UNIVERSITY OF LONDON 291 0325 ZB

External Programme

B. Sc. Examination 2008

COMPUTER INFORMATION SYSTEMS

2910325 [Eastern] Data Compression

Duration: 2 hours and 15 minutes

Date and time: Tuesday 20 May 2008: 2.30 – 4.45 pm

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, text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

THIS EXAMINATION PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM

Question 1

(a) A compression process is often said to be 'negative' if its compression factor value is less than 1. Explain and demonstrate, with an example, a negative-factor case for lossless compression. [5]

(b) Consider a binary source file consisting of characters A and B with probability of 0.2 for B. Discuss the cause of ineffciency of applying to this source the static Huffman compression algorithm. Demonstrate how the problem may be solved. [10]

(c) Given the probability distribution of two sources: $P = (p_0, p_1, \cdots, p_m)$ and $Q = (q_0, q_1, \cdots, q_m)$ with entropy H_p and H_q. Suppose that $q_i = p_i$, where $i = 2, 3, \cdots, m$, and $q_0 = q_1 = \frac{p_0+p_1}{2}$. Discuss whether $H_p < H_q$, and give your justification. [10]

Question 2

(a) Comment on the *truth* of the following statement in describing the absolute limits of lossless compression. [5]

"more than 99% of files cannot be compressed even by one byte."

(b) 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?

, apply the predictive encoding to the matrix below and demonstrate the difference made by the predictive encoding. [10]

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

(c) Following the approach of 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. Demonstrate the content changes of the main variables and the dictionary. [10]

Question 3

(a) Explain what is meant by an *optimal code* 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: [10]

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

(b) 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 three bytes are assigned to every pixel? If 247 distinct colours are required for an application, what is the smallest colour depth value required? Give your reasons. [5]

(c) Consider the task of sending a set of numbers (27, 28, 28, 26, 27, 28, 30, 31, 32, 34) over a mobile communication channel with as few number of bits as possible. Propose a coding scheme to achieve a good compression. Demonstrate all your compression and decompression work and evaluate your approach. Justify your choice of any standard compression methods. [10]

Question 4

(a) Decode the following string using the HDC algorithm. Explain the meaning of each control symbol used. What is the compression ratio? What is the entropy of the source? [10]

```
r4n1Ar2n6BB3322r31n30ABr3Cn2BC
```

(b) Derive the *Reflected Grey Code* for each of the colour codes in decimal below. Explain why Reflected Grey Codes are regarded as a better representation than normal binary codes for coding greyscale images. [5]

9	10
10	11

(c) Outline the Arithmetic decoding algorithm.

A binary sequence of length 4 (symbols) was encoded on the binary alphabet (B,W) using 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]

Question 5

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

(b) Given an alphabet of four symbols (A, B, C, D), discuss the possibility of finding a uniquely decodable binary code in which the codeword for A is of length 2, that for B of length 1 and for both C and D of length 3. Give your reasons and one example to justify your arguments. Demonstrate how a binary tree can be used to explore the unique decodability of a binary code and justify yourself with examples. [10]

(c) Consider part of a greyscale image with 16 shades of grey that is represented by the array A below:

```
0001 0010 1100 0110
0010 1000 1100 0110
0011 1100 1101 1011
```

Demonstrate how the image can be pre-processed by several bitplanes (bi-level images) and therefore may achieve better compression ratio. [5]