5.1 Camera Model †
5.2 Calibration †
5.3 Lens Distortion †
5.4 Rectification
5.5 Composite
5.6 Reflection
5.7 Shadow
5.8 Reflection Perception
5.9 Shadow Perception
5.10 Problem Set
5.11 Solutions
5.12 Readings

5.1 Camera Model †

Under an ideal pinhole camera model, the perspective projection of arbitrary points $\mathbf{P}$ in 3-D world coordinates is given, in homogeneous coordinates, by:

$$\begin{pmatrix} x \\ y \\ s \end{pmatrix} = \lambda \begin{pmatrix} \alpha f & \beta & c_1 \\ 0 & f & c_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R \mid \mathbf{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{5.1}$$

$$\mathbf{p} = \lambda K M \mathbf{P}, \tag{5.2}$$

where $\mathbf{p}$ is the 2-D projected point in homogeneous coordinates, which in non-homogeneous coordinates is $(\, x/s \quad y/s \,)$, λ is a scale factor, K is the intrinsic matrix, and M is the extrinsic matrix. Within the extrinsic matrix, R is a 3×3 rotation matrix, and $\mathbf{t}$ is a 3×1 translation vector. Within the intrinsic matrix, f is the focal length, α is the aspect ratio, (c_1, c_2) is the principle point (the projection of the camera center onto the image plane), and β is the skew.

For simplicity, it is typically assumed that the pixels are square ($\alpha = 1, \beta = 0$). This is a reasonable assumptions for most modern-day cameras. With this assumption, the intrinsic matrix simplifies to:

$$K = \begin{pmatrix} f & 0 & c_1 \\ 0 & f & c_2 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.3}$$

The camera model in Equation (5.2) specifies the perspective projection of arbitrary 3-D points from world to image coordinates. In the special case when the world points are constrained to a planar surface, the projection takes the form:

$$\begin{pmatrix} x \\ y \\ s \end{pmatrix} = \lambda \begin{pmatrix} f & 0 & c_1 \\ 0 & f & c_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 & r_2 & \mathbf{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{5.4}$$

$$\mathbf{p} = \lambda K M \mathbf{P} \tag{5.5}$$

$$\mathbf{p} = H \mathbf{P}, \tag{5.6}$$

where $\mathbf{p}$ is the 2-D projected point in homogeneous coordinates, and $\mathbf{P}$, in the appropriate coordinate system, is specified by 2-D coordinates in homogeneous coordinates. As before, λ is a scale factor, K is the intrinsic matrix, M is a now 3×3 extrinsic matrix in which $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_1 \times \mathbf{r}_2$ are the columns of the 3×3 rotation matrix that describes the transformation from world to camera coordinates, and as before, $\mathbf{t}$ is a 3×1 translation vector. The 3×3 matrix H, referred to as a homography, embodies the projection of a planar surface.

Example 5.1 Show that if Z is a constant in Equation (5.2), then this imaging model is the same as the model in Equation (5.6).

Recall from the previous section that a planar homography H is a scaled product of an intrinsic, K, and extrinsic, M, matrix: $H = \lambda KM$. It can be desirable to factor a homography into these components in order to determine the intrinsic camera parameters.

It is straightforward to show that $\mathbf{r}_1 = \frac{1}{\lambda} K^{-1}\mathbf{h}_1$ and $\mathbf{r}_2 = \frac{1}{\lambda} K^{-1}\mathbf{h}_2$ where $\mathbf{h}_1$ and $\mathbf{h}_2$ are the first two columns of the matrix H. The constraint that $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthogonal (they are columns of a rotation matrix) and have the same norm (unknown due to the scale factor λ) yields two constraints on the unknown intrinsic matrix K:

$$\begin{aligned} \mathbf{r}_1^T \mathbf{r}_2 &= 0 \\ \mathbf{h}_1^T (K^{-T} K^{-1}) \mathbf{h}_2 &= 0, \end{aligned} \tag{5.7}$$

and

$$\begin{aligned} \mathbf{r}_1^T \mathbf{r}_1 - \mathbf{r}_2^T \mathbf{r}_2 &= 0 \\ \mathbf{h}_1^T (K^{-T} K^{-1}) \mathbf{h}_1 - \mathbf{h}_2^T (K^{-T} K^{-1}) \mathbf{h}_2 &= 0. \end{aligned} \tag{5.8}$$

With only two constraints, it is possible to estimate the principal point (c_1, c_2) or the focal length f, but not both. If, however, the focal length is known, then it is possible to estimate the principal point.

For notational simplicity we solve for the components of $Q = K^{-T} K^{-1}$, which contain the desired coordinates of the principal point and the assumed known focal length:

$$Q = \frac{1}{f^2} \begin{pmatrix} 1 & 0 & -c_1 \\ 0 & 1 & -c_2 \\ -c_1 & -c_2 & c_1^2 + c_2^2 + f^2 \end{pmatrix}. \tag{5.9}$$

In terms of Q, the first constraint, Equation (5.7), takes the form:

$$\begin{aligned} h_1 h_2 + h_4 h_5 - (h_2 h_7 + h_1 h_8) c_1 - (h_5 h_7 + h_4 h_8) c_2 & \\ + h_7 h_8 (c_1^2 + c_2^2 + f^2) &= 0, \end{aligned} \tag{5.10}$$

Note that this constraint is a second-order polynomial in the coordinates of the principal point, which can be factored as follows:

$$(c_1 - \alpha_1)^2 + (c_2 - \beta_1)^2 = \gamma_1^2, \tag{5.11}$$

where:

$$\alpha_1 = (h_2 h_7 + h_1 h_8)/(2 h_7 h_8), \tag{5.12}$$

$$\beta_1 = (h_5 h_7 + h_4 h_8)/(2 h_7 h_8), \tag{5.13}$$

$$\gamma_1^2 = \alpha_1^2 + \beta_1^2 - f^2 - (h_1 h_2 + h_4 h_5)/(h_7 h_8). \tag{5.14}$$

Similarly, the second constraint, Equation (5.8), takes the form:

$$\begin{aligned} h_1^2 + h_4^2 + 2(h_2h_8 - h_1h_7)c_1 + 2(h_5h_8 - h_4h_7)c_2 \\ -h_2^2 - h_5^2 + (h_7^2 - h_8^2)(c_1^2 + c_2^2 + f^2) &= 0, \end{aligned} \quad (5.15)$$

or,

$$(c_1 - \alpha_2)^2 + (c_2 - \beta_2)^2 = \gamma_2^2, \quad (5.16)$$

where:

$$\alpha_2 = (h_1h_7 - h_2h_8)/(h_7^2 - h_8^2), \quad (5.17)$$

$$\beta_2 = (h_4h_7 - h_5h_8)/(h_7^2 - h_8^2), \quad (5.18)$$

$$\gamma_2^2 = \alpha_2^2 + \beta_2^2 - (h_1^2 + h_4^2 - h_2^2 - h_5^2)/(h_7^2 - h_8^2) - f^2. \quad (5.19)$$

Both constraints, Equations (5.11) and (5.16) are circles in the desired coordinates of the principal point c_1 and c_2, and the solution is the intersection of the two circles.

For certain homographies this solution can be numerically unstable. For example, if $h_7 \approx 0$ or $h_8 \approx 0$, the first constraint becomes numerically unstable. Similarly, if $h_7 \approx h_8$, the second constraint becomes unstable. In order to avoid these instabilities, an error function with a regularization term can be introduced. We start with the following error function to be minimized:

$$E(c_1, c_2) = g_1(c_1, c_2)^2 + g_2(c_1, c_2)^2, \quad (5.20)$$

where $g_1(c_1, c_2)$ and $g_2(c_1, c_2)$ are the constraints on the principal point given in Equations (5.10) and (5.15), respectively. To avoid numerical instabilities, a regularization term is added to penalize deviations of the principal point from the image center $(0, 0)$ (in normalized coordinates). This augmented error function takes the form:

$$E(c_1, c_2) = g_1(c_1, c_2)^2 + g_2(c_1, c_2)^2 + \Delta(c_1^2 + c_2^2), \quad (5.21)$$

where Δ is a scalar weighting factor. This error function is a nonlinear least-squares problem, which can be minimized using a Levenberg-Marquardt iteration. The image center $(0, 0)$ is used as the initial condition for the iteration.

If, on the other hand, we assume that the principal point is the image center $(0, 0)$, then the focal length f can be estimated. In this case, the intrinsic matrix simplifies to:

$$K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (5.22)$$

to yield a homography of the form:

$$H = \lambda \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} (\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}) . \tag{5.23}$$

Left-multiplying by K^{-1} yields:

$$\begin{pmatrix} \frac{1}{f} & 0 & 0 \\ 0 & \frac{1}{f} & 0 \\ 0 & 0 & 1 \end{pmatrix} H = \lambda (\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}) \tag{5.24}$$

$$\begin{pmatrix} \frac{1}{f} & 0 & 0 \\ 0 & \frac{1}{f} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} = \lambda (\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}) \tag{5.25}$$

As before, because $\mathbf{r}_1$ and $\mathbf{r}_2$ are the first two columns of a rotation matrix, their inner product, $\mathbf{r}_1^T \cdot \mathbf{r}_2$, is zero, leading to the following constraint:

$$\left[\begin{pmatrix} \frac{1}{f} & 0 & 0 \\ 0 & \frac{1}{f} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_4 \\ h_7 \end{pmatrix} \right]^T \cdot \left[\begin{pmatrix} \frac{1}{f} & 0 & 0 \\ 0 & \frac{1}{f} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h_2 \\ h_5 \\ h_8 \end{pmatrix} \right] = 0 \tag{5.26}$$

$$(h_1 \quad h_4 \quad h_7) \begin{pmatrix} \frac{1}{f^2} & 0 & 0 \\ 0 & \frac{1}{f^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h_2 \\ h_5 \\ h_8 \end{pmatrix} = 0. \tag{5.27}$$

The focal length is estimated by solving the above linear system for f:

$$f = \sqrt{-\frac{h_1 h_2 + h_4 h_5}{h_7 h_8}}. \tag{5.28}$$

The additional constraint that $\mathbf{r}_1$ and $\mathbf{r}_2$ are each unit length, $\mathbf{r}_1^T \cdot \mathbf{r}_1 = \mathbf{r}_2^T \cdot \mathbf{r}_2$, can also be used to estimate the focal length.

Example 5.2 The scale factor λ can be determined by exploiting the unit norm constraint on the columns of the rotation matrix. Describe how to estimate this scale factor.

If there is no relative rotation between the world and camera coordinate systems, then there is an inherent ambiguity between the world to camera translation in X and Y and the position of the principal point, and between the translation in Z (depth) and the focal length. As such, the factorization of the homography is not unique in the case of a fronto-parallel view.

5.3 *Lens Distortion*†

The imaging model described in the previous two sections assumes an idealized pinhole camera. In practice, however, cameras have multiple lenses that can deviate substantially from this model. Most significantly, lenses introduce geometric distortion whereby straight lines in the world appear curved in the image.

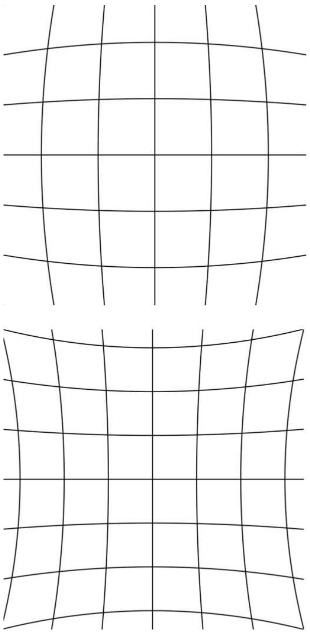

Figure 5.1 Barrel and pin cushion lens distortion.

Geometric lens distortions can be modeled with a one-parameter radially symmetric model. Given an ideal undistorted image $f_u(x, y)$, the distorted image is denoted as $f_d(\tilde{x}, \tilde{y})$, where the distorted spatial parameters are given by:

$$\tilde{x} = x + \kappa x r^2 \quad \text{and} \quad \tilde{y} = y + \kappa y r^2, \tag{5.29}$$

where $r^2 = x^2 + y^2$, and κ controls the amount of distortion. Shown in Figure 5.1 are the results of distorting a rectilinear grid with a negative (barrel distortion) and positive (pincushion distortion) value of κ.

This model assumes that the center of the image coincides with the principal axis of the lens. If, however, this is not the case, then it is necessary to add additional parameters to the model to account for a spatial offset of the distortion center. This new model takes the form:

$$\tilde{x} = x + \kappa(x - c_x)r^2 \quad \text{and} \quad \tilde{y} = y + \kappa(y - c_y)r^2, \tag{5.30}$$

where $r^2 = (x - c_x)^2 + (y - c_y)^2$, and (c_x, c_y) corresponds to the center of the distortion (i.e., the principal point).

Lens distortion should be removed when considering the geometric techniques described in this chapter so that the image formation more closely matches Equation (5.2). Lens distortion can be manually estimated and removed by distorting an image according to Equation (5.29) or (5.30) until lines that are known to be straight in the world appear straight in the image. For the one-parameter model, this is relatively easy to do. For the three-parameter model, however, this manual calibration can be difficult and should be automated. This can be done by specifying curved lines in the image that are known to be straight in the world, and searching the three model parameters until the curved lines are mapped to straight lines.

Note that, unlike the 3×4 projection matrix in Equation (5.2), the 3×3 planar homography in Equation (5.6) is invertible. This implies that if the homography H can be estimated, then the original world coordinates $\mathbf{P}$ can be determined from the projected image coordinates $\mathbf{p}$.

It is straight-forward to see that $\mathbf{p} \times H\mathbf{P} = \mathbf{0}$, where $\times$ denotes cross product. Specifically, the cross product is defined as:

$$\mathbf{a} \times \mathbf{b} \quad = \quad \mathbf{n}\|a\|\|b\|\sin(\theta), \tag{5.31}$$

where $\mathbf{n}$ is mutually orthogonal to $\mathbf{a}$ and $\mathbf{b}$, and θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. If $\mathbf{a} = \mathbf{b}$, then $\theta = 0$ and $\mathbf{a} \times \mathbf{b} = 0$. This identity yields the following:

$$\mathbf{p} \times H\mathbf{P} \quad = \quad \mathbf{0} \tag{5.32}$$

$$\begin{pmatrix} x \\ y \\ s \end{pmatrix} \times \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad = \quad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.33}$$

$$\begin{pmatrix} x \\ y \\ s \end{pmatrix} \times \begin{pmatrix} h_1X + h_2Y + h_3Z \\ h_4X + h_5Y + h_6Z \\ h_7X + h_8Y + h_9Z \end{pmatrix} \quad = \quad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{5.34}$$

where it can be assumed that $Z = 1$ and $s = 1$ because the homography will only be estimated to within an unknown scale factor. Evaluating the cross product on the left-hand side yields:

$$\begin{pmatrix} y(h_7X + h_8Y + h_9Z) - s(h_4X + h_5Y + h_6Z) \\ s(h_1X + h_2Y + h_3Z) - x(h_7X + h_8Y + h_9Z) \\ x(h_4X + h_5Y + h_6Z) - y(h_1X + h_2Y + h_3Z) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.35}$$

Re-ordering the terms yields a linear system $A\mathbf{h} = \mathbf{0}$, where the matrix A is:

$$\begin{pmatrix} 0 & 0 & 0 & -sX & -sY & -sZ & yX & yY & yZ \\ sX & sY & sZ & 0 & 0 & 0 & -xX & -xY & -xZ \\ -yX & -yY & -yZ & xX & xY & xZ & 0 & 0 & 0 \end{pmatrix}, \tag{5.36}$$

and:

$$\mathbf{h} = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 \end{pmatrix}^T \tag{5.37}$$

Given the known coordinates of a point, $\mathbf{P}$, on a plane in the world and its corresponding projected coordinates, $\mathbf{x}$, the above system seemingly provides three constraints in the nine unknowns of $\mathbf{h}$. Note, however, that the rows of the matrix A are not linearly independent (the third row is a linear combination of the first two rows). As such, this system provides only two constraints in the nine unknowns. Because the homography can only be estimated to within an unknown scale factor, the number of unknowns reduces from nine to eight.

As such, in order to solve for the projective transformation matrix H, four or more points with known coordinates $\mathbf{P}$ and $\mathbf{p}$ are required. The coordinates of these points are placed into the rows of matrix A to yield the following quadratic error function to be minimized:

$$E(\mathbf{h}) \quad = \quad \|A\mathbf{h}\|^2. \tag{5.38}$$

Note that minimizing this function using least-squares will lead to the degenerate solution $\mathbf{h} = 0$. In order to avoid this degenerate solution we constrain $\mathbf{h}$ to have unit sum $\|\mathbf{h}\|^2 = 1$ (hence the scale ambiguity in estimating the homography). This added constraint yields a total least-squares optimization. The optimal unit vector $\mathbf{h}$ is the minimal eigenvalue eigenvector of $A^T A$.

With a known projective transformation matrix H, an image can be warped according to H^{-1} to yield a rectified image, Figure 5.2.

Figure 5.2 A an original photo (top), a magnified view of the license plate (middle), and the planar rectified license plate (bottom).

Although the estimation of H is straight-forward, there are a few implementation details that should be considered. For the sake of numerical stability, it is recommended that the image coordinates $\mathbf{p}$ and world coordinates $\mathbf{P}$ are transformed so that their respective centroids are at the origin and that their respective mean distance from the origin is $\sqrt{2}$. In homogeneous coordinates, this transformation matrix takes the form:

$$T \quad = \quad \begin{pmatrix} \alpha & 0 & -\alpha c_1 \\ 0 & \alpha & -\alpha c_2 \\ 0 & 0 & 1 \end{pmatrix}, \tag{5.39}$$

where α is the multiplicative scaling and c_1 and c_2 are the additive offsets. The homography H estimated using these normalized coordinates is then transformed back to the original coordinates as $T_1^{-1} H T_2$, where T_1 and T_2 are the transformation matrices for the image and world coordinates, respectively.

Example 5.3 Show that after the image and world coordinates are transformed by T_1 and T_2, Equation (5.39), the estimated homography H should be transformed by $T_1^{-1} H T_2$.
