LINKÖPING UNIVERSITY

Written Exam in
Image and Audio Compression
TSBK38

27th May 2025 8:00 - 12:00

Location:	TER2, FE245, FE249
Examiner:	Harald Nautsch
Teacher:	Harald Nautsch, 1361
Department:	ISY
Module:	TEN1
Number of problems:	12
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 What are the properties of a discrete alphabet data sequence that makes it possible to compress it losslessly?

(2 p)

2 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)

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

(3 p)

4 a) In hybrid coding of video, motion compensated prediction is used. Explain how this works, both on the coder and the decoder side.

(2 p)

b) In most modern video coders, individual frames can be coded as either I, P or B frames. Explain the differences between the different types of frames and what type of information that needs to be transmitted for each frame.

(2 p)

5 a) When coding audio signals, a modified cosine transform (MDCT) is often used. Describe how the MDCT differs from a regular DCT and explain why the MDCT is a better choice for audio coding.

(1 p)

b) 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 A memoryless source has the alphabet $\mathcal{A} = \{a, b, c\}$ with the symbol probabilities $P(a) = 0.65$, $P(b) = 0.25$ and $P(c) = 0.1$.

Is it possible to code the output from the source losslessly at a rate lower than 1.3 bits/symbol?

If it is possible, construct such a code.

(3 p)

7 Given a gaussian input signal with zero mean and variance σ^2. We want to quantize the signal such that the resulting rate is 8 bits/sample. No source coding is performed, only fixed length coding.

a) What is the resulting signal-to-noise ratio (in dB) if the quantizer is a uniform quantizer, where the working range of the quantizer is chosen as $\pm 4\sigma$?

(2 p)

b) What is the resulting signal-to-noise ratio if we instead use a Lloyd-Max quantizer?

(1 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 A stationary memoryless source has the alphabet $\mathcal{A} = \{a, b, c\}$. The symbol probabilities are $P(a) = 0.6$, $P(b) = 0.3$ and $P(c) = 0.1$.

Code the sequence

$$aaabc$$

using arithmetic coding. Give both the resulting interval and the corresponding codeword.

(4 p)

10 A memoryless time discrete process with probability density function

$$f_X(x) = \begin{cases} 0.5 - 0.25x & ; \quad 0 \le x \le 2 \\ 0.5 + 0.25x & ; \quad -2 \le x \le 0 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

is Lloyd-Max quantized to three levels. Determine the decision borders and reconstruction levels. Also calculate the resulting average distortion.

(4 p)

11 An image is modelled as a stationary twodimensional zero mean normally distributed process $X_{i,j}$ (i and j are coordinates in the image). From a large set of data, the auto correlation function $R_{XX}(k,l) = E\{X_{i,j} \cdot X_{i+k,j+l}\}$ has been estimated as

$$R_{XX}(0,0) = 1104, \quad R_{XX}(0,1) = 1027$$

$$R_{XX}(1,0) = 1001, \quad R_{XX}(1,1) = R_{XX}(1,-1) = 988$$

The image is coded using a linear predictor of the form

$$p_{ij} = a_1 \cdot \hat{X}_{i-1,j} + a_2 \cdot \hat{X}_{i,j-1}$$

The prediction error is quantized uniformly and then source coded using a memoryless arithmetic coder.

How should the predictor coefficients a_1 and a_2 be chosen if we want to minimize the distortion of the coder at a given rate?

What is the lowest rate that can be used if we want to have a signal-to-noise ratio of at least 43 dB?

Compare your result to the lowest rate that can be achived by just using the quantizer and the memoryless arithmetic coder (no predictor) in order to reach 43 dB.

(5 p)

12 In the HEVC video coding standard, one of the transforms that can be used is a type VII Discrete Sine Transform (DST-VII). For a signal $x_i, i = 0, \ldots, N-1$ of length N, the DST-VII coefficents $\theta_j, j = 0, \ldots, N-1$ are defined as

$$\theta_j = \sqrt{\frac{2}{N+1/2}} \cdot \sum_{i=0}^{N-1} x_i \cdot \sin(\frac{\pi}{N+1/2}(i+1)(j+\frac{1}{2}))$$

For $N = 3$, this corresponds to using the transform matrix given by

$$\mathbf{A} = \frac{2}{\sqrt{7}} \begin{pmatrix} \sin\frac{\pi}{7} & \sin\frac{2\pi}{7} & \sin\frac{3\pi}{7} \\ \sin\frac{3\pi}{7} & \sin\frac{6\pi}{7} & \sin\frac{9\pi}{7} \\ \sin\frac{5\pi}{7} & \sin\frac{10\pi}{7} & \sin\frac{15\pi}{7} \end{pmatrix}$$

A signal is modelled as a stationary gaussian process X_n with zero mean and auto correlation function $R_{XX}(k)$

$$R_{XX}(k) = E\{X_n \cdot X_{n+k}\} = 4 \cdot 0.97^{|k|}$$

We want to code our signal using a 3-point DST-VII and Lloyd-Max quantization of the transform components so that the average rate is 2 bits/sample.

Allocate bits to the three transform components such that the average distortion is minimized and calculate the resulting signal-to-noise ratio.

(5 p)