# ELL793 Computer Vision
# Camera Calibration

Abhinava Sikdar 2017MT10724

Yashank Singh 2017MT10756

October 30, 2020

## 1 Introduction

Camera calibration refers to the estimation of all eleven camera parameters, out of which five are intrinsic and six are extrinsic. The intrinsic parameters from the name itself are invariant to the position of the camera coordinate axis with respect to the world coordinate axis. However, this is not true for the extrinsic parameters which vary along the movement of the camera. The perspective projection can be represented as

$$\mathbf{p} = \frac{1}{Z}\mathcal{M}\mathbf{P}$$

where $\mathbf{p} = (x, y, 1)^T$ and $\mathbf{P} = (X, Y, Z, 1)^T$ are the projection and world coordinates of a point P in the homogeneous system. $\mathcal{M}$ is the projection matrix of size $3 \times 4$. This matrix $\mathcal{M}$ can further be written as

$$\mathcal{M} = \mathcal{K}(\mathcal{R}\ t)$$

where $\mathcal{K}$ is the internal calibration matrix of dimensions $3 \times 3$, $\mathcal{R}$ is the rotation matrix of dimension $3 \times 3$ and $t$ is the translation vector. If we denote the rows of $\mathcal{M}$ as a column vector by using $\mathbf{m}_1, \mathbf{m}_2$ and $\mathbf{m}_3$, then

$$x = \frac{\mathbf{m}_1.\mathbf{P}}{\mathbf{m}_3.\mathbf{P}} \text{ and } y = \frac{\mathbf{m}_2.\mathbf{P}}{\mathbf{m}_3.\mathbf{P}} \tag{1}$$

Assume that we are given a set of $n$ fiducial points i.e. points for which both the image and world coordinates are known. Then for each of these point $P_i,\ i = 1. \dots, n$ by using Eqn(1), we get that

$$\mathcal{P}\mathbf{m} = \mathbf{0} \tag{2}$$

where

$$\mathcal{P} = \begin{pmatrix} \mathbf{P}_1^T & \mathbf{0}^T & -x_1\mathbf{P}_1^T \\ \mathbf{0}^T & \mathbf{P}_1^T & -y_1\mathbf{P}_1^T \\ \dots & \dots & \dots \\ \mathbf{P}_n^T & \mathbf{0}^T & -x_n\mathbf{P}_n^T \\ \mathbf{0}^T & \mathbf{P}_n^T & -y_n\mathbf{P}_n^T \end{pmatrix} \text{ and } \mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{pmatrix} \tag{3}$$

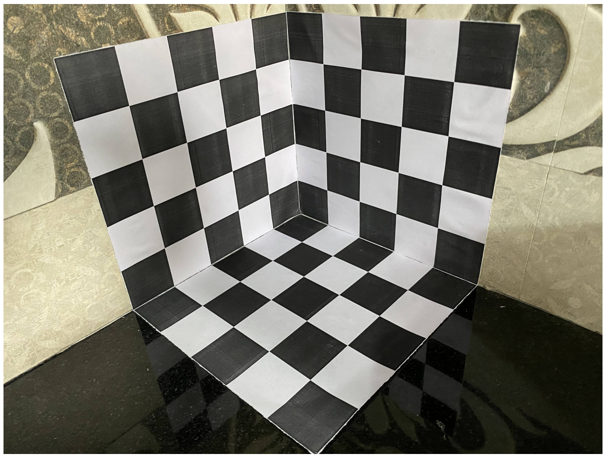

Figure 1: Image of three checkerboard prints pasted on three orthogonal planes

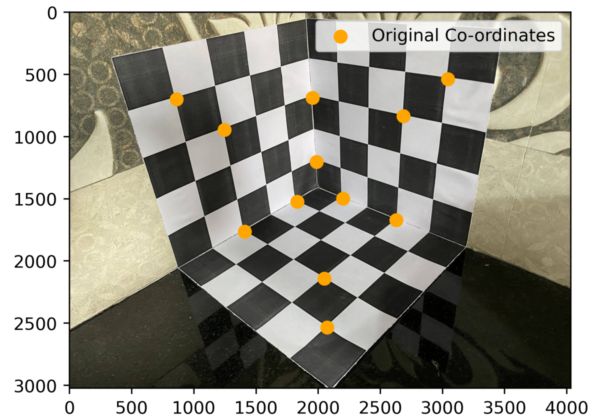


Figure 2: Visualisation of the 12 fiducial points in the image taken

This is a homogeneous system of equations in 12 variables. Hence at the least, 6 fiducial points are required. The solution to this i.e. $\mathbf{m}$ is given by the eigenvector corresponding to the smallest eigenvector of the matrix $\mathcal{P}^T\mathcal{P}$ upto a constant of multiplication $\rho$ [2].

# 2 Procedure Followed

As instructed in the problem, three checkerboards where printed and pasted on three orthogonal walls. The picture in Fig.1 was taken using a mobile camera. The world and image coordinates of the 12 points were noted. The image coordinates in pixels were obtained by using the MATLAB image viewer tool. They are presented in Table 1. These image coordinates were plotted on the picture taken for visualisation and can be seen in Fig.2.

Table 1: World and image coordinated of 12 fiducial points

| S.No. | X | Y | Z | x | y | S.No. | X | Y | Z | x | y |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 1831 | 1524 | 7 | 3 | 0 | 0 | 1408 | 1765 |
| 2 | 0 | 1 | 0 | 2200 | 1499 | 8 | 0 | 3 | 0 | 2627 | 1671 |
| 3 | 0 | 0 | 1 | 1988 | 1204 | 9 | 0 | 0 | 3 | 1955 | 688 |
| 4 | 3 | 3 | 0 | 2050 | 2142 | 10 | 3 | 0 | 3 | 1247 | 946 |
| 5 | 0 | 3 | 3 | 2686 | 836 | 11 | 4 | 4 | 0 | 2072 | 2536 |
| 6 | 4 | 0 | 4 | 861 | 699 | 12 | 0 | 4 | 4 | 3042 | 537 |

Before proceeding, we normalised the two sets of world and image coordinates. Let $(\bar{X}, \bar{Y}, \bar{Z}, 1)$ and $(\bar{x}, \bar{y}, 1)$ denote the centroids of the two systems. Similarly let $d_{world}$ and $d_{camera}$ denote the average distance of the points from the centroids in their respective system. Hence, these normalisation matrices are

$$\mathcal{U} = \frac{\sqrt{3}}{d_{world}} \begin{pmatrix} 1 & 0 & 0 & -\bar{X} \\ 0 & 1 & 0 & -\bar{Y} \\ 0 & 0 & 1 & -\bar{Z} \\ 0 & 0 & 0 & d_{world}/\sqrt{3} \end{pmatrix} \text{ and } \mathcal{T} = \frac{\sqrt{2}}{d_{camera}} \begin{pmatrix} 1 & 0 & -\bar{x} \\ 0 & 1 & -\bar{y} \\ 0 & 0 & d_{camera}/\sqrt{2} \end{pmatrix}$$

These also ensure that the average distance of points in the world and image coordinate systems are $\sqrt{3}$ and $\sqrt{2}$ respectively.

Hence we get, $\mathbf{P}_{world}^{normalised} = \mathcal{U}\mathbf{P}_{world}$ and $\mathbf{p}_{camera}^{normalised} = \mathcal{T}\mathbf{p}_{camera}$. We use these to get the normalised version of matrix $\mathcal{P}$ i.e. $\mathcal{P}^{normalised}$, solve Eqn.(2) to get the vector $\rho\mathbf{m}_{normalised}$ which essentially gives us $\rho\mathcal{M}_{normalised}$ through rearranging. We then determine the constant $\rho$ by using

$$\rho||\mathbf{a}_3|| = +1 \tag{4}$$

where $\mathbf{a}_3$ is the third row of matrix $\mathcal{A}$ when $\rho\mathcal{M}_{normalised}$ is written as $\rho(\mathcal{A}\ b)$. Note that the sign in Eqn.(4) is taken to be positive since the world coordinate system lies in front of the camera [2]. Finally, to get the un-normalised projection matrix $\mathcal{M}$,

$$\mathcal{M} = \mathcal{T}^{-1}\mathcal{M}_{normalised}\,\mathcal{U}$$

Now we write the un-normalised projection matrix $\mathcal{M}$ as $(\mathcal{A}\ b)$ where $\mathcal{A} = \mathcal{K}\mathcal{R}$ and $b = \mathcal{K}t$. Now note that $\mathcal{K}$ is an upper-triangular matrix whereas $\mathcal{R}$ is an orthogonal matrix. Hence we can use RQ decomposition such that

$$\mathcal{A} = (\eta\mathcal{K})\mathcal{R}$$

where $\eta$ is any scalar. However note that the form of the actual intrinsic matrix is as

$$\mathcal{K} = \begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & 1 \end{pmatrix}$$

Hence, we get that $\eta = K_{3,3}$ where $K_{3,3}$ is the third entry of third row of $\eta\mathcal{K}$. Note however, that this does not modify the camera in any way since the scale is arbitrary. Having determined the value of scalar $\eta$, we now uniquely have $\mathcal{K}$ and $\mathcal{R}$. Using the value of the thus obtained $\mathcal{K}$, we can retrieve $t$ as

$$t = \mathcal{K}^{-1}b$$

Furthermore the camera centre $(x_0, y_0)$ can be given as (where $\mathbf{a_i}$ is the $i^{th}$ row of matrix $\mathcal{A}$)

$$x_0 = \mathbf{a_1} \cdot \mathbf{a_3}$$

$$y_0 = \mathbf{a_2} \cdot \mathbf{a_3}$$

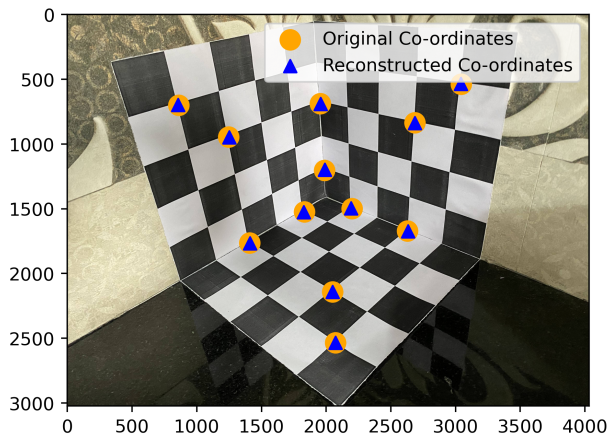


Figure 3: Projection of the original coordinates using the estimated projection matrix

# 3 Results and Conclusion

The aforementioned procedure of RQ decomposition results into the following matrices. The intrinsic parameter matrix

$$\mathcal{K} = \begin{pmatrix} -3052 & 41 & 2034 \\ 0 & -3038 & 1527 \\ 0 & 0 & 1 \end{pmatrix}$$

The rotation matrix

$$\mathcal{R} = \begin{pmatrix} 0.673 & -0.737 & 0.062 \\ -0.421 & -0.313 & 0.851 \\ -0.608 & -0.6 & -0.521 \end{pmatrix}$$

The translation vector

$$t = \begin{pmatrix} 0.042 & 0.144 & 4.145 \end{pmatrix}^T$$

The camera centre

$$(x_0, y_0) = (227.7, 171)^T \tag{5}$$

Having calculated all the intrinsic and extrinsic parameters, we used the thus obtained projection matrix $\mathcal{M}$ to project the set of points $\mathbf{P}$ onto the original image which can be seen in Fig.3.

We further calculate the MSE which comes out to be 9.607 approximately in pixel units which implies that the estimated projection matrix lies very close to the true estimation matrix in the matrix space.

### Why normalise matrix $\mathbf{P}_{\mathbf{world}}$ and $\mathbf{p}_{\mathbf{camera}}$?

This can be explained using Table 1. Clearly all the world coordinates have much smaller magnitudes like $0, 1, 2, \ldots$ whereas all the camera coordinates, measured in pixels are in thousands. This can cause the matrix $\mathcal{P}$ to become ill-conditioned or even singular. This can lead to a build up of a heap of numerical errors [1].

# References


[1] R. Burden, J. Faires, and A. Burden. *Numerical Analysis.* Cengage Learning, 2015.

[2] D. A. Forsyth and J. Ponce. *Computer Vision - A Modern Approach, Second Edition.* Pitman, 2012.