THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **CO3325 ZA**

BSc Examination

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING AND COMBINED DEGREE SCHEME

Data Compression

Thursday 3 May 2018: 14.30 – 16.45

Time allowed: 2 hours and 15 minutes

There are **THREE** questions on this paper. Candidates should answer all **THREE** questions. All questions carry equal marks and full marks can be obtained for complete answers to **THREE** questions. The marks for each part of a question are indicated at the end of the part in [.] brackets.

There are 75 marks available on this paper.

A handheld calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics, text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

Question 1

(a) Would it be possible to find a shorter prefix code than binary code (1, 01, 001, 010)? Give your reasons. [5]

(b) One way to improve the performance of the static Huffman compression algorithm on a small alphabet is to extend the alphabet before encoding. Demonstrate how a binary alphabet (A, B) with probability of 0.2 for A can be extended and how the compression efficiency will be improved. [7]

(c) Explain what is meant by *P picture* in the context of video compression. Draw a diagram to illustrate the concept of *P pictures* in the frame sequence `IPBBPBB` with time marks $0, 1, \cdots, 6$. [5]

(d) A binary tree (0-1 tree) can be used to represent a code containing a few codewords of variable length. Consider each of the four codes for characters A,B,C,D below and draw the binary trees for each code.

(i) (000, 0011, 110, 111)

(ii) (0, 110, 111, 1)

(iii) (0000, 0001, 1, 001)

(iv) (001, 0000, 0001, 1)

For each tree drawn, comment on whether the code being represented by the binary tree is a prefix code and give your reasons. [8]

Question 2

(a) According to Information Theory, the best that a lossless symbolic compression scheme can do is to encode the output of a source with an average number of bits equal to the entropy of the source.

Give the mathematical formula for the entropy of the source by Shannon and explain all the symbols used.

[5]

(b) Consider two commonly used colour representations RGB and LC, and the transform functions for mapping $RGB \rightarrow LC$: [10]

$$\begin{cases} Y & \approx & 0.3R + 0.6G + 0.1B \\ C_b & = & B - Y \\ C_r & = & R - Y \end{cases}$$

i. Explain what is meant by *transform* in the context of Data Compression.

ii. Given $(R, G, B) = (1, 2, 3)$, what are the corresponding values for: (Y, C_b, C_r)? Show all your working.

iii. Given $(Y, C_b, C_r) = (1, 2, 3)$, what are the corresponding RGB values after the detransform $LC \rightarrow RGB$? Show all your working.

(c) Demonstrate step by step how the basic LZW *encoding* and *decoding* algorithms maintain the same version of a dictionary without ever transmitting it in a separate file, using the small string, `AABBBEB`, as an example. At the end of your answer, conclude what can be seen from your demonstration with a list of the new entries in the dictionary. Assume that initially locations 1–256 of the dictionary are occupied by alphabet A, B, $\cdots$, *etc.*

[10]

Question 3

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

(b) Answer the following questions about the canonical minimum-variant Huffman coding algorithm:

i. Describe the main efficiency problem of the canonical minimum-variant Huffman coding algorithm that needs to maintain a frequency list. [2]
ii. Explain how the efficiency of the algorithm can be improved. [2]
iii. Demonstrate the improved algorithm using the following example: [6]

alphabet	A	B	C	D	E	F
frequency	8	6	5	4	3	3

(c) Outline the Arithmetic decoding algorithm.

A binary sequence of length 5 (symbols) was encoded on the binary alphabet (B,W) using the Arithmetic encoding algorithm. Suppose that the probability $Pr(B) = 0.2$ and the encoded output is 0.12. Demonstrate how the Arithmetic decoding algorithm would derive the original sequence of symbols step by step. [10]

END OF PAPER