

COMPUTER VISION
(CSEN 4236)

Time Allotted : 2½ hrs

Full Marks : 60

Figures out of the right margin indicate full marks.

*Candidates are required to answer Group A and
any 4 (four) from Group B to E, taking one from each group.*

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve:

12 × 1 = 12

Choose the correct alternative for the following

- (i) How does picture formation in the eye vary from image formation in a camera?
(a) Fixed focal length (b) Varying distance between lens and imaging plane
(c) No difference (d) Variable focal length
- (ii) Which transformation preserves parallel lines but does not necessarily preserve angles and lengths?
(a) Rigid transformation (b) Affine transformation
(c) Projective transformation (d) Non-linear transformation
- (iii) Which of the following operator is known as direction mask?
(a) Prewitt Operator (b) Sobel Operator
(c) Robinson Compass Masks (d) Kirsch Compass Masks
- (iv) Consider a 4-bit grey scale image with dimensions 16 x 32. What will be the range of values in its X-axis?
(a) 0 to 255 (b) 1 to 256 (c) 0 to 15 (d) -8 to 7
- (v) Which camera model is the most accurate for 3D reconstruction?
(a) Orthographic (b) Weak perspective
(c) Affine (d) Perspective
- (vi) Affine reconstruction differs from projective reconstruction because
(a) Affine reconstruction retains true angles and distances
(b) Affine reconstruction preserves parallelism but not depth
(c) Projective reconstruction is always more accurate than affine reconstruction
(d) Affine reconstruction requires calibrated cameras
- (vii) Radial lens distortion in cameras mainly causes:
(a) Changes in image brightness (b) Straight lines appearing curved in images
(c) Reduction in image resolution (d) Color distortion in images

- (viii) Regularization in motion estimation is used to
 - (a) Reduce motion blur in videos
 - (b) Improve stability by incorporating smoothness constraints in optical flow calculations
 - (c) Remove color artifacts from images
 - (d) Segment objects based on their speed
- (ix) The Kalman filter consists of two key steps: prediction and update. The update step is responsible for
 - (a) Propagating the motion model to the next state
 - (b) Refining the predicted state using the current observation
 - (c) Initializing the tracking system
 - (d) Ignoring noisy measurements
- (x) Which assumption is commonly made in the Horn–Schunck optical flow method?
 - (a) Motion is completely random across the image
 - (b) Optical flow varies smoothly across neighboring pixels
 - (c) Pixel intensity changes significantly between frames
 - (d) Motion exists only at object boundaries

Fill in the blanks with the correct word

- (xi) Histogram of Oriented Gradients (HOG) is primarily used for _____ detection and is commonly applied in pedestrian detection.
- (xii) In conic section fitting, the Euclidean distance minimizes the actual shortest distance from a point to the fitted curve, whereas the algebraic distance minimizes the _____ error.
- (xiii) In stereo vision, the process of estimating the 3D position of a point by using its projections in two or more images is called _____.
- (xiv) The _____ descriptor extracts distinctive keypoints that are invariant to scale and rotation.
- (xv) The _____ matrix describes the epipolar geometry between two calibrated cameras.

Group - B

2. (a) Define convolution in the context of image filtering. Explain how convolution is applied in image processing with an example of a 3×3 kernel. [[CO1](Understand/IOCQ)]
 (b) What is the difference between smoothing and differencing in image filtering? Illustrate real-world applications where each is used. [[CO1](Understand/LOCQ)]
 (c) Explain the concept of scale space in computer vision. How does it help in multi-scale image analysis? [[CO1](Understand/IOCQ)]

5 + 4 + 3 = 12
3. (a) A digital camera with a focal length of 50mm captures an image of an object located 2 meters (2000mm) away from the camera. The sensor has a width of 36mm and a resolution of 4000 pixels in width. Using the Thin Lens Equation, calculate the image distance (distance from the lens to the image sensor). [[CO2](Analyse/IOCQ)]

- (b) Calculate the pixel size (in mm) given that the sensor width is 36mm and the image resolution is 4000 pixels in width. [[C05](Remember/IOCQ)]
- (c) Explain how increasing the camera's aperture size (decreasing f-stop value) affects image brightness and radiometry principles. [[C03](Apply/IOCQ)]
- 5 + 4 + 3 = 12**

Group - C

4. (a) Explain the difference between edge detection and corner detection. What role do these techniques play in feature extraction? Provide an example where corner detection is preferable over edge detection. [[C01](Understand/IOCQ)]
- (b) What are SIFT and HOG descriptors? Compare them based on their robustness to scaling, rotation, and illumination changes. Which one is more suitable for pedestrian detection, and why? [[C01](Understand/LOCQ)]
- (c) What are active contours (snakes) in computer vision? How do they help in detecting object boundaries? Also, explain the concept of shape context descriptors and how they assist in object recognition. [[C05](Apply/IOCQ)]
- 4 + 4 + 4 = 12**
5. (a) Explain the difference between Hough Transform and Least Squares Line Fitting for detecting lines in an image. Which method is more suitable for noisy data, and why? [[C03](Analyse/LOCQ)]
- (b) Explain the difference between appearance-based methods and feature-based methods for object recognition. Which method is more suitable for recognizing objects under varying lighting conditions, and why? [[C03](Remember/LOCQ)]
- (c) Explain the concept of Shape Context Descriptors in feature detection. How do they help in object recognition, and what are their advantages over traditional edge-based descriptors? [[C01](Understand/LOCQ)]
- 4 + 4 + 4 = 12**

Group - D

6. (a) A stereo camera system is used for 3D depth estimation. The cameras have the following parameters
 Baseline distance between the two cameras: $B = 120$ mm
 Focal length of each camera $f = 1000$ pixels
 A 3D point appears at pixel coordinates
 Left image: $x_1 = 450$
 Right image: $x_2 = 410$
 Compute the disparity d between the two image points. [[C05](Apply/IOCQ)]
- (b) Using the disparity formula, calculate the depth Z of the point in mm. [[C05](Apply/LOCQ)]
- (c) If the focal length is doubled while keeping all other parameters constant, how does the depth change? Explain briefly. [[C05](Apply/IOCQ)]
- (d) If the baseline distance is reduced to 60 mm while keeping the focal length and disparity the same, what will be the new depth Z' . [[C05](Apply/IOCQ)]
- (e) If the disparity reduces to 20 pixels while keeping the original baseline and focal length constant, compute the new depth Z'' . [[C05](Apply/IOCQ)]
- 2 + 2 + 3 + 3 + 2 = 12**

7. (a) Explain the process of 3D reconstruction using projective geometry. How does it differ from Euclidean reconstruction? [[C02](Understand/IOCQ)]
 (b) What are the major challenges in reconstruction by triangulation? Describe three key factors that affect its accuracy. [[C02](Understand/LOCQ)]
 (c) Describe the role of the fundamental matrix in projective reconstruction. How does it help in recovering the structure of a scene? [[C02](Understand/IOCQ)]

4 + 4 + 4 = 12

Group - E

8. (a) Two well-known methods for computing optical flow are Lucas-Kanade and Horn-Schunck. Compare these two methods in terms of assumptions, advantages, and limitations. In which scenarios is Lucas-Kanade preferred over Horn-Schunck, and vice versa? [[C04](Analyse/IOCQ)]
 (b) Expectation-Maximization (EM) is used to separate moving objects from the background. Briefly describe how EM helps in motion segmentation. Name a real-world application where EM-based motion segmentation is used. [[C04](Analyse/IOCQ)]
 (c) The optical flow constraint equation (OFCE) is given by

$$I_x u + I_y v + I_t = 0$$

where I_x, I_y, I_t are partial derivatives of the image intensity, and (u, v) represent the flow velocities. Explain the significance of this equation in motion estimation. What is the aperture problem, and how does it limit the computation of optical flow? [[C05](Apply/IOCQ)]

4 + 3 + 5 = 12

9. (a) Explain how motion parallax helps in depth estimation. Derive the relationship between the depth of an object and its apparent motion in an image. [[C04](Analyse/IOCQ)]
 (b) Feature-based techniques like SIFT, SURF, and Harris corners help track motion in images. Explain how these features aid in motion tracking. Why are Harris corner features preferred over edge features for tracking motion? [[C04](Analyse/LOCQ)]
 (c) In an affine transformation, an object undergoes scaling, rotation, and translation. The transformation is represented as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

Define the role of each matrix component a, b, c, d, t_x, t_y in the affine transformation. Given the transformation matrix

$$A = \begin{bmatrix} 0.8 & 0.2 \\ -0.1 & 1.0 \end{bmatrix}, \quad T = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

and initial point $P(2,5)$, compute the new transformed point $P'(x', y')$. [[C05](Apply/IOCQ)]

3 + 4 + 5 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	31.25	68.75	0