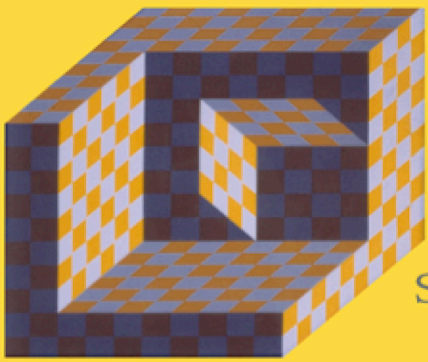
INTERDISCIPLINARY APPLIED MATHEMATICS

IMAGING, VISION, AND GRAPHICS

An Invitation to 3-D Vision

From Images to Geometric Models



Yi Ma
Stefano Soatto
Jana Kosecka
Shankar S. Sastry

Springer

6.5 Calibration with scene knowledge

In this section, we discuss additional techniques for camera calibration by exploiting scene knowledge. We first discuss, in Section 6.5.1, how to exploit orthogonality and parallelism constraints between vectors. In Section 6.5.2, we discuss how to use complete (Euclidean) knowledge of an object in the scene if available. More complete exposition of the constraints provided by partial scene knowledge follows more naturally from a multiple-view formulation[13] and will be covered in Chapter 10.

[12]For many restricted camera motion sequences, the camera parameters typically cannot be uniquely determined by any autocalibration schemes; see Exercise 6.16 and Section 8.5.

[13]Even for a single image.

6.5.1 *Partial scene knowledge*

In man-made environments, information such as *orthogonality* and *parallelism* among line features (e.g., edges of buildings, streets) can be safely assumed. For instance, Figure 6.13 shows three sets of lines that are image of lines in space that are likely pairwise parallel, with each set orthogonal to the other two.

Figure 6.13. Three sets of parallel lines: solid, dashed, and dotted. These sets are mutually orthogonal to each other. Image size: 400×300 pixels.

We say "likely" because there is no way to verify, based on images alone, whether such information is true or not. The reader should be aware that from images we can only *assume* that lines are parallel or perpendicular, and provide algorithms that exploit such assumptions. If the assumptions turn out to be violated (as for instance, Figure 6.1), the resulting calibration will be wrong, and we have no way of verifying that. We now demonstrate how such assumptions, when satisfied, can be used to calibrate the camera. A set of parallel lines intersect at the same point infinitely far. The projection of this point onto the image plane is called a vanishing point. If the two vectors $\boldsymbol{\ell}^1, \boldsymbol{\ell}^2 \in \mathbb{R}^3$ represent the (co)images of the two parallel lines,[14] then the corresponding vanishing point is given by

$$\boldsymbol{v} \sim \boldsymbol{\ell}^1 \times \boldsymbol{\ell}^2 = \widehat{\boldsymbol{\ell}^1}\boldsymbol{\ell}^2, \tag{6.51}$$

where $\sim$ indicates homogeneous equality up to a scalar factor. Since for an image such as that in Figure 6.14 the three sets of lines are mutually orthogonal, we may assume that, by a proper choice of the world coordinate frame, their 3-D directions coincide with the three principal directions: $e_1 = [1, 0, 0]^T$, $e_2 = [0, 1, 0]^T$, $e_3 = [0, 0, 1]^T$. In the image, the vanishing points corresponding to the three sets of parallel lines are respectively

$$\boldsymbol{v}_1 = KRe_1, \quad \boldsymbol{v}_2 = KRe_2, \quad \boldsymbol{v}_3 = KRe_3.$$

[14] We recall that we use the convention of using superscripts to enumerate different features in the scene and subscripts to enumerate different images of the same feature.

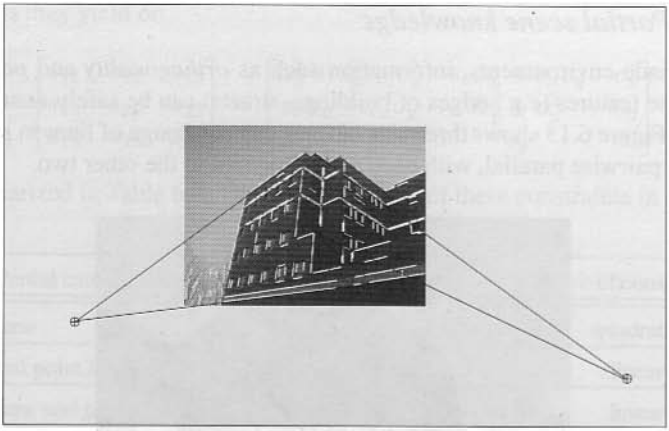


Figure 6.14. An example of two vanishing points associated with two out of the three sets of parallel lines in Figure 6.13.

Note that the coordinates of the vanishing points depend only on rotation and internal parameters of the camera, but not on translation. The orthogonality relations among e_1, e_2, e_3 readily provide constraints on the calibration matrix K. In particular, we have

$$\boldsymbol{v}_i^T S \boldsymbol{v}_j = \boldsymbol{v}_i^T K^{-T} K^{-1} \boldsymbol{v}_j = \boldsymbol{e}_i^T R^T R \boldsymbol{e}_j = \boldsymbol{e}_i^T \boldsymbol{e}_j = 0, \quad i \neq j,$$

where again $S = K^{-T}K^{-1}$ is the symmetric matrix associated with the uncalibrated camera

$$S = K^{-T}K^{-1} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

When three vanishing points are detected, they provide three independent constraints on the matrix S:

$$\boldsymbol{v}_1^T S \boldsymbol{v}_2 = 0, \quad \boldsymbol{v}_1^T S \boldsymbol{v}_3 = 0, \quad \boldsymbol{v}_2^T S \boldsymbol{v}_3 = 0.$$

In general, the symmetric matrix S has five degrees of freedom. Without additional constraints we can recover S only up to a two-parameter family of solutions from the three linear equations above. With the assumption of zero skew ($fs_\theta = 0$) and known aspect ratio ($fs_x = fs_y$), one may obtain a unique solution for K from a single image as above. As an example, the camera calibration for the image in Figure 6.13 is

$$K = \begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 409.33 & 0 & 177.46 \\ 0 & 409.33 & 165.75 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}. \quad (6.52)$$

A more detailed study of the degeneracies in exploiting these types of constraints can be found in [Liebowitz and Zisserman, 1999].

6.5.2 *Calibration with a rig*

Calibration with a rig is the method of choice for camera calibration when one has access to the camera and can place a known object in the scene.

Under these conditions, one can use an object with a distinct number of points, whose coordinates relative to some reference frame are known with high accuracy, as a *calibration rig*. Notice that a calibration rig could be an actual object manufactured primarily for the purpose of camera calibration (Figure 6.15), or simply an object in the scene with known geometry, for instance a golf ball with dots painted on it, or the rim of a car wheel whose alignment needs to be computed from a collection of cameras.

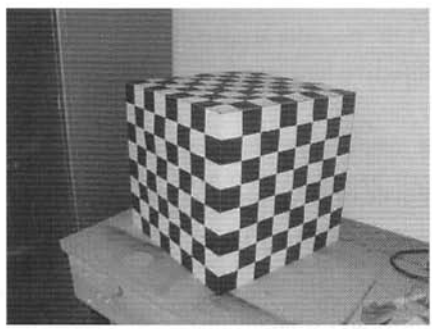


Figure 6.15. An example of a calibration rig for laboratory use: a checkerboard-textured cube.

Let $\boldsymbol{X} = [X, Y, Z, 1]^T$ be the coordinates of a point p on the rig. Its image has the pixel coordinates $\boldsymbol{x}' = [x', y', 1]^T$ that satisfy equation (6.1). If we let $\pi_1, \pi_2, \pi_3 \in \mathbb{R}^4$ be the three row vectors of the projection matrix $\Pi = K\Pi_0 g \in \mathbb{R}^{3\times 4}$, then equation (6.1) can be written for each point p^i on the rig as

$$\lambda^i \begin{bmatrix} x'^i \\ y'^i \\ 1 \end{bmatrix} = \begin{bmatrix} \pi_1^T \\ \pi_2^T \\ \pi_3^T \end{bmatrix} \begin{bmatrix} X^i \\ Y^i \\ Z^i \\ 1 \end{bmatrix}. \tag{6.53}$$

From the third row we get $\lambda^i = \pi_3^T \boldsymbol{X}^i$. Hence for each point we obtain the following two equations

$$\begin{aligned} x'^i(\pi_3^T \boldsymbol{X}^i) &= \pi_1^T \boldsymbol{X}^i, \\ y'^i(\pi_3^T \boldsymbol{X}^i) &= \pi_2^T \boldsymbol{X}^i. \end{aligned}$$

Unlike previous formulations, here X^i, Y^i, and Z^i are known, and so are x'^i, y'^i. We can therefore stack all the unknown entries of Π into a vector and rewrite the equations above as a system of linear equations

$$M\Pi^s = 0,$$

where Π^s is a stacked projection matrix Π,

$$\Pi^s = [\pi_{11}, \pi_{21}, \pi_{31}, \pi_{12}, \pi_{22}, \pi_{32}, \pi_{13}, \pi_{23}, \pi_{33}, \pi_{14}, \pi_{24}, \pi_{34}]^T \quad \in \mathbb{R}^{12},$$

and rows of M are functions of (x'^i, y'^i) and (X^i, Y^i, Z^i). A linear (suboptimal) estimate of Π^s can then be obtained by minimizing the least-squares criterion

$$\min \|M\Pi^s\|^2 \quad \text{subject to} \quad \|\Pi^s\|^2 = 1 \tag{6.54}$$

without taking into account the structure of the unknown vector Π^s. This can be accomplished, as usual, using the SVD. If we denote the (nonlinear) projection map from $\boldsymbol{X} = [X, Y, Z]^T$ to $\boldsymbol{x}' = [x', y']^T$ by

$$\boldsymbol{x}' \doteq h(\Pi \boldsymbol{X}),$$

where $h(\boldsymbol{X}) = [X,\ Y]/Z$, then the estimate can be further refined by a nonlinear minimization of the objective function

$$\min_{\Pi} \sum_i \left\| \boldsymbol{x}'^i - h(\Pi \boldsymbol{X}^i) \right\|^2.$$

After obtaining an estimate of the projection matrix

$$\Pi = K[R \mid T] = [KR \mid KT],$$

we can factor the first 3×3 submatrix into the calibration matrix $K \in \mathbb{R}^{3\times 3}$ (in its upper triangular form) and rotation matrix $R \in SO(3)$ using a routine QR decomposition

$$KR = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix},$$

which yields an estimate of the intrinsic parameters in the calibration matrix K and that of the rotational component of the extrinsic parameters. Estimating translation completes the calibration procedure:

$$T = K^{-1} \begin{bmatrix} \pi_{14} \\ \pi_{24} \\ \pi_{34} \end{bmatrix}.$$

This procedure requires that the points in space have known coordinates $\{\boldsymbol{X}^i\}$, and that they be in *general position*, i.e. that they do not lie on a set of measure zero, for instance a plane. However, the simplest and most common calibration rigs available are indeed planar checkerboard patterns! Therefore, we study the case of a planar calibration rig in detail below.

6.5.3 *Calibration with a planar pattern*

Although the approach just described requires only one image of a known object, it does not return a unique and well-conditioned estimate of the calibration if the object is planar. Since nonplanar calibration rigs are not easy to manufacture,

a more commonly adopted approach consists in capturing several images of a known planar object, such as a checkerboard like the one shown in Figure 6.16.

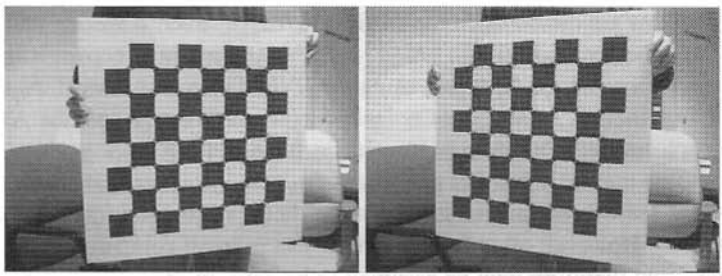


Figure 6.16. Two images of the checkerboard for camera calibration. The resolution of the images is 640×480.

Since we are free to choose the world reference frame, we choose it aligned with the board so that points on it have coordinates of the special form $\boldsymbol{X} = [X, Y, 0, 1]^T$. Notice that the center of the world frame needs to be *on* the board, and the Z-axis of the world frame is the normal vector. Then, with respect to the camera coordinate frame, the image $\boldsymbol{x}'$ of a point $\boldsymbol{X}$ on the board is given by equation (6.1), which for the given choice of coordinate frames simplifies to

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = K[r_1, r_2, T] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \tag{6.55}$$

where $r_1, r_2 \in \mathbb{R}^3$ are the first and second columns of the rotation matrix R. Notice that the matrix

$$H \doteq K[r_1, r_2, T] \quad \in \mathbb{R}^{3\times 3} \tag{6.56}$$

is a linear transformation of the homogeneous coordinates $[X, Y, 1]^T$ to the homogeneous coordinates $\boldsymbol{x}' = [x', y', 1]^T$; i.e. it is a homography between the checkerboard plane and the image plane.

Applying the common trick of multiplying both sides by the skew-symmetric matrix $\widehat{\boldsymbol{x}'}$ in order to eliminate λ (recall that $\widehat{\boldsymbol{x}'}\boldsymbol{x}' = 0$) yields

$$\widehat{\boldsymbol{x}'} H [X, Y, 1]^T = 0. \tag{6.57}$$

Notice that in the above equation, we know both $\boldsymbol{x}'$ (measured from the image) and $[X, Y, 1]^T$ (given from knowledge of the checkerboard). Hence H can be solved (up to a scalar factor) linearly from such equations if sufficiently many points on the checkerboard are given. We know from the previous chapter that at least four images of such points are needed in order to solve for the homography H up to a scalar factor.

Once we know H, we observe that its first two columns are simply $[h_1, h_2] \sim K[r_1, r_2]$. This is equivalent to $K^{-1}[h_1, h_2] \sim [r_1, r_2]$. Since r_1, r_2 are orthonormal vectors, we obtain two equations that the calibration matrix K has

to satisfy:

$$h_1^T K^{-T} K^{-1} h_2 = 0, \quad h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2. \tag{6.58}$$

These equations are quadratic in the entries of K. However, if we are willing to neglect the structure of $S = K^{-T}K^{-1} \in \mathbb{R}^{3\times 3}$, we can just recover it linearly from the equations above, as usual, using the SVD. Once S is known, K can be retrieved, in principle,[15] using the Cholesky factorization (see Appendix A).

From each image, we obtain two such (linear) equations in $S = K^{-T}K^{-1}$. The calibration matrix K has five unknowns $fs_x, fs_y, fs_\theta, o_x, o_y$; so does S. We then need at least three images to determine them, since each image produces two equations. However, often the skew term fs_θ is small compared to the other parameters, and therefore may be assumed to be zero. In that case, one has only four parameters in K (or S) to determine: fs_x, fs_y, o_x, o_y. Another way to see this is that in the zero-skew case, there is an extra linear constraint that the matrix S needs to satisfy:

$$e_1^T S e_2 = 0.$$

As an example, for the two images given in Figure 6.16, the calibration result given by the above scheme is

$$K = \begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 769.942 & 0 & 319.562 \\ 0 & 769.086 & 244.162 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

The method we have outlined above is merely conceptual, in the sense that it will not perform satisfactorily in the presence of noise in the measurements of $\boldsymbol{x}'$ and $\boldsymbol{X}$. In practice, some refinement of the solution based on nonlinear optimization techniques is necessary, and if needed, such techniques can also simultaneously estimate radial distortion.

Several free software packages are publicly available that perform camera calibration using a planar rig. For instance, Intel's OpenCV software library (`http://sourceforge.net/projects/opencvlibrary/`) provides code that is easy to use, accurate, and well documented.

[15] We say "in principle" because in practice, noise may prevent the matrix S recovered from data from being positive definite.