CS231a 2023 Midterm

Stanford University

02/13/2023

SOLUTIONS

Question	Points
T/F	
MC	
SA1	
SA2	
SA3	
SA4	
Total	

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 will 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.

1. Uses of Point Correspondences

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.

The epipoles stay the same, since they are a function of the camera origin and not points on the images. The epipolar lines corresponding to the points that got swapped will change, since the pixel coordinates of the points changed. Also acceptable to say all epipolar lines changes because the epipolar plane changes, or say epipolar lines dont change because those pixels locations stay the same and we just change the labels of the points.

Rubric:

- 3 Points for correct answer
- -1 Incorrect, vague, or confusing explanation for epipoles not changing or epipolar lines changing
- -1 Says things will change because the computed F will change - question asked if epipoles will change, not the computed epipoles
- -1.5 Incorrectly states epipoles would change or that epipolar lines will stay the same
- 0 Points if incorrect, extremely vague, or blank answer

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).

The equation $p^T F p' = 0$ holds because TFp' is normal to the epipolar plane, and therefore also normal to p^T which lies on the epipolar plane. The mismatched p' no longer lie on the true epipolar plane of P and p, meaning the normal property is no longer valid. Also valid to say p' no longer lies on l'

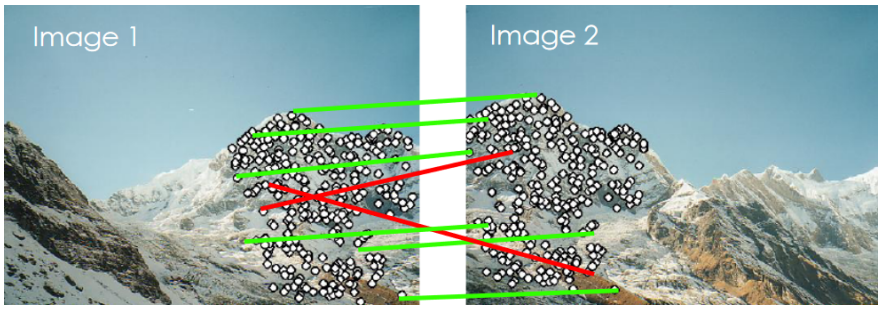
- 3 Points for correct answer
- -1.5 Partially correct or incomplete explanation
- -1.5 Partially correct, but does not discuss epipolar plane or Fp' no longer being the normal vector to pT
- -2.0 Vague explanation, eg just states variables change so the equation no longer holds
- 0 Points if incorrect, extremely vague, or blank answer

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.

The simplest method is to find the intersections of the lines from O_1 to P and O_2 and P (see slide 7 from lecture 5).

- 3 Points for correct answer
- -1.5 Provides valid answer, but not the simplest method (just computing intersection of two lines)
- -1.5 Mentions having M and/or triangulation, but does not explain beyond that
- 0 Points if incorrect, extremely vague, or blank answer

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.

We can use RANSAC to solve for F while avoiding the outliers. To do this, we sample random subsets of 8 matching points, solve for F with the eight point algorithm, and compute the number of other points for which this F is valid for N iterations.

- 3 Points for correct answer
- -1 Mentions RANSAC but not the steps involved
- -2 Says to use eight point algorithm or to solve correspondence issues, does not mention using RANSAC

- 0 Points if incorrect, extremely vague, or blank answer

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.

We can use space carving, if there is a method to find the silhouette of the statue in each view. The result would be voxels instead of a point cloud, and the degree of accuracy will depend on the number of voxels we choose to use. We can also use active stereo if we have a projector.

- 3 Points for correct answer
- -1.5 Says to use single view metrology, which does not do full 3D reconstruction
- -1.5 Valid approach but no or little discussion
- -1.5 Says to find point correspondences via window correlation and/or rectification and not a method that does not use point correspondences at all
- 0 Points if incorrect, extremely vague, or blank answer

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?

Rubric:

- 3.5 Points Correct (A. No. You would need information about the distance from camera to the tower and person)
- -1 Incorrect for not answering if current information is sufficient

- -1 If does not answer what additional info is needed
- -2 Does not correctly answer why cannot estimate height
- 0 Points if incorrect or blank answer

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).

Rubric:

- 4.5 Correct (A. relative magnification = 5.8/0.2 = 29, i.e. the person appears 29x bigger relative to the background tower)
- -0.5, Incorrectly gave magnification of tower relative to person
- -2, Incorrectly calculated relative magnification
- 0 Points if blank/incorrect

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).

- 4.5, Correct (A. 2m / X = 1 / 29, X = 58, or mentions similar triangles)
- -2, Incorrect/missing final calculation, but uses similar triangles
- 0 Points if Blank/incorrect

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?

- 2.5 Correct (A. 4 pairs of parallel lines, 3 only solves for zero-skew unit aspect ratio)
- -1. Guessed +/- 1 more/less lines than needed, but provided justification
- 0 Points if blank/incorrect

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 $\mathbf{K}$ 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 $\mathbf{K}$? What is the minimum number of correspondences we need to solve for the unknowns?

4 unknowns, 2 correspondences.

- -1 Incorrect number of unknowns.
- -1 Incorrect number of correspondences.
- 0 Points if both are incorrect or blank answer

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

$$\begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha x_i + c_x z_i \\ \beta y_i + c_y z_i \\ z_i \end{bmatrix}$$

Note this is in the homogeneous coordinate, to convert it to the euclidean coordinates, we should divide the vector by the last element z_i. Therefore,

$$\alpha \frac{x_i}{z_i} + c_x = u_i$$
$$\beta \frac{y_i}{z_i} + c_y = v_i$$

Then we can construct the linear system $\mathbf{Ax} = \mathbf{b}$ as follows

$$\mathbf{A} = \begin{bmatrix} \frac{x_1}{z_1} & 0 & 1 & 0 \\ 0 & \frac{y_1}{z_1} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_n}{z_n} & 0 & 1 & 0 \\ 0 & \frac{y_n}{z_n} & 0 & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \alpha \\ \beta \\ c_x \\ c_y \end{bmatrix}, \mathbf{b} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_n \\ v_n \end{bmatrix}$$

Variants of the same linear system are also acceptable.

- -4 if derived the equations for each correspondence with error and failed to formulate the correct matrices.
- -2 if derived the equations for each correspondence but failed to formulate in correct matrices.
- -3 if A is incorrect.
- -1 if A is partially correct.
- -1 if x is incorrect.
- -1 if b is incorrect.
- 0 Points if completely incorrect or blank answer

c) Now let's say you are given 5 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)

Select any 2 correspondences from the table and construct the linear system as above. Solve for the unknowns and we can get $\alpha = 1$, $\beta = 0.8$, $c_x = 0.5$, $c_y = 0.5$, so the camera matrix $\mathbf{K}$ is

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 0.8 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}$$

- -1 if 3 of 4 unknowns are correct.
- -2 if 2 of 4 unknowns are correct or listed the correct equations/linear system, but failed to solve it.
- -3 if 1 of 4 unknowns are correct or only listed equations with some error, and failed to solve it.
- 0 Points if all 4 unknowns are incorrect or blank answer.

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*$(30°) = \sqrt{3}$

A general camera matrix can be represented as

$$\begin{bmatrix} \alpha & -\alpha \cot\theta & c_x \\ 0 & \frac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

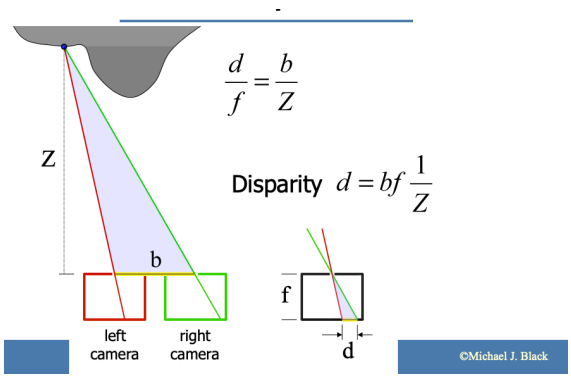
We have computed α, β, c_x, c_y in the previous part. By applying $\sin(30°) = \frac{1}{2}$, $\cot(30°) = \sqrt{3}$, we can get the new camera matrix.

$$\mathbf{K}' = \begin{bmatrix} 1 & -\sqrt{3} & 0.5 \\ 0 & 1.6 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}$$

- -1 if have some minor error or 1 of the values is incorrect.
- -2 if correctly include the skewness in the camera matrix but fails to compute the matrix.
- -3 if generally knows how to include skewness in the camera matrix with some minor error and fails to compute the matrix.
- 0 Points if completely incorrect or blank answer.

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 $\frac{1}{\text{depth}}$.



In a rectified stereo pair, disparity is described by the difference in x-axis of the projected points. By similar triangles, as shown in the figure, we have that $\frac{d}{f} = \frac{b}{Z}$, thus $d = \frac{bf}{Z}$, thus d is proportional to $\frac{1}{Z}$.

- -1.5 if
  - Correct reasoning but didn't state the formulation of the similar triangles.
  - Incorrect formulation of the similar triangles.
  - Correct formulation but missing explanation.
- -3 if completely incorrect or blank.

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.

We have $p = MP$, so $p \times (MP) = 0$. We could then formulate the following constraints from the cross-product:

$$x(M_3P) - (M_1P) = 0$$
$$y(M_3P) - (M_2P) = 0$$
$$x(M_2P) - y(M_1P) = 0$$

Doing the same for the other image, we obtain the linear equation $AP = 0$, where

$$A = \begin{bmatrix} xM_3 - M_1 \\ yM_3 - M_2 \\ x'M'_3 - M'_1 \\ y'M'_3 - M'_2 \end{bmatrix}$$

- Full credit for correct formulation of the linear equation and matrix A.
- -3 for incorrect formulation or blank answers.

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?

For affine SfM, we want $2mn \geq 8m + 3n - 8$. Given $m = 11$, we have $22n \geq 80 + 3n$, $n \geq \frac{80}{19}$, thus we need at least 5 points.

- -1 for minor mistakes (calculation error, incorrect formula for the ambiguity part, etc.).
- -2 for correct intuition in reasoning but no calculation for the final answer.
- -3 for major mistakes (misused m and n, or other major errors in calculation).
- -4 for completely incorrect or blank answers.

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.

No. $D = MS$ and $D = (MH)(H^{-1}S)$ are both valid decompositions. Examples of additional constraints that could help include absolute positions of some 3D points, or the measurements of certain objects in the scene, etc.

- -2 if didn't give example of ambiguity in decomposition.
- -2 if didn't give examples of additional constraints that can help derive a unique solution, or example still allows for similarity ambiguity.
- -5 for completely incorrect or blank answers.

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