Two hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Computer Vision

Date: Thursday 17th May 2012
Time: 09:45 - 11:45

Please answer any THREE Questions from the FOUR questions provided

Use a SEPARATE answer book for each Question.

For full marks your answers should be concise as well as accurate. Marks will be awarded for reasoning and method as well as being correct.

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are not programmable and do not store text.
1.

**Figure 1.** Left: A grayscale (0-255), 6 by 10 pixelated image. Note that the square grid-lines around pixels, and around the entire image, are for the purposes of illustration only. **Right:** The actual grayscale values for the top three rows of this image.

**Figure 2.** A 3 by 3 filter, with the centre as marked.

A COMP37212 student wishes to analyse the image in Figure 1, by performing a convolution with the 3 by 3 filter shown in Figure 2.

a) Compute the value of the convolution at the 3 places (A, B, and C) as marked in the above Figure. Make sure to include your working in your submitted answer. [3 marks]

b) Hence (or otherwise), describe the configurations of the 3 by 3 neighbourhood that would give:
   i. A zero response
   ii. A response of maximum modulus and positive sign
   iii. A response of maximum modulus and negative sign
   when convolved with the above operator. Give the reasoning behind your answer in each case. [3 marks]

c) By reorganising the values in the filter given above, construct a new filter which gives a maximum positive response to the structure shown at the right.

How would you describe the class of objects being detected here? [2 marks]
d) In considering a one-dimensional image, with image value \( f(n) \) at pixel \( n \), the student notes the following formula:

\[
[f(n+1) - f(n)] - [f(n) - f(n-1)] = f(n+1) - 2f(n) + f(n-1).
\]

Hence (or otherwise), explain how filters such as the one shown in Figure 2, along with suitable first-derivative filters such as the Sobel or Prewitt operators, can be used to construct an edge-detector suitable for use on two-dimensional images.

You should use the results of the convolution at A and B in Figure 1 (or similar) to explain how such a detector is capable of sub-pixel edge location accuracy.

[8 marks]

e) Explain briefly how the concepts of:
   i. Scale  
   ii. Non-maximum suppression
are incorporated into a practical edge-detector.

[2 marks]  [2 marks]
2.

A COMP37212 student is trying to develop a grasping robot that can lay tables. As a first step, she considers the image analysis problem of finding a single isolated fork on a table. See Figure 3 for some example images.

![Figure 3: Example images of single forks.](image)

After some thought, she decides that it may be useful to implement the Hough Transform for lines to help with this task.

a) Explain *briefly* the principle of the Hough Transform for lines. [4 marks]

b) What steps would you need to take in pre-processing these image before applying the Hough Transform? What features of the images shown in Figure 3 would you expect the Hough Transform to be useful in identifying? [2 marks]

The student is then given a large, suitably-annotated set of images of various single forks.

c) Describe in detail the steps involved in constructing a statistical shape model using this training set of annotated fork images. [8 marks]

The student is now presented with two classes of fork images, each containing several forks and several types of forks, but one set contains only non-overlapping, isolated forks, whilst the second set contains instances of overlapping forks.

d) Compare and contrast the predicted performance of:
   i. A suitably constructed and suitably initialized Active Shape Model (ASM),
   ii. A suitably initialized Active Contour Model (ACM),
   iii. The Hough Transform for lines,
when applied to the task of locating forks in each of these classes of images. You may find it useful to include sketches or diagrams to describe the expected results of each method. You should ensure that you explain how the expected results lead to the predicted performance in each case. [6 marks]
3.

You are asked to develop a computer vision system that can detect cars (but not motorbikes) from side views, such as in the images below (Figure 2).

![Figure 2](image)

a) Describe one method for detecting interesting feature points, which occur at a range of scales in an image (6 marks)

b) Describe one method for computing the “signature” representing the image structure around a point, assuming the scale and orientation have been estimated (6 marks)

c) Suppose that we have computed clusters of features from a training set, and determined how likely features in each cluster are to be part of a car. Describe how this information could be used in a “Bag of Features” car detector (8 marks)
4.

a. Explain what **calibrated camera** means. (2 marks)

b. Define **disparity** in stereo vision. (2 marks)

c. State the formula that relates disparity, focal length, inter-ocular separation and depth. Illustrate this relationship with a diagram. (5 marks)

d. What are the main differences between the data obtained from calibrated and uncalibrated stereo systems? Illustrate these differences with practical examples of use. (4 marks)

e. Given a pair of stereo images, what do we mean by the term **image rectification**? Why is it important? (3 marks)

f. Consider two ideal pinhole cameras with the following top view configuration:

![Diagram of two ideal pinhole cameras](image)

Draw the epipole and a few epipolar lines on the front view of the two 2D images. (4 marks)

**END OF EXAMINATION**