Time: 120 minutes Max marks: 25

Instructions:
- Do not plagiarize. Do not assist your classmates in plagiarism.
- Show your full solution for the questions to get full credit.
- Attempt all questions that you can. Each question carries 5 points each.
- In the unlikely case a question is not clear, discuss it with an invigilating TA. Please ensure that you clearly include any assumptions you make, even after clarification from the invigilator.

1. (5 points)

(a) (1+1 = 2 points) Given the detection results with corresponding confidence scores, compute the average precision. Compute the overall precision for the detection.

Solution:

Sample	1	2	3	4	5	6	7
confidence	0.63	0.77	0.92	0.86	0.88	0.58	0.91
Iou-based GT match	TP	TP	TP	FP	TP	TP	FP

Table 1: Detection result table for Q1(a)

Confidence	Cumulative TP	Cumulative FP	Precision
0.92	1	0	1
0.91	1	1	0.5
0.88	2	1	0.67
0.86	2	2	0.5
0.77	3	2	0.6
0.63	4	2	0.67
0.58	5	2	0.71

Table 2: Solution Table

The following methods could be used to calculate the average precision
(a)11 point method
(b) average precision= $\sum confidence_i * precision_i$
(c) average precision(vanilla)= (1+0.5+0.67+0.5+0.6+0.67+0.71)/7
Overall precision= 0.71

(b) (2+1 = 3 points) Sam Altman has installed a robotic arm and a voice synthesizer, which he used for his pet cat and dog. When one of them follows the command via the voice synthesizer, the robotic arm tosses a treat to that one. The cat seems to be quite content with the setup, however, the dog(through some super advanced AI communication) complains to Sam that the robotic arm is based against it and favors the cat. The arm relies on a camera that detects the cat and dog and categorizes their respective poses(sit, lie down, roll over, etc.) and signals the arm to toss a treat. We are sure that the pose classification is quite perfect, but the detector is quite questionable. (i) As part of your selection process at OpenAI, you are asked to design an experiment to establish the truth behind the claim of bias against the dog. (ii) Say you find that the model is indeed biased against the dog what observation would you expect to make?

Solution:

(i) Design of experiment
- Check the different categories of the detector.

- Estimate the overall accuracy/error say via mAP.
- Annotate images/frames for dog and cat separately.
- Compute class-wise metrics (e.g., AP, FNR, etc.) and check if the metrics are similar for the two classes.

(ii) Observation
If the errors have a large difference across the categories, the detector is biased.

2. (1×5 = 5 points)Provide short answers.

   (a) SIFT descriptors are preferred as they result in robust matches across viewpoints and do not generate outliers. Justify or refute.
   Solution: SIFT based detectors and descriptors are local features and their representation and provide robust matching across viewpoints. However, it performs thresholding to select matches and hence cannot guarantee correct matches and therefore will lead to outliers.

   (b) Why is a derivative of a Gaussian filter preferred as opposed to a finite difference operator(that takes the difference between adjacent pixels)?
   Solution: DOG helps smooth out the noise in an image whereas a finite difference operator will yield a high response for noise too.

   (c) Why is the second-order derivative filter useful?
   Solution: Blob or corner detection, basically wherever we need to detect zero-crossings at different scales.

   (d) Why is the Harris corner detector rotation invariant?
   Solution: Since it only considers the eigenvalues of the error coefficient(auto correlation) matrix and does not depend on the eigenvector (essentially the direction of the error for surface).

   (e) Why is the Harris corner detector not invariant to intensity changes?
   Solution: The error coefficient(auto correlation) matrix depends on the gradient magnitudes which inturn depends on intensity. Scaling of intensity results in a change in the gradient. Therefore, harris is only partially invariant to intensity changes.

3. (5 points)

   (a) (1+1=2 points) Given a pixel on a detection edge i.e. the boundary between two regions that have a sharp change in intensity, how would you compute the direction along the edge and the direction perpendicular to it? Hint: Can you use the image gradient?
   Solution: The gradient direction points in the direction of maximum change. Let

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} = [I_x \quad I_y]$$

   The edge is the boundary of the intensity change, i.e., the direction of minimum intensity change. Consequently, it will be orthogonal to the gradient vector. Therefore, it will be a vector proportional to

$$[-I_y \quad I_x]$$

   (b) (2+1=3 points) Adapt the Harris corner detector for detecting edges but not corners and write the algorithm along with the thresholding criteria that you would use. Propose a non-maximum suppression(NMS) strategy to thin the edges. Note that the NMS strategy should avoid suppressing pixels along the edge.
   Solution: Let $\begin{bmatrix} A & C \\ C & B \end{bmatrix}$ be the error coefficient matrix, λ_{max} and λ_{min} be the eigenvalues for detecting edges. We'd use $\lambda_{max} > \tau_{max}$ and $\lambda_{min} < \tau_{min}$

Following is the algo for NMS:

(a) Take a $k \times k$ patch around the edge pixel.

(b) find the direction of the edge as $[-I_y \quad I_x]$ as explained in the previous part.

(c) Ignore the pixels aligned with the edge directions. Alternatively, consider the pixels aligned with the gradient direction.

(d) Suppress the other pixels in the grid if and only if the gradient magnitude is lower than that of the central pixel.

4. (3+2 = 5 points)

(a) (1+1+1=3 points) Answer the following with an appropriate justification. Statements that are always true are true. All options must be selected to get credit.

(i) Given the Fundamental matrix and the intrinsic parameter matrices for two cameras, one can find the Essential matrix. State True or False.
Solution: True, as $F = (K^{-1})^T E K^{-1}$ assuming both cameras have the same intrinsics.

(ii) The Essential matrix has five degrees of freedom. State True or False.
Solution: True, 2 Translational Degree of freedom; 3 Rotational Degree of freedom.

(iii) You estimated the fundamental matrix F, but observed that the epipolar lines do not intersect. What seems to be the problem?
(A) F has a very high Frobenius norm (sum of squares of elements).
(B) F was estimated without normalization of the point correspondences.
(C) F is non-singular.
(D) F is pointless; we should have estimated the Essential matrix, E, instead
(E) The corresponding points were far too noisy.
Solution: C, When F is non-singular, it will have an empty null space, i.e. the epipole is not defined (i.e. the epipolar lines will not intersect).

(b) (2 points) If the relative translation between two cameras is aligned with the camera's Y-axis and there is no relative rotation, show that the X coordinate of the corresponding points in the two images will be identical.
Solution:
The essential matrix has the form:

$$E = [t]_\times R = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

Since $\mathbf{t} = [t_x \quad t_y \quad t_z]^T$ and the translation is in the Y-axis only. So, $t_x = t_z = 0$. Given that there is no relative rotation, i.e. $R = I$ and $\mathbf{t} = [t_y] = [0 \quad t_y \quad 0]$. So, the essential matrix is:

$$E = [t_y]_\times I = \begin{bmatrix} 0 & 0 & t_y \\ 0 & 0 & 0 \\ -t_y & 0 & 0 \end{bmatrix}$$

Let there be a point P_w in the world frame whose corresponding coordinates in both camera frames are P and P'. Let, $\mathbf{P} = [x \quad y \quad 1]$ and $\mathbf{P'} = [x' \quad y' \quad 1]$. Then, as per the epipolar constraints, for

$\mathbf{P}$, E and $\mathbf{P}'$ we have:

$$P^T E P' = 0$$

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & t_y \\ 0 & 0 & 0 \\ -t_y & 0 & 0 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = 0$$

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} t_y \\ 0 \\ -t_y x' \end{bmatrix} = 0$$

$$x t_y - t_y x' = 0$$

$$t_y (x - x') = 0$$

t_y is a non-zero constant, $x - x' = 0$ or $x = x'$. Therefore, the corresponding X coordinate in the two images will be identical.

5. (5 points)

   (a) (2 points) Given a set of 3D points(a point cloud), $\mathcal{X} = \{x_1, x_2, x_3, ..., x_n\}, \forall x_i \in \mathbb{R}^3, i = 1, 2, ..., n$, write the algorithm for finding the largest planar structure. Say the point cloud only has 30% points lying on the planar structure. How many iterations RANSAC would be needed to ensure that there is a 1 in 1000 chance (a 10^{-3} probability) that an inlier set would not be sampled during the RANSAC iterations.

   (b) (1+2=3 points) Given corresponding image points in pixel coordinates lying on the same plane, $\hat{p} = (x, y)^T$ and $\hat{p'}(x', y')$ are related via a Homographhby H as $p = Hp'$. Here, p and p' are the corresponding homogeneous representations. Write the expression using the Direct Linear Transformation (DLT) for homography estimation, and show that this corresponding pair only leads to two independent constraints.

Solution:

(a) Algorithm:

(1) Select a random sample of the minimum required size to fit the structure.

(2) Compute a putative model from a sample set.

(3) Compute the set of inliers to this structure from a set of points.

Repeat (1)-(3) until the structure with the most inliers over all samples is found.
Considering number of points in the minimal subset to be 3, and probability of 10^{-3} that only 30% points are lying on the planar structure, we have

$$10^{-3} = (1 - 0.3^3)^N$$

$$N = \frac{-3 \log(10)}{\log(1 - 0.3^3)}$$

(b)

$$p = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} p'$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$$

One way to show the constraint is via cross multiplication.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} h_1^T p' \\ h_2^T p' \\ h_3^T p' \end{bmatrix}$$

$$\begin{bmatrix} x/w \\ y/w \end{bmatrix} = \begin{bmatrix} h_1^T p' / h_3^T p' \\ h_2^T p' / h_3^T p' \end{bmatrix}$$

$$x h_3^T p' = w h_1^T p' \tag{1}$$

$$y h_3^T p' = w h_2^T p' \tag{2}$$

Another way to show equality of vectors via the cross product.

$$p \times H p' = 0$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \times \begin{bmatrix} h_1^T p' \\ h_2^T p' \\ h_3^T p' \end{bmatrix} = 0$$

$$\begin{bmatrix} 0 & -w & y \\ w & 0 & -x \\ -y & x & 0 \end{bmatrix} \begin{bmatrix} h_1^T p' \\ h_2^T p' \\ h_3^T p' \end{bmatrix} = 0$$

Rewriting this as $Ah = 0$ where $h = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$ will give a 3×9 matrix A, for which only 2 of the 3 rows are linearly independent. This can be shown by Gaussian elimination or any of your favorite methods for reducing the matrix to its row-echelon form.