Group 27

Computer Vision – HW3

A. Introduction

The goal of the assignment is to reconstruct 3D information with two 2D images. The method we apply in this assignment is called SfM.

B. Implementation Procedure

Step 1. Use SIFT to get the match points

1. Read two images and transform images to gray level.
2. Find the key points and descriptors of two images with SIFT in OpenCV.
3. Use Sum of Squared Differences (SSD) for any pair: (des1, des2)

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

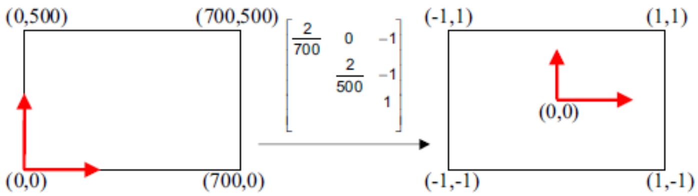
4. Store all the good match point pairs after ratio test (ratio = 0.5).

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

5. Show all the match point pairs.

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

1. Normalize image coordinates



• Correspondence Relation

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

2. Randomly choose 8 points to calculate fundamental matrix F.

$$X^{\mathrm{T}}FX' = 0$$

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

$$\mathbf{Af} = \begin{bmatrix} x_1'x_1 & x_1'y_1 & x_1' & y_1'x_1 & y_1'y_1 & y_1' & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n'x_n & x_n'y_n & x_n' & y_n'x_n & y_n'y_n & y_n' & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = \mathbf{0}$$

Do SVD on A to get F $\mathrm{A} = \mathrm{U} \times \mathrm{S} \times \mathrm{V}$ and the last column of V is F(reshape to 3 * 3)

Do SVD on F $\mathrm{F} = \mathrm{U} \times \mathrm{S} \times \mathrm{V}$ and we need to set S(3,3)=0

And the final F $= \mathrm{U} \times \mathrm{S} \times \mathrm{V}'$ is what we want.

3. For each pair of points, calculate whether $x_2' F x_1$ is smaller than the error threshold we set. If smaller, put it in the inliner group. Otherwise, just discard them.
4. For each iteration, do 2. and 3. And check which iteration 's the point number in the inliner group is max. That iteration's F is the best matrix we find.
5. De-normalize. $F_{de} = T'FT$

Step 3. draw the interest points

1. $\boldsymbol{Fx'}$ is the epipolar line associated with $\boldsymbol{x}$ ($\mathbf{l} = \boldsymbol{Fx'} = [\boldsymbol{i}\ \ \boldsymbol{j}\ \ \boldsymbol{k}]$)
   That is, for each point $\boldsymbol{x} = [\boldsymbol{a}\ \ \boldsymbol{b}]$, we got the equation
   $$[\mathbf{a}\ \ \mathbf{b}\ \ \mathbf{1}][\mathbf{i}\ \ \mathbf{j}\ \ \mathbf{k}]^{\boldsymbol{T}} = \mathbf{0}$$
2. We draw every line by two points on the equation where x coordinate located at 0 and width of image.

Step 4. get the new U and V

1. We use fundamental matrix and intrinsic parameters to get essential matrix.
   $$\mathrm{E} = \mathrm{K}_1^T F K_2$$
2. Do SVD on essential matrix and get $[\mathrm{U},\ \mathrm{S},\ V^T] = \mathrm{SVD}(\mathrm{E})$
   $$\mathrm{E} = \mathrm{U} * \mathrm{diag}(\mathrm{S}[0][0], \mathrm{S}[1][1], 0) * V^T$$

3. Do SVD on E of 2. and get the new U and V, $[\mathrm{U},\ \mathrm{S},\ V^T] = \mathrm{SVD(E)}$

Step 5. Apply triangulation to get 3D points

1. We set camera1 the origin of the world coordinate. Then, the camera matrix for camera1 $\mathrm{P1} = [\mathrm{I}\ |0]$
2. There are four possible P2

$$\mathbf{P}_2 = [\mathbf{UWV}^\top \mid +\mathbf{u}_3] \\ \mathbf{P}_2 = [\mathbf{UWV}^\top \mid -\mathbf{u}_3] \\ \mathbf{P}_2 = [\mathbf{UW}^\top\mathbf{V}^\top \mid +\mathbf{u}_3] \\ \mathbf{P}_2 = [\mathbf{UW}^\top\mathbf{V}^\top \mid -\mathbf{u}_3] \qquad \mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3. We can get world coordinate for points in 2D image by finding A.

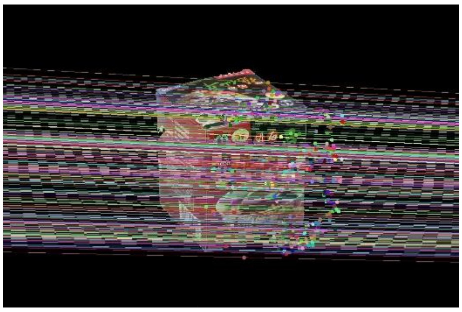
$$A = \begin{bmatrix} u\mathbf{p}_3^\top - \mathbf{p}_1^\top \\ v\mathbf{p}_3^\top - \mathbf{p}_2^\top \\ u'\mathbf{p}_3'^\top - \mathbf{p}_1'^\top \\ v'\mathbf{p}_3'^\top - \mathbf{p}_2'^\top \end{bmatrix}$$

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.

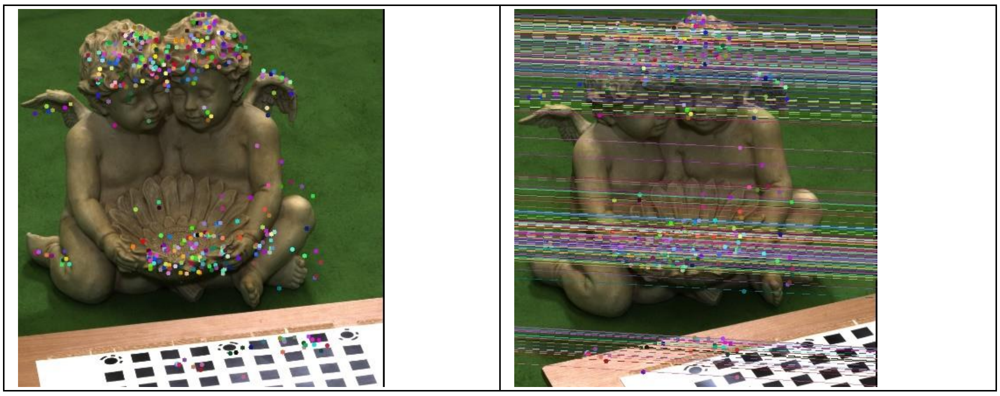
4. We can choose the most appropriate P2 by finding the one allows most points in front of the camera2. In other words, the P2 that generate most points that have z value in world coordinate larger than zero is the P2 we are finding.
5. We now have propriate P1 and P2 to change all points into world coordinate via the third process in this step.

C. Experimental Result

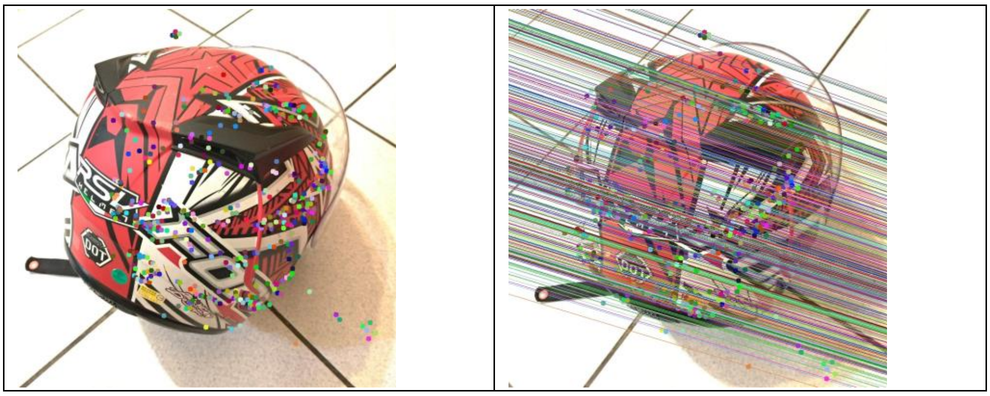
- Data provided by TA(epipolar line)

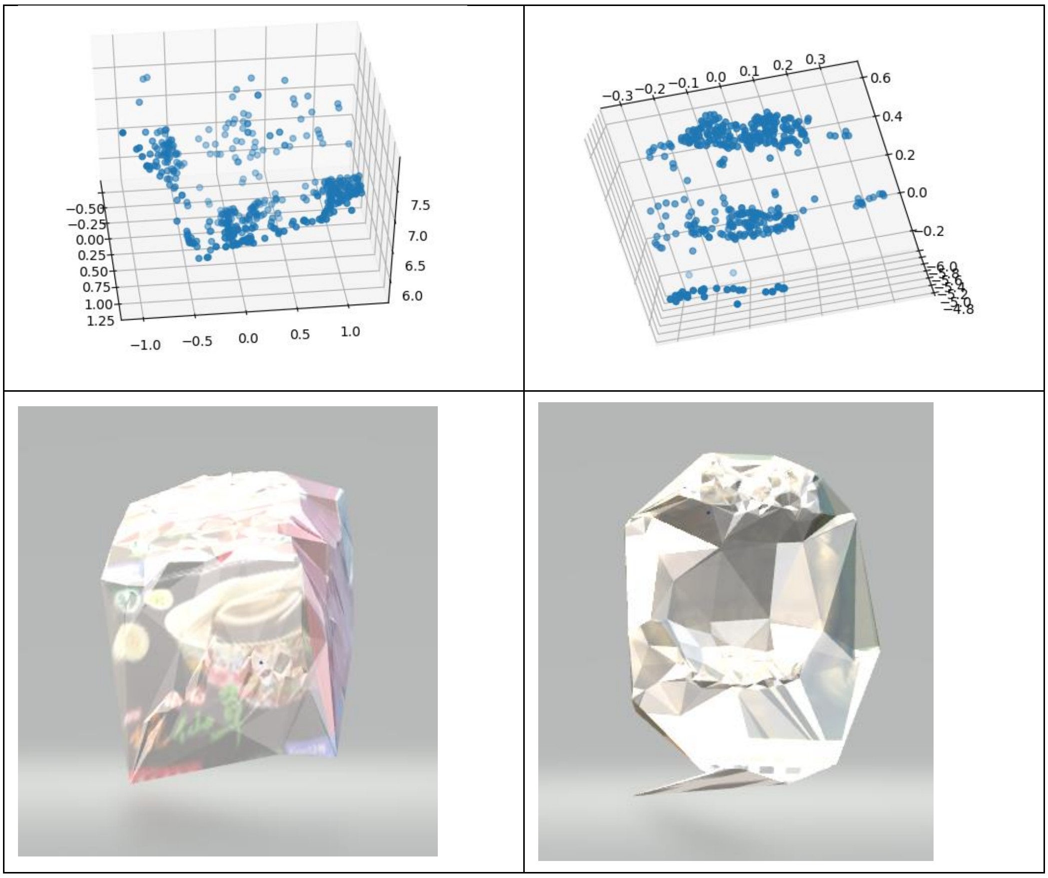
Group 27

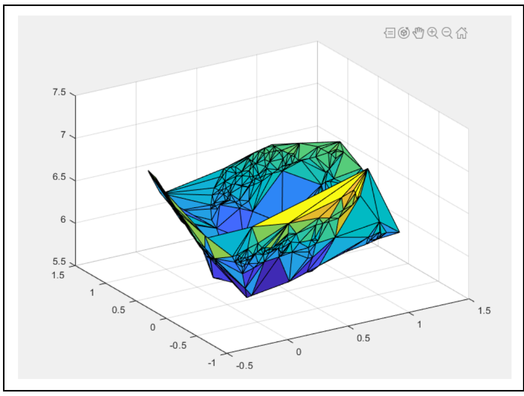


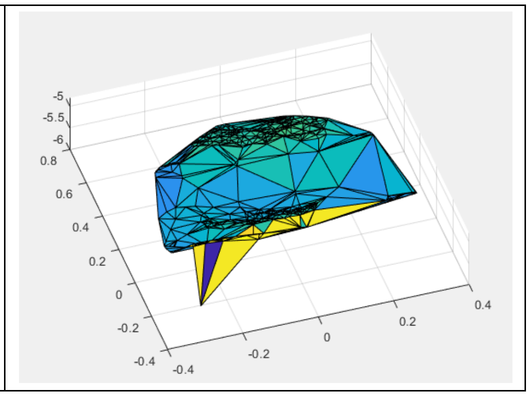
- Data provided by TA(epipolar line)



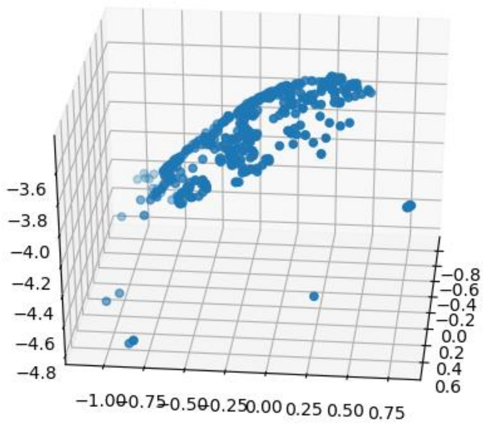
- Data provided by TA

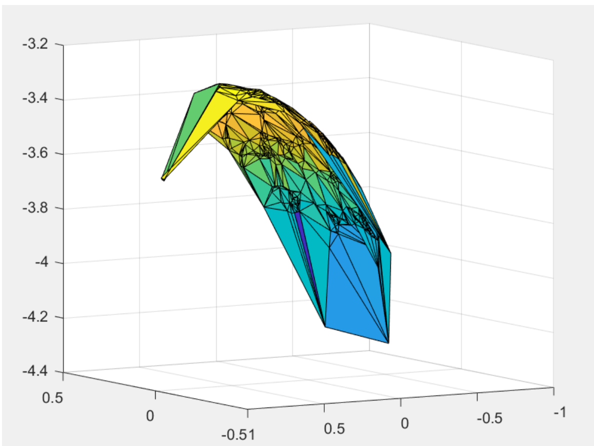


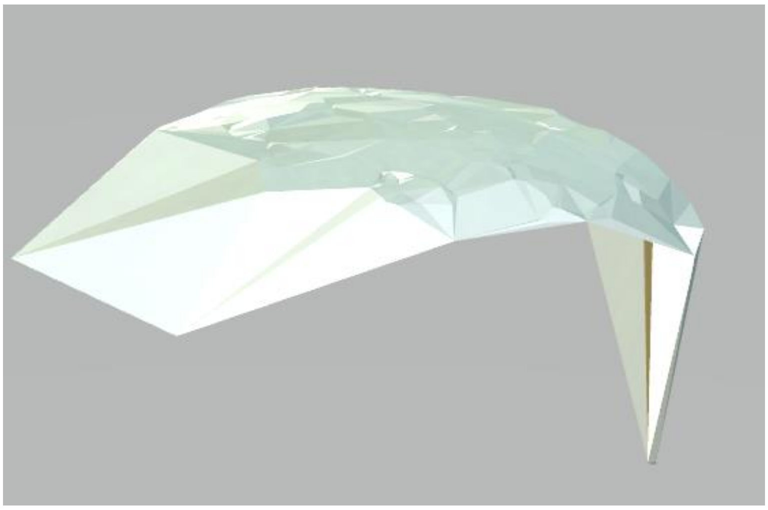

7.5
7
6.5
6
5.5
1.5
1
0.5
0
-0.5
-1
-0.5
0
0.5
1
1.5


-5
-5.5
-6
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.4
-0.2
0
0.2
0.4

- Our own data


−3.6
−3.8
−4.0
−4.2
−4.4
−4.6
−4.8
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6


-3.2
-3.4
-3.6
-3.8
-4
-4.2
-4.4
0.5
0
-0.51
0.5
0
-0.5
-1



D. Discussion

1. In this assignment, we need to implement RANSAC. The RANSAC use in this task is different from the one we implement in assignment3. We were confused at first. We thought it means that we have to choose better points like assignment3. However, the result looked weird. We, then, learn more about RANSAC.
2. This is the assignment we spend most time on so far. The procedure is more complex than previous assignments. Moreover, there are some instructions in the homework that is not that clear or with some minor error. As a result, we spend a lot of time to check teacher's pdf to fix our code.

E. Conclusion

1. It is important to choose threshold. Whether too many or to little points would lead to worse result.
2. The result shown in MATLAB is not really good. We believe this is due to the reason that there are not enough point pairs. There are some specific areas with little points or even no points since there is no obvious texture to locate interest point. This make that specific area with no information when reconstructing and looks bad. However, this is reasonable since it is not rich in texture.

F. Work Assignment

All of us wrote some parts of code and try to debug of the entire code. We discuss to each other to check weather our thoughts were correct. Thus, I would say that all of us has a great contribution to this assignment no matter in code or in report.

Group 27