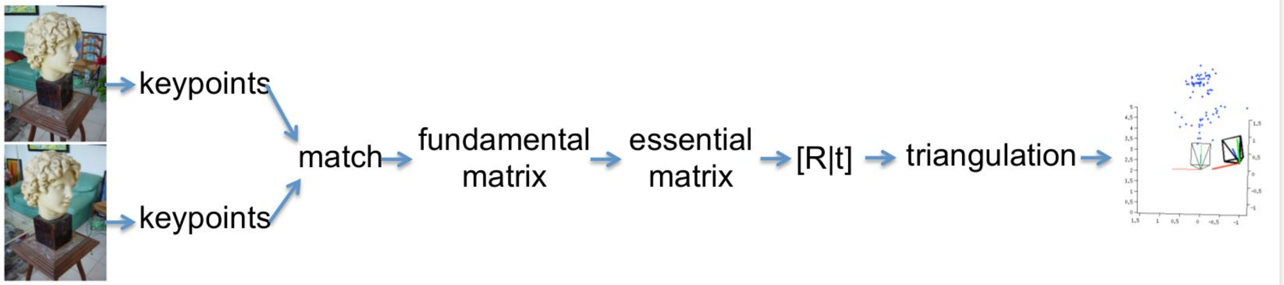
CV HW4 Report

Group 13

Introduction

Structure from motion (**SfM**) is a technique for estimating three-dimensional structures from two-dimensional image sequences that may be coupled with local motion signals. It is studied in the fields of computer vision and visual perception.

In this assignment, we calculate feature matching, fundamental matrix, essential matrix and camera pose. After we finish these calculation, we can use triangulation to reconstruct the 3D points.



Procedure

1. Feature Matching, Fundamental Matrix and RANSAC
2. Estimate Essential Matrix from Fundamental Matrix
3. Estimate Camera Pose from Essential Matrix
4. Check for Cheirality Condition using Triangulation

Procedure-Feature Matching, Fundamental Matrix and RANSAC

1.1. Feature matching - Hw3

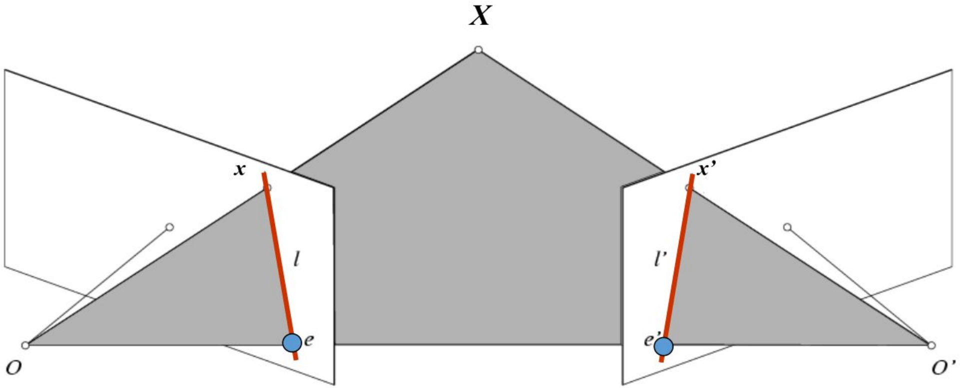
1.2. Epipolar Geometry

1.3. Estimating Fundamental Matrix F

1.4. Match Outlier Rejection via RANSAC

Epipolar Geometry

Based on Epipolar geometry, we can get the properties of the fundamental matrix and derive the below equation



$$x^T F x' = 0 \quad \text{with} \quad F = K^{-T} E K'^{-1}$$

Estimating Fundamental Matrix

We solve the fundamental matrix by 8-point algorithm. 8 point algorithm is derived from the equation of fundamental matrix properties. It uses 8 matching pair to construct the matrix A and solve f from Af=0 using SVD

$$X^\top F X' = 0$$

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

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

Matlab:

```
[U, S, V] = svd(A);
f = V(:, end);
F = reshape(f, [3 3])';
```

Estimating Fundamental Matrix

After we get the F, we need to make det(F) be equal to zero. Therefore, we resolve det(F)=0 constraint using SVD

Matlab:

```
[U, S, V] = svd(F);
S(3,3) = 0;
F = U*S*V';
```

Estimating Fundamental Matrix

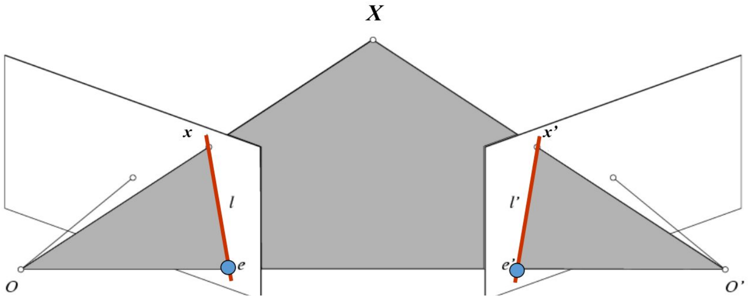
Because the difference of data magnitude, we need to normalized image coordinates to yields better results. We calculate the transformation matrix T and T', then multiply to the origin image 2D coordinates

$$\tilde{x} = Tx \quad \tilde{x}' = T'x'$$

RANSAC

Use RANSAC with 8-point to select the fundamental matrix that fit the most inliers. And because the normalized coordiantes, we need to de-normalize the fundamental matrix. $F = T'^{\top}\tilde{F}T$

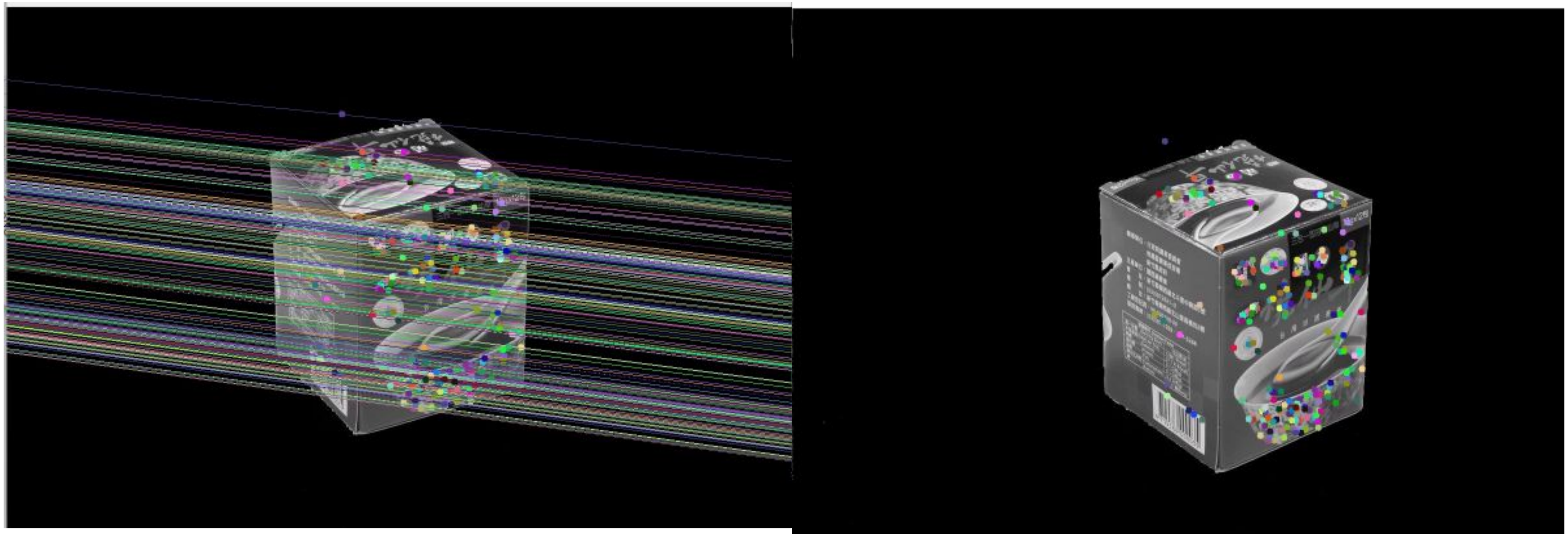
Finally, we get the fundamental matrix. So we can multiply it to 2D pixel coordinates to generate the epipolar line



$(l = F x')$

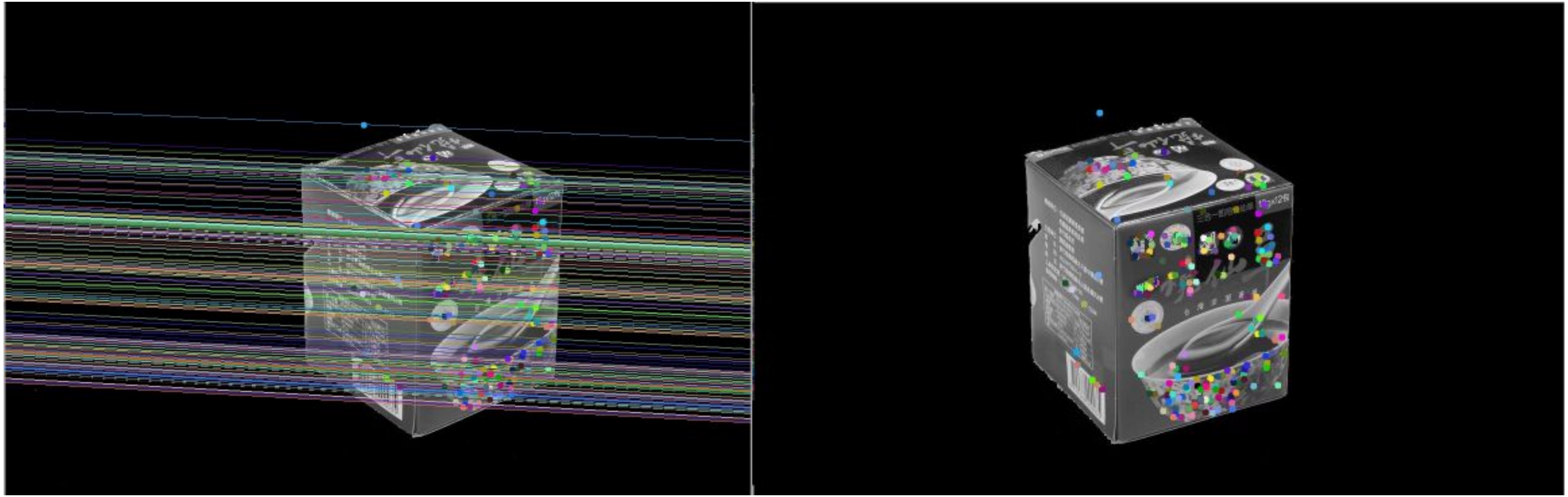
Result - Epipolar Line

Mesona-1



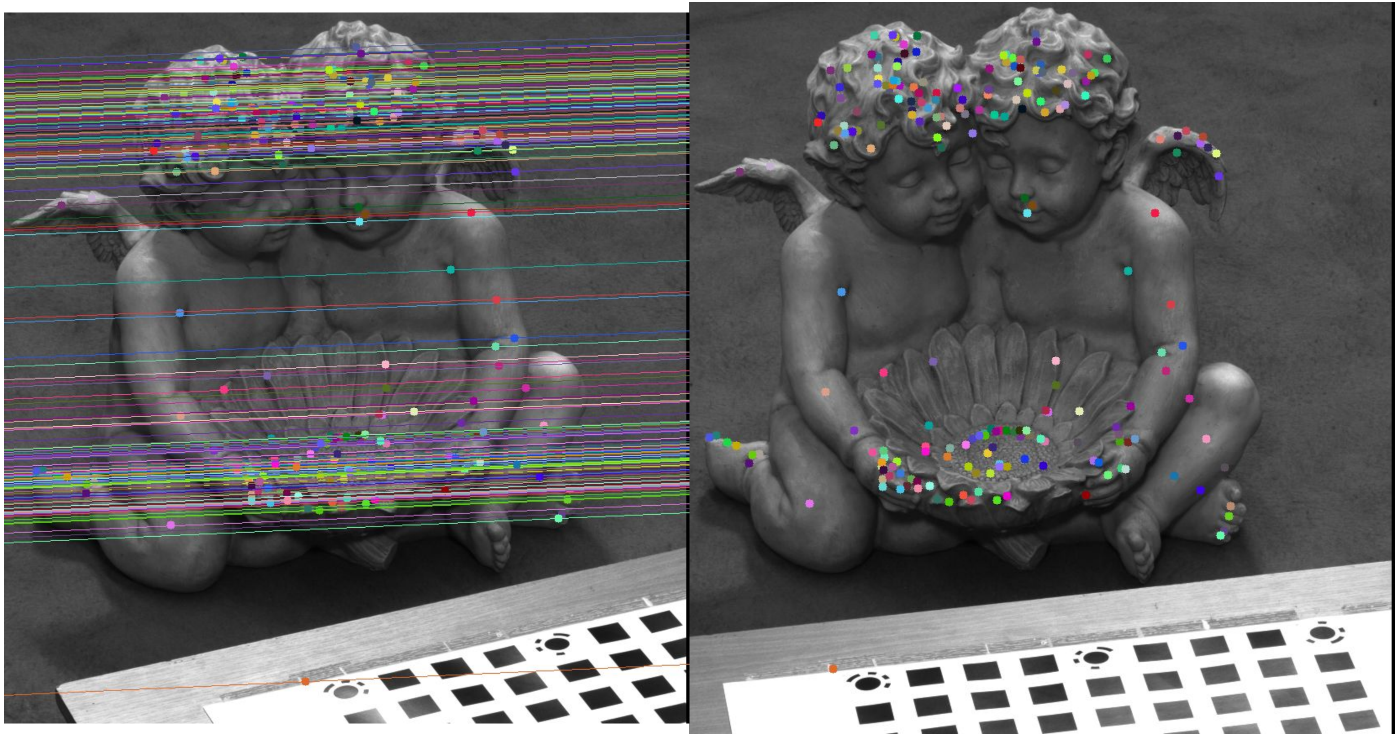
Result - Epipolar Line

Mesona-2



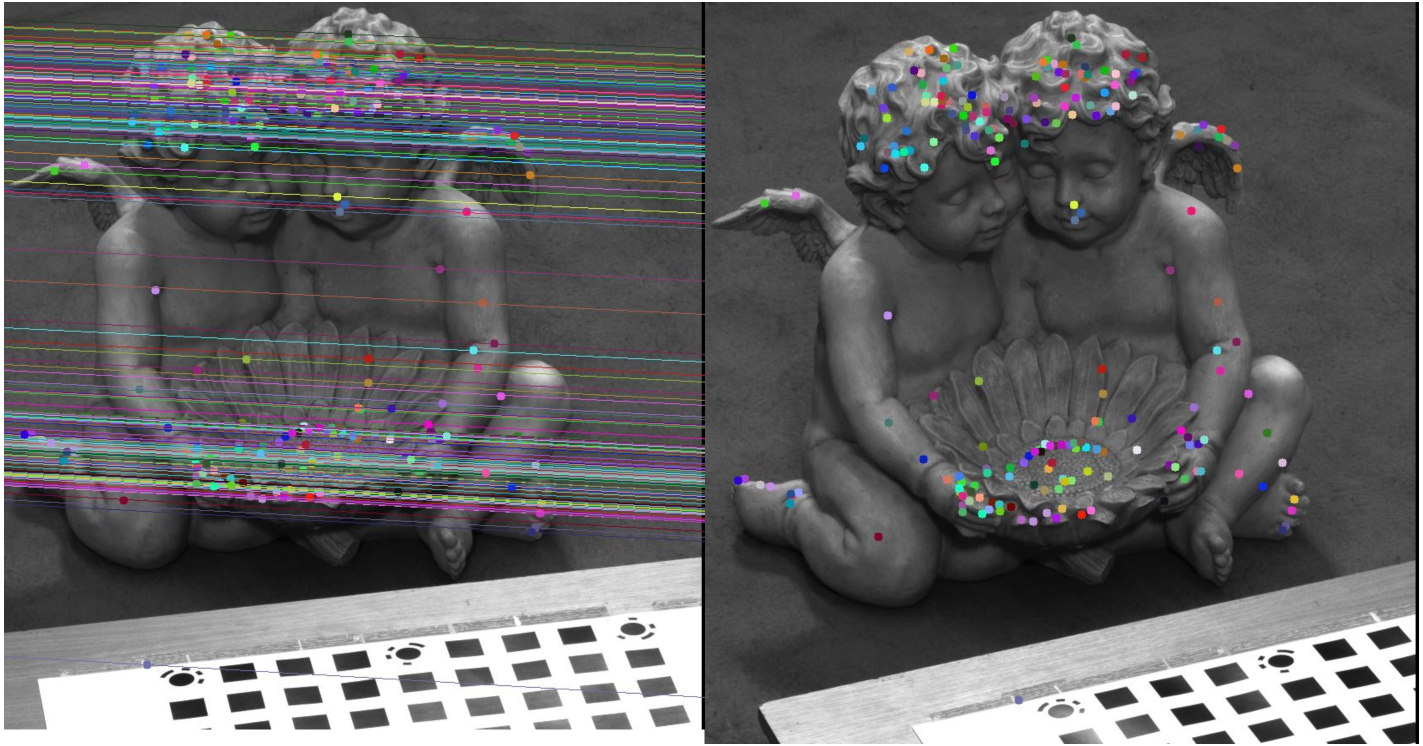
Result - Epipolar Line

Statue-1



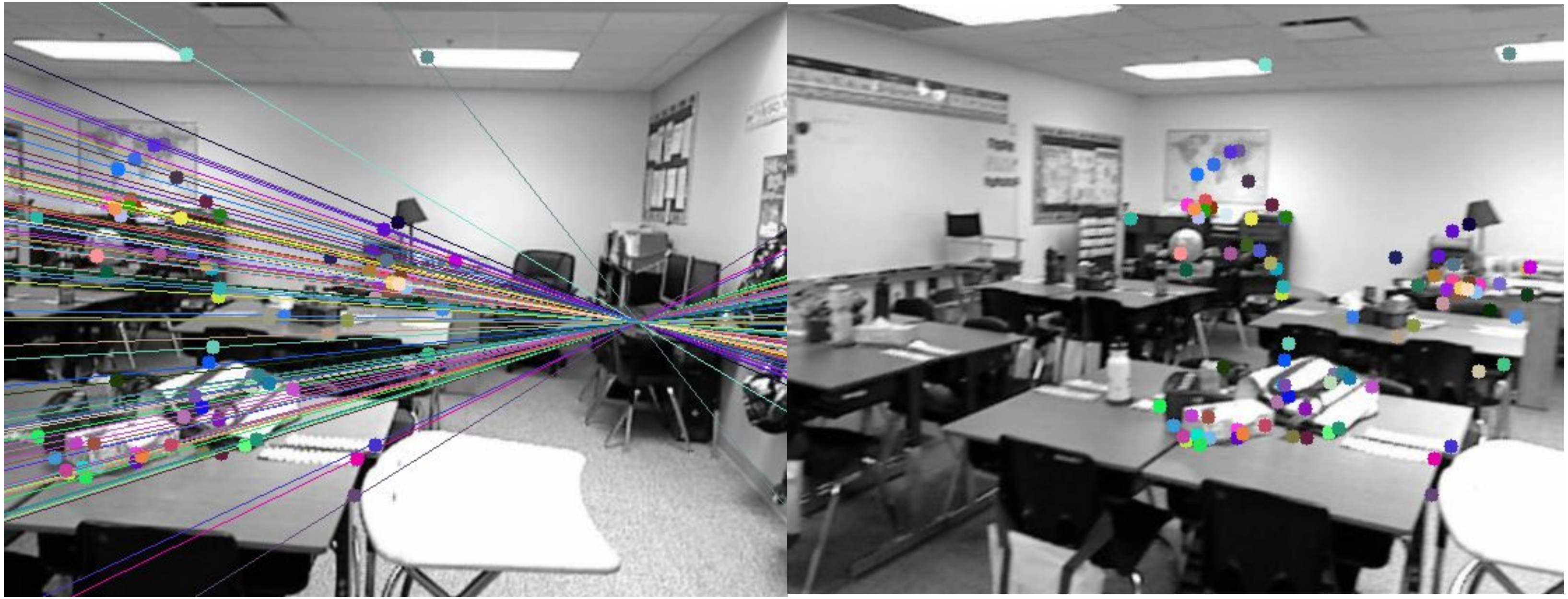
Result - Epipolar Line

Statue-2

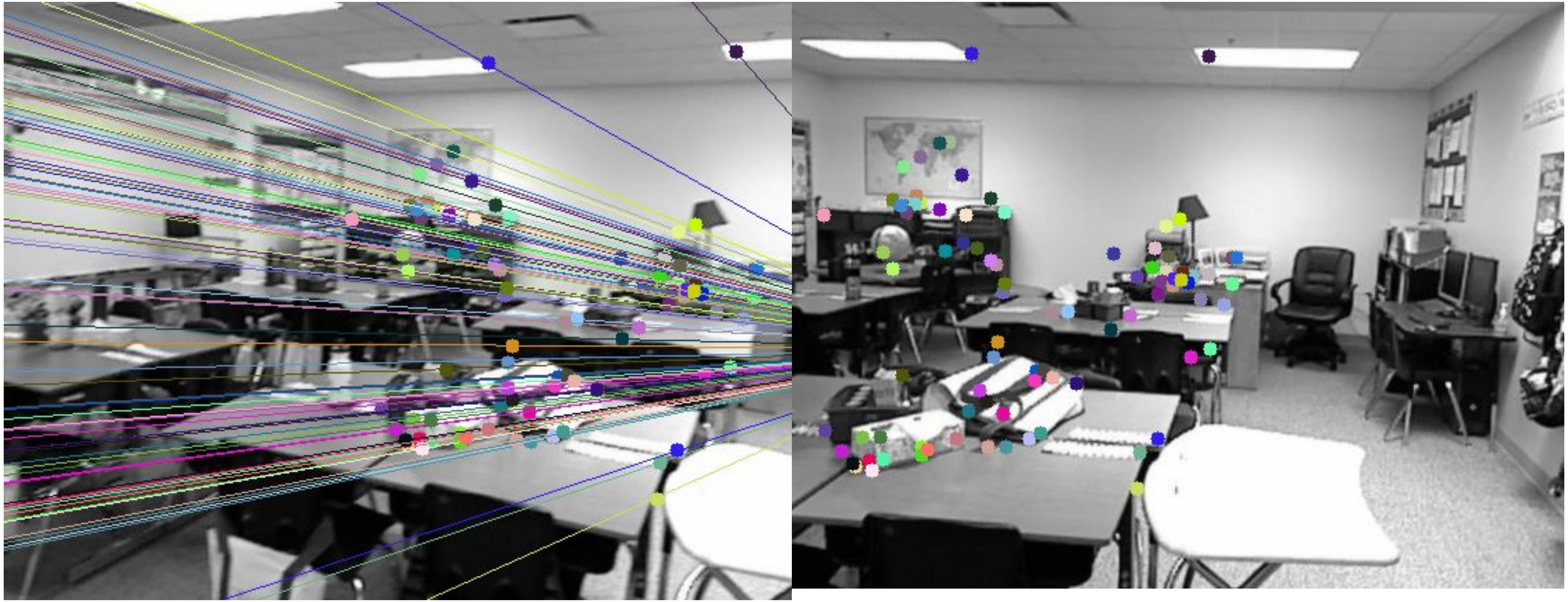


Result - our dataset

Result - our dataset



Result - our dataset



Estimate Essential Matrix from Fundamental Matrix

The camera pose consists of 6 degrees-of-freedom (DOF) Rotation (Roll, Pitch, Yaw) and Translation (X, Y, Z) of the camera with respect to the world. Since the E matrix is identified, the four camera pose configurations: (C1,R1),(C2,R2),(C3,R3) and (C4,R4)

Estimate Essential Matrix from Fundamental Matrix

where $C \in \mathbb{R}^3$ is the camera center and $R \in SO(3)$ is the rotation matrix,

can be computed. Thus, the camera pose can be written as: $P = KR[I_{3\times3} \; -C]$

There four configurations can be written as:

1. $C_1 = U(:,3)$ and $R_1 = UWV^T$
2. $C_2 = -U(:,3)$ and $R_2 = UWV^T$
3. $C_3 = U(:,3)$ and $R_3 = UW^TV^T$
4. $C_4 = -U(:,3)$ and $R_4 = UW^TV^T$

can be computed from E matrix

let $E = UDV^T$

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

Check for Cheirality Condition using Triangulation

We computed four different possible camera poses for a pair of images using essential matrix. Though, in order to find the *correct* unique camera pose, we need to remove the disambiguity. This can be accomplish by checking the **cheirality condition** *i.e. the reconstructed points must be in front of the cameras*. To check the cheirality condition, triangulate the 3D points (given two camera poses) using **linear least squares** to check the sign of the depth Z in the camera coordinate system w.r.t. camera center. A 3D point X is in front of the camera iff: $r_3 (X-C) > 0$ where r_3 is the third row of the rotation matrix (z-axis of the camera). Not all triangulated points satisfy this coniditon due of the presence of correspondence noise. The best camera configuration, (C, R, X) is the one that produces the maximum number of points satisfying the cheirality condition.

Check for Cheirality Condition using Triangulation

The triangulation linear solution is as following

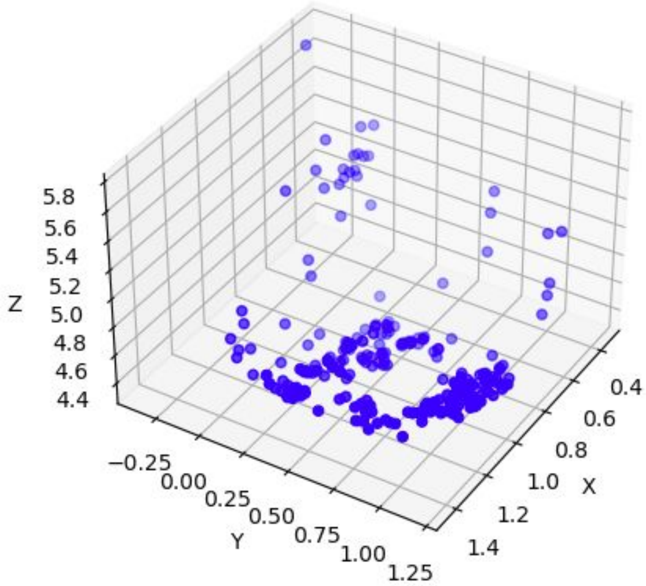
$$x = w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \qquad P = \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix} \qquad A = \begin{bmatrix} u p_3^T - p_1^T \\ v p_3^T - p_2^T \\ u' p_3'^T - p_1'^T \\ v' p_3'^T - p_2'^T \end{bmatrix}$$

$$x' = w\begin{bmatrix} u \\ v' \\ 1 \end{bmatrix}$$

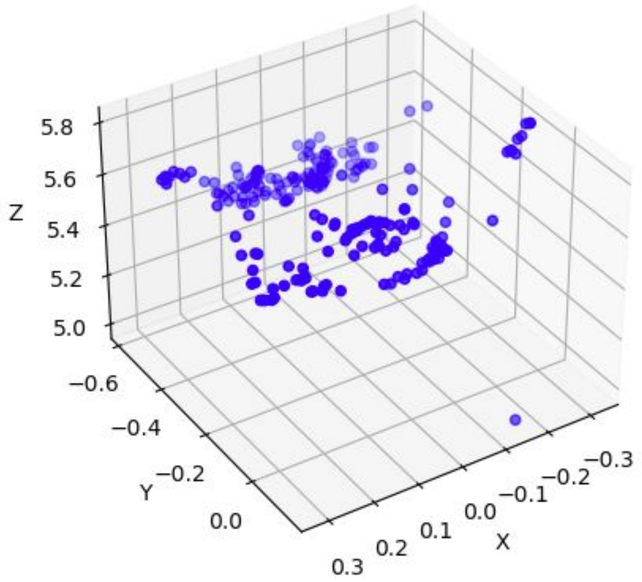
Given P, P', x, x'

1. precondition points and projection matrix
2. create matrix A
3. $[U, S, V] = \mathrm{svd}(A)$
4. $x = V(:, end)$

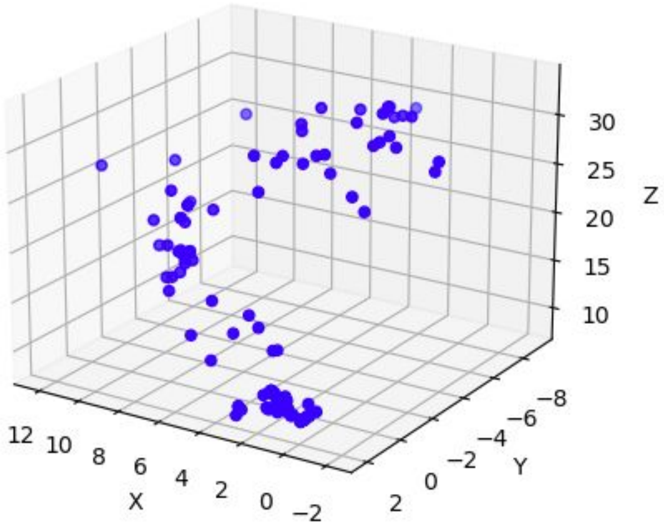
Result - Statue 3D points



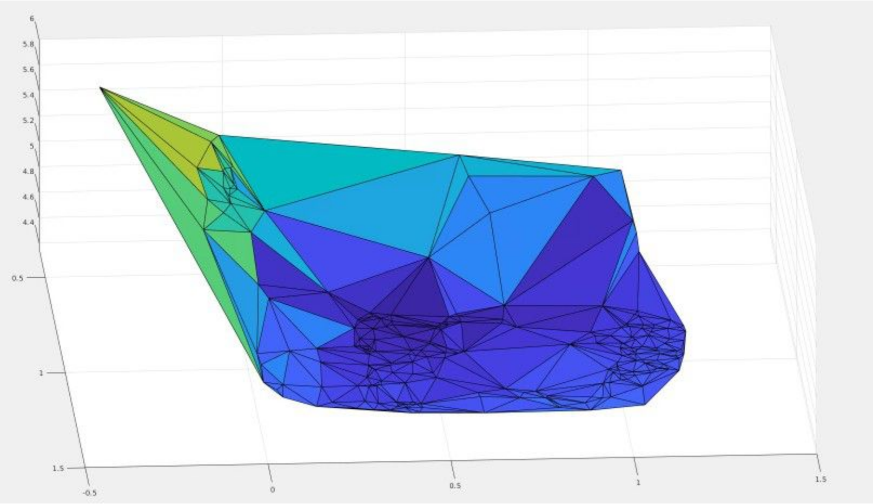
Result - Statue 3D points



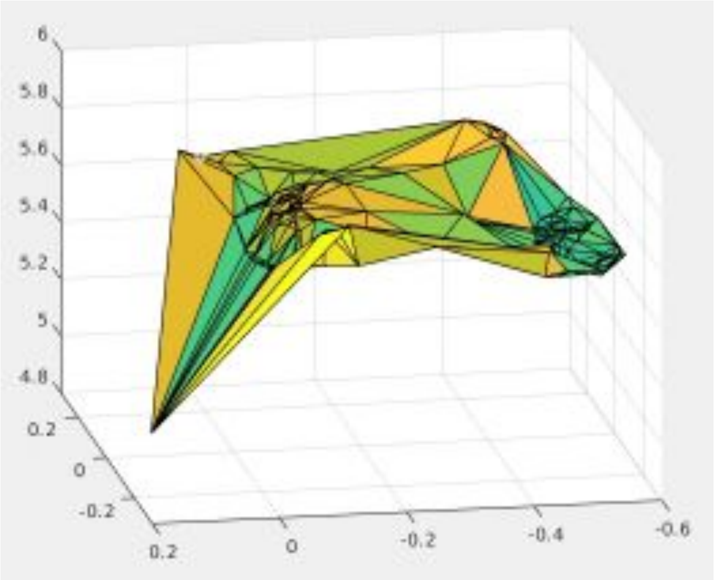
Result - our dataset



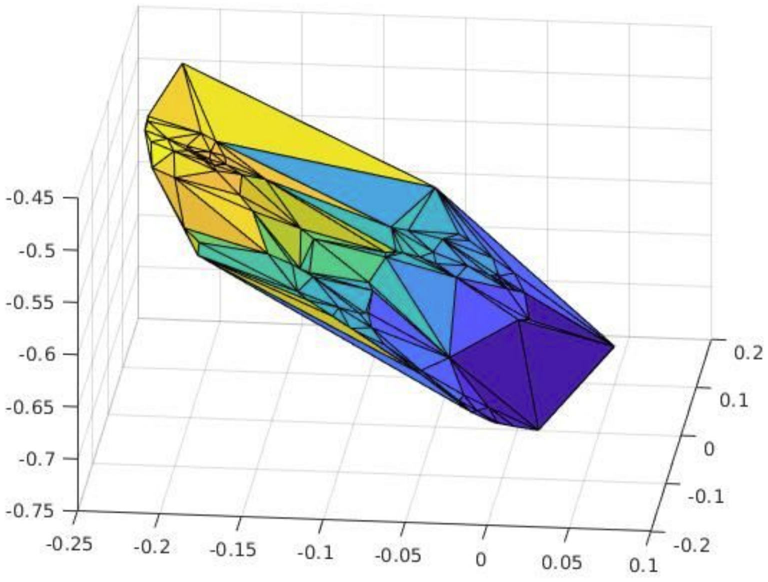
Result - Mesona 3D model



Result - Statue 3D model



Result - our dataset



Discussion & conclusion

We demonstrate and implement the Structure-from-Motion pipeline which works well and creates decent 3d points for the 2 images.

The experimental results did not reach satisfactory results in our dataset due to the scenario is more complicated.