Computer Vision
A Modern Approach

David A. Forsyth
University of California at Berkeley

Jean Ponce
University of Illinois at Urbana-Champaign

An Alan R. Apt Book

Prentice Hall
Upper Saddle River, New Jersey 07458

2

Geometric Camera Models

The fundamental laws of perspective projection were introduced in Chapter 1 in a camera-centered coordinate system. This chapter introduces the analytical machinery necessary to establish quantitative constraints between image measurements and the position and orientation of geometric figures measured in some arbitrary *external* coordinate system. We start by briefly recalling elementary notions of analytical Euclidean geometry, including homogeneous coordinates and matrix representations of geometric transformations. We then introduce the various physical parameters (the so-called *intrinsic* and *extrinsic* parameters) that relate the world and the camera coordinate frames and derive the general form of the perspective projection equation in this setting. We conclude with a brief presentation of *affine* models of the imaging process, that approximate pinhole perspective projection for distant objects, and include the orthographic and weak-perspective models briefly discussed in Chapter 1.

2.1 ELEMENTS OF ANALYTICAL EUCLIDEAN GEOMETRY

We assume that the reader has some familiarity with elementary Euclidean geometry and linear algebra. This section discusses useful analytical concepts such as coordinate systems, homogeneous coordinates, rotation matrices, and the like.

2.1.1 Coordinate Systems and Homogeneous Coordinates

We already used three-dimensional coordinate systems in chapter 1. This section introduces them a bit more formally. We assume throughout a fixed system of units, say meters or inches, so unit length is well defined.

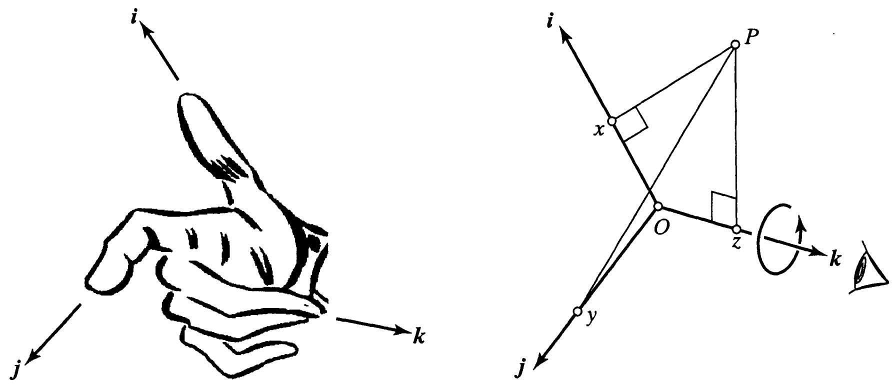

Figure 2.1 A right-handed coordinate system and the coordinates x, y, and z of a point P.

Picking a point O in the physical three-dimensional Euclidean space $\mathbb{E}^3$ and three unit vectors $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ orthogonal to each other defines an *orthonormal coordinate frame* (F) as the quadruple $(O, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$. The point O is the *origin* of the coordinate system (F), and $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are its *basis vectors*. We restrict our attention to *right-handed* coordinate systems, such that the vectors $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ can be thought of as being attached to fingers of your right hand, with the thumb pointing up, index pointing straight, and middle finger pointing left as shown in Figure 2.1.[1]

The coordinates x, y, and z of a point P in this coordinate frame are defined as the (signed) lengths of the orthogonal projections of the vector $\overrightarrow{OP}$ onto the vectors $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ (right side of Figure 2.1), with

$$\begin{cases} x = \overrightarrow{OP} \cdot \boldsymbol{i} \\ y = \overrightarrow{OP} \cdot \boldsymbol{j} \\ z = \overrightarrow{OP} \cdot \boldsymbol{k} \end{cases} \iff \overrightarrow{OP} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

The column vector

$$\boldsymbol{P} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

is called the *coordinate vector* of the point P in (F). We can also define the coordinate vector associated with any free vector $\boldsymbol{v}$ by the lengths of its projections onto the basis vectors of (F), and these coordinates are of course independent of the choice of the origin O. Let us now consider a plane Π, an arbitrary point A in Π, and a unit vector $\boldsymbol{n}$ perpendicular to the plane. The points lying in Π are characterized by

$$\overrightarrow{AP} \cdot \boldsymbol{n} = 0.$$

In a coordinate system (F), where the coordinates of the point P are x, y, and z and the coordinates of $\boldsymbol{n}$ are a, b, and c, this can be rewritten as $\overrightarrow{OP} \cdot \boldsymbol{n} - \overrightarrow{OA} \cdot \boldsymbol{n} = 0$, or

[1]This is the traditional way of defining right-handed coordinate systems. One of the authors, who is left-handed, has always found it a bit confusing and prefers to identify these coordinate systems using the fact that when one looks down the $\boldsymbol{k}$ axis at the $(\boldsymbol{i}, \boldsymbol{j})$ plane, the vector $\boldsymbol{i}$ is mapped onto the vector $\boldsymbol{j}$ by a *counterclockwise* 90° rotation (Figure 2.1). Left-handed coordinate systems correspond to *clockwise* rotations. Left- and right-handed readers alike may find this characterization useful as well.

$$ax + by + cz - d = 0, \tag{2.1}$$

where $d \stackrel{\text{def}}{=} \overrightarrow{OA} \cdot \boldsymbol{n}$ is independent of the choice of the point A in Π and is simply the (signed) distance between the origin O and the plane Π (Figure 2.2).

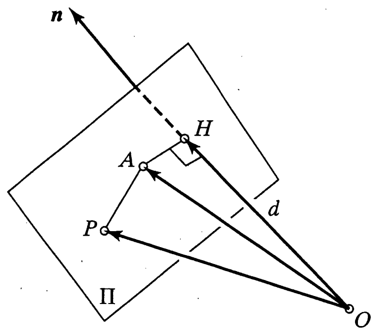

Figure 2.2 The geometric definition of the equation of a plane. The distance d between the origin and plane is reached at the point H where the normal vector passing through the origin pierces the plane.

At times, it is useful to use *homogeneous coordinates* to represent points, vectors, and planes. We formally justify their definition later in this book when we introduce affine and projective geometry in chapters 12 and 13, but for the time being let us note that Eq. (2.1) can be rewritten as

$$(a, b, c, -d) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = 0,$$

or, more concisely, as

$$\boldsymbol{\Pi} \cdot \boldsymbol{P} = 0, \quad \text{where} \quad \boldsymbol{\Pi} \stackrel{\text{def}}{=} \begin{pmatrix} a \\ b \\ c \\ -d \end{pmatrix} \quad \text{and} \quad \boldsymbol{P} \stackrel{\text{def}}{=} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}. \tag{2.2}$$

The vector $\boldsymbol{P}$ is called the *homogeneous coordinate vector* of the point P in the coordinate system (F), and it is simply obtained by adding a fourth coordinate equal to 1 to the ordinary coordinate vector of P. Likewise, the vector $\boldsymbol{\Pi}$ is the vector of homogeneous coordinates of the plane Π in the coordinate frame (F), and Eq. (2.2) is called the equation of Π in that coordinate system. Note that $\boldsymbol{\Pi}$ is only defined up to scale since multiplying this vector by any nonzero constant does not change the solutions of Eq. (2.2). We use the convention that homogeneous coordinates are only defined up to scale, whether they represent points or planes (this may appear a bit counterintuitive for points, but it is fully justified in chapter 13). To go back to the ordinary nonhomogeneous coordinates of points, one just divides all coordinates by the fourth one.

Before proceeding, let us point out that, although our presentation focuses on three-dimensional Euclidean geometry in this chapter, the concepts discussed throughout also apply to planar geometry: A coordinate frame (F) is defined in the plane by its origin o and a right-handed orthonormal basis $(\boldsymbol{i}, \boldsymbol{j})$; the coordinates of the point p in this frame are $x = \overrightarrow{op} \cdot \boldsymbol{i}$ and $y = \overrightarrow{op} \cdot \boldsymbol{j}$, and homogeneous coordinates can be defined as well; in particular, the equation of a line δ in the plane is

$$ax + by - d = 0 \Longleftrightarrow \delta \cdot \boldsymbol{p} = 0, \quad \text{where} \quad \delta = \begin{pmatrix} a \\ b \\ -d \end{pmatrix} \text{ and } \boldsymbol{p} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

and a, b, and d denote, respectively, the coordinates to the unit normal to δ in (F) and the signed distance from o to δ.

Let us go back to three-dimensional geometry and show that homogeneous coordinates can be used to describe more complex geometric figures than points and planes.[2] Consider, for example, a sphere S of radius R centered at the origin. A necessary and sufficient condition for the point P with coordinates x, y, and z to belong to S is of course that

$$x^2 + y^2 + z^2 = R^2,$$

which is equivalent to

$$(x, y, z, 1)^T \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -R^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = 0.$$

More generally, a *quadric surface* is the locus of the points P whose coordinates satisfy the equation

$$a_{200}x^2 + a_{110}xy + a_{020}y^2 + a_{011}yz + a_{002}z^2 + a_{101}xz + a_{100}x + a_{010}y + a_{001}z + a_{000} = 0,$$

and it is straightforward to check that this condition is equivalent to

$$\boldsymbol{P}^T \mathcal{Q} \boldsymbol{P} = 0, \quad \text{where} \quad \mathcal{Q} = \begin{pmatrix} a_{200} & \frac{1}{2}a_{110} & \frac{1}{2}a_{101} & \frac{1}{2}a_{100} \\ \frac{1}{2}a_{110} & a_{020} & \frac{1}{2}a_{011} & \frac{1}{2}a_{010} \\ \frac{1}{2}a_{101} & \frac{1}{2}a_{011} & a_{002} & \frac{1}{2}a_{001} \\ \frac{1}{2}a_{100} & \frac{1}{2}a_{010} & \frac{1}{2}a_{001} & a_{000} \end{pmatrix}. \tag{2.3}$$

In this equation, $\boldsymbol{P}$ denotes the homogeneous coordinate vector of P. Note that $\mathcal{Q}$ is a 4×4 symmetric matrix and, like the parameters a_{ijk}, it is only defined up to scale.

2.1.2 Coordinate System Changes and Rigid Transformations

When several different coordinate systems are considered at the same time, it is convenient to follow Craig (1989) and denote by ${}^F P$ (resp. ${}^F\boldsymbol{v}$) the coordinate vector of the point P (resp. vector $\boldsymbol{v}$) in the frame (F)—that is,

$${}^F P = {}^F\overrightarrow{OP} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Longleftrightarrow \overrightarrow{OP} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

Although the superscripts and subscripts preceding points, vectors, and matrices in Craig's notation may be awkward at first, the rest of this section clearly demonstrates their convenience. Let us now consider two coordinate systems: $(A) = (O_A, \boldsymbol{i}_A, \boldsymbol{j}_A, \boldsymbol{k}_A)$ and $(B) = (O_B, \boldsymbol{i}_B, \boldsymbol{j}_B, \boldsymbol{k}_B)$. The rest of this section shows how to express ${}^B P$ as a function of ${}^A P$. Let us

[2]The inquisitive reader may be wondering about lines in $\mathbb{E}^3$. A line can of course be defined as the intersection of two planes. More generally, lines in $\mathbb{E}^3$ can be defined in terms of *Plücker coordinates*, see Exercises.

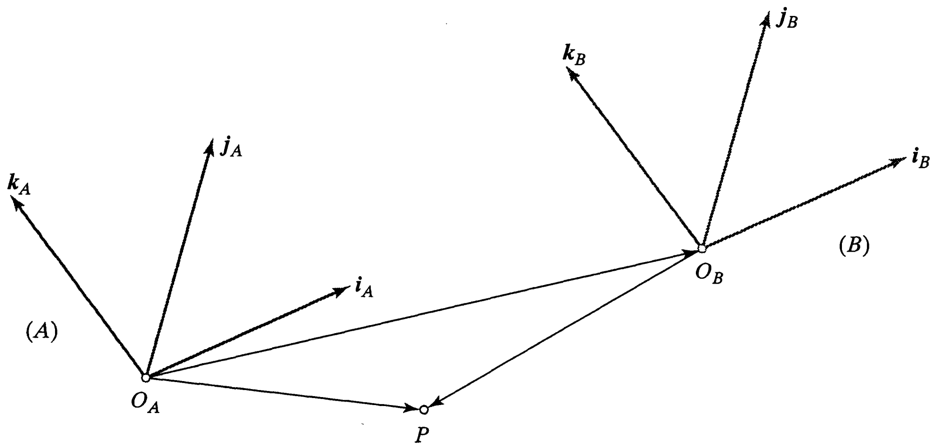

Figure 2.3 Change of coordinates between two frames: pure translation.

suppose first that the basis vectors of both coordinate systems are parallel to each other (i.e., $\boldsymbol{i}_A = \boldsymbol{i}_B$, $\boldsymbol{j}_A = \boldsymbol{j}_B$ and $\boldsymbol{k}_A = \boldsymbol{k}_B$), but the origins O_A and O_B are distinct (Figure 2.3).

We say in this case that the two coordinate systems are separated by a *pure translation*, and we have $\overrightarrow{O_B P} = \overrightarrow{O_B O_A} + \overrightarrow{O_A P}$, thus

$$^{B}P = {}^{A}P + {}^{B}O_A.$$

When the origins of the two frames coincide (i.e., $O_A = O_B = O$), we say that the frames are separated by a *pure rotation* (Figure 2.4). Let us define the *rotation matrix* ${}^{B}_{A}\mathcal{R}$ as the 3×3 array of numbers

$${}^{B}_{A}\mathcal{R} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{i}_A \cdot \boldsymbol{i}_B & \boldsymbol{j}_A \cdot \boldsymbol{i}_B & \boldsymbol{k}_A \cdot \boldsymbol{i}_B \\ \boldsymbol{i}_A \cdot \boldsymbol{j}_B & \boldsymbol{j}_A \cdot \boldsymbol{j}_B & \boldsymbol{k}_A \cdot \boldsymbol{j}_B \\ \boldsymbol{i}_A \cdot \boldsymbol{k}_B & \boldsymbol{j}_A \cdot \boldsymbol{k}_B & \boldsymbol{k}_A \cdot \boldsymbol{k}_B \end{pmatrix}.$$

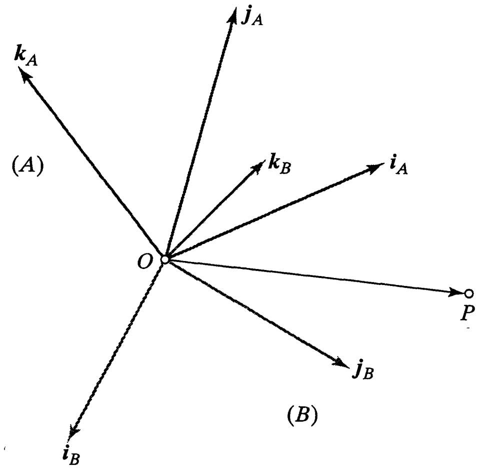

Figure 2.4 Change of coordinates between two frames: pure rotation.

Note that the first column of ${}^B_A\mathcal{R}$ is formed by the coordinates of $\boldsymbol{i}_A$ in the basis $(\boldsymbol{i}_B, \boldsymbol{j}_B, \boldsymbol{k}_B)$. Likewise, the third row of this matrix is formed by the coordinates of $\boldsymbol{k}_B$ in the basis $(\boldsymbol{i}_A, \boldsymbol{j}_A, \boldsymbol{k}_A)$, and so on. More generally, the matrix ${}^B_A\mathcal{R}$ can be written in a more compact fashion using a combination of three column vectors or three row vectors:

$$ {}^B_A\mathcal{R} = \begin{pmatrix} {}^B\boldsymbol{i}_A & {}^B\boldsymbol{j}_A & {}^B\boldsymbol{k}_A \end{pmatrix} = \begin{pmatrix} {}^A\boldsymbol{i}_B{}^T \\ {}^A\boldsymbol{j}_B{}^T \\ {}^A\boldsymbol{k}_B{}^T \end{pmatrix}. $$

It follows that ${}^A_B\mathcal{R} = {}^B_A\mathcal{R}^T$.

As noted earlier, all these subscripts and superscripts may be somewhat confusing at first. To keep everything straight, it is useful to remember that, in a change of coordinates, subscripts refer to the object being described, whereas superscripts refer to the coordinate system in which the object is described. For example, ${}^AP$ refers to the coordinate vector of the point P in the frame (A), ${}^B\boldsymbol{j}_A$ is the coordinate vector of the vector $\boldsymbol{j}_A$ in the frame (B), and ${}^B_A\mathcal{R}$ is the rotation matrix describing the frame (A) in the coordinate system (B).

Let us give an example of pure rotation: Suppose that $\boldsymbol{k}_A = \boldsymbol{k}_B = \boldsymbol{k}$, and denote by θ the angle such that the vector $\boldsymbol{i}_B$ is obtained by applying to the vector $\boldsymbol{i}_A$ a counterclockwise rotation of angle θ about $\boldsymbol{k}$ (Figure 2.5). The angle between the vectors $\boldsymbol{j}_A$ and $\boldsymbol{j}_B$ is also θ in this case, and we have

$$ {}^B_A\mathcal{R} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.4} $$

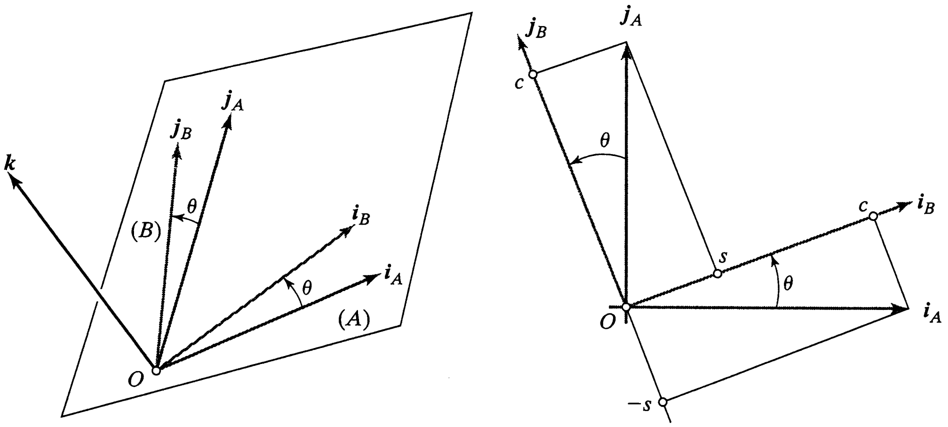

Figure 2.5 Two coordinate frames separated by a rotation of angle θ about their common $\boldsymbol{k}$ basis vector. As shown in the right of the figure, $\boldsymbol{i}_A = c\boldsymbol{i}_B - s\boldsymbol{j}_B$ and $\boldsymbol{j}_A = s\boldsymbol{i}_B + c\boldsymbol{j}_B$, where $c = \cos\theta$ and $s = \sin\theta$.

Similar formulas can be written when the two coordinate systems are deduced from each other via rotations about the $\boldsymbol{i}_A$ or $\boldsymbol{j}_A$ axes (see Exercises). In general, it can be shown that any rotation matrix can be written as the product of three elementary rotations about the $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ vectors of some coordinate system.

Let us go back to characterizing the change of coordinates associated with an arbitrary rotation matrix. Writing

$$\overrightarrow{OP} = \begin{pmatrix} \boldsymbol{i}_A & \boldsymbol{j}_A & \boldsymbol{k}_A \end{pmatrix} \begin{pmatrix} {}^A x \\ {}^A y \\ {}^A z \end{pmatrix} = \begin{pmatrix} \boldsymbol{i}_B & \boldsymbol{j}_B & \boldsymbol{k}_B \end{pmatrix} \begin{pmatrix} {}^B x \\ {}^B y \\ {}^B z \end{pmatrix}$$

in the frame (B) yields immediately

$$^B P = {}^B_A\mathcal{R}\, {}^A P,$$

since the rotation matrix ${}^B_B\mathcal{R}$ is obviously the identity. Note how the subscript matches the following superscript. This property remains true for more general coordinate changes, and it can be used after some practice to reconstruct the corresponding formulas without calculations.

It is easy to show (see Exercises) that rotation matrices are characterized by the following properties: (1) the inverse of a rotation matrix is equal to its transpose, and (2) its determinant is equal to 1. By definition, the columns of a rotation matrix form a right-handed orthonormal coordinate system. It follows from Properties (1) and (2) that their rows also form such a coordinate system.

It should be noted that the set of rotation matrices, equipped with the matrix product, forms a *group*, that is, (a) the product of two rotation matrices is also a rotation matrix (this is intuitively obvious and easily verified analytically); (b) the matrix product is associative—that is, $(\mathcal{R}\mathcal{R}')\mathcal{R}'' = \mathcal{R}(\mathcal{R}'\mathcal{R}'')$ for any rotation matrices $\mathcal{R}$, $\mathcal{R}'$ and $\mathcal{R}''$; (c) there is a unit element, the 3×3 identity matrix Id, that is indeed a rotation matrix and verifies $\mathcal{R}\,\mathrm{Id} = \mathrm{Id}\,\mathcal{R} = \mathcal{R}$ for any rotation matrix $\mathcal{R}$; and (d) every rotation matrix $\mathcal{R}$ admits an inverse $\mathcal{R}^{-1} = \mathcal{R}^T$ such that $\mathcal{R}\mathcal{R}^{-1} = \mathcal{R}^{-1}\mathcal{R} = \mathrm{Id}$. This group is not, however, commutative (i.e., given two rotation matrices $\mathcal{R}$ and $\mathcal{R}'$, the two products $\mathcal{R}\mathcal{R}'$ and $\mathcal{R}'\mathcal{R}$ are in general different).

When the origins and basis vectors of the two coordinate systems are different, we say that the frames are separated by a general *rigid transformation* (Figure 2.6), and we have

$$^B P = {}^B_A\mathcal{R}\, {}^A P + {}^B O_A, \tag{2.5}$$

where ${}^B_A\mathcal{R}$ and ${}^B O_A$ are defined as before. It should be clear that related formulas express coordinate changes for the homogeneous coordinate vectors of planes and the symmetric matrices associated with quadric surfaces (see Exercises).

Homogeneous coordinates can be used to rewrite Eq. (2.5) as a matrix product: Let us first note that matrices can be multiplied in blocks—that is, if

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix} \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} \\ \mathcal{B}_{21} & \mathcal{B}_{22} \end{pmatrix}, \tag{2.6}$$

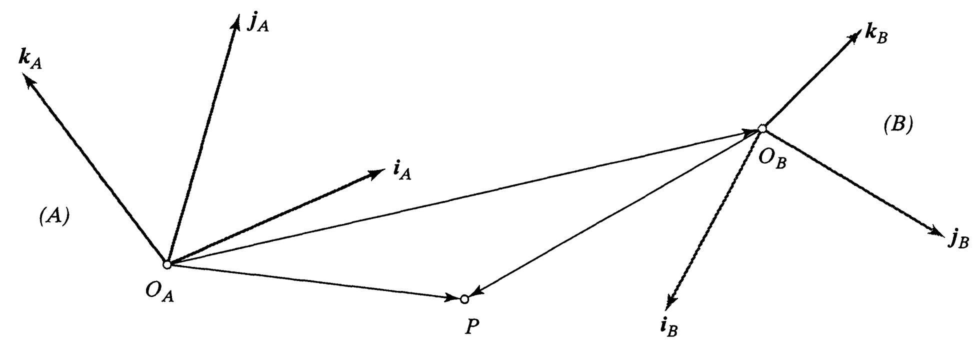

Figure 2.6 Change of coordinates between two frames: general rigid transformation.

where the number of columns of the submatrices $\mathcal{A}_{11}$ and $\mathcal{A}_{21}$ (resp. $\mathcal{A}_{12}$ and $\mathcal{A}_{22}$) is equal to the number of rows of $\mathcal{B}_{11}$ and $\mathcal{B}_{12}$ (resp. $\mathcal{B}_{21}$ and $\mathcal{B}_{22}$), then

$$\mathcal{A}\mathcal{B} = \begin{pmatrix} \mathcal{A}_{11}\mathcal{B}_{11} + \mathcal{A}_{12}\mathcal{B}_{21} & \mathcal{A}_{11}\mathcal{B}_{12} + \mathcal{A}_{12}\mathcal{B}_{22} \\ \mathcal{A}_{21}\mathcal{B}_{11} + \mathcal{A}_{22}\mathcal{B}_{21} & \mathcal{A}_{21}\mathcal{B}_{12} + \mathcal{A}_{22}\mathcal{B}_{22} \end{pmatrix}.$$

In particular, Eq. (2.6) allows us to rewrite the change of coordinates given by Eq. (2.5) as

$$\begin{pmatrix} {}^B P \\ 1 \end{pmatrix} = {}^B_A\mathcal{T} \begin{pmatrix} {}^A P \\ 1 \end{pmatrix}, \quad \text{where} \quad {}^B_A\mathcal{T} \stackrel{\text{def}}{=} \begin{pmatrix} {}^B_A\mathcal{R} & {}^B O_A \\ \mathbf{0}^T & 1 \end{pmatrix} \tag{2.7}$$

and $\mathbf{0} = (0, 0, 0)^T$. In other words, using homogeneous coordinates allows us to write a general change of coordinates as the product of a 4×4 matrix and a 4 vector. It is easy to show that the set of rigid transformations defined by Eq. (2.7), equipped with the matrix product operation, is also a group.

A rigid transformation maps a coordinate system onto another one. In a given coordinate frame (F), it can also be considered as a mapping between points—that is, a point P is mapped onto the point P' such that

$${}^F P' = \mathcal{R}\, {}^F P + \boldsymbol{t} \Longleftrightarrow \begin{pmatrix} {}^F P' \\ 1 \end{pmatrix} = \begin{pmatrix} \mathcal{R} & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} {}^F P \\ 1 \end{pmatrix}, \tag{2.8}$$

where $\mathcal{R}$ is a rotation matrix and $\boldsymbol{t}$ is an element of $\mathbb{R}^3$ (Figure 2.7). The set of rigid transformations considered as mappings of $\mathbb{E}^3$ onto itself and equipped with the law of composition is once again easily shown to form a group. It is also easy to show that rigid transformations preserve the distance between two points and the angle between two vectors. However, the 4×4 matrix associated with a rigid transformation depends on the choice of (F).

For example, let us consider the rotation of angle θ about the $\boldsymbol{k}$ axis of the frame (F). As shown in the exercises, this mapping can be represented by

$${}^F P' = \mathcal{R}\, {}^F P, \quad \text{where} \quad \mathcal{R} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In particular, if (F') is the coordinate system obtained by applying this rotation to (F), we have, according to Eq. (2.4), ${}^{F'} P = {}^{F'}_{F}\mathcal{R}\, {}^F P$ and $\mathcal{R} = {}^{F'}_{F}\mathcal{R}^{-1}$. More generally, the matrix

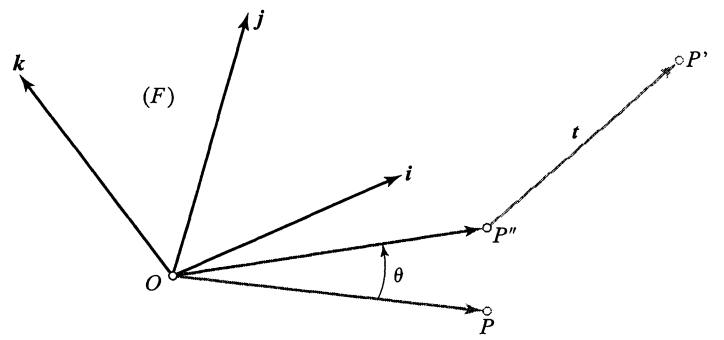

Figure 2.7 A rigid transformation maps the point P onto the point P'' through a rotation $\mathcal{R}$ before mapping P'' onto P' via a translation $\boldsymbol{t}$. In the example shown in this figure, $\mathcal{R}$ is a rotation of angle θ about the $\boldsymbol{k}$ axis of the coordinate system (F).

representing the change of coordinates between two frames is the inverse of the matrix mapping the first frame onto the second one.

What happens when $\mathcal{R}$ is replaced by an arbitrary nonsingular 3×3 matrix $\mathcal{A}$? Equation (2.8) still represents a mapping between points (or a change of coordinates between frames), but this time lengths and angles may not be preserved anymore (equivalently, the new coordinate system does not necessarily have orthogonal axes with unit length). We say that the 4×4 matrix

$$\mathcal{T} = \begin{pmatrix} \mathcal{A} & t \\ \mathbf{0}^T & 1 \end{pmatrix}$$

represents an *affine transformation*. When $\mathcal{T}$ is a nonsingular but otherwise arbitrary 4×4 matrix, we say that we have a projective transformation. Affine and projective transformations also form groups. They will be given a more thorough treatment in chapters 12 and 13.

2.2 CAMERA PARAMETERS AND THE PERSPECTIVE PROJECTION

We saw in chapter 1 that the coordinates x, y, and z of a scene point P observed by a pinhole camera are related to its image coordinates x' and y' by the perspective Eq. (1.1). In reality, this equation is only valid when all distances are measured in the camera's reference frame, and image coordinates have their origin at the principal point 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 principal point, and the position and orientation of the camera.

This section identifies these parameters. We distinguish the *intrinsic* parameters, which relate the camera's coordinate system to the idealized coordinate system used in chapter 1, from the *extrinsic* parameters, which 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 chapter the fact that, for cameras equipped with a lens, a point is only in focus when its depth and the distance between the optical center of the camera and its image plane obey the thin lens Eq. (1.6). Likewise, the nonlinear aberrations associated with real lenses are not taken into account by Eq. (1.1). We neglect these aberrations in this chapter, but revisit radial distortion in chapter 3 when we address the problem of estimating the intrinsic and extrinsic parameters of a camera (a process known as *geometric camera calibration*).

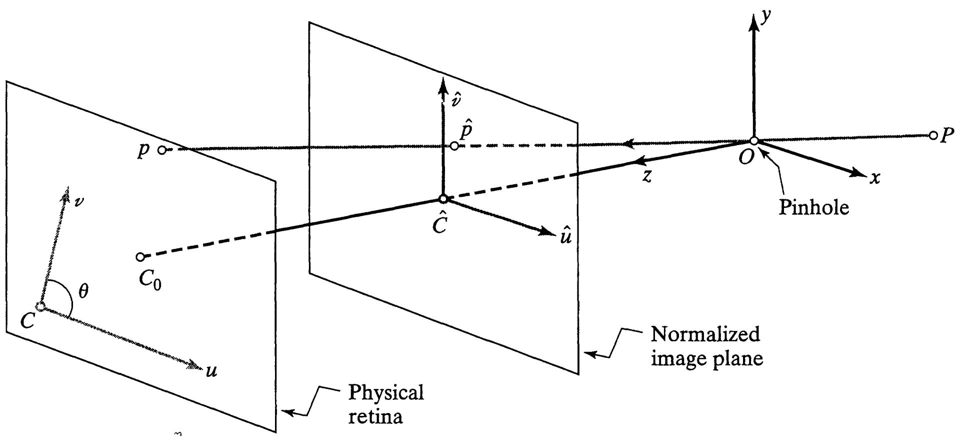

Figure 2.8 Physical and normalized image coordinate systems.

2.2.1 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 2.8). The perspective projection Eq. (1.1) can be written in this normalized coordinate system as

$$\begin{cases} \hat{u} = \dfrac{x}{z} \\ \hat{v} = \dfrac{y}{z} \end{cases} \iff \hat{\boldsymbol{p}} = \frac{1}{z}(\text{Id} \quad \mathbf{0})\begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix}, \tag{2.9}$$

where $\hat{\boldsymbol{p}} \stackrel{\text{def}}{=} (\hat{u}, \hat{v}, 1)^T$ is the vector of homogeneous coordinates of the projection $\hat{p}$ of the point P into the normalized image plane.

The physical retina of the camera is in general different (Figure 2.8): It is located at a distance $f \neq 1$ from the pinhole,[3] and the image coordinates (u, v) of the image 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

$$\begin{cases} u = kf\dfrac{x}{z}, \\ v = lf\dfrac{y}{z}. \end{cases} \tag{2.10}$$

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

In general, the origin of the camera coordinate system is at a corner C of the retina (e.g., in the case depicted in Figure 2.8, the lower left corner or sometimes the upper-left corner, when the image coordinates are the row and column indexes of a pixel) and not at its center, and the center of the CCD matrix usually does not coincide with the principal point C_0. This adds two parameters u_0 and v_0 that define the position (in pixel units) of C_0 in the retinal coordinate system, and Eq. (2.10) is replaced by

$$\begin{cases} u = \alpha\dfrac{x}{z} + u_0, \\ v = \beta\dfrac{y}{z} + v_0. \end{cases} \tag{2.11}$$

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°. In this case, it is easy to show that Eq. (2.11) transforms into

$$\begin{cases} u = \alpha\dfrac{x}{z} - \alpha\cot\theta\dfrac{y}{z} + u_0, \\ v = \dfrac{\beta}{\sin\theta}\dfrac{y}{z} + v_0. \end{cases} \tag{2.12}$$

[3]From now on, we assume that the camera is focused at infinity so the distance between the pinhole and image plane is equal to the focal length.

Combining Eqs. (2.9) and (2.12) now allows us to write the change in coordinates between the physical image frame and the normalized one as a planar affine transformation—that is,

$$p = \mathcal{K}\hat{p}, \quad \text{where} \quad p = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad \text{and} \quad \mathcal{K} \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha \cot\theta & u_0 \\ 0 & \dfrac{\beta}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.13}$$

Putting it all together, we obtain

$$p = \frac{1}{z}\mathcal{M}P, \quad \text{where} \quad \mathcal{M} \stackrel{\text{def}}{=} (\mathcal{K} \quad \mathbf{0}), \tag{2.14}$$

and $P = (x, y, z, 1)^T$ denotes this time the *homogeneous* coordinate vector of P in the camera coordinate system. In other words, homogeneous coordinates can be used to represent the perspective projection mapping by the 3×4 matrix $\mathcal{M}$.

Note that the physical size of the pixels and the skew are always fixed for a given camera and frame grabber, and in principle they can be measured during manufacturing (of course, this information may not be available—for example, in the case of stock film footage, or when the frame grabber's digitization rate is unknown). 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 also affects the magnification since it changes the lens-to-retina distance, but we continue to assume that the camera is focused at infinity and ignore this effect in the rest of this chapter.

2.2.2 Extrinsic Parameters

Let us now consider the case where the camera frame (C) is distinct from the world frame (W). Noting that

$$\begin{pmatrix} {}^C P \\ 1 \end{pmatrix} = \begin{pmatrix} {}^C_W\mathcal{R} & {}^C O_W \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} {}^W P \\ 1 \end{pmatrix}$$

and substituting in Eq. (2.14) yields

$$p = \frac{1}{z}\mathcal{M}P, \quad \text{where} \quad \mathcal{M} = \mathcal{K}(\mathcal{R} \quad t), \tag{2.15}$$

$\mathcal{R} = {}^C_W\mathcal{R}$ is a rotation matrix, $t = {}^C O_W$ is a translation vector, and $P = ({}^W x, {}^W y, {}^W z, 1)^T$ denotes the *homogeneous* coordinate vector of P in the frame (W).

This is the most general form of the perspective projection equation. We can use it to determine the position of the camera's optical center O in the world coordinate system. Indeed, as shown in the exercises, its *homogenous* coordinate vector $\boldsymbol{O}$ verifies $\mathcal{M}\boldsymbol{O} = 0$. (Intuitively, this is rather obvious since the optical center is the only point whose image is not uniquely defined.) In particular, if $\mathcal{M} = (\mathcal{A} \quad b)$, where $\mathcal{A}$ is a nonsingular 3×3 matrix and b is a vector in $\mathbb{R}^3$, then the *nonhomogeneous* coordinate vector of the point O is simply $-\mathcal{A}^{-1}b$.

It is important to understand that the depth z in Eq. (2.15) is *not* independent of $\mathcal{M}$ and $\mathcal{P}$ since, if m_1^T, m_2^T, and m_3^T denote the three rows of $\mathcal{M}$, it follows directly from Eq. (2.15) that $z = m_3 \cdot P$. In fact, it is sometimes convenient to rewrite Eq. (2.15) in the equivalent form:

$$\begin{cases} u = \dfrac{m_1 \cdot P}{m_3 \cdot P}, \\[2ex] v = \dfrac{m_2 \cdot P}{m_3 \cdot P}. \end{cases} \tag{2.16}$$

A projection matrix can be written explicitly as a function of its five intrinsic parameters (α, β, u_0, v_0, and θ) and its six extrinsic ones (the three angles defining $\mathcal{R}$ and the three coordinates of $\boldsymbol{t}$), namely,

$$\mathcal{M} = \begin{pmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + u_0 \boldsymbol{r}_3^T & \alpha t_x - \alpha \cot\theta t_y + u_0 t_z \\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T + v_0 \boldsymbol{r}_3^T & \dfrac{\beta}{\sin\theta} t_y + v_0 t_z \\ \boldsymbol{r}_3^T & t_z \end{pmatrix}, \tag{2.17}$$

where $\boldsymbol{r}_1^T$, $\boldsymbol{r}_2^T$, and $\boldsymbol{r}_3^T$ denote the three rows of the matrix $\mathcal{R}$ and t_x, t_y, and t_z are the coordinates of the vector $\boldsymbol{t}$. If $\mathcal{R}$ is written as the product of three elementary rotations, the vectors $\boldsymbol{r}_i$ ($i = 1, 2, 3$) can of course be written explicitly in terms of the corresponding three angles.

2.2.3 A Characterization of Perspective Projection Matrices

This section examines the conditions under which a 3×4 matrix $\mathcal{M}$ can be written in the form given by Eq. (2.17). Let us write without loss of generality $\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$, 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. (2.17), 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. (2.16) for some arbitrary $\lambda \neq 0$ does not change the corresponding image coordinates. This leads us in the rest of this book to consider projection matrices as *homogeneous objects*, only defined up to scale, whose canonical form of Eq. (2.17) can be obtained by choosing a scale factor such that $|\boldsymbol{a}_3| = 1$. Note that the parameter z in Eq. (2.15) 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. (2.17) 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 are fixed and may be known. In particular, we say that a 3×4 matrix is a *zero-skew perspective projection matrix* when it can be rewritten (up to scale) as Eq. (2.17) 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. (2.17) 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 $\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$ 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: According to Eq. (2.15), we have $\mathcal{A} = \mathcal{KR}$, thus the determinants of $\mathcal{A}$ and $\mathcal{K}$ are the same and $\mathcal{A}$ is nonsingular. Further, a simple calculation shows that the rows of $\mathcal{KR}$ in Eq. (2.17) satisfy the conditions of the theorem under the various assumptions imposed by its statement. The theorem conditions are proved to be sufficient in Faugeras (1993) and in the exercises.

2.3 AFFINE CAMERAS AND AFFINE PROJECTION EQUATIONS

When a scene's relief is small compared with the overall distance separating it from the camera observing it, *affine projection models* can be used to approximate the imaging process. These include the *orthographic* and *weak-perspective* projection models introduced in chapter 1 as well as the parallel and paraperspective models introduced in this section. Their name is justified in chapter 12.

2.3.1 Affine Cameras

Under orthographic projection, the imaging process is simply modeled as an orthogonal projection onto the image plane. This is a reasonable approximation of perspective projection for distant objects lying at a roughly constant distance from the cameras observing them. The *parallel projection* model subsumes the orthographic one and takes into account 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 are not necessarily perpendicular to the image plane.

The weak-perspective and paraperspective projection models generalize the orthographic and parallel projections models to allow for variations in the depth of an object relative to the camera observing it (Figure 2.9). Let O denote the optical center of this camera, and let R denote a scene reference point; the weak-perspective projection of a scene point P is constructed in two

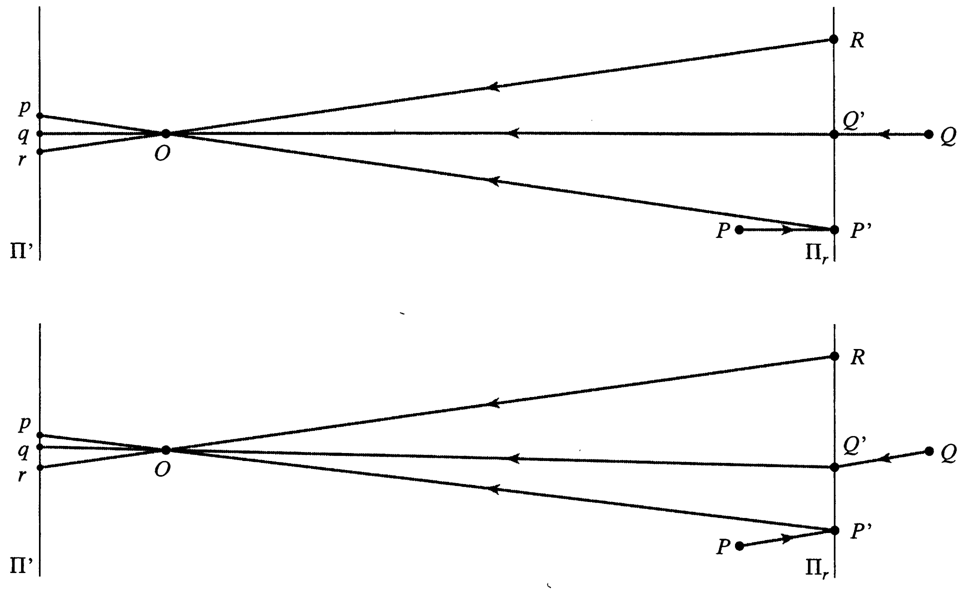

Figure 2.9 Affine projection models: (top) weak-perspective and (bottom) paraperspective projections.

steps: P is first projected orthographically 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 (top of Figure 2.9). Since Π_r is a fronto-parallel plane, the net effect of the second projection step is a scaling of the image coordinates. The paraperspective model 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 (bottom of Figure 2.9): 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.

2.3.2 Affine Projection Equations

Let us derive the weak-perspective projection equation. If z_r denotes the depth of the reference point R, the two elementary projection stages $P \rightarrow P' \rightarrow p$ can be written in the normalized coordinate system attached to the camera as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \longrightarrow \begin{pmatrix} x \\ y \\ z_r \end{pmatrix} \longrightarrow \begin{pmatrix} \hat{u} \\ \hat{v} \\ 1 \end{pmatrix} = \begin{pmatrix} x/z_r \\ y/z_r \\ 1 \end{pmatrix},$$

or, in matrix form,

$$\begin{pmatrix} \hat{u} \\ \hat{v} \\ 1 \end{pmatrix} = \frac{1}{z_r} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & z_r \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

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—that is,

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z_r} \mathcal{K} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & z_r \end{pmatrix} \begin{pmatrix} \mathcal{R} & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix}, \tag{2.18}$$

where $\boldsymbol{P}$ denotes as usual the *nonhomogeneous* coordinate vector of P in the world reference frame. Finally, noting that z_r is a constant and writing

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_2 & \boldsymbol{p}_0 \\ \boldsymbol{0}^T & 1 \end{pmatrix}, \quad \text{where} \quad \mathcal{K}_2 \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha \cot\theta \\ 0 & \dfrac{\beta}{\sin\theta} \end{pmatrix} \quad \text{and} \quad \boldsymbol{p}_0 \stackrel{\text{def}}{=} \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}$$

allows us to rewrite Eq. (2.18) as

$$\boldsymbol{p} = \mathcal{M} \begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix}, \quad \text{where} \quad \mathcal{M} = \begin{pmatrix} \mathcal{A} & \boldsymbol{b} \end{pmatrix}, \tag{2.19}$$

$\boldsymbol{p} = (u, v)^T$ is the *nonhomogeneous* coordinate vector of the point p, and $\mathcal{M}$ is a 2×4 projection matrix (compare to the general perspective Eq. [2.15]). 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_z 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 $u_0 = v_0 = 0$. 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

$$\mathcal{M} = \frac{1}{z_r}\begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathcal{R}_2 & \boldsymbol{t}_2 \end{pmatrix}, \tag{2.20}$$

where k and s denote, respectively, the aspect ratio and 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.

It is easy to show (see Exercises) that the paraperspective projection equations can also be written in the general affine form of Eq. (2.19) with

$$\mathcal{M} = \frac{1}{z_r}\begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\left(\begin{pmatrix} 1 & 0 & -x_r/z_r \\ 0 & 1 & -y_r/z_r \end{pmatrix}\mathcal{R} \quad \boldsymbol{t}_2\right), \tag{2.21}$$

where x_r, y_r, and z_r denote the coordinates of the reference point R in the normalized camera coordinate system. Note that Eq. (2.21) reduces (as expected) to the weak-perspective projection Eq. (2.20) when $x_r = y_r = 0$. According to Eq. (2.21), a paraperspective projection matrix is defined by two intrinsic parameters (k, s), five extrinsic parameters (the three angles defining $\mathcal{R}$ and the two coordinates of $\boldsymbol{t}_2$), and three structure parameters x_r, y_r, and z_r. In practice, the reference point R is often taken to be a point feature whose projection is observable in the image. Its coordinates x_r, y_r, and z_r cannot of course be measured in the image, but the coordinates u_r and v_r of its projection are readily available. It is easy to rewrite Eq. (2.21) as

$$\mathcal{M} = \frac{1}{z_r}\left(\begin{pmatrix} k & s & u_0 - u_r \\ 0 & 1 & v_0 - v_r \end{pmatrix}\mathcal{R} \quad \begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\boldsymbol{t}_2\right). \tag{2.22}$$

In this formulation, the paraperspective projection matrix is defined by four intrinsic parameters (k, s, u_0, and v_0), five extrinsic parameters (the three angles defining $\mathcal{R}$ and the two coordinates of $\boldsymbol{t}_2$), and a single structure parameter z_r.

The orthographic and parallel projection equations are obtained from the weak-perspective and paraperspective ones by fixing the value of z_r to be some constant (in practice, $z_r = 1$) in Eqs. (2.20), (2.21), or (2.22). When several different orthographic (resp. parallel) cameras observe the same scene (or, equivalently, when a zooming camera films an image sequence), the actual image magnifications become relevant, and the simplified calibration matrices used in Eq. (2.20) (resp. Eqs. [2.21] or [2.22]) must be replaced by $\mathcal{K}_2$.

2.3.3 A Characterization of Affine Projection Matrices

A 2×4 matrix $\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$, 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; also note that the matrix $\mathcal{A}$ associated with weak-perspective and paraperspective cameras has rank 2 by construction since, according to Eqs. (2.20), (2.21), and (2.22), it can be written as the product of rank-2 matrices.

Both weak-perspective and general affine projection matrices are defined by eight independent parameters. Paraperspective projection matrices, in contrast, have 10 degrees of freedom.

Weak-perspective and paraperspective projection 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 or a paraperspective one, but that the latter representation is not unique unless additional constraints are imposed on its form. This is confirmed by the following theorem.

Theorem 2. *An affine projection matrix can be written uniquely (up to a sign ambiguity) as a general weak-perspective projection matrix as defined by Eq. (2.20) or as a paraperspective projection matrix as defined by Eq. (2.21) or (2.22) with* $k = 1$ *and* $s = 0$.

This theorem is proven in Faugeras *et al.* (2001, Propositions 4.26 and 4.27) and the exercises. It shows that any affine projection can be written as a weak-perspective or paraperspective one, and that the geometric properties of these projection models apply to general affine projection. For example, as shown in chapter 12, weak-perspective projection preserves the parallelism of lines, and Theorem 2 implies that this property also holds for arbitrary affine projection. The fact that an arbitrary 2×4 matrix can always be written as a paraperspective projection matrix with $k = 1$ and $s = 0$ should *not* be interpreted as meaning that the aspect ratio of a paraperspective camera or its skew are irrelevant.

2.4 NOTES

Craig (1989) offers a good introduction to coordinate system representations and kinematics. Thorough presentations of geometric camera models can be found in Faugeras (1993), Hartley and Zisserman (2000), and Faugeras *et al.* (2001). The paraperspective projection model was introduced in computer vision by Ohta, Maenobu, and Sakai (1981), and its properties have been studied by Aloimonos (1990). The relationship between paraperspective and affine projection models is discussed in Basri (1996). Equations for the perspective projections of straight lines in terms of their *Plücker coordinates* are derived in Faugeras and Papadopoulo (1997) and the exercises below. The machinery introduced in this chapter is used in the next one to calibrate a camera (i.e., to compute its intrinsic and extrinsic parameters from the image positions of fiducial points). It is also a key to the methods for stereo vision and motion analysis presented in chapters 10 to 13. The main equations derived in this chapter have been collected in Table 2.1 for reference.

PROBLEMS

2.1. Write formulas for the matrices ${}^A_B\mathcal{R}$ when (B) is deduced from (A) via a rotation of angle θ about the axes $\boldsymbol{i}_A$, $\boldsymbol{j}_A$, and $\boldsymbol{k}_A$ respectively.

2.2. Show that rotation matrices are characterized by the following properties: (a) the inverse of a rotation matrix is its transpose and (b) its determinant is 1.

2.3. Show that the set of matrices associated with rigid transformations and equipped with the matrix product forms a group.

2.4. Let ${}^A\mathcal{T}$ denote the matrix associated with a rigid transformation $\mathcal{T}$ in the coordinate system (A), with

$$ {}^A\mathcal{T} = \begin{pmatrix} {}^A\mathcal{R} & {}^A\boldsymbol{t} \\ \boldsymbol{0} & 1 \end{pmatrix}. $$

Construct the matrix ${}^B\mathcal{T}$ associated with $\mathcal{T}$ in the coordinate system (B) as a function of ${}^A\mathcal{T}$ and the rigid transformation separating (A) and (B).

2.5. Show that if the coordinate system (B) is obtained by applying to the coordinate system (A) the transformation matrix $\mathcal{T}$, then ${}^BP = \mathcal{T}^{-1}\,{}^AP$.

TABLE 2.1 Reference card: geometric camera models.

Plane equation (homogeneous)	$\Pi \cdot \boldsymbol{P} = ax + by + cz - d = 0$
Quadric surface equation (homogeneous)	$\boldsymbol{P}^T \mathcal{Q} \boldsymbol{P} = 0$ with $\mathcal{Q} = \begin{pmatrix} a_{200} & \frac{1}{2}a_{110} & \frac{1}{2}a_{101} & \frac{1}{2}a_{100} \\ \frac{1}{2}a_{110} & a_{020} & \frac{1}{2}a_{011} & \frac{1}{2}a_{010} \\ \frac{1}{2}a_{101} & \frac{1}{2}a_{011} & a_{002} & \frac{1}{2}a_{001} \\ \frac{1}{2}a_{100} & \frac{1}{2}a_{010} & \frac{1}{2}a_{001} & a_{000} \end{pmatrix}$
Rotation matrix	${}^B_A\mathcal{R} = \begin{pmatrix} \boldsymbol{i}_A \cdot \boldsymbol{i}_B & \boldsymbol{j}_A \cdot \boldsymbol{i}_B & \boldsymbol{k}_A \cdot \boldsymbol{i}_B \\ \boldsymbol{i}_A \cdot \boldsymbol{j}_B & \boldsymbol{j}_A \cdot \boldsymbol{j}_B & \boldsymbol{k}_A \cdot \boldsymbol{j}_B \\ \boldsymbol{i}_A \cdot \boldsymbol{k}_B & \boldsymbol{j}_A \cdot \boldsymbol{k}_B & \boldsymbol{k}_A \cdot \boldsymbol{k}_B \end{pmatrix}$
Change of coordinates (nonhomogeneous)	${}^BP = {}^B_A\mathcal{R}\,{}^AP + {}^BO_A$
Perspective projection equation (homogeneous)	$\boldsymbol{p} = \frac{1}{z}\mathcal{M}\boldsymbol{P}$
Matrix of intrinsic parameters	$\mathcal{K} = \begin{pmatrix} \alpha & -\alpha\cot\theta & u_0 \\ 0 & \beta/\sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}$
Perspective projection matrix	$\mathcal{M} = \mathcal{K}(\mathcal{R} \quad \boldsymbol{t})$
Affine projection equation (nonhomogeneous)	$\boldsymbol{p} = \mathcal{M}\begin{pmatrix}\boldsymbol{P} \\ 1\end{pmatrix} = \mathcal{A}\boldsymbol{P} + \boldsymbol{b}$
Weak-perspective projection matrix	$\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b}) = \frac{1}{z_r}\begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}(\mathcal{R}_2 \quad \boldsymbol{t}_2)$
Paraperspective projection matrix I	$\mathcal{M} = \frac{1}{z_r}\begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\left(\begin{pmatrix} 1 & 0 & -x_r/z_r \\ 0 & 1 & -y_r/z_r \end{pmatrix}\mathcal{R} \quad \boldsymbol{t}_2\right)$
Paraperspective projection matrix II	$\mathcal{M} = \frac{1}{z_r}\left(\begin{pmatrix} k & s & u_0 - u_r \\ 0 & 1 & v_0 - v_r \end{pmatrix}\mathcal{R} \quad \begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\boldsymbol{t}_2\right)$

2.6. Show that the rotation of angle θ about the $\boldsymbol{k}$ axis of the frame (F) can be represented by

$$ {}^FP' = \mathcal{R}\,{}^FP, \quad \text{where} \quad \mathcal{R} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. $$

2.7. Show that the change of coordinates associated with a rigid transformation preserves distances and angles.

2.8. Show that when the camera coordinate system is skewed and the angle θ between the two image axes is not equal to 90°, then Eq. (2.11) transforms into Eq. (2.12).

2.9. Let $\boldsymbol{O}$ denote the *homogeneous* coordinate vector of the optical center of a camera in some reference frame, and let $\mathcal{M}$ denote the corresponding perspective projection matrix. Show that $\mathcal{M}(\boldsymbol{O}) = 0$.

2.10. Show that the conditions of Theorem 1 are necessary.

2.11. Show that the conditions of Theorem 1 are sufficient. Note that the statement of this theorem is a bit different from the corresponding theorems in Faugeras (1993) and Heyden (1995), where the condition $\text{Det}(\mathcal{A}) \neq 0$ is replaced by $\boldsymbol{a}_3 \neq 0$. Of course, $\text{Det}(\mathcal{A}) \neq 0$ implies $\boldsymbol{a}_3 \neq 0$.

2.12. If ${}^A\mathbf{\Pi}$ denotes the homogeneous coordinate vector of a plane Π in the coordinate frame (A), what is the homogeneous coordinate vector ${}^B\mathbf{\Pi}$ of Π in the frame (B)?

2.13. If ${}^A\mathcal{Q}$ denotes the symmetric matrix associated with a quadric surface in the coordinate frame (A), what is the symmetric matrix ${}^B\mathcal{Q}$ associated with this surface in the frame (B)?

2.14. Prove Theorem 2.

2.15. **Line Plücker coordinates.** The *exterior product* of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^4$ is defined by

$$\boldsymbol{u} \wedge \boldsymbol{v} \stackrel{\text{def}}{=} \begin{pmatrix} u_1v_2 - u_2v_1 \\ u_1v_3 - u_3v_1 \\ u_1v_4 - u_4v_1 \\ u_2v_3 - u_3v_2 \\ u_2v_4 - u_4v_2 \\ u_3v_4 - u_4v_3 \end{pmatrix}.$$

Given a fixed coordinate system and the (homogeneous) coordinates vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ associated with two points A and B in $\mathbb{E}^3$, the vector $\boldsymbol{L} = \boldsymbol{A} \wedge \boldsymbol{B}$ is called the vector of Plücker coordinates of the line joining A to B.

(a) Let us write $\boldsymbol{L} = (L_1, L_2, L_3, L_4, L_5, L_6)^T$ and denote by O the origin of the coordinate system and by H its projection onto L. Let us also identify the vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ with their non-homogeneous coordinate vectors. Show that $\overrightarrow{AB} = -(L_3, L_5, L_6)^T$ and $\overrightarrow{OA} \times \overrightarrow{OB} = \overrightarrow{OH} \times \overrightarrow{AB} = (L_4, -L_2, L_1)^T$. Conclude that the Plücker coordinates of a line obey the quadratic constraint $L_1L_6 - L_2L_5 + L_3L_4 = 0$.

(b) Show that changing the position of the points A and B along the line L only changes the overall scale of the vector $\boldsymbol{L}$. Conclude that Plücker coordinates are homogeneous coordinates.

(c) Prove that the following identity holds of any vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, and $\boldsymbol{t}$ in $\mathbb{R}^4$:

$$(\boldsymbol{x} \wedge \boldsymbol{y}) \cdot (\boldsymbol{z} \wedge \boldsymbol{t}) = (\boldsymbol{x} \cdot \boldsymbol{z})(\boldsymbol{y} \cdot \boldsymbol{t}) - (\boldsymbol{x} \cdot \boldsymbol{t})(\boldsymbol{y} \cdot \boldsymbol{z}).$$

(d) Use this identity to show that the mapping between a line with Plücker coordinate vector $\boldsymbol{L}$ and its image l with homogeneous coordinates $\boldsymbol{l}$ can be represented by

$$\rho \boldsymbol{l} = \tilde{\mathcal{M}} \boldsymbol{L}, \quad \text{where} \quad \tilde{\mathcal{M}} \stackrel{\text{def}}{=} \begin{pmatrix} (\boldsymbol{m}_2 \wedge \boldsymbol{m}_3)^T \\ (\boldsymbol{m}_3 \wedge \boldsymbol{m}_1)^T \\ (\boldsymbol{m}_1 \wedge \boldsymbol{m}_2)^T \end{pmatrix}, \tag{2.23}$$

and $\boldsymbol{m}_1^T$, $\boldsymbol{m}_2^T$, and $\boldsymbol{m}_3^T$ denote as before the rows of $\mathcal{M}$ and ρ is an appropriate scale factor.

Hint: Consider a line L joining two points A and B and denote by a and b the projections of these two points, with homogeneous coordinates $\boldsymbol{a}$ and $\boldsymbol{b}$. Use the fact that the points a and b lie on l, thus if $\boldsymbol{l}$ denote the homogeneous coordinate vector of this line, we must have $\boldsymbol{l} \cdot \boldsymbol{a} = \boldsymbol{l} \cdot \boldsymbol{b} = 0$.

(e) Given a line L with Plücker coordinate vector $\boldsymbol{L} = (L_1, L_2, L_3, L_4, L_5, L_6)^T$ and a point P with homogeneous coordinate vector $\boldsymbol{P}$, show that a necessary and sufficient condition for P to lie on L is that

$$\mathcal{L}\boldsymbol{P} = 0, \quad \text{where} \quad \mathcal{L} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & L_6 & -L_5 & L_4 \\ -L_6 & 0 & L_3 & -L_2 \\ L_5 & -L_3 & 0 & L_1 \\ -L_4 & L_2 & -L_1 & 0 \end{pmatrix}.$$

(f) Show that a necessary and sufficient condition for the line L to lie in the plane Π with homogeneous coordinate vector $\mathbf{\Pi}$ is that

$$\mathcal{L}^*\mathbf{\Pi} = 0, \quad \text{where} \quad \mathcal{L}^* \stackrel{\text{def}}{=} \begin{pmatrix} 0 & L_1 & L_2 & L_3 \\ -L_1 & 0 & L_4 & L_5 \\ -L_2 & -L_4 & 0 & L_6 \\ -L_3 & -L_5 & -L_6 & 0 \end{pmatrix}.$$