RBE 549: Homework 1 - AutoCalib

Mandeep Singh
M.S. Robotics Engineering
Worcester Polytechnic Institute (WPI)
Worcester, MA 01609
Email: msingh2@wpi.edu

***Abstract*—In this Homework we will be working on calculating camera calibration parameters. Camera calibration means estimating any camera's intrinsics, extrinsics and distortion parameters. Intrinsic parameters consists of focal length and principal point position and distortion parameters are coefficients of distortion. Camera calibration is one of the most important part of any computer vision project. In this homework we will be implementing a well known camera calibration technique proposed by Zhengyou Zhang.**

***Index Terms*—Camera Calibration, Intrinsic camera parameters, Distortion**

I. Introduction

We know that when we want to do a transformation between real world point with the one captured in image we have two kinds of transformation matrices involved. One that transforms the world point to image plane frame and the second which transforms the point in image plane frame to image pixel co-ordinate system. The former contains the extrinsic parameters and the latter having intrinsic parameters. Intrinsic camera matrix is shown in Fig.1

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Fig. 1: Intrinsic parameter matrix

Here (u0,v0) is the optical center or the principal point in pixel coordinates. (α and β), are scale factors in u and v axis and s is the skew coefficient in case the image plane frame and sensor is not perfectly aligned.

In Zhang's method firstly we will assume that the model plane is located on Z=0. Which means that the third column of Rt matrix i.e. r3 becomes redundant so we can remove it.

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

Fig. 2: Transformation between world point to image pixel co-ordinate.

II. Initial Parameters Estimation

First of all we will be needing data to start with our camera calibration process. Here, we are given with 13 images of checkerboard of known size taken from a Google Pixel XL smartphone with focal length locked. The actual size of checkerboard square side is 21.5mm. The checkerboard is of (7 x 10) size. But we will be neglecting the corner columns and rows and thus work with size of (6 x 9) checkerboard.

A. Approximating camera intrinsic matrix

We will be following the below written steps for coming with an initial guess of the intrinsic parameters.

- We will start with finding the corners of the checkerboard in world co-ordinates as well as pixel co-ordinates. Used cv2.checkerBoardCorners() function to find 54 corners in every image. And then calculated the corresponding real world co-ordinates with the given size of the square. Every corner found on the checker board will give us an equation of the form as shown in Fig. 2.

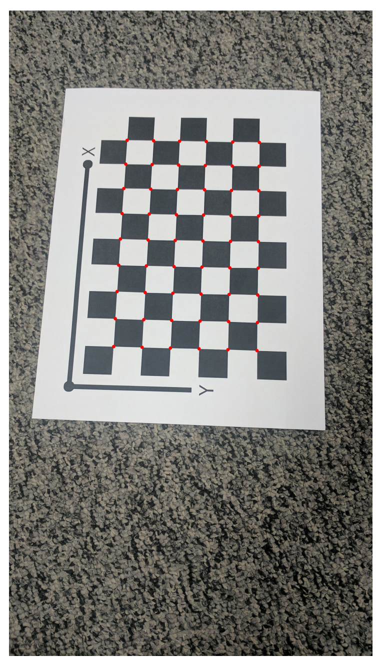

Fig. 3: 54 Corners found in the image 1.

- The next step is to calculate the overall transformation matrix or Homography matrix (consists of both A and Rt matrix) for each image.

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \underset{3\times 3}{\mathsf{H}} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix}$$

To calculate the homography, we will employ Singular value Decomposition (SVD) after rearranging the problem into Ax = 0 form. So, we can get our solution by solving the below set of equations.

$$\boldsymbol{a}_{x_i}^{\mathsf{T}} \boldsymbol{h} = 0$$
$$\boldsymbol{a}_{y_i}^{\mathsf{T}} \boldsymbol{h} = 0$$

$$\begin{aligned} \boldsymbol{a}_{x_i}^{\mathsf{T}} &= (-X_i, -Y_i, -1, 0, 0, 0, x_i X_i, x_i Y_i, x_i) \\ \boldsymbol{a}_{y_i}^{\mathsf{T}} &= (0, 0, 0, -X_i, -Y_i, -1, y_i X_i, y_i Y_i, y_i) \end{aligned}$$

But notice that we have 8 unknowns in vectorized Homography matrix h, so we will need minimum eight equations to solve all unknowns. To have 8 equations we need 4 corner points whose world and image co-ordinates are known to us. After SVD, normalize all the terms such that the last element of the h matrix becomes 1.

- After getting Homography matrix, our next task is to decompose this matrix into two individual matrices namely the camera intrinsic and extrinsic matrix. We will form another homogeneous linear system so that we can solve for the matrix A. To formulate such a system, we will exploit the constraints of the Rt matrix. We know that r1 and r2 in Rt matrix are orthonormal vectors means their dot product is zero and their norm is equal to 1. Using the above constraints we will get to below written equations which we have to solve.

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

Now, We will be defining a B matrix as follows:

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

Please note that B is a symmetric and positive definite matrix so it has DoF of only 6. We will define a linear homogenous system as follows:

$$\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T .$$

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

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

Because there are 6 unknowns we will need three different homography matrices to compute the above equations. To have three homographies we need minimum 3 images of the checker board in three different planes.

- Once, we get the B matrix we can easily get the K matrix by Cholesky's decomposition or we can just refer to Appendix B in Zhang's paper to find the individual elements of the K matrix and then construct K matrix out of them.
This is our initial estimate of the camera intrinsic matrix (K) as shown below which we will further optimize later.

$$\mathrm{K} = \begin{bmatrix} 2464.4 & -0.368 & 763.8 \\ 0 & 2444.1 & 1348.3 \\ 0 & 0 & 1 \end{bmatrix}$$

B. Approximating camera extrinsic matrix

The camera extrinsic parameters can be estimated by the following equations. Because of the noise in data the computed matrix Rt may not be able to satisfy properties of a Rotation matrix.

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

C. Approximating Distortion k

We are assuming that we have minimum camera distortion so we can take [0,0]T as our initial guess for the distortion coefficients. Please note that we are considering only the first two terms for calculating distortion and that's why only two parameters. We can take more parameters if need better results.

But later when we take distortion in consideration we use below equations to calculate the pixel co-ordinates. Please note that here we have assumed that $\gamma = 0$ i.e. there is no skewness in the image frame and the sensor.

$$\breve{u} = u + (u - u_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$
$$\breve{v} = v + (v - v_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] .$$

III. OPTIMIZATION

We have already estimated the required intrinsic, extrinsic and distortion parameters. The solution we obtained for intrinsic parameters is obtained by minimizing an algebraic distance which doesn't mean anything physically, so we will do Maximum likelihood estimation inference to get the exact solution. But we also have to update the distortion parameters as well. So, we can do Maximum Likelihood estimation on all set of parameters at once which is given by:

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

Here, we minimize the least squares distance between actual corner pixel co-ordinates and the projected co-ordinates (using estimated parameters). We use scipy.optimize.least_squares function for this task.

After optimizing the following final parameters are obtained:

$$K = \begin{bmatrix} 2464.4 & -0.368 & 763.8 \\ 0 & 2444.1 & 1348.3 \\ 0 & 0 & 1 \end{bmatrix} \quad k = \begin{bmatrix} 0.0125 \\ -0.0125 \end{bmatrix}$$

Re-projection error = 0.7535

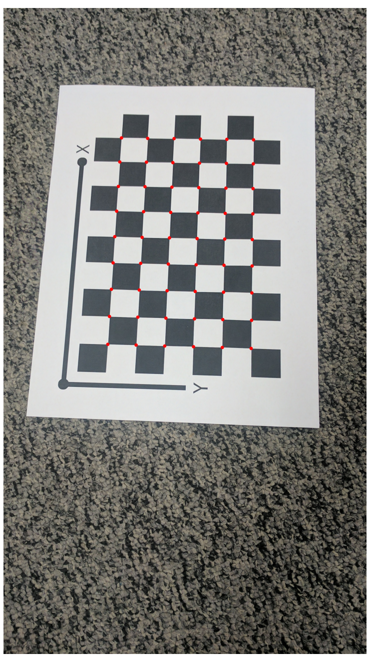

Fig. 4: Re-projected corners on undistorted image 1.

IV. References

1) Z. Zhang, "A flexible new technique for camera calibration," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, Nov. 2000, doi: 10.1109/34.888718.
2) http://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-22-Zhang-calibration.pptx.pdf
3) https://docs.opencv.org/4.x/dc/dbb/tutorial_py_calibration.html