EQ2330 Image and Video Processing

Re-Exam, April 15th, 2020, 08:00-13:00

This is a 5 hour, remote exam. The following aids are allowed: A scientific calculator, your course notes, course book, as well as an approved mathematical handbook[1].

Note: show that you understand the principles: do not spend much time on tedious numerical computations that provide little reward. Use reasonable approximations where this is appropriate. If the equations are not simple to derive or very important to the class, then they will be provided.

Start by browsing through the exam. The point value of each problem and subproblem is shown. These points are obtained if the solution is correct and clearly motivated. The maximum exam score is 100 points.

Responsible teacher:	Markus Flierl	08-790-7425
During the exam, questions are also answered by:	Linghui Zhou	

Good luck!

[1] Beta (Råde, Westergren); Taschenbuch der Mathematik (Bronstein, Semendjajew); Taschenbuch Mathematischer Formeln (Bartsch); Matematicke vzorce (Bartsch); Collection of Formulas in Signal Processing (Dept, S3, KTH); Mathematische Formeln—Erweiterte Ausgabe E (Sieber).

1. *True or False Statements (10p)*

   Are the statements below true or false? You get 1 point if your answer is correct, 1 point is deducted if your answer is wrong, and 0 points if no answer is given. The total score for this problem will not be smaller than zero. Answer only true, false, or no answer. Do not provide any motivation (only this problem).

   (a) ($\pm$1p) Consider median filtering in the spatial domain. By using two 1-D masks (one for the rows, and one for the columns) we can always achieve the same effect as using a 2-D median mask.

   (b) ($\pm$1p) DCT always achieves optimal energy concentration.

   (c) ($\pm$1p) Continuous histogram equalization is invertible.

   (d) ($\pm$1p) Two different images cannot have exactly the same histograms.

   (e) ($\pm$1p) Lloyd-Max algorithm always converges to a globally optimal quantizer.

   (f) ($\pm$1p) Biorthogonal wavelet transform are critically sampled.

   (g) ($\pm$1p) The distortion curve D(R) of a Gaussian source is always convex.

   (h) ($\pm$1p) Generally speaking, median filtering removes outliers whereas low pass filters smooth out the outlier over space.

   (i) ($\pm$1p) Huffman coding can result in a fixed length code.

   (j) ($\pm$1p) Rate distortion functions $R(D)$ are monotonically increasing for $D > 0$.

2. *Frequency Domain Processing (15p)*

The discrete Fourier transform of an image $f(x, y)$ of size $M \times N$ is given by

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M+vy/N)}. \tag{1}$$

The discrete convolution of two functions $f(x, y)$ and $h(x, y)$ of size $M \times N$ is defined by the expression

$$f(x, y) * h(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n). \tag{2}$$

The correlation of two functions $f(x, y)$ and $h(x, y)$ is defined as

$$f(x, y) \circ h(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f^*(m, n) h(x + m, y + n), \tag{3}$$

where f^* denotes the complex conjugate of f.

(a) (2p) Is the Fourier transform a linear operation? Explain your answer.

(b) (2p) Show that if $f(x, y)$ is separable, its Fourier transform is also separable, i.e., $F(u, v) = F(u)F(v)$, where $F(u)$ and $F(v)$ are one dimensional Fourier transforms.

(c) (3p) If a 2-D filter $g(x, y)$ has a size of $A \times B$, what is the minimum padded image size for the convolution of $f * g$ to avoid aliasing? Motivate.

(d) (1p) Write down the convolution theorem.

(e) (4p) Write down the equivalent expression for $f(x, y) \circ h(x, y)$ in the frequency domain. Prove your result. For simplicity, you can consider the one-dimensional case.

(f) (3p) Correlation methods are widely used for matching problems. If we want to determine whether f contains a particular object or region of interest, we let $h(x, y)$ be that object or region (normally called a template). Please explain how we can find a match using correlation.

3. *Noise and Linear Processing (15p)*

Consider a pixel f of a noisy image. To simplify the problem, we will look only at this pixel while neglecting the other pixels in the noisy image. Assume that the noisy pixel value f is caused by adding Gaussian noise n to the pixel value s of the clean image. Let σ_s^2 be the variance of the clean pixel s and let σ_n^2 be the variance of the additive Gaussian noise n. Further, the noise n is statistically independent from the clean pixel s. For simplicity, the mean of the clean pixel s and the mean of the noise n is zero.

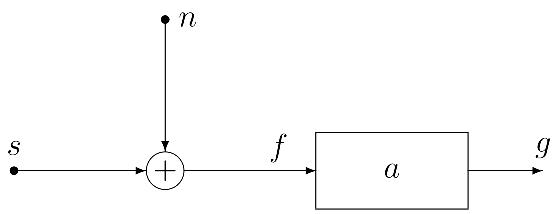

Figure 1: Linear processing of noisy signal.

Now, we like to reduce the impact of the noise in f and perform a linear attenuation by factor a to generate the recovered pixel value g. This process is shown in Fig. 1.

(a) (2p) Calculate the variance of the estimation error $\sigma_e^2 = E\{(g-s)^2\}$ as a function of the attenuation factor a and the given variances.

(b) (4p) The factor a can be chosen such that the variance of the estimation error is minimized. Determine a_o such that the variance of the estimation error is minimized.

(c) (4p) Determine the variance of the estimation error for the optimum a_o. How is it related to the variance of the recovered pixel g for the optimum a_o?

(d) (2p) Determine the correlation between the estimation error $g-s$ and the recovered pixel g for the optimum a_o.

(e) (1p) Now, consider an extended approach where the mean squared estimation error is minimized over the entire image. What name is usually used to refer to such a minimum mean squared error recovery?

(f) (2p) When considering the entire image, looking at the variance of the estimation error $e = g - s$ at one pixel only will not be sufficient in general. What will be useful to capture the second moment statistics of an entire image? In which case is it sufficient to look at the variance of one pixel only?

4. *Uniform Quantizer (15p)*

We use a uniform quantizer for lossy coding of the image x. $\hat{x}$ is the reconstructed image after quantization, as shown in Fig. 2. Let Δ be the quantizer step size.

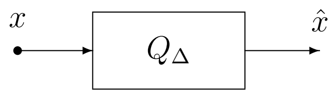

Figure 2: Uniform quantizer Q_Δ with quantizer step size Δ for the image x.

Let $e = \hat{x} - x$ be the quantization error. We assume that the quantization error is small and uniformly distributed in the interval $[-\Delta/2, \Delta/2]$, as depicted in Fig. 3.

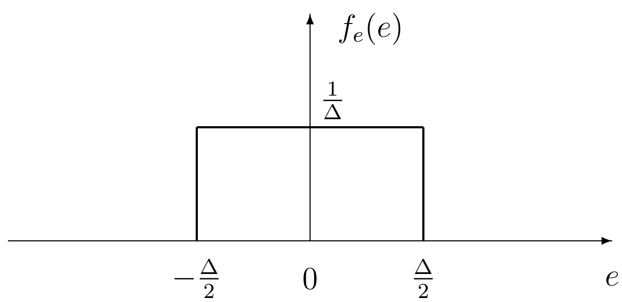

Figure 3: Probability density function $f_e(e)$ for the quantization error e.

We use the squared error to define the coding distortion $D = E\{e^2\}$ (squared error per pixel). In the following, we assume that the small distortion D decreases further with increasing data rate R (bits per pixel) for the entropy-coded quantizer indices

$$D(R) = \sigma^2 2^{-2R}, \tag{4}$$

where σ^2 denotes some signal variance. This allows us to relate the quantizer step size Δ to the slope of the tangent to the distortion rate curve.

(a) (3p) Determine the signal-to-noise ratio $\text{SNR} = 10\log_{10}(\sigma^2/D(R))$ as a function of the data rate R. Draw the graph and explain this result.

(b) (2p) For the uniform distribution of the quantization error in Fig. 3, determine the coding distortion $D = E\{e^2\}$ as a function of the quantizer step size Δ.

(c) (4p) Let μ be the negative slope of the tangent to the distortion rate curve, i.e.,

$$\mu = -\frac{dD}{dR}. \tag{5}$$

With above assumptions, determine μ as a function of the distortion D and, hence, as a function of the quantizer step size Δ. Plot μ over the quantizer step size Δ.

(d) (2p) Now, consider the Lagrangian cost function $J = D + \lambda R$ with positive Lagrangian multiplier λ. For a given λ, the Lagrangian costs J are minimized if distortion D and rate R are in equilibrium. Determine the relation between λ and the negative slope of the tangent μ if distortion and rate are in equilibrium.

(e) (2p) Explain how above relation between λ and the quantizer step size Δ can be helpful for rate distortion optimal coding mode decisions in sophisticated encoders that use uniform quantizers with step size Δ?

(f) (2p) Consider the case where the small distortion D decreases with increasing data rate R (bits per pixel) according to

$$D(R) = \frac{\sigma^2}{1 + R}, \tag{6}$$

where σ^2 denotes some signal variance. What is the relation between λ and the quantizer step size Δ in this case? Why would this relation not be helpful for practical rate distortion optimal coding mode decisions in encoders that use uniform quantizers?

5. *Source Coding (15p)*

(a) *Generic questions on Huffman codes*:

i. (2p) Can $l = (1, 2, 2)$ be the codeword lengths of a binary Huffman code ? What about $l = (2, 2, 3, 3)$?

ii. (1p) In general, what codeword lengths $l = (l_1, l_2, ..)$ can arise from binary Huffman coding ?

iii. (2p) Determine a binary Huffman code for a source with the probabilities $p = \{1/3, 1/3, 1/6, 1/6\}$. Calculate the redundancy of your code.

(b) We consider the following source X

X	$p_X(x)$	$C_1(x)$	$C_2(x)$
1	a	1	010
2	b	010	1
3	$\frac{1}{4}$	00	011
4	$\frac{1}{4}$	011	00

where C_1 and C_2 are two possible binary codes for X.

i. (2p) Give the necessary conditions on a and b so that p_X is a probability mass function.

ii. (1p) Compute, in terms of a, the entropy of X.

iii. (2p) Give the range of a for which C_1 is better than C_2?

iv. (2p) Is there any a such that C_1 is the Huffman code of X ? If yes, give such a.

v. (3p) Given $a = 1/4$, is any of the proposed codes optimal in terms of codeword length ? Please explain. If not, what is the optimal code?

6. *Multiresolution Processing (15p)*

Consider the following lifting implementation for multiresolution image processing as illustrated in Figure 4. Let $x_0, x_1, x_2, x_3, \ldots$ denote the pixel values of a line in an im-

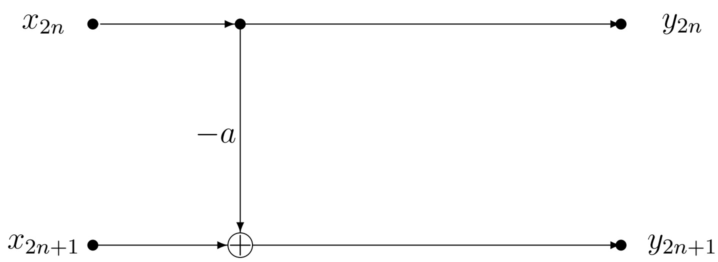

Figure 4: Lifting implementation for multiresolution image processing.

age. Consider a multiresolution processing technique where the analysis filters generate lowpass samples y_{2n} and highpass samples y_{2n+1} according to the transform matrix T

$$\begin{bmatrix} y_{2n} \\ y_{2n+1} \end{bmatrix} = T \begin{bmatrix} x_{2n} \\ x_{2n+1} \end{bmatrix}. \tag{7}$$

(a) (2p) Determine the transform matrix T for the lifting implementation in Figure 4 with parameter $a > 0$.

(b) (1p) Is the resulting wavelet transform orthonormal for any $a > 0$?

(c) (1p) Is the resulting wavelet transform energy preserving for any $a > 0$?

(d) (4p) If the resulting wavelet transform is orthonormal, construct the synthesis filters in the lifting implementation that allow for perfect reconstruction for any $a > 0$. If the resulting wavelet transform is not orthonormal, fix the lifting implementation in Figure 4 such that it gives orthogonal sub-bands and preserves energy for any $a > 0$. For your fixed lifting implementation, determine all necessary weighting and scaling operations as a function of a.

Now, we define $\mathbf{y} = T\mathbf{x}$, where $\mathbf{x} = [x_0 \ x_1]^{\mathrm{T}}$, $\mathbf{y} = [y_0 \ y_1]^{\mathrm{T}}$, and the orthonormal wavelet transform T decorrelates the intensity values of image pixel pairs (x_0, x_1). The autocorrelation matrix of the considered image pixel pairs is

$$R_{\mathbf{xx}} = \mathbb{E}\left[\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}^{\mathrm{T}} \right], \tag{8}$$

where the symbols $\mathbb{E}[\cdot]$ and $[\cdot]^{\mathrm{T}}$ denote the expectation and the transpose operation, respectively. Let

$$R_{\mathbf{yy}} = \mathbb{E}\left[\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix}^{\mathrm{T}} \right] \tag{9}$$

be the autocorrelation matrix of the wavelet coefficients $\mathbf{y}$.

(e) (2p) Determine the relation between $R_{\mathbf{xx}}$ and $R_{\mathbf{yy}}$ for a given orthonormal transform matrix T.

(f) (2p) State a general autocorrelation matrix $R_{\mathbf{yy}}$ for which the wavelet coefficients are decorrelated.

(g) (3p) Determine the autocorrelation matrix $R_{\mathbf{xx}}$ for the image pixel pairs that is decorrelated by above orthonormal wavelet transform T as a function of $a > 0$. For simplicity, assume for the second moment of the wavelet coefficients $\mathbb{E}[y_0^2] = 10$ and $\mathbb{E}[y_1^2] = 1$.

7. *Transforms & Rate Allocation (15p)*

We consider a communication system as shown in Fig. 5. A source outputs 2 Gaussian zero-mean random variables x_1 and x_2 with a statistical dependency described by the covariance matrix Σ, where

$$\Sigma = \frac{1}{2}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}. \tag{10}$$

The variables are transformed by means of a *linear* transform $\mathbf{T}$ and quantized *independently* at high rates with ideal *scalar* quantizers Q_1 and Q_2 yielding reconstructions $\hat{y}_1$ and $\hat{y}_2$ respectively. The reconstructed values $\hat{y}_1$ and $\hat{y}_2$ are transformed to the original domain by means of the inverse transformation $\mathbf{T}^{-1}$ yielding a final reconstruction $\hat{x}_1$ and $\hat{x}_2$.

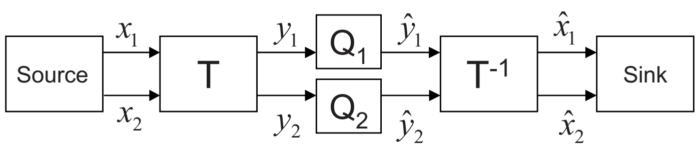

Figure 5: Communication system.

We make the following assumption about the quantizes. We assume that the i-th quantizer ($i = 1, 2$) achieves the following distortion-rate function

$$d_i(R_i) = \sigma_i^2 2^{-2R_i} \quad R_i > 0, \tag{11}$$

where $d_i = E\{(y_i - \hat{y}_i)^2\}$ and σ_i^2 is the variance of the random variable, the i-th quantizer is operating on.

(a) (2p) Assume that the variance of y_1 is σ_1^2 and the variance of y_2 is σ_2^2. Find the optimal bit-allocation between R_1 and R_2 as a function of σ_1^2 and σ_2^2. The optimal bit allocation should minimize the average *transform domain distortion* $d = \frac{1}{2}[d_1(R_1) + d_2(R_2)]$, given the total bit-rate $R_T = R_1 + R_2$.

(b) (2p) Assume that $\mathbf{T}$ is given by the following rotation matrix

$$\mathbf{T} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}, \tag{12}$$

where $\phi = \frac{\pi}{3}$. Assume that the optimal bit-allocation is used. Compute the average distortion in the *signal domain* $D = \frac{1}{2}[D_1 + D_2]$ as a function of R_T, where $D_i = E\{(x_i - \hat{x}_i)^2\}$, $i = 1, 2$.

(c) (3p) Find the KLT of Σ.

(d) (3p) Assume that $\mathbf{T}$ is the KLT transform. Assume that the optimal bit-allocation is used. Compute the average distortion in the signal domain $D = \frac{1}{2}[D_1 + D_2]$ as a function of R_T, where $D_i = E\{(x_i - \hat{x}_i)^2\}$, $i = 1, 2$.

(e) (2p) Assume that the system in (b) and the system in (d) operate at the same bit-rates. Which system achieves the lowest distortion in the signal domain? Motivate.

(f) (3p) It turns out that the rate allocation between R_1 and R_2 found in (a) remains optimal also when the distortion in the signal domain is concerned. Assume that you use some other invertible transformation $\mathbf{T}$, however, $\mathbf{T}^{-1} \neq \mathbf{T}^T$. Is the bit allocation derived in (a) a valid solution that minimizes the average distortion in the signal domain? Explain your answer.