2020

Sunday, 29 September 2024 7:55 pm

EN2550 Final Exa...

Index no.: | | | | | | | |

UNIVERSITY OF MORATUWA, SRI LANKA
Faculty of Engineering
Department of Electronic and Telecommunication Engineering
B.Sc. Engineering
Semester 4 Final Examination

EN2550—FUNDAMENTALS OF IMAGE PROCESSING AND MACHINE VISION

Time Allowed: 2 hours January 2020

ADDITIONAL MATERIAL

None.

INSTRUCTIONS TO CANDIDATES

- This paper contains 4 questions on 12 pages.
- Answer **all** the questions.
- This examination accounts for 70% of the module assessment.
- This is a closed-book examination.
- The symbols used in this paper have their usual meanings.
- Clearly state any assumptions that you may make.
- Answer the questions in the question paper itself. You may not attach extra sheets. Any extra sheet will be disregarded in assessing.
- Neat and orderly presentation is important.

Marks

Q01	
Q02	
Q03	
Q04	
Total	

Q1. (a) State an example each of [2]

i. Intensity transformations, and → Gamma, Identity, Negative, Intensity windowing

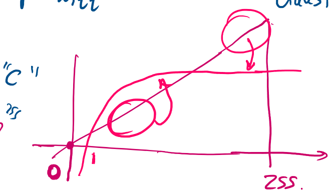
ii. Spatial filtering. → Smoothing (Average, Gaussian), Noise reduction (Median, Gaussian), Sharpning (Unsharp Mask Filter), edges (Central derivative, Gaussian derivative, Sobel, Prewitt, Canny), Corners (Harri's Corner detection), Blobs (laplacian of Gausian, DoG)

(b) Consider the intensity transformation given by

$$s = c\ln(r), \quad r \in [0, 255],$$

like a Gamma low Value

where r is the input intensity, s is the output intensity, and c is a constant to be estimated.



i. State the effect of this transformation. [2]

Compresses the Dynamic range
→ Increase brightness in dark regions, Maybe reducing brightness of lighter region

ii. In a program, what is the usual way of applying a transform of this nature? [2]

? S = cv.LUT(img, c*np.log(r/255)*255) or t = c*np.log(r/255)*255, S = t[r]

depends on "C"

Scaling is important.

iii. A student's implementation of this transformation resulted in Fig. Q1(b). What is the reason for the incorrectness? [3]

$\ln(0) \rightarrow -\infty$ undifined and closer values to zero is also negative → Use $s = c\ln(1+r)$ [0, 255]

High values

log transformation

Artifacts

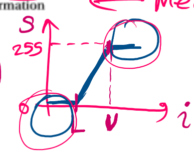
i. Original image. ii. Intensity transformed image.

Figure Q1(b): Intensity transformation **Q1**b.

(c) Figure Q1(c) shows the probability mass function $p_I(.)$ and cumulative distribution function of the intensities I of an 8-bit image. The darkest pixels (10%) and the brightest pixels (10%) are not important in this particular image. Consider the intensity transformation ← Medical Imaging

0 – 255

$$s(i) = \begin{cases} 0, & \text{if } i \le L, \\ \dfrac{(i-L)255}{U-L}, & \text{if } L < i < U, \\ 255, & \text{if } i \ge U, \end{cases}$$



where

$$\sum_{0}^{L} p_I(i) = 0.1, \quad \text{and} \quad \sum_{U}^{255} p_I(i) = 0.1.$$

i. Briefly explain the effect of this transformation. [3]

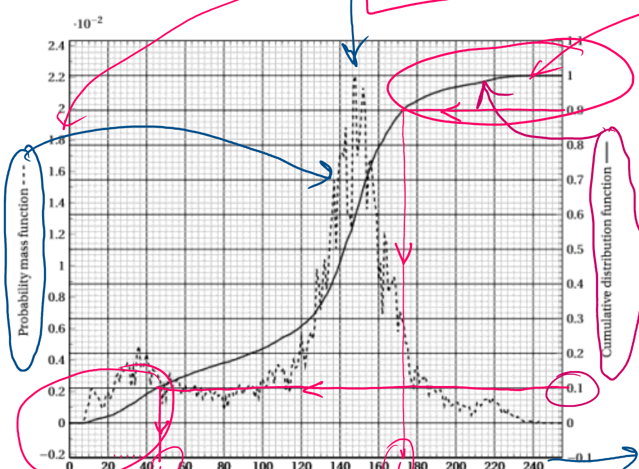
→ Unwanted Low & High 10% intensities will be saturated.
→ (L−U) scale intensities will go from (0−255)

$$n_k = (MN) \times P_r(r_k)$$

⇑

PDF $P_r(r_k)$

$$\text{CDF} = \int_0^r P_r(w)\,dw$$



$40 + \frac{20}{6} \times 2$

$L = 46.666$

$160 + \frac{20}{6} \times 4 = 173.33 = U$

Figure Q1(c): Intensity transformation **Q1**c.

Intensity

$40 + \frac{20}{6} \times 2$ $\qquad$ $160 + \frac{20}{6} \times 4 = 173.33$

$L = 46.666$ $\qquad$ $\parallel$ U

Figure Q1(c): Intensity transformation Q1c.

$U - L = 173.33 - 46.666$
$= 126.6666$

ii. Sate the values of L and U. [2]

$L = 46.6666$ $\qquad$ $U = 173.33$

iii. Compute the resulting values of pixels with intensity $I_1 = 100$ and $I_2 = 200$? [2]

$I_1 = \frac{255}{(U-L)}(i - L) = \frac{255}{126.666}(100 - 46.66) = 107.369$, $I_2 = 255$

(d) The Laplacian for a function $f(x, y)$ is

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

Discretization of this can result is a 3×3 Laplacian kernel.

i. Use the forward difference formula and backward difference formula together to obtain a symmetric expression for $\frac{\partial^2 f}{\partial x^2}$, if f is a digital image. [2]

$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) - 2f(x,y) + f(x-1, y)$

ii. Hence, write an expression for the Laplacian $\nabla^2 f$, if f is a digital image. [2]

$\nabla^2 f = [f(x+1,y) - 2f(x,y) + f(x-1,y)] + [f(x,y+1) - 2f(x,y) + f(x,y-1)]$

iii. Show this 3×3 Laplacian as a kernel. [2]

$0.f(x-1, y-1)$ $\qquad$ $-4f(x,y)$

$$\nabla^2 f = \begin{bmatrix} 0 + 1f(x, y+1) + 0 \\ f(x-1,y) - 4f(x,y) + f(x+1,y) \\ 0 + f(x, y-1) + 0 \end{bmatrix} = \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 1 & -4 & 1 \\ \hline 0 & 1 & 0 \\ \hline \end{array}$$

$(0,0)$

iv. Figure Q1(d) shows a different version of the 3×3 Laplacian kernel and an image. What is the resulting value at $(1, 1)$ of the Laplacian image? Top-left pixel is $(0, 0)$. [2]

$10 + 50 + 200 + 20 + 190 + 40 + 170 + 180 - 8(50) = 460$

v. If the image is zero padded, what would be the sum of values of the resulting Laplacian image? [1]

$(-8+5)[\ ] + (-8+3)[\ ] + (-8+8)[\ ] = -7790$

1	1	1
1	−8.0	1
1	1	1

i. Kernel.

3×3

$(0,0)$ $\quad$ $(1,1)$

0					
0	10	50	200	190	240
0	20	50	190	220	220
0	40	170	180	200	240
0	140	190	180	200	250
0	0	0			

ii. Image.

Figure Q1(d): Intensity transformation **Q1d**.

5×5 → 2 zero lines

7×7 $\quad$ 7/2 = 3 zero lines

$\frac{\partial f}{\partial x} = \frac{f(x+1, y) - f(x,y)}{1}$, $\quad \frac{\partial f}{\partial x} = \frac{f(x,y) - f(x-1, y)}{1}$

$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{f(x+1, y) - f(x,y)}{1} - \frac{f(x,y) - f(x-1,y)}{1}$

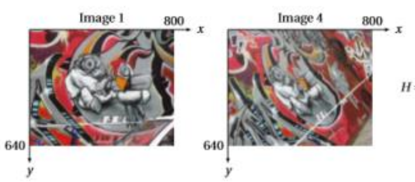
$= f(x+1, y) - 2f(x,y) + f(x-1, y)$

Q2. (a) Total least squares is useful in vision problems.

i. Show how to find the line parameters expressed as $ax + by = d$ using this method in terms of point coordinates (x_i, y_i), $i = 1, \dots, N$, and U, in regular notation. [3]

ii. In a particular case, $\bar{x} = 5.0$, $\bar{y} = 13.1$. The eigenvalues of $\mathrm{U}^\mathsf{T}\mathrm{U}$ are 21.2, 4324.1 and corresponding eigenvectors are $[-0.8974, 0.4411]^\mathsf{T}$, and $[-0.4411, -0.8974]^\mathsf{T}$. Find the gradient m and intercept c of the total-least-squares-fit line. [4]

(b) Figure Q2(b) shows the image 1 and 4 from the graffiti images, and the homography that maps image 1 to 4. Note the convention of axes.



$$H = \begin{bmatrix} 0.6638 & 0.68 & -31.2303 \\ -0.145 & 0.9713 & 148.7742 \\ 0.0004 & 0 & 1 \end{bmatrix}$$

Figure Q2(b): Image 1, image 4 and homography mapping image 1 to 4.

i. State the transformation that maps image 4 to image 1. [2]

ii. Compute the locations of the corners of image 4 when mapped onto image 1. [2]

iii. What is the size of the arrays needed to accommodate the stitched image after this mapping? [2]

iv. What are two additional steps to make the stitched result visually pleasing? [2]

(c) The scale-normalized Laplacian of Gaussian (LoG) for circularly symmetric blob detection is

$$\nabla^2_{\text{norm}} g = \sigma^2 \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right),$$

where g is the bi-variate Gaussian.

i. Obtain an expression for the radius of a black circle in a white background that would give the maximum response to this operator. [2]

ii. What is the suitable value for σ, if black circles of radius 10 pixels are to be detected. [2]

iii. Assume that two images of the scene—taken with two level of zoom—have features detected. How can the features be matched? [2]

(d) A conveyor belt carries *several types* wrapped chocolate bars. Images of reference wrapped chocolate bars are available. The factory has the following requirements:

- Counting the number of wrapped chocolate bars that pass along the belt.
- Determining if there are printing mistakes in comparison with references.

i. Propose a system for counting. [2]

ii. Propose a method for locating printing mistakes. [2]

Q3. (a) A camera calibration software gave the following parameters for a particular camera:

$$\alpha_x = 650, \quad \alpha_y = 650, \quad \beta_x = 303, \quad \beta_y = 242.$$

The camera rotation matrix and the translation matrix with respect to the world coordinate system are

$$\mathrm{R} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and,} \quad \tilde{C} = \begin{bmatrix} 0 & 0 & 2 \end{bmatrix}^{\mathsf{T}}.$$

i. Compute the camera matrix P. [2]

ii. Compute the Cartesian coordinates of the image of the point $\boldsymbol{X} = \begin{bmatrix} 2 & 3 & 10 & 1 \end{bmatrix}^{\mathsf{T}}$ [1]

iii. What will happen to the size of the image of an object when $\alpha = \alpha_x = \alpha_y$ increase? [1]

(b) Consider the epilopar geometry of two cameras. Assume that the first camera C is at the world origin with axes aligned with the world coordinate frame. The second camera C' has a rotation R and a translation $\boldsymbol{t}$ with respect to the first camera. The intrinsic camera calibration matrices are K and K$'$, respectively. $\boldsymbol{x}$ and $\boldsymbol{x}'$ are the images of the world point $\boldsymbol{X}$ in the first and the second camera, respectively.

i. Sketch the aforementioned epipolar geometry. [2]

ii. Obtain an expression for the essential matrix E. [1]

iii. Express epipolar lines associated with $\boldsymbol{x}$ and $\boldsymbol{x}'$ in terms of E. [1]

iv. Derive an expression for the Fundamental matrix F in terms of E. [1]

(c) Figure Q3(c) shows image I and I', and point $\boldsymbol{x}_1$ and $\boldsymbol{x}'_2$. The fundamental matrix relating the two cameras is

$$\mathrm{F} = \begin{bmatrix} 0 & -0.1935 & 171.0789 \\ 0.1935 & 0 & 6850.4876 \\ -171.0789 & -6850.4876 & 0 \end{bmatrix}$$

i. Compute the epipolar line $\boldsymbol{l}'_1$ and express it in the $y = mx + c$ form. [2]

ii. Compute the epipolar line $\boldsymbol{l}_2$ and express it in the $y = mx + c$ form. [2]

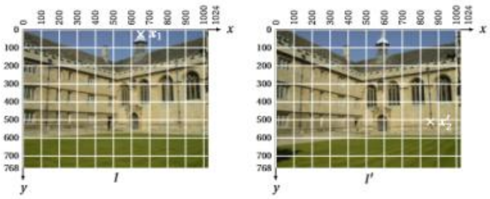


Figure Q3(c): Image I and I', and point $\boldsymbol{x}_1$ and $\boldsymbol{x}'_2$.

i. Thresholding.

	0	1	2	3
0	120	200	240	50
1	50	60	100	120
2	10	40	20	30

ii. k-means

	0	1	2	3
0	120	200	240	50
1	50	60	100	120
2	10	40	20	30

iii. k-means with spatial

	0	1	2	3
0	120	200	240	50
1	50	60	100	120
2	10	40	20	30

Figure Q3(d): Image and the grids for segmentation **Q3**d.

iii. Accurately plot these epipolar lines on the given images. [4]

(d) Figure Q3(d) shows a 3×4 image.

i. Segment this image by thresholding with the threshold of 80. [2]

ii. Carry out one iteration of k-means clustering in the intensity space. Use Manhattan distance and $K = 2$. Select $(0, 3)$ and $(2, 0)$ as initial cluster centers. [3]

iii. Carry out one iteration of k-means clustering in the intensity space and spatial space. Scale the spatial coordinated by 100. Use the parameters mentioned in **Q3**(d)ii. [3]

Show the results in the grids in Figure Q3(d)

Q4. (a) Sate the output of the following vision tasks: [5]

No.	Task	Output
1.	Image classification	
2.	Object detection	
3.	Semantic segmentation	
4.	Image captioning	
5.	Panoptic segmentation	

(b) Consider a dataset of three classes of images, say, dogs, cats, and bunnies. An image is represented by a 5-dimensional vector. A liner classifier (after the learning process) is represented by

$$W = \begin{bmatrix} -0.43 & 0.53 & -0.29 & -1.03 & 0.53 \\ -0.55 & -0.69 & 0.95 & -0.06 & 1.16 \\ 0.54 & -0.28 & -1.44 & -0.21 & -0.93 \end{bmatrix}, \text{ and}$$

$$\boldsymbol{b} = \begin{bmatrix} 2.64 & -0.64 & -3.27 \end{bmatrix}^T.$$

The first row of W and $\boldsymbol{b}$ is the one-vs-all classifier for dogs, the second for cats, and the third for bunnies.

i. State the steps for representing an image with a 5-dimensional feature vector. **[2]**

ii. Determine the class of a image with feature vector $\boldsymbol{x} = \begin{bmatrix} 5.31 & 1.51 & 1.61 & 8.1 & -8.45 \end{bmatrix}^T$. **[3]**

(c) Consider a neural network implementation for MNIST digit recognition. Assume that each digit is 28×28 in grayscale. There are 10 digits, from 0 to 9. The network flattens the input first. There is a dense layer of 128 nodes, a dropout layer of 20% dropout and a 10-node dense softmax output layer.

i. Sketch the network. **[3]**

ii. Compute the number of learnable parameters in this network. [2]

iii. If the network is altered to have, first, a convolutional layer of (3×3) convolutions with 64 filters, followed by a max-pooling layer of (2×2) pooling window, and then the aforementioned dense network, compute the number of learnable parameters. [3]

(d) Consider an in-vehicle vision system the recognizes speed-limit traffic signs along an expressway. There are several such speed limits, e.g., 50, and 100. A speed limit sign is circular of diameter 450 mm with a red border and black numbers in white background (e.g., Figure Q4(d)). Assume that the system has to detect these signs irrespective of the distance within a range of 20 m to a maximum dictated by the camera. The resolution of the camera is $2,048 \times 1,280$. Horizontal field of view is $\pm 30°$.

i. If the sign must occupy at least a width (or height) of 20 pixels for detection, compute the maximum distance for speed sign detection. [3]

Figure Q4(d): A example of a speed sign.

ii. Suggest a conventional method for detection. [3]

iii. How can the same camera be used for obstacle detection? [1]