16-720B Homework 4 Write-up

Gu, Qiao

November 9, 2019

Q1.1

Consider that the point $\mathbf{w}$ is the point in 3D where the principle axes of the two cameras intersect, and we can see that $\tilde{\mathbf{x}}_1 = [0, 0, 1]^T$ and $\tilde{\mathbf{x}}_2 = [0, 0, 1]^T$ are its projections in camera 1 and camera 2 respectively. Therefore

$$\tilde{\mathbf{x}}_2^T \mathbf{F} \tilde{\mathbf{x}}_1 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{13} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} \\ \mathbf{F}_{31} & \mathbf{F}_{32} & \mathbf{F}_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{F}_{33} = 0 \tag{1}$$

Q1.2

The translation and rotation from camera 1 to camera 2 are

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} t_x \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

And thus the essential matrix are

$$\mathbf{E} = \mathbf{t}_{\times}\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix} \tag{3}$$

Therefore for an epipolar line in camera 1 $\mathbf{l}_1^T \tilde{\mathbf{x}}_1 = 0$ and $\tilde{\mathbf{x}}_2^T \mathbf{E} \tilde{\mathbf{x}}_1 = 0$, where $\tilde{\mathbf{x}}_2$ is a fixed point on the image plane of camera 2 resulting from the ray corresponding to the epipolar line, then we can see that

$$\mathbf{l}_1^T = \tilde{\mathbf{x}}_2^T \mathbf{E} = \begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix} = \begin{bmatrix} 0 & t_x & -t_x y_2 \end{bmatrix} \tag{4}$$

Similarly we can see that any epipolar line in camera 1 has $\mathbf{l}_2^T = [0,\ -t_x,\ t_x y_1]$. Since the first elements in both $\mathbf{l}_1$ and $\mathbf{l}_2$ are zero, the epipolar lines are parallel to x axis.

Q1.3

Assume $(\mathbf{R}_i, \mathbf{t}_i)$ and $(\mathbf{R}_i, \mathbf{t}_i)$ are the rotation and translation from the world coordinate frame to the camera coordinate frame at time i and time j. And suppose $\mathbf{R}_{rel}$ and $\mathbf{t}_{rel}$ are the rotation and translation from camera at time i to the camera at time j. Then for a point $\mathbf{w}$ in the 3D world

$$
\begin{aligned}
&\lambda_i \tilde{\mathbf{x}}_i = \mathbf{R}_i \mathbf{w} + \mathbf{t}_i, \quad \lambda_j \tilde{\mathbf{x}}_j = \mathbf{R}_j \mathbf{w} + \mathbf{t}_j \\
\Rightarrow &\mathbf{w} = \mathbf{R}_i^T (\lambda_i \tilde{\mathbf{x}}_i - \mathbf{t}_i) \\
\Rightarrow &\lambda_j \tilde{\mathbf{x}}_j = \mathbf{R}_j \mathbf{R}_i^T (\lambda_i \tilde{\mathbf{x}}_i - \mathbf{t}_i) + \mathbf{t}_j \\
\Rightarrow &\lambda_j \tilde{\mathbf{x}}_j = \mathbf{R}_j \mathbf{R}_i^T \lambda_i \tilde{\mathbf{x}}_i - \mathbf{R}_j \mathbf{R}_i^T \mathbf{t}_i + \mathbf{t}_j \\
\Rightarrow &\lambda_j \tilde{\mathbf{x}}_j = \lambda_i \mathbf{R}_{rel} \tilde{\mathbf{x}}_i + \mathbf{t}_{rel} \qquad (5)
\end{aligned}
$$

Therefore

$$
\mathbf{R}_{rel} = \mathbf{R}_j \mathbf{R}_i^T, \quad \mathbf{t}_{rel} = \mathbf{t}_j - \mathbf{R}_j \mathbf{R}_i^T \mathbf{t}_i \qquad (6)
$$

Then the essential and fundamental matrix can be derived as

$$
\mathbf{E} = (\mathbf{t}_{rel})_\times \mathbf{R}_{rel} \qquad (7)
$$

$$
\mathbf{F} = (\mathbf{K}^{-1})^T \mathbf{F} \mathbf{K}^{-1} = (\mathbf{K}^{-1})^T (\mathbf{t}_{rel})_\times \mathbf{R}_{rel} \mathbf{K}^{-1} \qquad (8)
$$

Q1.4

Suppose the real world coordinate has its origin at the optical center of the camera and the mirror is orthogonal to a unit vector $\mathbf{v}$ pointing in to the mirror. Then for any point $\mathbf{w}$ in the world coordinate, the mirror produce its reflection $\mathbf{w}_2 = \mathbf{w}_1 + 2\alpha\mathbf{v}$, where is the dixed distance from $\mathbf{w}_1$ to the mirror.

These two points in 3D produce two point on the image plane as follows

$$\lambda_1\tilde{\mathbf{x}}_1 = \mathbf{w}_1 \tag{9}$$

$$\lambda_2\tilde{\mathbf{x}}_2 = \mathbf{w}_2 = \mathbf{w}_1 + 2\alpha\mathbf{v} \tag{10}$$

This is equivalent to a two-camera system where $\mathbf{R} = \mathbf{I}$ and $\mathbf{t} = 2\alpha\mathbf{v}$. Therefore

$$\mathbf{E} = \mathbf{t}_\times\mathbf{R} = 2\alpha\mathbf{v}_\times\mathbf{I} = 2\alpha\mathbf{v}_\times \tag{11}$$

is skew-symmetric as $\mathbf{v}_\times$ is skew-symmetric. Since there are only one camera with only one intrinsic $\mathbf{K}$, for fundamental matrix

$$\mathbf{F} = (\mathbf{K}^{-1})^T\mathbf{E}\mathbf{E}^{-1}\mathbf{F}^T = (\mathbf{K}^{-1})^T\mathbf{E}^T\mathbf{E}^{-1} = -(\mathbf{K}^{-1})^T\mathbf{E}\mathbf{E}^{-1} = -\mathbf{F} \tag{12}$$

Therefore the fundamental matrix $\mathbf{F}$ of this equivalent two-camera system is symmetric.

Q2.1

The fundamental matrix **F** given by the 8-point algorithm is

```
[[ 9.80213861e-10 -1.32271663e-07  1.12586847e-03]
 [-5.72416248e-08  2.97011941e-09 -1.17899320e-05]
 [-1.08270296e-03  3.05098538e-05 -4.46974798e-03]].
```

And visualization result is shown in Figure. 1

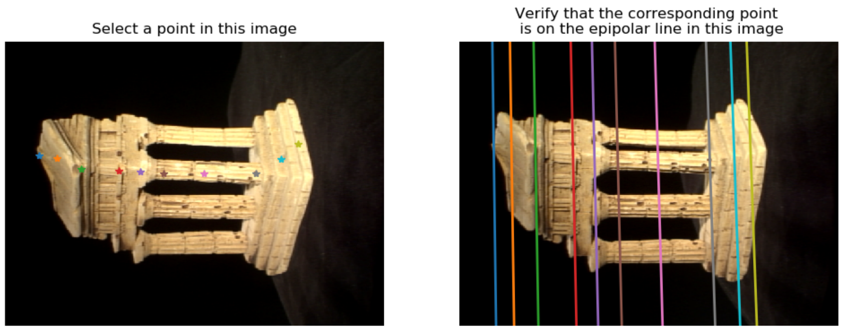


Figure 1: The visualization results showing the fundamental matrix given by 8-point algorithm.

Q2.2

To expedite the process of searching for the best F, I iteratively choose 7 points randomly and keep F with the minimum 2-norm difference from F given by 8-point algorithm. And the best reasonable F is given when I use the points at indices [10 3 92 108 41 30 99], which is

```
[[-1.29290341e-08  1.81485505e-07  8.27390734e-04]
 [-3.52772006e-07  1.05936161e-09  4.14750371e-05]
 [-7.87641850e-04 -1.90120686e-05 -4.69857949e-03]]
```

And the visualization result is shown in Figure. 2

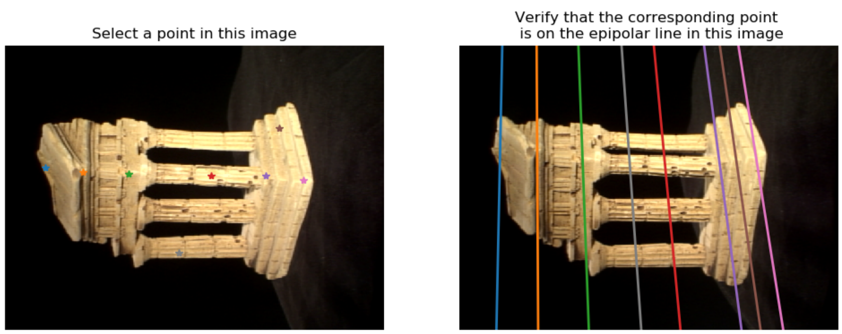


Figure 2: The visualization results showing the fundamental matrix given by 7-point algorithm.

Q3.1

By applying the equation $\mathbf{E} = \mathbf{K}_2^T \mathbf{F} \mathbf{K}_1$, we can get the $\mathbf{E} =$

```
[[ 2.26587820e-03 -3.06867395e-01  1.66257398e+00]
 [-1.32799331e-01  6.91553934e-03 -4.32775554e-02]
 [-1.66717617e+00 -1.33444257e-02 -6.72047195e-04]]
```

Q3.2

Suppose $\tilde{\mathbf{w}}$ is the homogenuous coordinate of the 3D point $\mathbf{w}$, and it projects $\tilde{\mathbf{x}}_{i1}$ and $\tilde{\mathbf{x}}_{i2}$ on camera 1 and camera 2 respectively, which means

$$\mathbf{C}_1\tilde{\mathbf{w}}_i = \lambda_1\tilde{\mathbf{x}}_{i1}, \quad \mathbf{C}_2\tilde{\mathbf{w}}_i = \lambda_2\tilde{\mathbf{x}}_{i2} \tag{13}$$

Now only consider $\mathbf{C}_1\tilde{\mathbf{w}} = \lambda_1\tilde{\mathbf{x}}_{i1}$. Suppose $\tilde{\mathbf{x}}_{i1} = [x_{i1}, y_{i1}, 1]^T$ and $\mathbf{C}_{11}^T, \mathbf{C}_{12}^T, \mathbf{C}_{13}^T$ are the first, the second, and the third row of the camera matrix $\mathbf{C}_1$. We get

$$\begin{cases} \lambda_1 x_{i1} & = \mathbf{C}_{11}^T\tilde{\mathbf{w}}_i \\ \lambda_1 y_{i1} & = \mathbf{C}_{12}^T\tilde{\mathbf{w}}_i \\ \lambda_1 & = \mathbf{C}_{13}^T\tilde{\mathbf{w}}_i \end{cases} \Rightarrow \begin{cases} x_{i1}\mathbf{C}_{13}^T\tilde{\mathbf{w}}_i & = \mathbf{C}_{11}^T\tilde{\mathbf{w}}_i \\ y_{i1}\mathbf{C}_{13}^T\tilde{\mathbf{w}}_i & = \mathbf{C}_{12}^T\tilde{\mathbf{w}}_i \end{cases} \Rightarrow \begin{bmatrix} x_{i1}\mathbf{C}_{13}^T - \mathbf{C}_{11}^T \\ y_{i1}\mathbf{C}_{13}^T - \mathbf{C}_{12}^T \end{bmatrix} \tilde{\mathbf{w}}_i = \mathbf{0} \tag{14}$$

We can get Similar constraints from the projection on camera 2, and by concatenate the constraints together we can get $\mathbf{A}_i\mathbf{w}_i = 0$ as follows:

$$\mathbf{A}_i\mathbf{w}_i = \begin{bmatrix} x_{i1}\mathbf{C}_{13}^T - \mathbf{C}_{11}^T \\ y_{i1}\mathbf{C}_{13}^T - \mathbf{C}_{12}^T \\ x_{i2}\mathbf{C}_{23}^T - \mathbf{C}_{21}^T \\ y_{i2}\mathbf{C}_{23}^T - \mathbf{C}_{22}^T \end{bmatrix} \tilde{\mathbf{w}}_i = \mathbf{0} \text{ and } \mathbf{A}_i = \begin{bmatrix} x_{i1}\mathbf{C}_{13}^T - \mathbf{C}_{11}^T \\ y_{i1}\mathbf{C}_{13}^T - \mathbf{C}_{12}^T \\ x_{i2}\mathbf{C}_{23}^T - \mathbf{C}_{21}^T \\ y_{i2}\mathbf{C}_{23}^T - \mathbf{C}_{22}^T \end{bmatrix} \tag{15}$$

Then $\tilde{\mathbf{w}}_i$ is in the null space of $\mathbf{A}_i$, which we can get by solving SVD decomposition of $\mathbf{A}_i$ and getting the last column of V.

Q4.1

Some detected correspondences are shown in Figure. 3

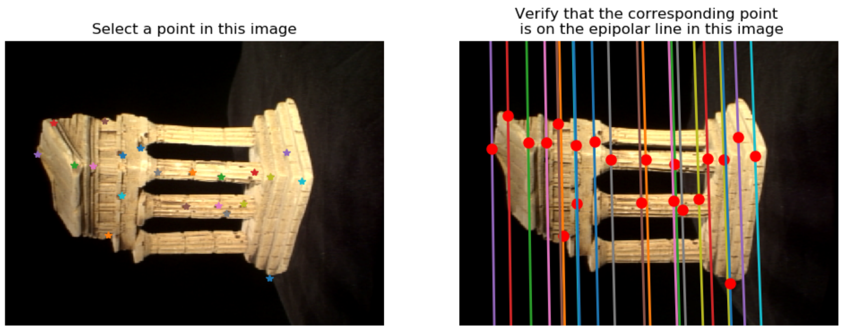
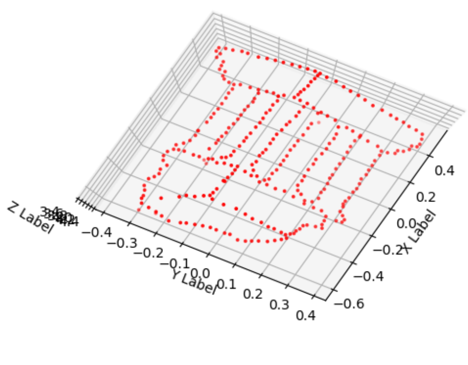


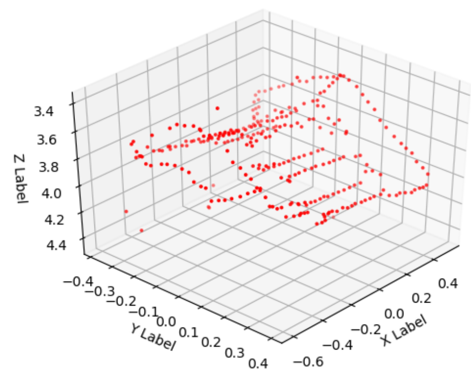
Figure 3: Visualization of some detected correspondences.

Q4.2

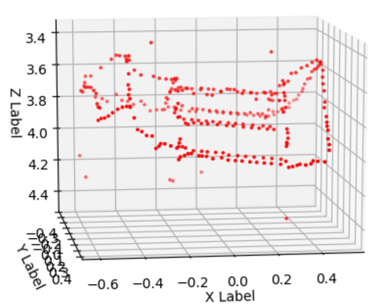
The recovered point could can be viewed in Figure. 4.



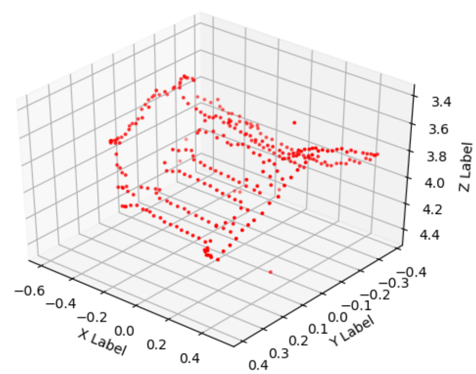
(a) view 1



(b) view 2



(c) view 3



(d) view 4

Figure 4: Different views of the point cloud recovered from `templeCoords`.

Q5.1

As suggested by the lecture notes: When the fundamental matrix is correct, the epipolar line induced by a point in the first image should pass through the matching point in the second image and vice-versa. Therefore, we used the distance of points to corresponding epipolar lines as the criterion in deciding inliers during RANSAC process. More specifically

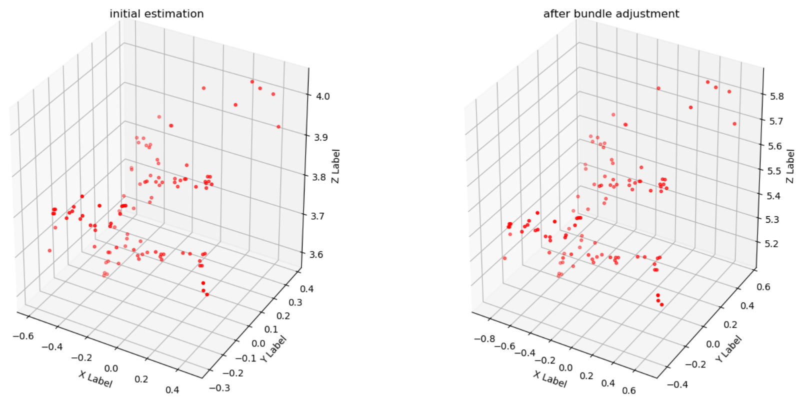
$$Err(\tilde{\mathbf{x}}_2, \mathbf{F}, \tilde{\mathbf{x}}_1) = dist^2(\tilde{\mathbf{x}}_2, \mathbf{F}\tilde{\mathbf{x}}_1) + dist^2(\tilde{\mathbf{x}}_1, \mathbf{F}^T\tilde{\mathbf{x}}_2), \tag{16}$$

$$\text{where } dist(\tilde{\mathbf{x}}_2, \mathbf{F}\tilde{\mathbf{x}}_1) = \frac{\tilde{\mathbf{x}}_2^T\mathbf{F}\tilde{\mathbf{x}}_1}{\sqrt{l_1^2 + l_2^2}}, \mathbf{F}\tilde{\mathbf{x}}_1 = [l_1, l_2, l_3]^T \tag{17}$$

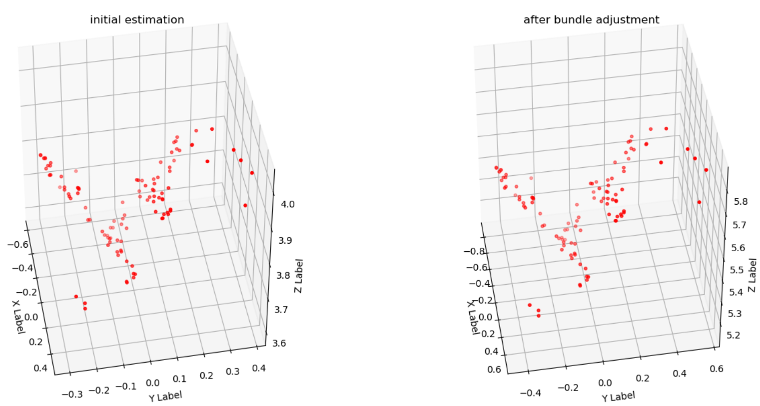
And a correspondence is considered an inlier if its error is below a threshold ϵ. To decide ϵ, I tried different values and found that when $\epsilon = 2$, the RANSAC yield around $140 * 0.75 = 105$ inliers, and thus decide the threshold to be 2.

Q5.3

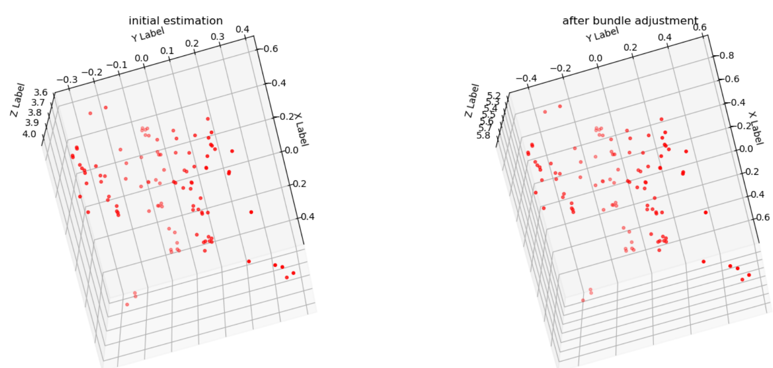
The images showing the 3D points before and after bundle adjustment are shown in Figure. 5. And in my experiement, the reprojection error before the bundle adjustment is 386.57 and the error after the bundle adjustment is 8.93.



(a) view 1



(b) view 2



(c) view 3

Figure 5: Different views of the point cloud before and after the bundle adjustment.