Intrinsic camera parameters (§2.2,§3.3)

- Recall the basic equation

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} \text{Transformation} \\ \text{representing} \\ \text{intrinsic parameters} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \text{Transformation} \\ \text{representing} \\ \text{extrinsic parameters} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

- Note that the special projection matrix can be ignored if we remember that by either (1) making the extrinsic parameter matrix 3x4 by dropping the bottom row, or (2) making the intrinsic parameter matrix 3x4 by adding a column of zeros

- Recall the meaning of α and β from the main slide sequence.

- We pick up the book's treatment after the authors give their version of α and β on page 29.

Intrinsic camera parameters (§2.2,§3.3)

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

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

a (e.g.,
, when
er, and
is adds
rdinate
(2.11)
cturing
fferent
(2.12)
$v = \dfrac{\beta}{\sin\theta} \dfrac{y}{z} + v_0.$

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,

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

Putting it all together, we obtain

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

and $\boldsymbol{P} = (x, y, z, 1)^T$ 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

$$\boldsymbol{p} = \frac{1}{z}\mathcal{M}\boldsymbol{P}, \quad \text{where} \quad \mathcal{M} = \mathcal{K}\begin{pmatrix}\mathcal{R} & \boldsymbol{t}\end{pmatrix}, \tag{2.15}$$

$\mathcal{R} = {}^C_W\mathcal{R}$ is a rotation matrix, $\boldsymbol{t} = {}^C O_W$ is a translation vector, and $\boldsymbol{P} = ({}^W x, {}^W y, {}^W z, 1)^T$ 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 $\mathcal{M} = \begin{pmatrix}\mathcal{A} & \boldsymbol{b}\end{pmatrix}$, where $\mathcal{A}$ is a nonsingular 3×3 matrix and $\boldsymbol{b}$ is a vector in $\mathbb{R}^3$, then the *nonhomogeneous* coordinate vector of the point O is simply $-\mathcal{A}^{-1}\boldsymbol{b}$.

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.

Intrinsic camera parameters (§2.2,§3.3)

3.2.2 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. [2.17] in chapter 2) can be used to recover these parameters as follows: We write as before $\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$, with $\boldsymbol{a}_1^T$, $\boldsymbol{a}_2^T$, and $\boldsymbol{a}_3^T$ denoting the rows of $\mathcal{A}$. We obtain

$$\rho(\mathcal{A} \quad \boldsymbol{b}) = \mathcal{K}(\mathcal{R} \quad \boldsymbol{t}) \Longleftrightarrow \rho \begin{pmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \boldsymbol{a}_3^T \end{pmatrix} = \begin{pmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + u_0 \boldsymbol{r}_3^T \\ \dfrac{\beta}{\sin\theta} \boldsymbol{r}_2^T + v_0 \boldsymbol{r}_3^T \\ \boldsymbol{r}_3^T \end{pmatrix},$$

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

$$\begin{cases} \rho = \varepsilon/|\boldsymbol{a}_3|, \\ \boldsymbol{r}_3 = \rho \boldsymbol{a}_3, \\ u_0 = \rho^2(\boldsymbol{a}_1 \cdot \boldsymbol{a}_3), \\ v_0 = \rho^2(\boldsymbol{a}_2 \cdot \boldsymbol{a}_3), \end{cases} \quad \text{where} \quad \varepsilon = \mp 1. \tag{3.11}$$

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

$$\begin{cases} \rho^2(\boldsymbol{a}_1 \times \boldsymbol{a}_3) = -\alpha \boldsymbol{r}_2 - \alpha \cot\theta \boldsymbol{r}_1, \\ \rho^2(\boldsymbol{a}_2 \times \boldsymbol{a}_3) = \dfrac{\beta}{\sin\theta}\boldsymbol{r}_1, \end{cases} \quad \text{and} \quad \begin{cases} \rho^2|\boldsymbol{a}_1 \times \boldsymbol{a}_3| = \dfrac{|\alpha|}{\sin\theta}, \\ \rho^2|\boldsymbol{a}_2 \times \boldsymbol{a}_3| = \dfrac{|\beta|}{\sin\theta}. \end{cases} \tag{3.12}$$

Thus,

$$\begin{cases} \cos\theta = -\dfrac{(\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)}{|\boldsymbol{a}_1 \times \boldsymbol{a}_3||\boldsymbol{a}_2 \times \boldsymbol{a}_3|}, \\ \alpha = \rho^2|\boldsymbol{a}_1 \times \boldsymbol{a}_3| \sin\theta, \\ \beta = \rho^2|\boldsymbol{a}_2 \times \boldsymbol{a}_3| \sin\theta, \end{cases} \tag{3.13}$$

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 part of Eq. (3.12) as

$$\begin{cases} \boldsymbol{r}_1 = \dfrac{\rho^2 \sin\theta}{\beta}(\boldsymbol{a}_2 \times \boldsymbol{a}_3) = \dfrac{1}{|\boldsymbol{a}_2 \times \boldsymbol{a}_3|}(\boldsymbol{a}_2 \times \boldsymbol{a}_3), \\ \boldsymbol{r}_2 = \boldsymbol{r}_3 \times \boldsymbol{r}_1. \end{cases} \tag{3.14}$$

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_z 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.