Two hours

UNIVERSITY OF MANCHESTER  
SCHOOL OF COMPUTER SCIENCE

Computer Vision

Date:  Monday 20th May 2013  
Time:  14:00 - 16:00

Please answer any THREE Questions from the FOUR Questions provided

Use a SEPARATE answer book for each QUESTION

This is a CLOSED book examination

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

[PTO]
1. Figure 1. Left: A grayscale (0-255) image. Right: The image histogram for this image.

A COMP37212 student is given the intensity image shown at the left of Figure 1, which shows the branching blood vessels in a human retina. They are also provided with the image histogram for this image, as shown at the right of Figure 1.

a) Give a brief explanation of the terms thresholding and window thresholding. Estimate the appropriate numerical value(s) that would be needed in order to segment the branching vessels from the image using one of these methods. Explain your choice of value(s).

[3 marks]

The student now applies three image analysis techniques to the original intensity image, and their results are shown in Figure 2.

Figure 2. From the Left: A close-up view of part of the image from Figure 1, and the results of processing this intensity image in three ways. The same portion of the whole image is shown in each case. Note that the lines in (B) and (C) are about one pixel wide.

Question 1 continues on the following page
Question 1 continued from the previous page

b) Name the two image analysis techniques you think were used to produce the results shown in Figure 2, subfigures (B), and (C). Explain your reasoning in each case. 

[2 marks]

By using the original image, and the results shown in Figure 2 (or otherwise), describe a strategy for:

c) Extracting the positions of all points where the large vessels branch or cross. 

[4 marks]

d) Measuring the widths of the blood vessels at any specified point 

e) Tracing the path of a vessel, and measuring the length of vessel between any two specified points. 

[3 marks]

In each case, you should describe in detail the steps involved, and explain how they achieve the desired result. You should discuss any problems that might arise with your proposed methods, and how you might overcome them.

f) Discuss whether a model-based method could be employed instead to locate the large branching vessels in the original intensity image (and in similar images). You should mention at least two model-based methods in your answer. 

[6 marks]

End of Question 1
2.
A COMP37212 student has decided to study faces and how they vary across a population. A sample from their training set of face images is shown in Figure 3. You may assume that all the faces in the training set are face-on, and have been suitably aligned, and also that a suitably-large training set was available.

![Face Images](image)

**Figure 3:** Three face images from a larger training set, and the face patches extracted from these images.

The student extracts face patches from each of the images, as in the example shown in the Figure.

- **a)** By using a simple one-dimensional example (or otherwise), explain why the student would have to use some form of interpolation when extracting these face patches, and why *linear* interpolation might be preferred over *nearest-neighbour* interpolation.  
  
  [4 marks]

- **b)** Explain in detail how these images could be used to create an Eigenfaces model, suitable for simple recognition, similar to that used by Turk and Pentland. You should also include a brief explanation of how the model could be used in recognition.  
  
  [10 marks]

The student now wishes to consider the problem of recognising faces when the images contain a face with either:

- A non-central location, or of a non-standard size, or
- The person not directly facing the camera, but looking slightly to the side, or
- A different facial expression.

*Question 2 continues on the following page*
Question 2 continued from the previous page

You may assume that the student can be provided with a suitable training set for each task.

c) How would the basic Eigenfaces model compare with either an Active Shape Model (ASM) or an Active Appearance Model (AAM) in terms of performance in these three cases? You need only make a comparison with one of these in each case, and you may choose a different model to compare with for different cases. You may also assume that the ASM or AAM has been suitably initialized. You do not need to describe the detailed construction of the ASM or AAM, but instead you should refer to the main properties of each model that lead to the predicted performance.

[6 marks]

End of Question 2
3.

a) The SIFT descriptor is a popular method for describing selected feature points based on local neighborhood properties so that they can be matched reliably across images. Assuming interest points (keypoints) have been previously detected, briefly describe the main steps of creating the SIFT keypoint descriptor at a given feature point. [8 marks]

b) Explain the role of the following parameters in the SIFT algorithm: [6 marks]
   i. Contrast threshold,
   ii. Curvature threshold and,
   iii. Dimensionality of feature vector.

c) Is the Harris corner detector a linear filter? Argue why or why not. [4 marks]

d) Assume that the Harris corner detector is applied to an unsmoothed image. What type of image would trigger the detector at places that clearly don’t contain a corner? [2 marks]

*End of Question 3*
4.

a) Using a drawing, indicate how the scene coordinates (X, Y, Z) of a point in space are related to the image coordinates (x, y) via the perspective transform in a simple pin-hole camera model, where the camera has focal length f. Show how two such cameras, with a common image plane (and therefore parallel optical axes) and identical focal lengths can be used to estimate the depth (Z - the distance from the image plane) of a scene point by calculating disparities. Make clear what is meant by disparity in this context.

[11 marks]

b) In a more general geometrical arrangement of cameras for stereo reconstruction (non-parallel optical axes, different image planes) what extrinsic camera parameters need to be determined in order to calculate 3D image coordinates from pairs of 2D views? Identify three or more intrinsic parameters that need to be known for each camera in order to make accurate depth measurements. In this general camera geometry, what are epipolar lines? What role do they play in calculating depth? What advantage can be obtained by using three calibrated cameras (trinocular stereo) rather than two (binocular stereo)?

[9 marks]