RBE/CS 549 Computer Vision HomeWork 1: Auto-Calib

Puneet Shetty
Robotics Engineering Department
Worcester Polytechnic Institute
Worcester, MA, USA
Email: ppshetty@wpi.edu

***Abstract*—Camera calibration refers to the process of estimating camera parameters such as focal length, distortion coefficients, and principal point. It is a crucial and time-consuming component of any computer vision study that uses three-dimensional geometry. In this assignment, we calibrate the camera by calculating its intrinsic and extrinsic parameters and modeling any potential image distortion using a radial-tangential model that is based on Zhang's study on camera calibration. We estimate the relationship between a 2D point in space and its associated image point based on the intrinsic, extrinsic, and distortion model of the camera. We then optimize the error between actual points and points following distortion correction.**

***Index Terms*— Calibration, Intrinsic, Extrinsic, Radial distortion, Focal length, Principle points, Optimization**

I. Introduction

Estimating a camera's properties is the process of camera calibration. In other words, we possess every piece of information required to accurately establish a connection between a 3D point in the real world and its corresponding 2D projection in the image captured by that calibrated camera, including parameters and coefficients.The process involves solving the problem in closed form first, then refining it non-linearly using the maximum likelihood criterion. The steps taken to execute this process are:

1) Create a pattern on paper and adhere it to a flat surface.
2) Move the model plane or the camera to capture a few shots of it in various positions.
3) Locate the image's feature spots (corners)
4) Calculate the Homography between the image feature points and the calibration target
5) Estimate all of the extrinsic parameters as well as the intrinsic parameters using the closed-form solution.
6) Determine the radial distortion coefficients by initially assuming that they are 0.
7) Use the optimization problem to minimize the geometric error, which will refine all the values.
8) Re-project the values and determine the Camera Calibration Matrix

II. Data Extraction

A. Calibration Dataset

The Zhang research uses a calibration target (checkerboard in our case) to estimate camera intrinsic parameters. The calibration target used is shown in Figure 1. This was printed on A4 paper, and each square measured 21.5mm. Thirteen photos were captured with the focus locked on a Google Pixel XL phone and will be used for calibration.
A 2D point is denoted by $\mathbf{m} = [u, v]^T$. A 3D point is denoted by $\mathbf{M} = [X, Y, Z]^T$. We use $\tilde{\mathbf{m}}$ to denote the augmented vector by adding 1 as the last element: $\tilde{\mathbf{m}} = [u, v, 1]^T$, and $\tilde{\mathbf{M}} = [X, Y, Z, 1]^T$. A camera is modeled by the usual pinhole: the relationship between a 3D point $\mathbf{M}$ and its image projection $\mathbf{m}$ is given by

$$s\tilde{\mathbf{m}} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\tilde{\mathbf{M}}$$

where s is an arbitrary scale factor, $[\mathbf{R}\mathbf{t}]$, called the extrinsic parameters, is the rotation and translation which relates the world coordinate system to the camera coordinate system, and $\mathbf{K}$, called the camera intrinsic matrix, is given by

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

with (u_0, v_0) being the coordinates of the principal point, α and β the scale factors in image u and v axes, and γ the parameter describing the skewness of the two image axes.

B. Corner Detection

Finding the pixel coordinates of each image's chess board corners is the first step. The cv2.findChessboardCorners method is used to find points. The number of inner corners that need to be identified is (9,6), which is the pattern size parameter. For every picture, a total of 54 corner points are located.

C. Homography Estimation

The homography between the world and picture coordinates is specified up to scale. The homogeneous system of linear equations is constructed using the homographies from each calibration picture. Eigen decomposition or singular value decomposition can be used to determine the system's solution. Our initial camera intrinsics estimations originate from this solution.

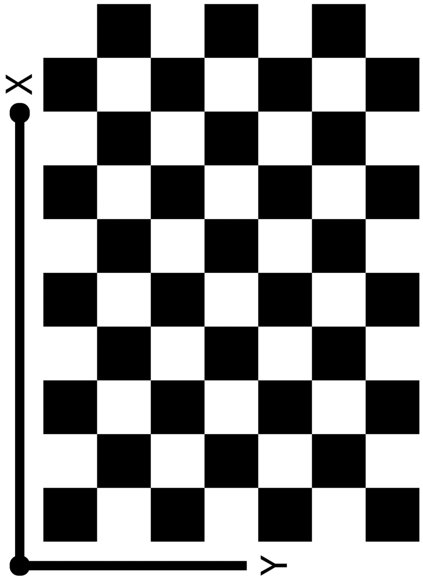


Fig. 1. CheckerBoard Pattern

For our calculations, we assume the model plane is on $Z = 0$ of the world coordinate system. We denote the i-th column of the rotation matrix R by r_i. From this, we get

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

$$= \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}.$$

By abuse of notation, we still use M to denote a point on the model plane, but $M = [X, Y]^T$ since Z is always equal to 0. In turn, $\tilde{M} = [X, Y, 1]^T$. Therefore, a model point M and its image m are related by a homography H:

$$s\tilde{m} = H\tilde{M} \text{ with } H = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}$$

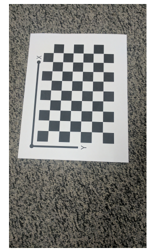
As is clear, the 3×3 matrix H is defined up to a scale factor.

Using cv2.findHomography, we get the Homography matrix from these locations. We will obtain thirteen H matrices in total after this step. The computed Homography matrix is then used to derive the intrinsic parameters.

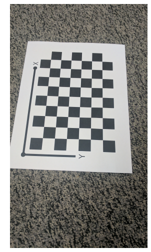
III. Parameter Estimation

We're attempting to obtain an accurate initial estimate of the parameters so that we can input it into the non-linear optimizer. Next, we'll define the parameters used in the code.

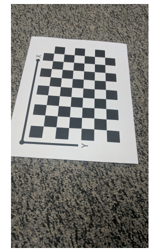
The image points are denoted by x, the world points by X, the radial distortion parameters by k=[k1,k2], the camera calibration matrix by K, and the rotation and translation of the camera in the world frame by R and t, respectively.



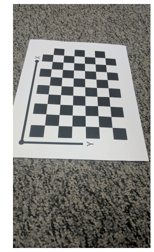
(a) Image 1



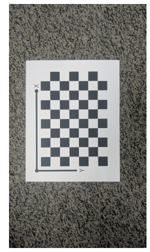
(b) Image 2



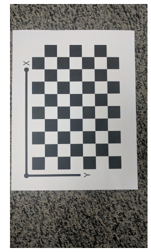
(c) Image 3



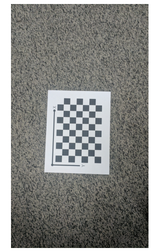
(d) Image 4



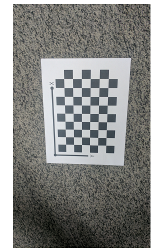
(e) Image 5



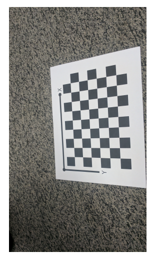
(f) Image 6



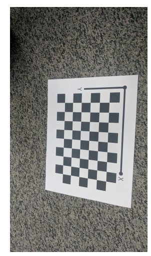
(g) Image 7



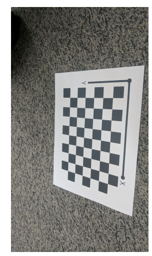
(h) Image 8



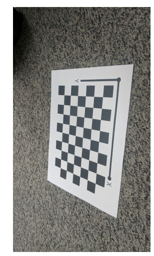
(i) Image 9



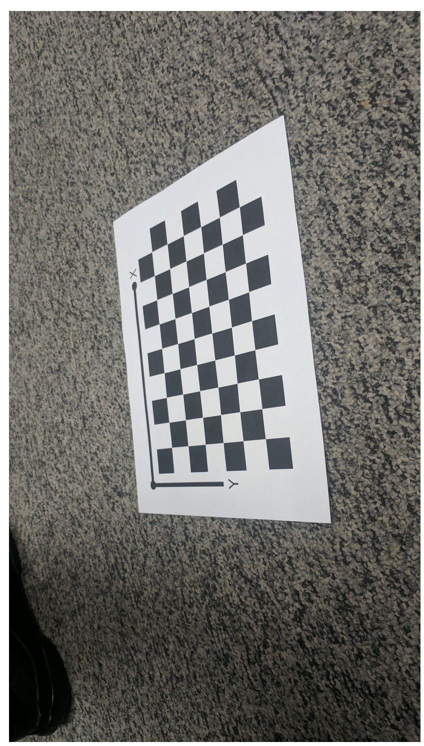
(j) Image 10



(k) Image 11

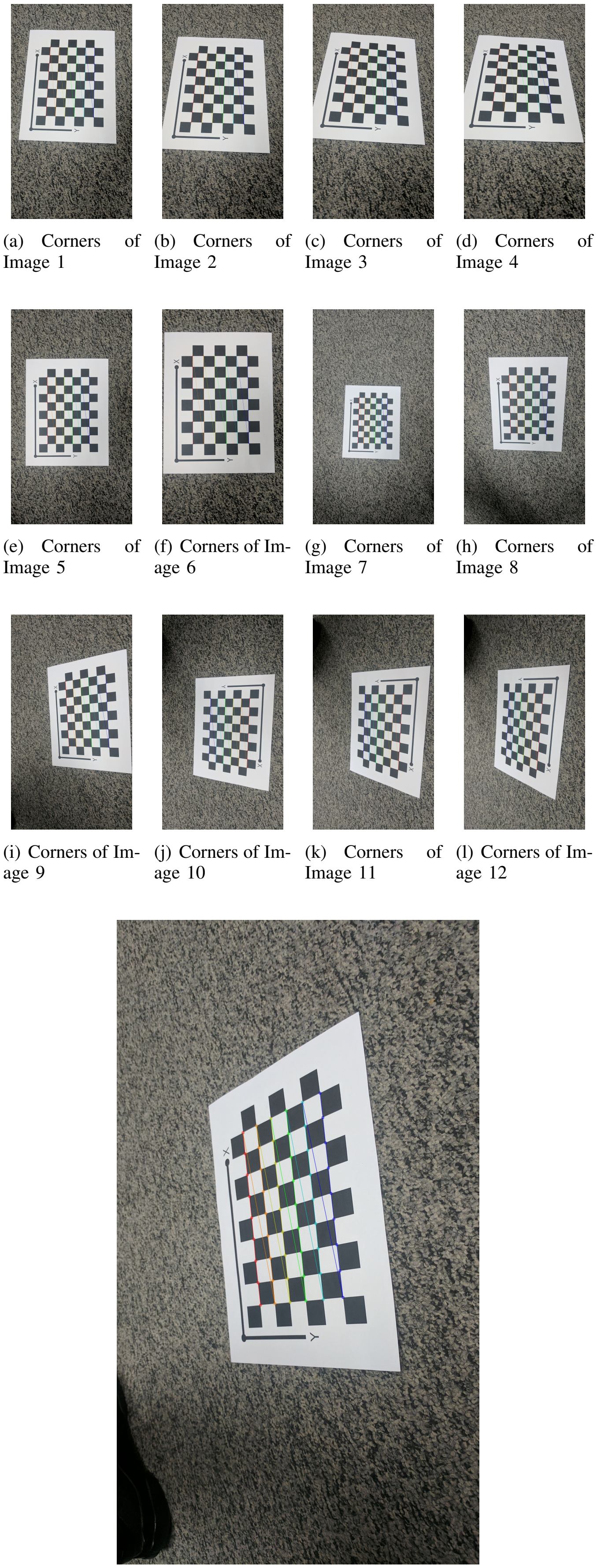


(l) Image 12



(m) Image 13

Fig. 2. Input Images



(a) Corners of Image 1
(b) Corners of Image 2
(c) Corners of Image 3
(d) Corners of Image 4
(e) Corners of Image 5
(f) Corners of Image 6
(g) Corners of Image 7
(h) Corners of Image 8
(i) Corners of Image 9
(j) Corners of Image 10
(k) Corners of Image 11
(l) Corners of Image 12
(m) Corners of Image 13

Fig. 3. Corners of Input Images

A. Closed-Form Solution

We can denote the columns if the homography matrix by $H = \begin{bmatrix} h1 & h2 & h3 \end{bmatrix}$. From the homography calculated, we can get:

$$\begin{bmatrix} h1 \\ h2 \\ h3 \end{bmatrix} = \lambda \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix},$$

where λ is an arbitrary scalar. Using the knowledge that $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal, we have:

$$h_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} h_2 = 0 \quad ,$$

$$h_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} h_1 = h_2^T \mathbf{K}^{-T} \mathbf{K}^{-1} h_2$$

Let

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

$$B_1 = \begin{bmatrix} 1/\alpha^2 \\ -\gamma/(\alpha^2\beta) \\ (v_0\gamma - u_0\beta)/(\alpha^2\beta) \end{bmatrix}$$

$$B_2 = \begin{bmatrix} -\gamma/(\alpha^2\beta) \\ (1/\beta^2) + (\gamma^2/(\alpha^2\beta^2)) \\ -\gamma(v_0\gamma - u_0\beta)/(\alpha^2\beta^2) - v_0/\beta^2 \end{bmatrix}$$

$$B_3 = \begin{bmatrix} (v_0\gamma - u_0\beta)/(\alpha^2\beta) \\ -\gamma(v_0\gamma - u_0\beta)/(\alpha^2\beta^2) - v_0/\beta^2 \\ (((v_0\gamma - u_0\beta)^2)/(\alpha^2\beta^2)) + ((v_0^2)/\beta^2) + 1 \end{bmatrix}$$

Note that B is symmetric,hence can be defined by a 6D vector

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

Let the i^{th} column vector of H be $h_i = \begin{bmatrix} h_{i1} \\ h_{i2} \\ h_{i3} \end{bmatrix}$. Then we have:

$$h_i^T B h_j = v_{ij}^T b$$

Let the ith column vector of H be $\mathbf{h}_i = \begin{bmatrix} h_{i1} \\ h_{i2} \\ h_{i3} \end{bmatrix}^T$.

With $\mathbf{v}_{ij} = \begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2} + h_{i2}h_{j1} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1} + h_{i1}h_{j3} \\ h_{i3}h_{j2} + h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix}$.

Therefore, the two fundamental constraints from a given homography, can be rewritten as 2 homogeneous equations in b:

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

If there are n images of the model plane are observed, by stacking such equations we get

$$V\mathbf{b} = 0.$$

B. Camera Intrinsic Parameters

Using the equation for v_{ij}, we compute the V matrix in order to determine the intrinsic parameters.The solution to $V\mathbf{b} = 0$. is well known as the eigenvector of V^TV associated with the smallest eigenvalue. When we solve for V using singular value decomposition(SVD),it yields $\mathbf{b}$.This $\mathbf{b}$ can be mapped to the matrix B. Next, $\alpha, \beta, \gamma, u_0, v_0$ are calculated using this B. Without difficulty, we can uniquely extract the intrinsic parameters from the matrix B.

$$v_0 = \frac{B_{12}B_{13} - B_{11}B_{23}}{B_{11}B_{22} - B_{12}^2}$$

$$\lambda = B_{33} - \frac{B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})}{B_{11}}$$

$$\alpha = \sqrt{\frac{\lambda}{B_{11}}}$$

$$\beta = \frac{\sqrt{\lambda B_{11}}}{B_{11}B_{22} - B_{12}^2}$$

$$\gamma = -\frac{B_{12}\alpha^2\beta}{\lambda}$$

$$u_0 = \frac{\gamma v_0}{\beta} - \frac{B_{13}\alpha^2}{\lambda}$$

The camera intrinsic matrix, is given by

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

C. Camera Extrinsic Parameters

The extrinsic parameters for every image can be readily determined if K is known. Each homography's extrinsic parameters are now calculated, and all of these values are appended to the extrinsic matrix.

$$\mathbf{r}_1 = \lambda \mathbf{A}^{-1}\mathbf{h}_1 \quad (9)$$

$$\mathbf{r}_2 = \lambda \mathbf{A}^{-1}\mathbf{h}_2 \quad (10)$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \quad (11)$$

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

with

$$\lambda = \frac{1}{\|\mathbf{A}^{-1}\mathbf{h}_1\|} = \frac{1}{\|\mathbf{A}^{-1}\mathbf{h}_2\|}$$

D. Distortion Parameters

We can assume that the distortion matrix, that is $\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ is equal to $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for a reasonable initial estimate since we assumed that the camera has negligible distortion.

IV. Error Optimization & Reprojection

The mathematical norm of the difference between the coordinates of the observed corners and the coordinates of the corners reprojected using the estimated parameters—also known as the geometric error—was selected as the error to be minimized. To do the non-linear minimization of this reprojection error, all of the various parameters that were to be optimized were combined into a single vector and supplied to the scipy library's optimize.minimize() function. We emply the Powell method in the minimize function to optimize our error.

$$\text{Geometric Error: } \sum_{i=1}^{N}\sum_{j=1}^{M}\|\mathbf{x}_{i,j} - \hat{\mathbf{x}}_{i,j}(\mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j, \mathbf{k})\|$$

Optimization Function:

$$\text{argmin}_{f_x, f_y, c_x, c_y, k_1, k_2} \sum_{i=1}^{N}\sum_{j=1}^{M}\|\mathbf{x}_{i,j} - \hat{\mathbf{x}}_{i,j}(\mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j, \mathbf{k})\|$$

V. Conclusion

Before the optimization, K matrix and distortion were:

$$K = \begin{bmatrix} 2.0652566 \times 10^3 & -2.9397471 & 7.6467615 \times 10^2 \\ 0 & 2.0534834 \times 10^3 & 1.3627693 \times 10^3 \\ 0 & 0 & 1 \end{bmatrix}$$

$$Distortion = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

After the optimization the values changed to:

$$K = \begin{bmatrix} 2.06105926 \times 10^3 & -2.82052649 & 7.64706888 \times 10^2 \\ 0 & 2.04942358 \times 10^3 & 1.36286775 \times 10^3 \\ 0 & 0 & 1 \end{bmatrix}$$

$$Distortion = \begin{bmatrix} 0.07122434 \\ -0.38982751 \end{bmatrix}$$

The errors for the 13 images reduced to:

Image	Initial Error	New Error
1	0.692	0.526
2	0.712	0.593
3	0.766	0.671
4	0.818	0.737
5	0.789	0.698
6	0.781	0.678
7	0.757	0.693
8	0.726	0.667
9	0.721	0.656
10	0.710	0.642
11	0.738	0.663
12	0.783	0.704
13	0.782	0.710

TABLE I
COMPARISON OF INITIAL AND NEW ERRORS FOR EACH IMAGE TAKING MEAN

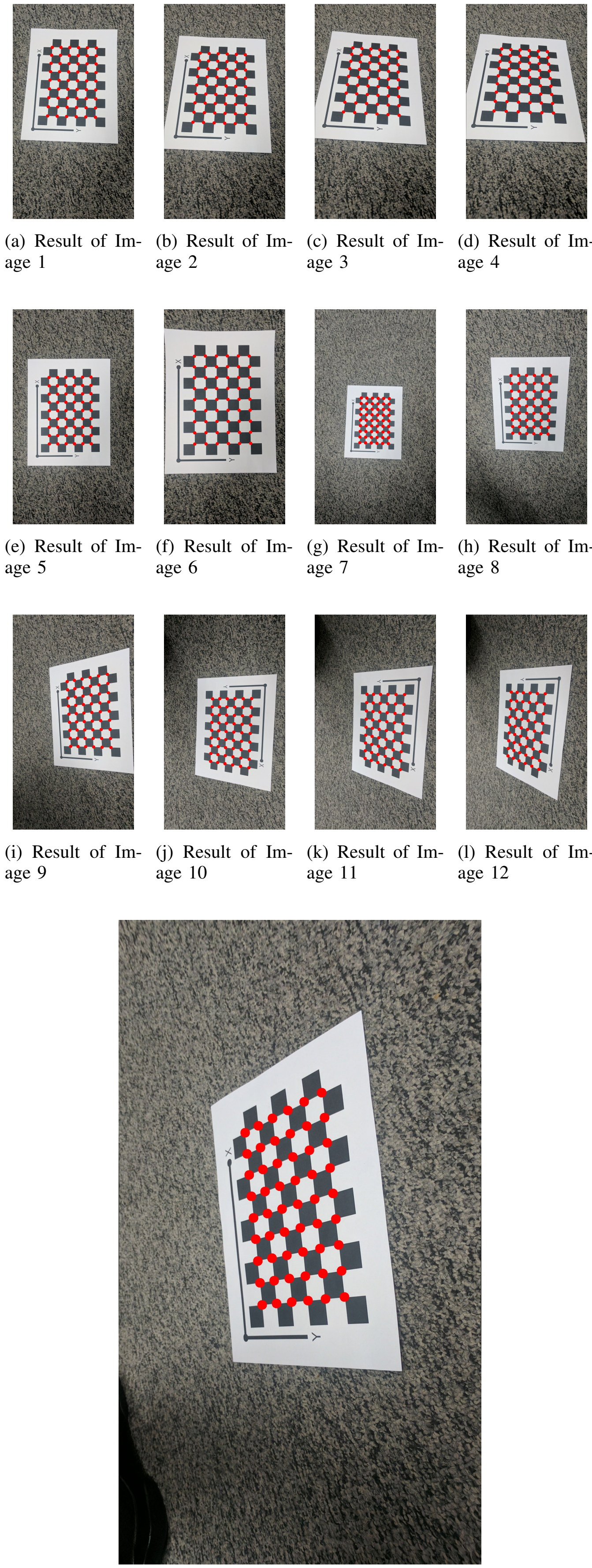
Image	Initial Error	New Error
1	37.37	28.93
2	76.86	64.60
3	124.10	109.85
4	176.61	160.81
5	213.00	189.52
6	253.01	221.16
7	286.14	261.47
8	313.69	286.70
9	350.25	317.55
10	383.43	345.02
11	438.25	391.75
12	507.19	453.75
13	548.78	494.96

TABLE II
COMPARISON OF INITIAL AND NEW ERRORS FOR EACH IMAGE TAKING SUMMATION

In this study, we employed Zhang's camera calibration method to initially calibrate the camera. The obtained results demonstrated a satisfactory calibration with reasonable accuracy. However, seeking further improvement, an optimization process was employed to refine the calibration parameters.

The optimization process led to noticeable enhancements in the calibration accuracy. The refined intrinsic parameters, distortion coefficients, and error values exhibited superior performance compared to the initial calibration. This outcome signifies the effectiveness of optimization techniques in fine-tuning camera calibration, resulting in a more precise mapping between 3D world coordinates and 2D image coordinates. The final reprojected results are shown in Fig. 4.

In conclusion, the combination of Zhang's method for the initial calibration and subsequent optimization yielded superior results, providing a calibrated camera with enhanced accuracy, which is crucial for various computer vision and image processing applications



(a) Result of Image 1 (b) Result of Image 2 (c) Result of Image 3 (d) Result of Image 4

(e) Result of Image 5 (f) Result of Image 6 (g) Result of Image 7 (h) Result of Image 8

(i) Result of Image 9 (j) Result of Image 10 (k) Result of Image 11 (l) Result of Image 12

(m) Result of Image 13

Fig. 4. Results of Input Images