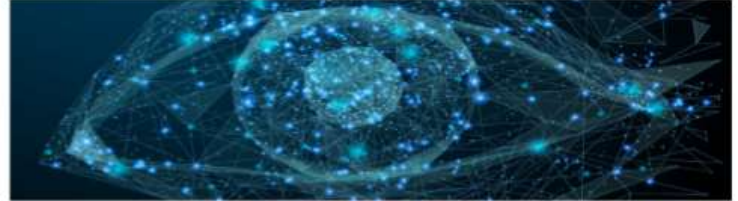




Computer Vision
16720-B Fall 2022



16720 (B) 3D Reconstruction - Assignment 5

Instructor: Kris

TAs: Arka, Jinkun, Rawal, Rohan, Sheng-Yu

Instructions

This section should include the visualizations and answers to specifically highlighted questions from P1 to P4. This section will need to be uploaded to gradescope as a pdf and manually graded (this is a separate submission from the coding notebooks).

1. Students are encouraged to work in groups but each student must submit their own work. Include the names of your collaborators in your write up. Code should **Not** be shared or copied. Please properly give credits to others by **LISTING EVERY COLLABORATOR** in the writeup including any code segments that you discussed, Please **DO NOT** use external code unless permitted. Plagiarism is prohibited and may lead to failure of this course.
2. **Start early!** This homework will take a long time to complete.
3. **Questions:** If you have any question, please look at Piazza first and the FAQ page for this homework.
4. All the theory question and manually graded questions should be included in a single writeup (this notebook exported as pdf or a standalone pdf file) and submitted to gradescope: pdf assignment.
5. **Attempt to verify your implementation as you proceed:** If you don't verify that your implementation is correct on toy examples, you will risk having a huge issue when you put everything together. We provide some simple checks in the notebook cells, but make sure you verify them on more complicated samples before moving forward.
6. **Do not import external functions/packages other than the ones already imported in the files:** The current imported functions and packages are enough for you to complete this assignment. If you need to import other functions, please remember to comment them out after submission. Our autograder will crash if you import a new function that the gradescope server does not expect.
7. Assignments that do not follow this submission rule will be **penalized up to 10% of the total score.**

Theory Questions (25 pts)

Before implementing our own 3D reconstruction, let's take a look at some simple theory questions that may arise. The answers to the below questions should be relatively short, consisting of a few lines of math and text (maybe a diagram if it helps your understanding).

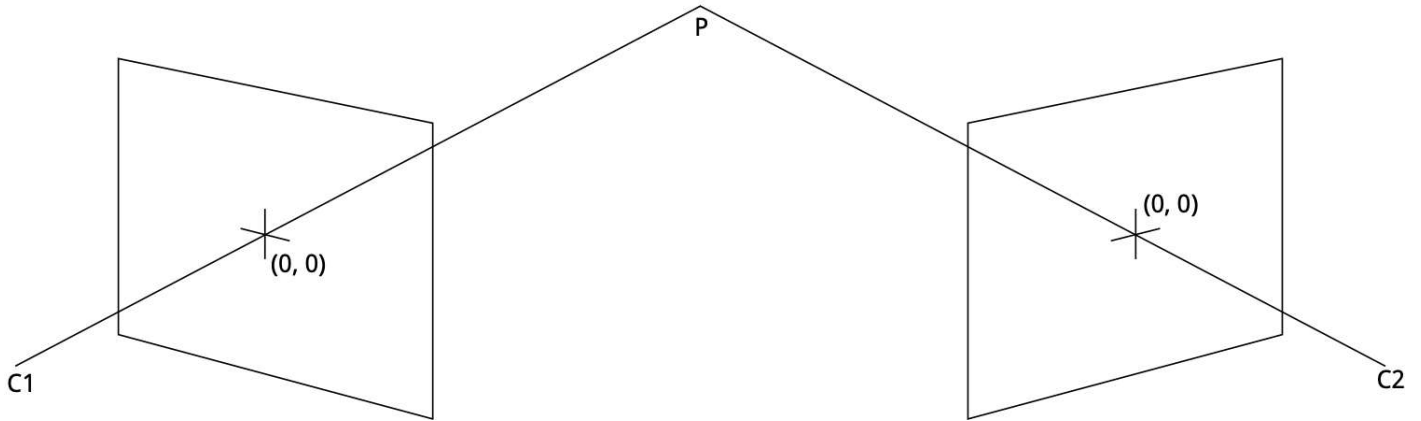


Figure1. Figure for Q1.1. $C1$ and $C2$ are the optical centers. The principal axes intersect at point w (P in the figure).

Q0.1

Suppose two cameras fixated on a point x (see Figure 1) in space such that their principal axes intersect at the point P . Show that if the image coordinates are normalized so that the coordinate origin $(0, 0)$ coincides with the principal point, the F_{33} element of the fundamental matrix is zero.

The fundamental matrix is given by $F = \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix}$

For a fundamental matrix, we know that $x^T F x = 0$.

Here $x^T = [0 \ 0 \ 1]$, since the coordinate origin is at principal point.

$$\begin{aligned}
 & x^T F x = 0 \\
 & [0 \ 0 \ 1] \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0 \\
 & [F_{31} \ F_{32} \ F_{33}] \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0 \\
 & F_{33} = 0
 \end{aligned}$$

Q0.2

Consider the case of two cameras viewing an object such that the second camera differs from the first by a pure translation that is parallel to the x -axis. Show that the epipolar lines in the two cameras are also parallel to the x -axis. Backup your argument with relevant equations.

Assume the translation vector is given by $t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$ and since the translation is parallel to the x-axis, we have

t_2 and $t_3 = 0$. Since we have pure translation, the rotation matrix R is an identity, that is,

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We know that for the essential matrix E are given by $x_2^T E x_1 = 0$ and $l_2 = E x_1$, where $E = T_X R$, where T_X

is the cross-product matrix of the translation vector, given by $T_X = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$. Since t_2 and t_3 are

zero (parallel translation to the x-axis), we get, $T_X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix}$. Now,

$$\begin{aligned} E &= T_X R \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \end{aligned}$$

The epipolar line l_2 in the second image plane corresponding to the point x_1 in the first image is given by,

$$\begin{aligned} l_2 &= E x_1 \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_1 \\ 0 & t_1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -t_1 \\ t_1 y_1 \end{bmatrix} \end{aligned}$$

The line equation of the epipolar line passing through any point x_2 is given by $x_2^T l_2 = 0$.

$$\begin{aligned} \begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -t_1 \\ t_1 y_1 \end{bmatrix} &= 0 \\ -t_1 y_2 + t_1 y_1 &= 0 \end{aligned}$$

We can clearly see from the above equation that the epipolar line is parallel to x-axis. Similarly it can also be proved to be the case for the epipolar line in the first image (l_1). Thus, the epipolar lines in the two cameras are also parallel to the x-axis for pure translation along the x-axis.

Q0.3

Suppose we have an inertial sensor which gives us the accurate extrinsics \mathbf{R}_i and \mathbf{t}_i (see Figure 2), the rotation matrix and translation vector of the robot at time i . What will be the effective rotation (\mathbf{R}_{rel}) and translation (\mathbf{t}_{rel}) between two frames at different time stamps? Suppose the camera intrinsics (\mathbf{K}) are known, express the essential matrix (\mathbf{E}) and the fundamental matrix (\mathbf{F}) in terms of \mathbf{K} , \mathbf{R}_{rel} and \mathbf{t}_{rel} .

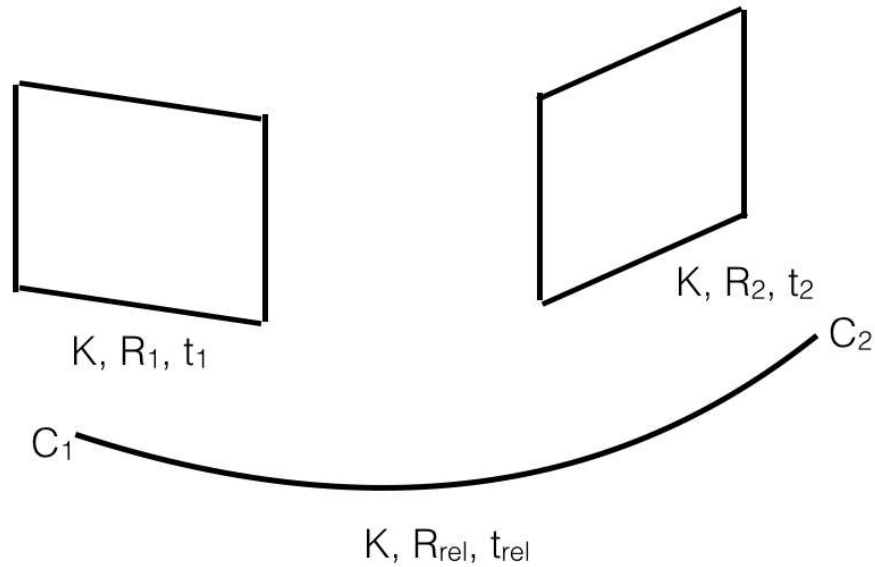


Figure 2. Figure for Q1.3. C_1 and C_2 are the optical centers. The rotation and the translation is obtained using inertial sensors. \mathbf{R}_{rel} and \mathbf{t}_{rel} are the relative rotation and translation between two frames.

Let X denote the point in the real world and x_i, y_i represent any point in the image. We know,

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = K(R_i | t_i) \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Consider two points x_1 and x_2 at the two time frames,

$$\begin{aligned} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} &= K(R_1 | t_1) \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= K \left(R_1 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t_1 \right) \\ \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} &= K \left(R_2 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t_2 \right) \\ \Rightarrow \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} &= R_2^{-1} \left(K^{-1} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} - t_2 \right) \end{aligned}$$

Substituting the above equation in the point equation for the first frame, we get,

$$\begin{aligned} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} &= K \left(R_1 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t_1 \right) \\ &= K \left(R_1 \left(R_2^{-1} \left(K^{-1} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} - t_2 \right) \right) + t_1 \right) \\ &= K R_1 R_2^{-1} K^{-1} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} - K R_1 R_2^{-1} t_2 + K R_1 t_1 \end{aligned}$$

We can see from the above equation, t_{rel} and R_{rel} are given by,

$$t_{rel} = -K R_1 R_2^{-1} t_2 + K R_1 t_1$$

$$R_{rel} = K R_1 R_2^{-1} K^{-1}$$

and the fundamental matrix F and essential matrix E are given by,

$$E = t_{relX} R_{rel}$$

$$F = K^{-1} E K$$

$$= K^{-1} t_{relX} R_{rel} K$$

Here, t_{relX} is the cross-product matrix of the translation vector t_{rel} .

Q0.4

Suppose that a camera views an object and its reflection in a plane mirror. Show that this situation is equivalent to having two images of the object which are related by a skew-symmetric fundamental matrix. You may assume that the object is flat, meaning that all points on the object are of equal distance to the mirror (**Hint:** draw the relevant vectors to understand the relationship between the camera, the object, and its reflected image.)

Assume 3D point P and its reflection P' in the mirror. There is no rotation, only translation between the two points. They are related by the homogeneous transformation

$$P = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} P'$$

The image points p and p' corresponding to P and P' must satisfy the epipolar correspondence.

$$p' F p = 0$$

Here $F = K E$, where E is the essential matrix

$$p' K E p = 0$$

Here $E = [t] R$ where t is the translation and R is the rotation

$$t = [t_x \quad t_y \quad t_z]'$$

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Rightarrow E = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

Taking the transpose of E

$$E^T = \begin{bmatrix} 0 & t_z & -t_y \\ -t_z & 0 & t_x \\ t_y & -t_x & 0 \end{bmatrix}$$

Since $E^T = -E$ E is a skew symmetric matrix

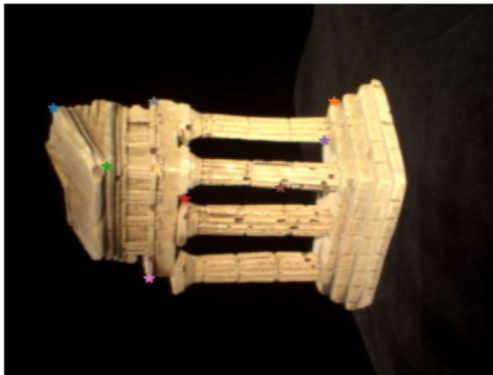
Coding Questions (30 pt)

Q1.1: The Eight Point Algorithm

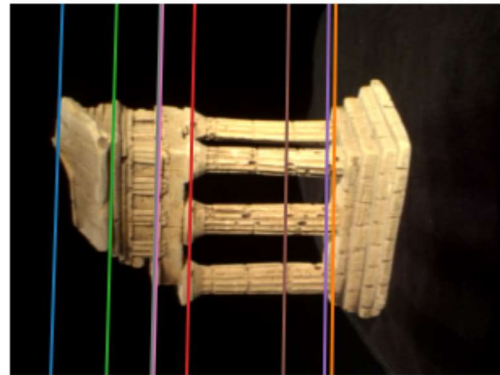
****Output:**** In your write-up: Write your recovered \mathbf{F} and include an image of some example outputs of `displayEpipolarF`.

$$F = \begin{bmatrix} -0. & 0. & -0.2811 \\ 0. & 0. & -0.0099 \\ 0.2713 & 0.0046 & 1. \end{bmatrix}$$

Select a point in this image



Verify that the corresponding point is on the epipolar line in this image

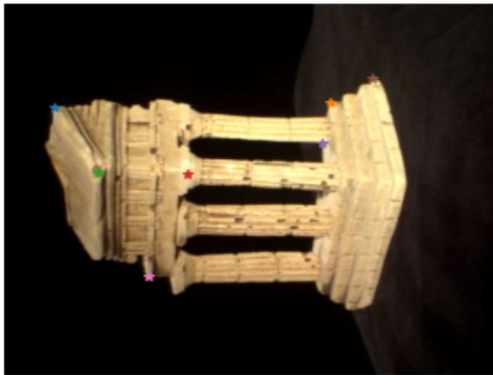


Q1.2: The Seven Point Algorithm

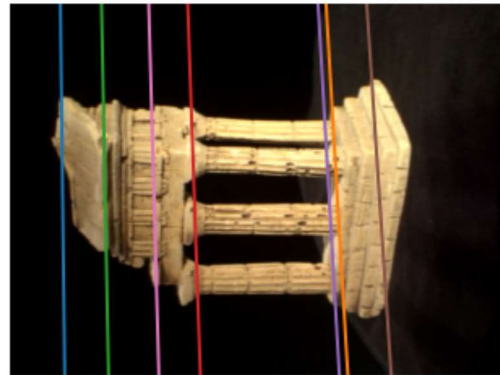
****Output:**** In your write-up: Print your recovered \mathbf{F} and include an image output of `displayEpipolarF`.

$$F = \begin{bmatrix} 0. & 0. & -0.201 \\ 0. & -0. & 0.0007 \\ 0.1922 & -0.0042 & 1. \end{bmatrix}$$

Select a point in this image



Verify that the corresponding point is on the epipolar line in this image



Q2.2 Triangulation and find M2

****Output:**** In your write-up: Write down the expression for the matrix \mathbf{A}_i

$$\mathbf{A}_i = \begin{bmatrix} y_i * (C13) - C12 \\ C11 - x_i * (C13) \\ y'_i * (C23) - C22 \\ C21 - x'_i * (C23) \end{bmatrix}$$

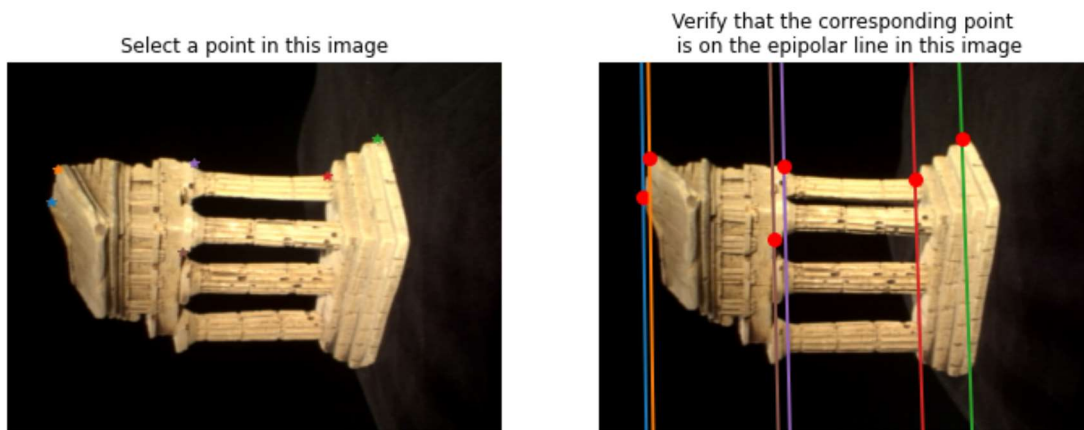
\mathbf{A}_i is a 4×4 matrix

where (x_i, y_i) and (x'_i, y'_i) are 2 corresponding points in im1 and im2 respectively

C_{ij} is the j^{th} row of the i^{th} camera

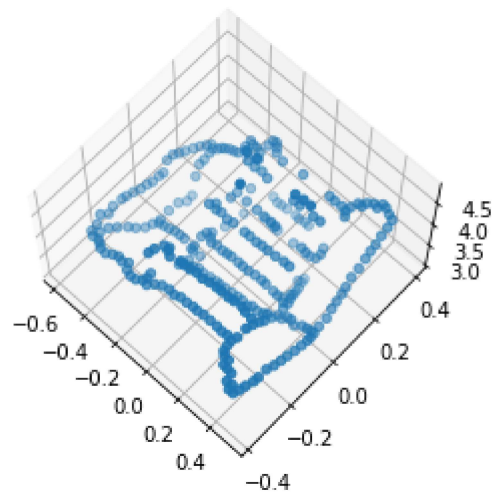
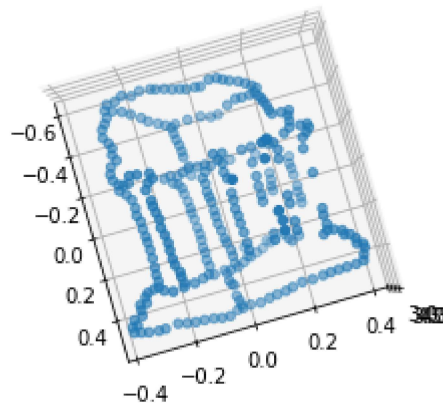
Q2.3 Epipolar Correspondence

****Output:**** In your write-up, include a screenshot of `epipolarMatchGUI` with some detected correspondences.



Q2.4 3D Visualization

****Output:**** In your write-up: Take a few screenshots of the 3D visualization so that the outline of the temple is clearly visible.



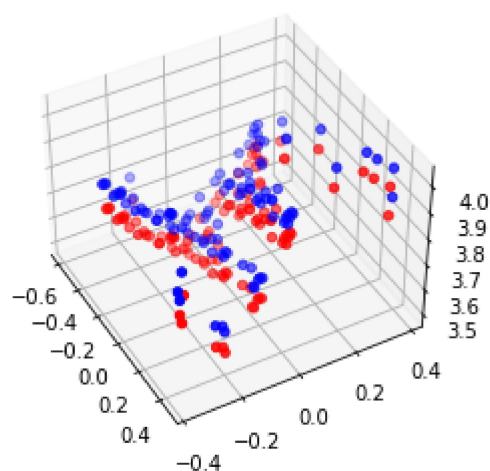
Q3.3 Bundle Adjustment

****Output:**** In your write-up: include an image of output of the `plot_3D_dual` function by passing in the original 3D points and the optimized points. Also include the before and after reprojection error for the `rodriguesResidual` function.

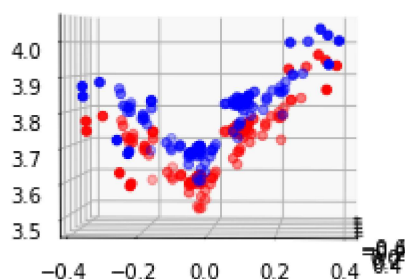
Reprojection error

Before: 352.8419, After: 10.8869

Blue: before; red: after



Blue: before; red: after



Blue: before; red: after

