CHAPTER 13

Using Camera Models

13.1 CAMERA CALIBRATION FROM A 3D REFERENCE

Camera calibration involves estimating the intrinsic parameters of the camera, and perhaps lens parameters if needed, from one or more images. There are numerous strategies, all using the following recipe: build a *calibration object*, where the positions of some points (*calibration points*) are known; view that object from one or more viewpoints; obtain the image locations of the calibration points; and solve an optimization problem to recover camera intrinsics and perhaps lens parameters. As one would expect, much depends on the choice of calibration object. If all the calibration points sit on an object, the extrinsics will yield the *pose* (for position and orientation) of the object with respect to the camera. We use a two step procedure: formulate the optimization problem, then find a good starting point.

13.1.1 Formulating the Optimization Problem

The optimization problem is relatively straightforward to formulate. Notation is the main issue. We have N reference points $\mathbf{s}_i = [s_{x,i}, s_{y,i}, s_{z,i}]$ with known position in some reference coordinate system in 3D. The measured location in the image for the i'th such point is $\hat{\mathbf{t}}_i = \left[\hat{t}_{x,i}, \hat{t}_{y,i}\right]$. There may be measurement errors, so the $\hat{\mathbf{t}}_i = \mathbf{t}_i + \xi_i$, where ξ_i is an error vector and $\mathbf{t}_i$ is the unknown true position of the image point. We will assume the magnitude of error does not depend on direction in the image plane (it is *isotropic*), so it is natural to minimize the squared magnitude of the error

$$\sum_i \xi_i^T \xi_i.$$

The main issue here is writing out expressions for ξ_i in the appropriate coordinates. Write $\mathcal{T}_i$ for the intrinsic matrix whose u, v'th component will be i_{uv}; $\mathcal{T}_e$ for the extrinsic transformation, whose u, v'th component will be e_{uv}. Recalling that $\mathcal{T}_i$ is upper triangular, and engaging in some manipulation, we obtain

$$\sum_i \xi_i^T \xi_i = \sum_i (t_{x,i} - p_{x,i})^2 + (t_{y,i} - p_{y,i})^2$$

where

$$\begin{aligned} p_{x,i} &= \frac{i_{11} g_{x,i} + i_{12} g_{y,i} + i_{13} g_{z,i}}{g_{z,3}} \\ p_{y,i} &= \frac{i_{22} g_{x,i} + i_{23} g_{z,i}}{g_{z,i}} \end{aligned}$$

and

$$\begin{aligned} g_{x,i} &= e_{11}s_{x,i} + e_{12}s_{y,i} + e_{13}s_{z,i} + e_{14} \\ g_{y,i} &= e_{21}s_{x,i} + e_{22}s_{y,i} + e_{23}s_{z,i} + e_{24} \\ g_{z,i} &= e_{31}s_{x,i} + e_{32}s_{y,i} + e_{33}s_{z,i} + e_{34} \end{aligned}$$

(which you should check as an exercise). This is a constrained optimization problem, because $\mathcal{T}_e$ is a Euclidean transformation. The constraints here are

$$\begin{array}{c} 1 - \sum_v e_{1v}^2 = 0 \text{ and } 1 - \sum_v e_{2v}^2 = 0 \text{ and } 1 - \sum_v e_{3v}^2 = 0 \\ \sum_v e_{1v}e_{2v} = 0 \text{ and } \sum_v e_{1v}e_{3v} = 0 \text{ and } 1 - \sum_v e_{2v}e_{3v} = 0 \end{array} .$$

We might just throw this into a constrained optimizer (review Section 32.2), but good behavior requires a good start point. This can be obtained by a little manipulation, which I work through in the next section. Some readers may prefer to skip this at first (or even higher) reading because it's somewhat specialized, but it shows how the practical application of some tricks worth knowing.

13.1.2 Setting up a Start Point

Write $\mathbf{C}_j^T$ for the j'th row of the camera matrix, and $\mathbf{S}_i = [s_{x,i}, s_{y,i}, s_{z,i}, 1]^T$ for homogeneous coordinates representing the i'th point in 3D. Then, assuming no errors in measurement, we have

$$\hat{t}_{x,i} = \frac{\mathbf{C}_1^T\mathbf{S}_i}{\mathbf{C}_3^T\mathbf{S}_i} \text{ and } \hat{t}_{y,i} = \frac{\mathbf{C}_2^T\mathbf{S}_i}{\mathbf{C}_3^T\mathbf{S}_i},$$

which we can rewrite as

$$\mathbf{C}_3^T\mathbf{S}_i\hat{t}_{x,i} - \mathbf{C}_1^T\mathbf{S}_i = 0 \text{ and } \mathbf{C}_3^T\mathbf{S}_i\hat{t}_{y,i} - \mathbf{C}_2^T\mathbf{S}_i = 0.$$

We now have two homogenous linear equations in the camera matrix elements for each pair (3D point, image point). There are a total of 12 degrees of freedom in the camera matrix, meaning we can recover a least squares solution from six point pairs. The solution will have the form $\lambda\mathcal{P}$ where λ is an unknown scale and $\mathcal{P}$ is a known matrix. This is a natural consequence of working with homogeneous equations, but also a natural consequence of working with homogeneous coordinates. You should check that if $\mathcal{P}$ is a projection from projective 3D to the projective plane, $\lambda\mathcal{P}$ will yield the same projection as long as $\lambda \neq 0$.

This is enough information to recover the focal point of the camera. Recall that the focal point is the single point that images to $[0,0,0]^T$. This means that if we are presented with a 3×4 matrix claiming to be a camera matrix, we can determine what the focal point of that camera is without fuss – just find the null space of the matrix. Notice that we do not need to know λ to estimate the null space.

Remember this: *Given a 3×4 camera matrix $\mathcal{P}$, the homogeneous coordinates of the focal point of that camera are given by $\mathbf{X}$, where $\mathcal{P}\mathbf{X} = [0, 0, 0]^T$*

There is an important relationship between the focal point of the camera and the extrinsics. Assume that, in the world coordinate system, the focal point can be represented by $\left[\mathbf{f}^T, 1\right]^T$. This point must be mapped to $[0, 0, 0, 1]^T$ by $\mathcal{T}_e$. Because we can recover $\mathbf{f}$ from $\mathcal{P}$ easily, we have an important constraint on $\mathcal{T}_e$, given in the box.

Remember this: *Assume camera matrix $\mathcal{P}$ has null space $\lambda\mathbf{u} = \lambda\left[\mathbf{f}^T, 1\right]^T$. Then we must have $\mathcal{T}_e\mathbf{u} = [0, 0, 0, 1]^T$, so we must have*

$$\mathcal{T}_e = \begin{bmatrix} \mathcal{R} & -\mathcal{R}\mathbf{f} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

This means that, if we know $\mathcal{R}$, we can recover the translation from the focal point. We must now recover the intrinsic transformation and $\mathcal{R}$ from what we know.

$$\lambda\mathcal{P} = \mathcal{T}_i \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{R} & -\mathcal{R}\mathbf{f} \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathcal{T}_i\mathcal{R} & -\mathcal{T}_i\mathcal{R}\mathbf{f} \end{bmatrix}$$

We do not know λ, but we do know $\mathcal{P}$. Now write $\mathcal{P}_l$ for the left 3×3 block of $\mathcal{P}$, and recall that $\mathcal{T}_i$ is upper triangular and $\mathcal{R}$ orthonormal. The first question is the sign of λ. We expect $\mathsf{Det}(\mathcal{R}) = 1$, and $\mathsf{Det}(\mathcal{T}_i) > 0$, so $\mathsf{Det}(\mathcal{P}_l)$ should be positive. This yields the sign of λ – choose a sign $s \in \{-1, 1\}$ so that $\mathsf{Det}(s\mathcal{P}_l)$ is positive.

We can now factor $s\mathcal{P}_l$ into an upper triangular matrix $\mathcal{T}$ and an orthonormal matrix $\mathcal{Q}$. This is an RQ factorization (Section 32.2). Recall we could not distinguish between scaling caused by the focal length and scaling caused by pixel scale, so that

$$\mathcal{T}_i = \begin{bmatrix} as & k & c_x \\ 0 & s & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

In turn, we have $\lambda = s(1/t_{33})$, $c_y = (t_{23}/t_{33})$, $s = (t_{22}/t_{33})$, $c_x = (t_{13}/t_{33})$, $k = (t_{12}/t_{33})$, and $a = (t_{11}/t_{22})$.

Procedure: 13.1 *Calibrating a Camera using 3D Reference Points*

For N reference points $\mathbf{s}_i = [s_{x,i}, s_{y,i}, s_{z,i}]$ with known position in some reference coordinate system in 3D, write the measured location in the image for the i'th such point $\hat{\mathbf{t}}_i = \left[\hat{t}_{x,i}, \hat{t}_{y,i}\right]$. Now minimize

$$\sum_i \xi_i^T \xi_i = \sum_i (\hat{t}_{x,i} - p_{x,i})^2 + (\hat{t}_{y,i} - p_{y,i})^2$$

where

$$\begin{aligned} p_{x,i} &= \frac{i_{11} g_{x,i} + i_{12} g_{y,i} + i_{13} g_{i,3}}{g_{i,3}} \\ p_{y,i} &= \frac{i_{22} g_{x,i} + i_{23} g_{i,3}}{g_{i,3}} \end{aligned}$$

and

$$\begin{aligned} g_{x,i} &= e_{11} s_{x,i} + e_{12} s_{y,i} + e_{13} s_{z,i} + e_{14} \\ g_{y,i} &= e_{21} s_{x,i} + e_{22} s_{y,i} + e_{23} s_{z,i} + e_{24} \\ g_{z,i} &= e_{31} s_{x,i} + e_{32} s_{y,i} + e_{33} s_{z,i} + e_{34} \end{aligned}$$

subject to:

$$1 - \sum_v e_{j,1v}^2 = 0 \text{ and } 1 - \sum_v e_{j,2v}^2 = 0 \text{ and } 1 - \sum_v e_{j,3v}^2 = 0$$

$$\sum_v e_{j,1v} e_{j,2v} = 0 \text{ and } \sum_v e_{j,1v} e_{j,3v} = 0 \text{ and } 1 - \sum_v e_{j,2v} e_{j,3v} = 0 \qquad .$$

Use the start point of procedure 13.2

Procedure: 13.2 *Calibrating a Camera using 3D Reference Points: Start Point*

Estimate the rows of the camera matrix $\mathbf{C}_i$ using at least six points and

$$\mathbf{C}_3^T \mathbf{S}_i \hat{t}_{x,i} - \mathbf{C}_1^T \mathbf{S}_i = 0 \text{ and } \mathbf{C}_3^T \mathbf{S}_i \hat{t}_{y,i} - \mathbf{C}_2^T \mathbf{S}_i = 0.$$

Write $\lambda \mathcal{P}$ for the 1D family of solutions to this set of homogeneous linear equations, organized into 3×4 matrix form. Compute the vector $\mathbf{n} = \left[\mathbf{f}^T, 1\right]$ such that $\mathcal{P}\mathbf{n}$. Write $\mathcal{P}_l$ for the left 3×3 block of $\mathcal{P}$. Choose $s \in \{-1, 1\}$ such that $\mathsf{Det}\,(s\mathcal{P}_l) > 0$. Use RQ factorization to obtain $\mathcal{T}$ and $\mathcal{Q}$ such that $s\mathcal{P}_l = \mathcal{T}\mathcal{Q}$. Then the start point for the intrinsic parameters is:

$$\left[\begin{array}{c} a \\ s \\ k \\ c_x \\ c_y \end{array}\right] = \left[\begin{array}{c} (t_{11}/t_{22}) \\ (t_{22}/t_{33}) \\ (t_{12}/t_{33}) \\ (t_{13}/t_{33}) \\ (t_{23}/t_{33}) \end{array}\right]$$

and for $\mathcal{T}_e$ is:

$$\left[\begin{array}{cc} \mathcal{Q} & -\mathcal{Q}\mathbf{f} \\ \mathbf{0} & 1 \end{array}\right].$$

13.2 CALIBRATING THE EFFECTS OF LENS DISTORTION

Now assume the lens applies some form of geometric distortion, as in Section 32.2. There are now strong standard models of the major lens distortions (Section 32.2). We will now estimate lens parameters, camera intrinsics and camera extrinsics from a view of a calibration object (as in Section 32.2; note the methods of Section 32.2 apply to this problem too). As in those sections, we use a two step procedure: formulate the optimization problem (Section 32.2), then find a good starting point (Section 32.2).

13.2.1 Modelling Geometric Lens Distortion

Geometric distortions caused by lenses are relatively easily modelled by assuming the lens causes (x, y) in the image plane to map to $(x+\delta x, y+\delta y)$ in the image plane. We seek a model for $\delta x, \delta y$ that has few parameters and that captures the main effects. A natural model of barrel distortion is that points are "pulled" toward the camera center, with points that are further from the center being "pulled" more. Similarly, pincushion distortion results from points being "pushed" away from the camera center, with distant points being pushed further (Figure **??**).

Set up a polar coordinate system (r, θ) in the image plane using the image center as the origin. The figure and description suggest that barrel and pincushion distortion can be described by a map $(r, \theta) \to (r + \delta r, \theta)$. We model δr as a polynomial in r. Brown and Conrady [] established the model $\delta r = k_1 r^3 + k_2 r^5$ as sufficient for a wide range of distortions, and we use $(r, \theta) \to (r + k_1 r^3 + k_2 r^5, \theta)$

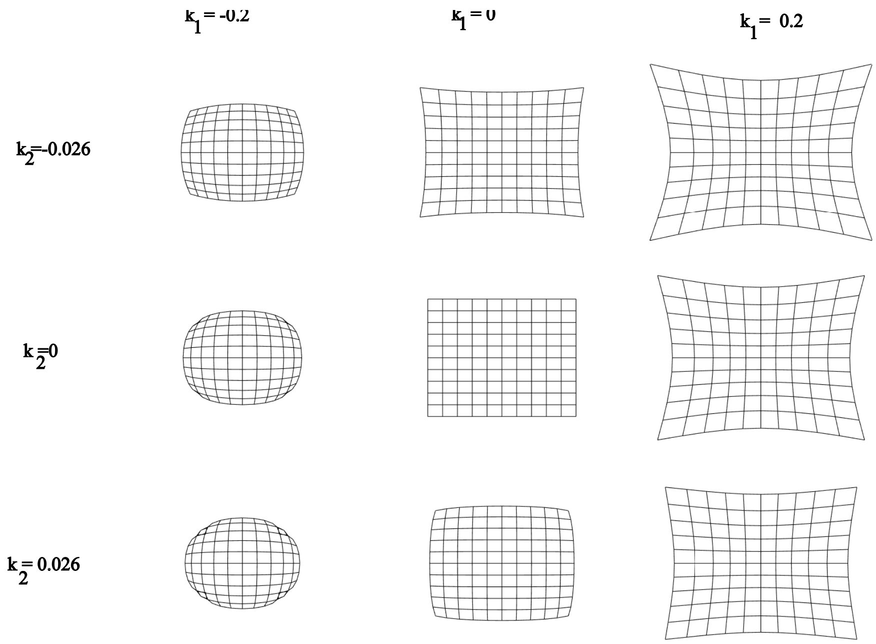

FIGURE 13.1: *The effects of k_1 and k_2 on a neutral grid (**center**), showing how the parameters implement various barrel or pincushion distortions. Notice how k_2 slightly changes the shape of the curves that k_1 produces from straight lines in the grid.*

for unknown k_1, k_2. We must map this model to image coordinates to obtain a map $(x, y) \rightarrow (x + \delta x, y + \delta y)$. Since $\cos\theta = x/r$, $\sin\theta = y/r$, we have $(x, y) \rightarrow (x + x(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2), y + y(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2))$. Figure 13.1 shows distortions resulting from different choices of k_1 and k_2. This model is known as a *radial distortion model.*

More sophisticated lens distortion models account for the lens being off-center. This causes *tangential distortion* (Figure 13.2). The most commonly used model of tangential distortion is a map $(x, y) \rightarrow (x + p_1(x^2 + y^2 + 2x^2) + 2p_2xy, y + p_2(x^2 + y^2 + 2y^2) + 2p_1xy)$ (derived from []; more detail in, for example []).

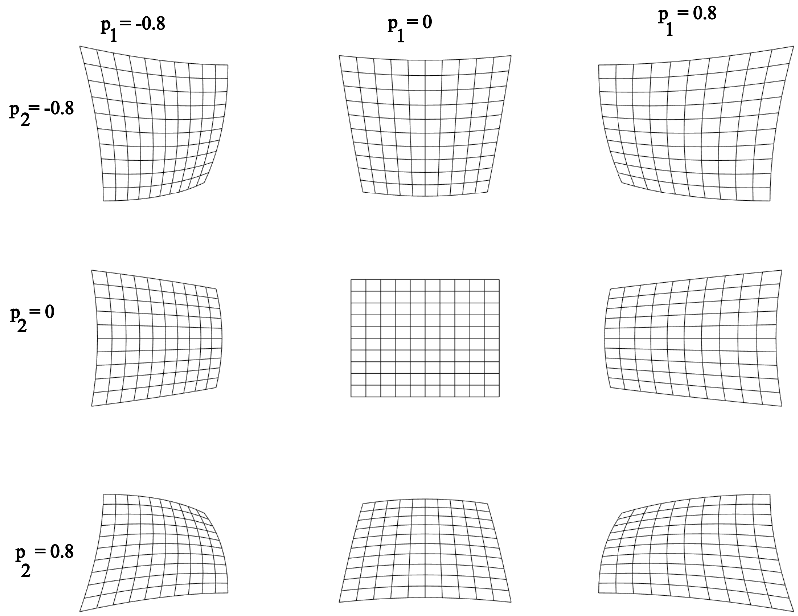

FIGURE 13.2: *The effects of p_1 and p_2 on a neutral grid (**center**), showing how the parameters implement various distortions. These parameters model effects that occur because the lens is off-center; note the grid "turning away" from the lens.*

> **Remember this:** *A full lens distortion model is*
>
> $$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} x + x(k_1(x^2+y^2) + k_2(x^2+y^2)^2) + p_1(x^2+y^2+2x^2) + 2p_2xy \\ y + y(k_1(x^2+y^2) + k_2(x^2+y^2)^2) + p_2(x^2+y^2+2y^2) + 2p_1xy \end{pmatrix}.$$
>
> *for k_1, k_2, p_1, p_2 parameters. It is common to ignore tangential distortion and focus on radial distortion by setting $p_1 = p_2 = 0$.*

13.2.2 Formulating the Optimization Problem

Again, the optimization problem is relatively straightforward to formulate. Write $\hat{\mathbf{t}}_i = [t_{x,i}, t_{y,i}]$ for the measured x, y position in the image plane of the i'th reference point. We have that $\hat{\mathbf{t}}_i = \mathbf{t}_i + \xi_i$, where ξ_i is an error vector and $\mathbf{t}_i$ is the true (unknown) position of the i'th point. Again, assume the error is isotropic, so it is natural to minimize

$$\sum_i \xi_i^T \xi_i.$$

We obtain expressions for $\xi_{i,j}$ in the appropriate coordinates as in Section 32.2, and using the notation of that section, but now accounting for the effects of the lens. We have

$$\sum_i \xi_i^T \xi_i = \sum_i (t_{x,i} - l_{x,i})^2 + (t_{y,i} - l_{y,i})^2$$

where

$$\begin{aligned} l_{x,i} &= p_{x,i} + p_{x,i}(k_1(p_{x,i}^2 + p_{y,i}^2) + k_2(p_{x,i}^2 + p_{y,i}^2)^2) + p_1(p_{x,i}^2 + p_{y,i}^2 + 2p_{x,i}^2) + 2p_2 p_{x,i} p_{y,i} \\ l_{y,i} &= p_{y,i} + p_{y,i}(k_1(p_{x,i}^2 + p_{y,i}^2) + k_2(p_{x,i}^2 + p_{y,i}^2)^2) + p_2(p_{x,i}^2 + p_{y,i}^2 + 2p_{y,i}^2) + 2p_1 p_{x,i} p_{y,i} \end{aligned}$$

(which models the effect of the lens on the imaged points). The imaged points are

$$\begin{aligned} p_{x,i} &= \frac{i_{11} g_{x,i} + i_{12} g_{y,i} + i_{13} g_{z,i}}{g_{z,i}} \\ p_{y,i} &= \frac{i_{22} g_{x,i} + i_{23} g_{z,i}}{g_{z,i}} \end{aligned}$$

and, as before, we have

$$\begin{aligned} g_{x,i} &= e_{11} s_{x,i} + e_{12} s_{y,i} + e_{13} s_{z,i} + e_{14} \\ g_{y,i} &= e_{21} s_{x,i} + e_{22} s_{y,i} + e_{23} s_{z,i} + e_{24} \\ g_{z,i} &= e_{31} s_{x,i} + e_{32} s_{y,i} + e_{33} s_{z,i} + e_{34}. \end{aligned}$$

(which you should check as an exercise). As before, this is a constrained optimization problem, because $\mathcal{T}_e$ is a Euclidean transformation. The constraints here are

$$1 - \sum_v e_{j,1v}^2 = 0 \text{ and } 1 - \sum_v e_{j,2v}^2 = 0 \text{ and } 1 - \sum_v e_{j,3v}^2 = 0$$
$$\sum_v e_{j,1v} e_{j,2v} = 0 \text{ and } \sum_v e_{j,1v} e_{j,3v} = 0 \text{ and } 1 - \sum_v e_{j,2v} e_{j,3v} = 0 \quad .$$

As in Section 32.2, simply dropping this problem into a constrained optimizer is not a particularly good approach. If we assume the lens distortion is minor, we can obtain a start point for the intrinsics and the extrinsics using Section 32.2, and use $k_1 = 0$, $k_2 = 0$, $p_1 = 0$ and $p_2 = 0$.