AutoCalib!
Computer Vision (RBE549) Homework 1

Prasham Soni
Masters of Science
Robotics Engineering
Worcester Polytechnic Institute
Worcester , Massachusetts 01609
Email: psoni@wpi.edu

***Abstract*—This paper presents an implementation of the well-known camera calibration technique introduced by Zhengyou Zhang [1] in "A Flexible New Technique for Camera Calibration." The approach uses one or more images of a planar calibration target, typically a checkerboard, observed from different viewpoints. By extracting corner correspondences between the planar pattern and the image, the method first computes homographies to estimate the initial intrinsic camera parameters, including focal lengths, principal point, and the skew coefficient. Next, the extrinsic parameters—rotation and translation with respect to the calibration target—are determined. Finally, a non-linear optimization step refines all parameters jointly, simultaneously estimating lens distortion coefficients to account for real-world imperfections. This flexible procedure requires only a flat, easily portable pattern, making it practical for diverse environments and camera setups, while delivering accurate calibration and distortion modeling that are crucial for many computer vision tasks.**

I. INTRODUCTION

Camera calibration is a fundamental process in computer vision that aims to determine a camera's internal parameters (intrinsic) and its orientation and position relative to a known scene (extrinsic). Accurate calibration is critical for tasks such as 3D reconstruction, object tracking, and augmented reality, where precise understanding of the camera's geometry directly influences performance and reliability. Among various methods available, the flexible approach proposed by Zhengyou Zhang has become a standard due to its simplicity and effectiveness. Zhang's method uses multiple views of a single planar pattern—a checkerboard—to reliably estimate both the camera's intrinsic parameters (focal length, principal point, skew) and lens distortion coefficients. Subsequently, these estimates are further refined through non-linear optimization, minimizing the reprojection error and increasing accuracy.

In this paper, we implement Zhang's calibration technique with the following key steps:
1. **Detect checkerboard corners**: Identify the correspondence between the known 2D pattern coordinates(world coordinates) and their projections in the image(image coordinates).
2. **Compute initial camera parameter**s: Estimate the camera's intrinsic matrix and extrinsic parameters using homographies derived from the detected corners.
3. **Refine parameters via non-linear optimization**: Simultaneously optimize intrinsic parameters (including distortion coefficients) and extrinsic parameters to minimize the overall reprojection error.
This procedure remains widely applicable and highly practical because it relies only on a flat, easily portable calibration target, making it versatile for diverse camera setups and real-world conditions.

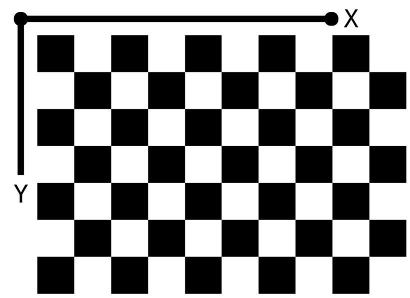

Fig. 1. Checkerboard pattern

A. System Setup: Data

Following Zhang's method, we acquired multiple images of a planar checkerboard target (fig:1) from diverse viewpoints to facilitate camera calibration. The checkerboard's known geometry—such as a 9×6 grid of internal corners and squares measuring 21.5mm per side—provides precise real-world dimensions. All images were taken under well-controlled lighting conditions(By Google Pixel XL phone with focus locked) to improve corner detection accuracy and ensure high contrast. By capturing the target at various angles and distances, the dataset covers a wide range of perspectives. Subsequently, the detected corner coordinates in each image are mapped to their corresponding positions on the checkerboard plane, enabling the estimation of both the intrinsic parameters (focal lengths, principal point, skew, and distortion) and the extrinsic parameters (rotation and translation) of the camera.

B. Detect checkerboard corners

After establishing the dataset of checkerboard images, we extract the corresponding image coordinates and world coordinates in two main steps. First, we utilize OpenCV's `findChessboardCorners` function to detect the internal corners in the grayscale images. Once the corners are located, they undergo sub-pixel refinement to ensure minimal error, which is especially critical for stable calibration results.

In parallel, we generate the real-world coordinates by iterating over the expected grid layout on the checkerboard plane. Each square's known physical size (e.g., 21.5,mm per side) provides a precise positional reference for each corner in the plane. Consequently, for each image, we obtain pixel coordinates (image coordinates) and their corresponding 2D planar coordinates (world coordinates), forming the core input for the subsequent estimation of intrinsic and extrinsic camera parameters.

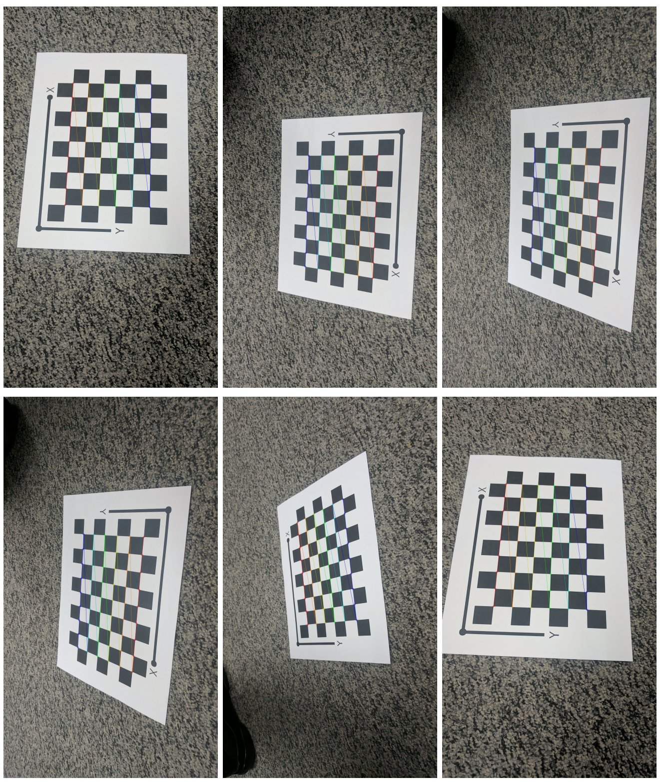

Fig. 2. Corners Detected by `findChessboardCorners`

C. Compute initial camera parameters

1) Defining Notations: Following Zhang's notation, we denote a 2D image point by

$$\mathbf{m} = \begin{bmatrix} u \\ v \end{bmatrix},$$

and a 3D world point by

$$\mathbf{M} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

To work conveniently with homogeneous coordinates, we augment these vectors by appending 1 as the last element:

$$\widetilde{\mathbf{m}} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \quad \widetilde{\mathbf{M}} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

A pinhole camera projects a point $\mathbf{M}$ in the world onto the image plane as $\mathbf{m}$, according to:

$$s\,\widetilde{\mathbf{m}} = \mathbf{A}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\widetilde{\mathbf{M}},$$

where s is a projective scale factor, and the pair $(\mathbf{R}, \mathbf{t})$ (the *extrinsic parameters*) defines the rotation and translation from the world coordinate system to the camera coordinate system. The matrix $\mathbf{A}$ (the *intrinsic matrix*) is given by:

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

where (u_0, v_0) is the principal point, α and β are scale factors along the u and v axes, respectively, and γ models any skew between those axes. For convenience, we abbreviate $\mathbf{A}^{-T}$ to mean $\left(\mathbf{A}^{-1}\right)^T$ or equivalently $\left(\mathbf{A}^T\right)^{-1}$.

When the target is a planar object (e.g., a checkerboard with $Z = 0$), the relationship between any point $\widetilde{\mathbf{M}}$ on that plane and its image $\widetilde{\mathbf{m}}$ can be expressed via a 3×3 homography $\mathbf{H}$. The subsequent sections detail how we use these notations and transformations to estimate $\mathbf{H}$ from matched corner correspondences.

2) Computing Homography: Assuming that the calibration plane is located at $Z = 0$ in the world coordinate system, any point on the model plane can be represented as $\mathbf{M} = [X,\, Y]^T$ with homogeneous coordinates $\widetilde{\mathbf{M}} = [X,\, Y,\, 1]^T$. Its corresponding image point is denoted as $\mathbf{m} = [u,\, v]^T$ with homogeneous coordinates $\widetilde{\mathbf{m}} = [u,\, v,\, 1]^T$.

Under the pinhole camera model, the projection is given by

$$s\,\widetilde{\mathbf{m}} = \mathbf{A}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\widetilde{\mathbf{M}},$$

where $\mathbf{A}$ is the intrinsic matrix and $(\mathbf{R}, \mathbf{t})$ are the extrinsic parameters. Since $Z = 0$ for all points on the calibration plane, the equation reduces to

$$s\,\widetilde{\mathbf{m}} = \mathbf{A}\begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}\end{bmatrix}\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix},$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the first two columns of the rotation matrix $\mathbf{R}$. This can be succinctly written as

$$s\,\widetilde{\mathbf{m}} = \mathbf{H}\widetilde{\mathbf{M}},$$

with the homography matrix defined by

$$\mathbf{H} = \mathbf{A}\begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}\end{bmatrix}.$$

It is important to note that $\mathbf{H}$ is determined only up to an arbitrary scale factor.

Estimating $\mathbf{H}$ via the Direct Linear Transform (DLT) To estimate $\mathbf{H}$ from a set of corresponding points, we first construct a linear system that encapsulates the relationship between the known world coordinates and their image projections. For each correspondence $(X, Y) \leftrightarrow (u, v)$, two linear equations are formulated that relate the elements of $\mathbf{H}$ to the coordinates. By assembling these equations into a matrix equation of the form

$$\mathbf{A}\,\mathbf{h} = \mathbf{0},$$

where $\mathbf{h}$ is a 9-dimensional vector representation of the homography matrix, the solution is obtained through Singular Value Decomposition (SVD). The singular vector corresponding to the smallest singular value is selected as the solution, since it minimizes the least-squares error. Finally, the homography is normalized (typically by fixing the scale such that the bottom-right element equals 1), resolving the inherent scale ambiguity.

This Direct Linear Transform (DLT) method is central to our calibration process, as it provides the homography that maps the known model points to their image projections without repetition of the underlying derivation. This homography matrix encodes both the intrinsic and extrinsic matrices, capturing the transformation from the 3D world to the 2D image plane

*3) **Extracting the Intrinsic Matrix**:* To compute the intrinsic matrix $\mathbf{A}$ from the estimated homographies, we follow a four-step procedure:

a) Step 1: Exploiting the Constraints: For each homography $\mathbf{H}$ (obtained from a view of the calibration plane with $Z = 0$), the first two columns of the rotation component must be orthonormal. This yields the constraints:

$$\mathbf{h}_1^T\,\mathbf{B}\,\mathbf{h}_2 = 0 \quad \text{and} \quad \mathbf{h}_1^T\,\mathbf{B}\,\mathbf{h}_1 = \mathbf{h}_2^T\,\mathbf{B}\,\mathbf{h}_2,$$

where $\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1}$ is a symmetric matrix encoding the intrinsic parameters.

b) Step 2: Defining the Constraint Matrix via v_{ij}: For each homography, we define a vector v_{ij} that collects specific products of the elements of the ith and jth columns of $\mathbf{H}$. In practice, the vectors v_{12} and the difference $v_{11} - v_{22}$ are computed to compactly express the above constraints as linear equations in the unknown entries of $\mathbf{B}$.

c) Step 3: Solving the Homogeneous System: Stacking the linear equations from all views leads to a homogeneous system:

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

where $\mathbf{b}$ is a 6-dimensional vector containing the independent entries of $\mathbf{B}$. This system is solved using Singular Value Decomposition (SVD), with the solution given by the singular vector corresponding to the smallest singular value. The vector $\mathbf{b}$ is then reshaped to form the symmetric matrix $\mathbf{B}$.

d) Step 4: Decomposing $\mathbf{B}$ to Recover $\mathbf{A}$: The intrinsic parameters α, β, γ, u_0, and v_0 are then extracted by expressing the elements of $\mathbf{B}$ in terms of these parameters. In particular, we compute:

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

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

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

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

Finally, the intrinsic matrix is assembled as:

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

This four-step procedure leverages the geometric constraints imposed by the homographies, uses the v_{ij} vectors to form and solve a linear system, and finally decomposes the resulting matrix to recover the camera's intrinsic parameters.

*4) **Estimation of Initial Extrinsic Parameters**:* Once the intrinsic matrix $\mathbf{A}$ has been determined and the homography $\mathbf{H}$ is known for a given view, the next step is to extract the extrinsic parameters—namely, the rotation and translation—that describe the pose of the calibration plane relative to the camera.

The process begins by computing the inverse of the intrinsic matrix, $\mathbf{A}^{-1}$. The columns of the homography $\mathbf{H}$ are denoted by $\mathbf{h}_1$, $\mathbf{h}_2$, and $\mathbf{h}_3$. To eliminate the unknown scale factor inherent in $\mathbf{H}$, a normalization factor λ is calculated as the inverse of the norm of $\mathbf{A}^{-1}\mathbf{h}_1$. Using this scale factor, the first two columns of the rotation matrix are recovered as:

$$\mathbf{r}_1 = \lambda\,\mathbf{A}^{-1}\mathbf{h}_1, \quad \mathbf{r}_2 = \lambda\,\mathbf{A}^{-1}\mathbf{h}_2.$$

The third rotation vector is then obtained by taking the cross product:

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

This cross product ensures that the estimated rotation matrix is as close as possible to an orthogonal matrix, a necessary property for a valid rotation.

Similarly, the translation vector is extracted from the homography by:

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

For further processing, the three rotation vectors $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ are assembled into a rotation matrix $\mathbf{R}$ by stacking them as rows (or columns, depending on the chosen convention). The translation vector $\mathbf{t}$ is reshaped into a column vector. This yields the complete extrinsic parameter set $(\mathbf{R}, \mathbf{t})$ for the view, which describes the pose of the calibration plane in the camera coordinate system.

This method of estimating extrinsic parameters provides an initial approximation that can later be refined using non-linear optimization techniques.

D. Refine Parameters via Non-Linear Optimization

To further improve the accuracy of our calibration, we refine the initial estimates of the intrinsic parameters A, the extrinsic parameters $\{R_i, t_i\}$ for each calibration image, and the lens distortion coefficients k_1 and k_2. Under the assumption that the camera exhibits minimal distortion, the initial distortion estimate is set to $k = [0, 0]^T$.

Given these initial estimates, the objective is to minimize the reprojection error, defined as the Euclidean distance between the observed ground truth corner coordinates $x_{i,j}$ and the reprojected corner coordinates $\hat{x}_{i,j}(A, R_i, t_i, X_j, d)$, where d represents the distortion parameters. The optimization problem is formulated as follows:

$$\min_{A,\{R_i,t_i\}_{i=1}^{N},k_1,k_2} \quad \sum_{i=1}^{N}\sum_{j=1}^{M} \left\| x_{i,j} - \hat{x}_{i,j}(A, R_i, t_i, X_j, d) \right\|^2 ,$$

where:

- N is the number of calibration images,
- M is the number of detected corners per image,
- $x_{i,j}$ is the observed position of the jth corner in the ith image, and
- $\hat{x}_{i,j}(A, R_i, t_i, X_j, d)$ is the corresponding reprojected point computed using the current estimates of the intrinsic parameters A, the extrinsic parameters (R_i, t_i) for the ith image, the world coordinates X_j of the jth corner, and the distortion parameters d.

This joint non-linear optimization adjusts all calibration parameters simultaneously to minimize the overall reprojection error, leading to a more accurate estimation of the camera's parameters. The reprojected corner points after optimization are depicted in *Figure 3 and Figure 4*, which demonstrates the improved alignment between the observed and computed positions.

E. Image Undistortion

After refining the calibration parameters, the next step is to apply these estimates to correct for lens distortion in the captured images. Using the estimated intrinsic matrix K and the distortion coefficients $d = [k_1, k_2]$, the undistortion process remaps the pixel coordinates of a distorted image to their correct positions, ensuring that straight lines in the scene appear straight in the resulting image.

This correction is typically achieved by computing a mapping based on the camera model and applying the inverse of the distortion effects. The undistortion process leverages the same distortion model used during calibration, thereby compensating for radial (and, if applicable, tangential) distortions. As a result, the output image is geometrically accurate and

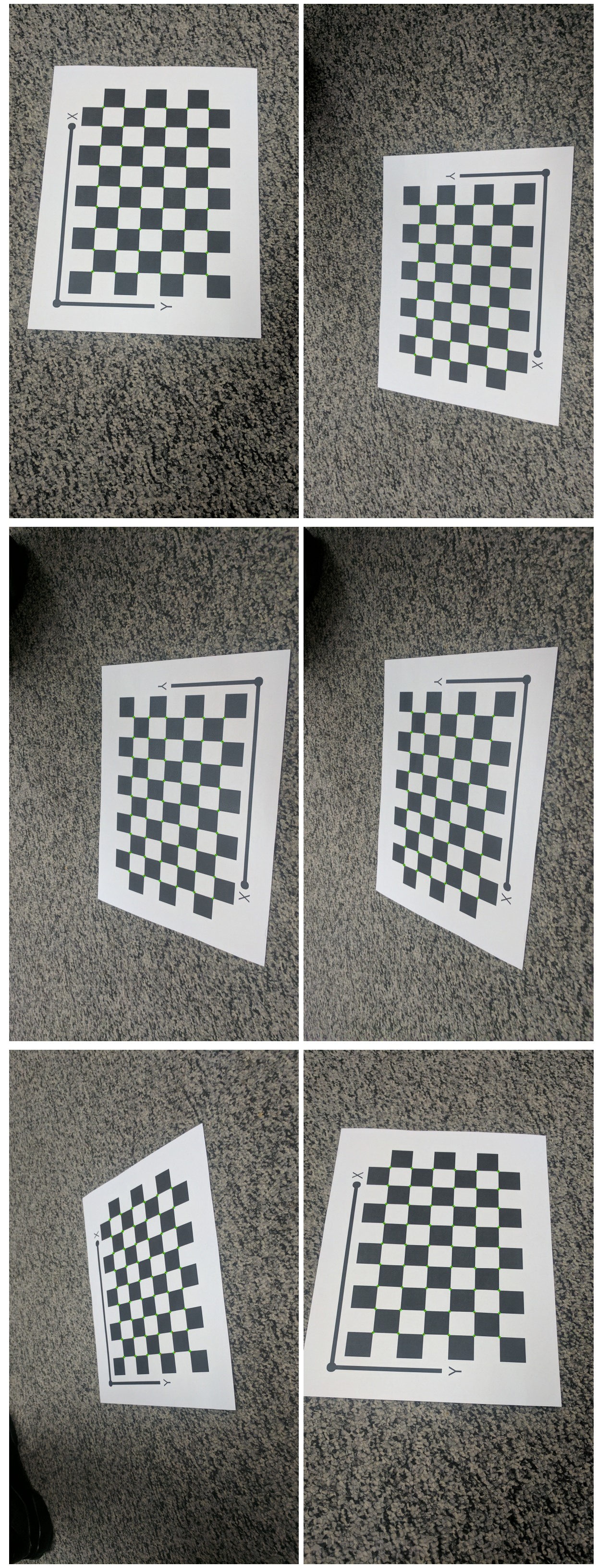

Fig. 3. Reprojected corners for images 1 to 6 after optimization.

suitable for subsequent computer vision tasks such as feature extraction, object detection, or 3D reconstruction.

The undistortion step not only serves as a validation of the calibration accuracy but also significantly enhances the quality and usability of the acquired images. Figure 5 shows several examples of images after undistortion.

II. Results

The calibration outcomes before and after refinement via non-linear optimization.

A. Intrinsic Parameter Estimates

The initial intrinsic parameters estimated for the camera are:

$$K_{\text{initial}} = \begin{bmatrix} 2.05637924 \times 10^3 & -1.12663319 & 761.434539 \\ 0 & 2.04100962 \times 10^3 & 1350.23624 \\ 0 & 0 & 1 \end{bmatrix}.$$

After joint non-linear optimization, the intrinsic matrix was refined to:

$$K_{\text{refined}} = \begin{bmatrix} 2.04853218 \times 10^3 & -1.83007623 & 758.7270 \\ 0 & 2.04074821 \times 10^3 & 1345.14372 \\ 0 & 0 & 1 \end{bmatrix},$$

with refined distortion coefficients of

$$k_1 = 0.173100, \quad k_2 = -0.753355.$$

B. Reprojection Error Analysis

The overall reprojection **error before optimization was 0.825 pixels**, which **improved to 0.720 pixels after optimization**. Table I presents a side-by-side comparison of the per-image reprojection errors (measured as Error) before and after refinement.

TABLE I
REPROJECTION ERRORS BEFORE AND AFTER OPTIMIZATION

Image	Error Before (pixels)	Error After (pixels)
1	0.768	0.595
2	0.672	0.544
3	0.932	0.790
4	1.144	0.928
5	0.881	0.940
6	0.790	0.702
7	0.963	0.915
8	1.112	1.069
9	0.643	0.421
10	0.806	0.664
11	0.369	0.278
12	0.550	0.463
13	0.751	0.584

These results demonstrate a consistent improvement in calibration accuracy after optimization, as evidenced by the reduction in reprojection errors across most images.

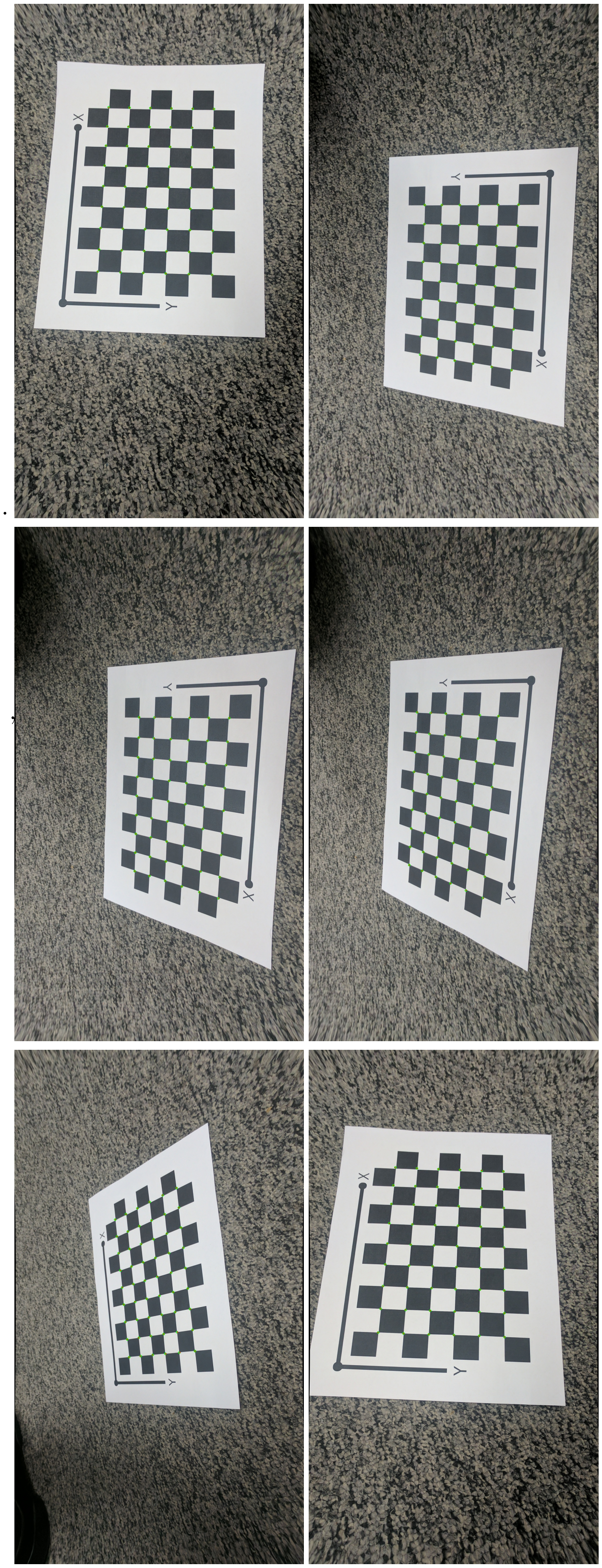

Fig. 4. Undistorted images after calibration, with reprojected corners overlayed on the checkerboard patterns.

III. Conclusion

In this paper, we presented an implementation of Zhang's camera calibration method using a planar checkerboard pattern. The process involved detecting checkerboard corners, computing homographies, and extracting both intrinsic and extrinsic parameters. We further refined these parameters through non-linear optimization, which significantly reduced the overall reprojection error. The final calibration results, as demonstrated by the reduced Error values across all images, confirm the robustness and accuracy of the method.

This work not only validates the effectiveness of Zhang's technique in practical settings but also highlights the importance of parameter refinement for achieving high-precision calibration. Future work may explore further improvements by incorporating more advanced distortion models and optimizing calibration for various camera types and conditions.

IV. References

- Zhang, Z. (2000). *A flexible new technique for camera calibration*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 1330–1334.[1]
- Stachniss, C. (Spring 2020). *Camera Calibration using Zhang's Method*. Slides. [Online]. Available: http://www.ipb.uni-bonn.de/html/te[2]