image lines, and can be created easily on a suitable calibration pattern. In fact, we can use the same calibration pattern (actually, the same image!) of Figure 6.1, already used for EXPL_PARS_CAL, so that EXPL_PARS_CAL and the new algorithm, IMAGE_CENTER_CAL, fit nicely together.

Algorithm IMAGE_CENTER_CAL

The input is an image of the calibration pattern in Figure 6.1, and the output of the first two steps of algorithm EXPL_PARS_CAL.

1. Compute the three vanishing points p_1, p_2, and p_3, determined by the three bundles of lines obtained in step 2 of EXPL_PARS_CAL.
2. Compute the orthocenter, O, of the triangle $p_1p_2p_3$.

The output are the image coordinates of the image center, O.

☞ It is essential that the calibration pattern is imaged from a viewpoint guaranteeing that no vanishing point lies much farther than the others from the image center; otherwise, the image lines become nearly parallel, and small inaccuracies in the location of the lines result in large errors in the coordinates of the vanishing point. This can happen if one of the three mutually orthogonal directions is nearly parallel to the image plane, a situation to be definitely avoided. Even with a good viewpoint, it is best to determine the vanishing points using several lines and least squares.

☞ To improve the accuracy of the image center estimate, you should run IMAGE_CENTER_CAL with several views of the calibration patterns, and average the results.

Experience shows that an accurate location of the image center is not crucial for obtaining precise estimates of the other camera parameters (see Further Readings). Be careful, however, as accurate knowledge of the image center *is* required to determine the ray in space identified by an image point (as we shall see, for example, in Chapter 8).

6.3 Camera Parameters from the Projection Matrix

We now move on to the description of a second method for camera calibration. The new method consists in two sequential stages:

1. estimate the projection matrix linking world and image coordinates;
2. compute the camera parameters as closed-form functions of the entries of the projection matrix.

6.3.1 Estimation of the Projection Matrix

As we have seen in Chapter 2, the relation between the 3-D coordinates (X_i^w, Y_i^w, Z_i^w) of a point in space and the 2-D coordinates (x, y) of its projection on the image plane

can be written by means of a 3×4 projection matrix, M, according to the equation

$$\begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix} = M \begin{pmatrix} X_i^w \\ Y_i^w \\ Z_i^w \\ 1 \end{pmatrix},$$

with

$$\begin{aligned} x &= \frac{u_i}{w_i} = \frac{m_{11}X_i^w + m_{12}Y_i^w + m_{13}Z_i^w + m_{14}}{m_{31}X_i^w + m_{32}Y_i^w + m_{33}Z_i^w + m_{34}} \\ y &= \frac{v_i}{w_i} = \frac{m_{21}X_i^w + m_{22}Y_i^w + m_{23}Z_i^w + m_{24}}{m_{31}X_i^w + m_{32}Y_i^w + m_{33}Z_i^w + m_{34}} \end{aligned} \tag{6.16}$$

The matrix M is defined up to an arbitrary scale factor and has therefore only 11 independent entries, which can be determined through a homogeneous linear system formed by writing (6.16) for at least 6 world-image point matches. However, through the use of calibration patterns like the one in Figure 6.1, many more correspondences and equations can be obtained and M can be estimated through least squares techniques. If we assume we are given N matches for the homogeneous linear system we have

$$A\mathbf{m} = 0, \tag{6.17}$$

with

$$A = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X_1 & -y_1Y_1 & -y_1Z_1 & -y_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -x_2X_2 & -x_2Y_2 & -x_2Z_2 & -x_2 \\ 0 & 0 & 0 & 0 & X_2 & Y_2 & Z_2 & 1 & -y_2X_2 & -y_2Y_2 & -y_2Z_2 & -y_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -x_NX_N & -x_NY_N & -x_NZ_N & -x_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -y_NX_N & -y_NY_N & -y_NZ_N & -y_N \end{bmatrix}$$

and

$$\mathbf{m} = [m_{11}, m_{12}, \ldots, m_{33}, m_{34}]^\top.$$

Since A has rank 11, the vector $\mathbf{m}$ can be recovered from SVD related techniques as the column of V corresponding to the zero (in practice the smallest) singular value of A, with $A = UDV^\top$ (see Appendix, section A.6). In agreement with the above definition of M, this means that the entries of M are obtained up to an unknown scale factor. The following is the detailed implementation of a method for estimating the matrix M:

Algorithm PROJ_MAT_CALIB

The input is an image of the calibration pattern described in the text (see Figure 6.1, for example).

1. Run the first two steps of EXPL_PARS_CAL.
2. Given N world-image matches, compute the SVD of A, system matrix of (6.17), $A = UDV^\top$. The solution $\mathbf{m}$ is the column of V corresponding to the smallest singular value of A.

The output is formed by the entries of the projection matrix, determined up to an unknown scale factor.

6.3.2 Computing Camera Parameters

We now want to express the intrinsic and extrinsic camera parameters as functions of the estimated projection matrix. To avoid confusion, we call $\hat{M}$ the projection matrix estimated through PROJ_MAT_CALIB (and hence $\hat{m}_{ij}$ the generic element of $\hat{M}$).

We first rewrite the full expression for the entries of M, or[4]

$$M = \begin{bmatrix} -f_x r_{11} + o_x r_{31} & -f_x r_{12} + o_x r_{32} & -f_x r_{13} + o_x r_{33} & -f_x T_x + o_x T_z \\ -f_y r_{21} + o_y r_{31} & -f_y r_{22} + o_y r_{32} & -f_y r_{23} + o_y r_{33} & -f_y T_y + o_y T_z \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix}. \quad (6.18)$$

☞ Notice that we are now using o_x, o_y, f_x, f_y as the four independent parameters, as opposed to $o_x, o_y, f_x, \alpha = s_y/s_x$ used in method 1.

☞ Notice also that not all 3×4 matrices can be written as functions of the extrinsic and intrinsic parameters as per 6.18. For details, see the Further Readings.

In what follows we also need the 3-D vectors

$$\mathbf{q}_1 = [\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13}]^\top,$$
$$\mathbf{q}_2 = [\hat{m}_{21}, \hat{m}_{22}, \hat{m}_{23}]^\top,$$
$$\mathbf{q}_3 = [\hat{m}_{31}, \hat{m}_{32}, \hat{m}_{33}]^\top,$$
$$\mathbf{q}_4 = [\hat{m}_{14}, \hat{m}_{24}, m_{34}]^\top.$$

Since M is defined up to a scale factor we can write

$$\hat{M} = \gamma M.$$

The absolute value of the scale factor, $|\gamma|$, can be obtained by observing that $\mathbf{q}_3$ is the last row of the rotation matrix, R. Hence,

$$\sqrt{\hat{m}_{31}^2 + \hat{m}_{32}^2 + \hat{m}_{33}^2} = |\gamma| \sqrt{r_{31}^2 + r_{32}^2 + r_{33}^2} = |\gamma|.$$

[4] These expressions were already written in Chapter 2 in order to link the perspective and weak-perspective camera models, but scaled by $1)f$, and under the simplifying assumptions $s_x = s_y = 1$, $o_x = o_y = 0$.

We now divide each entry of $\hat{M}$ by $|\gamma|$, and observe that the resulting, normalized projection matrix differs from M by, at most, a sign change. From now on, therefore, we indicate with $\hat{M}$ the normalized matrix. From the last row of (6.18) we have

$$T_z = \sigma \hat{m}_{34}$$

and

$$r_{3i} = \sigma \hat{m}_{3i}, \quad i = 1, 2, 3,$$

with $\sigma = \pm 1$. Taking the dot products of $\mathbf{q}_3$ with $\mathbf{q}_1$ and $\mathbf{q}_2$ we find

$$o_x = \mathbf{q}_1^\top \mathbf{q}_3$$

and

$$o_y = \mathbf{q}_2^\top \mathbf{q}_3.$$

Then we can recover f_x and f_y:

$$\begin{aligned} f_x &= \sqrt{\mathbf{q}_1^\top \mathbf{q}_1 - o_x^2} \\ f_y &= \sqrt{\mathbf{q}_2^\top \mathbf{q}_2 - o_y^2}, \end{aligned} \quad (6.19)$$

We can now compute the remaining extrinsic parameters as

$$\begin{aligned} r_{1i} &= \sigma(o_x \hat{m}_{3i} - \hat{m}_{1i})/f_x, \quad i = 1, 2, 3 \\ r_{2i} &= \sigma(o_y \hat{m}_{3i} - \hat{m}_{2i})/f_y, \quad i = 1, 2, 3 \\ T_x &= \sigma(o_x T_z - \hat{m}_{14})/f_x \\ T_y &= \sigma(o_y T_z - \hat{m}_{24})/f_y. \end{aligned} \quad (6.20)$$

As usual, the estimated rotation matrix, $\hat{R}$, is not really orthogonal, and we can find the closest orthogonal matrix as done in section 6.2.2.

☞ You may have noticed that M has 11 independent parameters, but there are only 6 extrinsic and 4 intrinsic parameters. The missing parameter is the angle, θ, formed by the axes of the image reference frame. Here we have exploited the fact that this angle is 90° within great accuracy in all commercial cameras, and is therefore not considered explicitly in the intrinsic parameters set. For details of both linear and nonlinear calibration methods estimating θ as well, see the Further Readings.

We are left to discuss how to determine the sign σ. The sign σ can be obtained from $T_z = \sigma \hat{m}_{34}$, because we know whether the origin of the world reference frame is in front of ($T_z > 0$) or behind ($T_z < 0$) the camera. We do not give the usual algorithm, which would merely restate the equations derived in this section.

6.4 Concluding Remarks

Is there any difference between the two calibration methods presented? Obviously, you should expect the same (in practice, very similar) parameter values from both! Method 2 is probably simpler, but the algorithm on which we based EXPL_PARS_CAL (see Further Readings) is a well-known technique in the computer vision community, and has been implemented and used by many groups. PROJ_MAT_CALIB is useful whenever the projection matrix is sufficient to solve a vision problem, and there is no need to make the individual parameters explicit; an example is referenced in the Further Readings.

We said that the precision of calibration depends on how accurately the world and image reference points are located. But *which* accuracy should one pursue? As the errors on the parameter estimates propagate to the results of the application, the answer depends ultimately on the accuracy requirements of the target application. For instance, inspection systems in manufacturing often require submillimeter accuracies, whereas errors of centimeters can be acceptable in vision systems for outdoors navigation. Locating image reference points or lines at pixel precision is generally unsatisfactory in the former case, but acceptable in the latter. A practical guideline is: *the effort going into improving calibration accuracy should be commensurate to the requirements of the application.*

6.5 Summary

After working through this chapter you should be able to:

- ❑ explain what calibration is and why it is necessary
- ❑ calibrate the intrinsic and extrinsic parameters of an intensity camera
- ❑ estimate the entries of the projection matrix
- ❑ design a calibration pattern, motivating your design choices

6.6 Further Readings

EXPL_PARS_CAL has been adapted from a well-known (but rather involved) paper by Tsai [10], which contains a proof that the rank of the matrix A of EXPL_PARS_CAL is 7 in the ideal case. The orthocenter property and the calibration of the image center through vanishing points is due to Caprile and Torre [2], who also suggest a neat method for calibrating the rotation matrix. Direct calibration of the projection matrix is described, for example, by Faugeras and Toscani [5]. The explicit computation of the calibration parameters from the projection matrix is taken from Faugeras's book [4], which also discusses the conditions under which this is possible (Chapter 3, Section 4). Ayache and Lustman [1] describe a stereo system which requires calibration of the projection matrix, but not explicit knowledge of the camera parameters. The influence on vision algorithms of erroneous estimates of the image center and other camera parameters is discussed by [3, 7, 6]. Thacker and Mayhew [9] describe a method for calibrating a stereo system from arbitrary stereo images.

Recent developments on calibration are presented, for instance, by Faugeras and Maybank [8] who introduce an elegant method based solely on point matches and able to obtain camera calibration without the use of calibration patterns.

6.7 Review

Questions

- ❑ *6.1* Describe the purpose of calibration and name applications in which calibration is necessary.
- ❑ *6.2* What is the relation between the translation vectors between two reference frames if the transformation between the frames is (a) first rotation and then translation and (b) *vice versa*? Verify your answer in the trivial case in which rotation is described by the identity matrix.
- ❑ *6.3* Why does algorithm EXPL_PARS_CAL determine only f_x and not f_y?
- ❑ *6.4* Discuss the construction requirements of a calibration pattern.
- ❑ *6.5* How would you estimate the accuracy of a calibration program?

Exercises

- ❍ *6.1* In the case of large fields of view, or if pixels very far from the center of the image are considered, the radial distortion cannot be neglected. A possible procedure for calibrating the distortion parameter k_1 is to rewrite (6.13) with $x_i(1+k_1r_i^2)$ in place of x_i with $r_i^2 = x_i^2 + \alpha^2 y_i^2$, or

$$x_i(1+k_1r_i^2)(r_{31}X_i^w + r_{32}Y_i^w + r_{33}Z_i^w + T_z) = -f_x(r_{11}X_i^w + r_{12}Y_i^w + r_{13}Z_i^w + T_x).$$

  The corresponding nonlinear system of equations for the unknowns f_x, T_z, and k_1 can be solved through gradient descent techniques using the output of EXPL_PARS_CAL as initial guess for f_x and T_z and 0 as initial guess for k_1. Write out the complete equations suggested, and devise a calibration algorithm including k_1 in the intrinsic parameters.
- ❍ *6.2* Prove the orthocenter theorem by geometrical arguments. (*Hint*[5]: Let h be the altitude from the vertex (and vanishing point) v to the side s, and O the projection center. Since both the segments h and vO are orthogonal to s, the plane through h and vO is orthogonal to s and hence to the image plane . . .)
- ❍ *6.3* Prove that the vanishing points associated to three coplanar bundles of parallel lines are collinear.
- ❍ *6.4* Estimate the theoretical error of the coordinates of the principal point as a function of the error (uncertainty) of the coordinates of the vanishing points. Can you guess the viewpoint that minimizes error propagation?

[5] Hint by Francesco Robbiano.