THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **291 0325 ZA**

BSc Examination
for External Students

COMPUTING AND INFORMATION SYSTEMS AND CREATIVE COMPUTING

Data Compression

Dateline: Friday 22 May 2009 : 2.30 – 4.45 pm

Duration: 2 hours 15 minutes

Candidates should answer **THREE** questions only. Full marks will be awarded for complete answers to **THREE** questions. There are 75 marks available on this paper.

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

Question 1

(a) 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)?

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

(b) Poderla Kindom Patent Office plans to grant a patent in which a new universal lossless compression algorithm is proposed. The algorithm is said to be able to reduce, by at least 1 byte, the size of any text files in digital form. In the worst case, the algorithm is capable of a lossless compression for at least 10% of all digital files. Explain why you would stand up against the decision. [5]

(c) Explain what is meant by *sampling* in the context of data compression. What is the so-called *sample rate*? Consider a sine wave signal $s(t) = 2sin(1000\pi t)$ What value should be set for the minimum sample rate in order to reconstruct the signal later? [5]

(d) Explain, with an example, the concept of *delay* in the context of audio compression. [5]

Question 2

(a) Consider a binary source $S_1 = (A, B)$, where A and B occur independently, and the extended alphabet $S_2 = (AA, AB, BA, BB)$. [10]

(i) Prove that the entropy of S_1 is half the value of the entropy of S_2.

(ii) Give an example where $S_1 = (A, B) = (0.2, 0.8)$ to demonstrate the entropy relationship between S_1 and S_2.

(b) Outline the LZW encoding algorithm in a flowchart. Demonstrate how an identical dictionary is constructed independently on either end of a compression and its decompression, using a small example `BBAAACA`. Assume that addresses $1, \cdots 62$ of the dictionary are filled initially with A, $\cdots$ Z, a, $\cdots$ z, and 0, $\cdots$ 9, respectively. Summarise your work in the format below: [15]

Encoding end:

step	x	word+x	word	output	new in dictionary
1	$\cdots$				

Decoding end:

step	token	element	word	output	new in dictionary
1	$\cdots$				

Question 3

(a) Explain, with the aid of a small example, what *B picture* means and how it works in the context of video compression. [5]

(b) Consider part of a greyscale image that is represented by the array A of entries (in decimal) below, where a, b have the values $2, 3$ respectively:

```
a b a 5
a b a a
4 a 4 5
```

Represent the entries of the matrix by 3-bit Reflected Gray codes.

Demonstrate how the image can be pre-processed using bitplanes (bi-level images) and therefore how a better compression ratio may be achieved. Show all your work and justify the compression results. [10]

(c) Explain, with an example of data or diagram, the concept of *data rate* in the context of audio compression. What does the data rate measure? [5]

(d) Explain what is used to represent the so-called *colour depth* in a common RGB colour model. What is the value of the colour depth in a representation where 2 bytes are assigned to every pixel? If 253 distinct colours are required for an application, what is the smallest colour depth value required? Give your reasons. [5]

Question 4

(a) Outline in a flowchart the adaptive Huffman algorithm for encoding. [5]

(b) Demonstrate how your adaptive Huffman algorithm works for encoding `BAAHA`. Trace the values of the `input`, `output`, `alphabet` and the `tree structure` on each step. Write down the encoding result and compute the compression factor. Assume that the fixed-length codes for characters $A, \cdots, Z$ are $01000001, \cdots, 01011010$, respectively, and that each new symbol is added to the front (left most position) of the alphabet before a stable sort. [10]

(c) Describe a simple way of checking if a given code is a prefix code. Give an example of a binary code containing 4 codewords. [5]

(d) One important step of the Arithmetic decoding algorithm is to update boundary. Identify an assignment error in the Arithmetic algorithm below and correct the error. [5]

```
1. L <- 0 and d <- 1
2. If x is within [L,L+d*p1)
3.     then output s1, leave L unchanged, and
4.          set d<-d*p1
5.     else if x is within [L+d*p1, L+d)
6.          then output s2, set L<- L*d+p2 and d<-d*p2
7. If the_number_of_decoded_symbols
8.             < the_required_number_of_symbols
9.     then go to step 2.
```

Question 5

(a) Consider the 4×4 matrix below that represents the decimal colour values of a segment of an image. Derive a canonical minimum-variance Huffman code for the matrix. [5]

```
A B C D
A B A A
B A A B
C C D E
```

(b) Explain the predictive rule of JPEG $x = Q + (S - T)/2$. Demonstrate how the rule can be applied in pre-processing to the matrix above, where $A, \cdots F$ have values $2, \cdots 6$, respectively. Assume the pixel layout

T	S
Q	x?

[5]

(c) Compare the compression approaches in the previous parts, i.e. the coding with and without preprocessing. Which approach is closer to optimal coding? Justify your answer. [6]

(d) Compute the *Average of the absolute difference matrix* [4]

(e) Consider the situation without preprocessing in the previous part. Would it be possible to find a prefix code whose lengths are (i) (1, 2, 3, 4, 5) and (ii) (1, 2, 2, 5, 5)? Why or why not? [5]