Homework 3 Questions

Instructions

- 6 questions.
- Write code where appropriate; feel free to include images or equations.
- Please make this document anonymous.
- This assignment is **fixed length**, and the pages have been assigned for you in Gradescope. As a result, **please do NOT add any new pages**. We will provide ample room for you to answer the questions. If you *really* wish for more space, please add a page *at the end of the document*.
- **We do NOT expect you to fill up each page with your answer.** Some answers will only be a few sentences long, and that is okay.

Questions

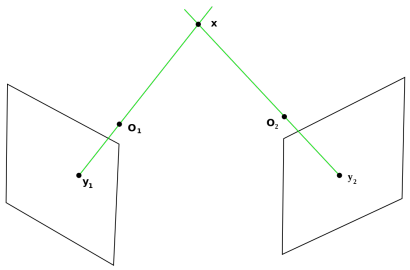
Q1: Given a stereo pair of cameras:

(a) Briefly describe triangulation (using images might be helpful).

(b) Why is it not possible to find an absolute depth for each point when we don't have calibration information for our cameras?

A1: Your answer here.

(a) Triangulation refers to the process of determining a point in 3D space given its projections onto two, or more, images. If a pair of corresponding points in two images can be found, it must be an intersection of 2 3D lines, which is a common 3D point.



The above picture describes x as intersection of 2 green lines and x is projected onto two 3D planes through O1, O2 focal points. The resulting image points are y1 and y2.

A1 (continued): Your answer here.

(b) Because any project is inherently ambiguous and there will be a space of possible solutions that can project a 3D point to a 2D image point. And Least Square Approximation only reduces the error when we have data points to calculate possible calibration matrices. Therefore, if we do not calibration information for our cameras, we can only approximate the possible solution of the projection process, but we cannot get the absolute depth for each point. Also, since the projection onto 2D image is depth scaled, if we do not have calibration information, we can not know the actual depth of each point in the image.

Q2: In two-view camera geometry and depth estimation:

(a) Why does rectification simplify matching features across our stereo image pair?

(b) What information do we need to know to rectify our image pair?

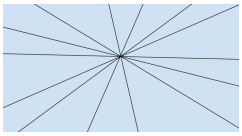
A2: Your answer here.

(a) Image rectification is a transformation process used to project images onto a common image plane. Image rectification warps both images such that they appear as if they have been taken with only a horizontal displacement and as a consequence all epipolar lines are horizontal.

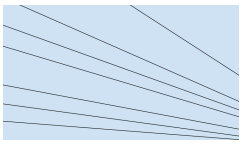
(b) We need to intrinsic parameters and extrinsic parameters of cameras. Intrinsic parameters describe a mapping between camera coordinates and pixel coordinates in the image frame, such as focal length. Extrinsic parameters define the location and orientation of the camera with respect to the world frame, such as baseline distance T in parallel camera case and R,t in non-parallel case.

Q3: In two-view camera geometry, what does it mean when the epipolar lines:

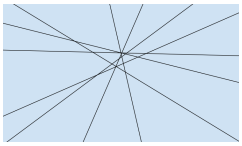
(a) radiate out of a point on the image plane,



(b) converge to a point outside of the image plane, and



(c) intersect at more than one point?



We highly recommend using this interactive demo to explore the different scenarios and get a better feel for epipolar geometry.

A3: Your answer here.

(a) It means that two cameras and the point in the scene lie on the same line. It happens when one view is forward of the other.

(b) It means that two cameras are not in the same plane. The point in the scene and 2 points on the image form a plane, and they land on the same line. Usually in the case of moving forward a camera while taking pictures.

(c) It means that two cameras are similar to the scenario of (b) but the camera moves horizontally while moving forward so that points in the scene and points in the camera does not all land on the same line.

Q4: Suppose that we have the following three datasets of an object of unknown geometry:

(a) A video circling the object;

(b) An stereo pair of calibrated cameras capturing two images of the object; and

(c) Two images we take of the object at two different camera poses (position and orientation) using the same camera but with different lens zoom settings.

1. For each of the above setups, decide if we are able to find/calculate the essential matrix, the fundamental matrix, or both.
   LaTeX: To fill in boxes, replace '\square' with '\blacksquare' for your answer.

(a)	Essential Matrix	☐
	Fundamental Matrix	■
	Both	☐
(b)	Essential Matrix	☐
	Fundamental Matrix	☐
	Both	■
(c)	Essential Matrix	☐
	Fundamental Matrix	■
	Both	☐

2. State an advantage and disadvantage of using each setup for depth reconstruction; and

3. Name an application scenario for each of the different setups.

A4: Your answer here.

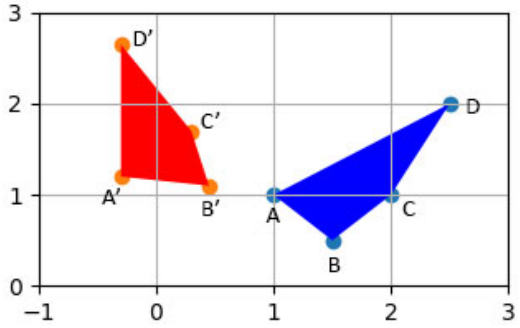
(a) It can recover 3D depth with actual depth but it needs lots of calculation between each frames, which costs lots of computation resources and time. It is also severly impacted by noices. SfM (Struction from motion) photogrammetry, which recover 3D landscapes.

(b) It can recover 3D with depth, which means the depth are relatively captured in reconstruction. But it can be fast and saves computation power. Photo tourism system is an application.

(c) It does not require calibrated parameters and can be computed very fast to match points and recover depth.It can be used for real-time 3d reconstruction.

A4 (continued): Your answer here.

Q5 (Linear algebra/numpy question): Suppose we have a quadrilateral $ABCD$ and a transformed version $A'B'C'D'$ as seen in the image below.



$$\begin{aligned} A &= (1, 1) & A' &= (-0.3, 1.3) \\ B &= (1.5, 0.5) & B' &= (0.5, 1.1) \\ C &= (2, 1) & C' &= (0.3, 1.8) \\ D &= (2.5, 2) & D' &= (-0.3, 2.6) \end{aligned} \tag{1}$$

Let's assume that each point in $ABCD$ was approximately mapped to its corresponding point in $A'B'C'D'$ by a 2×2 transformation matrix M.

e.g. if $A = \begin{pmatrix} x \\ y \end{pmatrix}$ and $A' = \begin{pmatrix} x' \\ y' \end{pmatrix}$, and $\boldsymbol{M} = \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix}$

then $\begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix} * \begin{pmatrix} x \\ y \end{pmatrix} \approx \begin{pmatrix} x' \\ y' \end{pmatrix}$

We would like to approximate $\boldsymbol{M}$ using least squares for linear regression.

(a) Rewrite the equation $\boldsymbol{M}x \approx x'$ into a pair of linear equations. We have provided you with a template of what they should look like below.

(b) Use the equations you wrote for part (a) and coordinate values for $ABCD$ and $A'B'C'D'$ to construct a matrix $\boldsymbol{Q}$ and column vector b that satisfy

$$\boldsymbol{Q} * \begin{pmatrix} m_{1,1} \\ m_{1,2} \\ m_{2,1} \\ m_{2,2} \end{pmatrix} = b \tag{2}$$

We have provided you with a template of what they should look like below.

Hint: You have a pair of equations for each x-x' correspondence, giving you 8 rows in $\boldsymbol{Q}$ and b. If you're having trouble, try writing out the equations for each pair of points like in part (a).

Note: Systems of linear equations are typically written in the form $\boldsymbol{A}x = b$, but since we have already defined A and x, we're writing it as $\boldsymbol{Q}m = b$

(c) Our problem is now over-constrained, so we want to find values for $m_{i,j}$ that minimize the squared error between approximated values and real values, or $||\boldsymbol{Q}m - b||_2$. To do this we use singular value decomposition to find the pseudoinverse of $\boldsymbol{Q}$, written as $\boldsymbol{Q}^\dagger$. We then multiply it by both sides, giving us $\boldsymbol{Q}^\dagger\boldsymbol{Q}m = \boldsymbol{Q}^\dagger b \quad \Rightarrow \quad m \approx \boldsymbol{Q}^\dagger b$.

Thankfully, the computer can do all of this for us! `numpy.linalg.lstsq()` takes in our $\boldsymbol{Q}$ matrix and b vector, and returns approximations for m. Plug the values you wrote in part (b) into that function and write the returned $\boldsymbol{M}$ matrix here.

A5: Your answer here.

(a) Replace each of the '__' below with x, y, x', y', or 0.

$$\begin{cases} xm_{1,1} + ym_{1,2} + 0m_{2,1} + 0m_{2,2} = x' \\ 0m_{1,1} + 0m_{1,2} + xm_{2,1} + ym_{2,2} = y' \end{cases} \tag{3}$$

(b) Replace each of the '__' below with a 0 or a coordinate value from $ABCD$ and $A'B'C'D'$.

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1.5 & 0.5 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 2.5 & 2 & 0 & 0 \\ 0 & 0 & 2.5 & 2 \end{pmatrix} * \begin{pmatrix} m_{1,1} \\ m_{1,2} \\ m_{2,1} \\ m_{2,2} \end{pmatrix} = \begin{pmatrix} -0.3 \\ 1.3 \\ 0.5 \\ 1.1 \\ 0.3 \\ 1.8 \\ -0.3 \\ 2.6 \end{pmatrix} \tag{4}$$

(c) Replace each of the '__' below with the value of $m_{i,j}$.

$$M = \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix} = \begin{pmatrix} 0.632 & -0.9408 \\ 0.528 & 0.6786 \end{pmatrix} \tag{5}$$

Q6: Goal: Be aware of some of the implicit assumptions that shape our views and decisions regarding photography technology and discuss how some of the challenges in photography are part of a longer history of technological development and social change.

A common problem in photography and videography is properly calibrating lighting for a camera to capture a wide variety of skin tones. This issue stems from an inherited bias built into photography that light skin as the norm. Watch this video on Kodak's Shirley Card and answer the following questions.

(a) What kind of harm can come to marginalized or minority populations that aren't considered when building imaging technology? Are we seeing any of these effects today? (3–4 sentences)

(b) Read about how some researchers are trying to take the racial bias out of computer vision applications here. If you were developing and launching a new camera, what steps would you take to eliminate possible racial bias? (3–4 sentences)

(c) It took chocolate and wood corporations complaining that their products couldn't be photographed correctly for Kodak to adjust its film to be more inclusive of other skin tones. Who should be involved in shaping the direction of new photography developments in the future (companies, developers, interest groups etc)? (3–4 sentences)

A6: Your answer here.

(a) If minority populations are not considered when building imaging technology, their aesthetics will be misled and they will feel self-abasement. Also, they will be misrepresented in the social media or other media forms. There were a AI technology called PULSE (Photo Upsampling via Latent Space Exploration) that was trained mostly on white population images accidentally modified black people's faces to white in its output.

(b) Firstly, I will make sure that the camera imaging technology treat different colors in the color map in the same general way without biases on any specific range of colors. Then I will invite people with different social backgrounds, sociologists, psychologists and other scientists in related fields to participate in the test of the effects that my camera shows on difference races. In the end, I may try to test my camera with certain machine learning algorithms to see if the algorithm show any bias toward a certain race.

(c) People from different social background should be involved in shaping the new photography technologies. Only programmers and designers involved will generate biases since they themselves have stereotypical views. Specifically, food industry, clothing industry and industries that color matters in their products should participate. Developers from different races and different communities should also participate.

A6 (continued): Your answer here.

Feedback? (Optional)

Please help us make the course better. If you have any feedback for this assignment, we'd love to hear it!