Homework 1 - AutoCalib

Niranjan Kumar Ilampooranan
MS Robotics Graduate Student
Worcester Polytechnic Institute

***Abstract*—The objective of the assignment is to estimate the intrinsic and extrinsic parameters, given a calibration target and a bunch of other images to be used for calibration. This method is inspired by the work of Zhang et al. [1]. After obtaining the initial estimates, it is further fine-tuned by minimizing the error using least squares. Finally, the mean value and the final parameters are reported.**

INTRODUCTION

The image captured by the camera of any object depends on various parameters, some intrinsic properties to the camera itself (the image axes skewness and scale factors), and some would be extrinsic - the pose of the camera. This information is essential in the complete mapping of the world (metric) co-ordinates [X , Y] to pixel coordinates [u , v]. This assignment focuses on obtaining good estimates of these parameters and is essentially done by following this paradigm.

- Obtain the target pattern - in this case, a checkerboard, along with its pictures in different orientations with the camera
- Estimate the homography between the target and the calibration images
- Using the above information, obtain an initial estimate of the intrinsic and the extrinsic parameters
- With the initial guess, obtain the radial distortion coefficients and better estimates of the parameters by minimizing the least square error

INTRINSIC PARAMETER ESTIMATION

Before estimating the camera intrinsics, the Homography between the target image and each image taken using the camera is derived. This can be obtained by finding the feature matches (minimum of 4 pairs) between the two images in question, here it would be the checkerboard corners (only the squares in the center region of 6×9 is considered). Finally, the homography can be obtained by singular value decomposition of the following matrix (x, y are the corner locations in the target image and x_p, y_p are the locations of the corners in the calibration images).

$$A = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -x(x_p) & -y(x_p) & -x_p \\ 0 & 0 & 0 & x & y & 1 & -x(y_p) & -y(y_p) & -y_p \end{bmatrix}$$

Stacking such rows to this matrix for each pair would give a minimum of 8 rows. Singular value decomposition of this matrix to the 9×9 orthogonal matrix V is key in obtaining the H matrix. The next step would be estimating the camera intrinsics. The camera intrinsics consist of the following parameters.

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

To obtain the individual parameters, symmetric matrix B is used ($\mathbf{B} = \mathbf{A}^T\mathbf{A}$). Some of the values in B ($\mathbf{b}$ = $[\mathbf{B}_{11}, \mathbf{B}_{12}, \mathbf{B}_{22}, \mathbf{B}_{13}, \mathbf{B}_{23}, \mathbf{B}_{33}]$, where the subscripts correspond to the row and column number respectively) can be used to solve for this parameters exclusively.

For each of the images, another matrix V can be built using specific linear combination of the homography matrix columns. This matrix is of size $2n \times 6$, where n is the number of images. Solving for $\mathbf{b}$ in $\mathbf{Vb} = \mathbf{0}$ can be done by taking the eigenvector of $\mathbf{V}^T\mathbf{V}$ that corresponds to the smallest eigenvalue. With b, the parameters can be solved for, to give the entire $\mathbf{A}$ matrix.

EXTRINSIC PARAMETER ESTIMATION

With the values of $\mathbf{A}$ matrix, the extrinsic parameters - rotation matrix and translation vector - can be obtained easily by leveraging the properties of rotation matrices. Given that the rotation matrix is orthonormal, the following can be equated to obtain $\mathbf{R}$ and $\mathbf{t}$.

$$\mathbf{r}_1 = \lambda \mathbf{A}^{-1}\mathbf{h}_1$$

$$\mathbf{r}_2 = \lambda \mathbf{A}^{-1}\mathbf{h}_2$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$$

$$\mathbf{t} = \lambda \mathbf{A}^{-1}\mathbf{h}_3$$

where $\mathbf{r}_i$ indicates column i of rotation matrix and $\mathbf{h}_i$ indicates column i of homography matrix. λ is obtained by averaging the norms of $\mathbf{r}_1$ and $\mathbf{r}_2$.

With $\mathbf{R}$ and $\mathbf{t}$ obtained, the next focus would be on obtaining the radial distortion coefficients and making better estimates of the parameters.

Projection Error and Final Estimates

With the initial estimates of the parameters along with the distortion coefficients set to 0, the following equation is used to obtain the error between the true and predicted value.

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

,where n indicates the number of images, m indicates the total number of corners in each image, and $\mathbf{m}_{ij}$ is the pixel coordinate of the corner j in image i.

While this iterative process continues, the pixel coordinate is updated using the distortion coefficients as shown below.

$$\breve{u} = u + (u - u_0)(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2)$$

$$\breve{v} = v + (v - v_0)(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2)$$

where $\breve{u}, \breve{v}$ indicate the real pixel coordinates (due to distortion), u_0, v_0 coordinates of the principal point, and u, v are the pixel coordinates in the ideal scenario of no distortion. Also, k_1, k_2 are the distortion coefficients.

After using least squares to minimise the error and converge to better estimates, the following values are obtained for comparison (Hats above indicate initial estimates).

$$\hat{\mathbf{A}} = \begin{bmatrix} 2057.93 & 2.31 & 750.89 \\ 0 & 2046.87 & 1350.25 \\ 0 & 0 & 1 \end{bmatrix}$$

Initial error obtained = 2.955

$$\mathbf{A} = \begin{bmatrix} 2048.28 & -0.02 & 761.94 \\ 0 & 2037.67 & 1356.54 \\ 0 & 0 & 1 \end{bmatrix}$$

$$k = \begin{bmatrix} 0.021 & -0.122 \end{bmatrix}$$

With these estimates, **2.65** was recorded as the mean re-projection error across all images. This was obtained using the equation below.

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

The following image subset serves as comparison between the original image with its corners marked, and its re-projected corners in the same image.

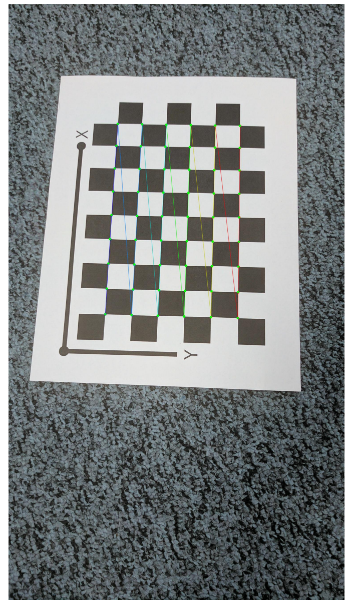

Fig. 1. Corners in Image 1

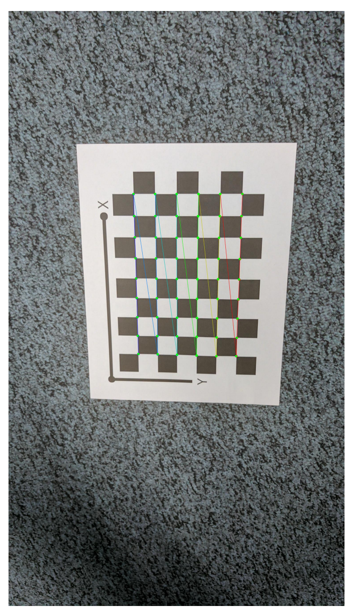

Fig. 2. Corners in Image 8

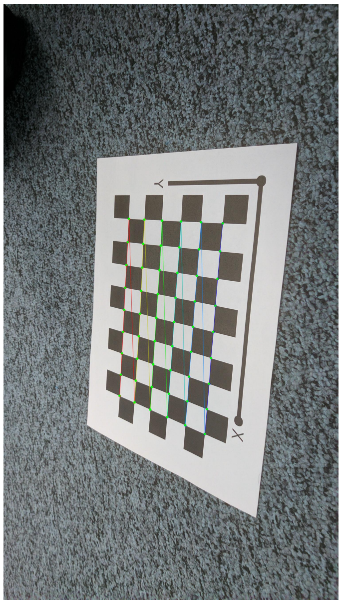

Fig. 3. Corners in Image 12

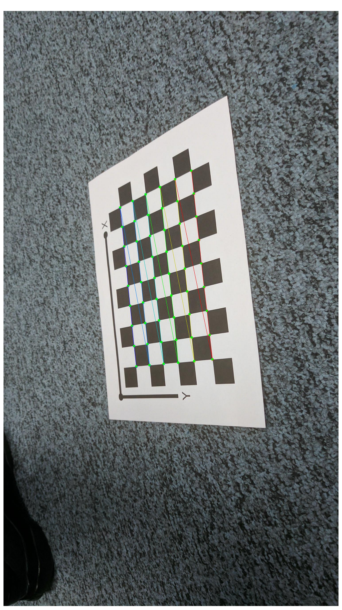

Fig. 4. Corners in Image 13

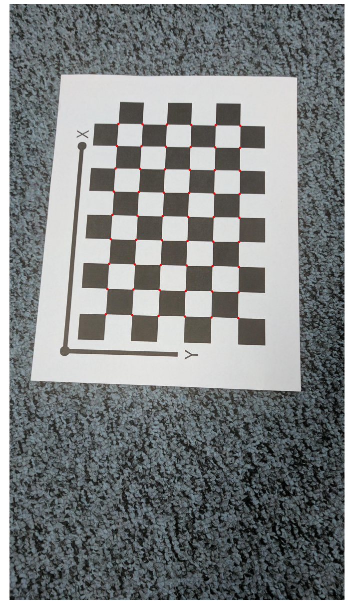

Fig. 5. Re-projected corners in Image 1

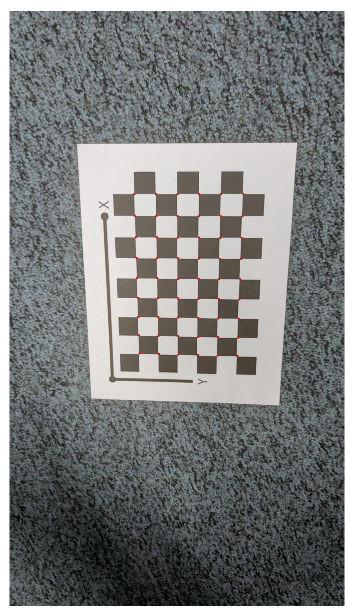

Fig. 6. Re-projected corners in Image 8

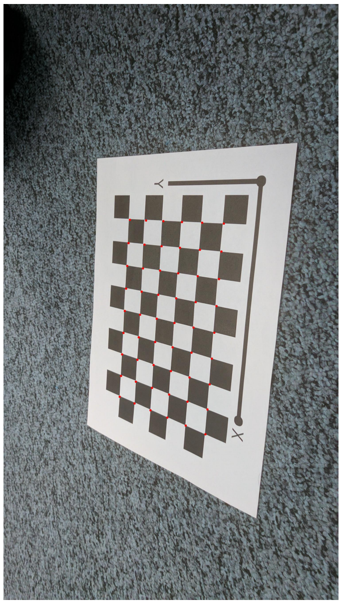

Fig. 7. Re-projected corners in Image 12

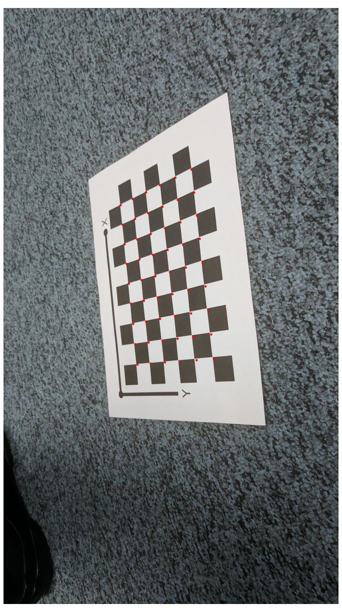

Fig. 8. Re-projected corners in Image 13

References

[1] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 1330 – 1334, 12 2000.