confusion in the same image plane (transverse chromatic aberration).

Aberrations can be minimized by aligning several simple lenses with well-chosen shapes and refraction indices, separated by appropriate stops. These *compound lenses* can still be modeled by the thick lens equations. They suffer from one more defect relevant to machine vision: Light beams emanating from object points located off-axis are partially blocked by the various apertures (including the individual lens components themselves) positioned inside the lens to limit aberrations (Figure 1.12). This phenomenon, called *vignetting*, causes the image brightness to drop in the image periphery. Vignetting may pose problems to automated image analysis programs, but it is not quite as important in photography, thanks to the human eye's remarkable insensitivity to smooth brightness gradients. Speaking of which, it is time to have a look at this extraordinary organ.

1.1.4 The Human Eye

Here we give a (brief) overview of the anatomical structure of the eye. It is largely based on the presentation in Wandell (1995), and the interested reader is invited to read this excellent book for more details. Figure 1.13 (left) is a sketch of the section of an eyeball through its vertical plane of symmetry, showing the main elements of the eye: the *iris* and the *pupil*, which control the amount of light penetrating the eyeball; the *cornea* and the crystalline *lens*, which together refract the light to create the retinal image; and finally the *retina*, where the image is formed. Despite its globular shape, the human eyeball is functionally similar to a camera with a field of view covering a 160° (width) × 135° (height) area. Like any other optical system, it suffers from various types of geometric and chromatic aberrations. Several models of the eye obeying the laws of first-order geometric optics have been proposed, and Figure 1.13(right) shows one of them, *Helmoltz's schematic eye*. There are only three refractive surfaces, with an infinitely thin cornea and a homogeneous lens. The constants given in Figure 1.13 are for the eye focusing at infinity (*unaccommodated eye*). This model is of course only an approximation of the real optical characteristics of the eye.

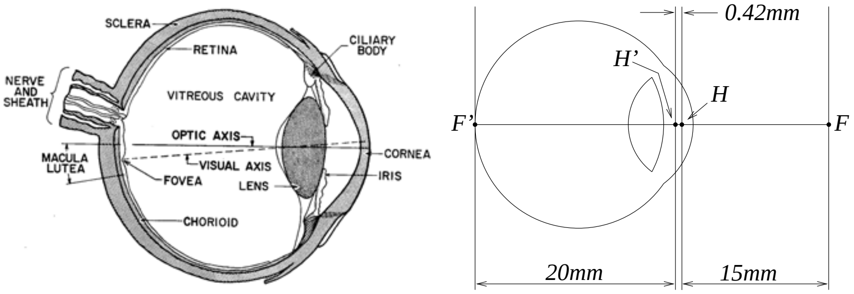

FIGURE 1.13: Left: the main components of the human eye. *Reproduced with permission, the American Society for Photogrammetry and Remote Sensing. A.L. Nowicki, "Stereoscopy." MANUAL OF PHOTOGRAMMETRY, edited by M.M. Thompson, R.C. Eller, W.A. Radlinski, and J.L. Speert, third edition, pp. 515–536. Bethesda: American Society of Photogrammetry, (1966).* Right: Helmoltz's schematic eye as modified by Laurance (after Driscoll and Vaughan, 1978). The distance between the pole of the cornea and the anterior principal plane is 1.96 mm, and the radii of the cornea, anterior, and posterior surfaces of the lens are respectively 8 mm, 10 mm, and 6 mm.

Let us have a second look at the components of the eye one layer at a time: the cornea is a transparent, highly curved, refractive window through which light enters the eye before being partially blocked by the colored and opaque surface of the iris. The pupil is an opening at the center of the iris whose diameter varies from about 1 to 8 mm in response to illumination changes, dilating in low light to increase the amount of energy that reaches the retina and contracting in normal lighting conditions to limit the amount of image blurring due to spherical aberration in the eye. The refracting power (reciprocal of the focal length) of the eye is, in large part, an effect of refraction at the the air–cornea interface, and it is fine tuned by deformations of the crystalline lens that accommodates to bring objects into sharp focus. In healthy adults, it varies between 60 (unaccommodated case) and 68 diopters (1 diopter = 1 m^{-1}), corresponding to a range of focal lengths between 15 and 17 mm.

The retina itself is a thin, layered membrane populated by two types of photoreceptors—*rods* and *cones*. There are about 100 million rods and 5 million cones in a human eye. Their spatial distribution varies across the retina: The *macula lutea* is a region in the center of the retina where the concentration of cones is particularly high and images are sharply focused whenever the eye fixes its attention on an object (Figure 1.13). The highest concentration of cones occurs in the *fovea*, a depression in the middle of the macula lutea where it peaks at $1.6 \times 10^5/\mathrm{mm}^2$, with the centers of two neighboring cones separated by only half a minute of visual angle. Conversely, there are no rods in the center of the fovea, but the rod density increases toward the periphery of the visual field. There is also a *blind spot* on the retina, where the ganglion cell axons exit the retina and form the optic nerve.

The rods are extremely sensitive photoreceptors; they are capable of responding to a single photon, but they yield relatively poor spatial detail despite their high number because many rods converge to the same neuron within the retina. In contrast, cones become active at higher light levels, but the signal output by each cone in the fovea is encoded by several neurons, yielding a high resolution in that area. As discussed further in chapter **??**, there are three types of cones with different spectral sensitivities, and these play a key role in the perception of color. Much more could (and should) be said about the human eye—for example how our two eyes verge and fixate on targets, and cooperate in stereo vision, an issue briefly discussed in chapter **??**.

1.2 INTRINSIC AND EXTRINSIC PARAMETERS

Digital images, like animal retinas, are spatially discrete, and divided into (usually) rectangular picture elements, or *pixels*. This is an aspect of the image formation process that we have neglected so far, assuming instead that the image domain was spatially continuous. Likewise, the perspective equation derived in the previous section is only valid when all distances are measured in the camera's reference frame, and image coordinates have their origin at the image center where the axis of symmetry of the camera pierces its retina. In practice, the world and camera coordinate systems are related by a set of physical parameters, such as the focal length of the lens, the size of the pixels, the position of the image center, and the position and orientation of the camera. This section identifies these parameters. We will distinguish the *intrinsic* parameters, that relate the camera's coordinate system to the idealized coordinate system used in Section 1.1, from the *extrinsic* parameters, that relate the camera's coordinate system to a fixed world coordinate system and specify its position and orientation in space.

We ignore in the rest of this section the fact that, for cameras equipped with a lens, a point will only be in focus when its depth and the distance between the optical center of the camera and its image plane obey Eq. (1.5). In particular, we assume that the camera is focused at infinity so $d = f$. Likewise, the nonlinear aberrations associated with real lenses are not taken into account by Eq. (1.1). We neglect these aberrations in this section, but revisit radial distortion in Section 1.3 when we address the problem of estimating the intrinsic and extrinsic parameters of a camera (a process known as *geometric camera calibration*).

Notation. This section features our first use of *homogeneous* coordinates to represent the position of points in two and three dimensions. Consider a point P whose position in some coordinate frame $(F) = (O, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ is given by

$$\overrightarrow{OP} = X\boldsymbol{i} + Y\boldsymbol{j} + Z\boldsymbol{k}.$$

We define the usual (nonhomogeneous) coordinate vector of P as the element $(X, Y, Z)^T$ of $\mathbb{R}^3$, and its homogeneous coordinate vector as the element $(X, Y, Z, 1)^T$

of $\mathbb{R}^4$. We use bold letters to denote (homogeneous and nonhomogeneous) coordinate vectors in this book, and always state which type of coordinates we use when it is not obvious from the context. We also use a superscript *on the left side of* coordinate vectors to indicate when necessary which coordinate frame a position is expressed in. For example ${}^F\boldsymbol{P}$ stands for the coordinate vector of the point P in the frame (F). Homogeneous coordinates are a convenient device for representing various geometric transformations by matrix products. For example, as shown in chapter **??**, the change of coordinates between two Euclidean coordinate systems (A) and (B) may be represented by a 3×3 rotation matrix $\mathcal{R}$ and a translation vector $\boldsymbol{t}$ in $\mathbb{R}^3$, and the corresponding *rigid transformation* can be written in nonhomogeneous coordinates as

$${}^A\boldsymbol{P} = \mathcal{R}\,{}^B\boldsymbol{P} + \boldsymbol{t}$$

where ${}^A\boldsymbol{P}$ and ${}^B\boldsymbol{P}$ are elements of $\mathbb{R}^3$. In homogeneous coordinates, we write instead

$${}^A\boldsymbol{P} = \mathcal{T}\,{}^B\boldsymbol{P}, \quad \text{where} \quad \mathcal{T} = \begin{bmatrix} \mathcal{R} & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{bmatrix},$$

and ${}^A\boldsymbol{P}$ and ${}^B\boldsymbol{P}$ are this time elements of $\mathbb{R}^4$.

As will be shown in the rest of this section, homogeneous coordinates also provide an algrebraic representation of the perspective projection process in the form of a 3×4 matrix $\mathcal{M}$, so that the coordinate vector $\boldsymbol{P} = (X, Y, Z, 1)^T$ of a point P in some fixed world coordinate system and the coordinate vector $\boldsymbol{p} = (x, y, 1)^T$ of its image p in the camera's reference frame are related by the *perspective projection equation*

$$\boldsymbol{p} = \frac{1}{Z}\mathcal{M}\boldsymbol{P}. \tag{1.6}$$

1.2.1 Intrinsic Parameters

It is possible to associate with a camera a *normalized image plane* parallel to its physical retina but located at a unit distance from the pinhole. We attach to this plane its own coordinate system with an origin located at the point $\hat{c}$ where the optical axis pierces it (Figure 1.14). Equation (1.1) can be written in this normalized coordinate system as

$$\begin{cases} \hat{x} = \dfrac{X}{Z} \\ \hat{y} = \dfrac{Y}{Z} \end{cases} \Longleftrightarrow \hat{\boldsymbol{p}} = \frac{1}{Z}\begin{bmatrix} \mathrm{Id} & \boldsymbol{0} \end{bmatrix}\boldsymbol{P}, \tag{1.7}$$

where $\hat{\boldsymbol{p}} \overset{\text{def}}{=} (\hat{x}, \hat{y}, 1)^T$ is the vector of homogeneous coordinates of the projection $\hat{p}$ of the point P into the normalized image plane, and $\boldsymbol{P}$ is as before the homogeneous coordinate vector of P in the world coordinate frame.

The physical retina of the camera is in general different (Figure 1.14): It is located at a distance $f \neq 1$ from the pinhole (remember that we assume that the camera is focused at infinity so the distance between the pinhole and the image plane is equal to the focal length), and the image coordinates (x, y) of the image

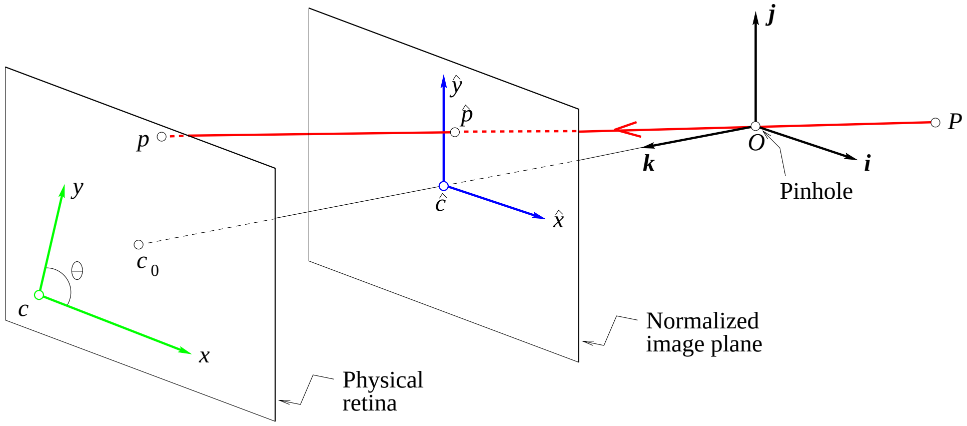

FIGURE 1.14: Physical and normalized image coordinate systems.

point p are usually expressed in pixel units (instead of, say, meters). In addition, pixels are normally rectangular instead of square, so the camera has two additional scale parameters k and l, and

$$\begin{cases} x = kf\dfrac{X}{Z} = kf\hat{x}, \\ y = lf\dfrac{Y}{Z} = lf\hat{y}. \end{cases} \tag{1.8}$$

Let us talk units for a second: f is a distance, expressed in meters for example, and a pixel will have dimensions $\frac{1}{k} \times \frac{1}{l}$, where k and l are expressed in pixel $\times\, \mathrm{m}^{-1}$. The parameters k, l and f are not independent, and they can be replaced by the magnifications $\alpha = kf$ and $\beta = lf$ expressed in pixel units.

Now, in general, the actual origin of the camera coordinate system is at a corner c of the retina (e.g., in the case depicted in Figure 1.14, the lower-left corner, or sometimes the upper-left corner, when the image coordinates are the row and column indices of a pixel) and not at its center, and the center of the CCD matrix usually does not coincide with the image center c_0. This adds two parameters x_0 and y_0 that define the position (in pixel units) of c_0 in the retinal coordinate system. Thus, Eq. (1.8) is replaced by

$$\begin{cases} x = \alpha\hat{x} + x_0, \\ y = \beta\hat{y} + y_0. \end{cases} \tag{1.9}$$

Finally, the camera coordinate system may also be skewed, due to some manufacturing error, so the angle θ between the two image axes is not equal to (but of course not very different from either) 90 degrees. In this case, it is easy to show that Eq. (1.9) transforms into

$$\begin{cases} x = \alpha\hat{x} - \alpha\cot\theta\hat{y} + x_0, \\ y = \dfrac{\beta}{\sin\theta}\hat{y} + y_0. \end{cases} \tag{1.10}$$

This can be written in matrix form as

$$\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}, \quad \text{where} \quad \boldsymbol{p} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \text{and} \quad \mathcal{K} \overset{\text{def}}{=} \begin{bmatrix} \alpha & -\alpha\cot\theta & x_0 \\ 0 & \dfrac{\beta}{\sin\theta} & y_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1.11}$$

The 3×3 matrix $\mathcal{K}$ is called the (internal) *calibration matrix* of the camera. Putting Eqs. (1.7) and (1.11) together, we obtain

$$\boldsymbol{p} = \frac{1}{Z}\mathcal{K}\begin{bmatrix}\text{Id} & \mathbf{0}\end{bmatrix}\boldsymbol{P} = \frac{1}{Z}\mathcal{M}\boldsymbol{P}, \quad \text{where} \quad \mathcal{M} \overset{\text{def}}{=} \begin{bmatrix}\mathcal{K} & \mathbf{0}\end{bmatrix} \tag{1.12}$$

which is indeed an instance of Eq. (1.6). The five parameters α, β, θ, x_0, and y_0 are called the *intrinsic parameters* of the camera.

Note that the physical size of the pixels and the skew are always fixed for a given camera, and they can in principle be measured during manufacturing (this information may of course not be available, in the case of stock film footage for example). For zoom lenses, the focal length may vary with time, along with the image center when the optical axis of the lens is not exactly perpendicular to the image plane. Simply changing the focus of the camera will also affect the magnification since it will change the lens-to-retina distance, but we will continue to assume that the camera is focused at infinity and ignore this effect in the rest of this chapter.

1.2.2 Extrinsic Parameters

Equation (1.12) is written in a coordinate frame (C) attached to the camera. Let us now consider the case where this frame is distinct from the world coordinate system (W). To emphasize this, we rewrite Eq. (1.12) as $\boldsymbol{p} = \frac{1}{Z}\mathcal{M}{}^C\boldsymbol{P}$, where ${}^C\boldsymbol{P}$ denotes the vector of homogeneous coordinates of the point P expressed in (C). The change of coordinates between (C) and (W) is a rigid transformation, and it can be written as

$${}^C\boldsymbol{P} = \begin{bmatrix} \mathcal{R} & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix} {}^W\boldsymbol{P},$$

where ${}^W\boldsymbol{P}$ is the vector of homogeneous coordinates of the point P in the coordinate frame (W). Taking $\boldsymbol{P} = {}^W\boldsymbol{P}$ and substituting in Eq. (1.12) finally yields

$$\boldsymbol{p} = \frac{1}{Z}\mathcal{M}\boldsymbol{P}, \quad \text{where} \quad \mathcal{M} = \mathcal{K}\begin{bmatrix}\mathcal{R} & \boldsymbol{t}\end{bmatrix}. \tag{1.13}$$

This is the most general form of the perspective projection equation, and indeed an instance of Eq. (1.6). The rotation matrix $\mathcal{R}$ is defined by three independent parameters (for example three *Euler angles*, see chapter **??**). Adding to these the three coordinates of the vector $\boldsymbol{t}$, we obtain a set of six *extrinsic parameters* that define the position and orientation of the camera relative to the world coordinate frame.

It is very important to understand that the depth Z in Eq. (1.13) is *not* independent of $\mathcal{M}$ and $\boldsymbol{P}$ since, if $\boldsymbol{m}_1^T$, $\boldsymbol{m}_2^T$ and $\boldsymbol{m}_3^T$ denote the three rows of $\mathcal{M}$, it follows directly from Eq. (1.13) that $Z = \boldsymbol{m}_3 \cdot \boldsymbol{P}$. In fact, it is sometimes convenient to rewrite Eq. (1.13) in the equivalent form:

$$\begin{cases} x = \dfrac{\boldsymbol{m}_1 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}}, \\ y = \dfrac{\boldsymbol{m}_2 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}}. \end{cases} \tag{1.14}$$

A perspective projection matrix can be written explicitly as a function of its five intrinsic parameters, the three rows $\boldsymbol{r}_1^T$, $\boldsymbol{r}_2^T$, and $\boldsymbol{r}_3^T$ of the matrix $\mathcal{R}$, and the three coordinates t_1, t_2, and t_3 of the vector $\boldsymbol{t}$, namely:

$$\mathcal{M} = \begin{bmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + x_0 \boldsymbol{r}_3^T & \alpha t_1 - \alpha \cot\theta t_2 + x_0 t_3 \\ \dfrac{\beta}{\sin\theta} \boldsymbol{r}_2^T + y_0 \boldsymbol{r}_3^T & \dfrac{\beta}{\sin\theta} t_2 + y_0 t_3 \\ \boldsymbol{r}_3^T & t_3 \end{bmatrix}. \tag{1.15}$$

When $\mathcal{R}$ is written as the product of three elementary rotations, the vectors $\boldsymbol{r}_i$ $(i = 1, 2, 3)$ can of course be written in terms of the corresponding three angles, and Eq. (1.15) gives an explicit parameterization of $\mathcal{M}$ in terms of all 11 camera parameters.

1.2.3 Perspective Projection Matrices

This section examine the conditions under which a 3×4 matrix $\mathcal{M}$ can be written in the form given by Eq. (1.15). Let us write without loss of generality $\mathcal{M} = \begin{bmatrix} \mathcal{A} & \boldsymbol{b} \end{bmatrix}$, where $\mathcal{A}$ is a 3×3 matrix and $\boldsymbol{b}$ is an element of $\mathbb{R}^3$, and let us denote by $\boldsymbol{a}_3^T$ the third row of $\mathcal{A}$. Clearly, if $\mathcal{M}$ is an instance of Eq. (1.15), then $\boldsymbol{a}_3^T$ must be a unit vector since it is equal to $\boldsymbol{r}_3^T$, the last row of a rotation matrix. Note, however, that replacing $\mathcal{M}$ by $\lambda\mathcal{M}$ in Eq. (1.14) for some arbitrary $\lambda \neq 0$ does not change the corresponding image coordinates. This will lead us in the rest of this book to consider projection matrices as *homogeneous objects*, only defined up to scale, whose canonical form, as expressed by Eq. (1.15), can be obtained by choosing a scale factor such that $|\boldsymbol{a}_3| = 1$. Note that the parameter Z in Eq. (1.13) can only rightly be interpreted as the depth of the point P when $\mathcal{M}$ is written in this canonical form. Note also that the number of intrinsic and extrinsic parameters of a camera matches the 11 free parameters of the (homogeneous) matrix $\mathcal{M}$.

We say that a 3×4 matrix that can be written (up to scale) as Eq. (1.15) for some set of intrinsic and extrinsic parameters is a *perspective projection matrix*. It is of practical interest to put some restrictions on the intrinsic parameters of a camera since, as noted earlier, some of these parameters will be fixed and may be known. In particular, we will say that a 3×4 matrix is a *zero-skew perspective projection matrix* when it can be rewritten (up to scale) as Eq. (1.15) with $\theta = \pi/2$, and that it is a *perspective projection matrix with zero skew and unit aspect-ratio* when it can be rewritten (up to scale) as Eq. (1.15) with $\theta = \pi/2$ and $\alpha = \beta$. A camera with *known* nonzero skew and nonunit aspect-ratio can be transformed into a camera with zero skew and unit aspect-ratio by an appropriate change of image coordinates. Are arbitrary 3 × 4 matrices perspective projection matrices? The following theorem answers this question.

Theorem 1. Let $\mathcal{M} = \begin{bmatrix}\mathcal{A} & \boldsymbol{b}\end{bmatrix}$ be a 3 × 4 matrix and let $\boldsymbol{a}_i^T$ $(i = 1, 2, 3)$ denote the rows of the matrix $\mathcal{A}$ formed by the three leftmost columns of $\mathcal{M}$.

- A necessary and sufficient condition for $\mathcal{M}$ to be a perspective projection matrix is that $\text{Det}(\mathcal{A}) \neq 0$.
- A necessary and sufficient condition for $\mathcal{M}$ to be a zero-skew perspective projection matrix is that $\text{Det}(\mathcal{A}) \neq 0$ and
$$(\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3) = 0.$$
- A necessary and sufficient condition for $\mathcal{M}$ to be a perspective projection matrix with zero skew and unit aspect-ratio is that $\text{Det}(\mathcal{A}) \neq 0$ and
$$\begin{cases} (\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3) = 0, \\ (\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_1 \times \boldsymbol{a}_3) = (\boldsymbol{a}_2 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3). \end{cases}$$

The conditions of the theorem are clearly necessary: By definition, given some perspective projection matrix $\mathcal{A}$, we can always write $\rho\mathcal{A} = \mathcal{K}\mathcal{R}$ for some nonzero scalar ρ, calibration matrix $\mathcal{K}$, rotation matrix $\mathcal{R}$, and vector $\boldsymbol{t}$. In particular, $\rho^3\text{Det}(\mathcal{A}) = \text{Det}(\mathcal{K}) \neq 0$ since calibration matrices are nonsingular by construction, so $\mathcal{A}$ is nonsingular. Further, a simple calculation shows that the rows of the matrix $\frac{1}{\rho}\mathcal{K}\mathcal{R}$ satisfy the conditions of the theorem under the various assumptions imposed by its statement. These conditions are proven to also be sufficient in Faugeras (1993).

1.2.4 Weak-Perspective Projection Matrices

As noted in Section 1.1.2, when a scene's relief is small compared to the overall distance separating it from the camera observing it, a weak-perspective projection model can be used to approximate the imaging process (Figure 1.15, top): Let O denote the optical center of the camera and let R denote a scene reference point; the weak-perspective projection of a scene point P is constructed in two steps: The point P is first projected orthogonally onto a point P' of the plane Π_r parallel to the image plane Π and passing through R; perspective projection is then used to map the point P' onto the image point p. Since π_r is a fronto-parallel plane, the net effect of the second projection step is a scaling of the image coordinates.

As shown in this section, the weak-perspective projection process can be represented in terms of a 2 × 4 matrix $\mathcal{M}$, so that the *homogeneous* coordinate vector $\boldsymbol{P} = (X, Y, Z, 1)^T$ of a point P in some fixed world coordinate system and the *non-homogeneous* coordinate vector $\boldsymbol{p} = (x, y)^T$ of its image p in the camera's reference frame are related by the *affine projection equation*

$$\boldsymbol{p} = \mathcal{M}\boldsymbol{P}. \tag{1.16}$$

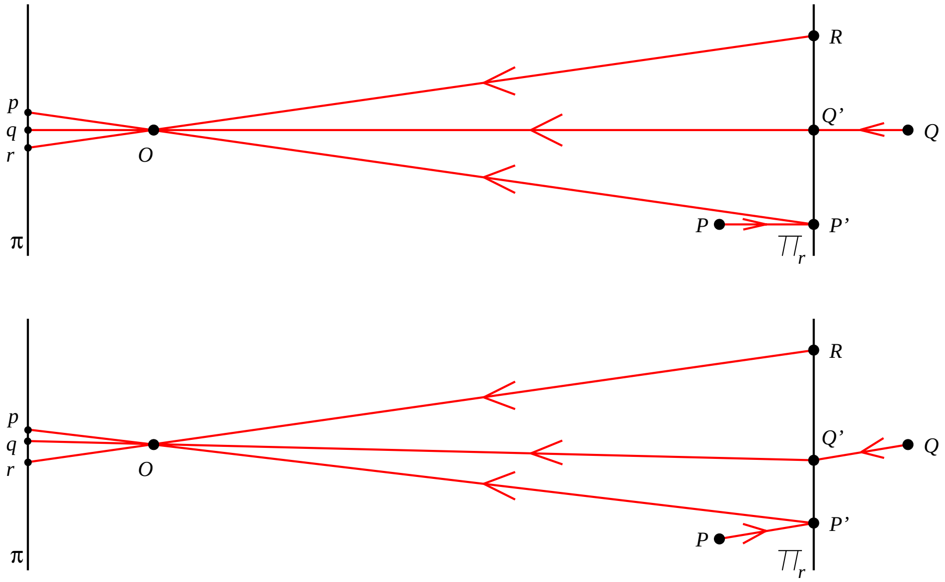

FIGURE 1.15: Affine projection models: (top) weak-perspective and (bottom) paraperspective projections.

It turns out that this general model accomodates various other approximations of the perspective projection process. These include the orthographic projection model discusses earlier, as well as the *parallel projection* model, that subsumes the orthographic one, and takes into account the fact that the objects of interest may lie off the optical axis of the camera. In this model, the viewing rays are parallel to each other but not necessarily perpendicular to the image plane. Paraperspective is another affine projection model that takes into account both the distortions associated with a reference point that is off the optical axis of the camera and possible variations in depth (Figure 1.15, bottom): Using the same notation as before, and denoting by Δ the line joining the optical center O to the reference point R, parallel projection in the direction of Δ is first used to map P onto a point P' of the plane Π_r; perspective projection is then used to map the point P' onto the image point p.

We will focus on weak perspective in the rest of this section. Let us derive the corresponding projection equation. If Z_r denotes the depth of the reference point R, the two elementary projection stages $P \to P' \to p$ can be written in the normalized coordinate system attached to the camera as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \longrightarrow \begin{bmatrix} Z \\ Y \\ Z_r \end{bmatrix} \longrightarrow \begin{bmatrix} \hat{x} \\ \hat{y} \\ 1 \end{bmatrix} = \begin{bmatrix} X/Z_r \\ Y/Z_r \\ 1 \end{bmatrix},$$

or, in matrix form,

$$\begin{bmatrix}\hat{x}\\ \hat{y}\\ 1\end{bmatrix} = \frac{1}{Z_r}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & Z_r\end{bmatrix}\begin{bmatrix}X\\ Y\\ Z\\ 1\end{bmatrix}.$$

Introducing the calibration matrix $\mathcal{K}$ of the camera and its extrinsic parameters $\mathcal{R}$ and $\boldsymbol{t}$ gives the general form of the projection equation, i.e.,

$$\boldsymbol{p} = \frac{1}{Z_r}\mathcal{K}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & Z_r\end{bmatrix}\begin{bmatrix}\mathcal{R} & \boldsymbol{t}\\ \boldsymbol{0}^T & 1\end{bmatrix}\boldsymbol{P}, \tag{1.17}$$

where $\boldsymbol{P}$ and $\boldsymbol{p}$ denote as before the homogeneous coordinate vector of the point P in the world reference frame, and the homogeneous coordinate vector of its projection p in the camera's coordinate system. Finally, noting that Z_r is a constant and writing

$$\mathcal{K} = \begin{bmatrix}\mathcal{K}_2 & \boldsymbol{p}_0\\ \boldsymbol{0}^T & 1\end{bmatrix}, \quad \text{where} \quad \mathcal{K}_2 \overset{\text{def}}{=} \begin{bmatrix}\alpha & -\alpha\cot\theta\\ 0 & \dfrac{\beta}{\sin\theta}\end{bmatrix} \quad \text{and} \quad \boldsymbol{p}_0 \overset{\text{def}}{=} \begin{bmatrix}x_0\\ y_0\end{bmatrix},$$

allows us to rewrite Eq. (1.17) as

$$\boldsymbol{p} = \mathcal{M}\boldsymbol{P}, \quad \text{where} \quad \mathcal{M} = \begin{bmatrix}\mathcal{A} & \boldsymbol{b}\end{bmatrix}, \tag{1.18}$$

where $\boldsymbol{p}$ is, this time, the *nonhomogeneous* coordinate vector of the point p, and $\mathcal{M}$ is a 2×4 projection matrix (compare to the general perspective case of Eq. [1.13]). In this expression, the 2×3 matrix $\mathcal{A}$ and the 2-vector $\boldsymbol{b}$ are respectively defined by

$$\mathcal{A} = \frac{1}{Z_r}\mathcal{K}_2\mathcal{R}_2 \quad \text{and} \quad \boldsymbol{b} = \frac{1}{Z_r}\mathcal{K}_2\boldsymbol{t}_2 + \boldsymbol{p}_0,$$

where $\mathcal{R}_2$ denotes the 2×3 matrix formed by the first two rows of $\mathcal{R}$ and $\boldsymbol{t}_2$ denotes the 2-vector formed by the first two coordinates of $\boldsymbol{t}$.

Note that t_3 does not appear in the expression of $\mathcal{M}$, and that $\boldsymbol{t}_2$ and $\boldsymbol{p}_0$ are coupled in this expression: The projection matrix does not change when $\boldsymbol{t}_2$ is replaced by $\boldsymbol{t}_2 + \boldsymbol{a}$ and $\boldsymbol{p}_0$ is replaced by $\boldsymbol{p}_0 - \frac{1}{Z_r}\mathcal{K}_2\boldsymbol{a}$. This redundancy allows us to arbitrarily choose $x_0 = y_0 = 0$. In other words, the position of the center of the image is immaterial for weak-perspective projection. Note that the values of Z_r, α and β are also coupled in the expression of $\mathcal{M}$, and that the value of Z_r is a priori unknown in most applications. This allows us to write

$$\mathcal{M} = \frac{1}{Z_r}\begin{bmatrix}k & s\\ 0 & 1\end{bmatrix}\begin{bmatrix}\mathcal{R}_2 & \boldsymbol{t}_2\end{bmatrix}, \tag{1.19}$$

where k and s denote respectively the aspect ratio and the skew of the camera. In particular, a weak-perspective projection matrix is defined by two intrinsic parameters (k and s), five extrinsic parameters (the three angles defining $\mathcal{R}_2$ and the two coordinates of $\boldsymbol{t}_2$), and one scene-dependent *structure* parameter Z_r.

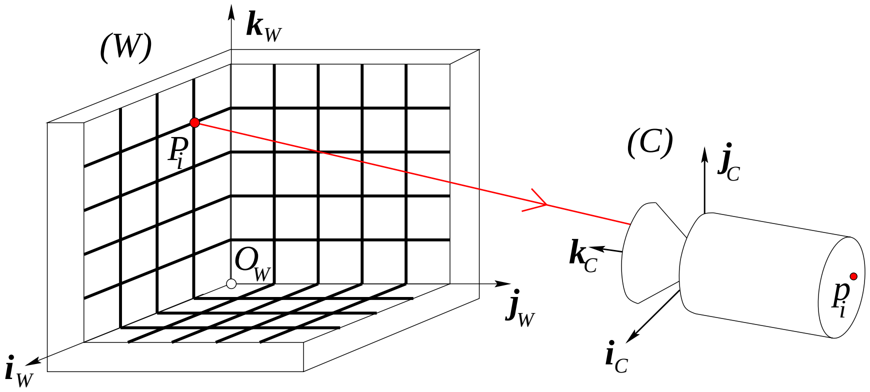

FIGURE 1.16: Camera calibration setup: In this example, the calibration rig is formed by three grids drawn in orthogonal planes. Other patterns could be used as well, and they may involve lines or other geometric figures.

A 2×4 matrix $\mathcal{M} = \begin{bmatrix} \mathcal{A} & \boldsymbol{b} \end{bmatrix}$ where $\mathcal{A}$ is an arbitrary rank-2 2×3 matrix and $\boldsymbol{b}$ is an arbitrary vector in $\mathbb{R}^2$ is called an *affine projection matrix*. The rank condition follows from the fact that a rank-1 matrix would project all scene points onto a single image line; note also that the matrix $\mathcal{A}$ associated with weak-perspective cameras has rank 2 by construction since, according to Eq. (1.19), it can be written as the product of rank-2 matrices. Both weak-perspective and general affine projection matrices are defined by 8 independent parameters. Weak-perspective matrices are, of course, affine ones. Conversely, a simple parameter-counting argument suggests that it should be possible to write an arbitrary affine projection matrix as a weak-perspective one. This is formally shown to be true in Faugeras *et al.* (2001) and the exercises.

1.3 GEOMETRIC CAMERA CALIBRATION

This section addresses the problem of estimating the intrinsic and extrinsic parameters of a camera from the image positions of scene features such as points of lines, whose positions are known in some fixed world coordinate system (Figure 1.16): In this context, camera calibration can be modeled as an optimization process, where the discrepancy between the observed image features and their theoretical positions is minimized with respect to the camera's intrinsic and extrinsic parameters.

Specifically, we assume that the image positions (x_i, y_i) of n fiducial points P_i $(i = 1, \ldots, n)$ with known homogeneous coordinate vectors $\boldsymbol{P}_i$ have been found in a picture of a calibration rig, either automatically or by hand. In the absence of modeling and measurement errors, geometric camera calibration amounts to finding the intrinsic and extrinsic parameters $\boldsymbol{\xi}$ such that

$$\begin{cases} x_i = \dfrac{\boldsymbol{m}_1(\boldsymbol{\xi}) \cdot \boldsymbol{P}_i}{\boldsymbol{m}_3(\boldsymbol{\xi}) \cdot \boldsymbol{P}_i}, \\ y_i = \dfrac{\boldsymbol{m}_2(\boldsymbol{\xi}) \cdot \boldsymbol{P}_i}{\boldsymbol{m}_3(\boldsymbol{\xi}) \cdot \boldsymbol{P}_i}, \end{cases} \tag{1.20}$$

where $\boldsymbol{m}_i^T(\boldsymbol{\xi})$ denotes the i^{th} row of the projection matrix $\mathcal{M}$, explicitly parameterized in this equation by the camera parameters. In the typical case where there are more measurements than unknowns (at least six points for 11 intrinsic and extrinsic parameters), Eq. (1.20) does not admit an exact solution, and an approximate one has to be found as the solution of a *least-squares* minimization problem (see chapter **??**). We present two least-squares formulations of the calibration problem in the rest of this section. The corresponding algorithms are illustrated with the calibration data shown in Figure 1.17.

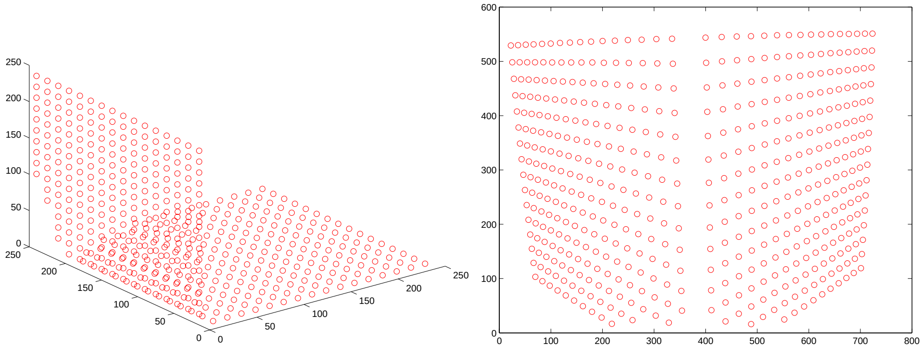

FIGURE 1.17: Camera calibration data. Left: A rendering of 491 3D fiducial points measured on a calibration rig. Right: The corresponding image points. Data courtesy of Janne Heikkila. Copyright ©2000 University of Oulu.

1.3.1 A Linear Approach to Camera Calibration

We decompose the calibration process into (1) the computation of the perspective projection matrix $\mathcal{M}$ associated with the camera, followed by (2) the estimation of the intrinsic and extrinsic parameters of the camera from this matrix.

Estimation of the Projection Matrix. Let us assume that our camera has nonzero skew. According to Theorem 1, the matrix $\mathcal{M}$ is not singular, but otherwise arbitrary. Clearing the denominators in Eq. (1.20) yields two *linear* equations in $\boldsymbol{m}_1$, $\boldsymbol{m}_2$ and $\boldsymbol{m}_3$ (we omit the parameters $\boldsymbol{\xi}$ from now on for the sake of conciseness), namely

$$\left\{\begin{array}{lcccc} (\boldsymbol{m}_1 - x_i\boldsymbol{m}_3)\cdot\boldsymbol{P}_i & = & \boldsymbol{P}_i^T\boldsymbol{m}_1 + \boldsymbol{0}^T\boldsymbol{m}_2 - x_i\boldsymbol{P}_i^T\boldsymbol{m}_3 & = & 0, \\ (\boldsymbol{m}_2 - y_i\boldsymbol{m}_3)\cdot\boldsymbol{P}_i & = & \boldsymbol{0}^T\boldsymbol{m}_1 + \boldsymbol{P}_i^T\boldsymbol{m}_2 - y_i\boldsymbol{P}_i^T\boldsymbol{m}_3 & = & 0. \end{array}\right.$$

Collecting the constraints associated with all points yields a system of $2n$ homogeneous linear equations in the twelve coefficients of the matrix $\mathcal{M}$, namely,

$$\mathcal{P}\boldsymbol{m} = 0, \tag{1.21}$$

where

$$\mathcal{P} \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{P}_1^T & \boldsymbol{0}^T & -x_1\boldsymbol{P}_1^T \\ \boldsymbol{0}^T & \boldsymbol{P}_1^T & -y_1\boldsymbol{P}_1^T \\ \ldots & \ldots & \ldots \\ \boldsymbol{P}_n^T & \boldsymbol{0}^T & -x_n\boldsymbol{P}_n^T \\ \boldsymbol{0}^T & \boldsymbol{P}_n^T & -y_n\boldsymbol{P}_n^T \end{bmatrix} \quad \text{and} \quad \boldsymbol{m} \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \boldsymbol{m}_3 \end{bmatrix} = 0.$$

When $n \geq 6$, homogeneous linear least-squares can be used to compute the value of the unit vector $\boldsymbol{m}$ (hence the matrix $\mathcal{M}$) that minimizes $|\mathcal{P}\boldsymbol{m}|^2$ as the eigenvector of the 12×12 matrix $\mathcal{P}^T\mathcal{P}$ associated with its smallest eigenvalue (see chapter **??**). Note the fact that any nonzero multiple of the vector $\boldsymbol{m}$ would have done just as well, reflecting the fact that $\mathcal{M}$ is only defined by 11 independent parameters).

Degenerate Point Configurations. Before showing how to recover the intrinsic and extrinsic parameters of the camera, let us pause to examine the *degenerate configurations* of the points P_i $(i = 1, \ldots, n)$ that may cause the failure of the camera calibration process. We focus on the (ideal) case where the positions $\boldsymbol{p}_i$ $(i = 1, \ldots, n)$ of the image points can be measured with zero error, and identify the *nullspace* of the matrix $\mathcal{P}$ (i.e., the subspace of $\mathbb{R}^{12}$ formed by the vectors $\boldsymbol{l}$ such that $\mathcal{P}\boldsymbol{l} = \boldsymbol{0}$).

Let $\boldsymbol{l}$ be such a vector. Introducing the vectors formed by successive quadruples of its coordinates—that is, $\boldsymbol{\lambda} = (l_1, l_2, l_3, l_4)^T$, $\boldsymbol{\mu} = (l_5, l_6, l_7, l_8)^T$, and $\boldsymbol{\nu} = (l_9, l_{10}, l_{11}, l_{12})^T$ allows us to write

$$\boldsymbol{0} = \mathcal{P}\boldsymbol{l} = \begin{bmatrix} \boldsymbol{P}_1^T & \boldsymbol{0}^T & -x_1\boldsymbol{P}_1^T \\ \boldsymbol{0}^T & \boldsymbol{P}_1^T & -y_1\boldsymbol{P}_1^T \\ \ldots & \ldots & \ldots \\ \boldsymbol{P}_n^T & \boldsymbol{0}^T & -x_n\boldsymbol{P}_n^T \\ \boldsymbol{0}^T & \boldsymbol{P}_n^T & -y_n\boldsymbol{P}_n^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_1^T\boldsymbol{\lambda} - x_1\boldsymbol{P}_1^T\boldsymbol{\nu} \\ \boldsymbol{P}_1^T\boldsymbol{\mu} - y_1\boldsymbol{P}_1^T\boldsymbol{\nu} \\ \ldots \\ \boldsymbol{P}_n^T\boldsymbol{\lambda} - x_n\boldsymbol{P}_n^T\boldsymbol{\nu} \\ \boldsymbol{P}_n^T\boldsymbol{\mu} - y_n\boldsymbol{P}_n^T\boldsymbol{\nu} \end{bmatrix}. \tag{1.22}$$

Combining Eq. (1.20) with Eq. (1.22) yields

$$\begin{cases} \boldsymbol{P}_i^T\boldsymbol{\lambda} - \dfrac{\boldsymbol{m}_1^T\boldsymbol{P}_i}{\boldsymbol{m}_3^T\boldsymbol{P}_i}\boldsymbol{P}_i^T\boldsymbol{\nu} = 0, \\ \boldsymbol{P}_i^T\boldsymbol{\mu} - \dfrac{\boldsymbol{m}_2^T\boldsymbol{P}_i}{\boldsymbol{m}_3^T\boldsymbol{P}_i}\boldsymbol{P}_i^T\boldsymbol{\nu} = 0, \end{cases} \quad \text{for} \quad i = 1, \ldots, n.$$

Thus, we finally obtain after clearing the denominators and rearranging the terms:

$$\begin{cases} \boldsymbol{P}_i^T(\boldsymbol{\lambda}\boldsymbol{m}_3^T - \boldsymbol{m}_1\boldsymbol{\nu}^T)\boldsymbol{P}_i = 0, \\ \boldsymbol{P}_i^T(\boldsymbol{\mu}\boldsymbol{m}_3^T - \boldsymbol{m}_2\boldsymbol{\nu}^T)\boldsymbol{P}_i = 0, \end{cases} \quad \text{for} \quad i = 1, \ldots, n. \tag{1.23}$$

As expected, the vector $\boldsymbol{l}$ associated with $\boldsymbol{\lambda} = \boldsymbol{m}_1$, $\boldsymbol{\mu} = \boldsymbol{m}_2$ and $\boldsymbol{\nu} = \boldsymbol{m}_3$ is a solution of these equations. Are there other solutions?

Let us first consider the case where the points P_i $(i = 1, \ldots, n)$ all lie in some plane Π, so $\boldsymbol{P}_i \cdot \boldsymbol{\Pi} = 0$ for some 4-vector $\boldsymbol{\Pi}$. Clearly, choosing $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$ equal to $(\boldsymbol{\Pi}, \boldsymbol{0}, \boldsymbol{0})$, $(\boldsymbol{0}, \boldsymbol{\Pi}, \boldsymbol{0})$, or $(\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{\Pi})$, or any linear combination of these vectors will yield a solution of Eq. (1.23). In other words, the nullspace of $\mathcal{P}$ contains the four-dimensional vector space spanned by these vectors and $\boldsymbol{m}$. In practice, this means that the fiducial points P_i should not all lie in the same plane.

In general, for a given nonzero value of the vector $\boldsymbol{l}$, the points P_i that satisfy Eq. (1.23) must lie on the curve where the two quadric surfaces defined by the corresponding equations intersect. A closer look at Eq. (1.23) reveals that the straight line where the planes defined by $\boldsymbol{m}_3 \cdot \boldsymbol{P} = 0$ and $\boldsymbol{\nu} \cdot \boldsymbol{P} = 0$ intersect lies on both quadrics. It can be shown that the intersection curve of these two surfaces consists of this line and of a *twisted cubic* curve Γ passing through the origin. A twisted cubic is entirely determined by six points lying on it, and it follows that seven points chosen at random will not fall on Γ. Since, in addition, this curve passes through the origin, choosing $n \geq 6$ random points will in general guarantee that the matrix $\mathcal{P}$ has rank 11 and that the projection matrix can be recovered in a unique fashion.

Estimation of the Intrinsic and Extrinsic Parameters. Once the projection matrix $\mathcal{M}$ has been estimated, its expression in terms of the camera intrinsic and extrinsic parameters (Eq. [1.15]) can be used to recover these parameters as follows: We write as before $\mathcal{M} = \begin{bmatrix} \mathcal{A} & \boldsymbol{b} \end{bmatrix}$ with $\boldsymbol{a}_1^T$, $\boldsymbol{a}_2^T$ and $\boldsymbol{a}_3^T$ denoting the rows of $\mathcal{A}$, and obtain

$$\rho \begin{bmatrix} \mathcal{A} & \boldsymbol{b} \end{bmatrix} = \mathcal{K} \begin{bmatrix} \mathcal{R} & \boldsymbol{t} \end{bmatrix} \Longleftrightarrow \rho \begin{bmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \boldsymbol{a}_3^T \end{bmatrix} = \begin{bmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + x_0 \boldsymbol{r}_3^T \\ \dfrac{\beta}{\sin\theta} \boldsymbol{r}_2^T + y_0 \boldsymbol{r}_3^T \\ \boldsymbol{r}_3^T \end{bmatrix},$$

where ρ is an unknown scale factor, introduced here to account for the fact that the recovered matrix $\mathcal{M}$ has unit Frobenius form since $|\mathcal{M}| = |\boldsymbol{m}| = 1$.

In particular, using the fact that the rows of a rotation matrix have unit length and are perpendicular to each other yields immediately

$$\begin{cases} \rho = \varepsilon / |\boldsymbol{a}_3|, \\ \boldsymbol{r}_3 = \rho \boldsymbol{a}_3, \\ x_0 = \rho^2 (\boldsymbol{a}_1 \cdot \boldsymbol{a}_3), \\ y_0 = \rho^2 (\boldsymbol{a}_2 \cdot \boldsymbol{a}_3), \end{cases} \tag{1.24}$$

where $\varepsilon = \mp 1$.

Since θ is always in the neighborhood of $\pi/2$ with a positive sine, we have

$$\begin{cases} \rho^2 (\boldsymbol{a}_1 \times \boldsymbol{a}_3) = -\alpha \boldsymbol{r}_2 - \alpha \cot\theta \boldsymbol{r}_1, \\ \rho^2 (\boldsymbol{a}_2 \times \boldsymbol{a}_3) = \dfrac{\beta}{\sin\theta} \boldsymbol{r}_1, \end{cases} \quad \text{and} \quad \begin{cases} \rho^2 |\boldsymbol{a}_1 \times \boldsymbol{a}_3| = \dfrac{|\alpha|}{\sin\theta}, \\ \rho^2 |\boldsymbol{a}_2 \times \boldsymbol{a}_3| = \dfrac{|\beta|}{\sin\theta}, \end{cases} \tag{1.25}$$

thus:

$$\begin{cases} \cos\theta = -\dfrac{(\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)}{|\boldsymbol{a}_1 \times \boldsymbol{a}_3| |\boldsymbol{a}_2 \times \boldsymbol{a}_3|}, \\ \alpha = \rho^2 |\boldsymbol{a}_1 \times \boldsymbol{a}_3| \sin\theta, \\ \beta = \rho^2 |\boldsymbol{a}_2 \times \boldsymbol{a}_3| \sin\theta, \end{cases} \tag{1.26}$$

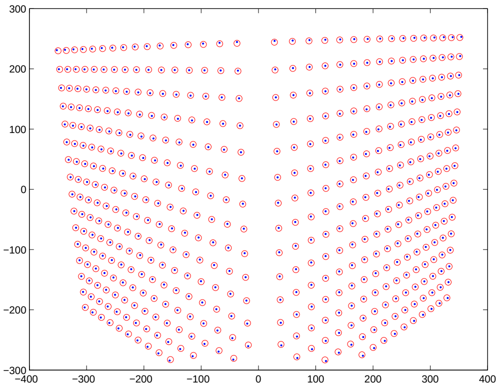

FIGURE 1.18: Results of camera calibration on the dataset shown in Figure 1.17. The original data points (circles) are overlaid with the reprojected 3D points (dots). The root-mean-squared image error is less than a pixel.

since the sign of the magnification parameters α and β is normally known in advance and can be taken to be positive.

We can now compute $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ from the second equation in Eq. (1.25) as

$$\begin{cases} \boldsymbol{r}_1 = \dfrac{\rho^2 \sin\theta}{\beta}(\boldsymbol{a}_2 \times \boldsymbol{a}_3) = \dfrac{1}{|\boldsymbol{a}_2 \times \boldsymbol{a}_3|}(\boldsymbol{a}_2 \times \boldsymbol{a}_3), \\ \boldsymbol{r}_2 = \boldsymbol{r}_3 \times \boldsymbol{r}_1. \end{cases} \tag{1.27}$$

Note that there are two possible choices for the matrix $\mathcal{R}$ depending on the value of ε. The translation parameters can now be recovered by writing $\mathcal{K}\boldsymbol{t} = \rho\boldsymbol{b}$ and hence $\boldsymbol{t} = \rho\mathcal{K}^{-1}\boldsymbol{b}$. In practical situations, the sign of t_3 is often known in advance (this corresponds to knowing whether the origin of the world coordinate system is in front or behind the camera), which allows the choice of a unique solution for the calibration parameters.

Figure 1.18 shows the results of an experiment with the dataset from Figure 1.17.

Bibliography

Faugeras, O. (1993), *Three-Dimensional Computer Vision*, MIT Press.

Faugeras, O., Luong, Q.-T. & Papadopoulo, T. (2001), *The Geometry of Multiple Images*, MIT Press.

Forsyth, D. & Ponce, J. (2003), *Computer Vision: A Modern Approach*, Prentice-Hall.

Navy, U. (1969), *Basic Optics and Optical Instruments*, Dover. Prepared by the Bureau of Naval Personnel.

Thompson, M., Eller, R., Radlinski, W. & Speert, J., eds (1966), *Manual of Photogrammetry*, American Society of Photogrammetry. Third Edition.

Wandell, B. (1995), *Foundations of Vision*, Sinauer Associates, Inc., Sunderland, MA.