2011 Paper 8 Question 5

Computer Vision

(*a*) Define the notion of the "semantic gap" in the context of systems for content-based image retrieval. [2 marks]

(*b*) Many systems for optical character recognition make use of *convolutional neural networks.*

(*i*) In what sense are such networks "convolutional", and to what extent do they recognise features independent of position? [3 marks]

(*ii*) Outline how such a network could be used to recognise the characters in a high resolution digital image of this examination question, and highlight which aspects of convolutional neural networks allow for efficient detection and recognition. Assume that the network was trained to recognise only isolated instances of each of the characters. [4 marks]

(*iii*) Consider a convolutional neural network with input image size 29×29 pixels where the first stage is a convolutional layer whose feature maps each have 37 weights. The second stage of the network implements spatial subsampling by a factor of 2 in each dimension, and this is followed by another convolutional layer (third stage) whose feature maps have 26 weights each. How many neurons are there in each of the feature maps of the third stage? [3 marks]

(*iv*) Why is handwriting recognition a more difficult problem than the recognition of printed text? [1 mark]

(*c*) A template-based face detector with a basic detector size of 20×20 pixels is to be applied to an image using a multi-scale sliding window approach. The detector has a hit rate of 99.99% and a false positive rate of 0.1%.

(*i*) Explain whether this detector is likely to yield good recognition performance and give a lower bound estimate of the likely number of false positives if the image has a resolution of 4 megapixels. [2 marks]

(*ii*) Briefly describe the detection mechanism of the Viola–Jones approach to face detection, and highlight **two** aspects of this approach that make it efficient as a sliding-window detector. [5 marks]