UNIVERSITY OF MARYLAND,COLLEGE PARK

PROJECT 3

ENPM 673 PERCEPTION FOR AUTONOMOUS ROBOTS

Kiran Ajith 119112397

April, 2023

Contents

1

1.1 Problem

In this problem, the task is to calibrate a camera.

1.2 Pipeline

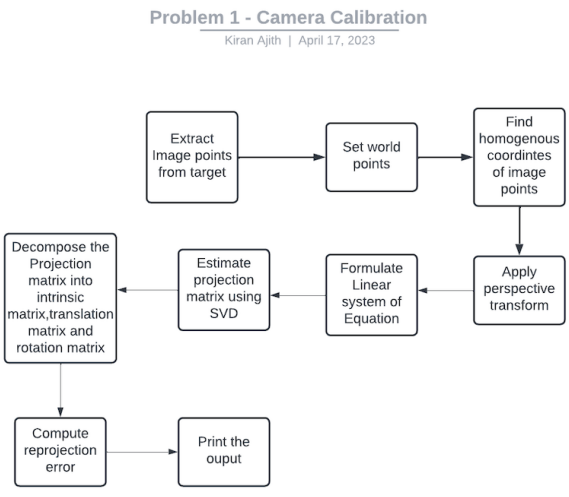

Figure 1.1: Flowchart representing the various processes involved

A point in space represented in 3D by a world coordinate frame can be transformed to a camera frame and then to an image plane using what is defined as a Projection matrix. Camera calibration involves estimating the projection matrix. Let there be a point P in space repre-

sented in world coordinate frame by,

$$X_W = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \tag{1.1}$$

This is transformed to the camera coordinate frame, and it becomes,

$$X_C = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \tag{1.2}$$

After the point in the camera coordinate system is obtained, perspective projection is applied to it to get the 2D coordinates on the image plane, which is written as follows,

$$X_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{1.3}$$

The relation between image plane coordinates and camera coordinates is given by,

$$x_i = \frac{fx_c}{z_c}, y_i = \frac{fy_c}{z_c} \tag{1.4}$$

Let u, v be the coordinates of the point in the image plane. They can be represented in by

$$u = m_x f \frac{x_c}{z_c} + O_x \tag{1.5}$$

$$v = m_y f \frac{y_c}{z_c} + O_y \tag{1.6}$$

where $f_x = m_x f$ and $f_y = m_y f$ are the effective focal lenghts in the x and y direction respectively.
The unknowns f_x, f_y, O_x, O_y are the intrinsic parameters of the camera and are dependent on the internal geometry of the camera. The extrinsic parameter of the camera depends on the transformation of the world and camera coordinate frames. The camera-centric location of a point in space is given by

$$X_c = R(X_w - C_w) \tag{1.7}$$

The above equation can be rewritten in terms of translation as follows:

$$X_c = R(X_w + t) \tag{1.8}$$

where R is the rotation matrix and t is the translation matrix.

$$X_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{1.9}$$

where
$X_c \rightarrow$Vector representing a point in space with respect to the camera frame
$X_w \rightarrow$ Vector representing a point in space with respect to the world frame
$C_w \rightarrow$ Vector pointing from the origin of the world frame to the origin of the camera frame.
The equation can be rewritten in homogeneous form as follows:

$$\tilde{X}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{1.10}$$

The Extrinsic matrix can be written as

$$M_{ext} = \begin{bmatrix} R_{3\times3} & t \\ 0_{3\times3} & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.11}$$

The relation between the coordinates of a point in the camera frame and pixel frame is given by:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{1.12}$$

which can be written as $\tilde{u} = M_{int}\tilde{X}_c$
The relation between the coordinates of a point in the world frame and camera frame is given by:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & O_x & 0 \\ 0 & f_y & O_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{1.13}$$

which can be written as $\tilde{X}_c = M_{ext}\tilde{X}_w$. The product of extrinsic and intrinsic matrix gives the projection matrix P. Hence a point in 3D space can be represented in the image plane in 2D

using the following relation:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{1.14}$$

In this case, we need a minimum of **6** points to solve this mathematically. This is because the rotation matrix has 3 degrees of freedom (yaw, pitch, and roll), the intrinsic matrix has 5 degrees of freedom (three for focal length and two for skew and principle point), and the translation vector has 3 degrees of freedom (x, y, and z). Therefore, in total, we need 11 parameters to determine the intrinsic and extrinsic matrices of the camera.

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & h_5 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n & -v_n \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \end{bmatrix}$$

By taking the Singular Value Decomposition (SVD) of the above matrix, the projection matrix is obtained. The Gram Schmidt process is applied to decompose the projection matrix into the rotation, translation and intrinsic matrices.

1.3 Result

The steps discussed in the previous section are used to calculate the projection matrix, intrinsic matrix, translation matrix and rotation matrix.

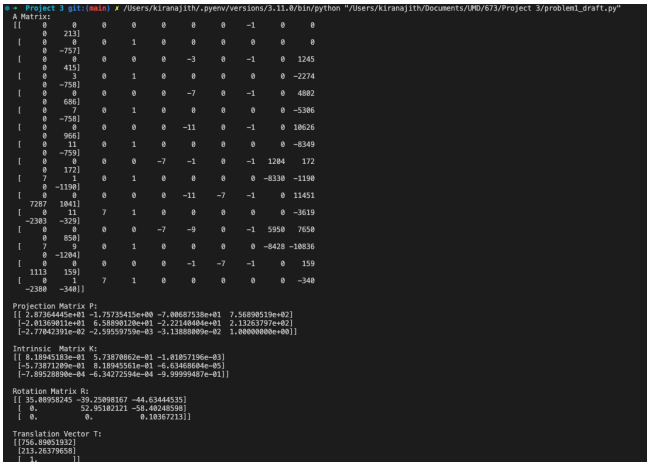

Figure 1.2: Output showing the P,K,T matrices

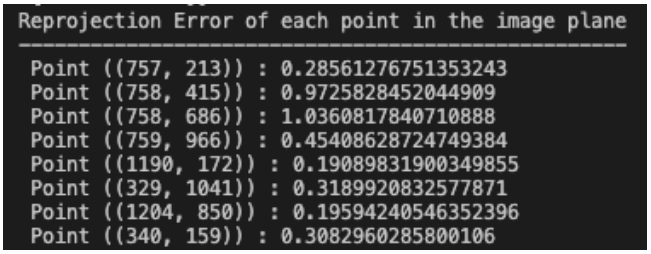

Figure 1.3: Output showing the reprojection error of each point

2

2.1 Problem

The task is to perform camera calibration. It is assumed that the given camera is a pinhole camera model, and radial distortion is ignored. Also given is a calibration target which is used to estimate the camera parameters.

2.2 Pipeline

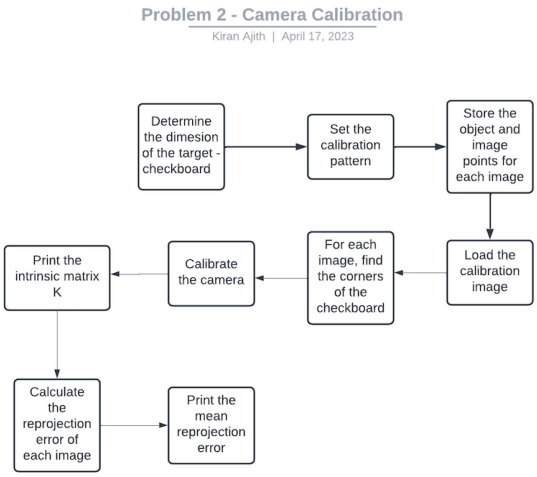

Figure 2.1: Flowchart representing the various processes involved

The checkboard size $(9, 6)$ and the dimension of the checkboard squares$(21.6mm)$ are given. The initial step involves setting up the checkboard calibration pattern by forming an array of 3D points that represent the corners of the checkboard. For each calibration image, we store the 3D object points and the 2D image points in an empty list. Then we iterate through each image, convert them to grayscale, find the corners in the image using the inbuilt openCV function $cv2.ChessboardCorners$. If corners are detected, they are stored in the 3D object point and 2D image point lists. Having obtained the coordinates of the corners, we then draw the corners on the image. After that, the camera is calibrated using the $cv2.calibrateCamera$ function for all the images. Then the reprojection error is calculated for each image, and the average reprojection error is obtained.

2.3 Result

Following are the intermediate and final outputs

Figure 2.2: Figure showing the K matrix

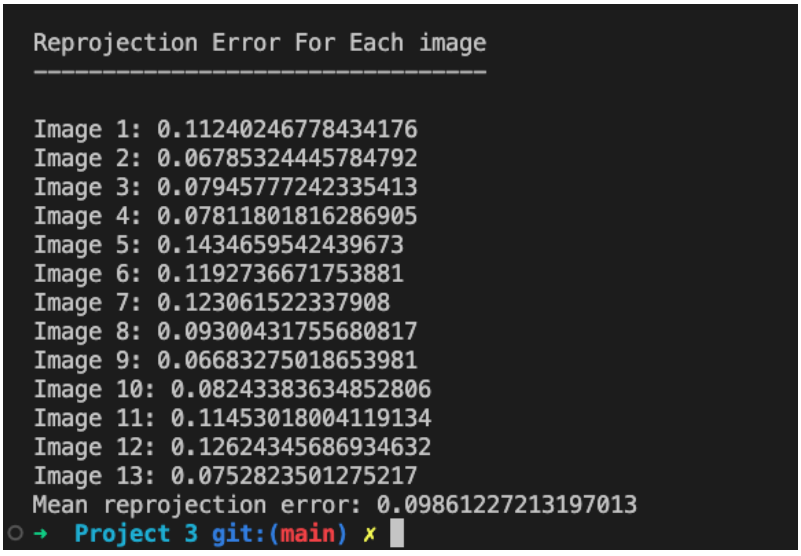

Figure 2.3: Figure showing the reprojection error for each image and the mean reprojection error

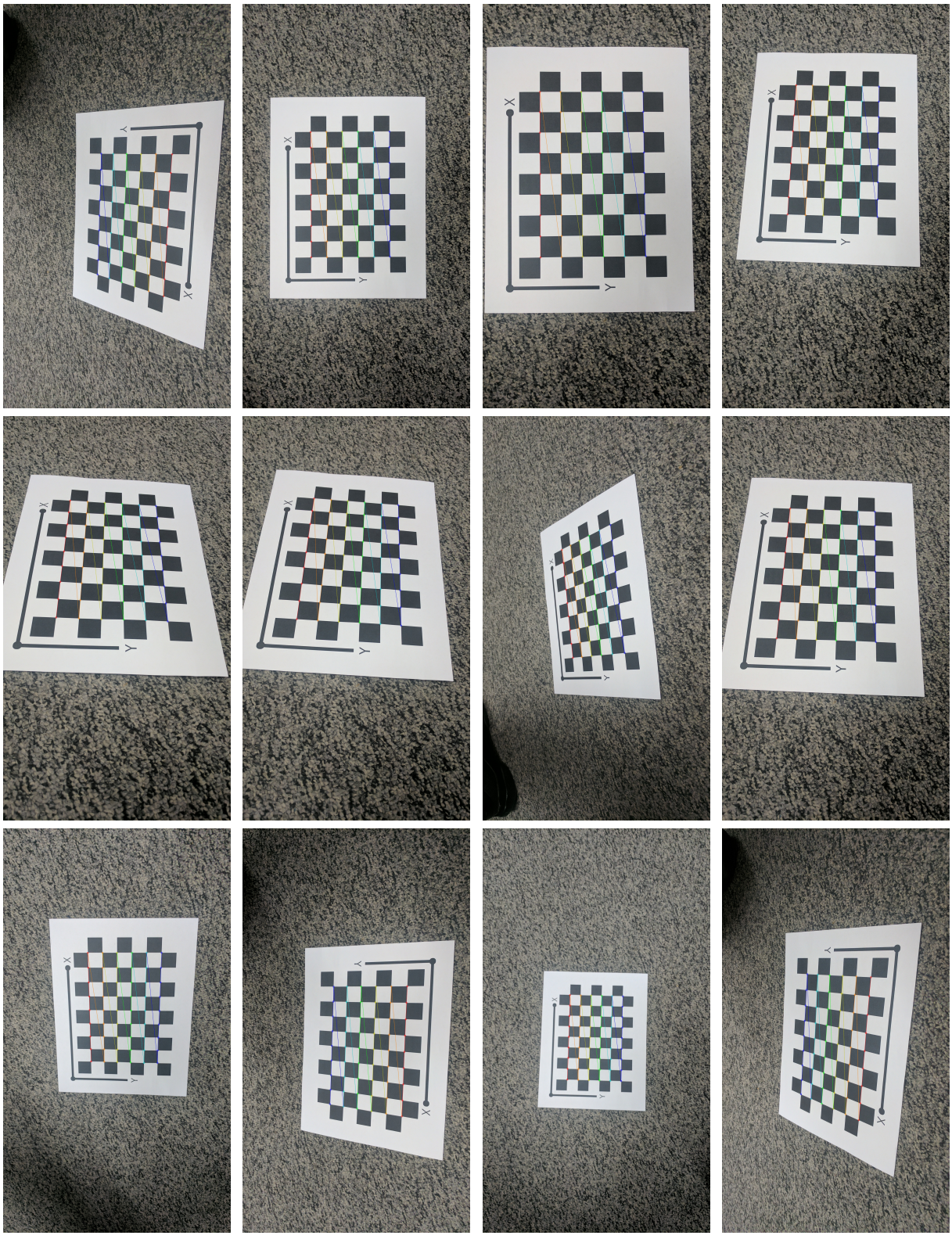

Figure 2.4: Output generated from the code

2.4 Improving the accuracy of the K matrix

Following are some of the ways by which the accuracy of the intrinsic matrix can be improved

- By varying the position and orientation of the calibration images, the distortion coefficients may be estimated more accurately, and the inaccuracy of the K matrix can be reduced.
- Using a calibration target that is more precise, with smaller squares and more regularly spaced corners, will increase the calibration accuracy and deliver more precise corner measurements.
- The use of a more sophisticated camera calibration method, such as the Zhang or Bouguet methods, accounts for lens and radial distortion. While requiring more calibration photos and processing time, these methods can produce more precise calibration results.