ELL 793: Computer Vision

Assignment 1

Camera Calibration

By Gauri Gupta (2017MT60207)

Objective

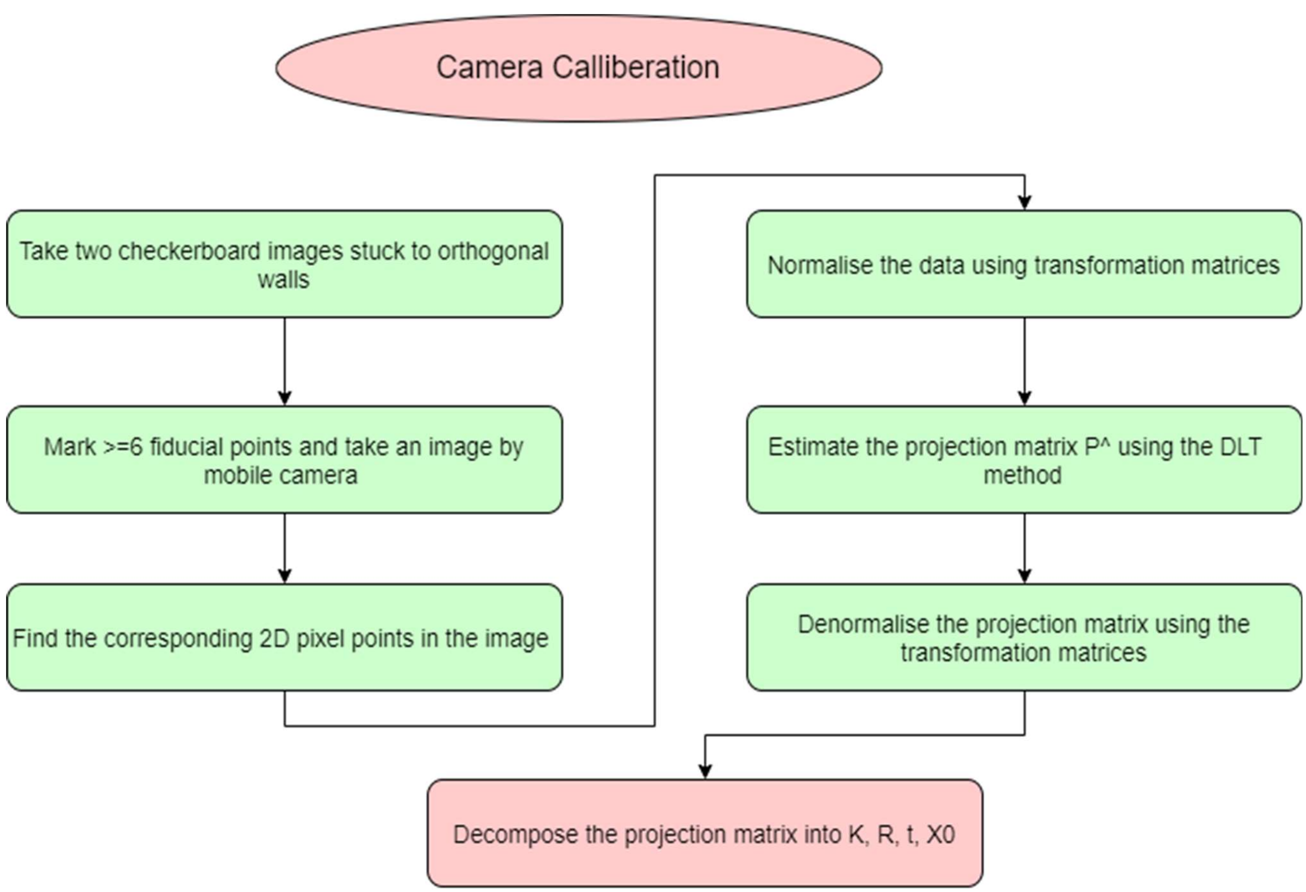
To find intrinsic and extrinsic camera calibration parameters of a mobile phone's camera.

Background

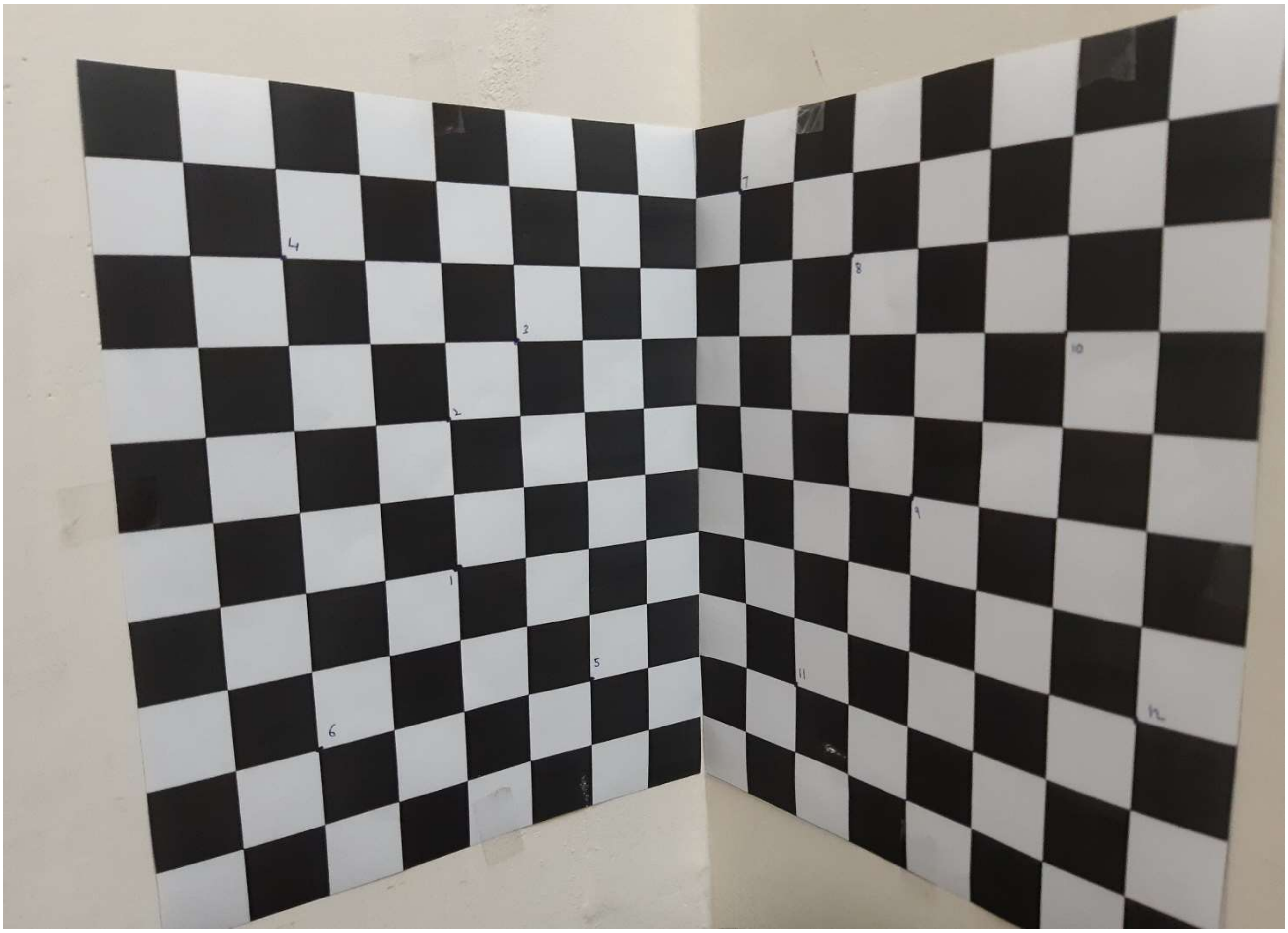
The projection matrix depends on 11 parameters: 5 intrinsic parameters and 6 extrinsic parameters (3 for rotation and 3 for translation). The process of calibration of a camera consists in estimating its intrinsic and/or extrinsic parameters.

1. **Internal parameters** of the camera/lens system. E.g. focal length, optical center, skew, aspect ratio etc.
2. **External parameters:** This refers to the orientation (rotation and translation) of the camera with respect to some world coordinate system.

Procedure: Explained using a flow chart



Data Preparation



1. Two print outs of a checkerboard image were taken and stuck to two orthogonal planes.
2. 12 points were marked on these checkerboards (6 on each of them).
3. A picture of this setting was taken by a mobile phone camera as shown above.
4. The pixel values in the mobile phone image corresponding to these points were noted.
5. A 3Xn matrix (p) and 4Xn matrix (P) of homogenized coordinates was formed.

P =

4	0	4
4	0	6
3	0	7
6	0	8
2	0	2
6	0	2
0	1	9
0	3	8
0	4	5
0	6	7
0	2	2
0	7	3

p =

219	269
213	198
244	160
133	120
280	321
150	354
351	88
402	119
432	233
505	155
377	323
538	340

Where P are the coordinates of the points in the world coordinate system and p are the corresponding points in the 2D image.

6. These matrices were then transformed using transformation matrices p_trans and P_trans so that the centroid of 2D and 3D points becomes the origin and the average Euclidean distance of 2D and 3D points from the origin is √2 and √3, respectively. Where transportation matrix is formed by P_trans = P_scale*P_shift.

```
P_scale =

Diagonal Matrix

   0.42783         0         0         0
         0   0.42783         0         0
         0         0   0.42783         0
         0         0         0   1.00000
```

```
P_shift =

   1.00000   0.00000   0.00000  -2.08333
   0.00000   1.00000   0.00000  -1.91667
   0.00000   0.00000   1.00000  -5.25000
   0.00000   0.00000   0.00000   1.00000
```

So, P_trans = P_scale*P_shift. And similarly, p_trans is calculated.

```
P_trans =

   0.42783   0.00000   0.00000  -0.89131
   0.00000   0.42783   0.00000  -0.82000
   0.00000   0.00000   0.42783  -2.24609
   0.00000   0.00000   0.00000   1.00000

p_trans =

   0.00945   0.00000  -3.02557
   0.00000   0.00945  -2.10940
   0.00000   0.00000   1.00000
```

Calculating Projection Matrix

The normalized matrices p_norm and P_norm were then used to estimate the normalized projection matrix M_norm using **Direct Linear Transform** as follows -

1. A 2nX12 matrix Q was formed as

```
for i=1:n
  Q(2*i-1,:) = [ P_norm(i,:) 0 -p_norm(i,1)*P_norm(i,:) ];
  Q(2*i,:) = [ 0 P_norm(i,:) -p_norm(i,2)*P_norm(i,:) ];
endfor
```

2. Eigenvalues and eigenvectors of Q^TQ were calculated and the eigenvector corresponding to the smallest eigenvalue was noted. This is equivalent to minimising the norm ||Qm||.
3. The 3X4 normalized projection matrix M_norm was then formed by arranging the first 4 coordinates of this eigenvector as the first row, the next 4 as the second row and the last 4 as

the third row.

```
M_norm =

  -0.4294083   0.3345344  -0.0102719  -0.0300208
   0.0539853   0.0533183  -0.5220003   0.0053234
  -0.0653379  -0.0874878  -0.0181749   0.6420009
```

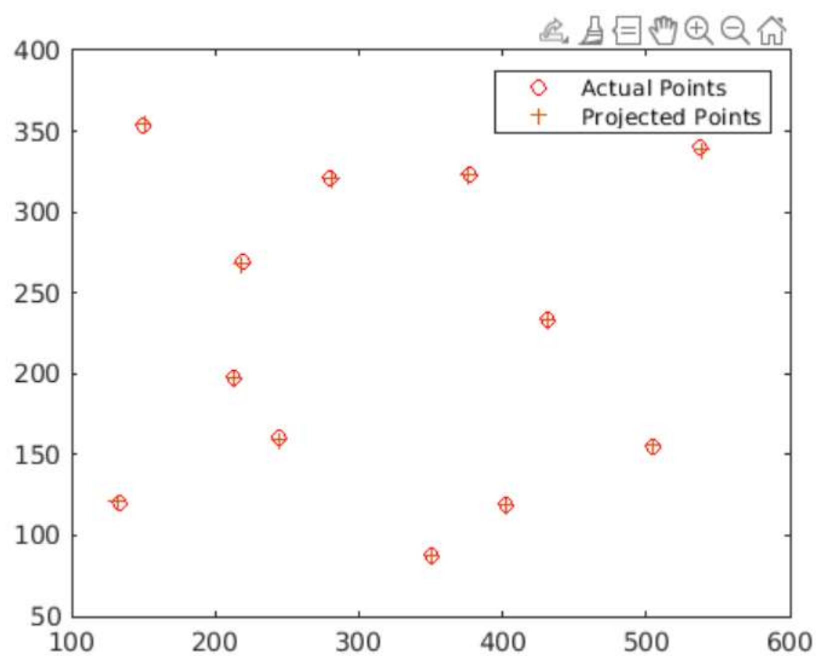
4. The projection matrix M was now calculated as: M = p_trans^{-1}*M_norm*P_trans

```
M =

 -2.8405e+01   3.1632e+00  -2.9561e+00   2.7111e+02
 -3.7976e+00  -5.9442e+00  -2.5381e+01   2.9650e+02
 -2.7953e-02  -3.7430e-02  -7.7757e-03   8.1280e-01
```

Estimated projected points

The estimated 2D projections of the marked 3D points were calculated and plotted along the 2D points marked on the image.



RMSE = 0.67112

To validate M, we compute the residual error between the projected 2d location of each 3d point and the actual location of that point in the 2d image. The residual is the Euclidean distance in the image plane (square root of the sum of squared differences in u and v). RMSE calculated for 12 points is 0.67 pixels for 612 X 454 size image, which is very small validating the calculated projection matrix.

Estimating Intrinsic and Extrinsic Parameters

The intrinsic and extrinsic parameters of the camera were calculated in the following manner using the decomposition of projection matrix. The calibration matrix K of the camera and its extrinsic parameters are R and t, where P and p denote as before the homogeneous coordinate vector of the point P in the world reference frame, and the homogeneous coordinate vector of its projection p in the camera's coordinate system. The objective is to find K, R and T from M, where M is the projection matrix.

$$p = \mathcal{M}P,$$

$$p = \frac{1}{Z_r}\mathcal{K}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & Z_r \end{pmatrix}\begin{pmatrix} \mathcal{R} & t \\ 0^T & 1 \end{pmatrix}P,$$

$$\overbrace{\mathbf{K}}^{\text{Intrinsic Matrix}} \times \overbrace{[\mathbf{R} \mid \mathbf{t}]}^{\text{Extrinsic Matrix}}$$

```
A = M(:,1:3);
R = zeros(3,3);
K = zeros(3,3);
rho = -1/norm(A(3,:),2);
R(3,:) = rho*A(3,:);
X0 = rho*rho*(A(1,:)*A(3,:).');
Y0 = rho*rho*(A(2,:)*A(3,:).');
cross1 = cross(A(1,:),A(3,:));
cross2 = cross(A(2,:),A(3,:));
n_cross1 = norm(cross1,2);
n_cross2 = norm(cross2,2);
theta = acos(-cross1/n_cross1*cross2.'/n_cross2);
alpha = rho*rho*n_cross1*sin(theta);
Beta = rho*rho*n_cross2*sin(theta);
R(1,:) = cross2/n_cross2;
R(2,:) = cross(R(3,:),R(1,:));
K(1,:) = [alpha -alpha*cot(theta) X0];
K(2,:) = [0 Beta/sin(theta) Y0];
K(3,:) = [0 0 1];
t = rho*inv(K)*M(:,4);
```

Results

1. Internal Calibration Matrix (K):

```
K =

   520.65399     0.22719   311.48234
     0.00000   504.39341   234.52529
     0.00000     0.00000     1.00000
```

2. Intrinsic Camera Parameters:

$$K = \begin{pmatrix} fk_u & -fk_u\cot(\theta) & u_0 \\ 0 & \dfrac{fk_v}{\sin(\theta)} & v_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here f is the focal length in mm, k_u and k_v are the number of pixels per millimeter (width and height, respectively), and u_0, v_0 are the center of projection. If we let α and β be f k_u and f k_v respectively by multiplying the focal length (f) in mm by k, given in mm/pixel, this allows us to work in pixel units. The ratio α/β is the aspect ratio and is often (but not always) one. The skew angle θ is almost always 90 degrees because modern camera-sensing elements are manufactured accurately.

```
u0 =  311.48
v0 =  234.53
theta =  1.5712
alpha =  520.65
Beta =  504.39
```

Hence, aspect ratio is 1.0322 and skew is 0.046 degrees, focal length is 26mm.

3. Extrinsic Camera Parameters:

```
R =

  -0.798920   0.601062  -0.021231
  -0.115467  -0.118642   0.986201
   0.590249   0.790347   0.164188

t =

   -0.72551
   -4.43240
  -17.16271
```

4. Camera center X0:

```
x0 =

    9.0389
   13.4747
    7.1737
```

Importance of Normalization

During the construction of matrix Q, we have rows of the form [Pi 0 -xi*Pi] and [Pi 0 -yi*Pi]. If we don't normalize the image and world coordinates, then the values of Xi, Yi, Zi in Pi and xi, yi can be quite large. With these large values, we'll also have 0 and 1(in Pi) in the matrix. So, the values in the matrix will vary greatly. This may make the matrix Q^TQ ill conditioned and the errors while calculating the eigenvalues and eigenvectors of Q^TQ will be large. This will result in a projection matrix M with large error.

Estimation of the fundamental matrix can be improved by normalizing the coordinates before computing the fundamental matrix. In real world images, the source and target coordinates are usually noisy, and could have large variance. A simple solution is to normalize the image coordinates before estimating the fundamental matrix. This provides a very well-balanced matrix Q and a much more stable and accurate results for the projection matrix M.