Written Exam in
Image and Audio Compression
TSBK38

5th January 2024 8:00 - 12:00

Location:	SP71
Examiner:	Harald Nautsch
Teacher:	Harald Nautsch, 1361
Department:	ISY
Module:	TEN1
Number of problems:	11
Number of pages:	5+formula collection
Permitted equipment:	Calculator, “Tables and Formulas for Image Coding and Data Compression”
Grades:	3 : 15+ points from part I 4 : 15+ points from part I, 25-32 total points 5 : 15+ points from part I, 33-40 total points
Other:	Answers can be given in English or Swedish.

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 15 out of 20 points from part I.

In addition, 25-32 total points gives grade 4 and 33-40 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 Describe in detail how modern hybrid coders and decoders for video signals work. H.264 and HEVC are examples of such coders.

(4 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 Describe how the coding works in each of these still image coding standards

a) GIF

(2 p)

b) PNG

(2 p)

c) JPEG

(2 p)

6 A memoryless source has the alphabet $\mathcal{A} = \{a, b, c\}$ with the symbol probabilities $P(a) = 0.7$, $P(b) = 0.2$ 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

7 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.

(3 p)

8 A source has the alphabet $\{a,\ b,\ c,\ d,\ e,\ f,\ g,\ h\}$. A long sequence of symbols from the source is coded using LZW. The resulting index sequence starts as

$$6,\ 0,\ 8,\ 10,\ 0,\ 1,\ 4,\ 14,\ 14,\ 3,\ 9,\ 12,\ \ldots$$

The starting dictionary is:

index	sequence	index	sequence
0	a	4	e
1	b	5	f
2	c	6	g
3	d	7	h

Decode the given index sequence as far as possible. Also give the resulting dictionary.

(3 p)

9 A stationary memoryless amplitude continuous and time discrete signal X_n has a triangular probability density funcion $f_X(x)$ given by

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

X_n is quantized uniformly with the stepsize $\Delta = 2^{-k}$, where k is a non-negative integer. The quantized signal $\hat{X}_n$ is source coded using a fixed length code.

Give the distortion (mean square error) D of the coder as a function of the rate R.

(4 p)

10 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

$$R_{XX}(0,0) = 1.70, \quad R_{XX}(0,1) = 1.58$$

$$R_{XX}(1,0) = 1.54, \quad R_{XX}(1,1) = R_{XX}(1,-1) = 1.52$$

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 uniformely and then coded using an 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 40 db?

(5 p)

11 A stationary time-discrete gaussian process X_n with auto correlation function $R_{XX}(k)$ is transform coded with a 4-point Discrete Walsh-Hadamard transform.

$$R_{XX}(k) = 17 \cdot 0.91^{|k|}$$

We want to Lloyd-Max quantize the transform coefficients so that the average rate is 2 bit/sample. Distribute bits to the quantizers for the transform coefficients so that the average distortion is minimized and calculate the resulting SNR.

How big is the SNR gain (in dB) of the transform coder, compared to if we just do Lloyd-Max quantization (without transform) of the original signal?

(5 p)