RBE 549- HW1 -Camera Calibration

Ankit Mittal
Department of Robotics Engineering
Worcester Polytechnic Institute
Email: amittal@wpi.edu

***Abstract*—This homework assignment delves into the critical field of camera calibration in computer vision. It primarily revolves around the estimation of a camera's intrinsic parameters (including focal length and principal point position), extrinsics, and distortion coefficients. Camera calibration is a cornerstone in computer vision, significantly influencing the accuracy and effectiveness of various applications. In this task, we will undertake the implementation of a well-established camera calibration method formulated by Zhengyou Zhang. This approach is pivotal for comprehensively understanding and accurately capturing the unique characteristics of any camera used in computer vision projects.**

I. Introduction

We understand that to transform a real-world point to its representation in an image, two distinct types of transformation matrices are employed. The first type of matrix converts the real-world point to the image plane frame, involving the extrinsic parameters of the camera. The second type transforms the point within the image plane frame to the image pixel coordinate system, incorporating the camera's intrinsic parameters. The camera intrinsic matrix is given by:-.

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_o \\ 0 & \beta & v_o \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

with (u_0, v_0) representing the coordinates of the principal point, α and β as the scale factors in the image u and v axes respectively, and γ as the parameter describing the skewness of the two image axes.

Without loss of generality, we assume the model plane is on $Z = 0$ of the world coordinate system. Let's denote the i-th column of the rotation matrix $\mathbf{R}$ by $\mathbf{r}_i$. From pinhole camera model we have, where s is the scale

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A}\left[\mathbf{r}_1 \mathbf{r}_2 \mathbf{r}_3 \mathbf{t}\right] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}$$

$$= \mathbf{A}\left[\mathbf{r}_1 \mathbf{r}_2 \mathbf{t}\right] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

II. Initial Parameter estimation

To begin our camera calibration process, we first require data. In this case, we have 13 images of a checkerboard of known dimensions, captured using a Google Pixel XL smartphone with a fixed focal length. The size of each square on the checkerboard is 21.5mm. The original dimensions of the checkerboard are 7x10 squares. However, for this calibration, we will disregard the outermost columns and rows, focusing instead on a 6x9 checkerboard configuration.

A. camera intrinsic matrix

We will initiate the process by identifying the corners of the checkerboard both in world coordinates and pixel coordinates. The function findChessboardCorners was utilized to detect 54 corners in each image. Subsequently, we determined the corresponding real-world coordinates based on the known size of each square on the checkerboard. Each corner located on the checkerboard provides us with an equation of the form depicted in Figure 1.

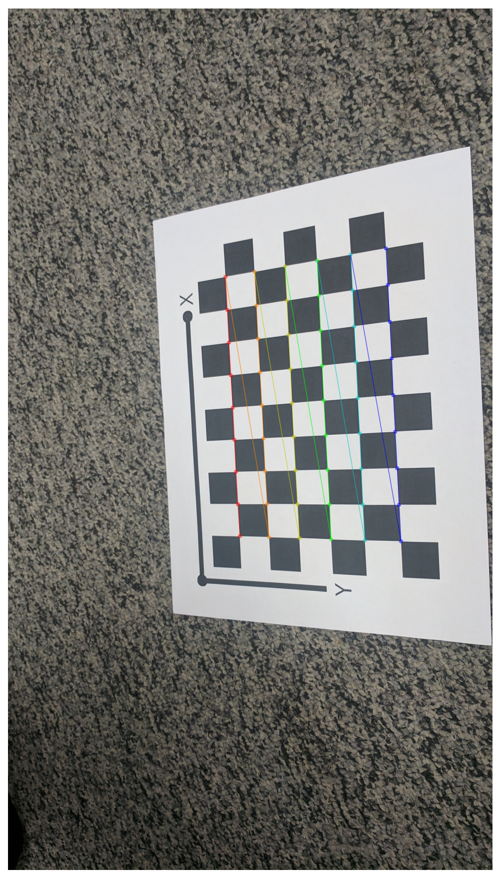

Fig. 1: Chess broad corners

The next step involves calculating the overall transformation matrix. Considering $Z = 0$, indicating that all world points lie on the same plane, we can define the product of matrix $\mathbf{A}$ and $\mathbf{Rt}$ as the homography matrix $\mathbf{H}$. This matrix $\mathbf{H}$ facilitates the transformation from 3D world points to image coordinates. Subsequently, we can decompose this homography matrix to extract the matrices $\mathbf{A}$ and $\mathbf{Rt}$ from $\mathbf{H}$.

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix}$$

To calculate the homography $\mathbf{H}$, we will employ Singular Value Decomposition (SVD) after rearranging the problem into the $\mathbf{Ax} = \mathbf{0}$ form using DLT. The solution can be obtained by solving the below set of equations.

$$\mathbf{a}_{x_i}^\top \mathbf{h} = 0$$
$$\mathbf{a}_{y_i}^\top \mathbf{h} = 0$$

where

$$\mathbf{a}_{x_i}^\top = (-X_i, -Y_i, -1, 0, 0, 0, x_i X_i, x_i Y_i, x_i)$$
$$\mathbf{a}_{y_i}^\top = (0, 0, 0, -X_i, -Y_i, -1, y_i X_i, y_i Y_i, y_i)$$

However, it is important to note that we have 8 unknowns in the vectorized Homography matrix $\mathbf{H}$, so a minimum of eight equations is required to solve for all unknowns. To obtain these 8 equations, we need the world and image coordinates of at least 4 corner points. After performing SVD, we should normalize all the terms such that the last element of the $\mathbf{h}$ matrix is set to 1. After acquiring the Homography matrix, the next step is to decompose it into two distinct matrices: the camera intrinsic matrix and the extrinsic matrix. This will involve setting up another homogeneous linear system to solve for the intrinsic matrix $\mathbf{A}$. To establish such a system, we utilize the constraints inherent to the $\mathbf{Rt}$ matrix. We know that $\mathbf{r}_1$ and $\mathbf{r}_2$ from the $\mathbf{Rt}$ matrix are orthonormal vectors, implying that their dot product equals zero and their respective norms are one. By employing the aforementioned constraints, we derive the subsequent set of equations that we must resolve.

$$\mathbf{h}_1^\top \mathbf{A}^{-\top} \mathbf{A}^{-1} \mathbf{h}_2 = 0$$
$$\mathbf{h}_1^\top \mathbf{A}^{-\top} \mathbf{A}^{-1} \mathbf{h}_1 = \mathbf{h}_2^\top \mathbf{A}^{-\top} \mathbf{A}^{-1} \mathbf{h}_2$$

Fom the above equation we can observe that $\mathbf{A}^{-\top}\mathbf{A}^{-1}$ is common in both equation containing camera intrinsic parameter hence we can consider an new matrix $\mathbf{B}$ as:

$$\mathbf{B} = \mathbf{A}^{-\top}\mathbf{A}^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}$$

Please note that $\mathbf{B}$ is a symmetric and positive definite matrix, so it has a degree of freedom (DoF) of only 6. We will define a linear homogeneous system as follows:

$$\mathbf{b} = \begin{bmatrix} B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33} \end{bmatrix}^\top .$$

Let the i-th column vector of $\mathbf{H}$ be $\mathbf{h}_i = \begin{bmatrix} h_{i1} & h_{i2} & h_{i3} \end{bmatrix}^\top$. Then, we have

$$\mathbf{h}_i^\top \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^\top \mathbf{b}$$

with,

$$\mathbf{v}_{ij} = \begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2} + h_{i2}h_{j1} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1} + h_{i1}h_{j3} \\ h_{i3}h_{j2} + h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix}^\top$$

writing it into Ax = 0 form we will have

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

Because there are 6 unknowns, we will require three different homography matrices to compute the above equations. To acquire three homographies, a minimum of 3 images of the checkerboard in three different orientations is needed. Once we obtain the matrix $\mathbf{B}$, we can readily derive the matrix $\mathbf{K}$ using Cholesky decomposition, or we may consult Appendix B in Zhang's paper to determine the individual elements of the $\mathbf{K}$ matrix and then construct the $\mathbf{K}$ matrix from these elements. This provides our initial estimate of the camera intrinsic matrix $\mathbf{K}$, which we will further refine through optimization.

$$\mathbf{K} = \begin{bmatrix} 2056.37858 & -1.12657309 & 761.43522 \\ 0.00000 & 2041.00896 & 1350.23618 \\ 0.00000 & 0.00000 & 1.00000 \end{bmatrix}$$

B. Camera extrinsic matrix

The camera extrinsic parameters can be estimated by the following equations. Due to noise in the data, the computed matrix $\mathbf{Rt}$ may not perfectly satisfy the properties of a rotation matrix.

$$\mathbf{r_1} = \lambda \mathbf{A}^{-1} \mathbf{h_1}$$
$$\mathbf{r_2} = \lambda \mathbf{A}^{-1} \mathbf{h_2}$$
$$\mathbf{r_3} = \mathbf{r_1} \times \mathbf{r_2}$$
$$\mathbf{t} = \lambda \mathbf{A}^{-1} \mathbf{h_3}$$

C. Distortion Parameter

We assume minimal camera distortion, allowing us to use $[0, 0]^\top$ as our initial estimate for the distortion coefficients. It's important to note that we are considering only the first two terms for distortion calculation, resulting in only two parameters. Additional parameters can be incorporated for improved accuracy if necessary. However, when accounting for distortion later, we utilize the following equations to compute pixel coordinates. It's worth mentioning that in this context, we assume $\gamma = 0$, signifying the absence of skewness in the image frame and the sensor.

$$u = u + (u - u_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$
$$v = v + (v - v_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$

III. Optimization

We have previously estimated the necessary intrinsic, extrinsic, and distortion parameters. The solution we obtained for the intrinsic parameters was obtained by minimizing an algebraic distance, which lacks physical significance. Therefore, we intend to perform Maximum Likelihood Estimation (MLE) inference to obtain the precise solution. However, it's important to note that we also need to update the distortion parameters. As a result, we will conduct Maximum Likelihood Estimation on the entire set of parameters simultaneously, as expressed by the following equations:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\|m_{ij} - \widetilde{m}(A, k_1, k_2, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j)\|^2$$

In this step, we minimize the least squares distance between the actual corner pixel coordinates and the projected coordinates (using the estimated parameters) by utilizing the '$scipy.optimize.leastsquares()$' function. After optimization, we obtain the final set of parameters as follows:

$$K = \begin{bmatrix} 2.05636702 \times 10^3 & -1.13967122 & 7.61244430 \times 10^2 \\ 0.00000000 & 2.04100449 \times 10^3 & 1.35023744 \times 10^3 \\ 0.00000000 & 0.00000000 & 1.00000000 \end{bmatrix}$$

$$k = \begin{bmatrix} 0.010159 \\ -0.0760 \end{bmatrix}$$

$$ReprojectionError = 1.095$$

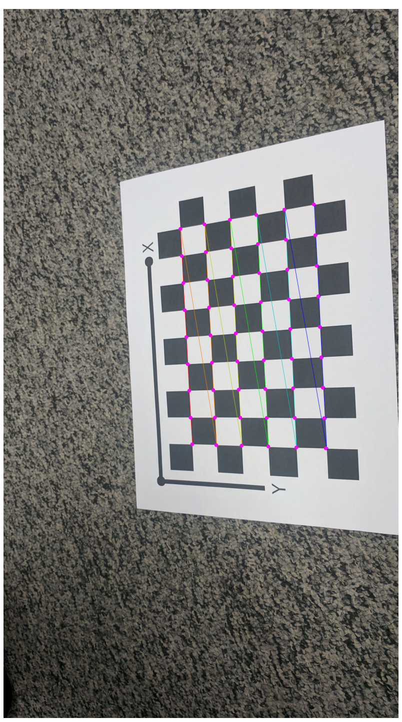

Fig. 2: undistorted Chessboard corners

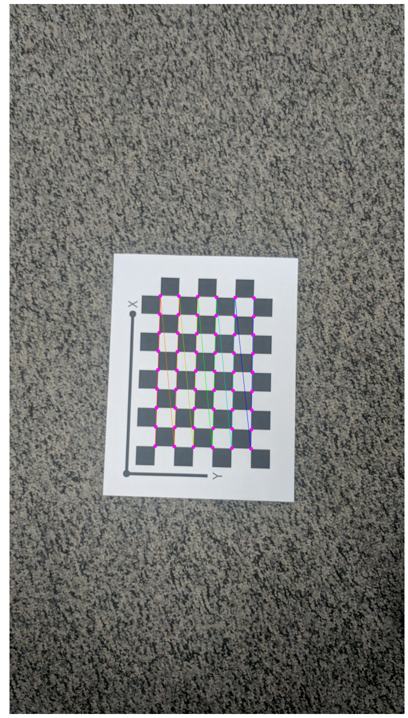

Fig. 3: undistorted Chessboard corners

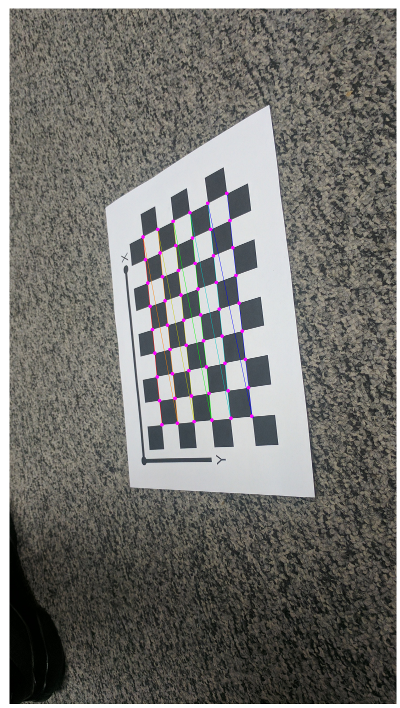

Fig. 4: undistorted Chessboard corners

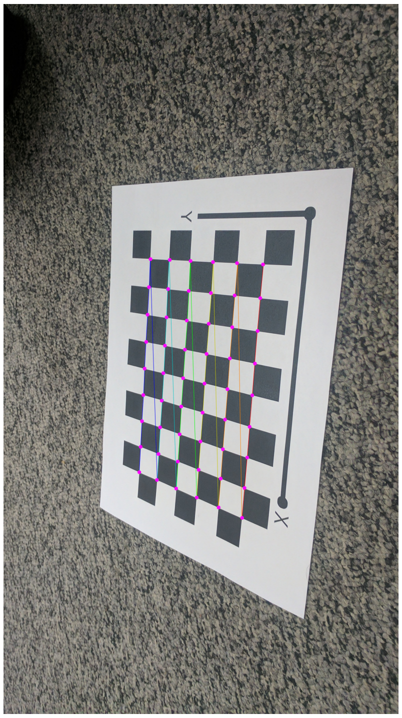

Fig. 5: undistorted Chessboard corners

IV. Acknowledgment

The author would like to thank Prof. Nitin Sanket and the TA of this course RBE549- Computer Vision.

References

[1] RBE549 - Computer Vision Website**Link**
[2] A Flexible New Technique for Camera Calibration-Zhengyou Zhang **Link**