PROBLEM SET 5

16822 GEOMETRY-BASED METHODS IN VISION (FALL 2022)
https://piazza.com/cmu/fall2022/16822

OUT: Nov. 8, 2022
DUE: Nov. 15, 2022 11:59 PM
Instructor: Shubham Tulsiani
TAs: Mosam Dabhi, Kangle Deng, Jenny Nan

START HERE: Instructions

- **Collaboration policy:** All are encouraged to work together BUT you must do your own work (code and write up). If you work with someone, please include their name in your write up and cite any code that has been discussed. If we find highly identical write-ups or code without proper accreditation of collaborators, we will take action according to university policies, i.e. you will likely fail the course. See the Academic Integrity Section detailed in the initial lecture for more information.
- **Late Submission Policy:** There are **no** late days for Problem Set submissions.
- **Submitting your work:**
  - We will be using Gradescope (https://gradescope.com/) to submit the Problem Sets. Please use the provided template. Submissions can be written in LaTeX. Regrade requests can be made, however this gives the TA the opportunity to regrade your entire paper, meaning if additional mistakes are found then points will be deducted. Each derivation/proof should be completed on a separate page. For short answer questions you **should** include your work in your solution.
- **Materials:** The data that you will need in order to complete this assignment is posted along with the writeup and template on Piazza. s

Instructions for Specific Problem Types

For "Select One" questions, please fill in the appropriate bubble completely:

Select One: Who taught this course?

- ● Shubham Tulsiani
- ○ Deepak Pathak
- ○ Fernando De la Torre
- ○ Deva Ramanan

For "Select all that apply" questions, please fill in all appropriate squares completely:

Select all that apply: Which are scientists?

- ■ Stephen Hawking
- ■ Albert Einstein
- ■ Isaac Newton
- □ None of the above

For questions where you must fill in a blank, please make sure your final answer is fully included in the given space. You may cross out answers or parts of answers, but the final answer must still be within the given space.

Fill in the blank: What is the course number?

$$\boxed{\text{16-822}}$$

1 Auto-calibration [10 pts]

1. **[8 pts]** How many images are required to compute a metric rectification in each of the following scenarios (please give brief 1-2 line explanations):

   (a) The principal point in all images is known to be at $(0, 0)$.

   (b) The cameras have zero skew.

   (c) The principal point in all images is known to be at $(0, 0)$, and the cameras have zero skew.

   (d) Camera intrinsics are known for all cameras.

   (e) Camera instrinsics for each camera are of the form $\begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 0 \end{bmatrix}$, but each camera can have a different and unknown focal length.

   (f) The intrinsics across all cameras are known to be the same.

   (g) (2pts) The intrinsics across all cameras are known to be the same and the plane at infinity is known in the 3D reconstruction.

Consider intrinsics matrix to be $K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$

Also, $\omega^* = P_n L P_n^T = KK^T$ to a scale

$$\therefore \begin{bmatrix} \omega^*_{11} & \omega^*_{12} & \omega^*_{13} \\ \omega^*_{21} & \omega^*_{22} & \omega^*_{23} \\ \omega^*_{31} & \omega^*_{32} & \omega^*_{33} \end{bmatrix} \equiv \begin{bmatrix} \alpha_x^2 + s^2 + x_0^2 & s\alpha_y + x_0 y_0 & x_0 \\ s\alpha_y + x_0 y_0 & \alpha_y^2 + y_0^2 & y_0 \\ x_0 & y_0 & 1 \end{bmatrix} \quad — (1)$$

Linear constraints in ω^* are also linear in L.

Need to find L to compute K & v.

L is a 4×4 symmetric matrix $\therefore$ 10 variables but 9 degrees of freedom

However, if we impose $\det(L) = 0$ then 8 degrees of freedom (2)

a) Principal point @ $(0, 0)$ i.e $x_0, y_0 = 0$

$\therefore \omega^*_{13} = 0$ and $\omega^*_{23} = 0$

i) contd.

a) $\therefore$ Constraints $\rightarrow$ $\omega^*_{13} = 0$ & $\omega^*_{23} = 0$

$\therefore$ 2 constraints per image

$\therefore$ 4 images needed to satisfy 8 dof.

b) Camera has zero skew i.e $s = 0$

$$\therefore \omega^*_{12} = \omega^*_{33} (x_0 y_0) = \omega^*_{33} \left(\frac{\omega^*_{13}}{\omega^*_{33}} \cdot \frac{\omega^*_{23}}{\omega^*_{33}} \right)$$

$$\therefore \omega^*_{12} \omega^*_{33} = \omega^*_{13} \omega^*_{23}$$

$\therefore$ 1 constraint per image

$\therefore$ 8 images needed to satisfy 8 dof.

c) principal pt. @ $(0,0)$ and skew $= 0$

$\therefore$ $|||^r$ to a) & b)

$\omega^*_{13} = 0$, $\omega^*_{23} = 0$ & $\omega^*_{12} \omega^*_{33} = \omega^*_{13} \omega^*_{23} = 0$

$\therefore$ 3 contraints per image

$\therefore$ 3 (or $8/3$) images needed to satisfy 8 dof.

d) $K_1 \ldots K_n$ Known

V can be directly computed from the equation without any images required

$\therefore$ 1 view will be sufficient.

e) $s = x_0 = y_0 = 0$ and $\alpha_x = \alpha_y = f$

$\therefore \omega^*_{12} = 0$, $\omega^*_{13} = 0$, $\omega^*_{23} = 0$

$\therefore \omega^*_{11} = \omega^*_{22}$

$\therefore$ 4 constraints per image.

$\therefore$ 2 images needed for 8 dof

f) $K_1 = K_2 \dots = K_n = K$

$$\therefore \frac{\omega^{*i}_{11}}{\omega^{*j}_{11}} = \frac{\omega^{*i}_{12}}{\omega^{*j}_{12}} = \frac{\omega^{*i}_{13}}{\omega^{*j}_{13}} = \frac{\omega^{*i}_{22}}{\omega^{*j}_{22}} = \frac{\omega^{*i}_{23}}{\omega^{*j}_{23}}$$

$\therefore$ 5 constraints between a pair of images.

$\therefore$ 2 pairs needed. So, 3 images needed for 8 dof

g) Since plane at infinity is known, v is known

$\therefore$ 5 dof remaining out of 8 remaining.

$\therefore$ 5 constraints similar to f) come from shared intrinsics

$\therefore$ 2 images needed.

2. **[2 pts]** To recover a metric rectification, we need to compute a Homography H of the form $\begin{bmatrix} K_1 & \mathbf{0} \\ \mathbf{v}^\top & \cancel{0}\ 1 \end{bmatrix}$. Defining $A = \begin{bmatrix} K_1 \\ \mathbf{v}^\top \end{bmatrix}$, our strategy is to first determine $L = \overset{AA^\top}{\cancel{A^\top} A}$ using constraints on intrinsic matrices, and then recover A. Given L, outline the approach for computing A.

$$L = AA^\top = \begin{bmatrix} K_1 \\ v^\top \end{bmatrix} \begin{bmatrix} K_1^\top & v \end{bmatrix} = \begin{bmatrix} K_1 K_1^\top & K_1 v \\ (K_1 v)^\top & v^\top v \end{bmatrix}$$

- Given L, take the top left 3x3 sub matrix and perform Cholesky decomposition to get K_1
- Next, we take the 1st 3 elements of the last column of L. This represents $K_1 v$. Since K_1 is known, we can multiply this with K_1^{-1} to get v.
- Hence, K_1 and v are known. So, A is known.

2 Non-linear Least Squares [4 pts]

3. **[4 pts]** Are these statements true or false?

(a) Gauss-Newton method is a second-order optimization method.

(b) If we optimize $\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2$ using Gauss-Newton, we will find the global optima in a single iteration.

(c) Gauss-Newton optimization can be used for minimizing any cost function $\min_{\mathbf{x}} C(\mathbf{x})$.

(d) When performing bundle adjustment, we have an objective that is typically of the form of multiple re-projection errors measuring the difference between the observed and re-projected 2D coordinates: $L = \sum \|\mathbf{x}_n^{obs} - \mathbf{x}_n^{proj}\|$, where this optimization is iteratively performed via a algorithm like Gauss-Newton. Is the following statement true or false – if we add a (positive) weight associated with each error i.e. $L = w_n \sum \|\mathbf{x}_n^{obs} - \mathbf{x}_n^{proj}\|$, then we can no longer use Gauss-Newton to optimize this objective.

a) False
b) True
c) False
d) False

3 Trifocal Tensor [5 pts]

4. **[2 pts]** Are these statements true or false?

(a) The trifocal tensfor corresponding to the triplet of cameras (P, P', P'') is the same as the trifocal tensfor for $(PH, P'H, P''H)$, where H is any invertible matrix.

(b) Assuming $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]$ is the trifocal tensor for (P, P', P''), then $\mathbf{T}_1^\top, \mathbf{T}_2^\top, \mathbf{T}_3^\top$ is the trifocal tensor for (P, P'', P').

a) True

b) True

5. **[3 pts]** Let us assume that $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]$ is the trifocal tensor for (P, P', P''). We then transform the camera P'' to $\bar{P}''$ by rotating it in place by R (i.e. camera center and intrinsics K'' are unchanged). Find the trifocal tensfor for the triplet $(P, P', \bar{P}'')$ in terms of $[\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3], K, R$.

Consider a point x'' on image 3 such that $x'' = H_{13}(l')x$ where x is a point on image 1.
After the new rotation, x'' becomes $\bar{x}''$ related by
$\bar{x}'' = Hx''$ where $H = K''RK''^{-1}$

contd.

4 Structure from Motion [6 pts]

6. **[3 pts]** Are these statements true or false?

(a) Given sufficient images of a static scene with accurate correspondences, one can recover a metric reconstruction of the underlying scene.

(b) When performing Hierarchical SfM, we first obtain independent reconstructions across clusters and then align them. If we know camera intrinsics for all cameras, two independently reconstructed clusters would need to be aligned via a euclidean transform.

(c) When performing incremental SfM, the initial two cameras selected for bootstrapping should ideally correspond to the closest possible pair.

a) False - Rectifies to a 3D homography ambiguity

b) True

c) False - The larger the triangulation angle, the more robust it is to noise.

5)

$$\bar{x}'' = K''RK''^{-1}[T_1^T\ T_2^T\ T_3^T]\, l' x \quad = [\bar{T}_1^T\ \bar{T}_2^T\ \bar{T}_3^T]\, l' x$$

$\therefore$ the new trifocal tensor is given by

$$\bar{T}_1^T = K''RK''^{-1} T_1^T$$

$$\therefore \bar{T}_1 = T_1 (K''RK''^{-1})^T$$

Similarly, $\bar{T}_2 = T_2 (K''RK''^{-1})^T$, $\bar{T}_3 = T_3 (K''RK''^{-1})^T$

$$\therefore [\bar{T}_1\ \bar{T}_2\ \bar{T}_3] = [T_1\ T_2\ T_3] (K''RK''^{-1})^T$$

7. **[3 pts]** Affine Triangulation: Assuming known affine cameras $\{\mathbf{P}_n\}$ and observed projections $\{\mathbf{x}_n\}$ for an (unknown) 3D point $\mathbf{X}$, describe an algorithm to compute $\mathbf{X}$ that minimizes the sum of squared reprojection error across all cameras.

Consider $x = PX = \begin{bmatrix} m_1^T & p_4^1 \\ m_2^T & p_4^2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$

$$x = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} m_1^T \tilde{X} + p_4^1 \\ m_2^T \tilde{X} + p_4^2 \\ 1 \end{bmatrix} \quad \text{where } \tilde{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$\therefore \tilde{x} = M\tilde{X} + t$ where $M = \begin{bmatrix} m_1^T \\ m_2^T \end{bmatrix}$, $t = \begin{bmatrix} p_4^1 \\ p_4^2 \end{bmatrix}$

$\therefore$ Reprojection error is given by

$$L = \sum_{i=1}^{n} \left\| \tilde{x}_i - (M_i\tilde{X} + t_i) \right\|^2$$

$\therefore$ To get $\tilde{X}$ that minimizes the above error, computing $\tilde{X}$ that sets $\frac{\partial L}{\partial \tilde{X}}$ to 0

$$\therefore \frac{\partial L}{\partial \tilde{X}} = 2\sum_{i=1}^{n} \underbrace{(-M_i^T)}_{3\times 2} \Big(\underbrace{\tilde{x}_i}_{2\times 1} - \big(\underbrace{\underbrace{M_i}_{2\times 3} \underbrace{\tilde{X}}_{3\times 1}}_{2\times 1} + \underbrace{t_i}_{2\times 1} \big) \Big)$$

$$\sum_{i=1}^{n} \underbrace{M_i^T M_i \tilde{X}}_{3\times 1} + \underbrace{M_i^T t_i}_{3\times 1} - \underbrace{M_i^T \tilde{x}_i}_{3\times 1} = 0$$

$$\underbrace{\left(\sum_{i=1}^{n} M_i^T M_i \right)}_{3\times 3} \tilde{X} = \underbrace{\sum_{i=1}^{n} M_i^T (\tilde{x}_i - t_i)}_{3\times 1}$$

Let $A = \sum_{i=1}^{n} M_i^T M_i$ a 3×3 matrix

$\therefore A^+ = (A^T A)^{-1} A^T$ be the pseudo inverse

$$\therefore \tilde{X} = A^+ \left(\sum_{i=1}^{n} M_i^T (\tilde{x}_i - t_i) \right)$$

$\therefore$ The closed form solution shown above should help compute $\tilde{X}$

Collaboration Questions Please answer the following:

1. Did you receive any help whatsoever from anyone in solving this assignment?

   ○ Yes

   ● No

   - If you answered 'Yes', give full details:
   - (e.g. Jane Doe explained to me what is asked in Question 3.4)

2. Did you give any help whatsoever to anyone in solving this assignment?

   ○ Yes

   ● No

   - If you answered 'Yes', give full details:
   - (e.g. I pointed Joe Smith to section 2.3 since he didnt know how to proceed with Question 2)

3. Did you find or come across code that implements any part of this assignment ?

   ○ Yes

   ● No

   - If you answered 'Yes', give full details: No
   - (book & page, URL & location within the page, etc.).