GENERAL SIR JOHN KOTELAWALA DEFENCE UNIVERSITY
Faculty of Engineering
Department of Electrical, Electronic and Telecommunication Engineering

B.Sc. Engineering Degree
Semester 5 Examination – May 2024
(Intake 39 – EE, ET)

ET 3112 – IMAGE PROCESSING AND MACHINE VISION

Time allowed: 3 Hours 06 May 2024

ADDITIONAL MATERIAL PROVIDED

None

INSTRUCTIONS TO CANDIDATES

This paper contains 4 questions on pages 3 to 8.

Answer ALL questions

This is a closed book examination

This examination accounts for 70% of the module assessment. The marks assigned for each question and parts thereof are indicated in square brackets

If you have any doubt as to the interpretation of the wordings of a question, make your own decision, but clearly state it on the script

Assume reasonable values for any data not given in or provided with the question paper, clearly make such assumptions made in the script

All examinations are conducted under the rules and regulations of the KDU

DETAILS OF ASSESSMENT

Learning Outcome (LO)	Questions that assess LO	Marks allocated (Total 100%)
LO1	Q1	7
LO2	Q2	7
LO3	Q1, Q3	36
LO4	Q2, Q3	10
LO5	Q4	10
LO6	Q2, Q4	30

This Page is Intentionally Left Blank

Question 1

(a) List one of the following image processing operations with one of the application areas to achieve the most appropriate paring: [05 marks]

Image processing operations:	Application areas:
i. Histogram equalization	i. Photo processing
ii. Segmentation	ii. Medical image processing
iii. Object detection	iii. Autonomous driving
iv. Re-identification	iv. Security and surveillance
v. Visual SLAM	v. Robotics

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

i. Judge the purpose of this transformation. [02 marks]

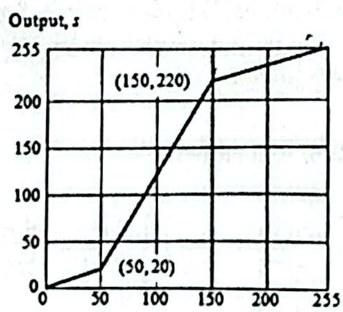
ii. Apply this transfer function to the input image and show the resulting pixel values in the grid shown, in Figure Q1(b) iii. [03 marks]

iii. The image in Q1(b)ii. is transformed using the operation represented by

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

where c is a constant. Discuss the effect of this. [02 marks]

$\frac{255-220}{255-150} = \frac{1}{3}x + 170$



2/5

i. Transfer function.

200	200	255
140	140	200
100	100	120
25	25	50

ii. Input image.

iii. Grid for output.

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

(c) Figure Q1(c) shows an image, a kernel, and a grid for displaying the results of spatial filtering.

10	50	50	40	20
20	50	50	40	30
40	170	200	100	60
30	40	50	30	20

i. Image to be filtered.

0.1	0.1	0.1
0.1	0.2	0.1
0.1	0.1	0.1

ii. Kernel.

iii. Grid for results.

Figure Q1(c): Image for spatial filtering and the kernel.

i. Carry out the filtering and show the results in the grid. Compute only for the region where the kernel can be placed without going out of the image. [04 marks]

ii. What is the reason for pixel (2, 2) in the result to be high? [02 marks]

(d) Fig. Q1(d) shows an image enhancement pipeline. The purpose is to emphasize the edges.

i. Suggest kernels for operations represented by $*k_1$ and $*k_2$? [02 marks]

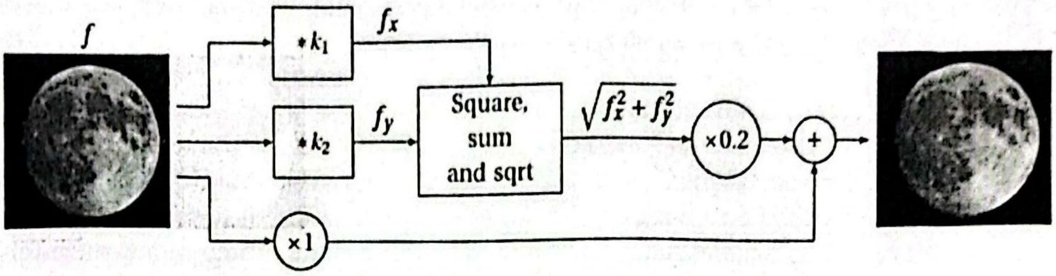


Figure Q1(d): Block diagram of an image enhancement pipelines.

ii. Sketch $\sqrt{f_x^2+f_y^2}$. [03 marks]

iii. What is the purpose of this pipeline? [02 marks]

Question 2

(a) Consider a projective camera. If the world frame to camera rotation matrix is R, the world frame to camera translation vector is t, the focal length in mm is f, principal point coordinates are (p_x , p_y), and the pixel-per-mm parameters of the imaging device are m_x and m_y, write an ex- pression for the camera matrix P. [05 marks]

(b) The following parameters are given in relation to a camera:

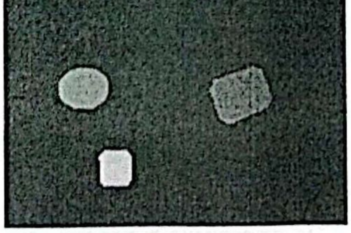
$$t=[0\quad 0\quad 0]^{T},\quad K=\begin{bmatrix}480 & 0 & 400.2\\ 0 & 480 & 299.7\\ 0 & 0 & 1\end{bmatrix},\quad R=\begin{bmatrix}0.707 & 0.707 & 0\\ -0.707 & 0.707 & 0\\ 0 & 0 & 1\end{bmatrix}$$

i. Compute the camera matrix P. [03 marks]

ii. Compute where the point $X=[0.5\quad 1.0\quad -100\quad 1]^{T}$ in the world frame is projected on the image plane [03 marks]

(c) Fig. Q2(c) shows an image with three gems and the corresponding connected components analysis result. The statistics returned the following pixel counts as areas of connected components: 36049, 1104, 758, and 489. Pixel size is 2.2 μm. The focal length of the lens is 4 mm. The distance between the lens and the imaging surface is 720 mm.

i. Earrings image.



ii. Intensity-indexed connected components

i. Compute the area of each gem in mm^2. [03 marks]

ii. What are the image processing steps involved to obtain the area in pixels? [03 marks]

iii. List three drawbacks of this method of size measurement. [03 marks]

(d) Fig. Q2(d) shows a 3-D point X and two cameras C_1 and C_2. The cameras are separated by a distance b. The point X is projected to the image planes of the cameras as x_1 and x_2.

i. Mark x_1 and x_2. [02 marks]

ii. If X is not known but x_1 is known, mark possible locations of x_2. [02 marks]

iii. Mathematically express a constraint that relates x_1 and x_2. [02 marks]

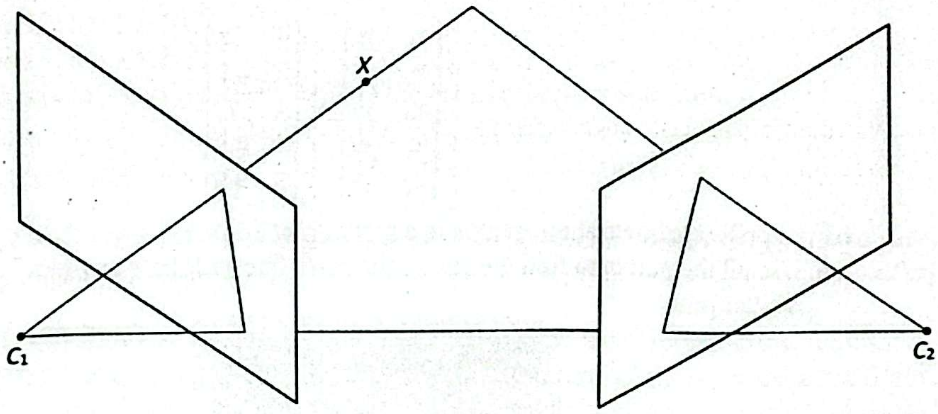


Figure Q2(d): Epipolar geometry

Question 3

(a) Name two applications of line fitting in image processing. [04 marks]

(b) Consider a dataset consisting of three classes: class A, class B and class C, respectively. After the learning process, a linear classifier is given as follows:

$$W = \begin{bmatrix} 2 & -2 \\ -1 & 0.25 \\ -2.5 & 2 \end{bmatrix}^T \in \mathbb{R}^{3\times 2}, \text{ and } b = [0.4 \quad 1.5 \quad -2.5]^T$$

Three data samples x_1, x_2 and x_3 are given by

$$\begin{bmatrix} 3 & -3 & -4 \\ -5 & 4 & -2 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

These data samples are fed to the liner classifier. Output of the linear classifier for i−th data sample is given by $y_i = f(W^T x_i + b) \in \mathbb{R}^{3\times 1}$. Here, f (.) is the sigmoid function.

i. Compute the output of the linear classifier for three data samples x_1, x_2, and x_3. [06 marks]

ii. Suppose that the output of the linear classifier is

$$Y = [y_1 \quad y_2 \quad y_3] = \begin{bmatrix} 0.80 & 0.012 & 0.008 \\ 0.15 & 0.75 & 0.10 \\ 0.06 & 0.04 & 0.90 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

Based on the above output, determine data samples x1, x2, and x3 are either within the same class or belong to different classes? Justify your answer. [05 marks]

(c) Figure Q3(b) shows an agriculture field. A dataset is formed by extracting the pixel positions of plants. It should be noted that the process of extracting the pixel positions is not perfect and some pixel locations may have been altered. These pixel points are known to form a line. The resulting dataset consists of a set of points with their corresponding x and y coordinates are given below.

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \\ x_5 & y_5 \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ 2 & 2.5 \\ 4 & 7 \\ 5 & 8.5 \\ 6 & 9.5 \end{bmatrix}$$

Figure Q3(c) shows above points in a scatter plot.

i. Use all the points to find the least-squares-fit line and show graphically in the same scatter plot. [03 marks]

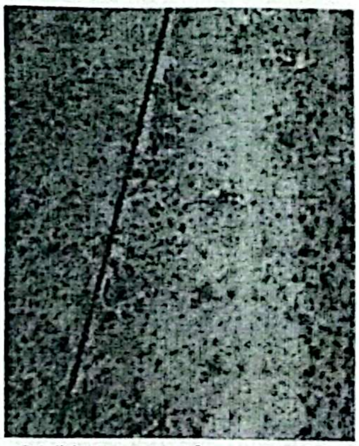


Figure Q3(b): Image of an agriculture field.

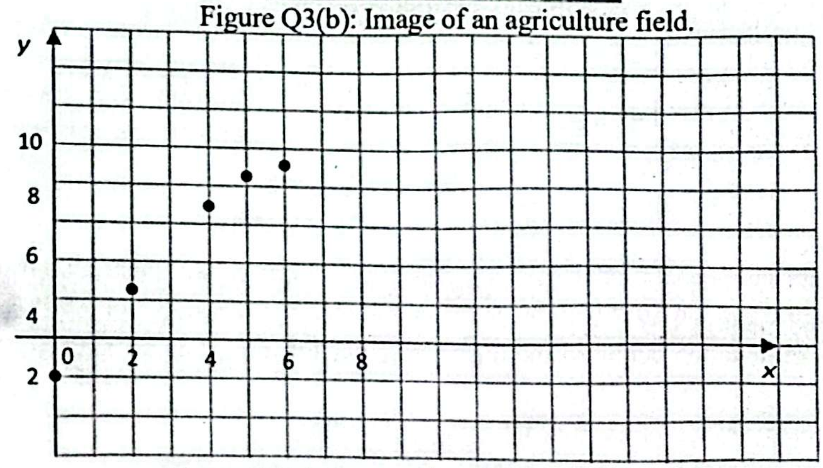


Figure Q3(c): Scatter plot of points for line fitting.

ii. Write down the values of the slope (m) and intercept (c) of the line based on you least-squares solution. [02 marks]

(d) Standard loss function of the least-squares is given by

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

here, y_i and $\hat{y}_i$ are true value and least-squares output, respectively. To reduce the impact of outliers, a modified loss function is introduced. It is given as

$$P_a(L) = \frac{L^2}{L^2 + a^2}$$

Here, a is a scale parameter. Three schemes are proposed for setting a,
Scheme 1: a = 1,
Scheme 2: a = 1×10^{-4},
Scheme 3: a = 1×10^{3},
which scheme do you select? Explain the reason for your selection.

Question 4

(a) Why are multilayer perceptrons (MLPs) not well-suited for structured grid-like data, such as images and sequences? [04 marks]

(b) Figure Q4(b) shows an image, a CNN kernel, and a grid for displaying the output.
i. Apply the convolution operation on the image using the kernel shown in the same figure. Here, both padding and stride are set to 1. Show the output in the given grid. [04 marks]

0	10	20	30
10	0	20	200
40	20	20	200
190	180	190	200

i Image

$\frac{0}{5}$	$\frac{0}{5}$	$\frac{1}{5}$
$\frac{-1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
$\frac{-1}{5}$	$\frac{-1}{5}$	$\frac{0}{5}$

ii CNN Kernel

iii Grid for CNN Kernel output

iv Grid for activation function output

Figure Q4(b): Image, CNN kernel, and output grid for Q4b.

ii. The activation function is defined as f (x) = max(15, x). Now, apply this activation function on the result. [02 marks]

(c) Consider the simple neural network shown in Figure Q4(c). In this network, x_1, x_2 and x_3 are the inputs, while h is the outputs of the hidden layer. The final output of the neural network is denoted as $\hat{y}_1$ and $\hat{y}_2$. Throughout each layer, the ReLU activation function (f(a) = max(0, a)) is applied. Write down the mathematical relationship between the inputs and the outputs. [05 marks]

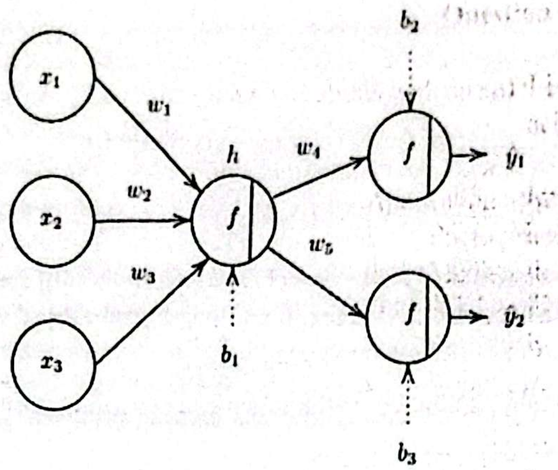


Figure Q4(c): Simple neural network.

(d) Backpropagation and stochastic gradient descent are the two main operations in training a neural network. Consider the following function,

$$f(w) = 3w_1^2 x_1 + 2w_2 x_2 + b.$$

Starting with initial values of $w_1 = 3$, $w_2 = 2$, $b = 0$, $x_1 = 2$, and $x_2 = 3$, perform one iteration of gradient descent with a learning rate of 0.2. What will be the updated value of w_1 and $f(w)$ after one iteration? [05 marks]

(e) Figure Q4(e)i shows the Hough parameter space corresponds to an image. Draw five pixels corresponds to each of Hough parameter pair in Figure Q4(e)ii. [08 marks]

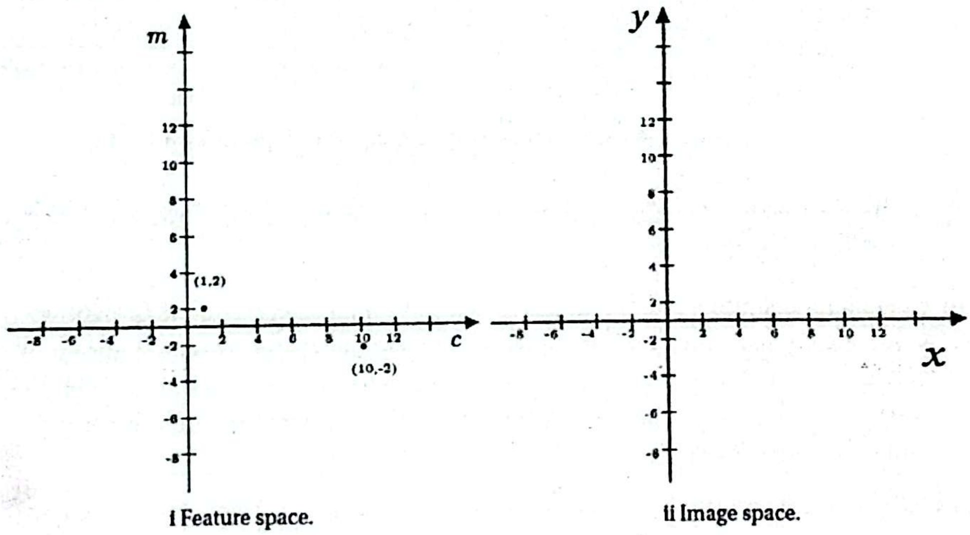


Figure Q4(e): Feature space and image space.

End of question paper