16-720A Computer Vision: Homework 4 3D Reconstruction

Radhika Mohanan(rmohanan)

1 Theory

Q1.1

It is given that both principal points of two image planes coincide with coordinate (0,0), thus the projected point x in the two image planes are $x_1^T = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ and $x_2^T = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$

According to the property of fundamental matrix when the points are at the origin we have,

$$x_2^T F x_1 = 0$$

$$\begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0 \tag{1}$$

$$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0$$

On simplifying the above matrices, all the terms that have the image point coordinates get cancelled at the origin and therefore, we get the following relation.

$$F_{33} = 0 \tag{2}$$

Q1.2

For pure translation along the x-axis, transformation matrices are as follows:

$$\begin{aligned} R &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ t &= \begin{bmatrix} t_1 \\ 0 \\ 0 \end{bmatrix} \end{aligned} \tag{3}$$

The essential matrix is given by,

$$E = t_\times R \tag{4}$$

$t_\times$ and essential matrix can be written as follows,

$$\begin{aligned} t_\times &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \\ E &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \end{aligned} \tag{5}$$

Let $\widetilde{x_1^T} = \begin{bmatrix} a_1 & a_2 & 1 \end{bmatrix}$ and $\widetilde{x_2^T} = \begin{bmatrix} b_1 & b_2 & 1 \end{bmatrix}$, we have

$$l_1^T = \widetilde{x_2^T} E \tag{6}$$

$$l_1^T = \begin{bmatrix} b_1 & b_2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & t_1 & -b_2 t_1 \end{bmatrix} \tag{7}$$

$$l_2^T = \widetilde{x_1^T} E \tag{8}$$

$$l_2^T = \begin{bmatrix} a_1 & a_2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & t_1 \\ 0 & -t_1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -t_1 & a_2 t_1 \end{bmatrix} \tag{9}$$

Thus, the epipolar line in the first camera is $t_1 y_1 - b_2 t_1 = 0$, and the epipolar line in the second camera is $-t_1 y_2 + a_2 t_1 = 0$. Both of them do not contain x components, so they are both parallel to the x-axis

Q1.3

Let P be 3D world coordinates of the point in the image, p_1 be 2D coordinates on the image plane at time frame i and p_2 be 2D coordinates on the image plane at time frame $i+1$ World coordinate and image plane can be related by:

$$P = t_1 + R_1 p_1 \tag{10}$$
$$p_1 = {R_1}^{-1}(P - t_1) \tag{11}$$
$$\tag{12}$$

Similarly,

$$p_2 = {R_2}^{-1}(P - t_2) \tag{13}$$

Combining the above equations:

$$p_2 = {R_2}^{-1}(t_1 + R_1 p_1 - t_2) \tag{14}$$
$$p_2 = {R_2}^{-1} R_1 p_1 + {R_2}^{-1}(t_1 - t_2) \tag{15}$$
$$\tag{16}$$

Comparing the above equation with $p_2 = R_{rel} p_1 + t_{rel}$

$$R_{rel} = {R_2}^{-1} R_1 \tag{17}$$
$$t_{rel} = {R_2}^{-1}(t_1 - t_2) \tag{18}$$
$$\tag{19}$$

Also from the equations for essential matrix and fundamental matrix:

$$E = [t_{rel}]_x R_{rel} \tag{20}$$
$$F = K^{-T}[t_{rel}]_x R_{rel} K^{-1} \tag{21}$$

Q1.4

Let C and C' be the camera in the real and virtual world respectively, its intrinsic matrix be K. Let P and x be the 3D point in real world and the point in image plane and P' and x' be its reflection in the in the mirror and this point in the image plane.Given the mirror is flat, the transformation between these two points is a pure translation.

$$
\begin{gathered}
R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \\
P' = \ P + \ t \\
\lambda_1 \ x = K \ P \\
\lambda_2 \ x' = K \ P'
\end{gathered} \tag{22}
$$

With the help of the above equations, relationship between the two points is as follows.

$$\lambda_2 K^{-1} \ x' = \lambda_1 K^{-1} \ x + \ t \tag{23}$$

We can simplify the equation and eliminate some terms by taking cross product with t on both sides, followed by dot product with x' to get the following.

$$
\begin{gathered}
x'^T K^{-T} \ t_\times K^{-1} \ x = 0 \\
t_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}
\end{gathered} \tag{24}
$$

We can see that, $t_\times$ is a skew-symmetric matrix here and we know the relation that $x'^T \ F \ x = 0$. Comparing this with above form, we get the following expression for F.

$$F = K^{-T} \ t_\times K^{-1} \tag{25}$$

Since, $t_\times$ is skew symmetric, it can be shown that F here will also retain the property of skew-symmetric for a given intrinsic matrix K.

$$F^T = - \ F \tag{26}$$

Therefore, we can conclude that the two images of the object are related by a skew-symmetric fundamental matrix.

2 Fundamental Matrix Estimation

Q2.1 The Eight Point Algorithm

The visualization of the algorithm is:

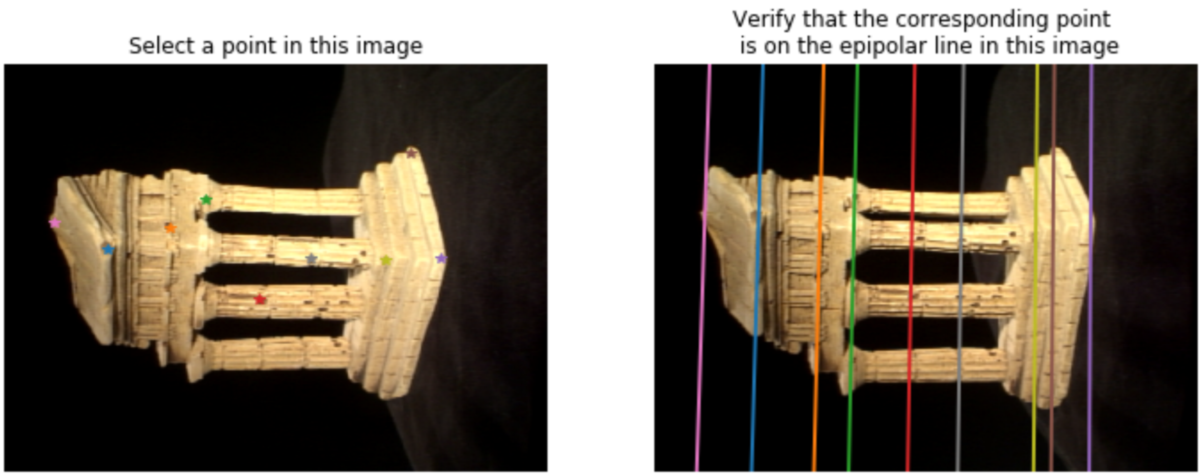

Figure 1: Eight-Point Algorithm Output

The recovered matrix F from eight point algorithm is

```
[[-8.33149229e-09  1.29538462e-07 -1.17187851e-03]
 [ 6.51358336e-08  5.70670058e-09 -4.13435037e-05]
 [ 1.13078765e-03  1.91823637e-05  4.16862080e-03]]
```

3 Metric Reconstruction

Q3.1 Essential Matrix

We know that, the essential and fundamental matrices are related by $K_2^T F K_1$. Therefore, the essential matrix from the fundamental matrix calculated in 2.1 is as follows:

```
[[-1.92592122e-02  3.00526429e-01 -1.73693252e+00]
 [ 1.51113724e-01  1.32873151e-02 -3.08885271e-02]
 [ 1.73986815e+00  9.11774760e-02  3.90697725e-04]]
```

Q3.2 Essential Matrix

Let C_{1i} be the i^{th} row of C_1 and C_{2i} be the i^{th} row of C_2. If W_i is a 4X1 vector of the 3D coordinates in the homogeneous form, we have

$$\begin{gathered} C_1 W_i = \widetilde{x_{i1}} \\ \begin{bmatrix} C_{11} \\ C_{12} \\ C_{13} \end{bmatrix} \times \begin{bmatrix} u_i \\ v_i \\ w_i \\ 1 \end{bmatrix} = \begin{bmatrix} x_{i1} \\ y_{i1} \\ 1 \end{bmatrix} \\ C_2 W_i = \widetilde{x_{i2}} \\ \begin{bmatrix} C_{21} \\ C_{22} \\ C_{23} \end{bmatrix} \times \begin{bmatrix} u_i \\ v_i \\ w_i \\ 1 \end{bmatrix} = \begin{bmatrix} x_{i2} \\ y_{i2} \\ 1 \end{bmatrix} \\ C_{11} W_i = x_{i1} \\ C_{12} W_i = y_{i1} \\ C_{13} W_i = 1 \\ C_{21} W_i = x_{i2} \\ C_{22} W_i = y_{i2} \\ C_{23} W_i = 1 \end{gathered} \tag{27}$$

On rearranging the terms we get,

$$\begin{gathered} (x_{i1} C_{13} - C_{11}) W_i = 0 \\ (y_{i1} C_{13} - C_{12}) W_i = 0 \\ (x_{i2} C_{23} - C_{21}) W_i = 0 \\ (y_{i2} C_{23} - C_{22}) W_i = 0 \end{gathered} \tag{28}$$

Thus A can be written as,

$$A_i = \begin{bmatrix} x_{i1} C_{13} - C_{11} \\ y_{i1} C_{13} - C_{12} \\ x_{i2} C_{23} - C_{21} \\ y_{i2} C_{23} - C_{22} \end{bmatrix} \tag{29}$$

Q3.3

The function has been implemented in findM2.py. The best M2 value and error is as follows:

```
[[ 0.9993695   0.03519757 -0.00466109  0.01780296]
 [-0.03283858  0.96623389  0.25556546 -1.        ]
 [ 0.01349899 -0.25525126  0.96678052  0.08705062]]
```

Best error:94.15843096298792

4 3D Visualization

Q4.1 Epipolar Correspondences

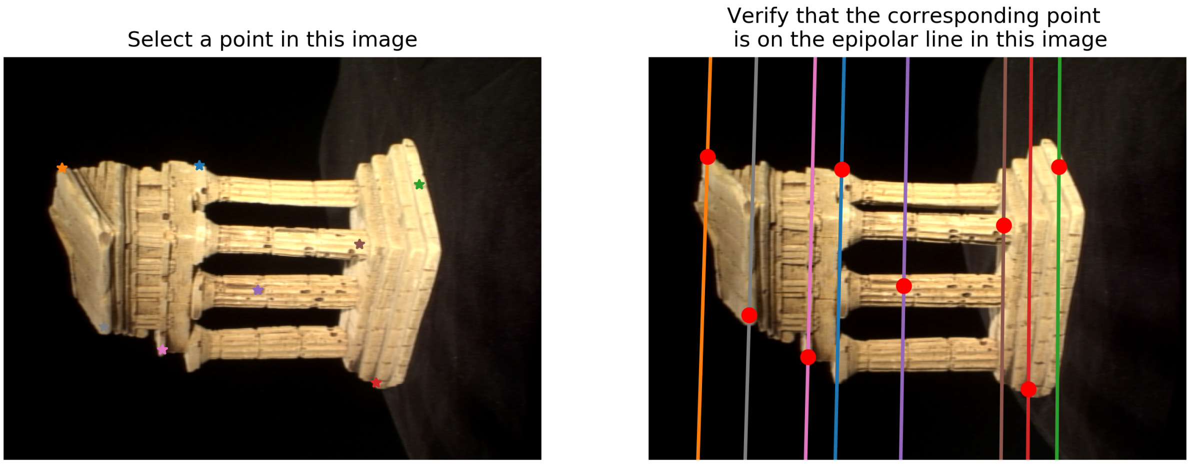

Figure 2: Epipolar Correspondences

From the figures it can be seen that the corresponding points in different perspectives matched well.

Q4.2 3D Visualization

We can see from the images given below that the triangulate function was able to convert the 2D points to 3D as the visualization of the temple points outline is clearly visible.

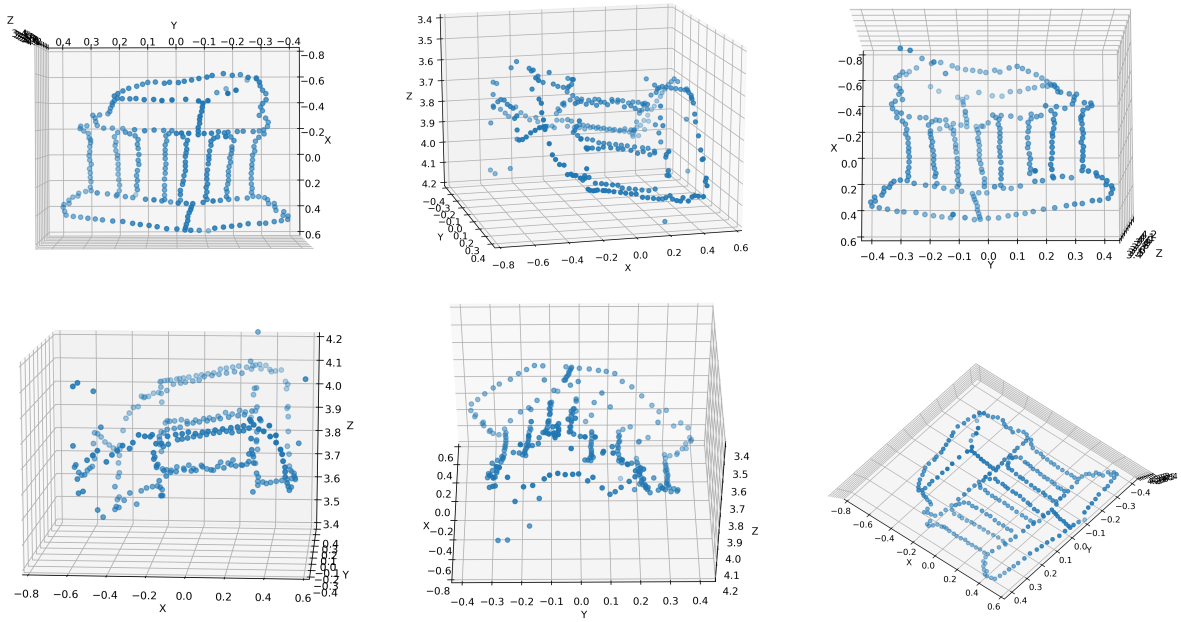

Figure 3: 3D Visualization of the Temple

5 Bundle Adjustment

Q5.1 The Eight Point Algorithm with RANSAC

Using "some_corresp_noisy.npz", without RANSAC, the visualization of the epipolar lines appears to be converging at a point.While with RANSAC, the visualization looks much better.

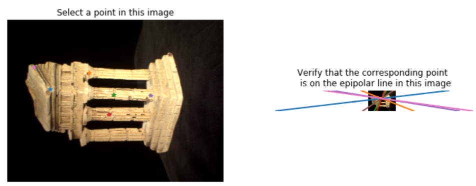

Figure 4: Eight-Point Algorithm without RANSAC Output

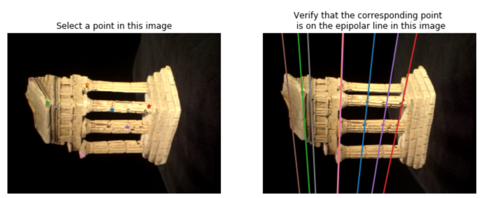

Figure 5: Eight-Point Algorithm with RANSAC Output

For the computed F_i in the ith iteration of RANSAC, an corresponding point pairs as X_{i1} and X_{i2} ,
The error is calculated as: $error_i = X_{i2}^T F_i X_{i1}$

If $error_i$ is smaller than a tolerance (i.e. 0.82), we consider this pair of points as an inlier.
Effect of number of iterations and tolerance on fundamental matrix.
1) As the number of iterations is increased the fundamental matrix becomes more accurate in determining the points as it is a non-deterministic algorithm that produces result only with a certain probability, the probability increases with iterations. But after a point of threshold iteration the results provided by fundamental matrix stops changing value.
2) With the increase in the tolerance value more inliers and thus data points gets included due to which estimation of fundamental matrix becomes more accurate. But higher tolerance will allow all the points to be included as inliers and defeat the purpose of RANSAc.

Q5.3 Optimization

```
Re-projection Error for Inlier Points (Before Optimization):1181.399812131348
Re-projection Error for Inlier Points (After Optimization): 6.937519730014351
```

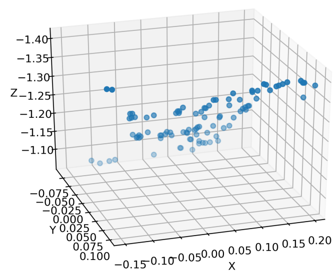

Figure 6: 3D Visualization of the Points Before optimization)

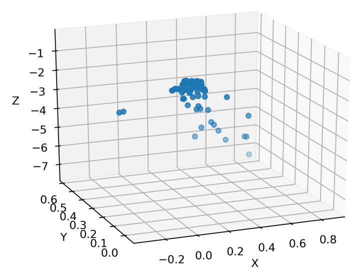

Figure 7: 3D Visualization of the Points after Optimization

From the above figures you can observe that after optimization the spread of random points has reduced.

6 Multiview Keypoint Reconstruction

Q6.1 Multiview Reconstruction

In order to compute the 3D location, triangulate function was extended to three views by changing the A matrix as follows:

$$A_i = \begin{bmatrix} x_{i1}C_{13} - C_{11} \\ y_{i1}C_{13} - C_{12} \\ x_{i2}C_{23} - C_{21} \\ y_{i2}C_{23} - C_{22} \\ x_{i3}C_{33} - C_{31} \\ y_{i3}C_{33} - C_{32} \end{bmatrix} \tag{30}$$

This modified A matrix is used to solve SVD and get the 3D location of points. From the given 2D points only the ones with confidence value greater than threshold is used for finding the 3D locations. After multiple iterations with the threshold, value of 140 gave the best results.

The keypoint reconstruction for the image 'cam1_time1.jpg' at time1 is as follows:

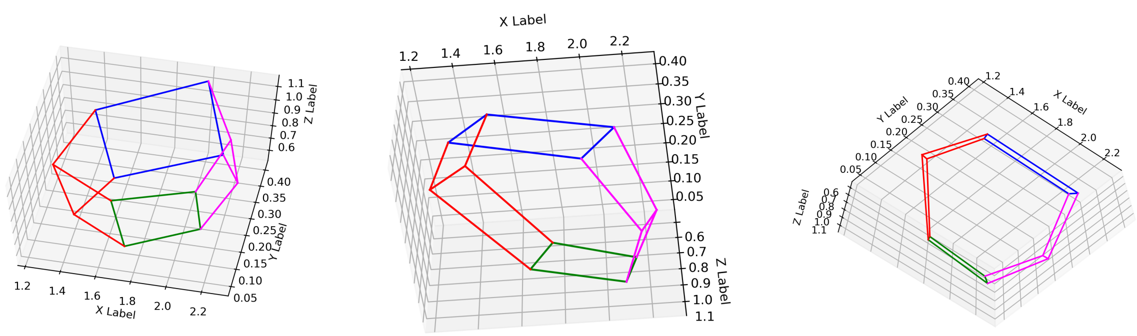

Figure 8: 3D Visualization of the KeyPoints

Q6.2 Extra Credit

After iteratively repeating the process of Q6.1 over time, we compute a spatio-temporal reconstruction of the car for all 10 time instances.The images from multiple views are as below.

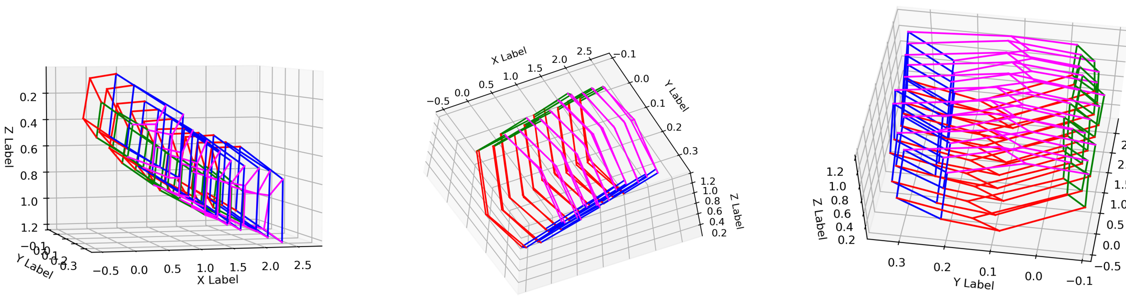

Figure 9: spatio-temporal reconstruction