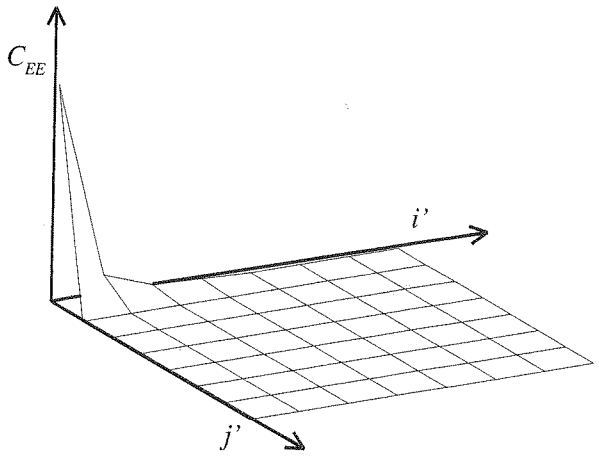

Figure 2.12 Autocovariance of the image of a uniform pattern for a typical image acquisition system, showing cross-talking between adjacent pixels along i'.

☞ The autocovariance should actually be estimated as the average of the autocovariance computed on many images of the same pattern. To minimize the effect of radiometric nonlinearities (see (2.13)), C_{EE} should be computed on a patch in the central portion of the image.

Figure 2.12 displays the graph of the average of the autocovariance computed on many images acquired by the same acquisition system used to generate Figure 2.11. The autocovariance was computed by means of (2.18) on a patch of 16×16 pixels centered in the image center. Notice the small but visible covariance along the horizontal direction: consistently with the physical properties of many CCD cameras, this indicates that the grey value of each pixel is not completely independent of that of its neighbors.

2.4 Camera Parameters

We now come back to discuss the geometry of a vision system in greater detail. In particular, we want to characterize the parameters underlying camera models.

2.4.1 Definitions

Computer vision algorithms reconstructing the 3-D structure of a scene or computing the position of objects in space need equations linking the coordinates of points in 3-D space with the coordinates of their corresponding image points. These equations are written in the camera reference frame (see (2.14) and section 2.2.4), but it is often assumed that

- the camera reference frame can be located with respect to some other, known, reference frame (the *world reference frame*), and

- the coordinates of the image points in the camera reference frame can be obtained from *pixel coordinates*, the only ones directly available from the image.

This is equivalent to assume knowledge of some camera's characteristics, known in vision as the camera's *extrinsic* and *intrinsic* parameters. Our next task is to understand the exact nature of the intrinsic and extrinsic parameters and why the equivalence holds.

Definition: Camera Parameters

The *extrinsic parameters* are the parameters that define the location and orientation of the camera reference frame with respect to a known world reference frame.

The *intrinsic parameters* are the parameters necessary to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame.

In the next two sections, we write the basic equations that allow us to define the extrinsic and intrinsic parameters in practical terms. The problem of estimating the value of these parameters is called *camera calibration*. We shall solve this problem in Chapter 6, since calibration methods need algorithms which we discuss in Chapters 4 and 5.

2.4.2 Extrinsic Parameters

The camera reference frame has been introduced for the purpose of writing the fundamental equations of the perspective projection (2.14) in a simple form. However, *the camera reference frame is often unknown*, and a common problem is determining the location and orientation of the camera frame with respect to some known reference frame, *using only image information*. The extrinsic parameters are defined as *any set of geometric parameters that identify uniquely the transformation between the unknown camera reference frame and a known reference frame*, named the *world reference frame*.

A typical choice for describing the transformation between camera and world frame is to use

- a 3-D translation vector, $\mathbf{T}$, describing the relative positions of the origins of the two reference frames, and
- a 3×3 rotation matrix, R, an orthogonal matrix ($R^\top R = RR^\top = I$) that brings the corresponding axes of the two frames onto each other.

The orthogonality relations reduce the number of degrees of freedom of R to three (see section A.9 in the Appendix).

In an obvious notation (see Figure 2.13), the relation between the coordinates of a point $\mathbf{P}$ in world and camera frame, $\mathbf{P}_w$ and $\mathbf{P}_c$ respectively, is

$$\mathbf{P}_c = R(\mathbf{P}_w - \mathbf{T}), \tag{2.19}$$

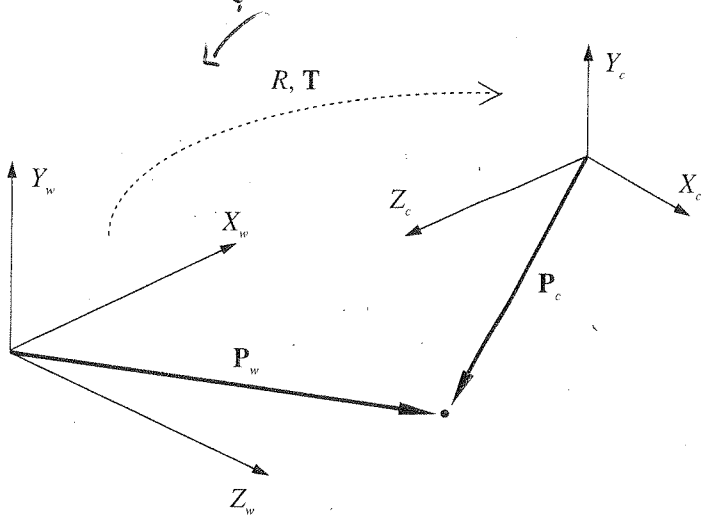

Figure 2.13 The relation between camera and world coordinate frames.

with

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}.$$

Definition: Extrinsic Parameters

The camera extrinsic parameters are the translation vector, $\mathbf{T}$, and the rotation matrix, R (or, better, its free parameters), which specify the transformation between the camera and the world reference frame.

2.4.3 Intrinsic Parameters

The intrinsic parameters can be defined as the set of parameters needed to characterize the optical, geometric, and digital characteristics of the viewing camera. For a pinhole camera, we need three sets of intrinsic parameters, specifying respectively

- the perspective projection, for which the only parameter is the focal length, f;
- the transformation between camera frame coordinates and pixel coordinates;
- the geometric distortion introduced by the optics.

From Camera to Pixel Coordinates. To find the second set of intrinsic parameters, we must link the coordinates (x_{im}, y_{im}) of an image point in pixel units with the coordinates (x, y) of the same point in the camera reference frame. The coordinates

(x_{im}, y_{im}) can be thought of as coordinates of a new reference frame, sometimes called *image reference frame*.

The Transformation between Camera and Image Frame Coordinates

Neglecting any geometric distorsions possibly introduced by the optics and in the assumption that the CCD array is made of a rectangular grid of photosensitive elements, we have

$$x = -(x_{im} - o_x)s_x$$
$$y = -(y_{im} - o_y)s_y \tag{2.20}$$

with (o_x, o_y) the coordinates in pixel of the image center (the principal point), and (s_x, s_y) the effective size of the pixel (in millimeters) in the horizontal and vertical direction respectively.

Therefore, the current set of intrinsic parameters is f, o_x, o_y, s_x, s_y.

☞ The sign change in (2.20) is due to the fact that the horizontal and vertical axes of the image and camera reference frames have opposite orientation.

In several cases, the optics introduces image distortions that become evident at the periphery of the image, or even elsewhere using optics with large fields of view. Fortunately, these distortions can be modelled rather accurately as simple *radial distortions*, according to the relations

$$x = x_d(1 + k_1 r^2 + k_2 r^4)$$
$$y = y_d(1 + k_1 r^2 + k_2 r^4)$$

with (x_d, y_d) the coordinates of the distorted points, and $r^2 = x_d^2 + y_d^2$. As shown by the equations above, this distortion is a radial displacement of the image points. The displacement is null at the image center, and increases with the distance of the point from the image center. k_1 and k_2 are further intrinsic parameters. Since they are usually very small, radial distortion is ignored whenever high accuracy is not required in all regions of the image, or when the peripheral pixels can be discarded. If not, as $k_2 << k_1$, k_2 is often set equal to 0, and k_1 is the only intrinsic parameter to be estimated in the radial distortion model.

☞ The magnitude of geometric distortion depends on the quality of the lens used. As a rule of thumb, with optics of average quality and CCD size around 500×500, expect distortions of several pixels (say around 5) in the outer cornice of the image. Under these circumstances, a model with $k_2 = 0$ is still accurate.

It is now time for a summary.

Intrinsic Parameters

The camera intrinsic parameters are defined as the focal length, f, the location of the image center in pixel coordinates, (o_x, o_y), the effective pixel size in the horizontal and vertical direction (s_x, s_y), and, if required, the radial distortion coefficient, k_1.

2.4.4 Camera Models Revisited

We are now fully equipped to write relations linking directly the pixel coordinates of an image point with the world coordinates of the corresponding 3-D point, *without explicit reference to the camera reference frame* needed by (2.14).

Linear Version of the Perspective Projection Equations. Plugging (2.19) and (2.20) into (2.14) we obtain

$$\begin{aligned} -(x_{im} - o_x)s_x &= f\frac{\mathbf{R}_1^\top(\mathbf{P}_w - \mathbf{T})}{\mathbf{R}_3^\top(\mathbf{P}_w - \mathbf{T})} \\ -(y_{im} - o_y)s_y &= f\frac{\mathbf{R}_2^\top(\mathbf{P}_w - \mathbf{T})}{\mathbf{R}_3^\top(\mathbf{P}_w - \mathbf{T})} \end{aligned} \tag{2.21}$$

where $\mathbf{R}_i$, $i = 1, 2, 3$, is a 3-D vector formed by the i-th row of the matrix R. Indeed, (2.21) relates the 3-D coordinates of a point in the world frame to the image coordinates of the corresponding image point, via the camera extrinsic and intrinsic parameters.

☞ Notice that, due to the particular form of (2.21), not all the intrinsic parameters are independent. In particular, the focal length could be absorbed into the effective sizes of the CCD elements.

Neglecting radial distortion, we can rewrite (2.21) as a simple matrix product. To this purpose, we define two matrices, M_{int} and M_{ext}, as

$$M_{int} = \begin{pmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$M_{ext} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & -\mathbf{R}_1^\top\mathbf{T} \\ r_{21} & r_{22} & r_{23} & -\mathbf{R}_2^\top\mathbf{T} \\ r_{31} & r_{32} & r_{33} & -\mathbf{R}_3^\top\mathbf{T} \end{pmatrix},$$

so that *the* 3×3 *matrix* M_{int} *depends only on the intrinsic parameters, while the* 3×4 *matrix* M_{ext} *only on the extrinsic parameters*. If we now add a "1" as a fourth coordinate of $\mathbf{P}_w$ (that is, express $\mathbf{P}_w$ in homogeneous coordinates), and form the product $M_{int}M_{ext}\mathbf{P}_w$, we obtain a linear matrix equation describing perspective projections.

The Linear Matrix Equation of Perspective Projections

$$\begin{pmatrix} x_h \\ y_h \\ \cdot \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = M_{int} M_{ext} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}.$$

What is interesting about vector $[x_1, x_2, x_3]^\top$ is that the ratios (x_1/x_3) and (x_2/x_3) are nothing but the image coordinates:

$$x_1/x_3 = x_{im}$$
$$x_2/x_3 = y_{im}.$$

Moreover, we have separated nicely the two steps of the world-image projection:

- M_{ext} performs the transformation between the world and the camera reference frame;
- M_{int} performs the transformation between the camera reference frame and the image reference frame.

☞ In more formal terms, the relation between a 3-D point and its perspective projection on the image plane can be seen as a linear transformation from the *projective space*, the space of vectors $[X_w, Y_w, Z_w, 1]^\top$, to the *projective plane*, the space of vectors $[x_1, x_2, x_3]^\top$. This transformation is defined *up to an arbitrary scale factor* and so that the matrix M has only 11 independent entries (see review questions). This fact will be discussed in Chapter 6.

The Perspective Camera Model. Various camera models, including the perspective and weak-perspective ones, can be derived by setting appropriate constraints on the matrix $M = M_{int} M_{ext}$. Assuming, for simplicity, $o_x = o_y = 0$ and $s_x = s_y = 1$, M can then be rewritten as

$$M = \begin{pmatrix} -fr_{11} & +fr_{12} & +fr_{13} & f\mathbf{R}_1^\top \mathbf{T} \\ -fr_{21} & -fr_{22} & -fr_{23} & f\mathbf{R}_2^\top \mathbf{T} \\ r_{31} & r_{32} & r_{33} & -\mathbf{R}_3^\top \mathbf{T} \end{pmatrix}.$$

When unconstrained, M describes the full-perspective camera model and is called *projection matrix*.

The Weak-Perspective Camera Model. To derive the form of M for the weak-perspective camera model, we observe that the image $\mathbf{p}$ of a point $\mathbf{P}$ is given by

$$\mathbf{p} = M \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \begin{pmatrix} f\mathbf{R}_1^\top (\mathbf{T} - \mathbf{P}) \\ f\mathbf{R}_2^\top (\mathbf{T} - \mathbf{P}) \\ \mathbf{R}_3^\top (\mathbf{P} - \mathbf{T}) \end{pmatrix}. \quad (2.22)$$

But $\|\mathbf{R}_3^\top(\mathbf{P}-\mathbf{T})\|$ is simply the distance of $\mathbf{P}$ from the projection center along the optical axis; therefore, the basic constraint for the weak-perspective approximation can be written as

$$\left|\frac{\mathbf{R}_3^\top(\mathbf{P}_i-\bar{\mathbf{P}})}{\mathbf{R}_3^\top(\bar{\mathbf{P}}-\mathbf{T})}\right| << 1, \tag{2.23}$$

where $\mathbf{P}_1, \mathbf{P}_2$ are two points in 3-D space, and $\bar{\mathbf{P}}$ the centroid of $\mathbf{P}_1$ and $\mathbf{P}_2$. Using (2.23), (2.22) can be written for $\mathbf{P}=\mathbf{P}_i$, $i=1,2$, as

$$\mathbf{p}_i \approx \begin{pmatrix} f\mathbf{R}_1^\top(\mathbf{T}-\mathbf{P}_i) \\ f\mathbf{R}_2^\top(\mathbf{T}-\mathbf{P}_i) \\ \mathbf{R}_3^\top(\bar{\mathbf{P}}-\mathbf{T}) \end{pmatrix}.$$

Therefore, the projection matrix M becomes

$$M_{wp} = \begin{pmatrix} -fr_{11} & -fr_{12} & -fr_{13} & f\mathbf{R}_1^\top\mathbf{T} \\ -fr_{21} & -fr_{22} & -fr_{23} & f\mathbf{R}_2^\top\mathbf{T} \\ 0 & 0 & 0 & \mathbf{R}_3^\top(\bar{\mathbf{P}}-\mathbf{T}) \end{pmatrix}.$$

The Affine Camera Model. Another interesting camera model, widely used in the literature for its simplicity, is the so-called *affine model*, a mathematical generalization of the weak-perspective model. In the affine model, the first three entries in the last row of the matrix M are equal to zero. All other entries are unconstrained. The affine model does not appear to correspond to any physical camera, but leads to simple equations and has appealing geometric properties. The affine projection does not preserve angles but does preserve parallelism.

The main difference with the weak-perspective model is the fact that, in the affine model, only the ratio of distances measured along parallel directions is preserved. We now move on to consider range images.

2.5 Range Data and Range Sensors

In many applications, one wants to use vision to measure distances; for example, to steer vehicles away from obstacles, estimate the shape of surfaces, or inspect manufactured objects. A single intensity image proves of limited use, as pixel values are related to surface geometry only indirectly; that is, through the optical and geometrical properties of the surfaces as well as the illumination conditions. All these are usually complex to model and often unknown. As we shall see in Chapter 9, reconstructing 3-D shape from a single intensity image is difficult and often inaccurate. Can we acquire images encoding shape *directly*? Yes: this is exactly what range sensors do.

Range Images

Range images are a special class of digital images. Each pixel of a range image expresses the *distance between a known reference frame and a visible point in the scene*. Therefore, a range image reproduces the 3-D *structure* of a scene, and is best thought of as a *sampled surface*.