HW1 : AutoCalib

Bansal, Mayank
Masters in Robotics
Worcester Polytechnic Institute
Worcester, MA 01609
mbansal1@wpi.edu

***Abstract*—The primary objective of this investigation is to ascertain the camera's intrinsic and extrinsic parameters, as well as its distortion coefficients, leveraging the framework delineated by Zhengyou Zhang in his influential paper, "A Flexible New Technique for Camera Calibration". The calibration protocol employed involves a checkerboard with 21.5 mm square dimensions serving as the calibration reference. The process commences with an initial approximation of the intrinsic matrix (A), succeeded by the derivation of the rotation (R) and translation (t) vectors. Further refinement of the intrinsic parameters alongside the distortion coefficients is achieved through a non-linear optimization process, with a focus on reducing geometric errors. Validation of the proposed methodology is conducted via analysis of real-world images featuring the calibration pattern, with outcomes confirming the method's accuracy in the precise determination of the camera parameters.**

I. Introduction

In the realm of computer vision research that delves into 3D geometry, the precise estimation of camera intrinsic parameters proves essential. These parameters span focal lengths, principal points, skew, and distortion coefficients. In a significant development in 1998, Zhengyou Zhang unveiled an automated and methodical approach for camera calibration through a calibration matrix, denoted by K. This exposition provides an overview of the practical implementation of Zhang's calibration technique. The intrinsic matrix, labeled A, is composed of scale factors and , skewness factor , and the principal point coordinates (u0, v0). Additionally, our model incorporates considerations for radial distortion, an inherent issue with non-ideal camera lenses, by determining distortion parameters k1 and k2 to amend the distortion observed in images.

II. Data Generation

Camera intrinsic parameter estimation is vital in computer vision research focused on 3D geometry. The checkerboard-based automatic calibration method introduced by Zhang is commonly adopted. In our investigation, we utilized a checkerboard with nine rows and six columns of inner squares, each square having a dimension of 21.5mm, printed on A4 paper, as the calibration tool for a Google Pixel XL smartphone camera. A series of thirteen images with a fixed focus were captured of the checkerboard to derive the camera's intrinsic parameters following Zhang's method. This method entails generating a homography matrix from a minimum of four image points and subsequent refinement of the intrinsic matrix and distortion coefficients through a process of non-linear optimization.

III. Methodology

A. Calibration Process

Initially, the algorithm identifies the corners in the checkerboard pattern image by employing the `cv2.findChessboardCorners` function. Subsequent to the corner detection, it calculates the Homography matrix (H) correlating the input image with a standard 9x6 grid. This process allows for the extraction of the array v_{ij} from the elements of the Homography matrix.

$$V_{ij} = \begin{bmatrix} h_{1i}h_{1j} & h_{1i}h_{2j} + h_{2i}h_{1j} & h_{2i}h_{2j} \\ h_{3i}h_{1j} + h_{1i}h_{3j} & h_{3i}h_{2j} + h_{2i}h_{3j} & h_{3i}h_{3j} \end{bmatrix}^T \tag{1}$$

Using v11, v12, v22, SVD is performed for calculating "b", based on this equation:

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} \mathbf{b} = 0 \tag{2}$$

The elements of "b" are as follows:

$$\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T \tag{3}$$

Finally, the intrinsic parameters are calculated as follows:

$$\begin{aligned} v_0 &= \frac{(B_{12}B_{13} - B_{11}B_{23})}{(B_{11}B_{22} - B_{12}^2)} \\ \lambda &= B_{33} - \left[B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})\right]/B_{11} \\ \alpha &= \sqrt{\lambda/B_{11}} \\ \beta &= \sqrt{\lambda B_{11}/(B_{11}B_{22} - B_{12}^2)} \\ \gamma &= -B_{12}\alpha^2\beta/\lambda \\ u_0 &= \gamma v_0/\beta - B_{13}\alpha^2/\lambda. \end{aligned} \tag{4}$$

Finally, we get intrinsic parameter matrix:

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

Using "A" matrix and the columns of "H", "R" and "t" matrices are computed as follows:

$$\begin{aligned} r_1 &= \lambda A^{-1}h_1, \\ r_2 &= \lambda A^{-1}h_2, \quad with \lambda = \frac{1}{\|A^{-1}h_1\|} = \frac{1}{\|A^{-1}h_2\|}, R = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} \\ r_3 &= r_1 \times r_2, \\ t &= \lambda A^{-1}h_3, \end{aligned}$$

k1 and k2, both are initialized to be equal to zero.

B. Optimization

Using A, R, and t matrices, the world point image (mesh grid) is projected to the image (pixel coordinates). The L2 norm between the image corner location and the projected point are calculated as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = -A \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \tag{7}$$

The calculation was modified to fix the radial distortion by using the formula given below:

$$\begin{aligned} \hat{u} &= u + (u - u_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \\ \hat{v} &= v + (v - v_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \end{aligned} \tag{8}$$

Here, u0 and v0 are calculated in the previous section through the matrix ”A”, (x, y) were obtained by multiplying transformation matrix (Rt) to the 3D world point coordinates. The loss is calculated over all the corners of all the images, which gives us this loss summation:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \| m_{ij} - \hat{m}(A, R_i, t_i, M_j) \|^2 , \tag{9}$$

This is a type of MLE (maximum likelihood estimation) where noise is considered to be equally distributed.

IV. Results

The results are shown in the fig. 1. The error before optimization is 0.00431011813919655 and after optimization is 0.004298870201304391. Final k1 and k2 values are 0.0012510872 and -0.004825358 respectively. The final K matrix is given below:

$$\begin{bmatrix} 2.05610659e+03 & -1.01709788e+00 & 7.61655247e+02 \\ 0.00000000e+00 & 2.04050404e+03 & 1.35130849e+03 \\ 0.00000000e+00 & 0.00000000e+00 & 1.00000000e+00 \end{bmatrix} \tag{10}$$

Fig. 2-14 show the distorted and undistorted checkboard images side by side.

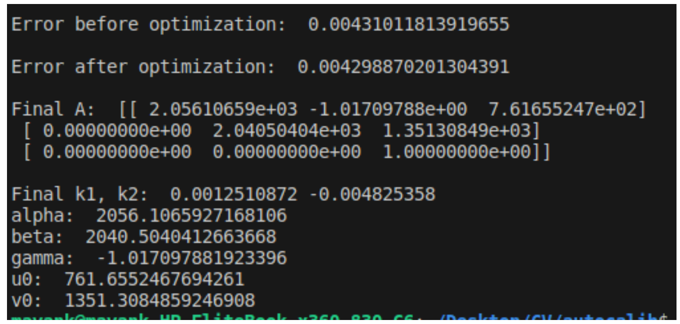

Fig. 1: Final calculated parameters

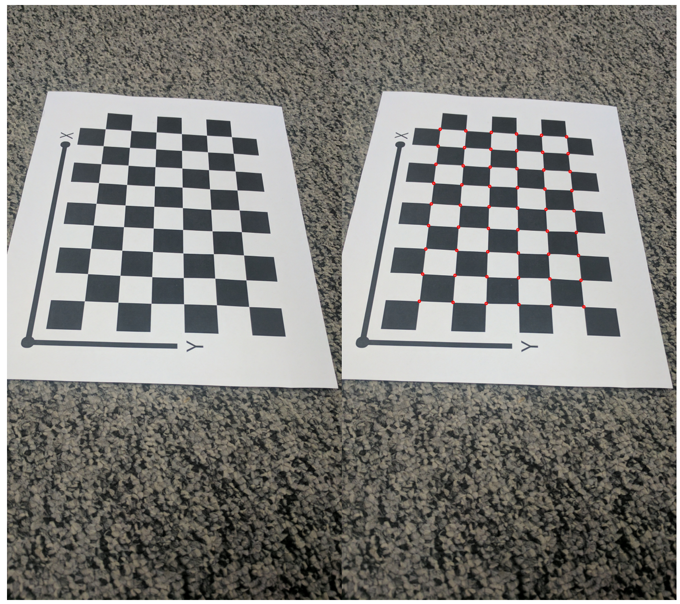

Fig. 2: Distorted and undistorted test image 1

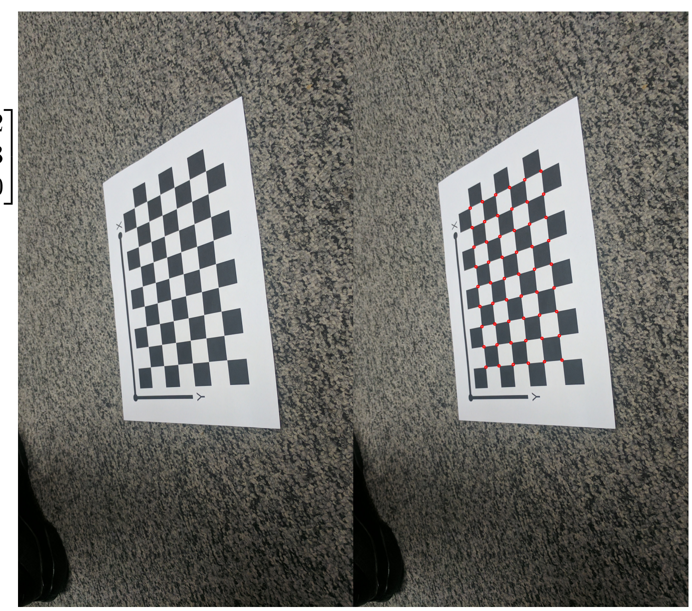

Fig. 3: Distorted and undistorted test image 2

References

[1] A Flexible New Technique for Camera Calibration: link

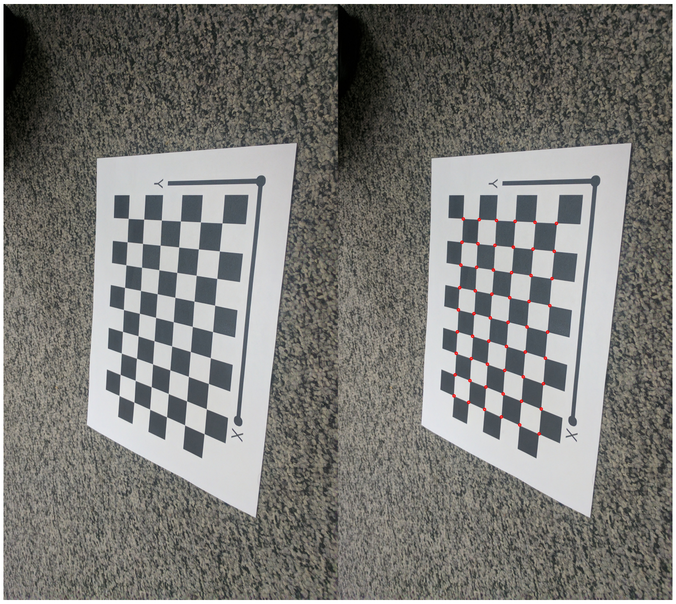

Fig. 4: Distorted and undistorted test image 3

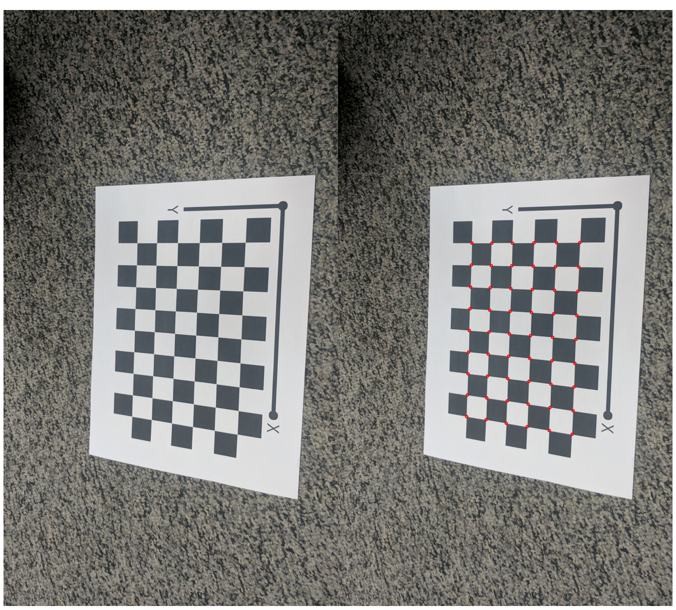

Fig. 5: Distorted and undistorted test image 4

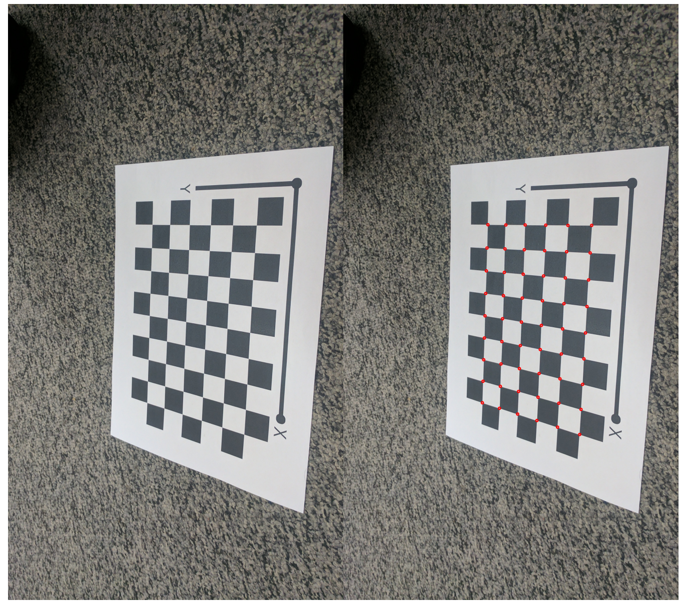

Fig. 6: Distorted and undistorted test image 5

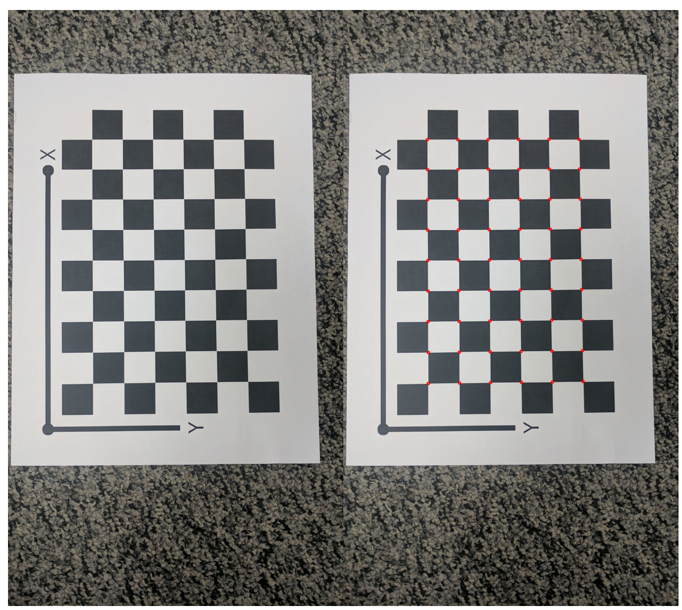

Fig. 7: Distorted and undistorted test image 6

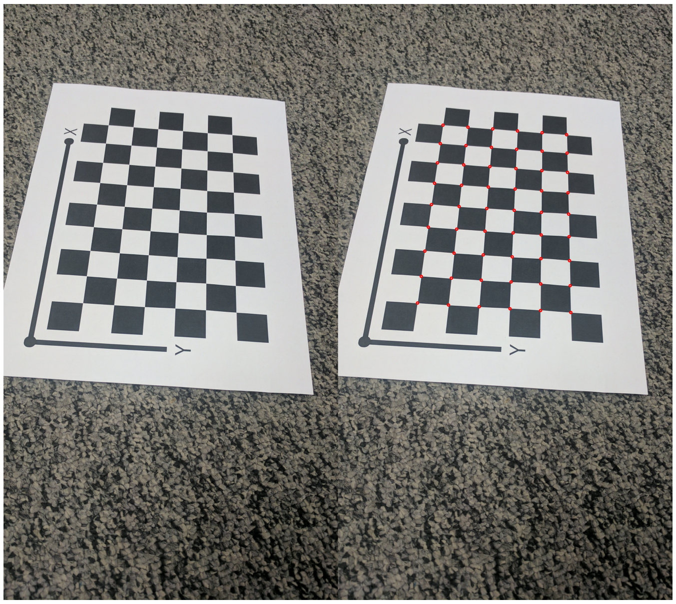

Fig. 8: Distorted and undistorted test image 7

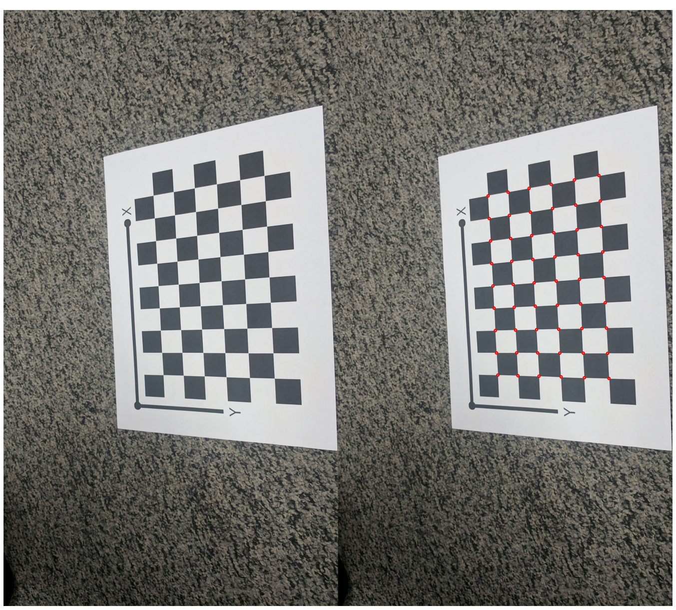

Fig. 9: Distorted and undistorted test image 8

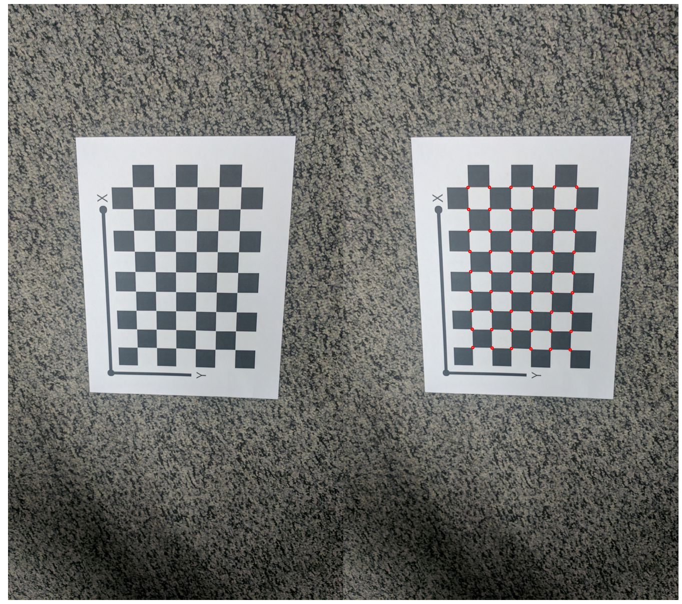

Fig. 10: Distorted and undistorted test image 9

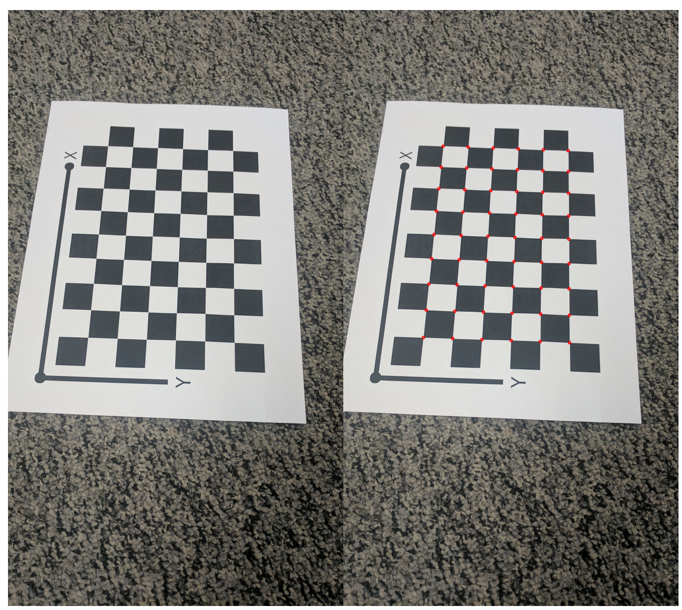

Fig. 11: Distorted and undistorted test image 10

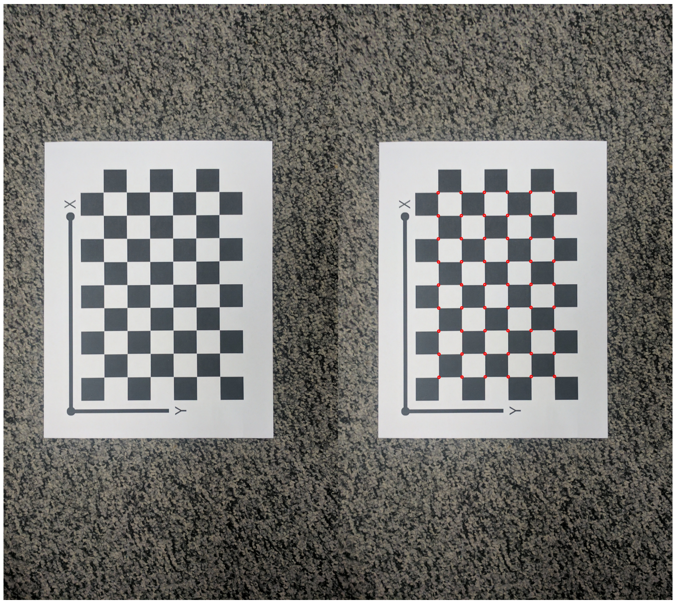

Fig. 12: Distorted and undistorted test image 11

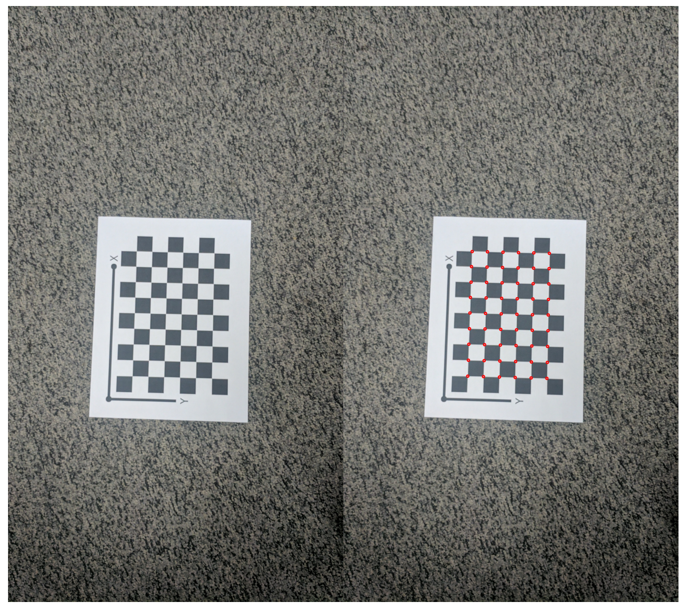

Fig. 14: Distorted and undistorted test image 13

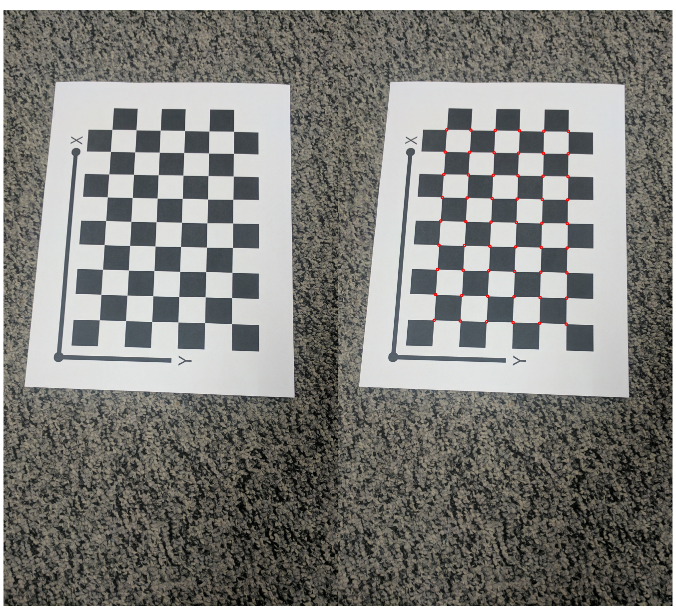

Fig. 13: Distorted and undistorted test image 12