Written Exam in

# Image and Audio Compression
# TSBK38

21st March 2024 14:00 - 18:00

| | |
|---|---|
| **Location:** | SP71 |
| **Examiner:** | Harald Nautsch |
| **Teacher:** | Harald Nautsch, 1361 |
| **Department:** | ISY |
| **Module:** | TEN1 |
| **Number of problems:** | 11 |
| **Number of pages:** | 6+formula collection |
| **Permitted equipment:** | Calculator, "Tables and Formulas for Image Coding and Data Compression" |
| **Grades:** | 3 : 15+ points from part I<br>4 : 15+ points from part I, 25-32 total points<br>5 : 15+ points from part I, 33-40 total points |
| **Other:** | Answers can be given in English or Swedish.<br>The teacher visits the exam room around 15.30. |

# Part I

1 Explain how Lempel-Ziv coding works. Describe both major variants (LZ77 and LZ78).

(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 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)

c) Briefly describe what improvements that have been made to make a modern hybrid coder (eg HEVC) much better than an old hybrid coding standard (eg MPEG2).

(1 p)

4 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)

c) Most music is stored in a stereo format, using two channels (left and right). Describe how the stereo information in an audio signal can be utilized to get a more efficient coding.

(1 p)

5 Two common methods for still image coding are PNG and JPEG. Explain how they work and in what situations you might prefer to use one method over the other.

(4 p)

6 A memoryless source has the alphabet $\mathcal{A} = \{a, b, c, d, e\}$ and symbol probabilities

$$P(a) = 0.4, \quad P(b) = 0.3, \quad P(c) = 0.15, \quad P(d) = 0.1, \quad P(e) = 0.05$$

a) What is theoretically lowest rate (in bits/symbol) we can get if we we want to code the source without distortion?

(1 p)

b) We want to perform lossless coding of the output of the source and keep the average rate as low as possible. Which of the following three codes should we choose?

| Symbol | Code 1 | Code 2 | Code 3 |
|---|---|---|---|
| $a$ | 0 | 0 | 1 |
| $b$ | 10 | 100 | 01 |
| $c$ | 110 | 101 | 00 |
| $d$ | 1110 | 110 | 010 |
| $e$ | 1111 | 111 | 001 |

(2 p)

# Part II

7 Describe how a CELP speech coder and decoder works.

(2 p)

8 A memoryless source has the alphabet $\mathcal{A} = \{a, b, c\}$ with the symbol probabilities $P(a) = 0.8$, $P(b) = 0.15$ and $P(c) = 0.05$.

a) Construct a Huffman code for the source that gives an average rate of at most 1 bit/symbol.

(2 p)

b) Code the sequence $a, c, a, a$ from the source using arithmetic coding. Give both the resulting interval and the binary codeword.

(2 p)

9 A memoryless time discrete process with probability density function

$$f_X(x) = \begin{cases} 1-x & ; \quad 0 \leq x \leq 1 \\ 1+x & ; \quad -1 \leq x \leq 0 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

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

(4 p)

10 An audio signal is modelled as a one-dimensional stationary gaussian process $Y_n$. The signal statistics have been estimated as:

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

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

$$R_{YY}(0) = 19.84\ , \quad R_{YY}(1) = 18.32$$
$$R_{YY}(2) = 15.53\ , \quad R_{YY}(3) = 12.69$$

We want to code the signal using linear predictive coding. The resulting average rate should be at most 5 bits per sample and the signal to noise ratio should be at least 37.5 dB.

Which is the smallest predictor (smallest number of predictor coefficients) that fulfills the demands? Give motivations for all assumptions and simplifications that you use.

(5 p)

11 An $N$ point transform matrix is constructed by taking polynoms of degree 0 to $N-1$ and then orthogonalizing and normalizing them. For a four point transform the matrix looks like

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 3/\sqrt{20} & 1/\sqrt{20} & -1/\sqrt{20} & -3/\sqrt{20} \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/\sqrt{20} & -3/\sqrt{20} & 3/\sqrt{20} & -1/\sqrt{20} \end{pmatrix}$$

Suppose that we want to code a onedimensional signal $X_i$ using this transform. $X_i$ is modelled as a gaussian process with mean 0 and autocorrelation function

$$R_{XX}(k) = E\{X_i \cdot X_{i+k}\} = 0.91^{|k|}$$

We want to quantize the transform components with scalar Lloyd-Max quantizers so that the average rate is 2 bits/sample and the average distorsion is minimized.

Allocate bits to the transform components to meet the demands and calculate the resulting signal to noise ratio in dB.

How much lower would the signal to noise ratio be if we instead used a four point Hadamard transform?

(5 p)