RKMVERI

MSc. Big Data Analytics/CSE, Year 2023
Mid-term Examination
Computer Vision
Maximum Marks 100

Time: 3 hours Date: 14/06/2023

Group-A

Answer any 20 questions (3 marks each):

1. Define field of view by mentioning the parameters.

   Field of view $\psi = tan^{-1}(\frac{d}{2f})$, where d is linear dimension (height or width) of the image plane and f is the focal length of the camera.

2. What is the objective of Gamma correction?

   Rectification of non-linear response of the photo sensor to the incident illumination to linear response.

3. What is epipolar line?

   An image point and the optical centre of a camera form a 3D line. This line when projected onto the image plane of another camera, this 2D line is called the epipolar line due to image point of first camera.

4. Which criterion does Otsu's method try to satisfy to select the threshold for image segmentation?

   Otsu's method selects the threshold for image segmentation by maximizing inter-class variance or minimizing intra-class variance.

5. According to Canny what are three criteria that a god edge detector should satisfy?

   (i) Good detection, (ii) Good localization, and (iii) Single response constraint.

6. Why is equation $x\ cos\theta + y\ sin\theta = p$ of straight line is preferred over the equation $y = mx + c$?

   The parameters 'p' and 'θ' are bounded, while the parameter 'm' is unbounded.

7. Define motion field.

   The motion field is the projection of the 3D scene motion onto the 2D image plane.

8. Define optical flow.

   Optical flow is the *apparent* motion of brightness patterns in the image.

9. What are three key assumptions for estimating optical flow?

   Three key assumptions are (i) brightness constancy, (ii) small motion and (iii) spatial coherence.

10. At which points can we estimate the optical flow? And in which direction?

    We can estimate optical flow at the points on the edge, and the optical flow is estimated along the direction perpendicular to the edge at that point.

11. In general, what is the main objective that mean shift used for?

    Mean shift is a tool for finding modes in a set of data samples, manifesting an underlying probability density function (PDF) in R^N.

12. Write down the two main steps of mean shift procedure.

    Two main steps are (i) Compute mean shift vector *m(x)*, and (ii) Translate the Kernel window by *m(x)*.

13. Suppose a Harris matrix is given as H = $\begin{bmatrix} 3 & 2 \\ 2 & 7 \end{bmatrix}$ and relevant empirical constant k=0.05. Calculate the single response (scalar) value for corner detection.

   Det(H)= 3x7 – 2x2= 17; Tr.(H)= 3+7= 10; Single response R= Det(H) – k(Tr.(H))^2= 17 - 0.05x100 = 12.

14. What is main idea or objective of SIFT?

    Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters.

15. What are the major steps of the SIFT algorithm?

    Major steps are (i) Scale-space extrema detection, (ii) Keypoint localization, (iii) Orientation assignment, and (iv) Keypoint description.

16. Mention two main problems of passive stereo vision that are solved by structured light method (active stereo vision).

    (i) Detecting key-points, and (ii) finding corresponding points (correspondence problem).

17. State a disadvantage of a light spot stereo system? Suggest a method to reduce the severity of the problem.

    Light spot stereo system is computationally and memory-wise intensive as one complete image is captured and processed for each laser spot. This may be reduced by using projecting light pattern.

18. What is meant by self-occlusion? How is correspondence problem solved in such cases?

    When some part of the object can be seen by a camera, but not by other camer(s), it is called self-occlusion problem. Correspondence problem, in such case, may be solved by projecting a series of binary light patters.

19. What are the major steps of the SIFT algorithm?

    Major steps are (i) Scale-space extrema detection, (ii) Keypoint localization, (iii) Orientation assignment and (iv) Keypoint description.

20. What is meant by epoch? When does number of epochs and number of iterations become equal?

    Passing (or using) entire training samples once to train the network is known as one epoch. When a batch contains all the training samples.

21. What will be the spatial size of feature map if 100x100 rgb image is convolved with a kernel of size 5x5x3? How this size reduction can be prevented?

    The spatial size of feature map will be 96x96. This size reduction can be prevented by zero padding.

22. State the main difference between principal component analysis and autoencoder.

    PCA attempts to discover a lower dimensional hyperplane which describes the original data, autoencoder learns nonlinear manifold. OR PCA applies linear transformation prior to data reduction, while AE uses non-linear transformation for representative learning.

23. How is denoising autoencoder differs from the basic representative autoencoder?

    While basic representative autoencoder tries to reconstruct the input image itself from its latent space (compressed) representation, the denoising autoencoder tries to reconstruct noise-free image from the latent space (compressed) representation of the corresponding noisy image.

Group-B

Answer any 4 questions (5 marks each):

24. Suppose there is a camera with focal length 2 cm and image plane of size 1.6 cm (horizontally) by 1.5 cm (vertically). Given that photo sensor density along horizontal direction is 2500 cm^{-1} and in vertical direction it is 2000 cm^{-1}, calculate the pixel coordinate corresponding to a scene point at (6 m, 4m, 40m). Assume the origin of the system is at centre of lens of the camera and the optical axis meets the image plane at its centre and the world coordinate axes coincide with that of the camera.

f = 2 cm, $\psi_x = 2500,\ \psi_y = 2000,\ d_x = \frac{1.6}{2}2500 = 2000,\ d_x = \frac{1.5}{2}2000 = 1500,$		
$x = f\dfrac{u\psi_x}{w} + d_x = 2\dfrac{6 \times 2500}{40} + 2000 = 2750$ $y = f\dfrac{v\psi_y}{w} + d_y = 2\dfrac{4 \times 2000}{40} + 1500 = 1900$	OR	$x = -f\frac{u\psi_x}{w} + d_x = -2\frac{6\times2500}{40} + 2000 = 1250$ $y = -f\frac{v\psi_y}{w} + d_y = -2\frac{4\times2000}{40} + 1500 = 1100$

25. A 3D scene point (u, v, w) forms its image at (250 pixels, 400 pixels) in a digital camera with focal length 2cm and photoreceptor density 1000pixels/cm. The camera is shifted back along the optical axis by a distance of 10 metre. The new coordinate of image point is (200 pixel, 320 pixel). Determine(u, v, w). Assume horizontal and vertical drifts $d_x = d_y = 0$.

Given $d_x = d_y = 0$

$$250 = f\frac{u\psi_x}{w} = 2\frac{u \times 1000}{w} \Rightarrow w = 8u \quad \cdots(1)$$

$$200 = 2\frac{u \times 1000}{w+10} \Rightarrow w + 10 = 10u \quad \cdots(12)$$

$Eqn.(2) - Eqn.(1)$ gives $10 = 2u \Rightarrow u = 5\,m$ and from $Eqn.(1)$ we get $w = 40\,m$

$$y = f\frac{v\psi_y}{w} \Rightarrow 400 = 2\frac{v \times 1000}{40} \quad \text{OR} \quad v = 8\,m$$

26. Suppose there are two normalized cameras. First camera is at location (0, 0, 0) and the other at (-√2, -3√2, -5√2). Second camera is rotated by 45° anti-clockwise about optical axis. Show that (-2, 3, √2) in the first camera is the corresponding point of the point (2, 12, 1) in the second camera.

Rotation matrix $R = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and translation vector $t = \begin{bmatrix} \sqrt{2} \\ 3\sqrt{2} \\ 5\sqrt{2} \end{bmatrix}$,

so the essential matrix $E = \begin{bmatrix} 0 & -5\sqrt{2} & 3\sqrt{2} \\ 5\sqrt{2} & 0 & -\sqrt{2} \\ -3\sqrt{2} & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} =$

$\begin{bmatrix} 5 & -5 & 3\sqrt{2} \\ 5 & 5 & -\sqrt{2} \\ -4 & -2 & 0 \end{bmatrix}$

Since $[2 \quad 12 \quad 1] \begin{bmatrix} 5 & -5 & 3\sqrt{2} \\ 5 & 5 & -\sqrt{2} \\ -4 & -2 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \\ \sqrt{2} \end{bmatrix} = [2 \quad 12 \quad 1] \begin{bmatrix} -19 \\ 3 \\ 2 \end{bmatrix} = 0$, these points are corresponding points.

27. Prove that magnitude of motion vector associated to an object point is inversely proportional to its depth.

To find image velocity $\boldsymbol{v}$, differentiate $\boldsymbol{p} = (x(t), y(t))$ with respect to t:

$$x = f\frac{X}{Z}, \text{ then } v_x = \frac{dx}{dt} = f\frac{ZV_X - V_Z X}{Z^2} = \frac{fV_X - V_Z x}{Z}$$

$$y = f\frac{Y}{Z}, \text{ then } v_y = \frac{dy}{dt} = f\frac{ZV_Y - V_Z Y}{Z^2} = \frac{fV_Y - V_Z y}{Z}$$

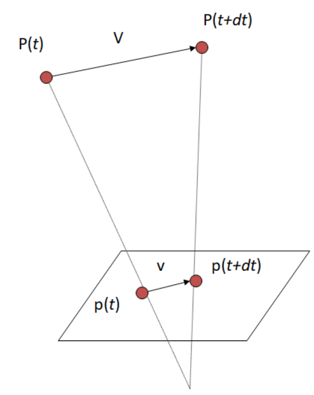

Image motion is a function of both the 3D motion ($\boldsymbol{V}$) and the depth of the 3D point (Z)

Assuming pure translation: $\boldsymbol{V}$ is constant everywhere.

$$\left.\begin{aligned} v_x &= \frac{fV_x - V_Z x}{Z} \\ v_y &= \frac{fV_y - V_Z y}{Z} \end{aligned}\right. \Rightarrow \boldsymbol{v} = \frac{1}{Z}(\boldsymbol{v}_0 - V_Z x), \quad where \;\; \boldsymbol{v}_0 = \left(v_x, v_y\right)^T$$

Thus, the magnitude of the motion vectors is inversely proportional to the depth Z.

28. Write down the steps of Horn and Schunck iterative algorithm for computing the optical flow.

- Set $k=0$
- Initialize all $u^k(x,y)$ and $v^k(x,y)$ with 0
- Until some error measure is satisfied, do

$$u^{(k+1)} = \bar{u}^{(k)} - I_x \frac{I_x \bar{u}^{(k)} + I_y \bar{v}^{(k)} + I_t}{\lambda^2 + I_x^2 + I_y^2}$$

$$v^{(k+1)} = \bar{v}^{(k)} - I_x \frac{I_x \bar{u}^{(k)} + I_y \bar{v}^{(k)} + I_t}{\lambda^2 + I_x^2 + I_y^2}$$

Answer any 2 questions (10 marks each)

29. State the assumptions and, hence, derive the 2D motion constraint equation $\nabla I \cdot \vec{v} = -I_t$ for optical flow. All variables have their usual meaning. At which points can we estimate the optical flow? {8+2]

30. Derive the Harris matrix. How is it used to detect corner points?
[7+3]

31. Derive the expression of 2D motion vector as proposed by Lucas and Kanade.
[5+5]
