CHAPTER 24

Using Camera Models

24.1 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 versions of 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.

24.1.1 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

$$\sum_i \xi_i^T \xi_i. \tag{24.1}$$

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

$$\sum_i \xi_i^T \xi_i = \sum_i (t_{x,i} - p_{x,i})^2 + (t_{y,i} - p_{y,i})^2 \tag{24.2}$$

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

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

(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 35.2), 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.

24.1.2 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

$$\hat{t}_{x,i} = \frac{\mathbf{C}_1^T\mathbf{S}_i}{\mathbf{C}_3^T\mathbf{S}_i} \text{ and } \hat{t}_{y,i} = \frac{\mathbf{C}_2^T\mathbf{S}_i}{\mathbf{C}_3^T\mathbf{S}_i}, \tag{24.3}$$

which we can rewrite as

$$\mathbf{C}_3^T\mathbf{S}_i\hat{t}_{x,i} - \mathbf{C}_1^T\mathbf{S}_i = 0 \text{ and } \mathbf{C}_3^T\mathbf{S}_i\hat{t}_{y,i} - \mathbf{C}_2^T\mathbf{S}_i = 0. \tag{24.4}$$

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*

$$\mathcal{T}_e = \begin{bmatrix} \mathcal{R} & -\mathcal{R}\mathbf{f} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{24.5}$$

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.

$$\lambda\mathcal{P} = \mathcal{T}_i \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{R} & -\mathcal{R}\mathbf{f} \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathcal{T}_i\mathcal{R} & -\mathcal{T}_i\mathcal{R}\mathbf{f} \end{bmatrix} \tag{24.6}$$

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 35.2). Recall we could not distinguish between scaling caused by the focal length and scaling caused by pixel scale, so that

$$\mathcal{T}_i = \begin{bmatrix} as & k & c_x \\ 0 & s & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{24.7}$$

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: 24.1 *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

$$\sum_i \xi_i^T \xi_i = \sum_i (\hat{t}_{x,i} - p_{x,i})^2 + (\hat{t}_{y,i} - p_{y,i})^2 \tag{24.8}$$

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 24.2

Procedure: 24.2 *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

$$\mathbf{C}_3^T\mathbf{S}_i\hat{t}_{x,i} - \mathbf{C}_1^T\mathbf{S}_i = 0 \text{ and } \mathbf{C}_3^T\mathbf{S}_i\hat{t}_{y,i} - \mathbf{C}_2^T\mathbf{S}_i = 0. \tag{24.9}$$

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:

$$\left[\begin{array}{c} a \\ s \\ k \\ c_x \\ c_y \end{array}\right] = \left[\begin{array}{c} (t_{11}/t_{22}) \\ (t_{22}/t_{33}) \\ (t_{12}/t_{33}) \\ (t_{13}/t_{33}) \\ (t_{23}/t_{33}) \end{array}\right] \tag{24.10}$$

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

$$\left[\begin{array}{cc} \mathcal{Q} & -\mathcal{Q}\mathbf{f} \\ \mathbf{0} & 1 \end{array}\right]. \tag{24.11}$$

24.2 CALIBRATING THE EFFECTS OF LENS DISTORTION

Now assume the lens applies some form of geometric distortion, as in Section 35.2. There are now strong standard models of the major lens distortions (Section 35.2). We will now estimate lens parameters, camera intrinsics and camera extrinsics from a view of a calibration object (as in Section 35.2; note the methods of Section 35.2 apply to this problem too). As in those sections, we use a two step procedure: formulate the optimization problem (Section 35.2), then find a good starting point (Section 35.2).

24.2.1 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 $(r, \theta) \rightarrow (r + \delta r, \theta)$. 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 $(r, \theta) \rightarrow (r + k_1 r^3 + k_2 r^5, \theta)$

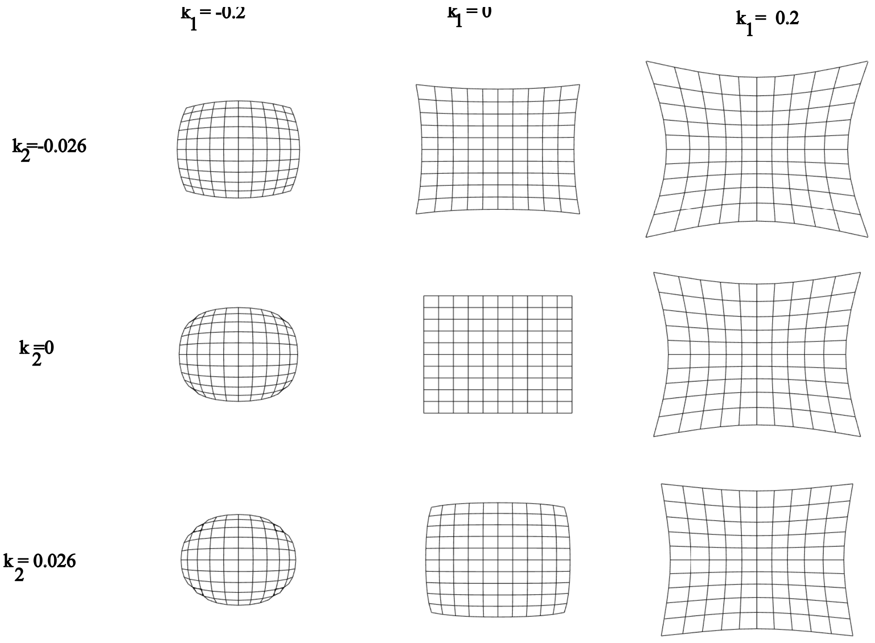

FIGURE 24.1: *The effects of k_1 and k_2 on a neutral grid (***center***), showing how the parameters implement various barrel or pincushion distortions. Notice how k_2 slightly changes the shape of the curves that k_1 produces from straight lines in the grid.*

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 24.1 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 24.2). 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 []).

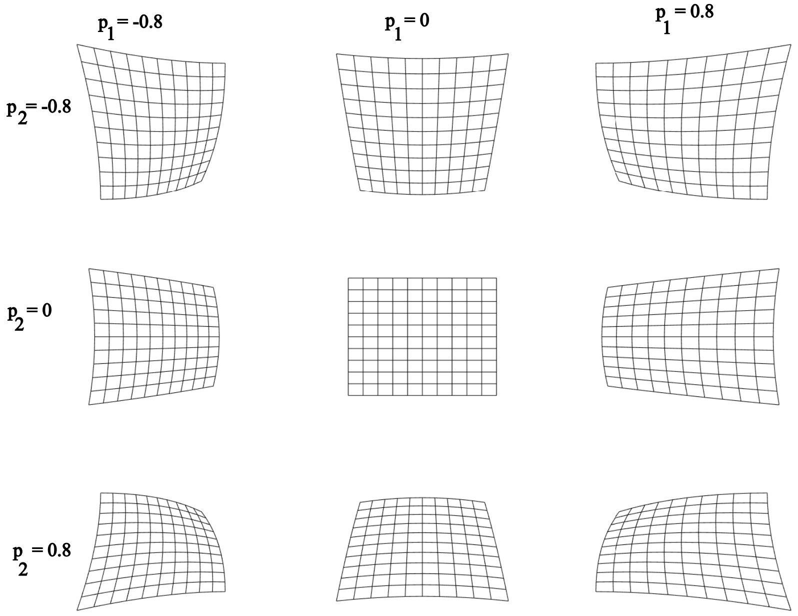

FIGURE 24.2: *The effects of p_1 and p_2 on a neutral grid (***center***), showing how the parameters implement various distortions. These parameters model effects that occur because the lens is off-center; note the grid "turning away" from the lens.*

Remember this: *A full lens distortion model is*

$$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} x + x(k_1(x^2+y^2) + k_2(x^2+y^2)^2) + p_1(x^2+y^2+2x^2) + 2p_2xy \\ y + y(k_1(x^2+y^2) + k_2(x^2+y^2)^2) + p_2(x^2+y^2+2y^2) + 2p_1xy \end{pmatrix}. \tag{24.12}$$

for k_1, k_2, p_1, p_2 parameters. It is common to ignore tangential distortion and focus on radial distortion by setting $p_1 = p_2 = 0$.

24.2.2 Lens Calibration: 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

$$\sum_i \xi_i^T \xi_i. \tag{24.13}$$

We obtain expressions for $\xi_{i,j}$ in the appropriate coordinates as in Section 35.2, and using the notation of that section, but now accounting for the effects of the lens. We have

$$\sum_i \xi_i^T \xi_i = \sum_i (t_{x,i} - l_{x,i})^2 + (t_{y,i} - l_{y,i})^2 \tag{24.14}$$

where

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

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

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

and, as before, we have

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

(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

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

As in Section 35.2, 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 35.2. We then use those parameters, together with $k_1 = 0$, $k_2 = 0$, $p_1 = 0$ and $p_2 = 0$, as a start point.

24.3 "FOUND" CALIBRATIONS

A useful set of camera intrinsics can be recovered from assumptions about the 3D world seen in a picture, assuming that vanishing points can be recovered in an image. One standard assumption is that we are viewing a *Manhattan world*. In this case, there are three cardinal directions. Each is normal to the other two (so you could make them be the x, y, and z directions in an appropriately chosen world coordinate system). Every line in the world is parallel to one of the three cardinal directions.

In a Manhattan world geometry, every image contains at most three families of line, one for each cardinal direction. Each family has a single vanishing point. Assume that there are lines from all families present in an image, and that the vanishing points can be found (Section 35.2). Then the vanishing points reveal camera intrinsics.

This is easily demonstrated in homogeneous coordinates. The three cardinal directions can be defined by their vanishing points *in 3D*. For simplicity, we assume that the cardinal directions are the x, y and z directions in the world coordinate system. In homogeneous coordinates, the collection of all lines parallel to the x axis can be written as $(t, a, b, 1)$, where t is a parameter along the line and a, b select the particular line. All such lines contain the point $\mathbf{E}_x = (1, 0, 0, 0)$ – this is the point at infinity where all of these lines intersect. Similarly, the point at infinity where all lines parallel to the y (resp. z) axis intersect is $\mathbf{E}_y = (0, 1, 0, 0)$ (resp. $\mathbf{E}_z = (0, 0, 1, 0)$.

There will be three vanishing points in the image, $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$. The coordinates of these points are known, because we have detected them, and we represent these points in homogeneous coordinates. Each is the image of one of the 3D points at infinity. We *choose* $\mathbf{e}_1$ as the image of $\mathbf{E}_x$, etc. Note that there are six possible choices. Three of them will result in rotations of the world coordinates, and the other three will result in improper rotations (a rotation, followed by flipping the z-axis direction). For the moment, we ignore the consequences of choosing.

Now recall the parametrization of a camera as $[\mathcal{KR}\mathbf{t}]$. Furthermore, notice that each of the $\mathbf{E}$'s has a zero in the fourth component. We have that

$$\mathbf{E}_x = \left(\begin{array}{c} (\mathcal{KR})^{-1}\mathbf{e}_1 \\ 0 \end{array} \right) \text{ and } \mathbf{E}_y = \left(\begin{array}{c} (\mathcal{KR})^{-1}\mathbf{e}_2 \\ 0 \end{array} \right) \text{ and } \mathbf{E}_z = \left(\begin{array}{c} (\mathcal{KR})^{-1}\mathbf{e}_3 \\ 0 \end{array} \right).$$

The cardinal directions are at right angles, so we have three constraints

$$\begin{array}{rcl} \mathbf{E}_x^T\mathbf{E}_y & = & \mathbf{e}_1\mathcal{K}^{-T}\mathcal{K}\mathbf{e}_2 \\ & = & 0 \\ \mathbf{E}_x^T\mathbf{E}_z & = & \mathbf{e}_1\mathcal{K}^{-T}\mathcal{K}\mathbf{e}_3 \\ & = & 0 \\ \mathbf{E}_y^T\mathbf{E}_z & = & \mathbf{e}_2\mathcal{K}^{-T}\mathcal{K}\mathbf{e}_3 \\ & = & 0 \end{array}$$

(notice that the rotations cancel because $\mathcal{R}^T = \mathcal{R}^{-1}$). Now assume that the aspect ratio a is 1, and the skew k is 0. In this case,

$$\mathcal{K} = \left(\begin{array}{ccc} s & 0 & c_x \\ 0 & s & c_y \\ 0 & 0 & 1 \end{array} \right).$$

Now write the known coordinates of the vanishing points $\mathbf{e}_1 = (e_{1x}, e_{2x}, 1)^T$ and so on. Some algebra yields the constraints

$$\begin{array}{rcl}
[e_{1x}e_{2x}+e_{1y}e_{2y}]+c_x\,[1-e_{1x}-e_{2x}]+c_y\,[1-e_{1y}-e_{2y}]+s^2+c_x^2+c_y^2 & = & 0\\
[e_{1x}e_{3x}+e_{1y}e_{3y}]+c_x\,[1-e_{1x}-e_{3x}]+c_y\,[1-e_{1y}-e_{3y}]+s^2+c_x^2+c_y^2 & = & 0\\
[e_{2x}e_{3x}+e_{2y}e_{3y}]+c_x\,[1-e_{2x}-e_{3x}]+c_y\,[1-e_{2y}-e_{3y}]+s^2+c_x^2+c_y^2 & = & 0
\end{array}$$

It is straightforward to extract c_x, c_y and s from these equations (exercises).

If one image vanishing point is at infinity (choose $\mathbf{e}_1$, so that $\mathbf{e}_1=(e_{1x},e_{2x},0)^T$), we have

$$\begin{array}{rcl}
[e_{1x}e_{2x}+e_{1y}e_{2y}]+c_x\,[-e_{1x}-e_{2x}]+c_y\,[-e_{1y}-e_{2y}] & = & 0\\
[e_{1x}e_{3x}+e_{1y}e_{3y}]+c_x\,[-e_{1x}-e_{3x}]+c_y\,[-e_{1y}-e_{3y}] & = & 0\\
[e_{2x}e_{3x}+e_{2y}e_{3y}]+c_x\,[1-e_{2x}-e_{3x}]+c_y\,[1-e_{2y}-e_{3y}]+s^2+c_x^2+c_y^2 & = & 0
\end{array}$$

and we can still solve for c_x, c_y and s.

24.4 MEASURING LENGTHS IN A SINGLE VIEW

In many very useful cases, we can measure lengths in a single *uncalibrated* image.

24.4.1 Exploiting a Ruler

In the simplest, the image shows an object on a ground plane, there is a ruler conveniently standing normal to the ground plane, and we wish to measure the height of the object. This is shown in Figure 24.3, and I use the notation of that figure. Construct the line bB (from the base of the ruler to the base of the object) and intersect that line with the horizon to get vanishing point V. Now construct the line VT from that vanishing point to the top of the object, and intersect that line with the ruler. The intersection point with the ruler shows the height of the object.

A ruler perpendicular to a ground plane can reveal the height of the camera above the ground plane, too. The horizon in the image is formed by the plane through the camera focal point and parallel to the ground plane. But this plane must intersect the ruler at the height of the focal point above the ground plane. This means that the horizon intersects the ruler at the height of the focal point above the ground plane.

24.4.2 Measuring with a Reference Object

A ruler has the special property that it has lengths marked on it. Imagine we now have a reference line segment that has known height, but doesn't have other heights marked on it. Using this reference object takes care, because (say) the midpoint of the reference object *in the image* may not lie halfway up the reference object in 3D (Figure 24.5).

This is because the transformation from the reference line segment in 3D to the image is not an affine transformation – it is a projective transformation. Figure 24.5 sketches the geometry. Parametrize the reference line segment in 3D using affine coordinates to get $\mathbf{p}+t\mathbf{d}$, where $\mathbf{d}$ is a unit vector (so a step of 1 in t is

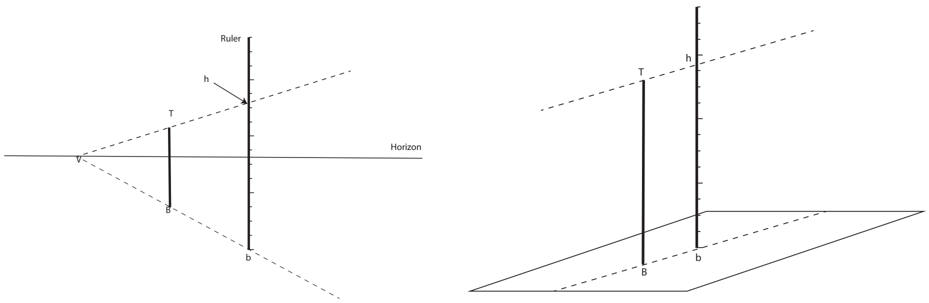

FIGURE 24.3: **Left**, *an image of a ruler and an object, which just happen to be standing perpendicular to a ground plane. In an* uncalibrated *image like this, we can measure the height of the object. Construct the line bB, and intersect that with the horizon to get the point V. The line from the top of the object T to the true height of the object on the ruler (h) is parallel in 3D to bB. In turn, the line Th must intersect the horizon at V. So if you construct VT, it will intersect the ruler at h yielding the height of the object.* **Right** *shows a 3D view; the line Th must be parallel to bB, and so in the image these two lines intersect at the horizon.*

a step of length 1 along the reference segment). Write c_{ij} for the i, j'th component of the 3×4 camera matrix. Then the homogeneous coordinates for the image line will be

$$\left(\begin{array}{c} (c_{11}p_1 + c_{12}p_2 + c_{13}p_3 + c_{14}) + t(c_{11}d_1 + c_{12}d_2 + c_{13}d_3 + c_{14}) \\ (c_{21}p_1 + c_{22}p_2 + c_{23}p_3 + c_{24}) + t(c_{21}d_1 + c_{22}d_2 + c_{23}d_3 + c_{24}) \\ (c_{31}p_1 + c_{32}p_2 + c_{33}p_3 + c_{34}) + t(c_{31}d_1 + c_{32}d_2 + c_{33}d_3 + c_{34}) \end{array}\right) = \left(\begin{array}{c} a + bt \\ c + dt \\ e + ft \end{array}\right).$$

SInce we know the image is a line, we can ignore one of these three homogeneous coordinates, so the transformation is a projective transformation. Now on the 3D reference line segment, the points $t = 0$ and $t = 1$ are the same distance apart as the points $t = 1$ and $t = 2$. But in the image line, using affine coordinates, these points are

$$\frac{a}{c}, \frac{a+b}{c+d}, \frac{a+2b}{c+2d}$$

which are not, in general, evenly spaced (check this with, for example, $a = 0$, $b = 1$, $c = 1$, $d = 1$).

A clever trick from projective geometry allows us to use a reference object to measure heights. Write $\mathbf{P}_1, \ldots, \mathbf{P}_4$ for the coordinates of four points on a projective line, written in homogeneous coordinates. Write $\mathcal{M}$ for a projective transformation of the line to itself (so a 2×2 matrix with non-zero determinant. Finally, write

$$d(\mathbf{P}_i, \mathbf{P}_j) = \det\left([\mathbf{P}_i \mathbf{P}_j]\right).$$

Notice that

$$\det\left([\mathcal{M}\mathbf{P}_i \mathcal{M}\mathbf{P}_j)]\right) = \det\left(\mathcal{M}[\mathbf{P}_i \mathbf{P}_j]\right) = \det\left(\mathcal{M}\right) \det\left([\mathbf{P}_i \mathbf{P}_j]\right)$$

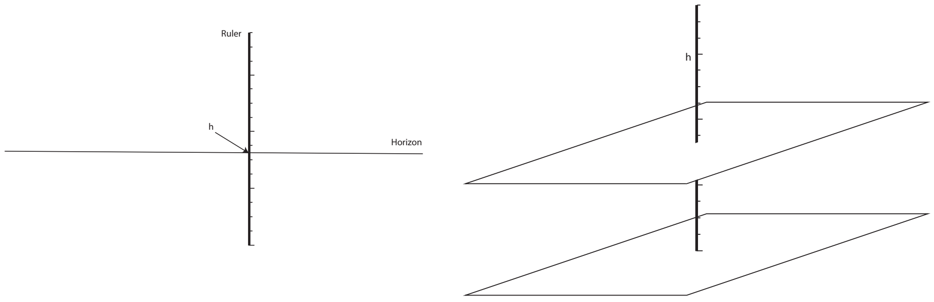

FIGURE 24.4: **Left**, *an image of a ruler which just happens to be standing perpendicular to a ground plane. In an* uncalibrated *image like this, we can measure the height of the camera focal point above the ground plane. The plane through the focal point parallel to the ground plan (and so the same height above the ground plane as the focal point) must form the horizon, so the intersection between horizon and ruler yields the height of the focal point.* **Right** *shows a 3D view; the bottom plane is the ground plane, and the top plane is the plane through the focal point parallel to the ground plane.*

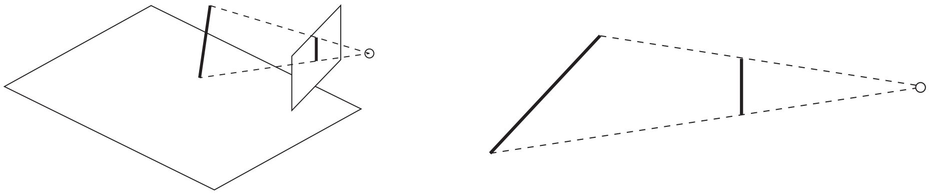

FIGURE 24.5: **Left**, *a perspective camera views a reference object perpendicular to a ground plane. This produces a line segment in the image plane.* **Right** *shows the reference object and the line segment in the image plane.*

which means that

$$\frac{d(\mathbf{P}_1, \mathbf{P}_2) d(\mathbf{P}_3, \mathbf{P}_4)}{d(\mathbf{P}_1, \mathbf{P}_3) d(\mathbf{P}_2, \mathbf{P}_4)}$$

is a *projective invariant* — computing the value of this *cross ratio* using $\mathbf{P}_1, \ldots, \mathbf{P}_4$ or using $\mathcal{M}\mathbf{P}_1, \ldots, \mathcal{M}\mathbf{P}_4$ will yield the same number, as long as $\mathcal{M}$ is a projective transformation.

Now check that the cross-ratio of the four points $(0, 1)$, $(a, 1)$, $(b, 1)$ and $(1, 0)$ is a/b (notice the last point is the point at infinity). We can use this observation to measure height relative to a reference object. Using the notation of Figure 24.6, we construct the line Bb from the base of the object to the base of the reference object. Produce this line to intersect the horizon at V. Now construct VT, which intersects the reference object at h. In 3D, the line VT is parallel to the ground plane, so that the point h in 3D is the same height above the ground plane as the point T in

3D. The vanishing point for the vertical lines (the object and the reference object) is at infinity in this image, *so we know where it lies on line bt.* Write P for this vanishing point, r for the height of the reference object and o for the height of the object. Then we have

$$\frac{d(b,h)d(t,P)}{d(b,t),b(h,P)}=\frac{r}{o}$$

but we *know* the height of the reference object and we can measure the cross ratio, so we can recover o.

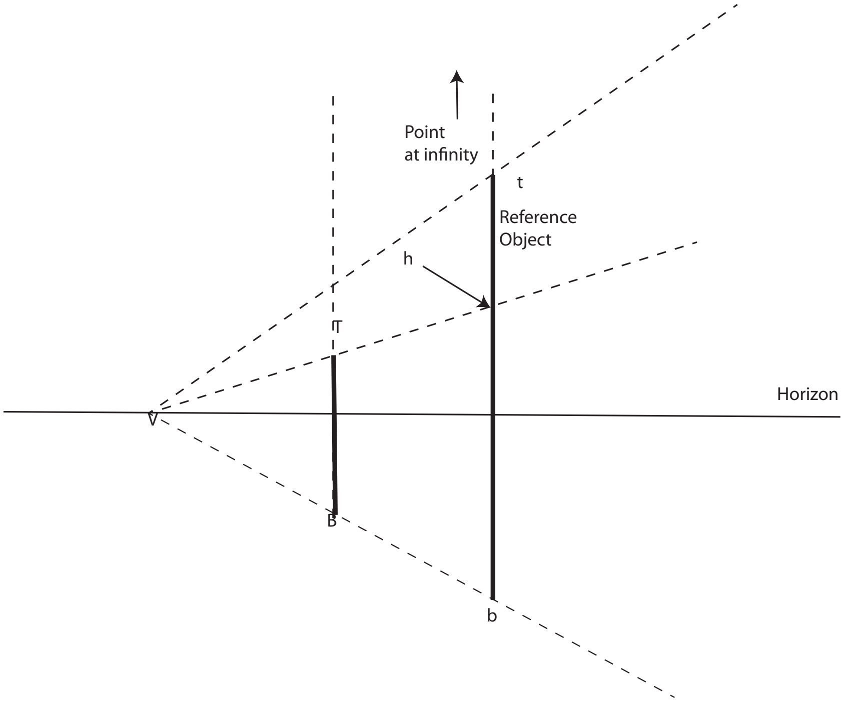

FIGURE 24.6: *A perspective camera views a reference object and another object perpendicular to a ground plane. This produces a line segment in the image plane. Constructing appropriate lines in the figure and taking a cross ratio yields the height of the object.*

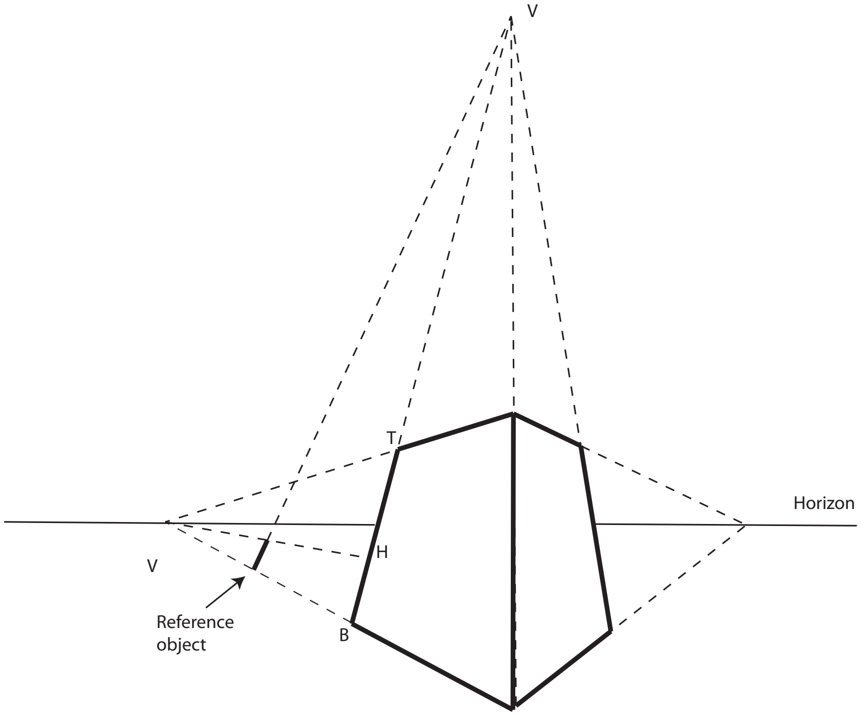

FIGURE 24.7: *A building and a person viewed in a more extreme perspective view than that of 24.6. The person has known height, and can act as reference object. The same construction as in that figure yields the height of the building relative to the person.*