Computer Vision
HW4 Report

Team 3， 0751231 曾揚 , 309505018 郭俊廷

Introduction:

Structure from motion(SfM) is a technique which utilizes a series of 2-dimensional images to reconstruct the 3-dimensional structure of a scene or object. We will show our result by using SfM to reconstruct the 3-dimensional structure both in TA's data and our data. Finally, we will discuss the merits and demerits of using SfM.

Implementation procedure:

1) Find out correspondence across images:

   We can use the SIFT algorithm to find the important keypoint and descriptor of two images. The reason why can't use camera calibration to find intrinsic parameters is that it needs four images at least. The advantages of the SIFT algorithm are invariant to rotation, translation, and scale.

   After getting the keypoint of two images, we have to match their corresponding keypoint. For each keypoint in the first image, we will iterate all keypoint in the second image, and find two best matching keypoint.The best keypoint means that it has minimum L2 norm distance. After matching keypoint between two images, we will discard some points which $ratio \times distance$ is too large in order to eliminate ambiguous matching.

$$x = K[R|t]X \qquad \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

   At least 8 keypoint of two images: $x_1 \Leftrightarrow x_2$

2) Estimate the fundamental matrix across images (normalized 8 points):

   We use the keypoints of two images to find the fundamental matrix by 8-points algorithm with RANSAC. First, we randomly choose 8 points from key points, and the centroid of the points are at the origin. After getting normalize points we need to get an A matrix by 2 sets of 8 points from normalize keypoints of two images. Since we need to find the solution of equation $AF = 0$, we use SVD to decompose the A matrix to U, Σ, V matrix.Now we use F and keypoints to select indices with accepted points, and Sampson distance as error.

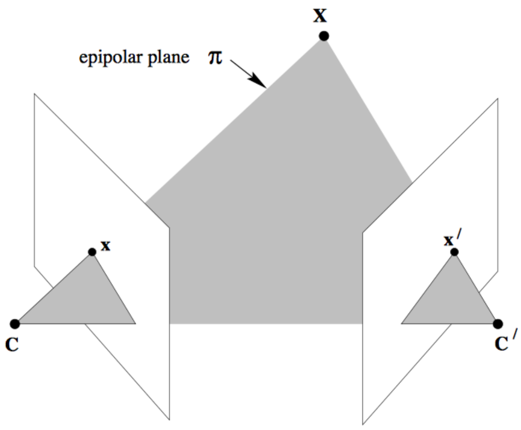
$$(x')^T F x = 0 \qquad \begin{bmatrix} x_1' & y_1' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0$$

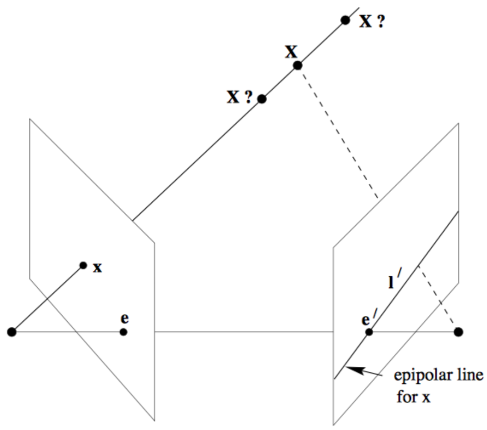
Direct Linear Transformation:

$$AF = 0 \qquad \begin{bmatrix} x_1x_1' & x_1y_1' & x_1 & y_1x_1' & y_1y_1' & y_1 & x_1' & y_1' & 1 \\ x_2x_2' & x_2y_2' & x_2 & y_2x_2' & y_2y_2' & y_2 & x_2' & y_2' & 1 \\ x_3x_3' & x_3y_3' & x_3 & y_3x_3' & y_3y_3' & y_3 & x_3' & y_3' & 1 \\ x_4x_4' & x_4y_4' & x_4 & y_4x_4' & y_4y_4' & y_4 & x_4' & y_4' & 1 \\ x_5x_5' & x_5y_5' & x_5 & y_5x_5' & y_5y_5' & y_5 & x_5' & y_5' & 1 \\ x_6x_6' & x_6y_6' & x_6 & y_6x_6' & y_6y_6' & y_6 & x_6' & y_6' & 1 \\ x_7x_7' & x_7y_7' & x_7 & y_7x_7' & y_7y_7' & y_7 & x_7' & y_7' & 1 \\ x_8x_8' & x_8y_8' & x_8 & y_8x_8' & y_8y_8' & y_8 & x_8' & y_8' & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

Decompose A by using SVD: $A = U\Sigma V^T$
each column of V represents a solution $AF = 0$

3) Draw the interest points on your found in the first image and the corresponding epipolar lines in another.





4) Get four possible solutions of essential matrix from fundamental matrix:
We can simply use the fundamental matrix to get the essential matrix by multiplying the intrinsic matrix of camera 1, camera 2,and fundamental matrix. Then we need to find the 4 possible answers of the essential matrix. We decompose the essential matrix to U, D, V, to get 4 possible answers.

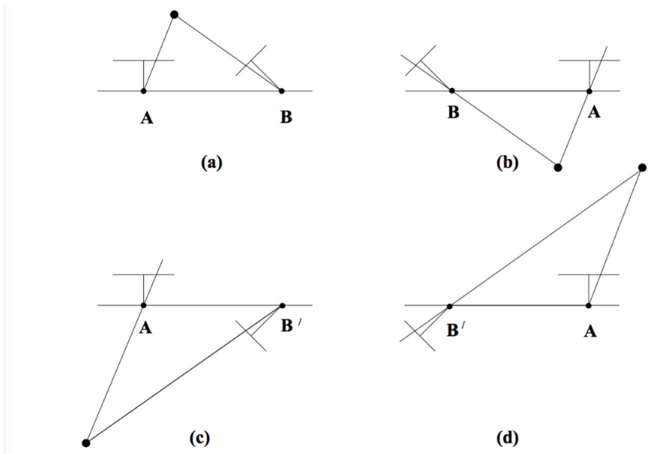
Camera 2 posed: $(C_1, R_1), (C_2, R_2), (C_3, R_3), (C_4, R_4)$

Essential matrix: $E = K_1^T F K_2 = UDV^T \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Four possible answers of camera 2:

1. $C_1 = U(:,3)$ and $R_1 = UWV^T$
2. $C_2 = -U(:,3)$ and $R_1 = UWV^T$
3. $C_3 = U(:,3)$ and $R_1 = UW^TV^T$
4. $C_4 = -U(:,3)$ and $R_1 = UW^TV^T$

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



5) Find out the most appropriate solution of essential matrix:

We have four possible solutions now, and we need to find the most appropriate answer from them. So, we try to use each possible solution to get 3D points triangulation, and pick the solution with most of the 3D points in front of the camera.The following image shows the method of finding the A matrix by using SVD method. We can decompose the matrix to the U, S, V matrix, and the least column of the V matrix is the answer that we want. (P.128)

$$x = w \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = PX = \begin{bmatrix} P_1^T X \\ P_2^T X \\ P_3^T X \end{bmatrix} \qquad x = PX \qquad x' = P'X$$

Because $x \times PX = 0$, then

$$u(P_3^T X) - (P_1^T X) = 0$$
$$v(P_3^T X) - (P_2^T X) = 0$$
$$(P_3^T X) - (P_3^T X) = 0$$

$$A = \begin{bmatrix} uP_3^T - P_1^T \\ vP_3^T - P_2^T \\ u'P_3'^T - P_1'^T \\ v'P_3' - P_2'^T \end{bmatrix} \qquad AX = 0$$

Decompose A by using SVD: $A = USV^T$

each column of V represents a solution $AX = 0$

6) Use texture mapping to get a 3D model:

    The solutions of matrix R, t aren't precise. Then we can use Bundle Adjustment algorithm to solve R,t. Bundle Adjustment algorithm is a non-linear optimization method, and the purpose is to minimize the reconstruction error. By adjusting the POSE and 3D points in order to minimize the back projection error.

Bundle Adjustment:

$$min \sum_i \sum_j (x_i^j - K[R_i|t_i]X^j)^2$$

Finally, we use the 3D points in front of the camera and match the texture of each point.

Experimental result :

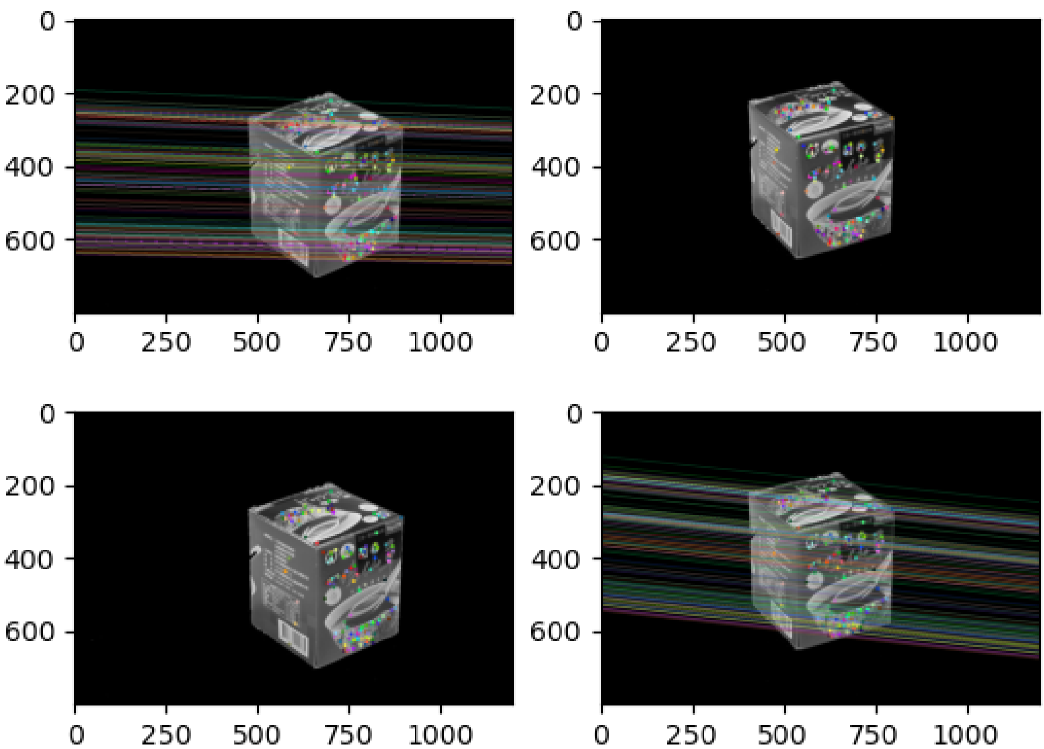
Mesona1.jpg Mesona2.jpg

Correspondence Points: 463
Inliers: 157
Fundamental matrix:

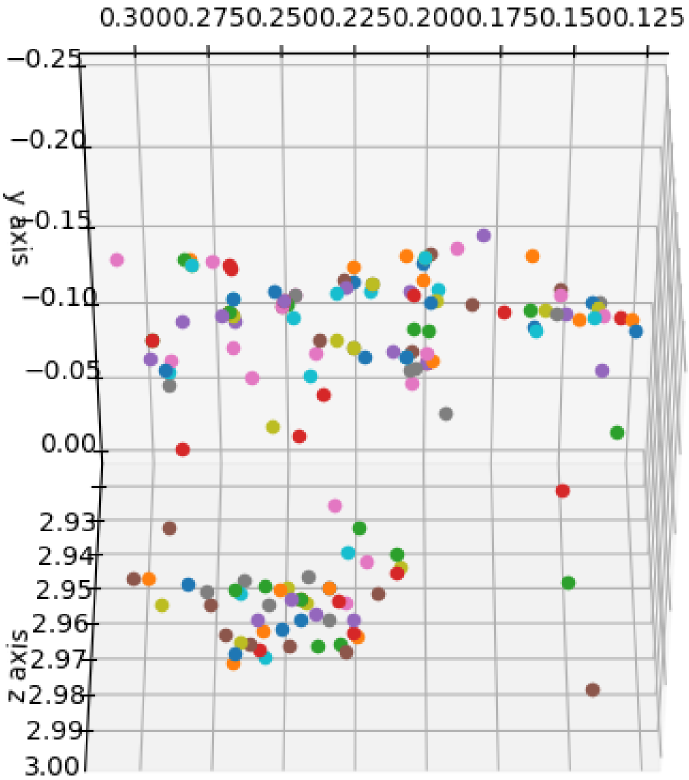
```
[[ 9.58368538e-08 -8.27728149e-07  8.46250446e-04]
 [-2.02796005e-07 -8.33010938e-08 -1.72920484e-02]
 [-1.89585775e-03  1.84741063e-02  1.00000000e+00]]
```

Essential matrix:

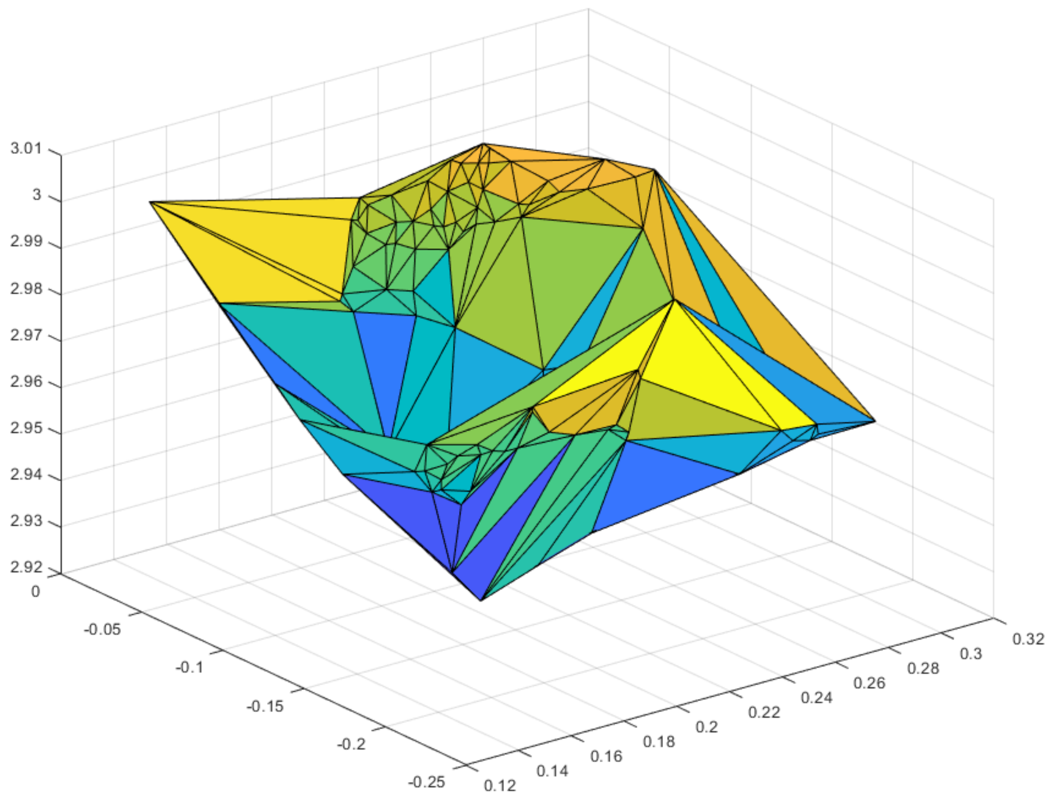
```
[[  2.82216805 -24.35885548   2.0060857 ]
 [ -5.96773013  -2.45370556 -94.48328978]
 [-10.8304876   98.4112585    0.54664437]]
```



Mesona 3D points



Mesona 3D Visualize using Matlab



Statue1.bmp

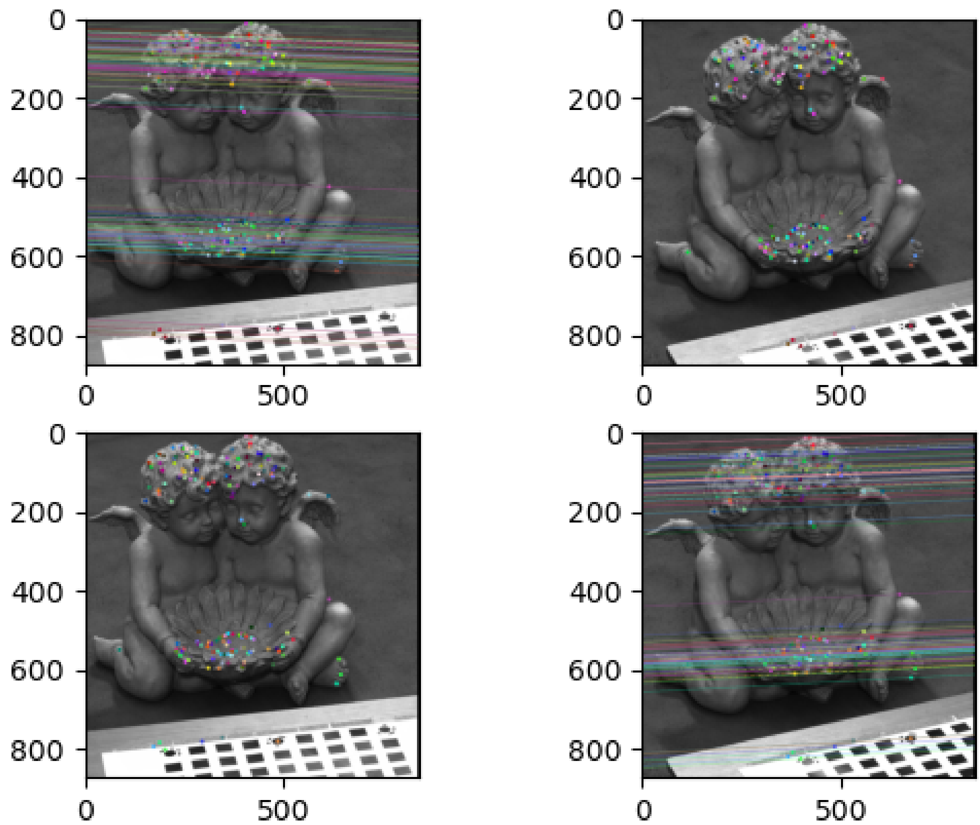
Statue2.bmp

Correspondence Points: 434
Inliers: 177
Fundamental matrix:

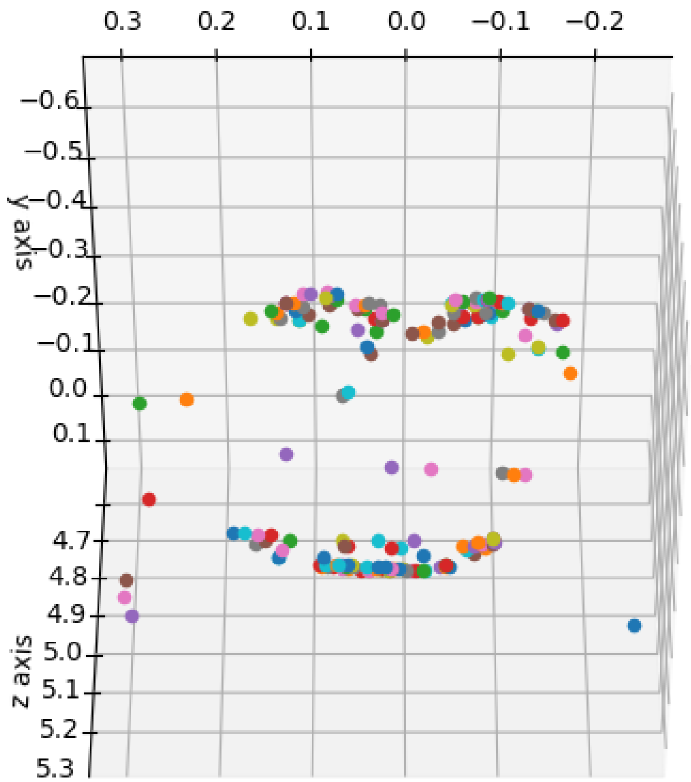
```
[[ 1.34633484e-08 -3.26839161e-07 -1.08993947e-03]
 [-6.56950964e-07  1.75708284e-07  2.81798273e-02]
 [-1.05558303e-03 -2.76486573e-02  1.00000000e+00]]
```

Essential matrix:

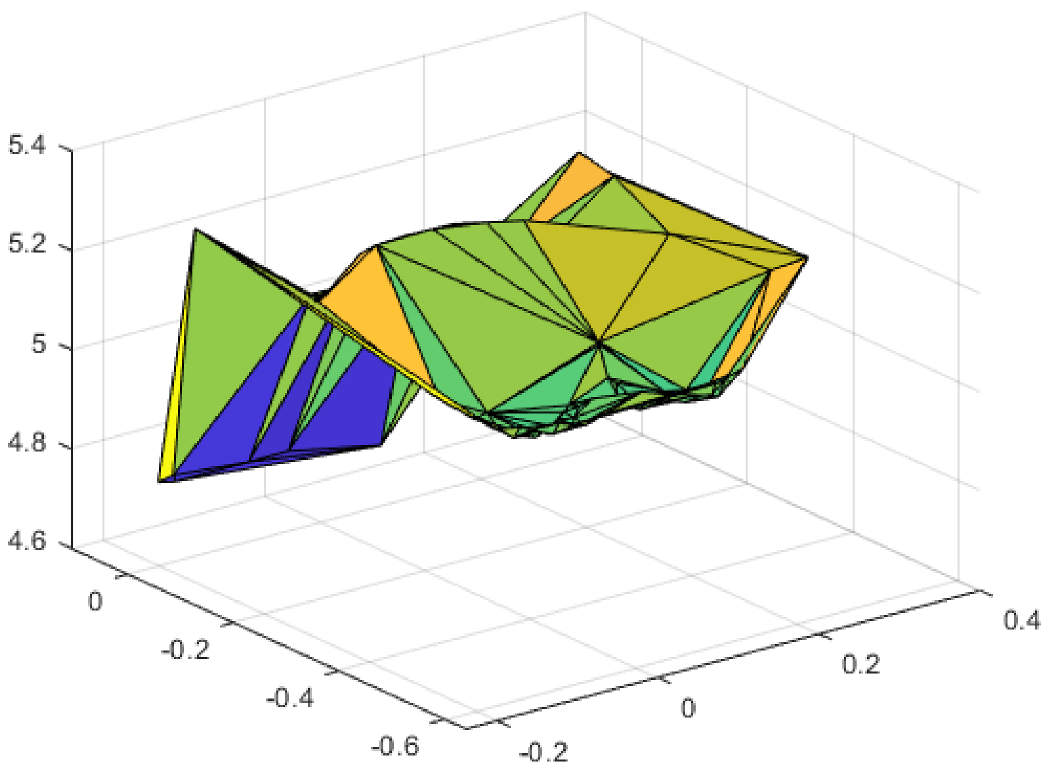
```
[[   0.39646368   -9.6184987    -6.98705872]
 [ -19.33335753    5.16402911  152.03315622]
 [  -8.01757436 -149.9100719     1.19968926]]
```



Statue 3D points



Statue 3D Visualize using Matlab



Discussion:

There are many parameters that we can fine tune. Such as the distance ration of SIFT, and threshold of inliers in RANSAC.If the distance ratio is too big, we will find more feature points than we need. So, setting the distance ratio to 0.7 is the best fit in this case.If the threshold of inliers in RANSAC is too big, it will increase more false inliers. So, setting the threshold of inliers to 0.01 is the best fit in this case.

Conclusion:

We implemented the Structure from Motion(SfM) in this assignment. We use SIFT to find the interesting points and 8-points algorithm to find the fundamental matrix. Using the fundamental matrix to get the essential matrix and then we get the four possible result.Then we choose the best one to be our SfM algorithm result. Finally, we use Matlab to map image texture to the points and get out models.