Table of Contents

Camera Calibration

The key idea behind calibration is to use the projection equations relating the known coordinates of a set of 3D positions and their projections in order to determine the camera parameters. In order to know the 3D coordinates of some points, we use calibration patterns which usually are planar surfaces with calibration patterns printed on.

Consider a 3D point in world coordinates $\vec{P}=(X,Y,Z)^T$. We assume the world reference system is known. It may coincide with the center of projection of the camera (but it does not have to in general). Let $\vec{P}_c=(X_c,Y_c,Z_c)^T$ be the coordinates of the same point, this time in the camera reference frame (with $Z_c>0$ if the point is to be visible). The origin of the camera frame is its center of projection and the Z axis is the optical axis. The extrinsic parameters of the camera are then the translation vector and the rotation matrix that effect the transformation from the world point to the same point in the frame of reference of the camera:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \vec{T}$$

This equation can be written out as:

$$\begin{aligned} X_c &= r_{11}X + r_{12}Y + r_{13}Z + T_x \\ Y_c &= r_{21}X + r_{22}Y + r_{23}Z + T_y \\ Z_c &= r_{31}X + r_{32}Y + r_{33}Z + T_z \end{aligned}$$

Assuming that radial distortion may be neglected, we can write the point in image plane reference frame as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \dfrac{-f X_c}{s_x Z_c} + o_x \\ \dfrac{-f Y_c}{s_y Z_c} + o_y \end{bmatrix}$$

We need to calibrate for the extrinsic parameters, which are R , a rotation matrix, and $\vec{T}$, a translation vector, relating the camera frame to the world coordinate system. Additionally we also need to calibrate for the intrinsic parameters, which are

- $f_x = \frac{f}{s_x}$, the focal length in effective horizontal pixel size units
- $\alpha = \frac{s_x}{s_y}$, the aspect ratio
- (o_x, o_y) , the coordinates of the image center
- k_i , the radial distortion coefficients

Combining the equations, we obtain:

$$\begin{aligned} x - o_x &= -f_x \frac{r_{11} X + r_{12} Y + r_{13} Z + T_x}{r_{31} X + r_{32} Y + r_{33} Z} \\ y - o_y &= -f_y \frac{r_{21} X + r_{22} Y + r_{23} Z + T_y}{r_{31} X + r_{32} Y + r_{33} Z} \end{aligned}$$

where $f_y = \frac{f}{s_y}$. Assuming that the location of the image center (o_x, o_y) is known and that radial distortion can be neglected, the problem is to estimate f_x , α , R , and $\vec{T}$ from image points $(x_i, y_i)^T$ which are the projection of N known world points $\vec{P}_i = (X_i, Y_i, Z_i)^T$ obtained from the calibration pattern, in world coordinates. Since the last two equations share the same denominator, one can write:

$$x_i f_y (r_{21} X_i + r_{22} Y_i + r_{23} Z_i + T_y) = y_i f_x (r_{11} X_i + r_{12} Y_i + r_{13} Z_i + T_x)$$

And since $\alpha = \frac{s_x}{s_y}$, the above equation can be thought of as linear in the eight unknowns $\vec{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_8)$, where

$$\begin{aligned} \gamma_1 &= r_{21} & \gamma_5 &= \alpha r_{11} \\ \gamma_2 &= r_{22} & \gamma_6 &= \alpha r_{12} \\ \gamma_3 &= r_{23} & \gamma_7 &= \alpha r_{13} \\ \gamma_4 &= T_y & \gamma_8 &= \alpha T_x \end{aligned}$$

A homogeneous linear system of equations can thus be formed for the N 3D calibration points as $A\vec{\gamma} = 0$, where

$$A = \begin{bmatrix} x_1 X_1 & x_1 Y_1 & x_1 Z_1 & x_1 & -y_1 X_1 & -y_1 Y_1 & -y_1 Z_1 & -y_1 \\ x_2 X_2 & x_2 Y_2 & x_2 Z_2 & x_2 & -y_2 X_2 & -y_2 Y_2 & -y_2 Z_2 & -y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N X_N & x_N Y_N & x_N Z_N & x_N & -y_N X_N & -y_N Y_N & -y_N Z_N & -y_N \end{bmatrix}$$

is an $N \times 8$ matrix.

Now we must determine the scale factor ϕ (and hence the calibration parameters) from the solution $\hat{\gamma}=\phi(r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x)$. Recalling that R is a rotation matrix, we have $r_{21}^2+r_{22}^2+r_{23}^2=1$ and we obtain, from the first 3 components of $\hat{\gamma}$:

$$\sqrt{\hat{\gamma}_1^2+\hat{\gamma}_2^2+\hat{\gamma}_3^2} \;=\; \sqrt{\phi^2(r_{21}^2+r_{22}^2+r_{23}^2)} \;=\; |\phi|$$

Similarly, since $r_{11}^2+r_{12}^2+r_{13}^2=1$ and $\alpha>0$, then from the 5th through the 7th component of $\hat{\gamma}$ we obtain:

$$\sqrt{\hat{\gamma}_5^2+\hat{\gamma}_6^2+\hat{\gamma}_7^2} \;=\; \sqrt{\phi^2\alpha^2(r_{11}^2+r_{12}^2+r_{13}^2)} \;=\; \alpha|\phi|$$

These equalities allow us to solve for $|\phi|$ and the aspect ratio α . Finally, we determine the sign of ϕ and finalize the estimates of the parameters. For every 3D calibration point, $Z_c>0$ and therefore x and $r_{11}X+r_{12}Y+r_{13}Z+T_x$ must have opposite signs. If $x(r_{11}X+r_{12}Y+r_{13}Z+T_x)>0$ (or y and $y(r_{21}X+r_{22}Y+r_{23}Z+Ty)>0$ for that matter) then the signs of of the first two rows of the rotation matrix R and the first two components of $\vec{T}$ must be reversed. No further changes are required otherwise.

Two parameters remain to be estimated: T_z and f_x . They both can be obtained with least-squares. For each $(x_i, y_i)^T$, we can write

$$x_i(r_{31}X_i+r_{32}Y_i+r_{33}Z_i+T_z) \;=\; -f_x(r_{11}X_i+r_{12}Y_i+r_{13}Z_i+T_x)$$

and solve an over-constrained system of N linear equations:

$$A\begin{pmatrix}T_z\\ f_x\end{pmatrix} \;=\; \vec{b}$$

where

$$A \;=\; \begin{bmatrix} x_1 & (r_{11}X_1+r_{12}Y_1+r_{13}Z_1+T_x) \\ x_2 & (r_{11}X_2+r_{12}Y_2+r_{13}Z_2+T_x) \\ \vdots & \vdots \\ x_N & (r_{11}X_N+r_{12}Y_N+r_{13}Z_N+T_x) \end{bmatrix}$$

and

$$\vec{b} \;=\; \begin{bmatrix} -x_1(r_{31}X_1+r_{32}Y_1+r_{33}Z_1) \\ \vdots \\ -x_N(r_{31}X_N+r_{32}Y_N+r_{33}Z_N) \end{bmatrix}$$

with least-squares solution:

$$\begin{pmatrix} T_z \\ f_x \end{pmatrix} = (A^T A)^{-1} A^T \vec{b}$$

This simple approach just seen can be cumbersome in practice because of the requirement to explicitly know the position of the 3D calibration points in the world coordinate system. Calibration techniques that are easier to use have appeared. In particular, that from Zhang which is used in the OpenCV suite of computer vision tools.

Lens Distortions

The ideal lens does not introduce distortions. In reality however, it is impossible to manufacture such a lens. In addition, lenses and CCD array are hard to align perfectly. The lenses of real cameras often noticeably distort the location of image features near the CCD array. This bulging effect is known as radial distortion. At the image center (once calibrated for), radial distortion is null, and generally increases as we get further away from it. Radial distortion is modeled by the following equations:

$$\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \begin{pmatrix} x(1+k_1 r^2+k_2 r^4+k_3 r^6) \\ y(1+k_1 r^2+k_2 r^4+k_3 r^6) \end{pmatrix}$$

where $(x, y)^T$ is the original location of the distorted point in image coordinates, $(x_c, y_c)^T$ is the new location as a result of the correction, and $r^2 = x^2 + y^2$.

Manufacturing and assembly defects can create what is known as tangential distortion. This type of distortion occurs when the lens and the CCD array are not exactly parallel to each other. Tangential distortion is characterized with two additional parameters:

$$\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \begin{pmatrix} x+2 p_1 y + p_2 (r^2+2x^2) \\ y + p_1 (r^2+2y^2)+2 p_2 x \end{pmatrix}$$

There are five lens distortion parameters that require estimation. There are other factors of distortion which occur in imaging systems (such as chromatic aberration for instance) but it is generally sufficient to only calibrate for radial and tangential distortion.

OpenCV Calibration Process

It would be convenient to calibrate cameras with a planar calibration surface with a simple pattern on it, without having an explicit knowledge of the 3D location of points lying on it. OpenCV implements Zhang's calibration method, which does just this.

A planar homography is a projective mapping from one plane to another. Given two points in homogeneous coordinates $\vec{P}=(X,Y,Z,1)^T$ and $\vec{p}=(x,y,1)^T$ then, a homography is expressed as $\vec{p}=sH\vec{P}$ where s is an arbitrary scale factor expressing the fact the homography is defined only up to a scale factor, conventionally factored out of the transformation matrix H . Observe that the matrix encompasses a transformation and a projection. The transformation part involves a rotation matrix and a translation vector (extrinsic parameters), as before. They can be expressed by a single matrix transformation in homogeneous coordinates that relates the calibration plane to the image plane in the following way:

$$W=\begin{bmatrix} R & \vec{T} \end{bmatrix}$$

Let the camera matrix be

$$M=\begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

Then, the relationship between a 3D point on the calibration plane and its image coordinates are give by $\vec{p}=sMW\vec{P}$. If we consider the coordinate $\vec{P}'$ defined only on the calibration plane, then a simplification occurs. Without loss of generality we can choose to define the plane so that $Z=0$. By breaking the rotation matrix in three column vectors as $R=[\vec{r}_{1,}\vec{r}_{2,}\vec{r}_3]$ then we can write $\vec{p}=sMW\vec{P}$ as:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = sM\begin{bmatrix} \vec{r}_1 & \vec{r}_2 & \vec{r}_3 & \vec{T} \end{bmatrix}\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = sM\begin{bmatrix} \vec{r}_1 & \vec{r}_2 & \vec{T} \end{bmatrix}\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

The homography matrix H that maps a planar object's points onto the image plane is completely described by $H=sM[\vec{r}_{1,}\vec{r}_{2,}\vec{T}]$, where $\vec{p}=sH\vec{P}'$. The matrix is now 3×3 .

The homography matrix H relates the positions of the points on the calibration image plane to the imaging plane of the camera with the following equations:

$$\vec{p}_d=H\,\vec{p}_s \qquad \vec{p}_s=H^{-1}\,\vec{p}_d$$

where:

$$\vec{p}_d = \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} \qquad \vec{p}_s = \begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix}$$

in which $\vec{p}_s$ and $\vec{p}_d$ are the source and destination points, respectively.

Calibration Steps

We first ignore the distortion in the camera while solving for the other parameters. For each view of the calibration plane, we collect a homography H and write it in column vectors $H = [\vec{h}_1 \quad \vec{h}_2 \quad \vec{h}_3]$. Then, as we have just seen, after including the scale factor s , we can write:

$$H = \left[\vec{h}_1 \quad \vec{h}_2 \quad \vec{h}_3\right] = sM\left[\vec{r}_1 \quad \vec{r}_2 \quad \vec{T}\right]$$

Consequently, we have:

$$\vec{h}_1 = sM\vec{r}_1 \qquad \vec{r}_1 = \frac{1}{s}M^{-1}\vec{h}_1$$
$$\vec{h}_2 = sM\vec{r}_2 \qquad \vec{r}_2 = \frac{1}{s}M^{-1}\vec{h}_2$$
$$\vec{h}_3 = sM\vec{T} \qquad \vec{T} = \frac{1}{s}M^{-1}\vec{h}_3$$

The rotation vectors are orthogonal to each other and since the scale is extracted it follows that they are also orthonormal. This means that their dot product is zero and that their magnitudes are equal. Thus we have:

$$\vec{r}_1^{\,T}\vec{r}_2 = \vec{r}_1 \cdot \vec{r}_2 = 0$$

Using $(\vec{a}\,\vec{b})^T = \vec{b}^T \vec{a}^T$, we substitute for $\vec{r}_1$ and $\vec{r}_2$ and derive the following constraint:

$$\vec{h}_1^{\,T}(M^{-1})^T M^{-1}\vec{h}_2 = 0$$

We also know that the magnitudes of the rotation vectors are equal:

$$\|\vec{r}_1\| = \|\vec{r}_2\| \qquad \vec{r}_1^{\,T}\vec{r}_1 = \vec{r}_2^{\,T}\vec{r}_2$$

Substituting for $\vec{r}_1$ and $\vec{r}_2$, we obtain a second constraint:

$$\vec{h}_1^{\,T}(M^{-1})^T M^{-1}\vec{h}_1 = \vec{h}_2^{\,T}(M^{-1})^T M^{-1}\vec{h}_2$$

We now set $B = (M^{-1})^T M^{-1}$, which yields:

$$B = \begin{bmatrix} \frac{1}{f_x^2} & 0 & \frac{-o_x}{f_x^2} \\ 0 & \frac{1}{f_y^2} & \frac{-o_y}{f_y^2} \\ \frac{-o_x}{f_x^2} & \frac{-o_y}{f_y^2} & \frac{o_x^2}{f_x^2}+\frac{o_y^2}{f_y^2}+1 \end{bmatrix}$$

With this matrix, both constraints have the form $\vec{h}_i^T B \vec{h}_j$. Since this matrix is symmetric, it can be written as a one six-dimensional vector dot product. Arranging the necessary elements of B into vector $\vec{b}$, we have:

$$\vec{h}_i^T B \vec{h}_j = \vec{v}_{ij}^T \vec{b} = \begin{bmatrix} \vec{h}_{i1}\vec{h}_{j1} \\ \vec{h}_{i1}\vec{h}_{j2}+\vec{h}_{i2}\vec{h}_{j1} \\ \vec{h}_{i2}\vec{h}_{j2} \\ \vec{h}_{i3}\vec{h}_{j1}+\vec{h}_{i1}\vec{h}_{j3} \\ \vec{h}_{i3}\vec{h}_{j2}+\vec{h}_{i2}\vec{h}_{j3} \\ \vec{h}_{i3}\vec{h}_{i3} \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{12} \\ B_{22} \\ B_{13} \\ B_{23} \\ B_{33} \end{bmatrix}^T$$

Using this definition for $\vec{v}_{ij}^T$, the two previous constraints are now written as:

$$\begin{bmatrix} \vec{v}_{12}^T \\ (\vec{v}_{11}-\vec{v}_{22})^T \end{bmatrix} \vec{b} = 0$$

If we collect N calibration images, then we can stack as many as these equations together to form $V\vec{b}=0$ where V is a $2N\times 6$ matrix.

The intrinsic parameters for the camera are obtained directly from matrix B as:

$$\begin{aligned} f_x &= \sqrt{\frac{\gamma}{B_{11}}} \\ f_y &= \sqrt{\frac{\gamma B_{11}}{(B_{11}B_{22}-B_{12}^2)}} \\ o_x &= \frac{-B_{13} f_x^2}{\gamma} \\ o_y &= \frac{(B_{12}B_{13}-B_{11}B_{23})}{B_{11}B_{22}-B_{12}^2} \end{aligned}$$

where

$$\gamma = B_{33} - \frac{(B_{13}^2 + o_y(B_{12}B_{13} - B_{11}B_{23}))}{B_{11}}$$

The extrinsic parameters are solved using the homography equations as:

$$\begin{aligned}
\vec{r}_1 &= \frac{1}{s}M^{-1}\vec{h}_1 \\
\vec{r}_2 &= \frac{1}{s}M^{-1}\vec{h}_2 \\
\vec{r}_3 &= \vec{r}_1 \times \vec{r}_2 \\
\vec{T} &= \frac{1}{s}M^{-1}\vec{h}_3
\end{aligned}$$

The scale parameter s is determined from the orthonormality condition:

$$\gamma = \frac{1}{s} = \frac{1}{\left\| M^{-1}\vec{h}_1 \right\|}$$

By solving for R in this way, it is almost certain that $R^T R = R R^T$ will not hold, as it should for a pure rotation matrix. This can be addressed with computing the SVD decomposition of the matrix and renormalize it.

We now turn to solving for the intrinsic parameters of the camera. The projected points onto the image plane are distorted owing to radial and tangential distortions. Let $(x_p, y_p)^T$ be the location of the point as if the optics did not introduce any distortion, and let $(x_d, y_d)^T$ be its distorted location. Then:

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} f_x \frac{X}{Z} + o_x \\ f_y \frac{Y}{Z} + o_y \end{bmatrix}$$

We use the results of the calibration without distortion via the following substitution:

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} x_d \\ y_d \end{bmatrix} + \begin{bmatrix} 2 p_1 x_d y_d + p_2(r^2 + 2 x_d^2) \\ p_1(r^2 + 2 y_d^2) + 2 p_2 x_d y_d \end{bmatrix}$$

A large list of these equations are collected and solved to find the distortion parameters, after which the intrinsic and extrinsic parameters are re-estimated.