COMP5421 Computer Vision

Homework Assignment 2
3D Reconstruction

Hartanto Kwee Jeffrey
jhk@connect.ust.hk — SID: 20851871

1 Theory

Conventions

In this solution, we will define the fundamental matrix as

$$x^{'T} F x = x^{'T} K^{'-T} \left(R \left[t_\times\right]\right) K^{-1} x = 0$$

where the camera coordinates systems are related by

$$x' = R(x - t)$$

and the epipolar lines are $l' = Fx$ nd $l = F^T x'$ respectively.
Our convention: If we define $x' = R(x - t)$, then

$$\begin{aligned} x' &= R\left(x - t\right) \\ R^T x' &= x - t \\ t \times R^T x' &= x \times t - t \times t \\ \left[t_\times\right] R^T x' &= x \times t \\ x^T \left[t_\times\right] R^T x' &= x^T \left(x \times t\right) = 0 \\ \left(x^T \left[t_\times\right] R^T x'\right)^T &= 0 \\ x^{'T} R \left[t_\times\right] x &= 0 \end{aligned}$$

Alternative conventions: defining $x' = Rx + t$ would give

$$\begin{aligned} x' &= Rx + t \\ t \times x' &= t \times Rx + t \times t \\ x^{'T} \left(t \times x'\right) &= x^{'T} \left[t_\times\right] Rx \\ x^{'T} \left[t_\times\right] Rx &= 0 \end{aligned}$$

Hence, we should be extra cautious about the definition.

Q1.1

The epipolar constraint suggests that

$$x^{'T} F x = 0$$

Since $x' = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ and $x = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ satisfy this equation, we have

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

Q1.2

Note that $E = R\left[t_\times\right]$ is the essential matrix. With a pure translation in the x-axis, we have $R = I_{3\times 3}$, $t = \begin{bmatrix} t_x & 0 & 0 \end{bmatrix}^T$ and $t_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix}$. Hence $E = [t_\times]$, and the epipolar line for the first camera is

$$l = Ex' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -t_x \\ t_x v' \end{bmatrix}$$

is a horizontal line since any point x on it satisfies $l^T x = 0$, and since $l^T x = \begin{bmatrix} 0 & -t_x & t_x v' \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$, we have

$$v = v'$$

which is the equation of a horizontal line (the vertical coordinate v is equal to a constant v'), i.e. parallel to the x-axis. Similar calculations show that $l' = E^T x$ is also a horizontal line.

Q1.3

Suppose the next timestamp following i is j. The relationship between camera and world coordinates in both frames are $X_{ci} = R_i X_W + t_i$ and $X_{cj} = R_j X_W + t_j$. We invert the latter to get $X_W = R_j^{-1}(X_{cj} - t_j)$, which we substitute into the former to get

$$X_{ci} = R_i R_j^{-1}\left(X_{cj} - t_j\right) + t_i = R_i R_j^{-1} X_{cj} + \left(t_i - R_i R_j^{-1} t_j\right)$$

Inverting gives

$$X_{cj} = R_j R_i^{-1}\left[X_{ci} - \left(t_i - R_i R_j^{-1} t_j\right)\right]$$

Remember that our convention is

$$X_{cj} = R_{rel}\left(X_{ci} - t_{rel}\right)$$

Hence, $R_{rel} = R_j R_i^{-1}$ and $t_{rel} = t_i - R_i R_j^{-1} t_j$, and the fundamental matrix is

$$F = K^{-T} R_{rel}\left[t_{rel\times}\right] K^{-1}$$

Q1.4

Denote p_r and p_i to be the real and imaginary image of the object as viewed by our camera. Set the world coordinate system at the mirror plane. Then, the world coordinates of the objects can be let as $w_r = a + b$ and $w_i = a - b$, where a is a vector lying on the mirror plane, and b is a vector perpendicular to the mirror plane. Note that a is variable and b is constant as the object is “flat”. If the camera is offset by rotation R and translation t, then the camera coordinates of the objects are $c_r = R\left(a+b\right) + t = Ra + (t + Rb)$ and $c_i = R\left(a-b\right) + t = Ra + (t - Rb)$ and the image coordinates are $p_r = Kc_r$ and $p_i = Kc_i$. Hence, this set up can also be viewed as having two cameras with the same rotation matrix R and intrinsic matrix K but different translations $t_r = t + Rb$ and $t_i = t - Rb$ viewing the same object. Applying this to the result of Q1.3, we have $R_{rel} = RR^{-1} = I$ and $t_{rel} = t + Rb - (t - Rb) = 2Rb$. The fundamental matrix of this setup is

$$F = K^{-T} R_{rel}\left[t_{rel\times}\right] K^{-1} = K^{-T}\left[t_{rel\times}\right] K^{-1}$$

Noting that $[t_\times] = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$ is skew symmetric since $[t_\times]^T = -[t_\times]$, we see that

$$F^T = \left(K^{-T}\left[t_{rel\times}\right] K^{-1}\right)^T = K^{-T}\left[t_{rel\times}\right]^T K^{-1} = -K^{-T}\left[t_{rel\times}\right] K^{-1} = -F$$

Hence, the fundamental matrix is skew-symmetric.

2 Fundamental matrix estimation

Q2.1

Recovered F:

```
[[-8.33149234e-09  1.29538462e-07 -1.17187851e-03]
 [ 6.51358336e-08  5.70670059e-09 -4.13435037e-05]
 [ 1.13078765e-03  1.91823637e-05  4.16862079e-03]]
```

Sample output:

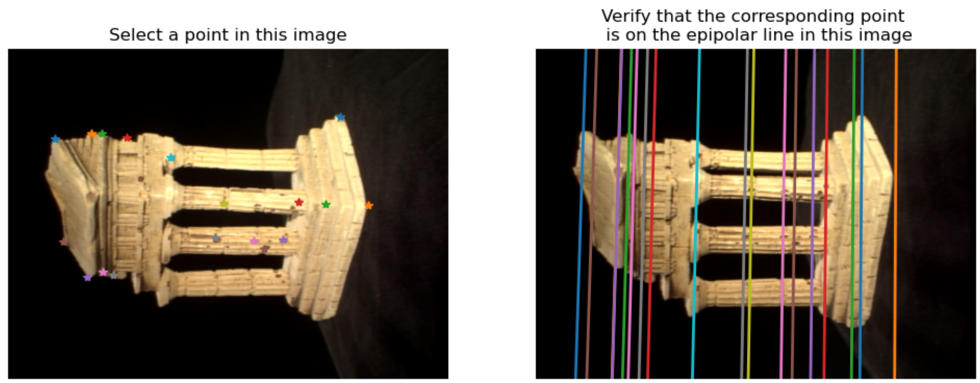


Figure 1: The Eight Point Algorithm

Q2.2

Selected points indices: `[50, 55, 25, 22, 61, 47, 70]`.
Recovered F:

```
[[-6.97773335e-08 -1.27649875e-07  1.13655676e-03]
 [-5.99367719e-08 -1.87605070e-08  1.14829050e-04]
 [-1.03106995e-03 -1.13298037e-04 -9.66454193e-03]]
```

Sample images:

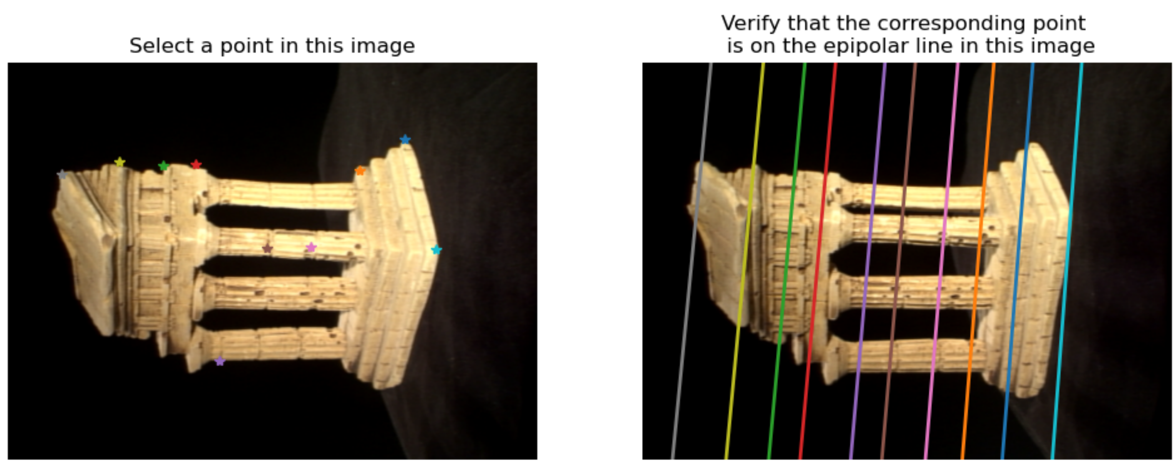


Figure 2: The Seven Point Algorithm using hand-selected points

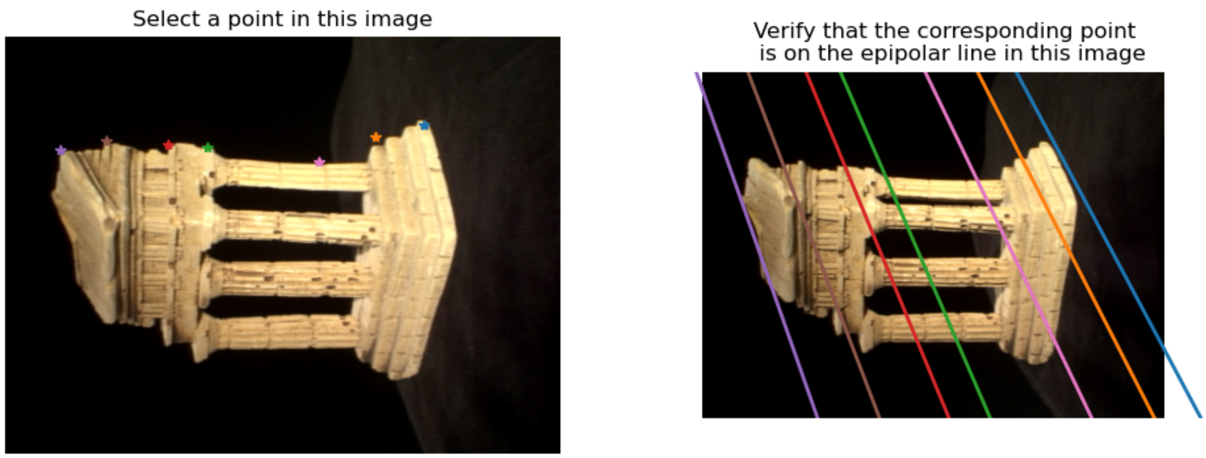


Figure 3: Failed example output 1

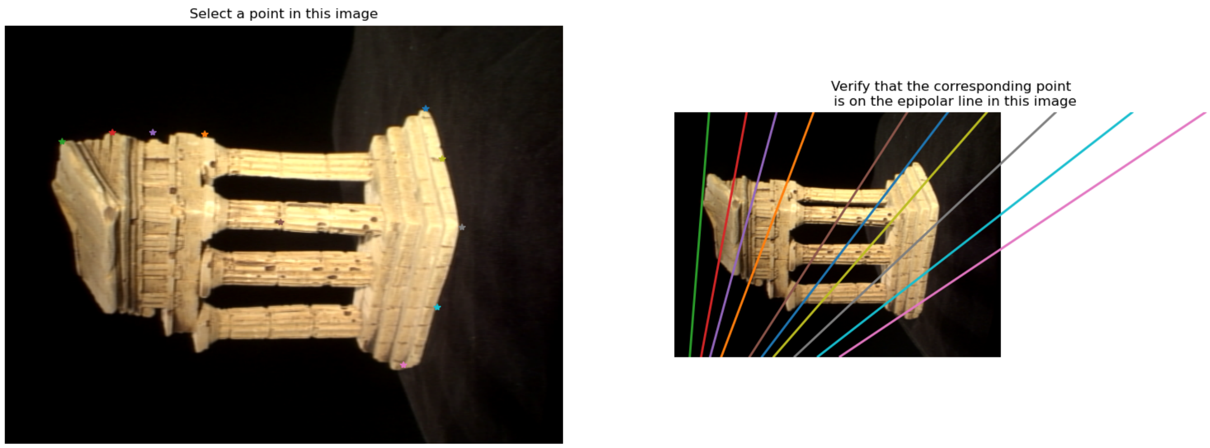


Figure 4: Failed example output 2

3 Metric Reconstruction

Q3.1

By our conventions,

$$F = K_2^{-T} E K_1^{-1}$$

Hence

$$E = K_2^T F K_1$$

Evaluated E:

```
[[-1.92592123e-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

Note: we omit the index i in the following derivation. The relationship between the camera matrix C_1, 3D object point P and 2D image point $\widetilde{x_1}$ is

$$\lambda_1 \widetilde{x_1} = C_1 P$$

Then, we have

$$\lambda_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} -C_{11}^T- \\ -C_{12}^T- \\ -C_{13}^T- \end{bmatrix} P$$

where C_{1i} is a column vector containing the elements of the ith row of C_1. We obtain three linear equations from this

$$\begin{aligned} \lambda_1 u_1 &= C_{11}^T P \\ \lambda_1 v_1 &= C_{12}^T P \\ \lambda_1 &= C_{13}^T P \end{aligned}$$

Substituting the third equality into the first two, we have

$$\begin{aligned} C_{11}^T P - C_{13}^T P u_1 = 0 \\ C_{12}^T P - C_{13}^T P v_1 = 0 \end{aligned}$$

In matrix form,

$$\begin{bmatrix} C_{11}^T - u_1 C_{13}^T \\ C_{12}^T - v_1 C_{13}^T \end{bmatrix} P = 0$$

We can get a similar matrix from the second camera:

$$\begin{bmatrix} C_{21}^T - u_2 C_{23}^T \\ C_{22}^T - v_2 C_{23}^T \end{bmatrix} P = 0$$

Therefore, the homogenous system we want to find is $AP = 0$, where

$$A = \begin{bmatrix} C_{11}^T - u_1 C_{13}^T \\ C_{12}^T - v_1 C_{13}^T \\ C_{21}^T - u_2 C_{23}^T \\ C_{22}^T - v_2 C_{23}^T \end{bmatrix}$$

Q3.3

The script is written in `findM2.py`. After triangulating with all four possible `M2`s, we selected the one with the most point that satisfy the cheirality condition. The reprojection of the final `M2` is 94.158436.

4 3D Visualization

Q4.1

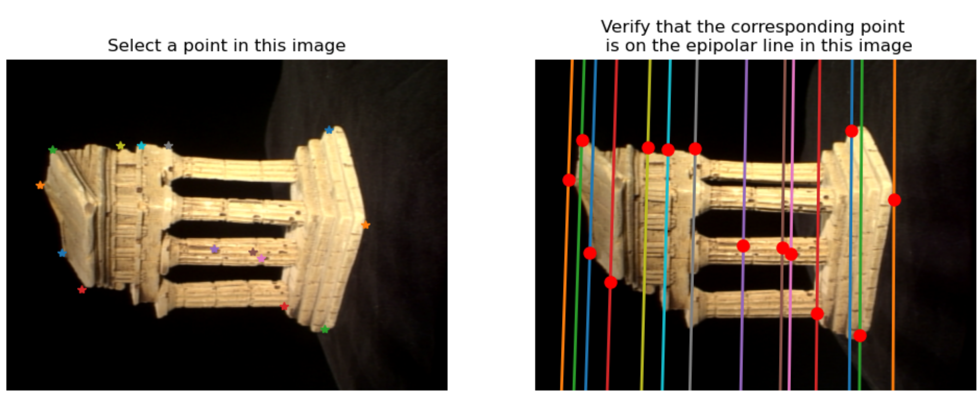


Figure 5: Epipolar Correspondences

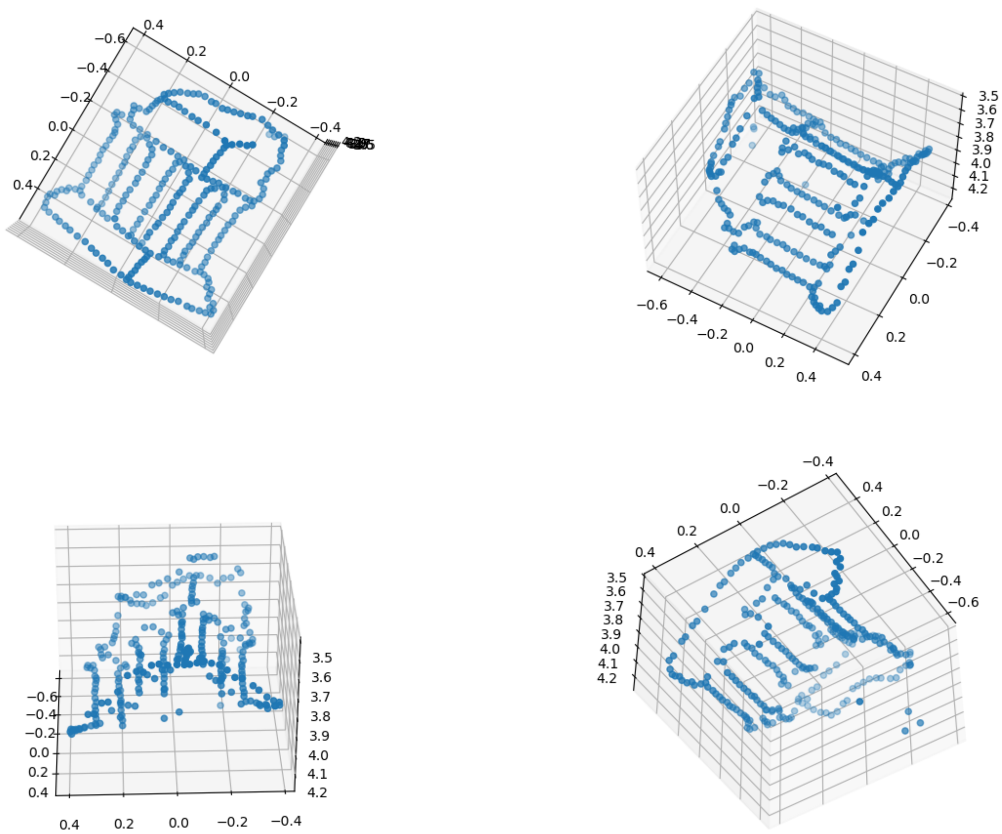


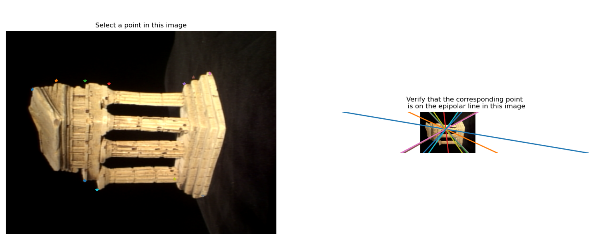
Figure 6: 3D Visualization

Q4.2

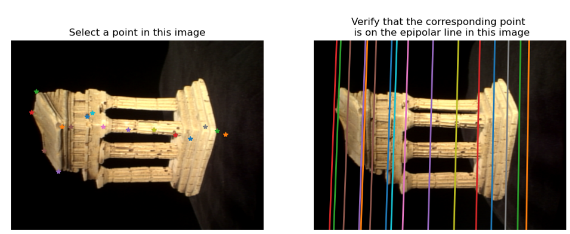
5 Bundle Adjustment

Q5.1

Comparing the epipolar lines generated from the estimated fundamental matrix from `eightpoint` and `ransacF`, it is obvious that noisy correspondences will throw off the estimation greatly, and RANSAC can prevent that by selecting correct correspondences.



(a)



(b)

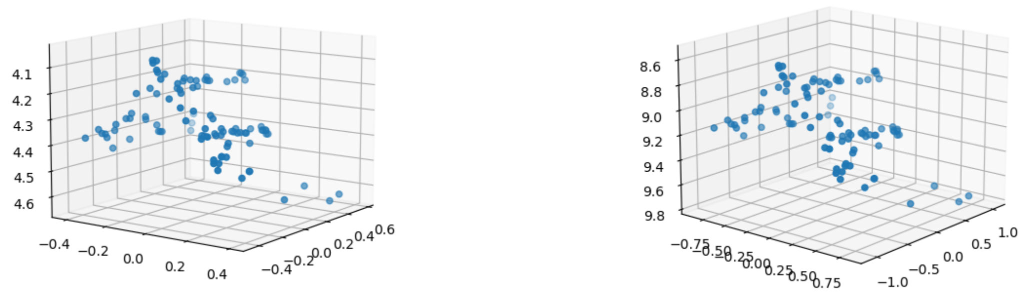
Figure 7: Comparison of fundamental matrix estimation performance by (a) `eightpoint` and (b) `ransacF`.

The error metric used is err $= |x_2^T F x_1|$, and a point is an inlier if err is less than a certain threshold. Since refining the estimated fundamental matrix in `sevenpoint` takes some time, `multiprocessing` is used to speed up the process.

Q5.3

The initial projection error is 104056.861470, and the projection error after bundle adjustment is 9.747948. Refer to the images for the 3D points.

Bundle adjustment was also run on `templeCoords.npz`. Point correspondences from `visualize.py` were optimized, and the reprojection error decreased from 203.3929 to 13.264073 after bundle adjustment.



(a) without bundle adjustment (b) with bundle adjustment

Figure 8: Comparison of noisy correspondences triangulation with and without bundle adjustment.

(a) without bundle adjustment (b) with bundle adjustment

Figure 9: Comparison of `templeCoords.npz` triangulation with and without bundle adjustment.