General Sir John Kotelawala Defence University, SRI LANKA
Faculty of Engineering
Department of Electrical and Electronic Engineering
B.Sc. Engineering
Semester 5 Final Examination

ET3112—IMAGE PROCESSING AND MACHINE VISION

Time Allowed: 2 hours April 2021

ADDITIONAL MATERIAL

None.

INSTRUCTIONS TO CANDIDATES

- This paper contains 4 questions on 8 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.
- Neat and orderly presentation is important.

Q1. (a) State two current applications each of [4]

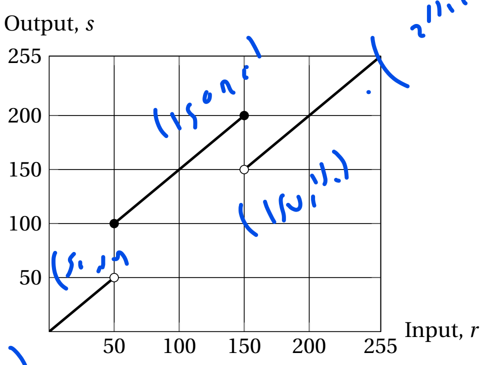
i. image processing, and → noise filter, blurring, segmentation

ii. computer vision. → img, AI, ...

(b) Figure Q1(b) shows an intensity transfer function and an input image.

i. Apply this transfer function to the input image and show the resulting values in the grid shown. [6]

ii. Suggest a use of this operation. [2] high lit (50-150)



i. Transfer function.

150	200	220
100	100	140
50	120	100
10	10	50

ii. Input image.

200	200	220
150	150	190
50	[illegible]	150
10	10	50

iii. Grid for output.

Figure Q1(b): Intensity transfer function, input image, and grid for output.

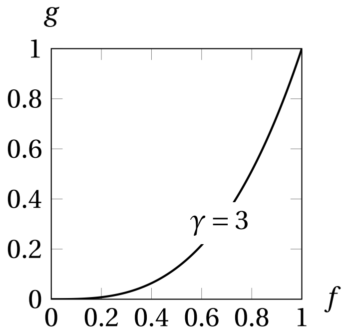
(c) Consider the gamma correction designated by

$$g = f^{\gamma},$$

where f is the input image, and g is the output image. $g, f \in [0,1]$ and $\gamma = 3$. Figure Q1(c) shows this transfer function, an input image, and a grid for showing the output.

i. What is the effect of this transformation? [2]

ii. Show the gamma-corrected output in the gird shown. [4]



i. Transfer function.

200	200	255
100	100	200
50	50	100
10	10	50

ii. Input image $f(i,j)$.

		255
		4.57
		16.32
0.016		2.04

iii. Grid for output.

Figure Q1(c): Grama correction function, input image, and grid for output.

(d) Histograms are important in characterizing images.

i. Figure Q1(d)i. shows two images. Sketch the histograms of these and compare. [4]

ii. Figure Q1(d)ii. shows an image and a grid. Carry out histogram equalization and show the resulting image in the grid. [3]

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

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

i. Image for histogram comparison.

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

3	2					2	3
3	2						
2							
2							
2							3
				6	1		3
				(	6	9	5
				(	6	7	6
				6	6	2	6
7	7	7	7	7	7	7	7

ii. Image for histogram operations and grid for showing the results.

Figure Q1(d): Images for **Q1**d.

Q2. (a) State the effect of the spatial filtering kernels shown in Figure Q2(a) would achieve. [**4**]

vert'l horiz'l

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

avg

i.

−1	−1	−1
0	0	0
1	1	1

ii.

−1	0	1
−1	0	1
−1	0	1

iii.

0	−1	0
−1	5	−1
0	−1	0

iv.

sharpening

SV

Figure Q2(a): Kernels.

SH

(b) Figure Q2(b) shows an image, a kernel, and a grid for displaying the output.

i. Apply the edge detection operation on the image using the kernel shown in the same figure. Show the output in the given grid. [**4**]

ii. Apply a 2×2 max-pooling operation (stride 2) on the result. A max-pooling operation selects the maximum value in the region of concern and places it where the (0,0) location of the kernel is in the image. [**2**]

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

i. Image

SH

−1	0	1
−2	0	2
−1	0	1

ii. Kernel

	670	510	160	
	680	740	210	

iii. Grid for edge output

	180			

iv. Grid for max-pool output

Figure Q2(b): Image, kernel, and output grid for **Q2**b.

0.8825		

Figure Q2(c): A part of un-normalized 3×3 Gaussian kernel.

(c) Figure Q2(c) shows a part of the un-normalized 3×3 Gaussian kernel with $\sigma = 2$.

i. Fill the other entries in the kernel. [**3**]

ii. Compute the normalized kernel. [**2**]

iii. Why are all the nine entries nearly the same? [**3**]

(d) Consider an image with several types of coins. How can a spatial filtering approach be used to detect the type of each coin? [**7**]

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

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

$P = K[R|t]$

$t = -R\tilde{C}$

$$K = \begin{bmatrix} \alpha_x & 0 & \beta_x \\ 0 & \alpha_y & \beta_y \\ 0 & 0 & 1 \end{bmatrix}$$

i. Compute the camera matrix P. [**2**]

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

iii. What will happen to the size of the image of an object when $\alpha = \alpha_x = \alpha_y$ increase? [**2**]

(b) Consider an images sensor of size 1200×1600 made of square pixels. Pixel size is 4.4 μm. The focal length of the lens is 25 mm. The distance between the lens and the imaging surface is 100 mm.

i. Compute the field of view in mm (vertical, and horizontal). [**4**]

ii. If a coin of diameter 20 mm is kept on the imaging surface, compute the diameter on the image of the coin in pixels. [**3**]

(c) Segment the image in Figure Q3(c)i. using thresholding. Choose $\boxed{175}$ as the threshold. [**6**]

(d) Segment the image in Figure Q3(c)i. using region growing starting at the seed at (7,7). Choose a threshold of 11 and a 4-connected neighborhood. Show the result in Figure Q3(d). [**6**]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	7	27	46	59	77	93	101	111	107	99	106	93	86	71	48	20
1	27	46	66	85	97	107	111	94	77	77	74	82	97	81	75	48
2	52	66	85	83	82	86	83	72	68	68	68	68	91	89	80	87
3	68	85	97	95	111	97	90	97	152	164	139	105	77	119	107	96
4	75	97	98	109	104	87	137	173	192	195	195	190	104	112	120	117
5	83	117	122	93	72	133	165	216	221	234	230	226	167	79	129	132
6	88	124	125	100	115	119	213	225	237	236	231	228	182	88	91	121
7	104	201	230	142	98	145	222	235	217	235	232	229	184	101	121	119
8	113	211	223	228	233	235	235	241	239	235	229	227	181	109	103	136
9	103	208	207	199	119	124	223	239	242	240	233	213	176	112	92	125
10	77	127	198	163	144	114	152	235	243	241	235	205	154	120	97	119
11	80	116	124	156	142	90	134	172	227	233	226	137	171	102	110	111
12	79	108	134	140	104	82	89	170	130	107	98	88	100	115	108	91
13	53	183	218	157	143	99	82	134	161	129	151	132	130	116	88	66
14	34	179	222	105	124	127	146	152	121	152	154	138	119	87	66	44
15	20	40	66	85	111	127	128	127	129	119	126	103	84	61	37	22

d < 11

segment

i. Image for segmentation

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0																
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																

ii. Grid for thresholding result.

Figure Q3(c): Image for up-scaling and grid for showing results.

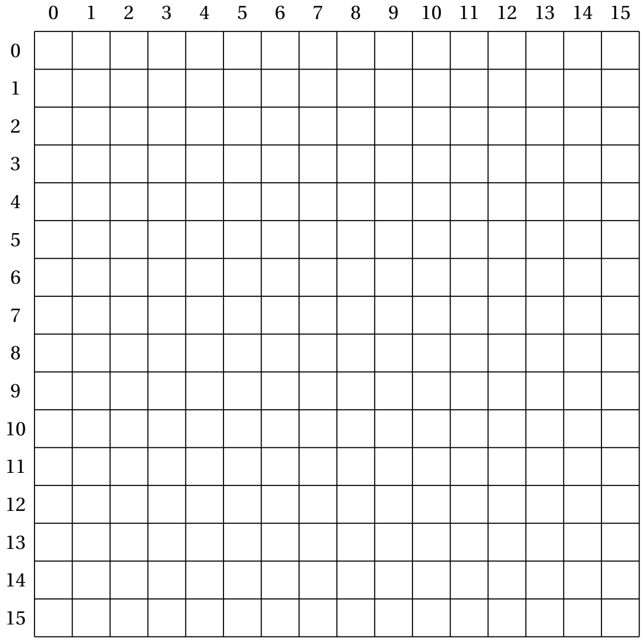


Figure Q3(d): Grid for region growing result for **Q3**d.

Q4. (a) In regard to a typical convolutional neural network for image classification, answer the following questions with reasons:

i. In what part of the network are most of the learnable parameters? [**2**]

ii. What part of the network is most heavy in computations? [**2**]

(b) Convolutional neural networks (and other image processing systems) sometimes need an image to be upscaled. An image and a grid are shown in Figure Q4(b). This image is to be scaled up by a factor of 1.5 using bilinear interpolation to generate a 12×12 image. Compute the pixel value at location $(9, 10)$ of the upscaeld image. [**4**]

(c) i. If 50,000 images of size $32 \times 32 \times 3$ (e.g., CIFAR10) are available for training, and linear classification is to be done using vectorized images themselves, briefly describe the procedure for training such a classifier. [**2**]

ii. What would the weights of each classifiers look like when seen as an image? [**2**]

iii. 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

$$W = \begin{bmatrix} -0.3 & 0.6 & -0.3 & 0.5 & 0.4 \\ 0.5 & 0.3 & -0.2 & -0.6 & -0.5 \\ 0.6 & -0.3 & 0.7 & -0.6 & 0.1 \end{bmatrix}, \text{ and}$$

$$\boldsymbol{b} = \begin{bmatrix} 0.2 & 2.2 & 3.2 \end{bmatrix}^{\mathrm{T}}.$$

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. Determine the class of an image with the feature vector $\boldsymbol{x} = \begin{bmatrix} 1.3 & -1.2 & 0.8 & 3.1 & 2.0 \end{bmatrix}^{\mathrm{T}}$. [**3**]

dense

(d) Consider a neural network implementation for digit recognition. There are 10 digits, from 0 to 9. Assume that each digit is 32×32 in grayscale. The network flattens the input first. There are two dense layers one connecting to 100 nodes and the other connecting to 20 nodes. These 20 nodes connect to the output nodes.

i. How many nodes must be there in the output? 10 [**2**]

100 20 10

ii. Sketch the network. [**2**]

iii. Compute the number of learnable parameters in this network. [**3**]

199	205	208	214	198	179	134	130
201	206	210	215	161	55	58	61
202	208	212	204	187	101	57	62
203	208	213	205	230	61	66	49
203	207	213	103	114	100	98	138
201	207	213	139	107	102	71	65
201	206	211	215	218	51	58	83
198	205	209	213	218	63	57	60

i. Image

	0	1	2	3	4	5	6	7	8	9	10	11
0												
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												

ii. Grid for result.

Figure Q4(b): Image for up-scaling and grid for showing results. Compute the value at $(10, 11)$ only.

iv. Consider a convolutional neural network for the same purpose with, first, a convolutional layer of 3 × 3 convolutions with 64 filters, followed by a max-pooling layer of 2 × 2 pooling window. After flattening, ~~the nodes densely connect to~~ 64 nodes, and these 64 nodes densely connect to the output nodes. Compute the number of learnable parameters. [**3**]

$64 \times 30 \times 30 \rightarrow 64 \times 15 \times 15 \rightarrow 64 \rightarrow 10$

$64 \times (1 \times 3 \times 3 + 1)$