Homework 3 - Computer Vision, 2018 Spring
Team35

Teammates:0416088鄭甯遠,0650249劉岑陽子,0660819司菲Siffi Singh

The task is to perform Structure from Motion (SfM), on the images given to our team. We are also given information about the intrinsic matrix for the cameras through which the images The tasks can be further broken down into following subtasks:
1. Find out correspondence across images.
2. Estimate the fundamental matrix across images.
3. Draw the interest points found in step 1 in one image and the corresponding epipolar lines in another.
4. Get possible solutions of essential matrix from fundamental matrix.
5. Find out the most appropriate solution of fundamental matrix.
6. Apply triangulation to get 3D points.

In this report, we will be explaining each part in the following way:
1. Steps to run the file.
2. Explanation of Code Logic
3. Results
4. Inferences from Result

1. Steps to run file

In this section, we will explain how to run the code of performing 3D reconstruction using structure from motion. For our code, you need to put the images and specify the path of the images in the pathname correctly. After doing so, you need to run **main.m** which will create model1.obj
To view the result, you should have installed Meshlab and import the model1.obj to view the 3D structure.

2. Explanation of Code Logic

I. Find out correspondence across images

For the first step, one must find correspondence across images. The technique is identical to the last homework where one first find feature points then use RANSAC algorithm to match up. The idea is to find the corresponding interest points in the two images that we are taking in as a data at a time. To do so, we read the image first, convert it to grayscale, detect and extract relevant features from it, and once we have the valid points and features from Image1 and Image2, we match the features using the features from the two images. To visualize the result we can show the matched features and view the correspondence.

II. Estimate the fundamental matrix across images

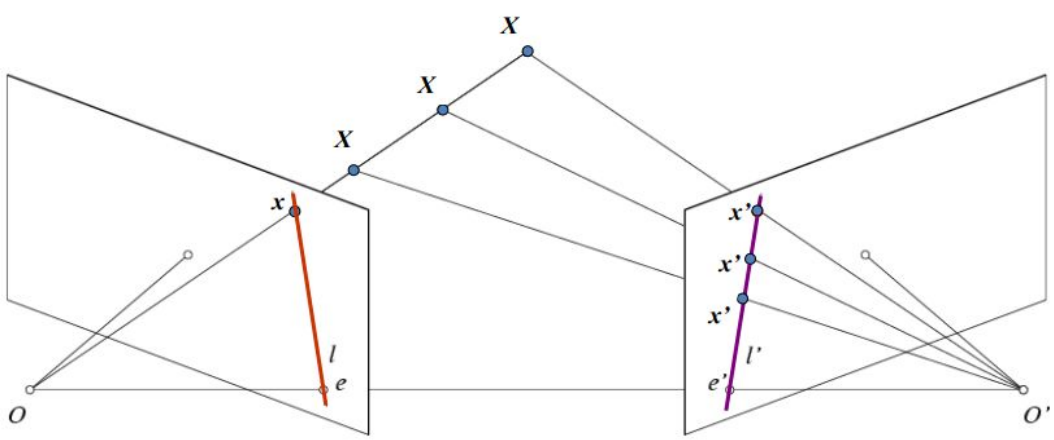
Using normalized 8 points algorithm to find fundamental matrix. The basic eight-point algorithm can in principle be used also for estimating the fundamental matrix F. The defining constraint for F is $(\mathbf{y}')^T \mathbf{F} \mathbf{y} = 0$ where y,y' are the homogeneous representations of corresponding image coordinates . This means that it is possible to form a matrix Y in a similar way as for the essential matrix and solve the equation:

$$\mathbf{f}^T \mathbf{Y} = \mathbf{0}$$

Using the above formula we can estimate the fundamental matrix across images.

III. Draw the interest points and the corresponding epipolar lines

For this part, we need to find the interest point, the interest points are well-defined positions in space. The local image structure around the interest point is rich in terms of local information contents (e.g. **significant 2D texture**), such that the use of interest points simplify further processing in the vision system. It is stable under local and global perturbations in the image domain as illumination/brightness variations, such that the interest points can be reliably computed with high degree of repeatability.



The end goal is to recover the 3D information from the two given 2D images. This is where drawing epipolar lines come handy. The potential matches for Image 1 lie on the corresponding line to the other image.

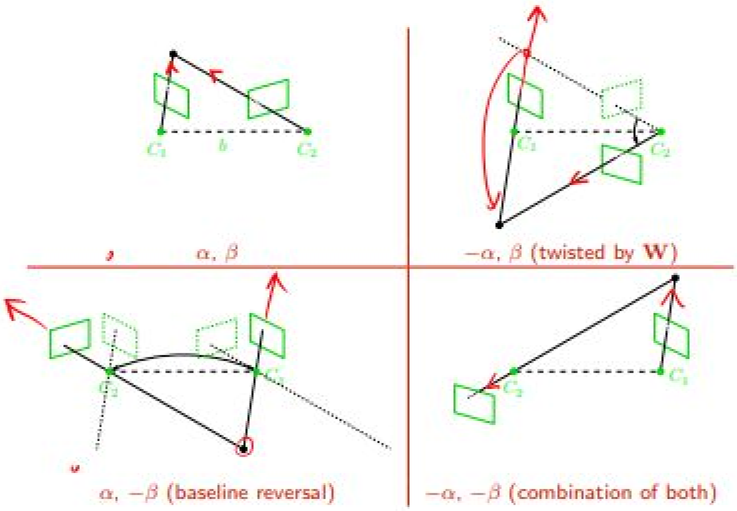
Initially we have the **baseline**, that is the line connecting the two camera centres. We have the **epipolar plane** from this, which consists of the plane containing baseline.
From the intersections of baseline with image planes, and from the projections of the other camera center, we can retrieve the **Epipoles**.
Intersections of epipolar plane with image planes, give us the **Epipolar lines.**

IV. Get possible solutions of essential matrix from fundamental matrix

The essential matrix is the specialization of the fundamental matrix to the case of normalized image coordinates. In our current problem we have, 4 possible solutions that we have drawn from the combinations of Translational and Rotational Matrices.
Consider a camera matrix decomposed as P = K[R | t], and let x = PX be a point in the image. Since, the calibration matrix K is known, then we may apply its inverse to the point x to obtain the point $\hat{x} = K^{-1}x$. Then $\hat{x} = [R \mid t]X$, where $\hat{x}$ is the image point expressed in normalized coordinates. It may be thought of as the image of the point X with respect to a camera [R | t] having the identity matrix I as calibration matrix. The camera matrix K−1P = [R | t] is called a normalized camera matrix, the effect of the known calibration matrix having been removed. The essential matrix has fewer degrees of freedom, and additional properties, compared to the fundamental matrix. Out of these 4 essential matrix, we will explain how we have chosen the best solution and that chosen solution would help us do the further calculation. Above is the formulation for essential matrix from the fundamental

$$E = K'^{\top} F K.$$

matrix where K is the fundamental matrix. The four possible matrice are the combinations of R1, R2 and T1, T2. where, R1 = U * W * V'; R2 = U * W' * V'; T1 = U(:, 3); T2 = -U(:, 3); We use, (R1, T1), (R2, T1), (R1, T2) and (R2, T2) as



described above in the images.

V. Find out the most appropriate solution of fundamental matrix

Next task is to find appropriate solution from the generated solutions. The solution that best represents the fundamental matrix is the one that relates pixel coordinates in the two views. It's a more general form than essential matrix, here is where we remove the need to know intrinsic parameters.

```
[U,S,V] = svd(E);
m = (S(1,1)+S(2,2))/2;
E = U*[m,0,0;0,m,0;0,0,0]*V';
[U,S,V] = svd(E);
W = [0,-1,0;1,0,0;0,0,1];
```

We used the hint provided by professor and calculated the appropriate solution for fundamental matrix through the following. In order to determine an optimal F, the reprojection error has to be minimized.This is achieved through e.g. the following problem formulation:

$$\min_{\mathbf{u}} \sum_{i=1}^{m} d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2$$

$$\text{s.t.} \quad \hat{\mathbf{x}}_i'^{\top} \mathrm{F} \hat{\mathbf{x}}_i = 0, \; i = 1, \dots, m$$

$$\|\mathrm{F}\|_F = 1, \det(\mathrm{F}) = 0,$$

Given an algorithm to solve,

$$\min_{\mathbf{u}} \frac{1}{2} f(\mathbf{u})^T W f(\mathbf{u})$$
$$\text{s.t. } c(\mathbf{u}) = 0$$

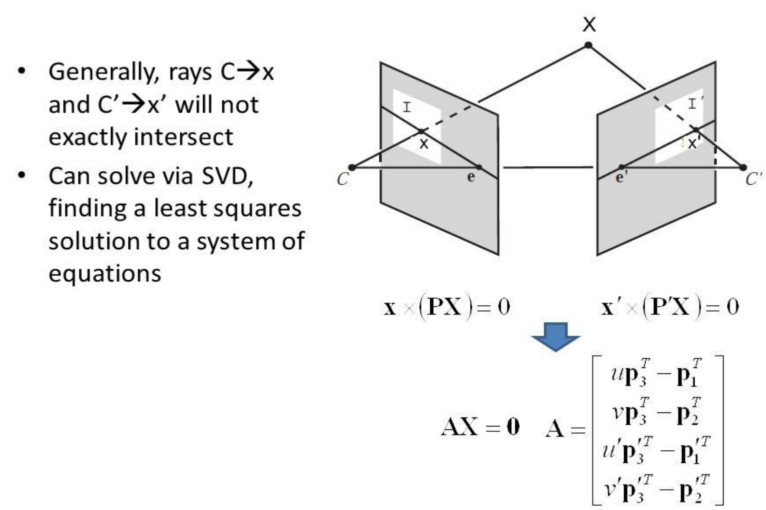
we get,

$$f(\mathbf{u}) = \begin{bmatrix} \hat{\mathbf{x}}_1 - \mathbf{x}_1 \\ \vdots \\ \hat{\mathbf{x}}_m - \mathbf{x}_m \\ \hat{\mathbf{x}}'_1 - \mathbf{x}'_1 \\ \vdots \\ \hat{\mathbf{x}}'_m - \mathbf{x}'_m \end{bmatrix} \text{ and } c(\mathbf{u}) = \begin{bmatrix} \hat{\mathbf{x}}_1'^{\top} \mathbf{F} \hat{\mathbf{x}}_1 \\ \vdots \\ \hat{\mathbf{x}}_m'^{\top} \mathbf{F} \hat{\mathbf{x}}_m \\ \|\mathbf{F}\|_F - 1 \\ \det(F) \end{bmatrix}$$

Hence, by using the above formulation, we calculate F that minimizes the algebraic error.

VI. Apply triangulation to get 3D points

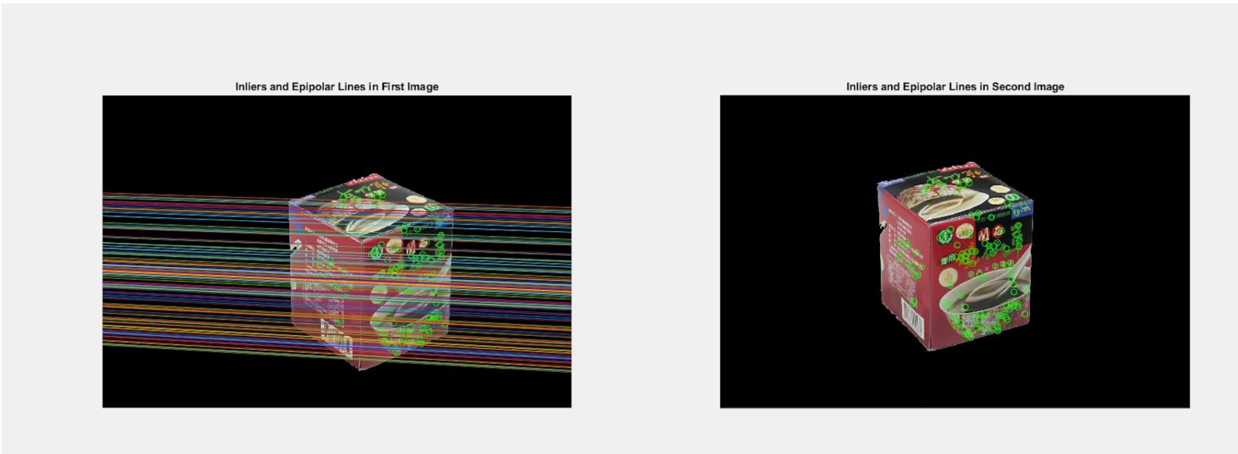
Given cameras P1, P2 and a correspondence x ↔ y compute a 3D point X projecting to x and y. Triangulation, sometime called reconstructions is a process that map a 2d point back to its 3d coordinate. It can be done using the essential matrix found in step 5. We make use of the Singular value decomposition method to perform triangulation and get 3D points.



The Pros and Cons of triangulation could be that it works for any number of corresponding images and that it is not projectively invariant respectively.

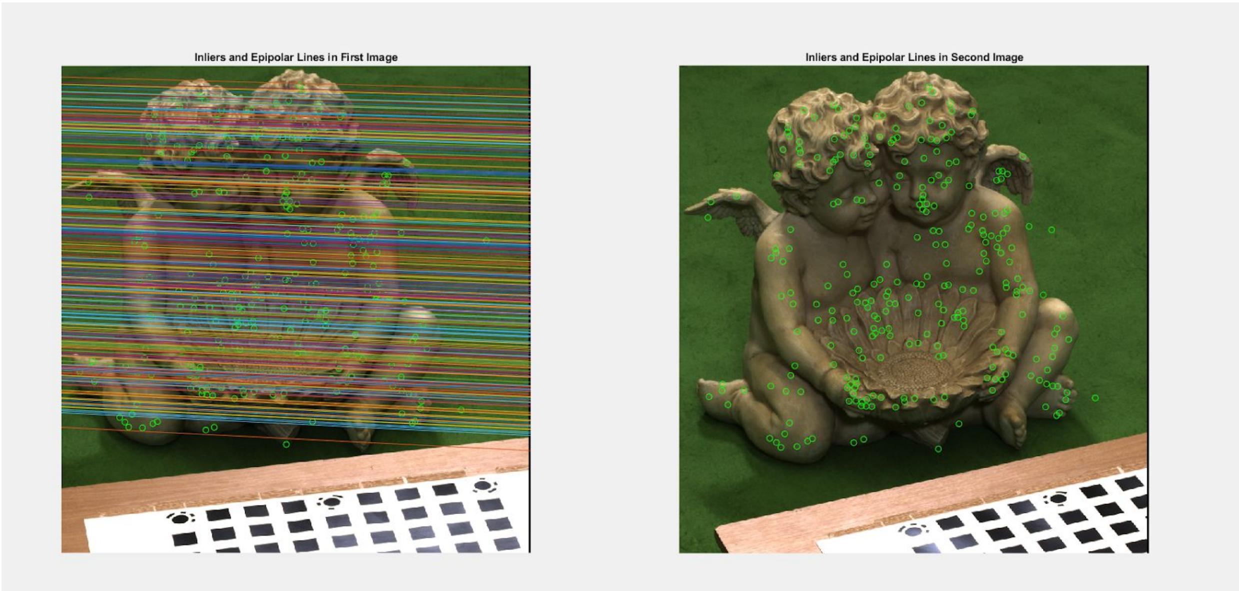
3. Results

The results of epipolar lines drawn for Mesona is below.

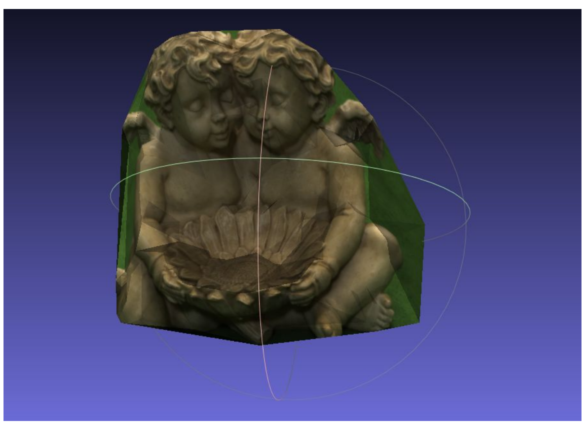


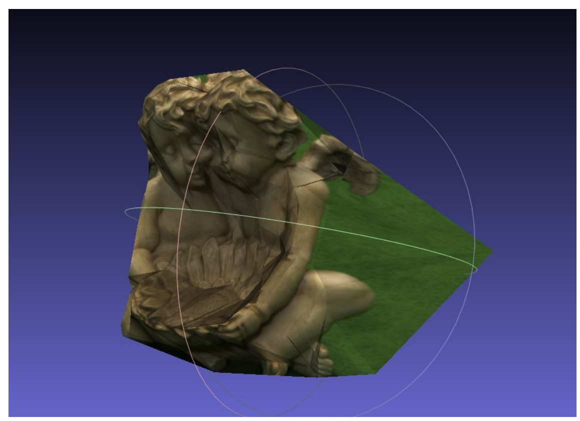
Below is the result for 3D reconstruction of the image for Mesona1 and Mesona2 from a few angels captured and attached.

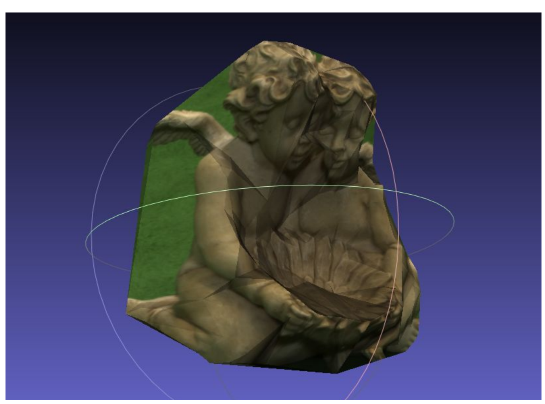
The results of epipolar lines drawn for Statue is below.



Below is the result for 3D reconstruction of the image for Staue1 and Statue2 from a few angels captured and attached.







4 . Inferences from Result

- Matched points make a huge difference in results. Once we found the correct matched points, the result is pretty good.
- Also, there should be matched points on each sides of the mesona box. Otherwise, we can not recover the 3D model. It would just be a plane.