MANIPAL INSTITUTE OF TECHNOLOGY
BENGALURU
(A constituent unit of MAHE, Manipal)

VI Semester MIDTERM TEST
Computer Vision (IT_4031)

Time Duration: 2 Hours **Date: 21-03-2024** **Max marks: 30 MARKS**

S. No.	Question	Marks	CO (Mapping)	Blooms Taxonomy Level
1.	What is the name of the process that moves a filter mask over the image, followed by calculating the sum of products? **(a) Correlation** (b) Convolution (c) Linear spatial filtering (d) Non-linear spatial filtering	**1**	**CO2**	**L2**
2.	Which method is used to represent colors in digital images? (a) RGB (Red, Green, Blue) (b) HSV (Hue, Saturation, Value) (c) CMYK (Cyan, Magenta, Yellow, Black) **(d) All of the above**	**1**	**CO2**	**L2**
3.	How to carry out an array function together with one or more images? (a) Row by Row (b) Column by Column **(c) Pixel by Pixel** (d) Array by Array	**1**	**CO2**	**L2**
4.	Which technique is commonly used for edge detection in images? (a) Fourier Transform (b) Gaussian blur **(c) Canny Edge Detection** (d) Histogram Equalization	**1**	**CO2**	**L2**
5.	Which feature detection algorithm is known for its resistance to noise and is used for detecting key points in images? (a) Sobel Operator **(b) Harris Corner Detector** (c) Gabor Filter (d) Median Filtering	**1**	**CO2**	**L2**
6.	Which concept is used to represent the local features of an image that are invariant to scale and rotation? (a) Hough transform (b) Difference of Gaussian (DoG) (c) Eigenfaces **(d) Scale-invariant feature transform**	**1**	**CO2**	**L2**
7.	What is the primary purpose of the Canny edge detector in image processing? (a) Noise Reduction (b) Image Resizing **(c) Edge Detection** (d) Color Correction	**1**	**CO2**	**L2**
8.	Which technique is used to detect and track objects in a sequence of frames? **(a) Optical flow** (b) Template matching (c) Harris corner detection (d) Scale-invariant feature transform	**1**	**CO3**	**L2**
9.	Which computer vision task involves figuring out the pose or placement of objects or individuals inside a scene? (a) Image Classification **(b) Object Tracking** (c) Optical Character Recognition (d) Semantic Segmentation	**1**	**CO3**	**L2**
10.	Which object tracking method is based on estimating the motion of points in an image by comparing brightness patterns? (a) Mean shift method **(b) Lukas-kanade Optical flow** (c) Histogram Equalization (d) Template Matching	**1**	**CO3**	**L2**

<table>
<tr><th>S. No.</th><th>Question</th><th>Marks</th><th>CO (Mapping)</th><th>Blooms Taxonomy Level</th></tr>
<tr><td>11.</td><td></td><td></td><td></td><td></td></tr>
<tr><td>A.

Sol:</td><td>

Identify and mention the requirements of a good feature detector. Also discuss the mathematics of Harris corner detector.

- This approach gives a mathematical approach to detect a corner in an image.
- We know that the shift in (u,v) is represented by:

$$\sum [I(x+u, y+v) - I(x,y)]^2$$

- When we rewrite the equation on matrix form:

$$\sum [u \; v] \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

[1M]

- We now have the equation where W is windowing function (usually 1)

$$M = \sum w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- We further calculate the trace and determinant of M to find R.
- K is empirically determined constant - $R = \det M - k(\mathit{trace}\, M)^2$

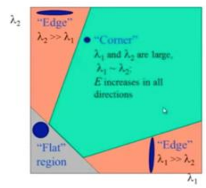
$$\det M = \lambda_1 \lambda_2$$

$$\mathit{trace}\, M = \lambda_1 + \lambda_2$$

[2M]

1. |R| is small, which happens when λ1 and λ2 are small, the region is flat.
2. R<0, which happens when λ1>>λ2 or vice versa, the region is edge.
3. R is large, which happens when λ1 and λ2 are large and λ1~λ2, the region is a corner.

[1M]



[1M]

</td><td>5</td><td>CO2</td><td>L3</td></tr>
<tr><td>B.

Sol:</td><td>

Discuss Correlation and Convolution filters and its operation.
Also for the given input image:

Input Image

<table>
<tr><td>2</td><td>4</td><td>6</td><td>8</td></tr>
<tr><td>10</td><td>12</td><td>14</td><td>16</td></tr>
<tr><td>18</td><td>20</td><td>22</td><td>24</td></tr>
<tr><td>26</td><td>28</td><td>30</td><td>32</td></tr>
</table>

Mask

<table>
<tr><td>1</td><td>2</td><td>3</td></tr>
<tr><td>4</td><td>5</td><td>6</td></tr>
<tr><td>7</td><td>8</td><td>9</td></tr>
</table>

Compute the correlation value for the pixel (1, 2) and convolution value for the pixel (0,1).

Cross-correlation

Let F be the imag H be the kernel (of size 2k+1 x 2k+1), an G be the output image

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v] F[i+u, j+v]$$

This is called a **cross-correlation** operation:

$$G = H \otimes F$$

- Can think of as a "dot product" between local neighborhood and kernel for each pixel

[1M]

Convolution

- Same as cross-correlation, except that the kernel is "flipped" (horizontally and vertically)

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v] F[i-u, j-v]$$

This is called a **convolution** operation:

$$G = H * F$$

- Convolution is **commutative** and **associative**

[1M]

Correlation pixel (1,2)

4 6 8   1 2 3
12 14 16   4 5 6
20 22 24   7 8 9

= [4*1+6*2+8*3+12*4+14*5+
16*6+20*7+22*8+24*9]
= [4+12+24+48+70+96+
140+176+216] = 1044

Convolution pixel (0, 1)

0 0 0   1 2 3
2 4 6   4 5 6
10 12 14   7 8 9

= [2*4+4*5+6*6+10*7+
12*8+14*9]
= [8+20+36+70+96+126]
= 356

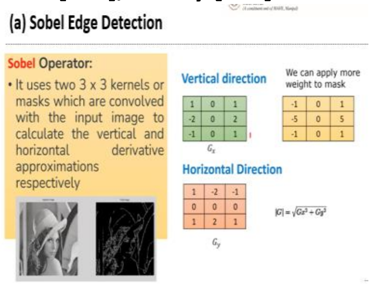
[1M]

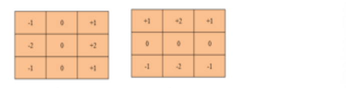
</td><td>3</td><td>CO2</td><td>L3</td></tr>
</table>

C.	Compare Sobel Vs Canny techniques for edge detection in Computer Vision.	**2**	**CO3**	**L3**

Sol:

Sobel [1M], Canny [1M]





Sobel Edge detection operator consists of 3x3 convolution kernels. Gx is a simple kernel and Gy is rotated by 90°

These Kernels are applied separately to input image because separate measurements can be produced in each orientation i.e Gx and Gy.

Following is the gradient magnitude:

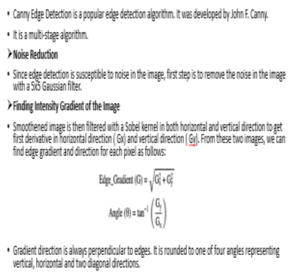
$|G| = \sqrt{Gx^2 + Gy^2}$

As it is much faster to compute An approximate magnitude is computed:

$|G| = |Gx| + |Gy|$

The angle of orientation of the edge (relative to the pixel grid) giving rise to the spatial gradient is given by:

$\theta = \arctan(Gy/Gx)$



• Canny Edge Detection is a popular edge detection algorithm. It was developed by John F. Canny.

• It is a multi-stage algorithm.

➤Noise Reduction

• Since edge detection is susceptible to noise in the image, first step is to remove the noise in the image with a 5x5 Gaussian filter.

➤Finding Intensity Gradient of the Image

• Smoothened image is then filtered with a Sobel kernel in both horizontal and vertical direction to get first derivative in horizontal direction (Gx) and vertical direction (Gy). From these two images, we can find edge gradient and direction for each pixel as follows:

$$Edge\_Gradient\,(G) = \sqrt{G_x^2 + G_y^2}$$

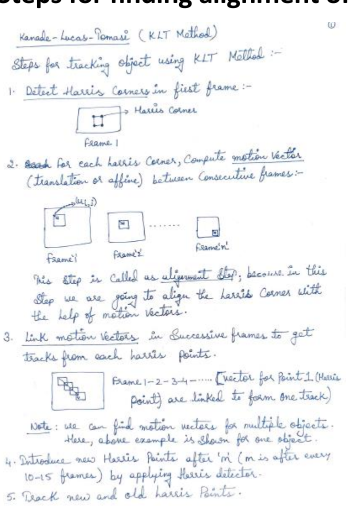
$$Angle\,(\theta) = \tan^{-1}\left(\frac{G_y}{G_x}\right)$$

• Gradient direction is always perpendicular to edges. It is rounded to one of four angles representing vertical, horizontal and two diagonal directions.

12.				
A.	Derive the motion estimation parameters required to align successive frames in a video using KLT tracking method.	**5**	**CO3**	**L3**

Sol:

Steps for tracking object using KLT method [2M]
Diagram: [1.5 M],
Steps for finding alignment of Lukus-Kanade [1.5M]



Kanade-Lucas-Tomasi (KLT Method)

Steps for tracking object using KLT Method :-

1. Detect Harris Corners in first frame :-

Harris Corner

Frame 1

2. For each harris Corner, Compute motion vector (translation or affine) between Consecutive frames :-

Frame1 Frame2 ...... Frame n

This step is Called as alignment step; because in this step we are going to align the harris Corner with the help of motion vectors.

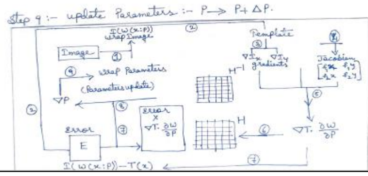
3. Link motion vectors in Successive frames to get tracks from each harris points.

Frame 1-2-3-4-...... [vector for Point 1 (Harris Point) are linked to form one track]

Note: we can find motion vectors for multiple objects. Here, above example is shown for one object.

4. Introduce new Harris Points after 'm' (m is after every 10-15 frames) by applying Harris detector.

5. Track new and old harris Points.



Finding Alignment of Lucas - Kanade :-

Goal :- Given template $T(x) \rightarrow$ find 'P' to minimize :

$$\sum_x \left[I(W(x;P)) - T(x) \right]^2$$

Step 1 :- Wrap the image with initial estimate (I with $W(x:P)$)

Step 2 :- Subtract from the template $T(x)$

$Error = I(W(x:P)) - T(x)$

Step 3 :- Compute the gradient ∇T

Step 4 :- Evaluate Jacobian : $\partial W / \partial P$.

Step 5 :- Multiply Jacobian with gradient : $\nabla T \cdot \frac{\partial W}{\partial P}$

Step 6 :- Compute Inverse Hessian Matrix : H^{-1}

Step 7 :- Multiply with error :-

$$\sum_x \left[\nabla T \frac{\partial W}{\partial P} \right]^T \left[I(W(x:P)) - T(x) \right]$$

Step 8 :- Compute ∇P :

$$\nabla P = H^{-1} \sum_x \left[\nabla T \cdot \frac{\partial W}{\partial P} \right]^T \cdot \left[I(W(x:P)) - T(x) \right]$$

Step 9 :- update Parameters :- $P \rightarrow P + \Delta P$.

B.	Consider the following input image:	**5**	**CO3**	**L3**

0	1	0	2	7
2	7	7	4	0
5	6	4	3	3
1	1	0	7	5
5	4	2	2	5

Calculate the new value of the pixel (2, 2) if smoothing is done using a (3 * 3) neighbourhood and find the following:
(i) Mean Filter
(ii) Weighted Average Filter [Mask: [1 1 1; 1 2 1; 1 1 1]
(iii) Median Filter
(iv) Min Filter
(v) Max Filter
[Each 1M]

Sol:

(a) Mean / box filter :-

$$\begin{matrix} 7 & 7 & 4 \\ 6 & 4 & 3 \\ 1 & 0 & 7 \end{matrix} * \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \frac{1}{9}\left[7+7+4+6+4+3+1+7\right] = \frac{39}{9} = \underline{4.3}$$

(b) weighted average filter :-

$$\begin{matrix} 7 & 7 & 4 \\ 6 & 4 & 3 \\ 1 & 0 & 7 \end{matrix} * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \frac{1}{9}\left[7+7+4+6+8+3+1+0+7\right] = \underline{43}$$

$$= 43/9 = \underline{4.78}.$$

(c) Median :- Arrange in ascending order

0 1 3 4 4 6 7 7 7

Median = 4

(d) Min filter = 0

(e) Max filter = 7

Note: BL refers to Bloom's Taxonomy Level.