# ME491(B)
# Homework 3 - Camera Calibration

Herman Kolstad Jakobsen
20196493

April 14, 2020

# Contents

# 1 Introduction

A camera is described by several parameters. Given a number of points with known 3D coordinates and known image projection, it is possible to estimate these camera parameters. This is known as geometric camera calibration or camera resectioning. Knowing the parameters enables the ability of correcting lens distortion, measuring the size of an object in world units or determining the position of the camera in the scene. These applications are much used in computer vision to detect and measure objects, and in 3D scene reconstruction.

In this homework, the extrinsic and intrinsic parameters of the camera will be estimated. This includes translation $\mathbf{t}$ of the optical center from the origin of the world coordinates, rotation $\mathbf{R}$ of the image plane, focal length $f$, principle point offset $(p_x, p_y)$ and pixels per meter $(m_x, m_y)$. The procedure of camera calibration will be described first, before presenting the results.

# 2 Procedure

The first step of camera calibration is to estimate the projection matrix $\mathbf{P}$. This matrix relates the world coordinates to the image projections in the following way

$$\mathbf{x}_i = \mathbf{P}\mathbf{X}_i \tag{1}$$

where $\mathbf{X}_i$ is the world coordinate of an arbitrary point and $\mathbf{x}_i$ is its corresponding image projection, both given in homogeneous coordinates. The elements of the projection matrix contains useful information which can be used to estimate the mentioned camera parameters.

In order to estimate the projection matrix, several points with known 3D coordinates and known image projection are required. These points were obtained from fig. 1. The world coordinate of each point was obtained combining the illustrated world frame and the length of each square. The image projection coordinate was found as pixel coordinates with the origin in the lower left corner. A total of 28 points were chosen arbitrary from the figure.

Each point corresponds to two linearly independent equations which can be solved to obtain the projection matrix $\mathbf{P}$. A minimal solution would therefore require at least 6 points, since the projection matrix has 11 degrees of freedom (5 from intrinsic parameters, 3 from $\mathbf{R}$ and 3 from the camera center position in the world frame $\mathbf{C}$). The chosen amount of 28 points satisfies the thumb rule from camera calibration of having at least 5 times as many equations as unknowns. Excessive equations reduces the impact of noise.

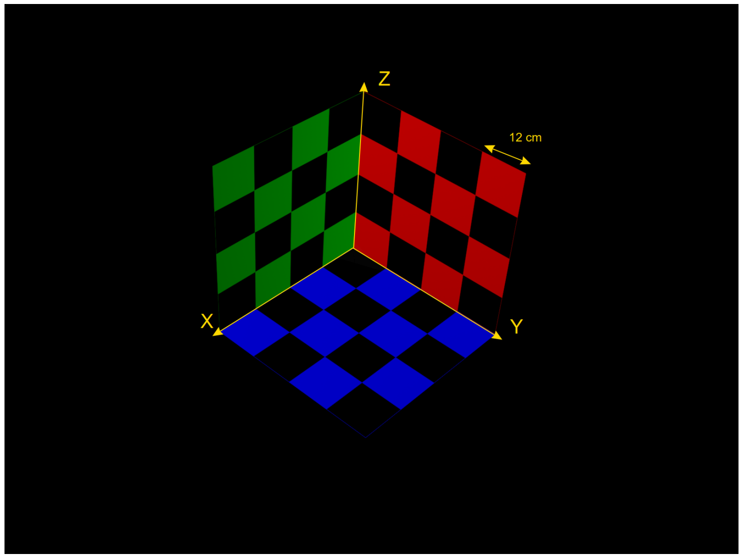


Figure 1: Rig used to calibrate camera. The world frame and the width and height of each square is illustrated.

Having obtained a sufficient amount of points and their corresponding coordinates, the projection matrix can be estimated using the Direct Linear transform (DLT) algorithm [1, Chapter 4.1]. The algorithm results in the following optimization problem

$$\min \|\mathbf{A}\mathbf{p}\| \quad \text{s.t.} \quad \|\mathbf{p}\| = 1 \tag{2}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{0}^T & -w_i\mathbf{X}_i^T & y_i\mathbf{X}_i^T \\ w_i\mathbf{X}_i^T & \mathbf{0}^T & -x_i\mathbf{X}_i^T \end{bmatrix} \tag{3}$$

and

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{p}^3 \end{bmatrix} \tag{4}$$

The $\mathbf{A}$ matrix is of size $2n \times 2$ where $n$ is the number of points. The vector $\mathbf{p}$ is of size $12 \times 1$ where $\mathbf{p}^i$ is a row vector of the projection matrix $\mathbf{P}$. The solution to the optimization problem is the unit singular vector $\mathbf{p}$ corresponding to the smallest singular value of $\mathbf{A}$. That is, the last column of $\mathbf{V}$, where $\mathbf{A} = \mathbf{UDV}^T$ is the singular value decomposition (SVD) [2] of $\mathbf{A}$.

Due to the use of the DLT algorithm, normalization of the data is essential [1, Chapter 4.4.4]. Hence, normalization of the coordinates should be carried out before applying the algorithm. The reason for the need of normalization is related to the condition number of the set of equations in $\mathbf{A}$, and how the image projection coordinates are affected by a magnitude change in either $x_i$, $y_i$ or $w_i$. In addition, normalization of data improves accuracy of the results.

After obtaining an estimation of the projection matrix, the intrinsic and extrinsic parameters can be found by decomposing the matrix. The projection matrix can be decomposed into

$$\mathbf{P} = \mathbf{KR}\left[\mathbf{I}_3 | -\tilde{\mathbf{C}}\right] \tag{5}$$

Here, $\mathbf{K}$ is an upper triangular calibration matrix where its elements corresponds to the intrinsic parameters. The $\mathbf{R}$ matrix is a rotation matrix, while $\tilde{\mathbf{C}}$ is the non-homogeneous coordinates of the camera center in the world frame.

A useful property of the projection matrix is that its null space is the camera center position in the world frame $\mathbf{C}$. Hence,

$$\mathbf{C} = Null(\mathbf{P}) \tag{6}$$

The null space can, for instance, be found by applying SVD and then choose the unit singular vector of $\mathbf{P}$ corresponding to the smallest singular value.

Then, define the left $3 \times 3$ submatrix of $\mathbf{P}$ as

$$\mathbf{M} = \mathbf{KR} \tag{7}$$

The matrix $\mathbf{M}$ is non-singular and can therefore be decomposed into the product of an upper-triangular matrix $\mathbf{K}$ and an orthogonal matrix $\mathbf{R}$ using the RQ factorization. The intrinsic parameters can be extracted from $\mathbf{K}$.

Finally, the translation can be obtained as

$$\mathbf{t} = -\mathbf{R}\tilde{\mathbf{C}} \tag{8}$$

# 3 Results

The calibration matrix $\mathbf{K}$ was estimated to be

$$\mathbf{K} = \begin{bmatrix} \alpha_x & 0 & \beta_x \\ 0 & \alpha_y & \beta_y \\ 0 & 0 & 1 \end{bmatrix} \approx \begin{bmatrix} -2111.9 & -1.3239 & 746.1273 \\ 0 & -2100.5 & 556.5570 \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

where

$$\alpha_i = m_i f \quad \text{and} \quad \beta_i = m_i p_i \quad \text{for} \quad i = x, y \tag{10}$$

Here, $f$ is the focal length, $p_i$ is the principal point offset and $m_i$ is pixels per meter. The rotation matrix $\mathbf{R}$ was estimated to be

$$\mathbf{R} = \begin{bmatrix} -0.6802 & 0.7313 & 0.0504 \\ -0.4549 & -0.4750 & 0.7533 \\ 0.5748 & 0.4895 & 0.6557 \end{bmatrix} \tag{11}$$

With the estimated position of the camera center in the world frame being

$$\tilde{\mathbf{C}} = \begin{bmatrix} 1.5898 \\ 1.3439 \\ 1.6231 \end{bmatrix} \tag{12}$$

the following translation vector $\mathbf{t}$ was obtained

$$\mathbf{t} = -\mathbf{R}\tilde{\mathbf{C}} = \begin{bmatrix} 0.0169 \\ 0.1388 \\ -2.6360 \end{bmatrix} \tag{13}$$

In order to check the sanity of the code used to estimate these quantities, the ground truth of the position of the camera center in the world frame was handed out

$$\tilde{\mathbf{C}}_{truth} = \begin{bmatrix} 1.6620 \\ 1.4146 \\ 1.7008 \end{bmatrix} \tag{14}$$

It is clearly seen that $\tilde{\mathbf{C}}$ differs a significant amount from $\tilde{\mathbf{C}}_{truth}$. Hence, it can be concluded that the estimations of the quantities are sub-optimal. For instance, the calibration matrix in eq. (9) has a non-zero value at an element that should be zero. It should also be noted that the negative values on the diagonal implies a negative focal length.

To find the reason for the poor estimation, it was attempted to use the data points from the handed out code. The handed out data points resulted in a considerable better accuracy. Hence, the importance of the chosen data points can be questioned. It should be mentioned that the data points were manually read - a procedure that often leads to noise and error.

In order to improve the result, the projection matrix was optimized with the use of non-linear minimization. The optimization resulted in a matrix mapping the world frame coordinates onto the image plane in an improved fashion. However, the further calculations using the optimized projection matrix resulted in incomprehensible results. This might be due to lack of constraints on the optimized projection matrix. Ultimately, the optimized projection matrix was scrapped.

# References


[1] R. Hartley and A. Zisserman. *Multiple View Geometry in computer vision.*

[2] *Singular Value Decomposition (SVD) tutorial.* URL: `https://web.mit.edu/be.400/www/SVD/Singular_Value_Decomposition.htm` (visited on 04/14/2020).