國立陽明交通大學

113-1 電腦視覺 HW-4

Group 9 313551130 鍾名捷/313551133 陳軒宇/313553036 葉軒宇

1. Introduction

The purpose of this assignment is to implement an SfM (Structure from Motion) system. It leverages techniques such as SIFT, RANSAC, and bundle adjustment to compute the relationships between camera matrices and reconstruct 3D scenes from 2D images.

2. Implementation procedure

Step 1. Feature Detection and Matching

First, we find the key points by using SIFT feature detector, interest points were extracted from pairs of images. Then descriptors for these keypoints were matched based on their SSD ratio.

$$SSD = \sum_{i=0}^{127} (des1[i] - des2[i])^2$$

The feature matching process uses a ratio-based approach to find reliable correspondence. So we applies a ratio test(threshold=0.5) to filter ambiguous matches, and return matched point pairs in both images.

$$\frac{\|f_1 - f_2\|}{\|f_1 - f_2'\|} < ratio$$

Step 2. Fundamental Matrix Estimation

The fundamental matrix is estimated using the normalized 8-point algorithm within a RANSAC framework. First, normalized the point using the following matrix.

$$\begin{bmatrix} 2/img\_width & 0 & -1 \\ 0 & 2/img\_width & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Then doing RANSAC iteration to find the best fundamental matrix.

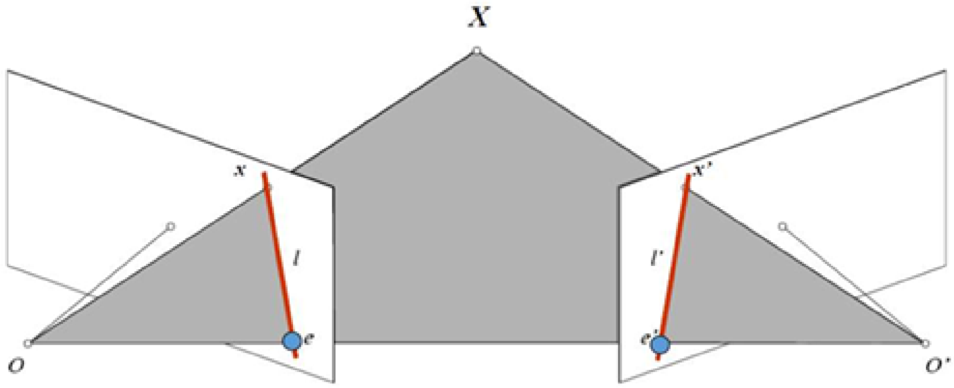


Figure 1

As the figure 1 shown, given a fundamental matrix F, the epipolar line l in left image can be obtained by $F^T x'$and the epipolar line l' in right image can be obtained by Fx. The fundamental matrix F can be solves by the following equation.

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

The equation can be expanded into

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

Then using 8-point algorithm to find out the fundamental matrix. Each time, we sample 8 pairs of match key points and form the following equation.

$$Af = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x'_8x_8 & x'_8y_8 & x'_8 & y'_8x_8 & y'_8y_8 & y'_8 & x_8 & y_8 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

After that, enforcement of rank-2 constraint through SVD, and denormalization to get the final fundamental matrix.

Step 3. Drawing Epipolar Lines

Fx's the epipolar line associated with $\boldsymbol{x}$ (**l** = $\boldsymbol{Fx'}$= [$\boldsymbol{i}$ $\boldsymbol{j}$ $\boldsymbol{k}$]) That is, for each point $\boldsymbol{x}$ = [$\boldsymbol{a}$ $\boldsymbol{b}$], we got the equation

$$[a \quad b \quad 1][i \quad j \quad k]^T = 0$$

We draw every line by two points on the equation where x coordinate located at 0 and width of image.

Step 4. Essential Matrix and Pose Estimation

The essential matrix was computed using the intrinsic calibration matrix and the fundamental matrix.

$$E = K_1^T F K_2$$

After we get the essential matrix E, the second camera matrix P_2 could have four possible solutions. Then do SVD on it to get U and V. The essential matrix E has two part $[R|t]$. The possible rotation matrix R could be UWV or $UW^T T$ and the translation vector could be $\pm u3$.

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

And the four possible solutions of the second camera matrix are following equations.

$$P_2 = [UWV|u_3]$$

$$P_2 = [UWV|-u_3]$$

$$P_2 = [UW^T V|u_3]$$

$$P_2 = [UW^T V|-u_3]$$

Step 5. Triangulation and Getting 3D point

After we get the second camera matrix, we use triangulation to predict the reconstruct 3D points. For key point pair x1, x2 and the camera matrices P_1, P_2

$$x_1 = w \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix}, x_2 = w \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix}$$

$$P_1 = [p_{11} \quad p_{12} \quad p_{13}], P_2 = [p_{21} \quad p_{22} \quad p_{23}]$$

Then create matrix A

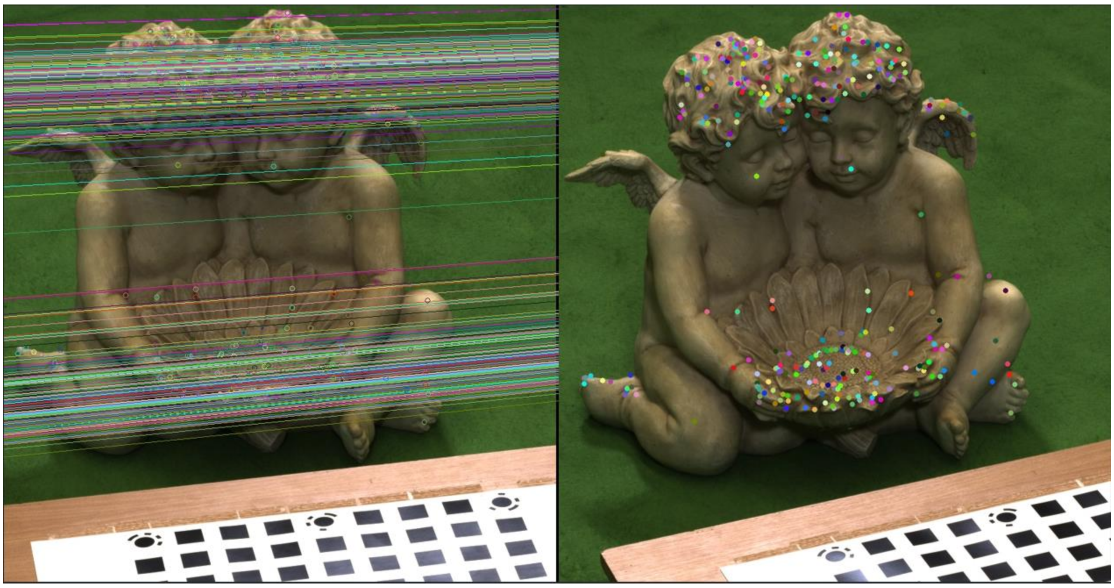
$$A = \begin{bmatrix} u_1 p_{13} - p_{11} \\ v_1 p_{13} - p_{12} \\ u_2 p_{23} - p_{21} \\ v_1 p_{23} - p_{22} \end{bmatrix}$$

Do SVD to matrix A. We find world coordinate of points by extracting the last row of V found in SVD. If the z value>0, it means that this point is in front of the camera.

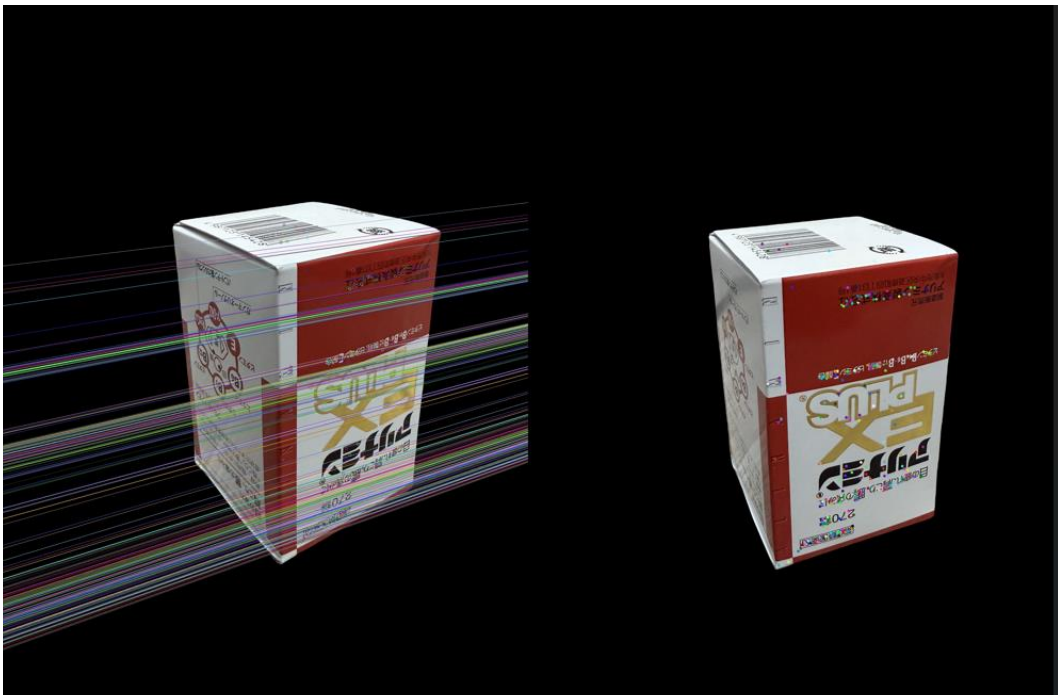
3. Experimental results

A. Epipolar line:

Data provided by TA:

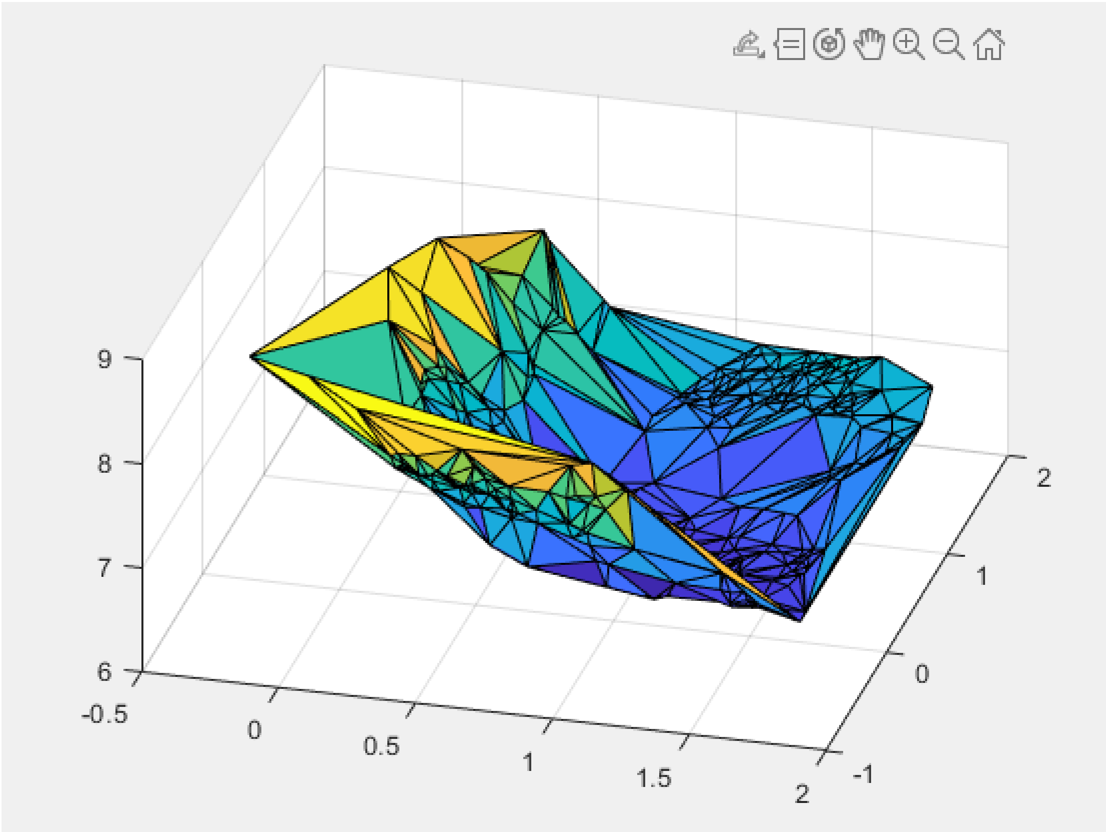


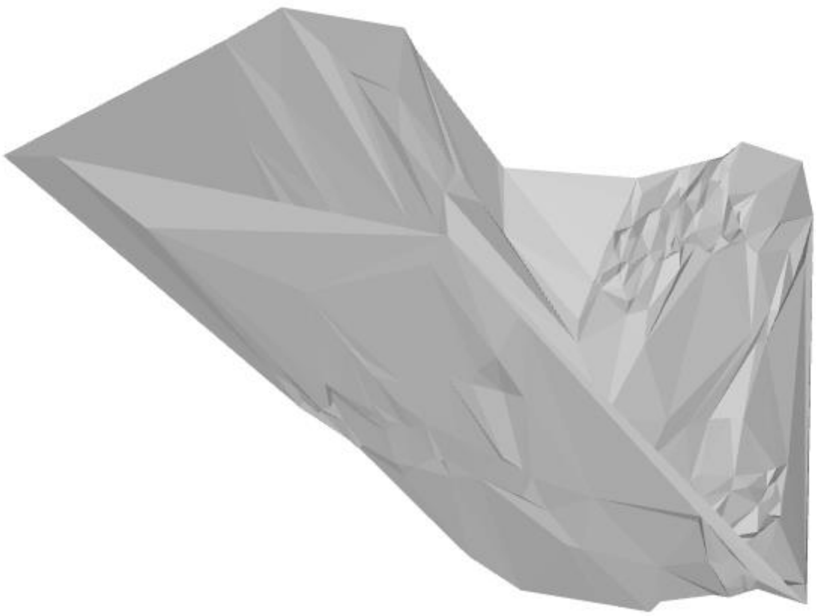
Our data:

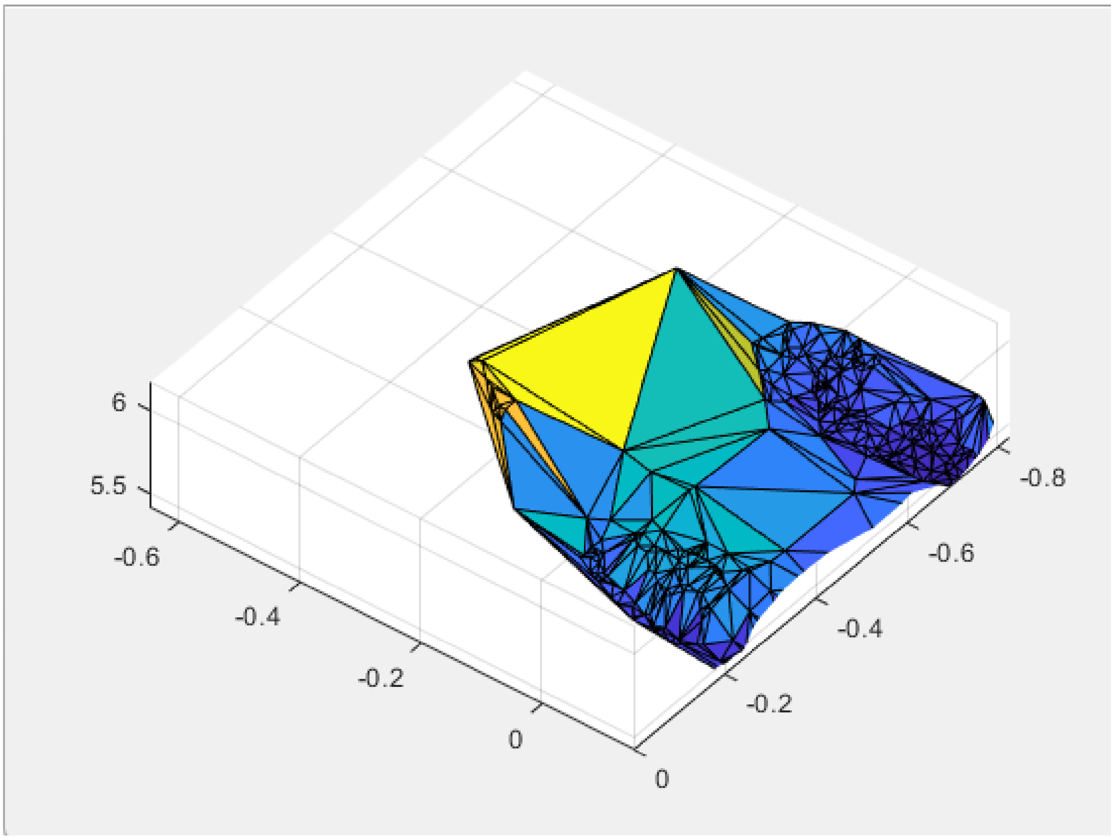


B.3D Model:

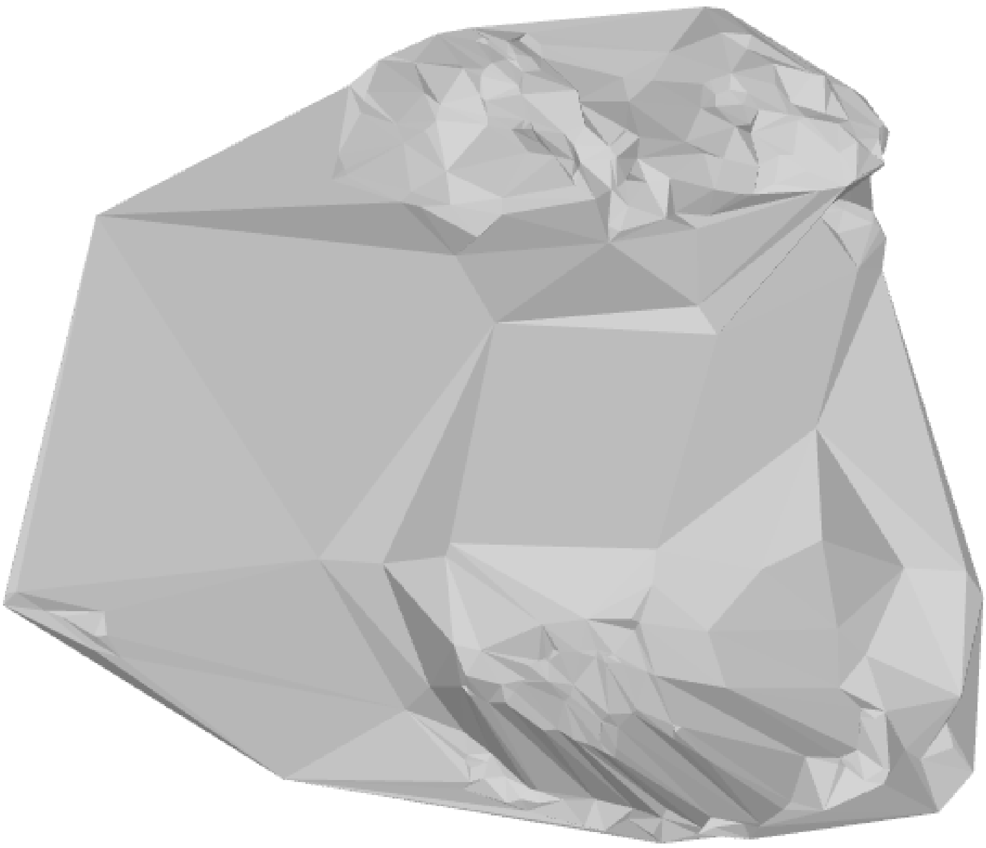
Data provided by TA:



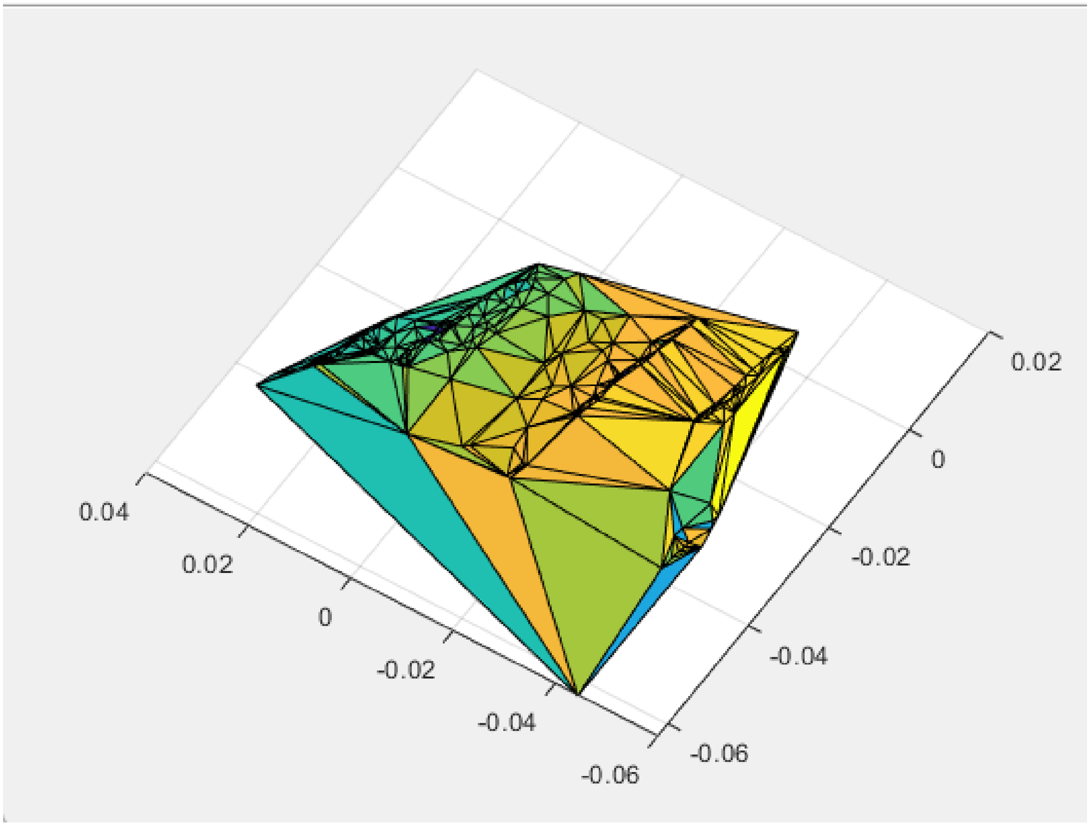


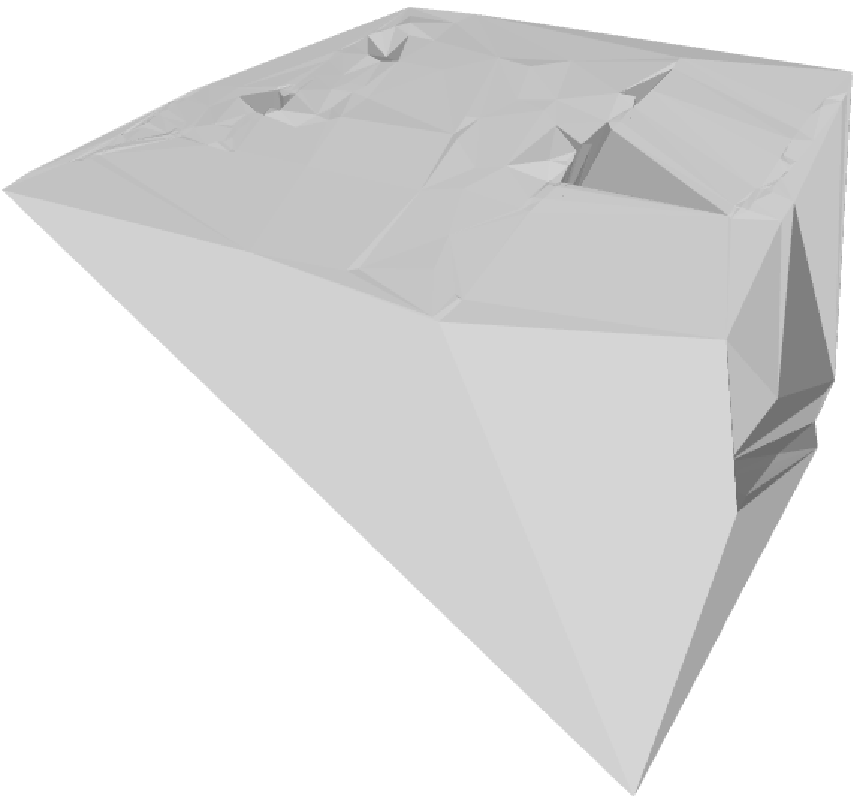


6
5.5
-0.6
-0.4
-0.2
0
0
-0.2
-0.4
-0.6
-0.8



Our data:





4. Discussion

We encountered significant challenges while working with our own captured images. The custom images often lacked well-defined features or suffered from poor lighting, resulting in reduced matching accuracy and reconstruction quality. After several times of taking new photo and adjusting the threshold, finally get some good results.

5. Conclusion

For different kinds of case, the choose of the threshold is very important. In some of the cases, too many or too little point would lead to the bad result, cause it cant calculate the right matrix we need. And it is very difficult to show good results on whatever matlab or online 3d viewer, due to the picture taken by us are not enough point pairs. If want to have nice result, the object you choose to shoot would we very important.

6. Work assignment plan

This project was completed through collaborative efforts among all team members. First, we discussed about the detail of implementing SIFT and snatching problem. After clearly understanding what we should de, we started to take some picture and modify the parameter to complete the whole task. Finally, we make a conclusion about we’ ve encountered within this project.