Homework 1: AutoCalib!

Rutwik S. Kulkarni
Department of Robotics Engineering
Worcester Polytechnic Institute
Email: rkulkarni1@wpi.edu

***Abstract*—This project addresses the crucial aspect of camera calibration, a key process in the field of computer vision. It involves accurately determining a camera's intrinsic features, such as the focal length and the position of the principal point, as well as its extrinsic parameters and distortion coefficients. Camera calibration is fundamental to the effectiveness and precision of various computer vision tasks. In this endeavor, we will implement the acclaimed calibration methodology developed by Zhengyou Zhang, a technique indispensable for the detailed comprehension and accurate depiction of the unique properties of cameras used in computer vision applications.**

I. Introduction

Camera calibration is a key process in computer vision, crucial for accurately mapping the three-dimensional world to a two-dimensional image plane. This process corrects for systematic camera errors, such as lens distortion, and is vital for applications requiring precision in image interpretation, such as 3D reconstruction, computer graphics, robotics, and augmented reality. By calibrating a camera, one can ensure a more accurate representation of the real world in images, a critical requirement for many computer vision tasks.

In this calibration, we use Zhang's method, known for its practicality and ease of implementation. This method requires just a planar pattern, like a checkerboard, displayed in various orientations and positions before the camera, making it a straightforward approach compared to traditional methods that need complex setups. Camera calibration involves determining intrinsic parameters, which include the camera's focal length, optical center, and lens distortion characteristics, and extrinsic parameters, which define the camera's position and orientation in the world. These parameters are essential for the correct interpretation of images, as they dictate how a camera projects 3D world points onto its 2D image sensor.

II. Data

The dataset for our camera calibration consists of thirteen images of a 10x7 checkerboard pattern, captured with a Google Pixel XL smartphone. During the capture process, the camera's focus was fixed to ensure uniformity across all images. For camera calibration, the full checkerboard grid is not utilized. Instead, only a smaller section, specifically the inner 9x6 grid, is considered. This selective approach mitigates the potential distortions and inaccuracies typically found at the checkerboard's edges. Each square in the checkerboard measures 21.5mm, providing a precise scale for calibration. The exact dimensions of these squares are crucial, as they offer a scale reference essential in determining the camera's intrinsic parameters, such as focal length. The accuracy in these measurements is fundamental in computing spatial relationships in the images, forming a solid base for effective camera calibration.

III. Solving for Camera Intrinsics and Extrinsics

A. Detecting Checkerboard Corners

The calibration begins with the detection of corners on a checkerboard pattern using the `findChessboardCorners` function in OpenCV. For a 10x7 grid checkerboard, excluding the outer grid, there are 54 corners in each image. These corners are vital for establishing the relationship between 2D image points (on the checkerboard) and their corresponding 3D world points.

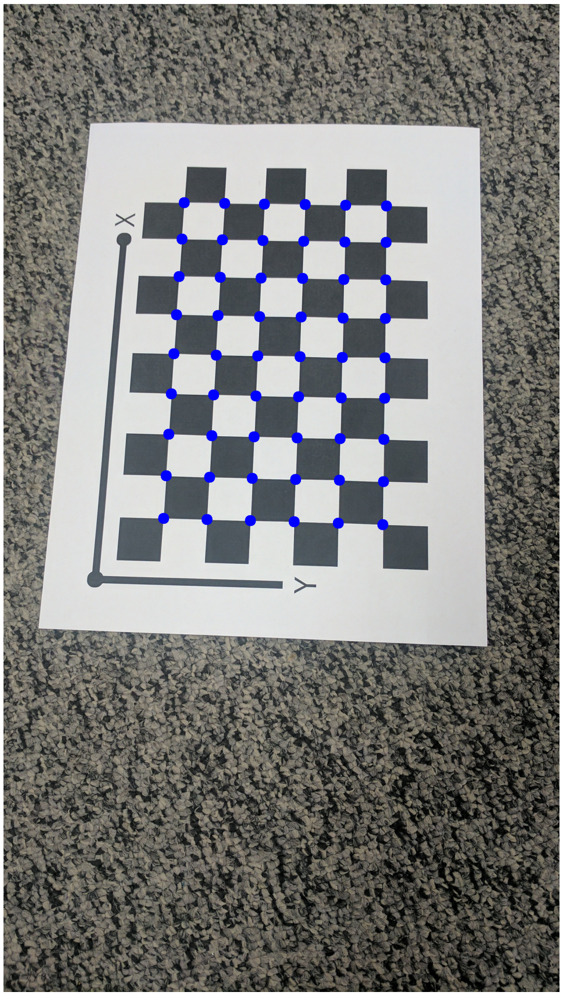

Fig. 1: Image 1 corners with distortions

B. Intrinsic Camera Matrix K

The intrinsic camera matrix, denoted as K, includes the camera's internal parameters and is represented as:

$$K = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

In this matrix, f_x and f_y are the focal lengths in pixels in the horizontal and vertical directions, respectively, c_x and c_y represent the coordinates of the principal point (optical center), and γ is the skew coefficient, which is typically zero in cameras with square pixels.

C. Homography Matrix Calculation

For each image, a homography matrix H is computed. This matrix relates the 3D world coordinates to the 2D image coordinates and is given by:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

The relationship between the image coordinates (x, y) and world coordinates (X, Y) is expressed as:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = H \cdot \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}$$

This equation is then reformulated into a linear system $Ah = 0$, where h is the flattened homography matrix and A is constructed from the corresponding points. The solution for h is found using Singular Value Decomposition (SVD).

D. Employing SVD for $Ah = 0$ Form

The intrinsic parameters are solved by rearranging the equations into the form $Ah = 0$ using SVD. This involves decomposing the homography matrix and formulating linear equations that relate the homography to the camera's intrinsic parameters.

E. B Matrix Computation

For each homography matrix H, six elements are computed using the function $V_{ij}(H)$, defined as:

$$V_{ij}(H) = \begin{pmatrix} h_{1i}h_{1j} \\ h_{1i}h_{2j} + h_{2i}h_{1j} \\ h_{2i}h_{2j} \\ h_{3i}h_{1j} + h_{1i}h_{3j} \\ h_{3i}h_{2j} + h_{2i}h_{3j} \\ h_{3i}h_{3j} \end{pmatrix}$$

The matrix B is then obtained by stacking the $V_{ij}(H)$ vectors for all homographies and solving $VB = 0$ using SVD, where V is the stacked matrix of V_{ij} vectors.

F. K Matrix Computation

From the matrix B, which is related to the intrinsic matrix as $B = K^{-T}K^{-1}$, the elements of the intrinsic matrix K are extracted using the following relationships:

$$\lambda = B_{22} - \frac{(B_{02}^2 + v_0(B_{01}B_{02} - B_{00}B_{12}))}{B_{00}}$$

$$\alpha = \sqrt{\frac{\lambda}{B_{00}}}$$

$$\beta = \sqrt{\frac{\lambda B_{00}}{B_{00}B_{11} - B_{01}^2}}$$

$$\gamma = -\frac{B_{01}\alpha^2\beta}{\lambda}$$

$$u_0 = \frac{\gamma v_0}{\beta} - \frac{B_{02}\alpha^2}{\lambda}$$

Intrinsic Matrix (K):

$$K = \begin{bmatrix} 2055.77 & -0.306 & 763.77 \\ 0.00 & 2038.8 & 1348.308 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

G. Decomposition into Intrinsic and Extrinsic Matrices

Once the homography matrices are computed, they are decomposed to separate the intrinsic matrix K and the extrinsic parameters (rotation R and translation T). This is crucial for distinguishing the camera's internal characteristics from its spatial position and orientation.

H. Extrinsics Computation

The extrinsic parameters (rotation R and translation t) for each image are determined as follows:

$$R := K^{-1}H, \quad \text{where } R \in \mathbb{R}^{3\times3}$$

$$t := \frac{R_{:,3}}{\|R_{:,1}\|}, \quad \text{where } t \in \mathbb{R}^{3\times1}$$

These parameters signify the camera's position and orientation in the world.

IV. Approximate Distortion k

A. Estimation of Distortion Coefficients

The distortion in a camera lens is characterized by a set of distortion coefficients, collectively denoted as k. These coefficients account for radial and tangential distortions that occur due to the lens geometry and manufacturing imperfections. In the calibration process, we aim to estimate these coefficients to correct for such distortions.

The distortion coefficients are approximated using the detected chessboard corners and the intrinsic matrix K. The process involves:

1) Mapping the observed 2D points to the undistorted 3D space using the intrinsic matrix K.
2) Applying an iterative optimization algorithm to minimize the reprojection error, which is the difference between the observed 2D points and the projected 2D points from the 3D model, considering radial and tangential distortions.
3) The resulting coefficients $k = [k_1, k_2, p_1, p_2, \ldots]$ represent the radial and tangential distortion parameters, where k_1, k_2 are radial distortion coefficients and p_1, p_2 are tangential distortion coefficients.

V. Non-Linear Geometric Error Minimization

A. Refining Camera Calibration

The process of minimizing non-linear geometric errors is essential for refining the camera calibration. This step aims to reduce the discrepancies between the observed image points and the projected points from the estimated camera model.

B. Error Minimization Formulation:

The optimization is formulated as the minimization of the sum of squared differences between the observed image points $x_{i,j}$ and the reprojected points $\hat{x}_{i,j}(K, R_i, t_i, X_j, k)$. Mathematically, this is represented as:

$$\sum_{i=1}^{N}\sum_{j=1}^{M} \|x_{i,j} - \hat{x}_{i,j}(K, R_i, t_i, X_j, k)\|^2$$

where N is the number of images, M is the number of points per image, K is the intrinsic matrix, R_i and t_i are the rotation and translation for image i, X_j are the coordinates of the 3D point corresponding to the j-th point in the world space, and k represents the distortion coefficients.

C. Optimization Objective:

The objective of the optimization process is to find the intrinsic parameters f_x, f_y, c_x, c_y and distortion coefficients k_1, k_2 that minimize the reprojection error. This is expressed as:

$$\text{argmin}_{f_x,f_y,c_x,c_y,k_1,k_2} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{M} \|x_{i,j} - \hat{x}_{i,j}(K, R_i, t_i, X_j, k)\|^2 \right\} \quad (1)$$

D. Implementation:

The implementation of this optimization involves:

1) Employing a non-linear optimization technique, typically the Levenberg-Marquardt algorithm, to iteratively adjust the camera parameters (intrinsic and extrinsic) and distortion coefficients.
2) The optimization is performed by minimizing the sum of squared differences between the observed image points and the reprojected points using the current estimates of camera parameters and distortion coefficients.
3) The optimization continues until the change in error between iterations falls below a predefined threshold, indicating that the model has converged to a solution.

E. Outcome:

The outcome of this process is a refined set of camera parameters and distortion coefficients that provide a more accurate representation of the camera's characteristics, leading to improved performance.

F. Final Parameter Estimation

Post optimization, the final set of parameters achieved is as follows:

Intrinsic Matrix (K):

$$K = \begin{bmatrix} 2055.77 & -0.31 & 763.77 \\ 0.00 & 2038.79 & 1348.32 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

Distortion Coefficients (k):

$$k = \begin{bmatrix} 0.012570 \\ -0.089845 \end{bmatrix}$$

Reprojection Error:

$$ReprojectionError = 0.75141$$

G. Results

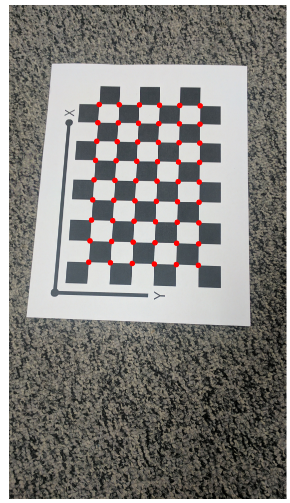

Fig. 2: Image 1 corners without distortions

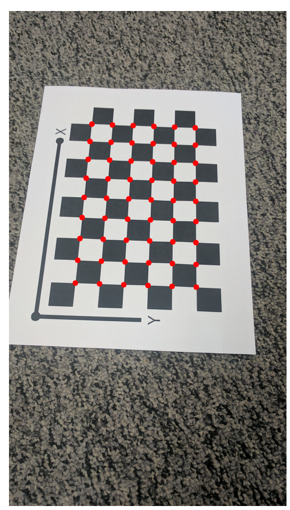

Fig. 3: Image 2 corners without distortions

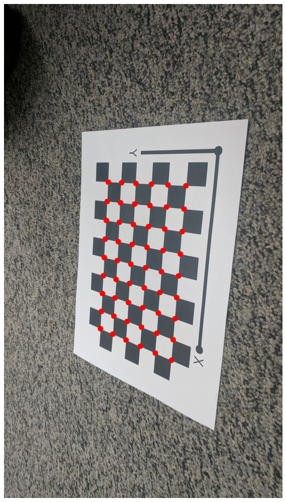

Fig. 4: Image 12 corners without distortions

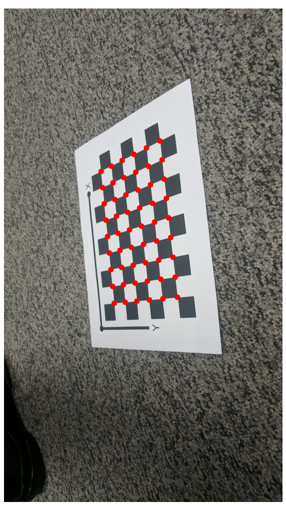

Fig. 5: Image 13 corners without distortions

H. Conclusion of Optimization

The optimization results yield a modest improvement in the camera model, evident from the updated parameters and reprojection error. This progress, while beneficial, highlights the need for more extensive distortion parameters beyond $k1$ and $k2$. Future work should aim to expand the distortion model to achieve greater precision in high-stakes computer vision applications, thereby enhancing the overall effectiveness of the camera calibration process.

VI. 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 Link**