UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

B.Sc. Examination 2016

Computing

IS53010A Resit Data Compression

Duration: 2 hours and 15 minutes

Date and time:

There are five questions in this paper. You should answer no more than three questions. Full marks will be awarded for complete answers to a total of three questions. Each question carries 25 marks. 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.

Electronic calculators must not be programmed prior to the examination. Calculators which display graphics, text or algebraic equations are not allowed.

THIS PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM

Question 1

(a) i. Explain why Reflected Gray Codes are regarded as a better representation than normal binary codes for coding greyscale images.

ii. Derive the *Reflected Gray Code* for each of the four colour codes in decimal below.

[5]

9	10
10	11

(b) What will be the output if the HDC algorithm is applied to the sequence below? Explain the meaning of each control symbol that you use. [6]

`KKH␣␣␣␣␣␣␣␣␣T␣␣UU5555555␣␣BBBAA`

(c) Explain what an *optimal code* is in the context of data compression. Are Huffman codes optimal? Comment and justify, with the aid of an example, the truth of the following statement: [14]

"Huffman codes for text compression are not optimal in general but optimal for video compression."

Question 2

(a) 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 a small string `AGGAGAGAG` as an example. [15]

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

Question 3

(a) Discuss the absolute limit of lossless compression by showing why more than 99% of files cannot be compressed even by one byte. [5]

(b) Outline the Arithmetic decoding algorithm for a binary source in a flowchart.

A binary sequence of length 4 (symbols) was encoded on the binary alphabet (B, W) using the Arithmetic encoding algorithm. Suppose that the probability $Pr(B)$ is computed based on a previous input sequence `BWWWWBWWWW` and the encoded output is 0.34.

Demonstrate, with the aid of a diagram or a table, how the Arithmetic decoding algorithm derives the original sequence of symbols step by step. [10]

(c) Describe the main idea of *predictive encoding*. Suppose the matrix below represents the pixel values (in decimal) of part of a grayscale image. Using the predictor $x = (Q + S)/2$ in JPEG

T	S
Q	x?

, illustrate step by step how the predictive encoding algorithm may be applied to the matrix: [10]

```
1 1 1 1
5 1 1 1
5 5 5 5
7 9 5 5
```

Question 4

(a) Explain each of the following terms in the context of Data Compression. Provide an example of the entity described by each term. [8]

i. variable-to-variable model

ii. bi-level image

(b) Consider the alphabet of four symbols (A, B, C, D). Discuss the possibility of finding

i. a uniquely decodable binary code in which the codeword for A is of length 2, for B is of length 1, and for C or D is of length 3.

ii. a shorter variable length prefix code than the one described in part (b)i.

Give your reasons and one example to justify your argument. [7]

(c) Compare and contrast, with the aid of the example text "`BAGHABGHGGGAAGH`", the static Huffman encoding algorithm with the Shannon Fano encoding algorithm. You should outline both encoding algorithms first. [10]

Question 5

(a) Keith claims that the binary code (1, 01, 001, 010) is a prefix code since it satisfies the Kraft inequality. Check if the code indeed satisfies the Kraft inequality and explain what is wrong with Keith's claim. [7]

(b) Consider part of a grayscale image with 16 shades of gray that is represented by the array `A` below:

```
0011 0010 1100 0111
0011 0001 1100 0110
0111 1100 1101 1011
```

Demonstrate how the image can be pre-processed by several bitplanes (bi-level images) and therefore may achieve a better compression ratio. Provide a complete solution to the instance including any main compression algorithm after the pre-process. [8]

(c) Demonstrate, with the aid of an example, how to improve the compression efficiency of the static Huffman algorithm on a small alphabet with an imbalanced probability distribution. [10]