Camera Calibration Report

Minghe Liu

2025.7.11

1   Abstract

This report details the implementation and results of a camera's calibration which is installed on a 6-degree-of-freedom manipulator. The calibration framework integrates camera intrinsic and extrinsic parameter estimation to achieve precise alignment between the robot's coordinate system and the vision sensor. Utilizing the pinhole camera model, intrinsic parameters (focal length f_x, f_y, principal point (c_x, c_y), skew coefficient s, and radial distortion coefficients (k_1, k_2, k_3)) were characterized. Extrinsic parameters, defining the camera's pose relative to the robot base, were computed via rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$.

2   Introduction

There are two principal calibration methodologies in all. Tsai's 3D target-based approach and Zhang's planar grid technique. Tsai's method leveraged the Direct Linear Transform (DLT) algorithm to solve the perspective projection equation using ≥ 6 non-coplanar control points, followed by Singular Value Decomposition (SVD) to recover initial parameters. Zhang's method utilized homography estimation from multiple views of a planar checkerboard, requiring only ≥ 4 coplanar points per view, significantly enhancing practicality. Both methods incorporated non-linear refinement via the Levenberg-Marquardt algorithm to minimize reprojection error—quantified as the Euclidean distance (pixels) between observed and reprojected points—while jointly optimizing lens distortion parameters.

Due to the complexity of Tsai's calculation process, the requirement for more specialized calibration tools, and the method being somewhat outdated, our experimental procedure employed the newer method proposed by Zhang: approximately 25 images of a 2D calibration board were captured by the camera from various angles and distances. To prevent distortion of the black and white squares caused by the flexibility of the calibration paper, which could affect calibration accuracy, the squares were imported into the computer and set to maximum brightness to facilitate the algorithm in distinguishing the vertices within different squares.

3   Camera Parameters and Calibration

3.1   Pinhole Camera Model

The whole calibration procedure is based on the pinhole camera model. It describes the image formation process in an ideal camera without lenses. Light rays from a 3D point pass through a single point (the pinhole) and project onto a 2D image plane.

The model is characterized by:
Focal Point (O): The center of projection (pinhole).
Image Plane (Retinal Plane): Where the image is formed, located at a distance of f.

And there are two common representations, real Image which its plane is behind the pinhole and virtual Image which its plane is in front of the pinhole (upright image). We use the virtual image model for simplicity.

3.2 Coordinate Systems

We define three coordinate systems:

- **World Coordinate System** (X_w, Y_w, Z_w): Arbitrary 3D reference frame.

- **Camera Coordinate System** (X_c, Y_c, Z_c): Origin at the pinhole O, Z_c-axis along the optical axis.

- **Image Coordinate System** (u, v): 2D coordinates in pixels on the image plane.

The transformation from world to camera coordinates is given by a rigid-body transformation (rotation and translation).

3.3 Projection Equations and Pixel Coordinates

Consider a 3D point $P = (Xc, Yc, Zc)$ in the camera coordinate system. The projection of P onto the image plane is given by similar triangles:

$$x = \frac{fX_c}{Z_c}, \quad y = \frac{fY_c}{Z_c}$$

where (x, y) are the coordinates in the image plane (in metric units, e.g., millimeters).
The image plane is sampled by a sensor array. We convert (x, y) to pixel coordinates (u, v) and the transformation is:

$$u = \frac{f_x X_c}{Z_c} + \frac{s Y_c}{Z_c} + c_x$$

$$v = \frac{f_y Y_c}{Z_c} + c_y$$

In homogeneous coordinates, the projection is:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}$$

The matrix K is the intrinsic matrix:

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

3.4 Extrinsic Parameters

The transformation from world coordinates to camera coordinates is:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Where R is a 3×3 rotation matrix and t is a 3×1 translation vector.

3.5 Full Projection Model

Combining intrinsic and extrinsic parameters:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

where $\lambda = Z_c$ is the depth.

3.6 Lens Distortion

Real lenses introduce radial and tangential distortion. The radial distortion model is:

$$\begin{bmatrix} x_{\text{distorted}} \\ y_{\text{distorted}} \end{bmatrix} \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 \right) \begin{bmatrix} x \\ y \end{bmatrix}$$

(x, y) are normalized image coordinates:

$$x = \frac{X_c}{Z_c}, y = \frac{Y_c}{Z_c}$$

With $r^2 = x^2 + y^2$ and k_1, k_2, k_3 are radial distortion coefficients. The distorted coordinates are then transformed to pixel coordinates:

$$u = f_x x_{\text{distorted}} + s y_{\text{distorted}} + c_x, v = f_y y_{\text{distorted}} + c_y$$

4 Zhang's Algorithm

4.1 Homography Estimation (DLT Algorithm)

The calibration target lies on the world plane $Z_w = 0$. The projection equation is:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \triangleq \mathbf{H} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix}$$

$\mathbf{H} = \mathbf{K}[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}]$ is the homography matrix and λ is a scale factor.

4.2 Solving the Homography Matrix

For each pair of corresponding points, we have:

$$\begin{bmatrix} u & v & 1 \end{bmatrix} \times \left(\mathbf{H} \begin{bmatrix} X_w & Y_w & 1 \end{bmatrix} \right) = 0$$

Expanding this, we obtain two equations:

$$X_w \mathbf{h}_1^T \mathbf{p} - u X_w \mathbf{h}_3^T \mathbf{p} = 0$$
$$Y_w \mathbf{h}_1^T \mathbf{p} - u Y_w \mathbf{h}_3^T \mathbf{p} = 0$$
$$X_w \mathbf{h}_2^T \mathbf{p} - v X_w \mathbf{h}_3^T \mathbf{p} = 0$$
$$Y_w \mathbf{h}_2^T \mathbf{p} - v Y_w \mathbf{h}3^T \mathbf{p} = 0$$

where ($\mathbf{p} = \begin{bmatrix} X_w & Y_w & 1 \end{bmatrix}$). However, in practice, we typically use the following two independent equations (derived from the cross product):

$$\begin{bmatrix} \mathbf{p}^T & \mathbf{0}^T & -u\mathbf{p}^T & \mathbf{0}^T & \mathbf{p}^T & -v\mathbf{p}^T \end{bmatrix} 2 \times 9 \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = 0$$

Here, $mathbf{p} = \begin{bmatrix} X_w & Y_w & 1 \end{bmatrix}^T$. For n points, We construct a $2n \times 9$ matrix $\mathbf{A}$ as follows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{p}_1^T & \mathbf{0}^T & -u_1\mathbf{p}_1^T & \mathbf{0}^T & \mathbf{p}_1^T & -v_1\mathbf{p}_1^T \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{p}_n^T & \mathbf{0}^T & -u_n\mathbf{p}_n^T & \mathbf{0}^T & \mathbf{p}_n^T & -v_n\mathbf{p}_n^T \end{bmatrix}$$

Then we solve the equation (A h = 0), where $\mathbf{h} = \begin{bmatrix} \mathbf{h}_1^T & \mathbf{h}_2^T & \mathbf{h}_3^T \end{bmatrix}$. Using Singular Value Decomposition (SVD), $\mathbf{h}$ is the eigenvector corresponding to the smallest singular value of $\mathbf{A}^T\mathbf{A}$.(i.e., the last column of $\mathbf{V}$)

4.3 Intrinsic Parameters from Homographies

Let $\mathbf{H}i = [\mathbf{h}i1, \mathbf{h}i2, \mathbf{h}i3]$ be the homography matrix for the i-th image. From the orthogonality of $\mathbf{r}1$ and $\mathbf{r}2$, we have the following constraints:

$$\mathbf{h}i1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}i2 = 0$$

$$\mathbf{h}i1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}i1 = \mathbf{h}i2^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}i2$$

Let $\mathbf{B} = \mathbf{K}^{-T}\mathbf{K}^{-1}$, which is symmetric. Define $\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T$. Then the constraints can be rewritten as:

$$\begin{bmatrix} h_{i1}h_{j1} & h_{i1}h_{j2} + h_{i2}h_{j1} & h_{i2}h_{j2} & h_{i3}h_{j1} + h_{i1}h_{j3} & h_{i3}h_{j2} + h_{i2}h_{j3} & h_{i3}h_{j3} \end{bmatrix}^T \mathbf{b} = 0$$

For each image, we obtain two equations (from the two constraints). Stacking m images yields the system $\mathbf{V}\mathbf{b} = \mathbf{0}$, where $\mathbf{V}$ is a $2m \times 6$ matrix. The solution is obtained via Singular Value Decomposition (SVD); $\mathbf{b}$ is the last column of $\mathbf{V}^T$.

4.4 Computing the Intrinsic Matrix and Extrinsic Parameters

From $\mathbf{b}$, we recover $\mathbf{B}$, and then obtain $\mathbf{K}$ via Cholesky decomposition $\mathbf{B} = \mathbf{K}^T\mathbf{K}^{-1}$. The specific steps are as follows:

Compute the elements of $\mathbf{B}$:

$$\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}$$

Use Cholesky decomposition to find $\mathbf{K}^{-1}$, and then compute $\mathbf{K}$ by taking its inverse.

For each image, compute the extrinsic parameters from $\mathbf{H}$ and $\mathbf{K}$:

$$\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t}_i \end{bmatrix} = \begin{bmatrix} \lambda\mathbf{K}^{-1}\mathbf{h}_1 & \lambda\mathbf{K}^{-1}\mathbf{h}_2 & \mathbf{r}_1 \times \mathbf{r}_2 & \lambda\mathbf{K}^{-1}\mathbf{h}_3 \end{bmatrix}$$

where $\lambda = \frac{1}{\|\mathbf{K}^{-1}\mathbf{h}_i 1\|}$. Then, orthogonalize $\mathbf{R} = [\mathbf{r}_i 1, \mathbf{r}_i 2, \mathbf{r}_i 3]$.

4.5 Nonlinear Optimization

Use the Levenberg-Marquardt algorithm to optimize all parameters (intrinsic parameters, extrinsic parameters, and distortion coefficients). Minimize the reprojection error:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \|p_{ij} - \pi(\mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{d}, \mathbf{P}_j)\|^2$$

where $\mathbf{d}$ is the distortion coefficient vector, and π is the projection function (including the distortion model).