16-720 Computer Vision
Spring 2020

Homework 4

Zongwen Mu, Andrew ID: zongwenm

1 Theory

Q1.1

For given point $\mathbf{x}$, we have $\tilde{x_1} = \tilde{x_2} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\mathbf{T}}$. Therefore, we have:

$$\tilde{x_2}^{\mathbf{T}} F \tilde{x_1} = 0 \tag{1}$$

$$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0 \tag{2}$$

$$F_{33} = 0 \tag{3}$$

Q1.2

For pure translation parallel to x-axis:

$$t = \begin{bmatrix} t_1 \\ 0 \\ 0 \end{bmatrix} \tag{4}$$

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

$$E = tR \tag{6}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \tag{7}$$

Suppose that $\tilde{x_1}^T = \begin{bmatrix} a_1 & a_2 & 1 \end{bmatrix}$ and $\tilde{x_2}^T = \begin{bmatrix} b_1 & b_2 & 1 \end{bmatrix}$, we have:

$$l_1^T = \tilde{x_2}^T E \tag{8}$$

$$= \begin{bmatrix} b_1 & b_2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \tag{9}$$

$$= \begin{bmatrix} 0 & t_1 & -b_2 t_1 \end{bmatrix} \tag{10}$$

$$l_2^T = \tilde{x_1}^T E^T \tag{11}$$

$$= \begin{bmatrix} a_1 & a_2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & t_1 \\ 0 & -t_1 & 0 \end{bmatrix} \tag{12}$$

$$= \begin{bmatrix} 0 & -t_1 & a_2 t_1 \end{bmatrix} \tag{13}$$

Thus, we could acquire epipolar lines in the two cameras: $t_1 y_1 - b_2 t_1 = 0, -t_1 y_2 + a_2 t_1 = 0$. Both do not contain x component, so they are parallel to the x-axis.

Q1.3

Assume that the coordinate of the object in 3D world is $\begin{bmatrix} u & v & w \end{bmatrix}^T$, and let $\begin{bmatrix} x_i & y_i \end{bmatrix}^T$ be the position at time i. Thus we have:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = K \left(R_1 \begin{bmatrix} u \\ v \\ w \end{bmatrix} + t_1 \right) \tag{14}$$

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = R_1^{-1} \left(K^{-1} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} - t_1 \right) \tag{15}$$

$$= R_1^T K^{-1} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} - R_1^T t_1 \tag{16}$$

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = K \left(R_2 \begin{bmatrix} u \\ v \\ w \end{bmatrix} + t_2 \right) \tag{17}$$

$$= K \left(R_2 \left(R_1^T K^{-1} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} - R_1^T t_1 \right) + t_2 \right) \tag{18}$$

$$= K R_2 R_1^T K^{-1} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} - K R_2 R_1^T t_1 + K t_2 \tag{19}$$

Therefore:

$$R_{rel} = K R_2 R_1^T K^{-1} \tag{20}$$

$$t_{rel} = -K R_2 R_1^T t_1 + K t_2 \tag{21}$$

$$E = t_{rel} \times R_{rel} \tag{22}$$

$$F = \left(K^{-1}\right)^T E K^{-1} \tag{23}$$

$$= \left(K^{-1}\right)^T \left(t_{rel} \times R_{rel}\right) K^{-1} \tag{24}$$

Q1.4

Assume that the distance between object and mirror is d, thus the distance between two images is $2d$ and there is pure translation:

$$R_{rel} = \mathbf{I} \tag{25}$$

$$t_{rel} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix} \tag{26}$$

$$F = \left(K^{-1}\right)^T \left(t_{rel} \times R_{rel}\right) K^{-1} \tag{27}$$

$$= \left(K^{-1}\right)^T \begin{bmatrix} 0 & t_z & -t_y \\ -t_z & 0 & t_x \\ t_y & -t_x & 0 \end{bmatrix} K^{-1} \tag{28}$$

Thus, fundamental matrix $\mathbf{F}$ is a skew-symmetric matrix.

2 Fundamental Matrix Estimation

Q2.1

The recovered matrix $\mathbf{F}$ is:

$$\mathbf{F} = \begin{bmatrix} 9.78833286e-10 & -1.32135929e-07 & 1.12585666e-03 \\ -5.73843315e-08 & 2.96800276e-09 & -1.17611996e-05 \\ -1.08269003e-03 & 3.04846703e-05 & -4.47032655e-03 \end{bmatrix} \tag{29}$$

and example output image is shown below:

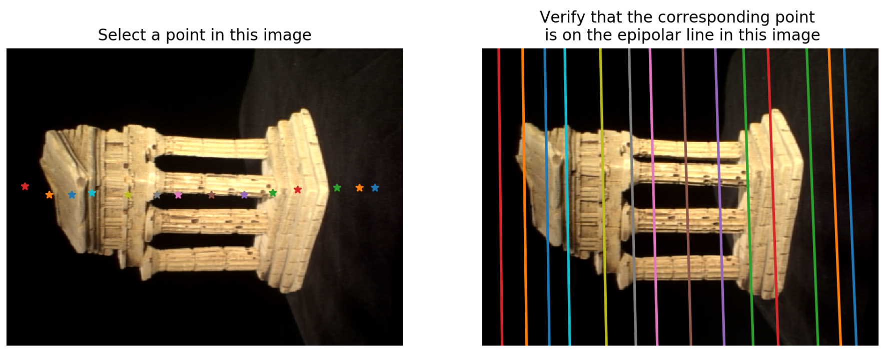


Figure 1: Visualized Epipolar Lines

3 Metric Reconstruction

Q3.1

$\mathbf{E}$ can be calculated by $\mathbf{K_2^T F K_1}$. Using the eight-point algorithm, the estimated $\mathbf{E}$ is:

$$\mathbf{E} = \begin{bmatrix} 2.26268684e-03 & -3.06552495e-01 & 1.66260633e+00 \\ -1.33130407e-01 & 6.91061098e-03 & -4.33003420e-02 \\ -1.66721070e+00 & -1.33210351e-02 & -6.72186431e-04 \end{bmatrix} \tag{30}$$

Q3.2

Suppose $\mathbf{C}_{1i}$ and $\mathbf{C}_{2i}$ is the ith row for $\mathbf{C}_1$ and $\mathbf{C}_2$. If $\tilde{\mathbf{w}}_i$ is a 4×1 vector of the 3D coordinate in the homogeneous form, we have:

$$\mathbf{C}_1 \tilde{\mathbf{w}}_i = \widehat{\mathbf{x}_{1i}} \tag{31}$$

$$\begin{bmatrix} \mathbf{C}_{11} \\ \mathbf{C}_{12} \\ \mathbf{C}_{13} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ w_i \\ 1 \end{bmatrix} = \begin{bmatrix} x_{1i} \\ y_{1i} \\ 1 \end{bmatrix} \tag{32}$$

$$\mathbf{C}_2 \tilde{\mathbf{w}}_i = \widehat{\mathbf{x}_{2i}} \tag{33}$$

$$\begin{bmatrix} \mathbf{C}_{21} \\ \mathbf{C}_{22} \\ \mathbf{C}_{23} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ w_i \\ 1 \end{bmatrix} = \begin{bmatrix} x_{2i} \\ y_{2i} \\ 1 \end{bmatrix} \tag{34}$$

Therefore, we have:

$$\mathbf{A}_i = \begin{bmatrix} x_{1i}\mathbf{C}_{13} - \mathbf{C}_{11} \\ y_{1i}\mathbf{C}_{13} - \mathbf{C}_{12} \\ x_{2i}\mathbf{C}_{23} - \mathbf{C}_{21} \\ y_{2i}\mathbf{C}_{23} - \mathbf{C}_{22} \end{bmatrix} \tag{35}$$

4 3D Visualization

Q4.1

Choose window size 20, the matched result is shown below:

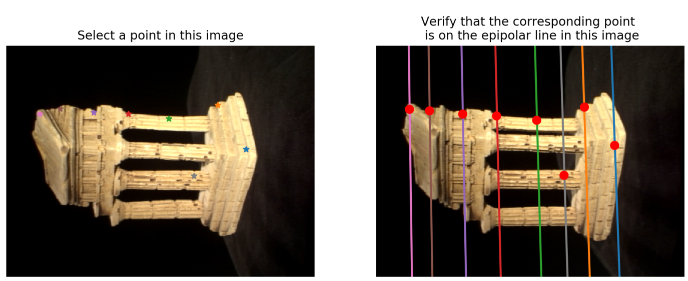
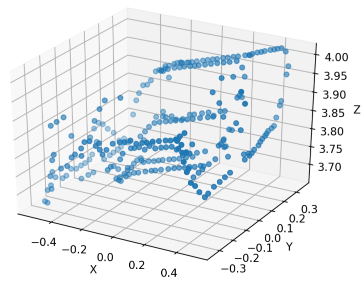
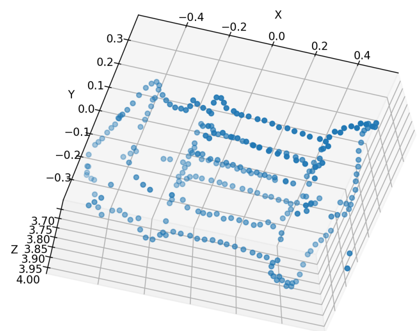


Figure 2: Corresponding Points

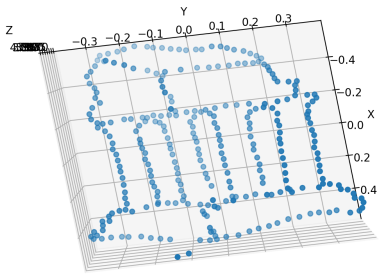
Q4.2 The results for $3D$ visualization is shown below:



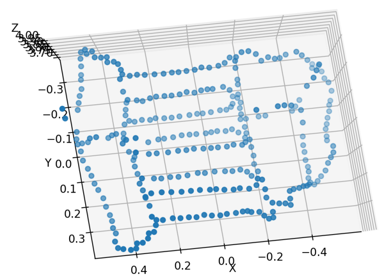
(a)



(b)



(c)



(d)

Figure 3: 3D Visualization

5 Bundle Adjustment

Q5.1

Using the noisy correspondences, without RANSAC, the visualization of epipolar lines is like:

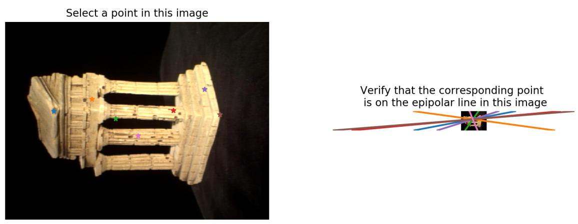


Figure 4: Visualization without RANSAC

With RANSAC implemented, the result looks much better:

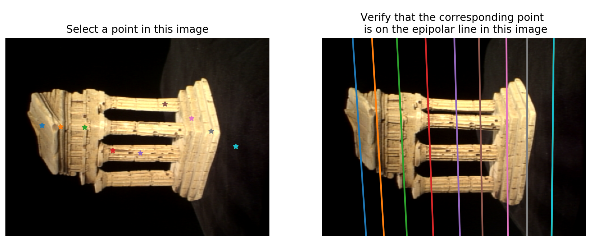


Figure 5: Visualization with RANSAC

Obtained fundamental matrix $\mathbf{F}$ is:

$$\mathbf{F} = \begin{bmatrix} 1.44845968e-08 & -3.12242026e-07 & 1.07078108e-03 \\ 1.67046292e-07 & 1.00324202e-08 & -8.51505084e-05 \\ -1.04288634e-03 & 9.23538949e-05 & -2.07767875e-03 \end{bmatrix} \tag{36}$$

For fundamental matrix, we have:

$$\tilde{\mathbf{x}_2}^T \mathbf{F} \tilde{\mathbf{x}_1} = 0 \tag{37}$$

So the error metrics used to determine if point i is an inlier is:

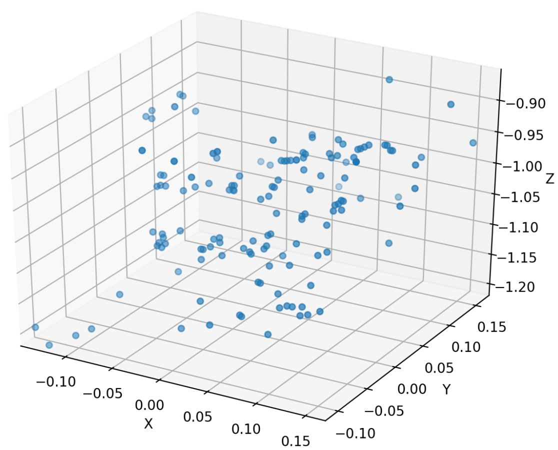
$$err = abs\left(\tilde{\mathbf{x}_{2i}}^T \mathbf{F} \tilde{\mathbf{x}_{1i}}\right) \tag{38}$$

Set tolerance as 0.8 and after 100 iterations, the **ransacF** function was able to find an ideal enough matrix $\mathbf{F}$.

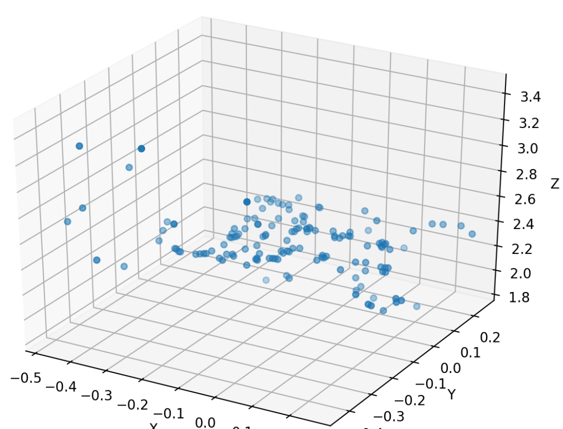
While tuning the parameters, turning the tolerance to a smaller number would decrease the inlier number, which would cause lower accuracy for RANSAC, and with more iterations, RANSAC would be able to find a better solution.

Q5.3

The resulting images are shown below:



(a) without bundle adjustment



(b) with bundle adjustment

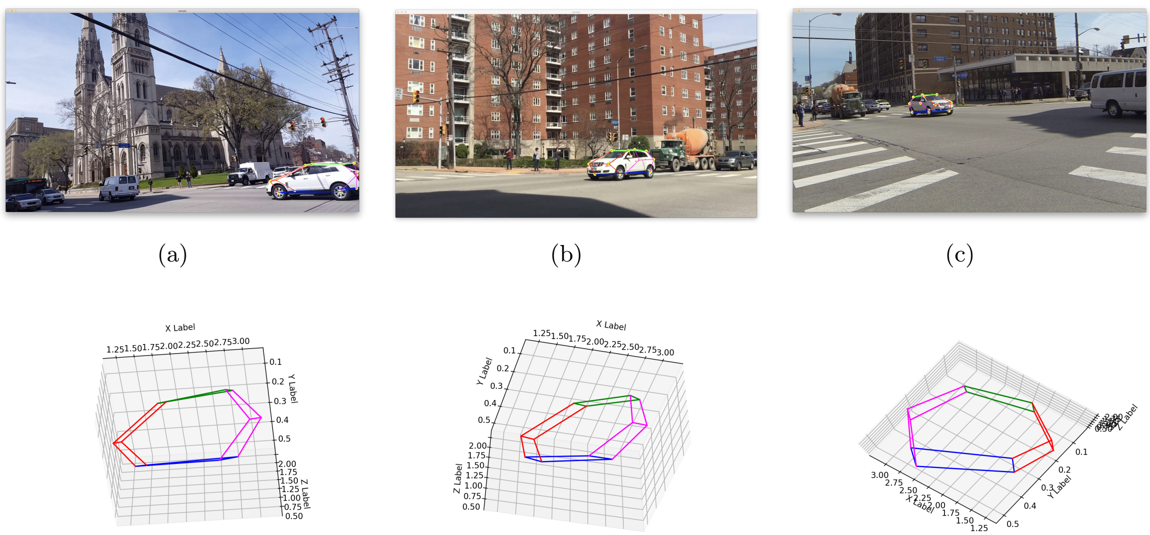
Figure 6: 3D Reconstruction with Noisy Data

Without bundle adjustment, the reprojection error is 51.484053051875, while with bundle adjustment the error is 32.15606321697, which significantly decreased.

6 Multi View Keypoint Reconstruction

Q6.1

In this case, I used the triangulate function I've written before to calculate 3 sets of $\begin{bmatrix}\mathbf{w} & err\end{bmatrix}$, and compared the errors to decide the one $\mathbf{w}$ with the smallest error, the chose this $\mathbf{w}$ as the one used in reconstruction. An example resulting image is shown below:



(a) (b) (c)

(d) (e) (f)

Figure 7: Detections and the Reconstructions from Multiple Views

Tuning the parameter threshold would influence the accuracy in keypoints detection, with lower threshold would lead to more accurate detection and reconstruction.