INDIAN STATISTICAL INSTITUTE

First-Semestral Examination: 2025-2026

Computer Vision

M.Tech.(Computer Science)

Date: 14.11.2025 **Full Marks: 100** **Time: 3 Hours**

Answer any **ten** questions. All questions carry equal marks.

1. Consider that $u(x)$ is the noise-free signal of interest and $g(x)$ is a filter for edge detection.

   (a) Prove that $g(x)$ has to maximize the following quantity to achieve good signal-to-noise ratio:

   $$\mathcal{S} = \left\{\int_{-\infty}^{+\infty} g(x)u(-x)dx\right\}\left\{\int_{-\infty}^{+\infty}[g(x)]^2dx\right\}^{-\frac{1}{2}}.$$

   (b) Prove that $g(x)$ is optimal with respect to good locality measure of an edge, if $g(x)$ maximizes the following quantity:

   $$\mathcal{L} = \left\{\int_{-\infty}^{+\infty} g''(x)u(-x)dx\right\}\left\{\int_{-\infty}^{+\infty}[g'(x)]^2dx\right\}^{-\frac{1}{2}}$$

   where $g'(x)$ and $g''(x)$ denote the first and the second derivatives of $g(x)$, respectively.

   (c) Prove that $g(x)$ has to maximize the following quantity to reduce the number of false local maxima due to noise:

   $$\mathcal{C} = \sqrt{\left\{\int_{-\infty}^{+\infty}[g'(x)]^2dx\right\}\left\{\int_{-\infty}^{+\infty}[g''(x)]^2dx\right\}^{-1}}$$

   where $g'(x)$ and $g''(x)$ denote the first and the second derivatives of $g(x)$, respectively.

   [3+4+3=10]

2. (a) Define perspective-n-point (PnP) problem.

   (b) Considering homogeneous coordinates, show that the PnP problem can be formulated as a minimum direction problem.

   (c) How many points will be required to solve the minimum direction problem of (b)?

   (d) Find extrinsic parameters of a camera by solving the minimum direction problem of (b).

   [2+4+1+3=10]

3. (a) Show that the fundamental matrix F can capture the geometric relationship between two cameras with arbitrary intrinsic matrices Λ_1 and Λ_2.

   (b) Show that the problem of estimating the fundamental matrix between two cameras can be formulated as a minimum direction problem.

   (c) Write down the steps to estimate the fundamental matrix between two cameras using RANSAC (RANdom SAmple Consensus) algorithm.

   (d) Derive the rotation matrix Ω and translation vector τ between two cameras when the fundamental matrix F between two cameras is known.

   [3+2+3+2=10]

4. (a) How can the coefficients of Laplacian pyramid be computed at a particular level?

   (b) Show that the Laplacian pyramid is a linear invertible image representation.

   (c) Write down the steps to generate static images that change in interpretation as a function of the viewing distance.

[3+4+3=10]

5. (a) Prove that in order for the Laplacian of Gaussian (LoG) and Difference-of-Gaussians (DoG) to have same zero crossings, the value of σ for the LoG, and σ_1, σ_2 for the DoG must be selected based on the following relation:

$$\sigma^2 = \frac{{\sigma_1}^2{\sigma_2}^2}{{\sigma_1}^2 - {\sigma_2}^2} \ln\left[\frac{{\sigma_1}^2}{{\sigma_2}^2}\right] \quad \text{where } \sigma_1 > \sigma_2.$$

   (b) Prove that in scale-space analysis, the value of the DoG function D at the extremum $\hat{\mathrm{x}}$ is given by

$$D(\hat{\mathrm{x}}) = D + \frac{1}{2}(\nabla D)^T \hat{\mathrm{x}}$$

   where ∇ is the gradient operator and T denotes transpose.

   (c) Prove that the cornerness of a pixel can be described by the structure tensor of the pixel.

   (d) Define a response function for corner detection.

[3+3+3+1=10]

6. (a) Consider a weighted graph G with four vertices $\{V_1, V_2, V_3, V_4\}$ and weight matrix

$$W = \begin{pmatrix} 0 & 2 & 1 & 2 \\ 2 & 0 & 3 & 1 \\ 1 & 3 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{pmatrix}$$

      i. Compute the Laplacian matrix L of the graph G.

      ii. If the vertices are partitioned into two sets $A = \{V_1, V_4\}$ and $B = \{V_2, V_3\}$, calculate the normalized cut between two sets A and B.

   (b) Let $\{(1, 1), (2, 5), (3, 1)\}$ be the set of three consecutive vertices on a discrete contour. Compute both elastic energy and curvature energy at the vertex (2, 5).

   (c) Show that the total energy of the snake model can be expressed as pairwise interactions of vertices.

   (d) What is the computational complexity of Viterbi algorithm?

[(2+2)+3+2+1=10]

7. (a) Describe the architecture of U-Net for semantic segmentation.

   (b) Define loss functions for pixel level and region level semantic segmentation.

   (c) Explain max-unpooling with example.

   (d) What are the main differences between U-Net and SegNet?

[4+2+2+2=10]

8. (a) Define the class activation map (CAM), and explain how it is related to the softmax layer of a CNN architecture trained for image classification task.

   (b) Define gradient-weighted CAM (grad-CAM).

   (c) Show that the grad-CAM is a generalisation of the CAM.

[(1+3)+2+4=10]

9. (a) What is convolutional block attention module (CBAM)?

   (b) Draw the diagram of each submodule of the CBAM and explain the corresponding computation process.

[2+(4+4)=10]

10. (a) Suppose a block, consisting of a 7×7 convolution layer (with padding 3 and stride 2) and a 2×2 max-pooling layer (with stride 2), takes a three channel 448×448 color image as input and produces 512 output feature maps.

   i. What is the number of parameters to be trained for this block?

   ii. What is the dimension of output feature maps for this block?

   (b) Define a scale and translation invariant transform to correct region proposals of R-CNN.

   (c) How does region proposal network (RPN) of faster R-CNN work?

   (d) Explain the loss function used to train the RPN.

[(2+1)+2+3+2=10]

11. (a) Describe the vanilla YOLO model for object detection.

   (b) Explain the multi-part loss function used to train the YOLO model.

   (c) How can feature pyramid network be used for multi-scale object detection?

[4+3+3=10]

12. (a) Let I_x and I_y be the 2D image gradient along x and y directions, respectively, while u and v be the corresponding optical flow along these two directions. Prove that the vectors $< u, v >$ and $< I_x, I_y >$ are parallel to each other.

   (b) Considering both optical flow and smoothness constraints, prove that

$$I_x \Delta v = I_y \Delta u$$

   where Δu and Δv are Laplacians of optical flow along x and y directions, respectively.

   (c) How does multi-resolution Lucas-Kanade method address the aperture problem when the motion of a pixel is not small?

   (d) Describe the architecture and loss function of FlowNet for learning the optical flow.

[2+2+3+(2+1)=10]