LINKÖPING UNIVERSITY

Written Exam in

Image and Audio Compression TSBK38

27th May 2024 8:00 - 12:00

Location:	U2
Examiner:	Harald Nautsch
Teacher:	Harald Nautsch, 1361
Department:	ISY
Module:	TEN1
Number of problems:	13
Number of pages:	6+formula collection
Permitted equipment:	Calculator, “Tables and Formulas for Image Coding and Data Compression”
Grades:	3 : 14+ points from part I 4 : 14+ points from part I, 24-31 total points 5 : 14+ points from part I, 32-40 total points
Other:	Answers can be given in English or Swedish. The teacher visits the exam room around 9.30.

Exam structure

The exam is split into two parts, with maximum 20 points in each. In order to get a passing grade (3) you will need to get at least 14 out of 20 points from part I.

In addition, 24-31 total points gives grade 4 and 32-40 total points gives grade 5.

Part I

1 Describe the difference between *lossy* and *lossless* compression. In what situations would we prefer one type of compression over the other?

(3 p)

2 Many of todays media coding algorithms are based on some form of transform coding. A couple of transforms that can be used are the Karhunen-Loève transform (KLT), the discrete cosine transform (DCT) and the discrete Walsh-Hadamard transform (DWHT).

a) Describe some of the most important properties that a good transform should have.

(2 p)

b) Give some advantages and disadvantages of chosing KLT, DCT or DWHT as your transform.

(2 p)

3 When coding colour still images and video signals, the colour space used is usually YCbCr instead of RGB. Explain how these colour spaces differ and why the YCbCr colour space is preferred.

(2 p)

4 Describe in detail how modern hybrid coders and decoders for video signals work. H.264 and HEVC are examples of such coders.

(4 p)

5 Two psychoacoustic phenomena are *frequency masking* and the *hearing threshold*. Explain what these are and how they can be utilized when coding audio signals.

(2 p)

6 Give a fairly detailed description of how JPEG coding of still images work.

(3 p)

7 A memoryless source has the alphabet $\mathcal{A} = \{a, b, c\}$ with the symbol probabilities $P(a) = 0.5,\ P(b) = 0.4$ and $P(c) = 0.1$.

Construct a Huffman code for pairs of symbols from the source and calculate the average rate (in bits/symbol) of the code.

(2 p)

Part II

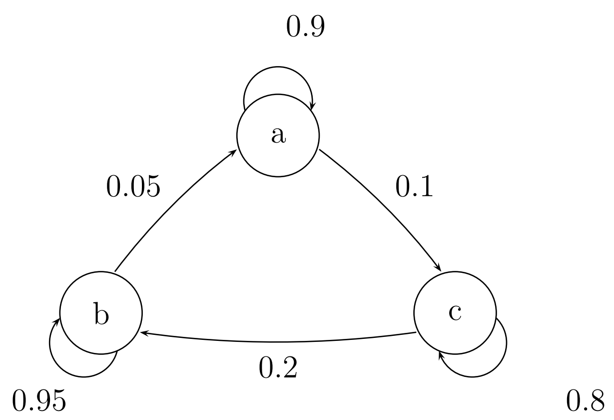
8 When coding speech signals, a relatively simple model of how human speech is generated is often used. Describe this model and how it can be used in the coding and decoding process.

(2 p)

9 Explain how arithmetic coding works.

(2 p)

10 A Markov source of order 1 X_n with alphabet $\{a, b, c\}$ is given by the following state diagram



What is, on average, the lowest possible rate (in bits/symbol) that the output from the source can be coded to?

(3 p)

11 A memoryless time discrete gaussian signal with variance σ^2 and zero mean is quantized using a 12 bit uniform quantizer. The edges of the quantizer are set at $\pm 5\sigma$. The quantized signal is coded using a fixed length code.

What is the resulting signal-to-noise ratio (in dB)?

What would the resulting signal-to-noise ratio be if we used Lloyd-Max quantization instead of uniform quantization?

(3 p)

12 A speech signal is modelled as a one-dimensional stationary gaussian process Y_n. The signal statistics have been estimated as:

$$E\{Y_n\} = 0$$

$$R_{YY}(k) = E\{Y_n \cdot Y_{n+k}\}$$

$$R_{YY}(0) = 9.07 \ , \quad R_{YY}(1) = 7.56 \ , \quad R_{YY}(2) = 4.09 \quad R_{YY}(3) = 0.08$$

We want to code the signal with no more than 5 bits per sample (on average) and a signal to noise ratio that is at least 35 dB.

Construct a predictive coder that fulfills the requirements. All assumptions and simplifications must be motivated.

(5 p)

13 A three channel audio signal consists of a left channel L_i, a right channel R_i and a center channel C_i. The following mean values have been measured:

$$E\{L_i\} = E\{R_i\} = E\{C_i\} = 0$$

$$E\{L_i^2\} = E\{R_i^2\} = E\{C_i^2\} = 1$$

$$E\{L_iR_i\} = 0.9 \ , \quad E\{L_iC_i\} = E\{R_iC_i\} = 0.93$$

The three channels can be modelled as a three-dimensional normal distribution.

The signal is coded by forming vectors $[L_i \;\; C_i \;\; R_i]^T$ that are transformed using a three point DCT. The transform components are Lloyd-Max quantized such that the average rate is 2 bits/sample/channel.

Allocate bits to the transform components such that that average distortion for the three channels is minimized and calculate this distortion.

What average distortion would we get if we instead did Lloyd-Max quantization of the three channels directly to 2 bits each, without using a transform?

(5 p)