RBE 549: Homework 1 - AutoCalib (Using 1 late day)

Rigved Sanku
M.S Robotics Engineering
Worcester Polytechnic Institute (WPI)
Worcester, MA 01609
Email: rsanku@wpi.edu

***Abstract*—In this Homework we will be working on calculating camera calibration parameters. Camera calibration means estimating any camera's intrinsics, extrinsics and distortion parameters. Intrinsic parameters consists of focal length and principal point position and distortion parameters are coefficients of distortion. Camera calibration is one of the most important part of any computer vision project. In this homework we will be implementing a well known camera calibration technique proposed by Zhengyou Zhang.**

Index Terms—Camera Calibration, Intrinsic camera parameters, Distortion

I. Introduction

This work aims to calibrate a camera using a checkerboard pattern with a square size of 21.5 mm. The calibration process involves two key steps:

1) **Intrinsic parameter estimation**: This determines the camera's internal characteristics, including the principal point coordinates, scale factors in image axes, and distortion coefficients. These parameters are estimated through a non-linear optimization process that minimizes the projection error between real-world points and their corresponding image locations.
2) **Extrinsic parameter estimation**: This defines the camera's position and orientation in the world coordinate system. It involves estimating the rotation and translation vectors that relate the world and camera coordinate systems.

The initial estimates for both intrinsic and extrinsic parameters are crucial for the optimization process to converge efficiently. The checkerboard pattern serves as a reference grid during calibration, allowing accurate correspondences between real-world points and their image pixels.

II. Estimation of the Intrinsic Parameter Matrix

This section describes the process of estimating the camera's intrinsic parameter matrix, which encapsulates its internal characteristics. The matrix contains:

- (u_0, v_0): Coordinates of the principal point (optical center) in the image plane.
- α, β: Scale factors in the horizontal and vertical image axes, respectively.
- γ: Parameter describing the skewness between the image axes.

Homography and Chessboard Pattern:

- **Corner Detection**: We begin by detecting the corner points of the checkerboard pattern in the captured image using OpenCV's `cv2.findChessboardCorners` function.
- **Real-World Coordinates**: Knowing the physical size of each square on the checkerboard, we calculate the corresponding real-world coordinates of the detected corners.
- **Homography Estimation**: Using the detected corner points in both image and real-world space, we compute the homography matrix using `cv2.findHomography`. This matrix combines both intrinsic and extrinsic camera parameters.

Deriving Intrinsic Parameters: By utilizing the orthogonality property and specific relationships within the homography matrix, we can derive equations to solve for the intrinsic parameters: u_0, v_0, α, β, and γ. This requires careful manipulation and algebraic steps based on the known properties of homography and camera models.

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

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{r_3} & \boldsymbol{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \tag{2}$$

$$= \boldsymbol{A} \begin{bmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{3}$$

$$H = \boldsymbol{A} \begin{bmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t} \end{bmatrix} \tag{4}$$

$$\boldsymbol{h_1^T A^{-T} A^{-1} h_2} = 0 \tag{5}$$

$$\boldsymbol{h_1^T A^{-T} A^{-1} h_1} = \boldsymbol{h_2^T A^{-T} A^{-1} h_2} \tag{6}$$

$$\boldsymbol{B} = \boldsymbol{A^{-T} A^{-1}} \equiv \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \tag{7}$$

B is a symmetric and positive definite matrix, thereby possessing only 6 degrees of freedom (DoF). We define a linear homogeneous system in the following manner:

$$\boldsymbol{b} = \begin{bmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{bmatrix}^T \tag{8}$$

Hence, from homography, homogeneous equations can be written as:

$$\boldsymbol{v_{ij}} = \begin{bmatrix} h_{i1}, & h_{i1}h_{j2} + h_{i2}h_{j1}, & h_{i2}h_{j2}, \\ & h_{i3}h_{j1} + h_{i1}h_{j3}, & h_{i3}h_{j2} + h_{i2}h_{j3}, & h_{i3}h_{j3} \end{bmatrix}^T \tag{9}$$

$$\begin{bmatrix} \boldsymbol{v_{12}}^T \\ \boldsymbol{v_{11}} - \boldsymbol{v_{22}}^T \end{bmatrix} \boldsymbol{b} = 0 \tag{10}$$

From above, we can calculate b vector and using this vector intrinsic parameters can be calculated as follows:

$$\begin{gathered} \boldsymbol{V_b} = 0 \\ \boldsymbol{\lambda} = (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2) \\ \boldsymbol{\alpha} = \sqrt{\alpha/B_{11}} \\ \boldsymbol{\beta} = \sqrt{\alpha B_{11}/(B_{11}B_{22} - B_{12}^2)} \\ \boldsymbol{\beta} = (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2) \\ \boldsymbol{\gamma} = -B_{12}\alpha^2\beta/\lambda \\ \boldsymbol{u_0} = \gamma v_0/\beta - B_{13}\alpha^2/\lambda \end{gathered}$$

III. Estimating the Camera Extrinsic Matrix

This section describes the process of estimating the camera's extrinsic parameters, which define its orientation and position in the world coordinate system. The extrinsic parameters are typically captured in a transformation matrix denoted by $\mathbf{R}_t$.

$$\begin{gathered} \boldsymbol{r_1} = \alpha \boldsymbol{A}^{-1} h_1 \\ \boldsymbol{r_2} = \alpha \boldsymbol{A}^{-1} h_2 \\ \boldsymbol{r_3} = r_1 \times r_3 \\ \boldsymbol{t} = \alpha \boldsymbol{A}^{-1} h_3 \end{gathered}$$

However, due to potential noise in the data used for estimation, the resulting $\mathbf{R}_t$ matrix may not perfectly adhere to the mathematical properties expected of a pure rotation matrix.

IV. Modeling Distortion with Minimal Assumptions

To begin, we assume minimal camera distortion and initialize the distortion coefficient vector $\mathbf{k}$ to $[0, 0]^\top$. This simplifies the distortion model by considering only the first two terms, resulting in two parameters. While using more parameters could potentially improve accuracy, this initial approach prioritizes simplicity.

Later, when incorporating distortion into the calculations, we employ the provided equations to estimate pixel coordinates. It's important to note that these equations assume no skewness in the image frame or sensor ($\gamma = 0$).

$$\begin{gathered} \hat{u} = u + (u - u_0)[k_1(x^2 + y^2) = k_2(x^2 + y^2)^2] \\ \hat{v} = v + (v - v_0)[k_1(x^2 + y^2) = k_2(x^2 + y^2)^2] \end{gathered}$$

Here, (u, v) are the ideal (non-observable distortion-free) pixel image coordinates, $(\tilde{u}, \tilde{v})$ are the corresponding real observed image coordinates, and k_1 and k_2 are the coefficients of the radial distortion.

V. Refining Parameter Estimates through Optimization

Having obtained initial estimates for intrinsic, extrinsic, and distortion parameters, we now seek to refine them further.

While the algebraic distance minimization used previously provided initial values, it lacks clear physical interpretation. Therefore, we employ Maximum Likelihood Estimation (MLE), a statistically rigorous method, to obtain more accurate solutions.

MLE inherently accounts for distortion. This allows us to simultaneously update all parameter sets (intrinsic, extrinsic, and distortion) in a single optimization process. The MLE objective function can be represented as:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} ||\boldsymbol{m_{ij}} - \boldsymbol{\hat{m}}(\boldsymbol{A}, k_1, k_2, \boldsymbol{R_i}, \boldsymbol{t_i}, \boldsymbol{M_j})||^2 \tag{11}$$

By minimizing this function, we obtain parameter estimates that maximize the likelihood of observing the real-world features given the captured image data. This statistically sound approach is expected to yield more accurate and meaningful camera model parameters.

To achieve more accurate parameter estimates, we employ a least squares minimization approach. This minimizes the squared difference between the actual pixel coordinates of detected corners and their projected coordinates based on the current parameter estimates. To perform this optimization efficiently, we leverage the `scipy.optimize.least_squares` function.

VI. Quantitative Results

The calibration matrix before and after optimization is as below:

$$\mathbf{A} = \begin{bmatrix} 2065.25 & -2.939 & 764.7 \\ 0 & 2053.48 & 1362.7 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{A}_{\text{optimum}} = \begin{bmatrix} 2065.25 & -2.940 & 763.1 \\ 0 & 2053.4 & 1351.8 \\ 0 & 0 & 1 \end{bmatrix}$$

The distortion coefficients obtained after optimization are:

$$\mathbf{K}_{\text{optimum}} = \begin{bmatrix} 0.016 \\ -0.116 \end{bmatrix}$$

Re-projection error = 0.785

VII. Qualitative Results

The detected corners and re-projected points for given images are illustrated in Fig.01 to Fig.07.

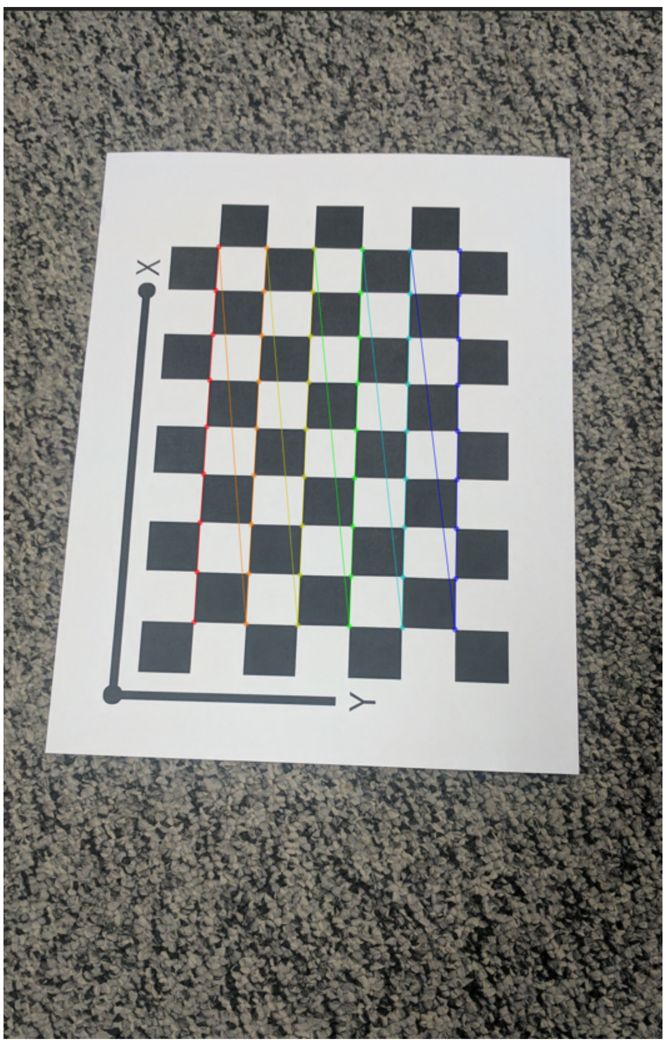

Fig. 1. Re-projected corners on undistorted image 1

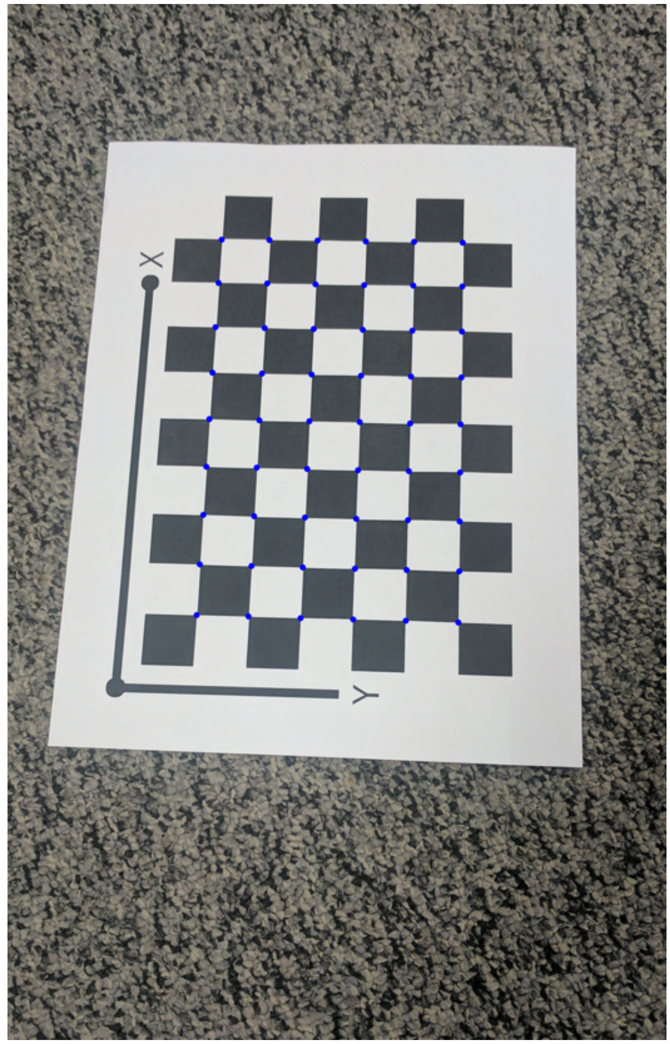

Fig. 2. Re-projected corners on undistorted image 1

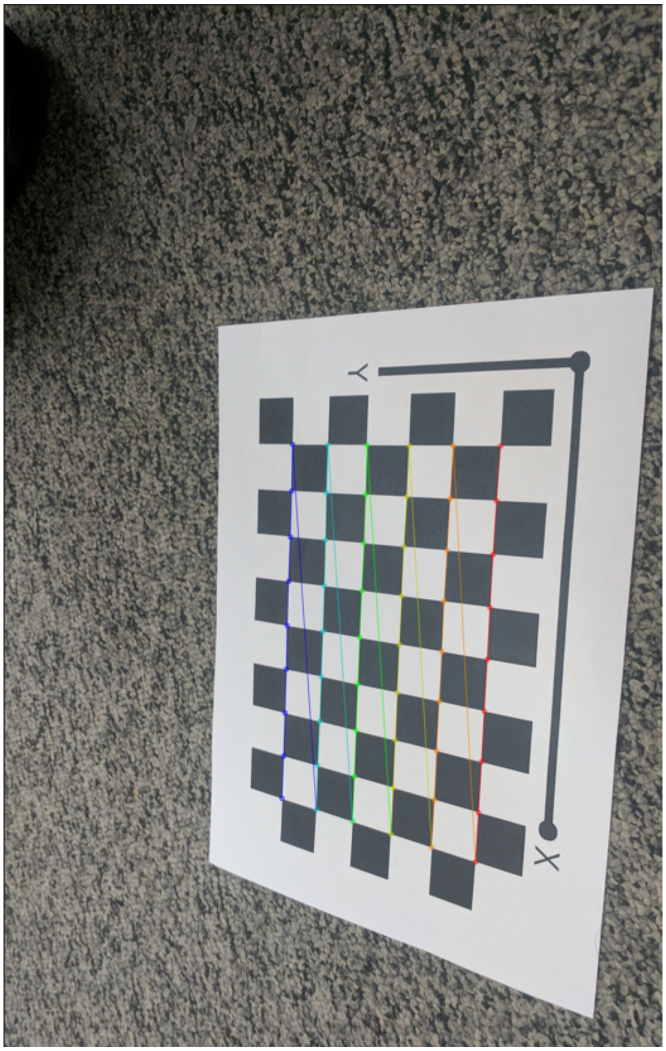

Fig. 3. 54 Corners found in the image 11

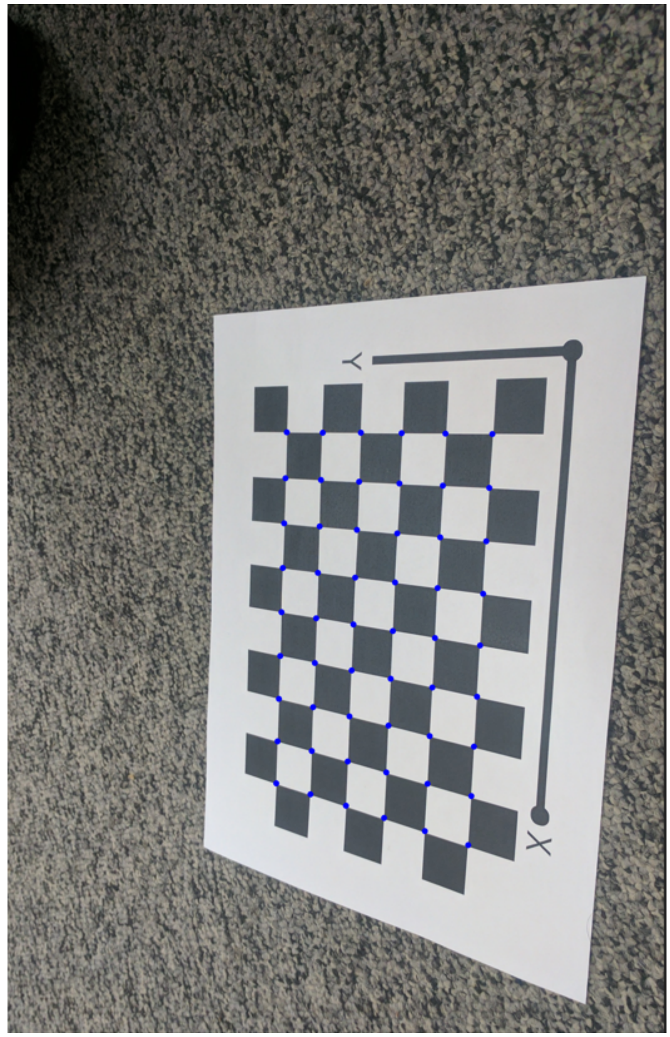

Fig. 4. Re-projected corners on undistorted image 11

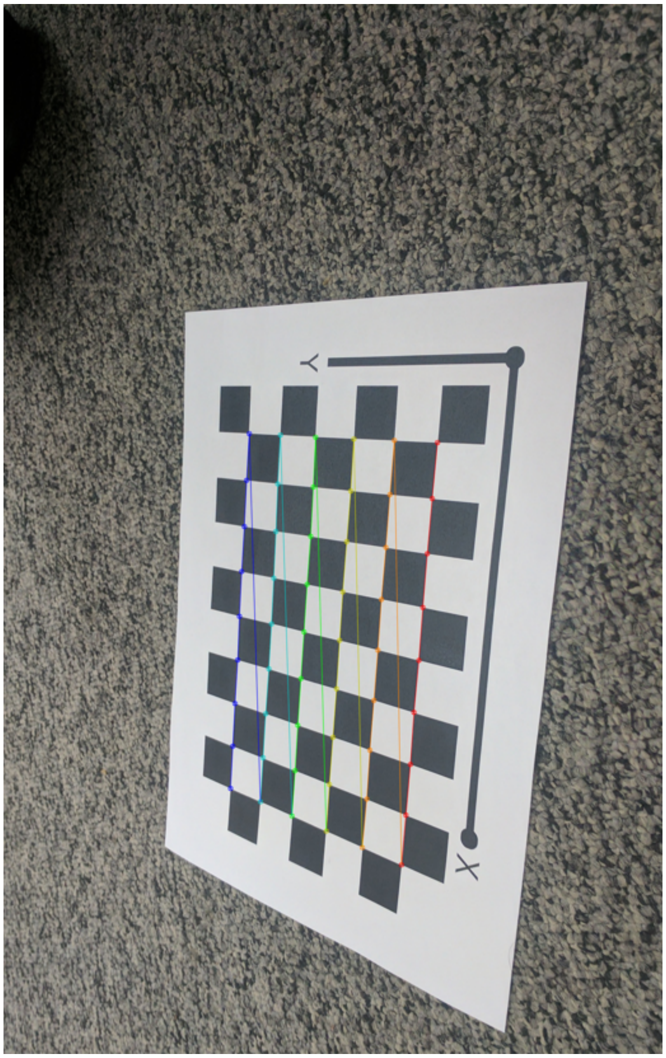

Fig. 5. 54 Corners found in the image 12

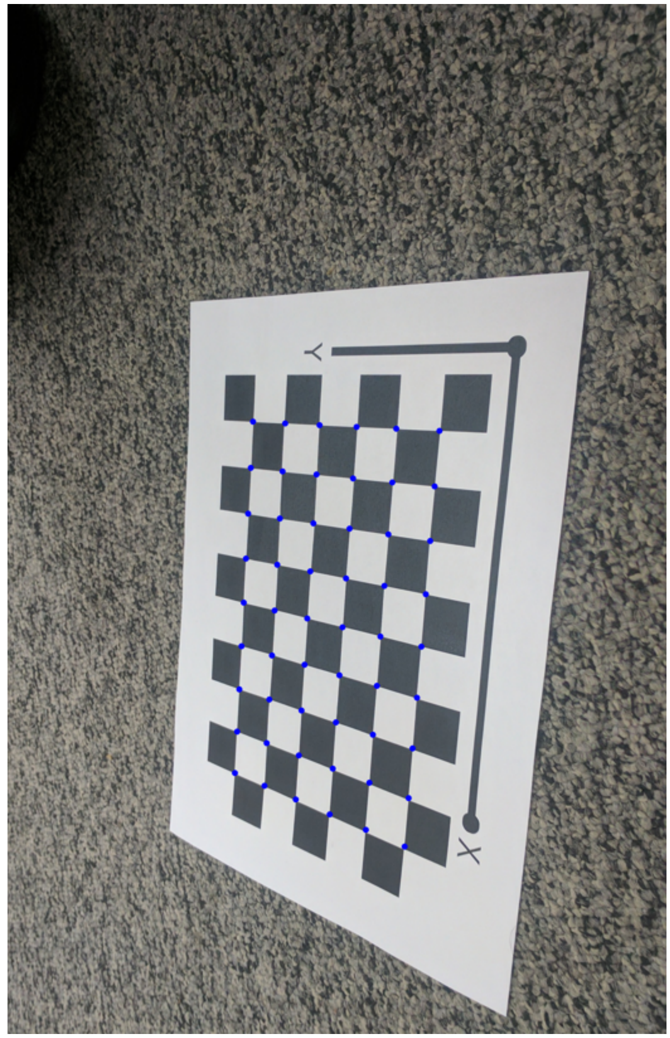

Fig. 6. Re-projected corners on undistorted image 12

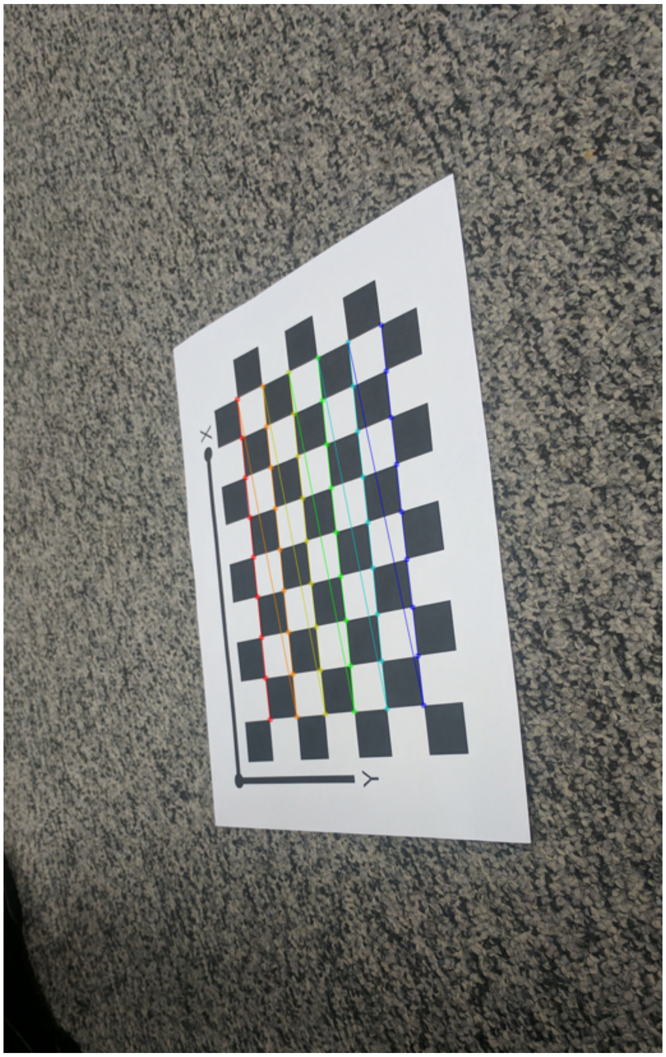

Fig. 7. 54 Corners found in the image 13

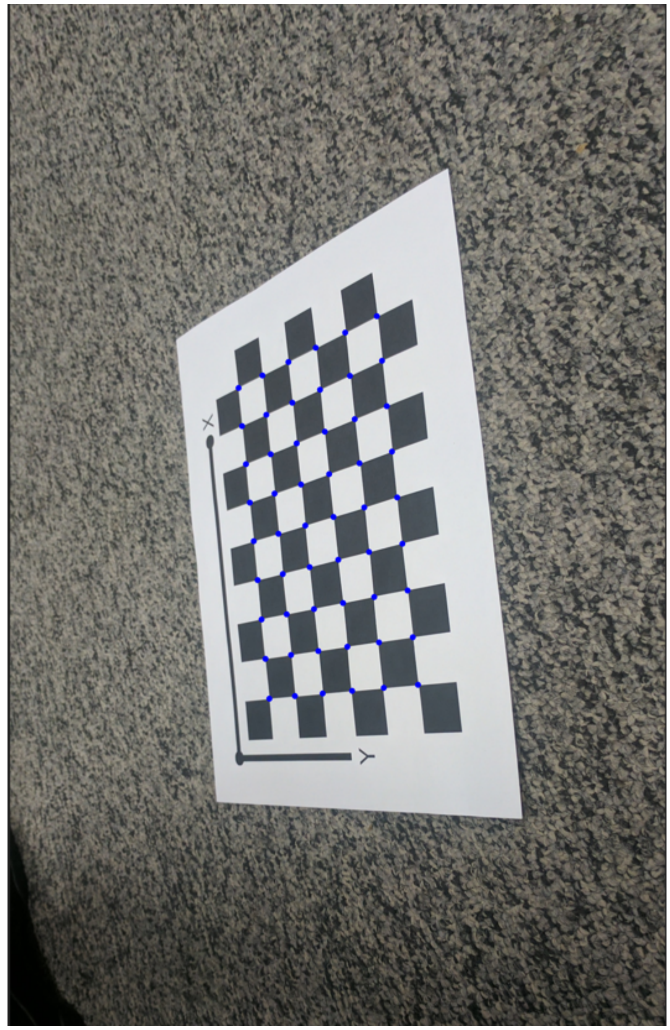

Fig. 8. Re-projected corners on undistorted image 13

REFERENCES

[1] Z. Zhang, *A flexible new technique for camera calibration*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, 2000, doi: 10.1109/34.888718.