# 16-720 B Assignment4

Sajal Maheshwari
sajalm@andrew.cmu.edu

November 4, 2019

## 1 Part 1 - Theory

### 1.1

The equation of the Fundamental matrix ($F$) w.r.t to corresponding points $\mathbf{x}^1$ and $\mathbf{x}^2$ are :

$$\begin{bmatrix} x_1^2 & x_2^2 & 1 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} x_1^1 \\ x_2^1 \\ 1 \end{bmatrix} = 0 \tag{1}$$

Substituting $\mathbf{x}^1$ as $[0, 0, 1]^T$ and $\mathbf{x}^2$ also as $[0, 0, 1]^T$ in Equation 1 we get $F_{33} = 0$.

### 1.2

Assuming the intrinsic matrices to be Identity matrices(Post), the equation of the Fundamental matrix as in Equation 1 becomes :

$$\begin{bmatrix} x_1^2 & x_2^2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathcal{T}_x \\ 0 & \mathcal{T}_x & 0 \end{bmatrix} \begin{bmatrix} x_1^1 \\ x_2^1 \\ 1 \end{bmatrix} = 0 \tag{2}$$

Now the epipolar line for a point $\mathbf{x}$ is $\mathbf{l}$, where the relation between the line and the point is given by $\mathbf{l}^T\mathbf{x} = 0$. This is similar to the equation of the fundamental matrix in Equations 1 and 2. The epipolar line is therefore given by Equation: 3, which reduces to $[0, 1, -x_2^2]^T = 0$, which is a line parallel to the x-axis.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathcal{T}_x \\ 0 & -\mathcal{T}_x & 0 \end{bmatrix} \begin{bmatrix} x_1^1 \\ x_2^1 \\ 1 \end{bmatrix} = 0 \tag{3}$$

### 1.3

Suppose we have a reference frame $F_0$, and the two other reference frames $F_i$ and $F_j$ w.r.t. the orientation and position of the camera at times i and j.

Now, a point $P_0$ in $F_0$ becomes $P_i$ in $F_i$ and $P_j$ in $F_j$. So, the relation between $P_i$, $P_j$ and $P_0$ is given by the equations 5:

$$\begin{aligned} P_i &= R_i P_0 + t_i \\ P_j &= R_j P_0 + t_j \end{aligned} \tag{4}$$

These equations can be written in homogeneous coordinates as :

$$\begin{aligned} P_i &= S_i P_0 \\ P_j &= S_j P_0 \end{aligned} \tag{5}$$

where $S = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$. Now, we calculate the value of $P_0$ using the first equation and substitute it in the second equation to get $P_j = S_j S_i^{-1} P_i$, which is representation of $P_j$ in terms of $P_i$. We find $S_i^{-1}$ which is equal to $S = \begin{bmatrix} R_i^{-1} & R_i^{-1} t_i \\ 0 & 1 \end{bmatrix}$. Therefore, $S_{rel} = S_j S_i^{-1}$ is equal to $\begin{bmatrix} R_j R_i^{-1} & t_j - R_j R_i^{-1} t_i \\ 0 & 1 \end{bmatrix}$, which makes $R_{rel} = R_j R_i^{-1}$ and $t_{rel} = t_j - R_j R_i^{-1} t_i$.

The essential matrix therefore becomes $E = t_{rel} \times R_{rel}$ and the Fundamental matrix becomes $F = (K_2^{-1})^T E K_1^{-1}$, where $\times$ denotes the cross-product.

### 1.4

Let a 3D point on the object be denoted by $X$. Now since the object is such that all its points are parallel to the reflection plane, the reflected point $X$ corresponding to $X$ can be written as $X' = X + t$, where $t$ is the translation vector.

Let the image of these two points in homogenous corordinates for a camera with projection matrix C be $x$ and $x'$. Then $x_h = CX_h$ and $x'_h = CX'_h$, where the subscript $h$ denotes the point in homogeneous coordinates. Now substituting the value of $X_h$ in $X'_h$, we get $X'_h = \begin{bmatrix} I & t \\ 0 & 1 \end{bmatrix} X_h$.

Substituting this in the projection equation $x'_h = CX'_h$, we get, $x'_h = C \begin{bmatrix} I & t \\ 0 & 1 \end{bmatrix} X_h$

This is equal to having another camera with projection matrix as $C' = C \begin{bmatrix} I & t \\ 0 & 1 \end{bmatrix}$. This implies that the two images can be thought of as those from two cameras with the projection matrices $C$ and $C'$. As can be seen from the relation, $C$ and $C'$ are related by just translation with rotation being equal to Identity matrix. Now the matrix $\mathcal{T}_x$ corresponding to the translation $t$ is skew-symmetric by nature. Since the rotation is identity, this makes the Fundamental matrix as skew symmetric (the camera intrinsics are equal, as both $C$ and $C'$ in reality are the same cameras.)

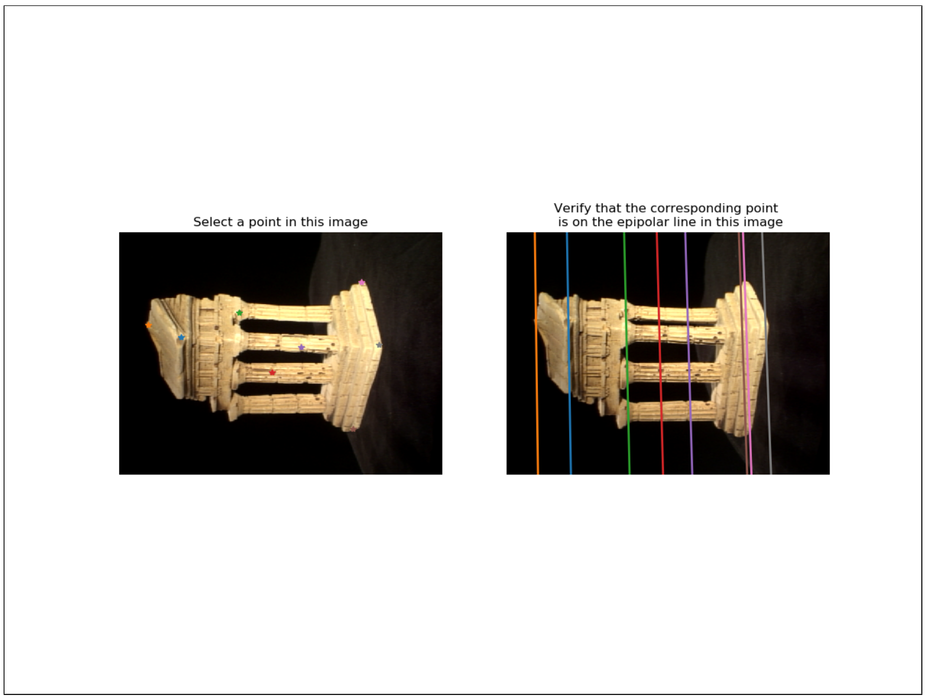
## 2 Part 2 - Practice

### 2.1 Eight Point Algorithm

The recovered value of F is :

```
array([[ 9.80213860e-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]])
```

An example output is :

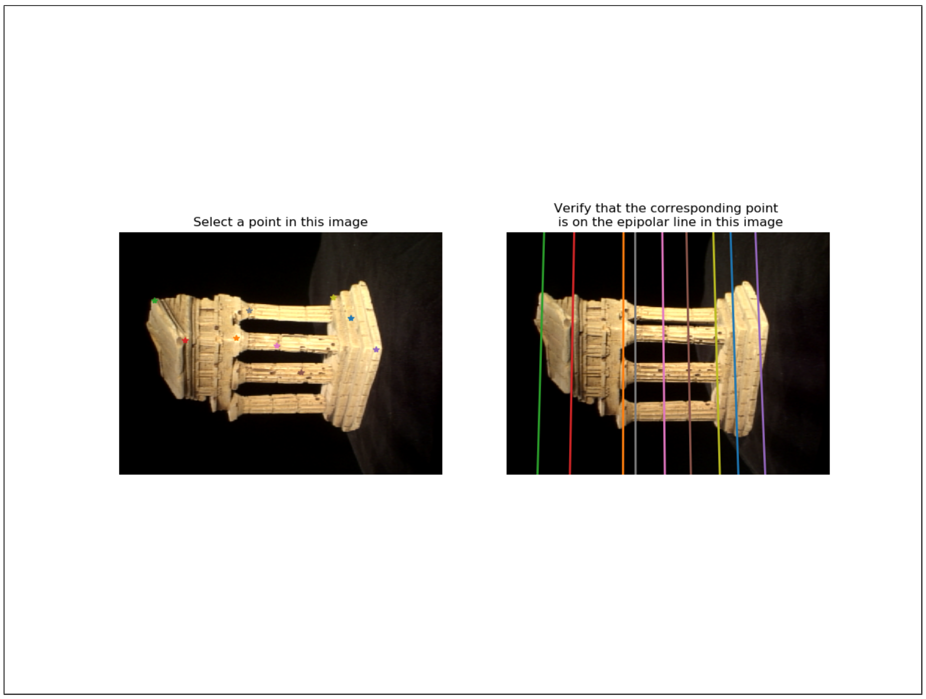


## 2.2 Seven Point Algorithm

The recovered value of F is :

```
array([[ 2.64586668e-05, -4.48737227e-05,  3.81024483e-01],
       [-5.83889067e-05, -1.59194479e-06,  1.45551623e-02],
       [-3.67753104e-01,  1.04001113e-03, -2.40789040e+00]])
```

An example output is :



## 3 Metric reconstruction

### 3.1

The essential matrix comes out to be :

```
[[ 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]]
```

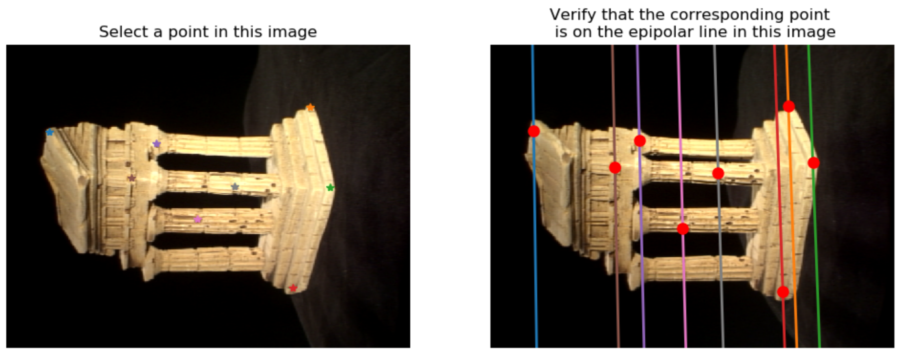
### 3.2

If we are given corresponding points $[x, y]^T$ and $[x', y']^T$ for two cameras for a 3D point, the 3D point can be found in the homogenous coordinates using the SVD of a matrix A denoted as : $\begin{bmatrix} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{bmatrix}$ which can be denoted in a compact form as : $\begin{bmatrix} - & \boldsymbol{p}_1^\top - \\ - & \boldsymbol{p}_2^\top - \\ - & \boldsymbol{p}_3^\top - \end{bmatrix}$

The expression for A in terms of these notations is : $\begin{bmatrix} y\boldsymbol{p}_3^\top - \boldsymbol{p}_2^\top \\ \boldsymbol{p}_1^\top - x\boldsymbol{p}_3^\top \\ y'\boldsymbol{p}_3'^\top - \boldsymbol{p}_2'^\top \\ \boldsymbol{p}_1'^\top - x'\boldsymbol{p}_3'^\top \end{bmatrix}$

# 4 3D Visualization

## 4.1



## 4.2

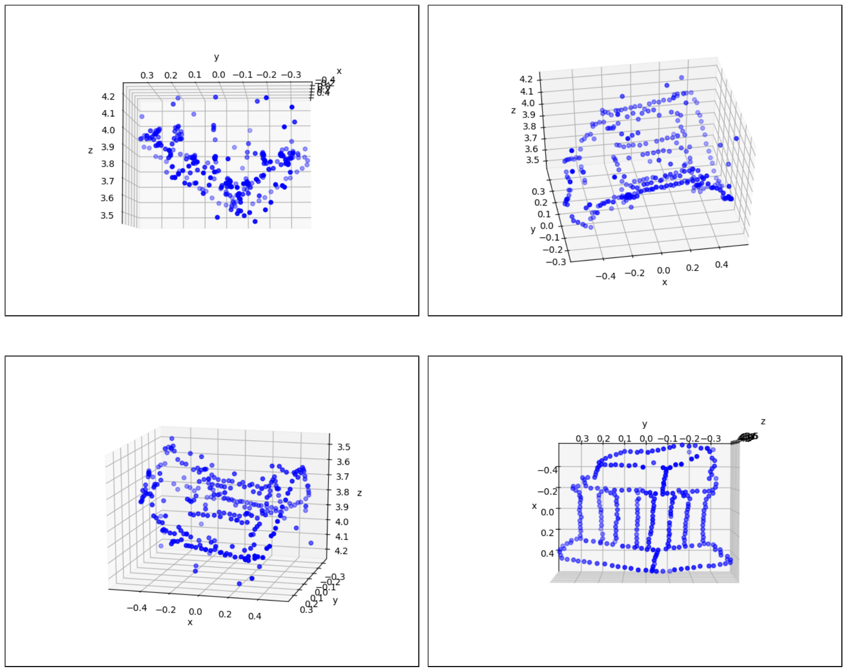


Figure 1: This figure shows the 3d point cloud reconstructed for the given temple coordinates from multiple angles.

# 5 Bundle Adjustment

## 5.1

The results of the RANSAC almost accurately generate the Fundamental matrix on noisy points, as RANSAC helps us to get the inlier points. In contrast, the epipolar lines for the noisy correspondences generate results that are clearly inaccurate.

The error metric used to get the inlier points is as follows :
For some randomly selected points, the seven point algorithm is run which generates the Fundamental matrix. This fundamental matrix($F$) is used to get the absolute value of the expression $(\mathbf{x}^2)^T F \mathbf{x}^1$, which should be less than a set threshold (set to 0.001 empirically). The number of pairs satisfying this condition are chosen to be the inlier pairs.

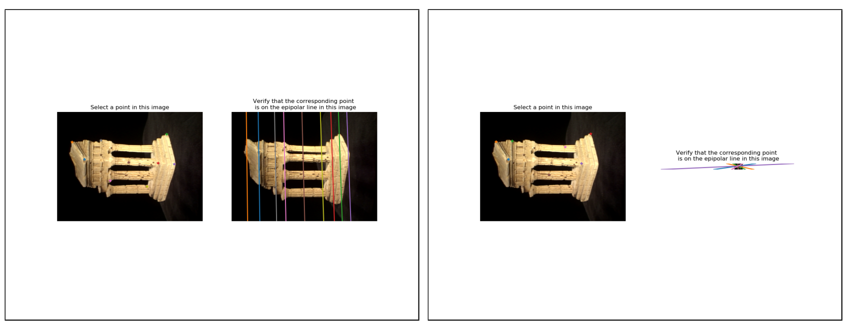



Figure 2: The figure shows the epipolar lines generated on noisy correspondances with and without inlier points generated using RANSAC

### 5.2

### 5.3

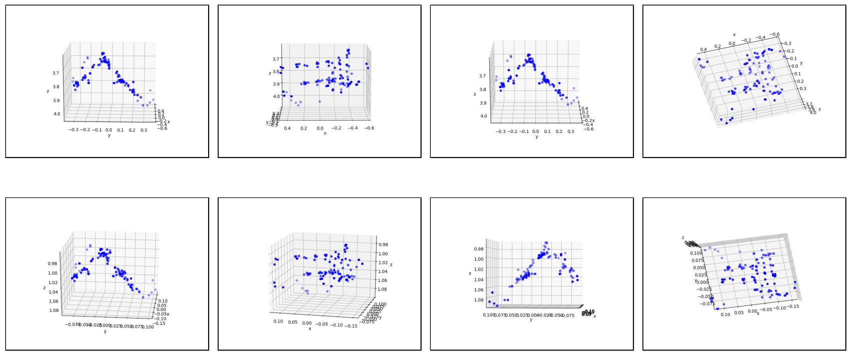


Figure 3: This figure shows the 3d point cloud reconstructed for the noisy correspondences. Although bundle adjustment reduces the error significantly, there is not much of a difference to be seen visually.

The initial extrinsic matrix was equal to :

```
[[ 0.9994253   0.03318864  0.00689856 -0.02526612]
 [-0.03384679  0.9658456   0.25689816 -1.        ]
 [ 0.00186316 -0.25698401  0.96641386  0.08568071]]
```

which gave an error of : 66.86972844043878.

The final extrinsic matrix after bundle adjustment was equal to :

```
[[ 1.00000000e+00  6.78210399e-02  3.68312861e-02 -6.74311240e-03]
 [ 9.68531835e-04  1.00000000e+00  2.92238381e-01 -2.69511708e-01]
 [ 3.19582856e-02 -2.23448809e-01  1.00000000e+00  3.25619865e-02]]
```

which gave an error of : 9.512796718911861