16-720 : Computer Vision: Homework 4 (Spring 2020)
3D Reconstruction

Han-Ting (Quentin) Cheng
Andrew id: hantingc

April 3, 2020

1 Theory

Q1.1

The fundamental matrix describing the two cameras is given as the following:

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

$$x_2^T \mathbf{F} x_1 = 0 \tag{2}$$

In this case both x_2 amd x_1 are $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and when multiplied out, we are left with $\mathbf{F_{33}} = 0$.

Q1.2

Starting from the general equation of the fundamental matrix, we have

$$\begin{gathered} \mathbf{E} = [\mathbf{t_{\times rel}}]\mathbf{R} \\ l_2 = \mathbf{E} x_1 \end{gathered} \tag{3}$$

We know that $t = \begin{bmatrix} t_1 \\ 0 \\ 0 \end{bmatrix}$ because there's only an x direction translation, and R is the identity matrix.
Therefore, the epipolar line is as following,

$$\begin{gathered} [\mathbf{t_{\times rel}}] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \\ \mathbf{E} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \end{gathered} \tag{4}$$

$$\begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -t_1 \\ t_1 y_1 \end{bmatrix} = 0 \tag{5}$$

$$-t_1 y_2 + t_1 y_1 = 0 \tag{6}$$

We can see from this equation, the line is parallel to the x-axis. Same goes when we reverse the two points for the epipolar line in image 1.

Q1.3

The homogeneous transformations measured by the sensors with respect to the inertial frames are as follows.

$$\begin{gathered} \mathbf{H_1^0} = [\mathbf{R_1}|\mathbf{t_1}] \\ \mathbf{H_2^0} = [\mathbf{R_2}|\mathbf{t_2}] \end{gathered} \tag{7}$$

The relative transformation from frame 1 to frame 2 is thus,

$$\begin{gathered} \mathbf{H_2^1} = [\mathbf{R_{rel}}|\mathbf{t_{rel}}] \\ \mathbf{H_2^1} = \left(\mathbf{H_1^0}\right)^{-1} \mathbf{H_2^0} \\ \mathbf{H_2^1} = \begin{bmatrix} \mathbf{R_1^T} & -\mathbf{R_1^T t_1} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R_2} & \mathbf{t_2} \\ \mathbf{0} & 1 \end{bmatrix} \\ \mathbf{H_2^1} = \begin{bmatrix} \mathbf{R_1^T R_2} & \mathbf{R_1^T t_2} - \mathbf{R_1^T t_1} \\ \mathbf{0} & 1 \end{bmatrix} \end{gathered} \tag{8}$$

From this, we can see that $\mathbf{R_{rel}}$ and $\mathbf{t_{rel}}$ are as follows:

$$\begin{gathered} \mathbf{R_{rel}} = \mathbf{R_1^T R_2} \\ \mathbf{t_{rel}} = \mathbf{R_1^T}(\mathbf{t_2} - \mathbf{t_1}) \end{gathered} \tag{9}$$

Finally, we express E and F in terms of $\mathbf{K}$, $\mathbf{R_{rel}}$, and $\mathbf{t_{rel}}$.

$$\begin{gathered} \mathbf{E} = [\mathbf{t_{\times rel}}]\mathbf{R_{rel}} \\ \mathbf{F} = \mathbf{K^{-T} E K} \\ \mathbf{F} = \mathbf{K^{-T}}[\mathbf{t_{\times rel}}]\mathbf{R_{rel} K} \end{gathered} \tag{10}$$

Q1.4

Let $\mathbf{P}$ be the 3D point and $\mathbf{x}$ be the point in the image plane and $\mathbf{P'}$ and $\mathbf{x'}$ be its reflection in the in the mirror and this point in the image plane. Since the object and the mirror are both flat, we know that the rotational matrix is the identity matrix, and there is pure translation.

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

$$\boldsymbol{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{11}$$

$$\mathbf{P'} = \mathbf{P} + \mathbf{t}$$

$$\lambda_1 \mathbf{x} = K\mathbf{P}$$

$$\lambda_2 \mathbf{x'} = K\mathbf{P'}$$

Plugging it in gives the following.

$$\lambda_2 K^{-1}\mathbf{x'} = \lambda_1 K^{-1}\mathbf{x} + \mathbf{t} \tag{12}$$

We take the cross product with $\mathbf{t}$ on both sides to eliminate the + t term.

$$\lambda_2 [\mathbf{t}_\times] K^{-1}\mathbf{x'} = \lambda_1 [\mathbf{t}_\times] K^{-1}\mathbf{x} \tag{13}$$

To eliminate the left side, we take the dot product with $\mathbf{P'}$ and divide out the $\lambda_1\lambda_2$ constants.

$$\mathbf{x'}^T K^{-T}[\mathbf{t}_\times]K^{-1}\mathbf{x} = 0$$

$$\mathbf{t}_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \tag{14}$$

We know $\mathbf{x'}^T\mathbf{F}\mathbf{x} = 0$, so

$$\mathbf{F} = K^{-T}\mathbf{t}_\times K^{-1} \tag{15}$$

Since, $\mathbf{t}_\times$ is skew symmetric, no matter what intrinsic matrix $\mathbf{K}$ we plug in, F will still be a skew semmetric matrix. 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 Eight Point Algorithm

$$F = \begin{bmatrix} 8.59559382\text{e}{-08} & -1.52797237\text{e}{-05} & 2.96376328\text{e}{-01} \\ -3.46886205\text{e}{-05} & 1.78422901\text{e}{-07} & 2.54661757\text{e}{-03} \\ -2.85058631\text{e}{-01} & 2.60516758\text{e}{-03} & -1.16584373\text{e}{+00} \end{bmatrix} \tag{16}$$

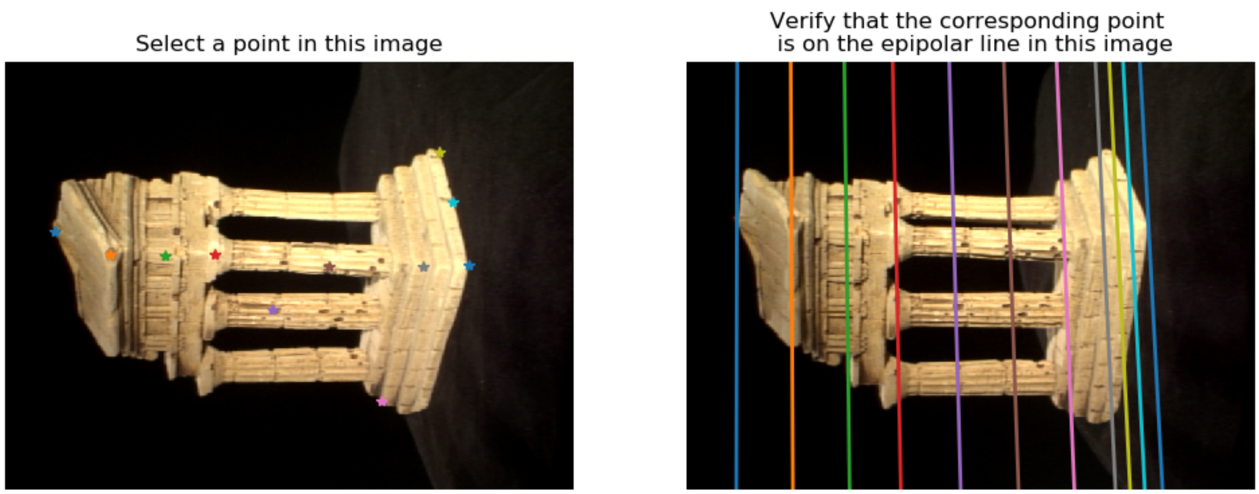

Figure 1: Results from displayEpipolarF

3 Metric Reconstruction

Q3.1 Essential Matrix from Fundamental Matrix

$$E = \begin{bmatrix} 1.98697136\text{e}{-01} & -3.54486283\text{e}{+01} & 4.44914969\text{e}{+02} \\ -8.04768489\text{e}{+01} & 4.15434674\text{e}{-01} & -1.20491151\text{e}{+01} \\ -4.46383700\text{e}{+02} & -3.00625374\text{e}{+00} & -1.83050707\text{e}{-01} \end{bmatrix} \tag{17}$$

Q3.2 Triangulate 3D point from 2D correspondences

Given camera matrix $\mathbf{C}$ and 3D point $\boldsymbol{w_i}$, we get 2D point $\tilde{\boldsymbol{x}}_i$.

$$\tilde{\boldsymbol{x}}_i = C\boldsymbol{w_i}$$

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{18}$$

If we have correspondence of two camera points, we can write the relationship in the following form with A matrix, where C_1^2 means the second row of the first C matrix, and solve for the 3D points using SVD.

$$\boldsymbol{A_i w_i} = \boldsymbol{0}$$

$$\boldsymbol{A_i} = \begin{bmatrix} C_1^1 - x_{i1}C_1^3 \\ C_1^2 - y_{i1}C_1^3 \\ C_2^1 - x_{i2}C_2^3 \\ C_2^2 - y_{i2}C_2^3 \end{bmatrix} \tag{19}$$

Q3.3 findM2

The script findM2.py is implemented.

4 3D Visualization

Q4.1 Epipolar Correspondence

Given two images, the fundamental matrix F, and a point in image 1, we can find the corresponding point in image 2 with the epipolar line and image intensity matching. Here a window size of 30 is used.

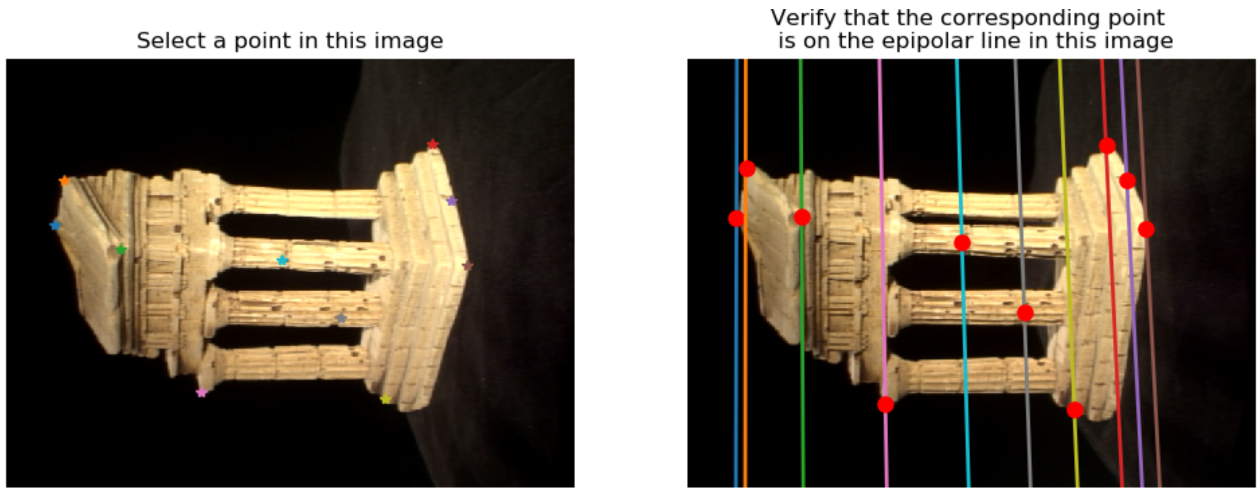

Figure 2: Results from epipolarCorrespondence

Q4.2 3D Visualization

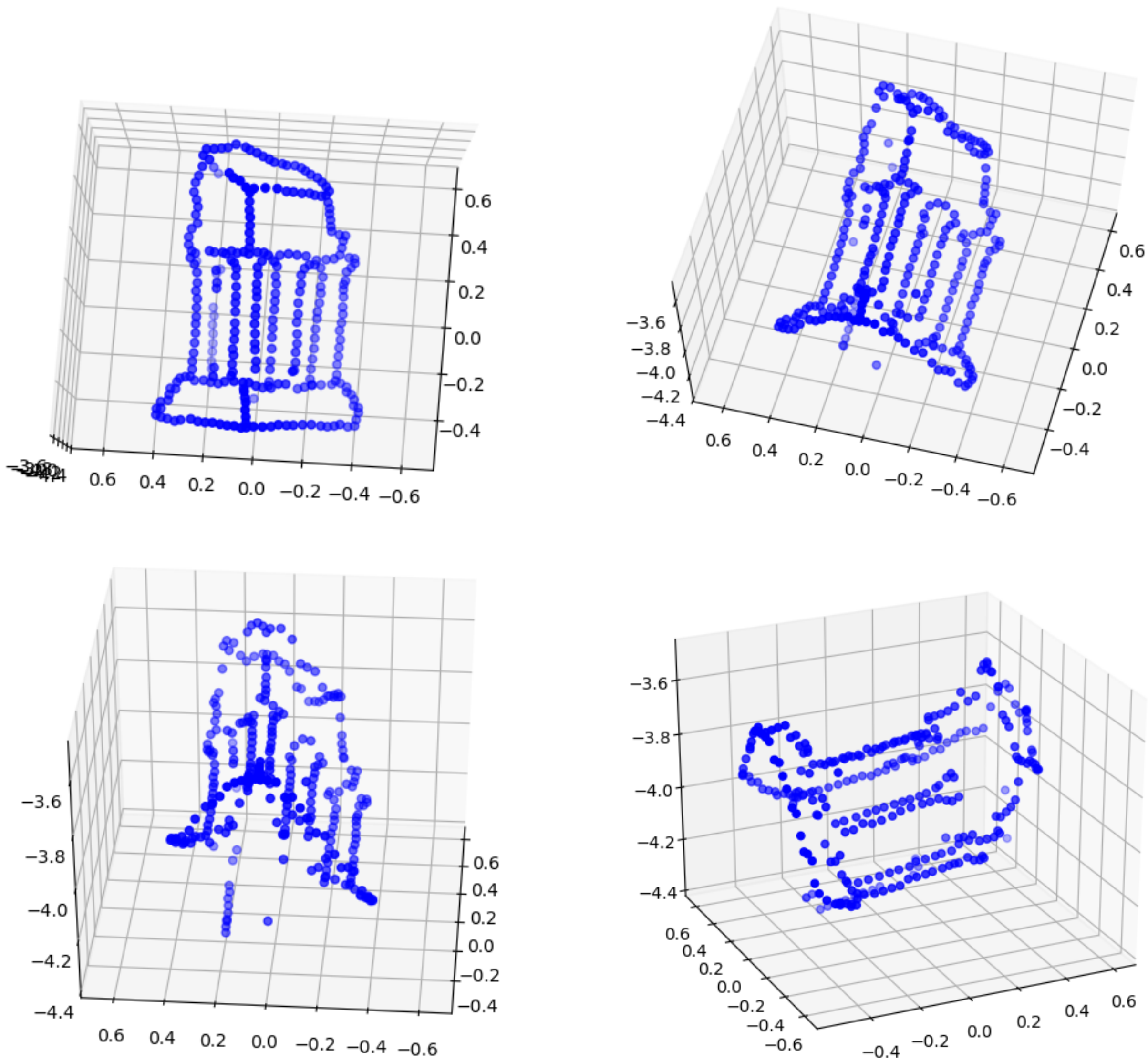

Figure 3: Temple 3D Visualization

5 Bundle Adjustment

Q5.1 RANSAC

Data from som_corresp_noisy.npz has 25% noise in it, which throws the 8-point algorithm from section 2.1 off depending on how drastic the noises are. In contrast, the RANSAC algorithm simply estimates the F based on the number of inliers, disregarding how far off the outliers are.

Given a point in img1 and the estimated Fundamental matrix F, the error metric is the sum of differences (Δx and Δy) between the corresponding point from img2 and the estimated point using the

Fundamental matrix F. If $\Delta x + \Delta y$ is less than the tolerance for that point, it is considered an inlier.

When looking at the results, the 8-point algorithm is thrown off quite a lot, while RANSAC is able to find a good F after having tuned the tolerance and number of iterations. When the tolerance is too high, it thinks every F is equally good, and produces poor results. When it's too low, good results are neglected. The number of iterations is usually better higher, but computational cost increases as a trade-off. A tolerance of 0.3 and nIters of 30 was tested to be decent.

Q5.2 Rodrigues

The two functions are implemented in submission.py.

6 Multi View Keypoint Reconstruction

To compute the 3D locations, the same method as triangulate in section 3.2 is used. Only difference here would be that we take into account the threshold, and use three images instead of two when we can. We create an A matrix where $A \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0$. When all three points pass the threshold, A is 6x4. When only two A is 4x4, and when less than two we skip that point.

Manipulating the threshold value, as long as the threshold is low enough, in this case lower than 350, the results were good. Only when it is too high do we get problematic reconstructions. The reprojection error when the result is good is around 279.35.

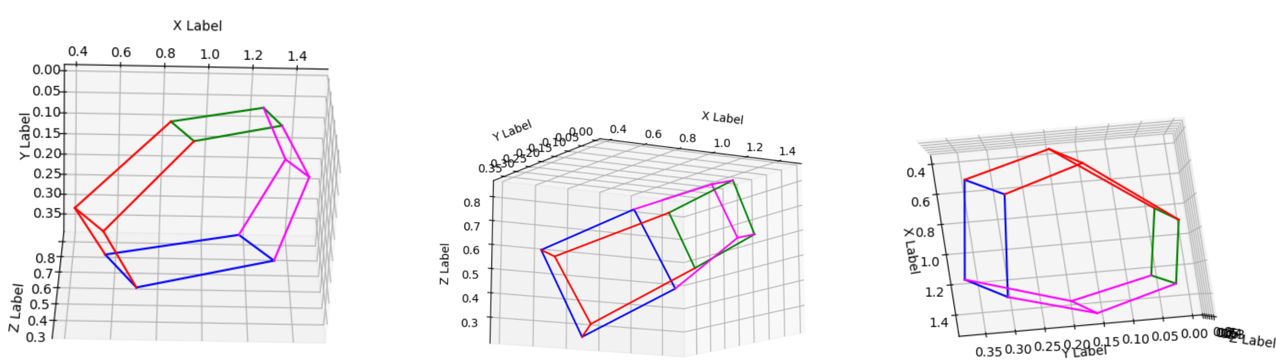

Figure 4: Car Keypoint Reconstruction