THE UNIVERSITY OF NEW SOUTH WALES

Sample Class Test

Data xx.xx.xxx
COMP9517

Computer Vision

Time Allowed: XX mins
Total Number of Questions: **5**
Maximum marks **40**
All questions must be answered in the space provided on the question paper.
This paper **may not be retained** by the candidate.

Note: **This is a sample test paper only and NOT a full test paper, in terms of time and marks.**

Question 1 (10 marks total)

Multiple Choice, **choose one answer only by a tick in front:**

1) For feature selection in pattern recognition, which one of the following is INCORRECT:
   a) select features that are translation, rotation and scale variant in images
   b) select features that can handle occlusion and projective distortion for 3-D objects in images
   c) select features that are invariant to translations in time and changes in amplitude
   d) select features that can handle non-rigid deformations common in 3-D vision

2) Which of the following statements on human vision are CORRECT:
   a) human vision is provided by light-sensitive receptors called rods and cones
   b) rods and cones operate day and night to provide colour vision
   c) rods and cones are equally distributed on the retina
   d) both luminance and chrominance are equally informative

3) One objective of image segmentation is to:
   a) decompose the image into parts for further analysis
   b) recognise the object in image
   c) improve the quality of the image
   d) classify the image into a category

4) For reliable feature detection and matching, which of the following statements is INCORRECT:
   a) the detector should detect the same point independently in both images
   b) the detector may detect different interest points as the image undergoes changes
   c) the descriptor should be distinctive to reliably match the feature points
   d) the transformation between images should be robust to erroneous correspondences.

5) Which one of the following statements about the Hough transform is INCORRECT:
   a) it is based on a voting scheme
   b) a line in the image corresponds to a point in Hough space
   c) a point in the image space maps to a line in Hough space
   d) it is easy to pick a grid size for the bins in Hough space

….

Answers in red

Question 2 (3 marks)

3

Consider imaging a scene and processing it to produce a binary image. Can this be achieved for the following scenes? Also explain why or why not.

i. The input image is a typed sheet of paper that is scanned using a page scanner. The objective is to recognise the typed characters and make an ASCII file that can be edited using a word processor.

ii. The input image is a satellite image of UNSW campus taken during winter. By using suitable algorithms, the objective is to create an image where a 0 indicates the presence of a camellia bush and a 1 indicates the absence of a camellia bush.

(i) Yes

Reason:

- Typed character recognition with high accuracy is feasible today, as there are well defined fonts, sizes, colours, with much less variation than handwritten characters for example.
- Once each character is recognised, it is easy to assemble them as a text file for further processing by a word processor.

(ii) No in general

Reason:

- To recognise pixels corresponding to a camellia bush, good segmentation and highly accurate object recognition is necessary.
- In winter, camellia is in full bloom, the tree is full of leaves and flowers- this is background knowledge. Colour segmentation may work upto a point, but a mix of green and multi colours for flowers make recognition of a camellia bush fairly challenging.
- there are also many other objects in the scene, such as buildings, roads, car parks, other trees, etc. Object recognition of multiple objects is not easy.

Question 3 (3 marks)

The results obtained by a single pass through an image of some two-dimensional masks can also be achieved by two passes of a one-dimensional mask.

Show that the result of using a 3 x 3 smoothing mask with coefficients 1/9 can also be obtained by first passing through an image the mask [1 1 1], followed by a pass of its vertical counterpart, and finally scaling the result by 1/9.

The operation of concern is convolution. Consider the 3x3 subimage I below:

$$I = \begin{matrix} f(x-1, y-1) & f(x-1, y) & f(x-1, y+1) \\ f(x, y-1) & f(x, y) & f(x, y+1) \\ f(x-1, y+1) & f(x+1, y) & f(x+1, y+1) \end{matrix}$$

By convolving the 3x3 smoothing mask with coefficients 1/9 with the above subimage, the central value becomes

$$1/9 (f(x-1, y-1) + f(x-1, y) + f(x-1, y+1) + f(x, y-1) + f(x, y) + f(x, y+1) + f(x-1, y+1) + f(x+1, y) + f(x+1, y+1)) \quad \text{-------- (1)}$$

Now pass the mask [1 1 1] through the subimage I, the central values become:

$$I1 = \begin{matrix} f(x-1, y-1) + f(x-1, y) + f(x-1, y+1) \\ f(x, y-1) + f(x, y) + f(x, y+1) \\ f(x-1, y+1) + f(x+1, y) + f(x+1, y+1) \end{matrix}$$

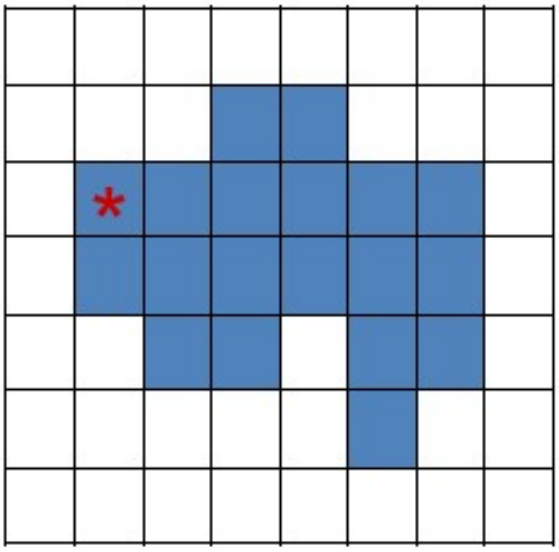
Then pass the mask $\begin{matrix} 1 \\ 1 \\ 1 \end{matrix}$ through I1:

$$I2 = f(x-1, y-1) + f(x-1, y) + f(x-1, y+1) + f(x, y-1) + f(x, y) + f(x, y+1) + f(x-1, y+1) + f(x+1, y) + f(x+1, y+1)$$

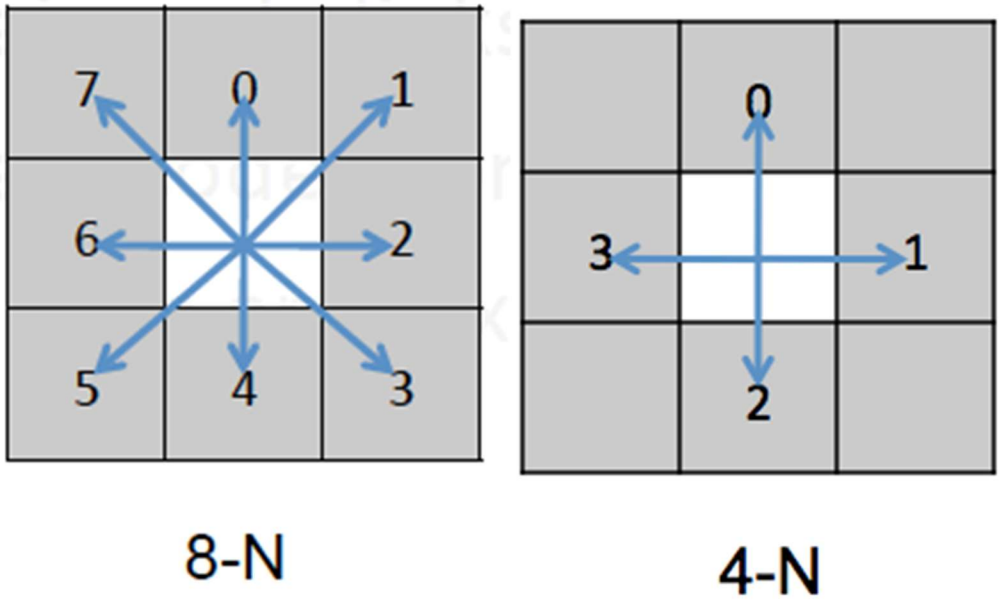
Clearly, (1) = 1/9 x I2
This is true for every pixel in the original image.

Question 4 (4 marks)

Provide 4N and 8N boundary coding solution for the following segmented image. Use the illustrated encoding schemas. Consider the square with a red star as the starting point. Please specify whether you traversed through clockwise or counter-clockwise direction.



Encoding schemas to be used are illustrated below:



Chain code representation using 8N

clockwise – 21232445075670

(or counter-clockwise – 43213410067656)

Chain code representation using 4N

clockwise – 11012112232003323030

(or counter-clockwise –21210112201003303233)

Note: all the above may be illustrated using diagrams, which would also make it easier to check.

Question 5 (5 marks)

a) Briefly describe the three recognition problems in computer vision discussed in the class. Explain their differences and the main approaches for each task.

b) You were asked to use a histogram in the assignment for finding the most frequent local pixel value.

i. Explain what a histogram is.
ii. How do you use the intensity histogram to find the most frequent local pixel value?

a)

Object detection: know what we are looking for; Quickly scan an image to determine where a match may occur
Instance recognition: Have a specific rigid object to recognize: Search for characteristic feature points and verify that they align in a geometrically plausible way
General category recognition: Recognising instances of extremely varied objects, such as animals or furniture; Techniques rely on the presence of features ("bag of words" model), and their relative positions (part-based models), usually requires segmenting the image into semantically meaningful regions

b)

i. A histogram is a graphical representation of the distribution of numerical data. To construct a histogram, the first step is to "bin" the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent and are usually of equal size. If the bins are of equal size, a rectangle is erected over the bin with height proportional to the frequency, the number of cases in each bin.
ii. Find the bin with maximum count among all bins, and then find the pixel value corresponding to that bin.