EQ2330 Image and Video Processing

Exam, January 9th, 2018, 14:00-19:00, E34, E53

This is a 5 hour, **closed-book** exam. The following aids are allowed: a calculator with its memory erased and 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. This exam sheet will not be returned with the solutions. Solutions will be available within one week at STEX.

Responsible teacher: Markus Flierl 08-790-7425

During the exam, questions are also answered by: Baptiste Cavarec

Good luck!

[1] The following handbooks are approved: 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). Handwritten notes are *not* allowed. The course book is *not* allowed.

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) Two different images cannot have the same histogram.

   (b) ($\pm$1p) The Karhunen-Loéve transform only requires first order statistics.

   (c) ($\pm$1p) P frames are encoded independently of the other frames.

   (d) ($\pm$1p) The Hough transform of a point, using the line detection model of the template (m, c) such that $y = mx + c$, is a point.

   (e) ($\pm$1p) Histogram equalization always yields uniform *pmf*.

   (f) ($\pm$1p) The iterative Lloyd-Max quantizer design method always converges to the global optimum.

   (g) ($\pm$1p) A McMillan sum of one is a sufficient condition for decodability.

   (h) ($\pm$1p) The Wiener filter minimizes the MSE of the reconstruction error.

   (i) ($\pm$1p) The intensity component contains no information about the color.

   (j) ($\pm$1p) Consider lossy coding of a memoryless Gaussian source using the MSE distortion. An increase in the bit-rate by 1 bit can increase the SNR by at most 6.02 dB.

2. *Histogram (15p)*

   In this problem, we consider transforms that equalize images.

   We represent the pixel values of an image by a continuous Beta random variable $x_{\alpha,\beta} \sim \mathrm{B}(\alpha, \beta)$ of pdf:

$$p_{x,\alpha,\beta}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,(1-x)^{\beta-1}\,x^{\alpha-1};\; x \in [0,1],\; \alpha \geq 0,\; \beta \geq 0. \tag{1}$$

   where Γ is the Gamma function, $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t$, and $\forall n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$.

   In this problem, the continuous histogram transform is denoted T and the discrete histogram transform H.

   (a) Show that $\forall(\alpha,\beta) \in \mathbb{R}_+^2,\ x_{\alpha,\beta} \sim \mathrm{B}(\alpha,\beta): \quad y = T(x_{\alpha,\beta}),\ y \sim \mathrm{B}(1,1)$.(2p)

   (b) Consider $a = 1$, $b = 2$. Roughly draw the histogram of a random variable $x_{a,b} \sim \mathrm{B}(a,b)$. Describe an image having such histogram.(2p)
   What if $a = 2$, $b = 1$?(1p)

   (c) We consider the following two bits quantizer:

$$Q(x) = \begin{cases} \frac{1}{8} \text{ for } 0 \leq x < \frac{1}{4} \\ \frac{3}{8} \text{ for } \frac{1}{4} \leq x < \frac{1}{2} \\ \frac{5}{8} \text{ for } \frac{1}{2} \leq x < \frac{3}{4} \\ \frac{7}{8} \text{ for } \frac{3}{4} \leq x \leq 1 \end{cases} . \tag{2}$$

   What is the *pmf* of the discrete random variable $y_{1,2} = Q(x_{1,2})$? (2p)

   (d) Give the discrete histogram equalization $y = H(y_{1,2})$.(2p)

   (e) We now reverse the order if the operations and define $y' = Q(T(x_{1,2}))$.
   What is the *pmf* of y'? (2p)

   (f) Compute the entropy of y and y'.(2p)

   (g) Comment on the differences between y and y'.(2p)

3. *Frequency Domain Processing (15p)*

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

$$F(u, v) = \mathcal{F}\{f\}(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M + vy/N)}. \tag{3}$$

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

$$f * h\,(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n). \tag{4}$$

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

$$f \circ h\,(x, y) = \mathbb{E}[f^*(m, n) h(x + m, y + n)], \tag{5}$$

where f^* denotes the complex conjugate of f. Eventually we write the power spectral density

$$\Phi_{f,h}(u, v) = \mathcal{F}\{f \circ h\}\,(u, v). \tag{6}$$

We consider the distortion model of Fig.1, where f is the original uncorrupted image, η is a noise signal, q a convolutive kernel representing a distortion, eventually g is the observed signal from which we wish to recover the original image f.

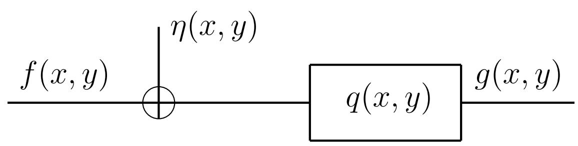

Figure 1: Distortion model

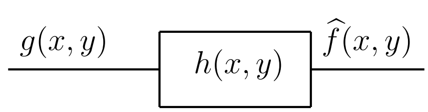

Figure 2: Recovery filter

(a) Recall and demonstrate the convolution theorem.(3p)

(b) Show that $G(u, v) = Q(u, v)F(u, v) + Q(u, v)\eta(u, v)$.(1p)

(c) As mentionned we want to recover the original image f from the observed image g, while minimizing the mean square error $\mathbb{E}\left[\left(f - \widehat{f}\right)^2\right]$.

i. Express $\Phi_{g,g}$ as a function of $Q(u, v)$, $\Phi_{f,f}$ and $\Phi_{\eta,\eta}$.(3p)

ii. We define the error signal $e = f - \widehat{f}$ where $\widehat{f}$ is obtained as in Fig.2 from a recovery filter h. Express the power spectral density $\Phi_{e,e}$ as a function of $H(u,v)$, $Q(u,v)$, $\Phi_{f,f}$ and $\Phi_{\eta,\eta}$.(3p)

iii. We recall that the Wiener filter minimizes the mean square error and is obtained as

$$H(u,v) = \frac{\Phi_{f,g}}{\Phi_{g,g}}. \tag{7}$$

Express H in terms of $Q(u,v)$, $\Phi_{f,f}$ and $\Phi_{\eta,\eta}$.(3p)

iv. What happens if there exists a pair $(\omega_{x_0}, \omega_{y_0})$ such that $Q(\omega_{x_0}, \omega_{y_0}) = 0$? (1p)

v. Propose a better degradation model of an image. (1p)

4. *Multiresolution Processing (15p)*

Consider the following lifting implementation for multiresolution image processing as illustrated in Figure 3.

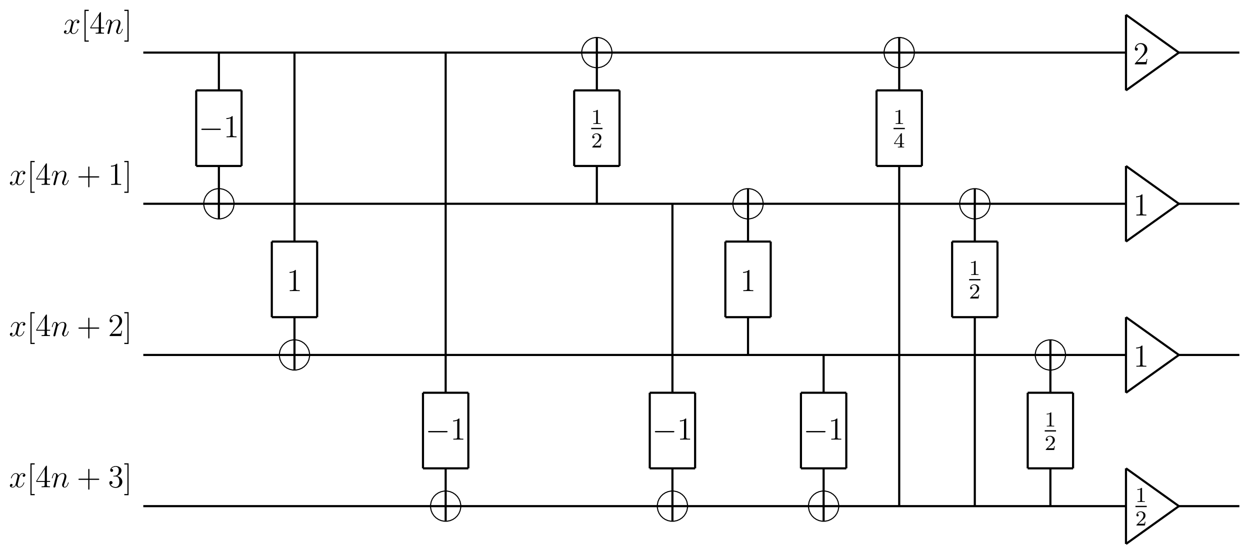

Figure 3: Multiresolution process

(a) Is the transform critically sampled? (2p)

(b) Derive the transform matrix associated to this lifting scheme. (3p)

Hint: Your transform should be of the type $\frac{1}{\sqrt{a^2+b^2+c^2+d^2}} \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}$.

(c) Is the following transform energy preserving?(1p) Is it unitary?(2p)

(d) Build the synthesis operation.(5p)

(e) Would you have perfect reconstruction if you exchange the order of the analysis and synthesis filters.(2p)

5. *Lossless Image Coding (15p)*

We want to compress the following source X:

X	P_i
1	0.25
2	0.5
3	0.125
4	0.125

(a) What is the entropy of the source? (2p)

(b) Determine the Huffman code of the source. (2p)

(c) Is the Kraft-McMillan inequality satisfied? (1p)

(d) Consider the following coding technique where we use the modified CDF:

$$\bar{F}(i) = \sum_{j \leq i} \mathrm{P}_j - \frac{1}{2}\mathrm{P}_i. \tag{8}$$

i. Show that the value of $\bar{F}$ can be used for uniquely encode the source realisation x, i.e that $\forall x, y \quad \bar{F}(x) \neq \bar{F}(y) \Leftrightarrow x \neq y$. (1p)

ii. Assume that we truncate the value of $\bar{F}(x)$ by using $\ell(x)$ bits, we denote the obtained value $\lfloor \bar{F}(x) \rfloor_{\ell(x)}$. Using the fact $\bar{F}(x) - \lfloor \bar{F}(x) \rfloor_{\ell(x)} < \frac{1}{2^{\ell(x)}}$, show that to use $\ell(x) = \lceil -\log_2(\mathrm{P}_x) \rceil + 1$ bits is sufficient to uniquely encode x. (3p)

iii. Show that, for any source $\mathcal{X}$, the average length of such code is $L < H(\mathcal{X}) + 2$. (1p)

iv. What is the average length of the code for the above source? Is the Kraft-McMillan inequality satisfied? (2p)

v. Give the code table for the above source. (3p)
Hint: Recall the binary representation of a number $z \in [0, 1[$

$$z = \sum_{i=1}^{\ell(z)} b_i 2^{-i}; \qquad b_i \in 0, 1. \tag{9}$$

Then we write $z := \{b_1 b_2 \cdots b_{\ell(z)}\}_2$, as an example $0.75 = 2^{-1} + 2^{-2} = \{11\}_2$.

6. *Coding Mode Decision (15p)*

Consider an image source with continuous pixel values X and variance σ^2. We observe that our image source is characterized by two types of pixels. Type one can be efficiently encoded with encoder E_1, whereas type two with encoder E_2. The coder control investigates each pixel whether type one or type two is currently present. The outcome of this investigation will be transmitted to the decoder with a control rate of $R_c = 1$ bit/pixel. The total rate $R = R_c + R_i$ of our coding system is the sum of control rate R_c and the rate R_i of the currently chosen encoder E_i.

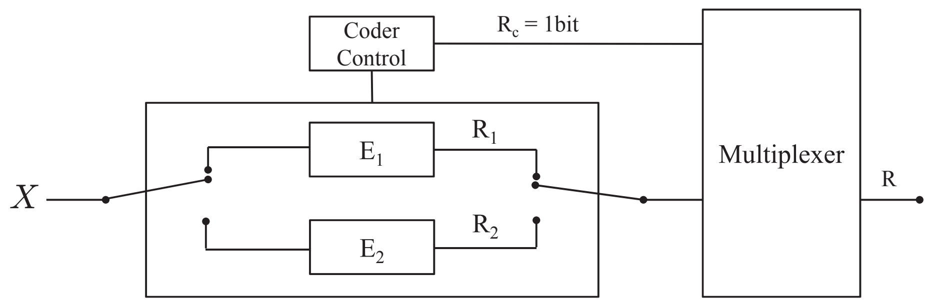

Figure 4: For each pixel of image X, the coder control chooses between encoder E_1 and E_2.

Let the distortion rate function of encoder E_1 be

$$D_1(R_1) = -2\frac{\sigma^2}{e^2}\left(R_1 - \frac{3}{2}\right) \quad \text{for} \quad 0 \leq R_1 \leq 1, \tag{10}$$

and that of encoder E_2 be

$$D_2(R_2) = \sigma^2 \exp(-2R_2) \quad \text{for} \quad R_2 \geq 0. \tag{11}$$

Note, encoder E_1 cannot generate a bitrate higher than one bit per pixel. For higher rates, E_1 cannot be used and its distortion rate function is not defined.

(a) (3p) We are interested in the optimal overall distortion rate performance D(R) of the given coding system. For that, define a rate distortion cost function for each encoder while using the same Lagrange multiplier $\lambda \geq 0$. Explain the decision that the coder control has to make for optimal overall distortion rate performance.

(b) (4p) Determine the optimal overall distortion rate function $D(R)$ of the given coding system for $R \geq 0$. Sketch the optimal function $D(R)$.

(c) (2p) Is the optimal function $D(R)$ convex? If so, explain why.

(d) (3p) Relate the multiplier λ from (a) to a property of the optimal function $D(R)$. Show that this relation is a necessary condition for optimal coder control.

(e) (3p) Consider an alternate system where all pixels are encoded with the same encoder. Its distortion rate function is $D_a(R) = 9\sigma^2 \exp(-2R)$ for $R \geq 0$. Which coding system has the better coding efficiency for rates higher than one bit per pixel? Compare both in the distortion rate plane and explain.

7. *Image Segmentation (15p)*

Let us assume that an image contains only two principal gray level regions. Suppose that through a statistical analysis of the image, we find that the probability density function is given by

$$p(x) = P_{fg} f_{fg}(x) + P_{bg} f_{bg}(x),$$

where x denotes the gray level values, P_{fg} the probability that a pixel belongs to the foreground, $f_{fg}(x)$ the probability density function for pixels belonging to the foreground, P_{bg} the probability that a pixel belongs to the background, and $f_{bg}(x)$ the probability density function for pixels belonging to the background. The conditional probability density functions are shown in Fig.5. Since any given pixel belongs either to the foreground or to the background, we therefore have $P_{fg} + P_{bg} = 1$. In this problem, we consider *supervised thresholding* which classifies all image pixels with gray level greater than the threshold T as foreground.

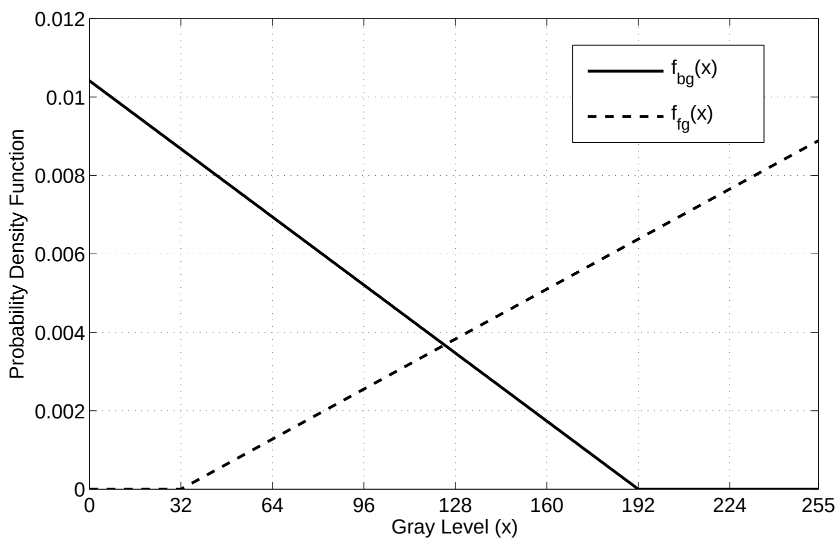

Figure 5: Gray level probability density functions of background and foreground regions in the image.

(a) (3p) Give the mathematical expressions for the background pixel probability density function $f_{bg}(x)$ and the foreground pixel probability density function $f_{fg}(x)$ by using Fig. 5.

(b) (4p) Optimal supervised thresholding with threshold T aims at minimizing the probability of misclassification, which is a sum of two misclassification probabilities. For example, consider $T = 96$ and mark the regions that represent the two misclassification probabilities. Explain these two terms shortly and give the corresponding expressions.

(c) (4p) We want to select the threshold T to minimize the probability of misclassification. What is the optimal segmentation threshold in the two limiting cases $P_{bg} = 0$ and $P_{fg} = 0$, respectively?

(d) (4p) For general probabilities P_{bg}, determine the expression to find the optimal threshold T that minimizes the probability of misclassification by using your expressions of $f_{bg}(x)$ and $f_{fg}(x)$. Give also the result for the special case $P_{bg} = P_{fg} = 1/2$.