RBE549 HW1 AutoCalib Report

Peter Dentch
BS/MS Robotics Engineering
Worcester Polytechnic Institute
Worcester, MA
pdentch@wpi.edu

***Abstract*—This paper details the software implementation of a camera calibration algorithm. The utilized method is described in "A flexible new technique for camera calibration" by Z. Zhang**

I. Introduction

The localization of objects in 3D space viewed by a camera has many important applications in computer vision and robotics engineering. This requires two reference frame transformations, first to transform between the 2D pixel coordinates on the image sensor plane to the camera's 3D frame, then from the camera frame to the 3D world coordinates. The matrices representing these transforms are known as the camera intrinsic and extrinsic matrices respectively, shown in (1).

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R_W^C \mid T_W^C \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

A method for deriving the parameters which make up these matrices, generally known as camera calibration, presented in [1] is widely used in current industry and has been implemented in this project using Python. First an analytical solution to calculating these parameters is performed, then with the computed parameters acting as initial conditions, an optimization method will be employed for more precisely determining the values.

II. Initial Parameter Estimation

A checkerboard pattern of known grid size is printed and secured to a flat surface. By photographing this pattern with the camera at different positions and orientations and identifying the pixel coordinates of the detected corners, these corner locations can be compared to those of an expected grid without lens distortion. An expected grid, defined simply by an array of the point locations in millimeters beginning at the picture origin, is related to a sample image taken of the checkerboard by a homography. "Fig. 1" shows a sample checkerboard image transformed by its computed homography to align with the expected grid.

This project used 13 checkerboard pattern images, giving 13 different homograpy matrices which can be used to compute the intrinsic and extrinsic camera parameters.

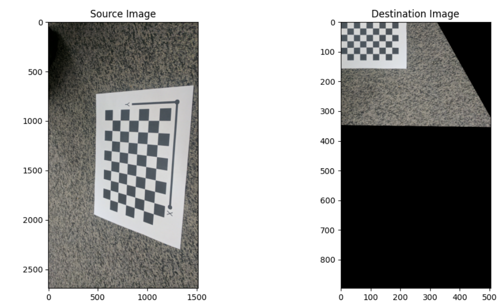

Fig. 1. Example raw and transformed image of the checkerboard pattern

A. Estimating Intrinsic Parameters

The intrinsic parameters make up the K matrix, also known as A, and are defined in (2).

$$K = A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

From [1] a vector b is defined as:

$$b = \begin{bmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{bmatrix}^T$$

the elements of which can be used to derive the intrinsic parameters. To compute the b vector, the following relation is used:

$$Vb = 0 \tag{3}$$

with V as a 2n x 6 matrix, n being the number of images or homography samples from the checkerboard patters. V is given as:

$$V = \begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix}$$

$$v_{ij} = \begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2} + h_{i1}h_{j2} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1} + h_{i1}h_{j3} \\ h_{i3}h_{j2} + h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix}$$

with h_i as the ith column of the homography matrix H. Utilizing single value decomposition to solve for b using relation (3), the A matrix can be formed from (2) after the intrinsic parameters are computed using the b vector.

B. Estimating Extrinsic Parameters

Once the intrinsic parameters are known, they can be used to compute the extrinsic parameters, the matrix $\left[R_W^C \mid T_W^C\right]$ in (1). From [1] they are given:

$$r_1 = \lambda A^{-1} h_1$$

$$r_2 = \lambda A^{-1} h_2$$

$$t = \lambda A^{-1} h_3$$

with $\lambda = 1/\|A^{-1} h_1\|$

III. Non-linear Geometric Error Minimization

The computed extrinsic and intrinsic parameters can then be used to project the expected checkerboard grid to match the captured image. The projected checkerboard corner coordinates can then be compared to the actual detected corner coordinates on each image using a cost function. The error between the expected and actual points is a result of the lens distortion, represented using the radial distortion coefficients k_1 and k_2. This cost function can then be minimized using the least squares method, treating the distortion error as noise and resulting in an optimal estimate for the coefficients.

IV. Results

The outputs generated from the code are as follows. The A matrix from the analytical computation is:

$$A = \begin{bmatrix} 2.06894637e+03 & -3.06883377e+00 & 7.61085218e+02 \\ 0.00000000e+00 & 2.05743292e+03 & 1.36295247e+03 \\ 0.00000000e+00 & 0.00000000e+00 & 1.00000000e+00 \end{bmatrix}$$

Then using the best estimate from error minimization we have:

$$A = \begin{bmatrix} 2.06894649e+03 & -3.06867955e+00 & 7.61086249e+02 \\ 0.00000000e+00 & 2.05743299e+03 & 1.36295339e+03 \\ 0.00000000e+00 & 0.00000000e+00 & 1.00000000e+00 \end{bmatrix}$$

$$k_1 = 0.00020586838307057412$$

$$k_2 = -0.0002318193726762886$$

with an average re-projection error of:

$$e = 0.8100222555464107$$

The checkerboard pattern images were then corrected for the lens distortion using the radial distortion coefficients.

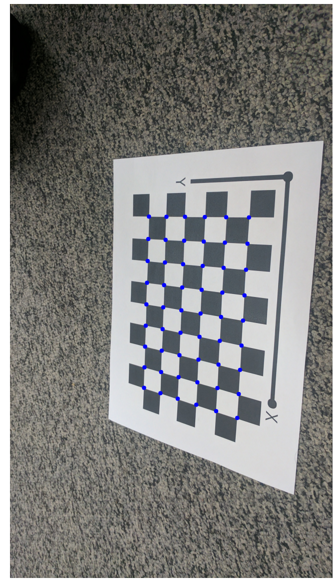

Fig. 2. Example checkerboard pattern image corrected for distortion

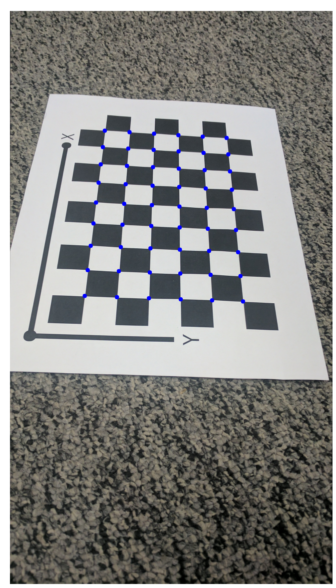

Fig. 3. Example checkerboard pattern image corrected for distortion

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.