THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

# UNIVERSITY OF LONDON

**291 0325 ZA**

## BSc/Diploma Examination for External Students

## Data Compression

Friday 21 May 2010 : 2.30 – 4.45 pm

**Duration: 2 hours and 15 minutes**

---

*Answer ALL questions.*

*Full marks will be awarded for complete answers to ALL the THREE questions.*

*There are 75 marks available on this paper.*

*Electronic calculators may be used. The make and model should be specified on the script. The calculator must not be programmed prior to the examination. Calculators which display graphics, test or algebraic equations are not allowed.*

## Question 1

(a) Consider the following incomplete list of milestones in data compression techniques. For each entry of the milestones, indicate the year in which each technique emerged. [3]

i. Morse code
ii. JPEG
iii. Shorthand
iv. FAX
v. Huffman codes
vi. Dictionary-based compression

(b) What is the main difference between a *lossless* compression and *lossy* compression? What does a lossy compression usually aim to do? Give an example of real life data that is suitable for lossy compression. [5]

(c) Describe and contrast a static compression system and an adaptive compression system with the aid of a diagram and an example. Show the model-coder structure in your discussion. [7]

(d) Consider a data source drawn from an alphabet (A, B, C, D) with probability distribution (0.3, 0.4, 0.2, 0.1). [10]

i. Derive a fixed length binary code and a Huffman code for the source.
ii. Comment on the difference in shape of a *binary tree* representing the fixed length code and the *Huffman tree* for the source.
iii. Discuss which code is closer to the optimal. Show all your work to justify your answer.

## Question 2

(a) John claims that he can find a uniquely decodable binary code for any four symbols (A, B, C, D) with codeword lengths (2, 1, 3, 3) respectively. Explain why we should or should not believe John. Give an example of such a uniquely decodable binary code to support your argument. [5]

(b) Consider a segment of an image and the tasks below. Show all your work to justify your answer.

| | | | | |
|---|---|---|---|---|
| A | B | A | A | A |
| B | C | C | D | A |
| D | D | A | B | A |
| C | D | D | C | A |

i. Derive a canonical minimum-variance Huffman code for the segment step by step. [5]

ii. Construct the smallest Reflected Gray Code for this image segment. [3]

iii. Explain and demonstrate how to separate the coded image segment into bit-planes for further compression. [7]

iv. Suppose that the user at the decompression end would like to have certain control over the decompression process. Demonstrate how Pyramid coding works using the image segment as an example. [5]

## Question 3

(a) Draw a flowchart to outline the adaptive Huffman algorithms for decoding. [5]

(b) Demonstrate how you would encode the string BABCAA using the adaptive Huffman algorithm. Trace the states (or values) of the input, output, alphabet and the tree structure on each step. Start with step 0: the initial state as follows: [5]

(0) input: nil
output: nil
alphabet: (#), where "#" represents "SHIFT (⇑)"
tree:

```
  root
0/
 #
```

(c) Following the approach of the LZW algorithm, decode the tokens (`1, 1, 2, 1, 3, 3, 258, 259, 257, 261, 3`) step by step. Assume that the dictionary initially contains single characters A–Z and occupies cells at 1–256 only. [5]

(d) Explain the predictive rule of JPEG $x = Q + S - T$. Demonstrate, with the aid of a small example, how the rule can be applied in pre-processing. Assume the pixel layout:

| | |
|---|---|
| T | S |
| Q | x? |

[5]

(e) As an example, apply the above rule to the array below and compute the *peak signal to noise ratio.* [5]

$$\begin{matrix} 2 & 3 & 4 & 5 \\ 2 & 3 & 2 & 2 \\ 3 & 2 & 2 & 3 \\ 4 & 4 & 5 & 6 \end{matrix}$$