CS231a 2023 Midterm

Stanford University

02/13/2023

Name: ______________________________

SUNet ID: ______________________@stanford.edu

Question	Points
T/F	
MC	
SA1	
SA2	
SA3	
SA4	
Total	

Instructions:

1. This exam contains 8 pages, including this page.
2. There are 10 True/False questions (2 points each), 5 multiple choice questions (4 points each), and 4 short answer questions (15 points each), totaling 100 points.
3. You have **80 minutes** to complete the midterm.
4. The midterm is open book and open notes, but closed internet. The use of the internet or any online sources during the exam is strictly forbidden.
5. If you are unsure about a problem statement when taking the exam, state your assumptions. We will take all reasonable assumptions into account when grading.
6. Please sign the below Honor Code statement.

In recognition of and in the spirit of the Stanford University Honor Code, I certify that I will neither give nor receive unpermitted aid on this examination.

Signature: ______________________________

True/False

Answer the below questions by filling in the circle for either T or F.

1. ◯ T ◯ F — Similarity transformations preserve the distances between points.
2. ◯ T ◯ F — Projective transformations map points at infinity to points no longer at infinity.
3. ◯ T ◯ F — When solving the structure-from-motion problem with the factorization method, one limitation is that the reconstruction could be different from the correct reconstruction by a similarity transformation. Bundle adjustment is a non-linear method that addresses this limitation.
4. ◯ T ◯ F — In the correspondence problem, having a small ratio of baseline/z-value is always desirable as it helps reduce errors in depth estimation.
5. ◯ T ◯ F — The factorization method first centers the image points with respect to the center of the image plane, then centers the 3D points with respect to the center of the world reference system.
6. ◯ T ◯ F — It is possible to set up a linear system of equations to calibrate a camera while accounting for radial distortion.
7. ◯ T ◯ F — The result of applying transformations to a vector in the order of rotation, scale, and translation is the same as the result of applying transformations in the order of translation, scale, and rotation.
8. ◯ T ◯ F — Space carving requires knowing the camera intrinsics and extrinsics.
9. ◯ T ◯ F — Space carving produces conservative 3D reconstructions (no smaller than the actual 3D shape).
10. ◯ T ◯ F — Least squares fitting is not robust to outliers.

Multiple Choice

Answer the below questions by filling in one or more squares that are applicable.

1. Which of the following statements do not hold under projective transformation?
   - ☐ Parallel lines remain parallel.
   - ☐ Ratio of areas remains the same.
   - ☐ Collinear points remain collinear.
   - ☐ Ratio of lengths of two parallel line segments remains the same.
2. Given p, p' and the fundamental matrix F, select the thing(s) you can compute.
   - ☐ The epipolar lines l and l'.
   - ☐ The epipoles e and e'.
   - ☐ The camera matrices M and M'.
   - ☐ The 3D point P corresponding to p and p'.

3. Which of the following is/are true about the factorization method?

   - ☐ The measurement matrix has rank 3.
   - ☐ $M = U_3\sqrt{\Sigma_3}$ and $S = \sqrt{\Sigma_3}V_3^T$ is the unique optimal decomposition of the measurement matrix.
   - ☐ Adding more camera views can help recover the scale of the scene.
   - ☐ The 3D points do not need to be centered since the factorization method can only give reconstructions up to a similarity transformation.

4. Select the limitation(s) of bundle adjustment.

   - ☐ Requires 8 or more key points in each camera view.
   - ☐ Does not work for more than two cameras.
   - ☐ Requires all labeled points to be visible in all cameras.
   - ☐ Requires solving a large nonlinear optimization problem.

5. Select the flaw(s) of RANSAC.

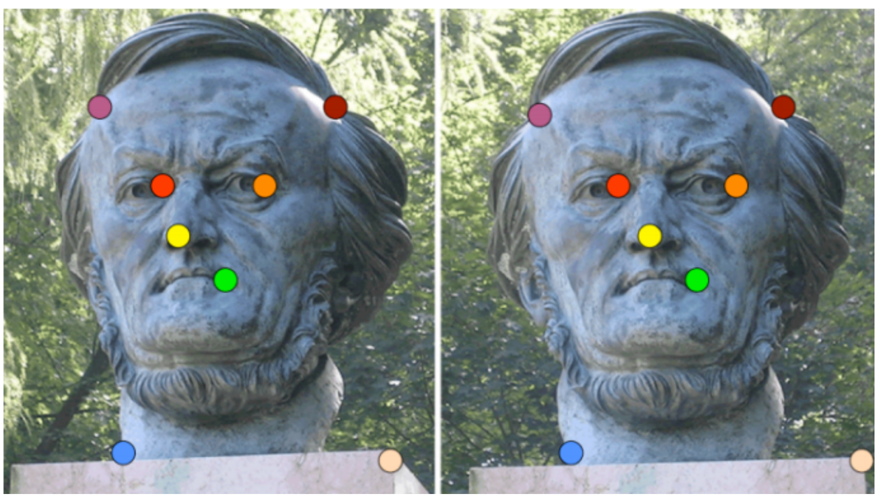
   - ☐ It has multiple parameters to manually tune.
   - ☐ Achieving a better solution requires running it for longer.
   - ☐ It is hard to predict how many iterations will likely lead to a good solution.
   - ☐ It is hard to implement.

Short Answer

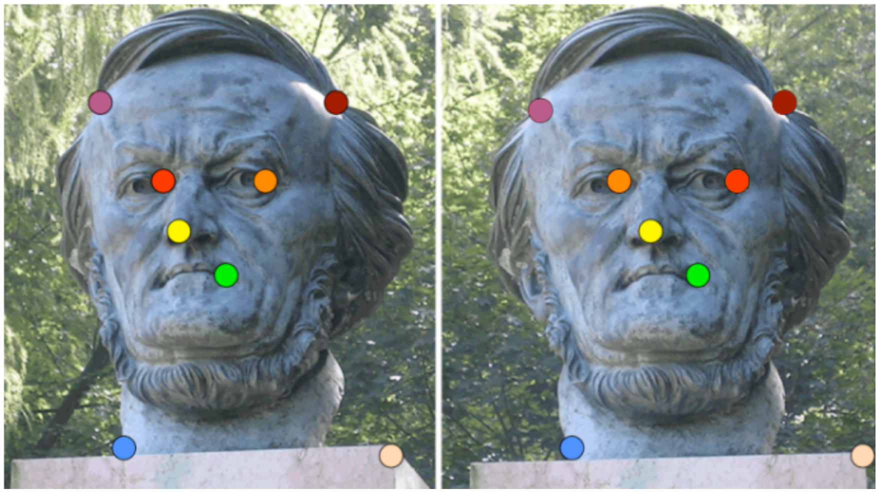
Write your answer to each question in the space below it. You may only need to write a short answer that does not fill up the space.

1. Uses of Point Correspondences

Let's recall the image with 8 corresponding points from the lecture:



Now let's imagine that two pairs of the points were incorrectly matched, as with the orange and red points below (the points on the right and left eyes of the statue):

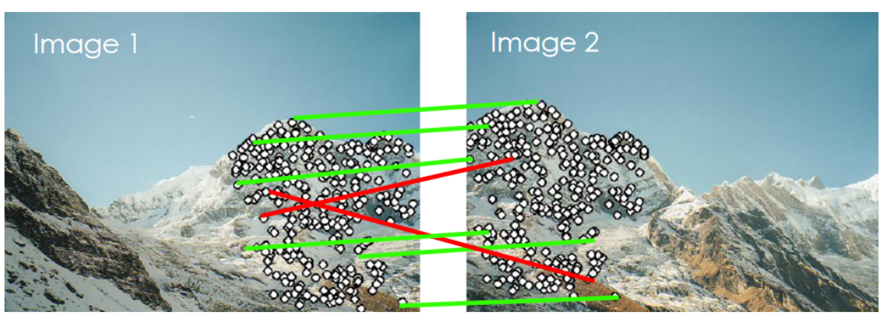


a) Would the epipoles of the images stay the same? How about the epipolar lines of each of the 4 points in these two pairs of incorrect matches? Explain why or why not.

b) We know the eight points algorithm is built around the constraint that $p^T F p' = 0$. Why would this constraint no longer be true for the two incorrectly matched pairs of points? Explain in terms of P, p, and p', and the epipolar plane (you can use plain english, equations are not required).

c) Now let's assume all the correspondences are correct, as in the first image, and we know the camera's extrinsics and intrinsics for each photo. Could we find these points' locations in 3D space? If so, how? Describe the simplest method, assuming there is no noise.

Now let's take a look at two images with many more pairs of points, where some are correctly matched and some are incorrectly matched, as with the green and red lines below:



d) How could you go about deriving the fundamental matrix in this case? Which algorithm would you use? Explain the steps that would be involved.

e) In PSET 2, we create a 3D reconstruction of a statue from several images with knowledge of corresponding points in pairs of the images. What is another approach you could have used to create a 3D reconstruction, without needing point correspondences? State your assumptions, and briefly discuss how the results would differ.

2. Single-View Metrology Reconstruction

Suppose you are given one camera image of a person standing next to the leaning tower of Pisa. The person's height is unknown but the standing height of the leaning tower (the tallest point to the ground) is known to be 55.86 m. The height of one of the levels (highlighted as A) is known to be 5.8m (about 5 "Tuscan arms").

a) Given only one image and the absolute height of the tower and its levels, is it possible to reconstruct the height of the person standing next to it? What additional information, if any, would be needed to compute the person's height?

b) Suppose the intrinsics of the camera to be zero-skew and unit aspect ratio (that is, square pixels). Flipping the problem now, assume the length of the the person's hand to be 20cm (0.2m). Given their hand is the same length as the height of one of the levels in the image (5.8m), what is the relative amount of magnification between the plane the person is standing at relative to the plane of the tower? (Hint, use the weak-perspective camera model).

c) Construct the formula for computing the distance to the tower from the camera knowing the distance of the person to the camera (2m). (You may assume the camera center to be C, and that the camera has a focal length of 1).

d) Suppose the camera does not have the canonical zero-skew unit aspect ratio. How many parallel lines from the image (and their corresponding vanishing points) are needed to compute the camera's intrinsic parameters? Why?

3. Camera Models and Camera Calibration

Suppose the world reference system is the same as the camera reference system and we want to solve for the camera matrix **K** as shown below. We are given n correspondences; each correspondence consists of a 3D scene point (x_i, y_i, z_i) and its image pixel coordinates (u_i, v_i) for i in $1 \cdots n$.

$$\mathbf{K} = \begin{bmatrix} \alpha & 0 & c_x \\ 0 & \beta & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

a) How many unknowns do we have in the camera matrix **K**? What is the minimum number of correspondences we need to solve for the unknowns?

b) Set up a linear system in the form of $\mathbf{Ax} = \mathbf{b}$ to solve for the unknowns in **K**. Please write down the matrices **A**, **x** and **b** in terms of x_i, y_i, z_i, u_i, and v_i (you may use vertical "..." as in the lecture).

c) Now let's say you are given 6 correspondences listed in the table below. Solve for the unknowns and write down the camera matrix $\mathbf{K}$. *Hint: you may not need all 6 correspondences to solve for the unknowns.*

(x_i, y_i, z_i)	(u_i, v_i)
(6, 10, 10)	(1.1, 1.3)
(1, 10, 1)	(1.5, 8.5)
(2, 5, 2)	(1.5, 2.5)
(10, 10, 5)	(2.5, 2.1)
(20, 10, 1)	(20.5, 8.5)

d) We have a new camera that has the same intrinsic parameters as the one you just computed, but with **a skew of** 30°. What is the camera matrix $\mathbf{K}'$ for this new camera? *Hint:* $sin(30°) = \frac{1}{2}$, $cot(30°) = \sqrt{3}$

4. Stereo Systems and Multi-View Geometry

a) Recall that in a stereo camera system, disparity is the difference in pixels positions of the projections of a 3D point onto each camera. Briefly explain why in a rectified stereo pair, disparity is proportional to $\frac{1}{\text{depth}}$.

b) Triangulation gives an estimate of the position of a 3D point $P = [X, Y, Z]$ given its projections $p = [x, y, 1]$, $p' = [x', y', 1]$, and corresponding camera matrices M and M'. Derive the system of linear equations that can be used to find an estimate of P.

c) In the Structure-from-Motion problem, we can reconstruct the 3D geometry of the scene and camera parameters using multi-view correspondences. Given 11 camera views, what is the minimum number of points needed to solve the affine Structure-from-Motion problem?

d) Does solving the SfM problem using the Tomasi and Kanade algorithm give a unique set of camera parameters and scene geometry? If yes, describe the steps to compute the camera parameters and scene geometry. If not, give a counterexample that decomposes the same measurement matrix D into different motion and structure matrices, and describe any additional constraints that can help derive a unique solution.

That's it, you're done! !(ˆˆ)!
Feel free to use this space to doodle.