EQ2330 Image and Video Processing

Exam, January 9th, 2020, 14:00-19:00, M31

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.

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

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) A spatial filter has a mean value of zero. After applying this spatial filter to an image, the mean value of the output is dependent on the input image.

   (b) ($\pm$1p) The Wiener filter minimizes the mean squared estimation error.

   (c) ($\pm$1p) Image sharpening can be achieved in the frequency domain by a highpass filtering process.

   (d) ($\pm$1p) The FWT is critically sampled.

   (e) ($\pm$1p) Unitary transforms that maximize energy concentration generally increase the dependencies among coefficients.

   (f) ($\pm$1p) A Lloyd-Max quantizer assigns wider quantization cells to regions with high probability of data.

   (g) ($\pm$1p) "Ideal" frequency-domain filters are rarely used because they cause artifacts in the spatial domain.

   (h) ($\pm$1p) It is possible to reduce the average bit-rate per pixel of a given image below the entropy of scalar pixel values of that image.

   (i) ($\pm$1p) We have a filter mask with the dimensions $H \times H$. The dimensions of the image are $M \times M$. We will perform frequency-domain filtering. Then, padding the filter to the size $(M + H) \times (M + H)$ is the minimum size to avoid aliasing.

   (j) ($\pm$1p) The Hough transform of a point in the image space is a point in the Hough parameter space.

2. *Histogram (15p)*

The following picture shows two continuous probability density functions (pdf) $p_f(f)$ and $p_g(g)$. The transformation function for histogram equalization is given as

$$s = T_f(f) = \int_0^f p_f(\alpha)\ d\alpha. \tag{1}$$

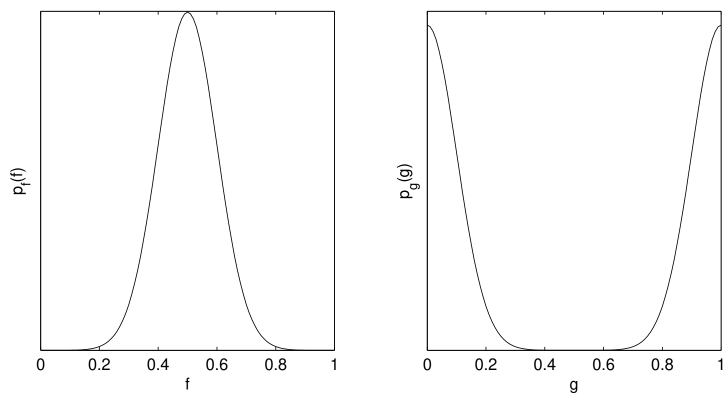


Figure 1: Probability density functions.

(a) (2p) Apply histogram equalization to the two pdfs $p_f(f)$ and $p_g(g)$. Sketch the two equalized histograms.

(b) (4p) Sketch roughly and explain the two transformation functions $T_f(f)$ and $T_g(g)$ for equalizing $p_f(f)$ and $p_g(g)$, respectively.

(c) (2p) Sketch roughly the cumulative distribution function (cdf) for g. What is the relation between the cdf of g and $T_g(g)$?

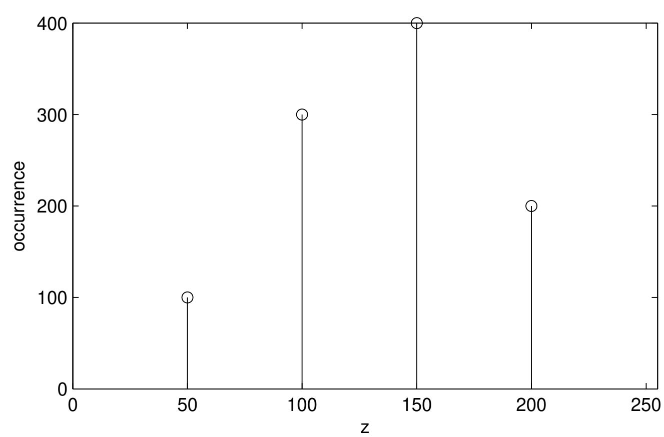


Figure 2: Histogram of the digital image.

(d) (4p) Now we consider a digital image with a histogram as shown in Fig. 2. We apply discrete histogram equalization to this histogram. Compute the equalized histogram. Compare to the continuous case and explain the difference. Note, the intensities of the digital image range between 0 and 255.

(e) (3p) An 8-bit pixel x can be represented as $x = b_0 2^0 + b_1 2^1 + \ldots + b_7 2^7$, where $b_0, \ldots, b_7$ are binary values. By extracting the binary values b_i of all pixels in the image at level i, where $i = 0, \ldots, 7$, we obtain the i-th bit plane of the image. For example, the bit plane zero is generated by extracting b_0 from each pixel. Now, let the bit plane zero be set to zero. What is the effect on the histogram of the modified image? What do we observe in the histogram if the seventh bit plane is set zero?

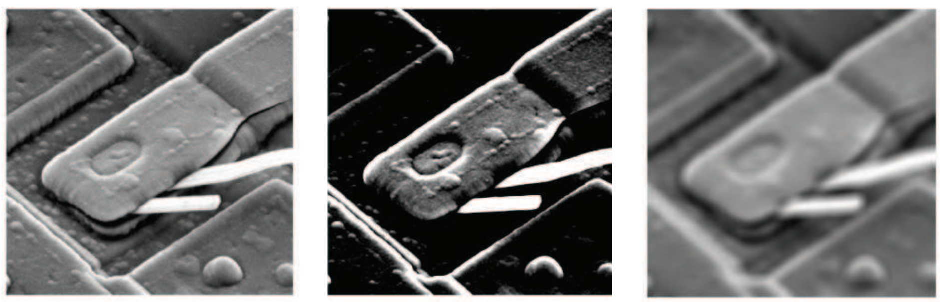
3. *Discrete Fourier Transform (15p)*

This problem considers the Discrete Fourier transform. The discrete one-dimensional Fourier transform of one variable, $f(x)$, $x \in \{0, 1, \ldots, M-1\}$ is given by

$$F(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(x) \exp(-j2\pi(ux/M)). \tag{2}$$

For given $F(u)$, $u \in \{0, 1, \ldots, M-1\}$, the discrete one-dimensional inverse Fourier transform is given by

$$f(x) = \sum_{u=0}^{M-1} F(u) \exp(j2\pi(ux/M)). \tag{3}$$



(a) Original Image. (b) Filtered Image 1. (c) Filtered Image 2.

Figure 3: Original and filtered images.

(a) (3p) The discrete convolution of two functions $f(x)$ and $h(x)$ is denoted by

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

Let $F(u)$ and $H(u)$ represent the discrete Fourier transforms of $f(x)$ and $h(x)$, respectively. Show that $f(x) * h(x)$ and $F(u)H(u)$ form a Fourier transform pair. *Hint: Use the translation property of the Fourier transform.*

(b) (2p) Give an expression for the discrete two-dimensional Fourier transform of an $M \times N$ image $f(x, y)$, where $x \in \{0, 1, \ldots, M-1\}$ and $y \in \{0, 1, ..., N-1\}$.

(c) (3p) Explain the concept of separability for a two-dimensional transform. Show that the two-dimensional Fourier transform is separable.

(d) (3p) Consider a filter in the frequency domain that has the form

$$H(u, v) = Ae^{-(u^2/\sigma_u^2 + v^2/\sigma_v^2)}. \tag{5}$$

Find the equivalent filter in the spatial domain.

(e) (2p) In Fig. 3, two images, Fig. 3(b) and Fig. 3(c), illustrate the effect of some filtering on the original image as shown in Fig. 3(a). Now, which of the two filtered images as shown in Fig. 3 is obtained by applying the resulting spatial domain filter from (d) to the original image 3(a)? Sketch the transfer function of the resulting spatial filter from (d).

(f) (2p) Provide the advantages and disadvantages of the Discrete Fourier Transform (DFT) compared to the Discrete Cosine Transform (DCT) for filtering.

4. *Image Restoration(15p)*

This question considers the phenomenon of motion blur as part of the image acquisition process. During exposure, a clean image $s(x, y)$ with autocorrelation function

$$\phi_{ss}(x, y) = \sigma_s^2 \exp\{-\frac{1}{2}(|x| + |y|)\} \tag{6}$$

undergoes translatory motion in x–direction with constant speed v_x and in y–direction with constant speed v_y. Moreover, the motion-blurred image is further degraded by additive noise $n(x, y)$. Images are captured by the camera during the entire moving period.

(a) (3p) For a 2–D autocorrelation function $\phi(x, y)$, the power spectral density is

$$\Phi(\omega_x, \omega_y) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \phi(x, y) \exp\{-j(\omega_x x + \omega_y y)\}. \tag{7}$$

Derive the power spectral density of the clean image $\Phi_{ss}(\omega_x, \omega_y)$ and sketch its approximate shape. You may find the following identity useful:

$$\sum_{x=-\infty}^{\infty} \exp(-a|x|)(\cos(bx) - j\sin(bx)) = \frac{e^{2a} - 1}{e^{2a} - 2e^a \cos(b) + 1} \tag{8}$$

(b) (2p) Assume that the additive noise $n(x, y)$ is white Gaussian noise with zero mean and variance σ_n^2. Write down the autocorrelation function of this noise. Compute the power spectral density of the noise $\Phi_{nn}(\omega_x, \omega_y)$.

(c) (3p) Assume that shutter opening and closing takes place instantaneously and that the optical imaging process is perfect. For general translatory motion, the transfer function $B(\omega_x, \omega_y)$ of the resulting motion blur is

$$B(\omega_x, \omega_y) = \sum_{t \in \mathcal{T}} \exp\{-j(\omega_x x_0(t) + \omega_y y_0(t))\}, \tag{9}$$

where $x_0(t)$ and $y_0(t)$ are the time-dependent translation components in x– and y–direction, respectively. For simplicity, the exposure happens at discrete time instances $\mathcal{T} = \{-2, -1, 0, 1, 2\}$ and we assume that v_x and v_y are integer-valued. Give a simple expression for the transfer function of the motion blur that is parametrized by v_x and v_y.

(d) (2p) Sketch a block diagram of a filter technique which recovers images that have been degraded by above motion blur and noise.

(e) (3p) We use the Wiener filter for image restoration

$$H(\omega_x, \omega_y) = \frac{\Phi_{sf}(\omega_x, \omega_y)}{\Phi_{ff}(\omega_x, \omega_y)},$$

where f is the image that is degraded by motion blur and additive noise. Express the Wiener filter in terms of the known spectra $\Phi_{ss}(\omega_x, \omega_y)$ and $\Phi_{nn}(\omega_x, \omega_y)$.

(f) (2p) Determine the Wiener filter as a function of the parameters v_x and v_y. Sketch the resulting transfer function of the Wiener filter in its approximate shape for $v_x = 1$ and $v_y = 0$.

5. *Multiresolution Processing (15p)*

Let $x_0, x_1, x_2, x_3, \dots$ denote the pixel values of a line in an image. Consider a multiresolution processing technique where the analysis filters generate lowpass samples l_n and highpass samples h_n according to

$$\begin{bmatrix} l_n \\ h_n \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_{2n} \\ x_{2n+1} \end{bmatrix}. \tag{10}$$

(a) (2p) Is the resulting multiresolution representation critically sampled? Explain.

(b) (2p) Sketch a practical implementation of above analysis filters for multiresolution processing of a line in an image.

(c) (4p) Construct synthesis filters that are biorthogonal to above analysis filters.

(d) (2p) Sketch a practical implementation of your synthesis filters.

(e) (2p) Does your analysis-synthesis pair allow for perfect reconstruction? Explain your answer.

(f) (3p) Does energy conservation hold for above analysis filters?

$$\sum_k x_k^2 = \sum_\mu l_\mu^2 + \sum_\nu h_\nu^2 \tag{11}$$

Explain your answer. Extend your discussion to non-biorthogonal pairs.

6. *Lossless Image Coding (15p)*

Consider two neighboring pixels of an image with intensities x_1 and x_2 which can have four values only: 80, 100, 120, 140. The joint probabilities of the pixel pairs f_{X_1,X_2} having intensity values x_1 and x_2 are given in Tab. 1.

f_{X_1,X_2}	80	100	120	140
80	$\frac{1}{8}$	$\frac{1}{12}$	0	0
100	$\frac{1}{12}$	$\frac{1}{8}$	$\frac{1}{12}$	0
120	0	$\frac{1}{12}$	$\frac{1}{8}$	$\frac{1}{12}$
140	0	0	$\frac{1}{12}$	$\frac{1}{8}$

Table 1: Joint probabilities of two pixel values x_1 and x_2.

(a) (2p) Compute the probability mass function (pmf) $f_{X_1}(x_1)$ for X_1. Determine the minimum average codeword length per pixel with which we are able to encode X_1.

(b) (2p) Compute the joint entropy $H(X_1, X_2)$ and determine the minimum average codeword length per pixel that is needed to encode X_1 and X_2 jointly. Explain your result by comparing to (a).

(c) (4p) Now, assume a coding scheme that uses the conditional entropy. Calculate the conditional pmf $f_{X_2|X_1}(x_2|x_1)$ and the conditional entropy $H(X_2|X_1)$. Explain the coding scheme for the pixel pairs, determine its achievable minimum average codeword length per pixel, and explain your result by comparing to (b).

(d) (7p) Determine the Huffman codes for the four conditional pmfs $f_{X_2|X_1}(X_2|X_1 = x_1)$ as well as for the marginal pmf $f_{X_1}(x_1)$. Determine the average codeword length per pixel for the coding scheme that uses your Huffman codes to encode the pixel pairs. Compare your result to (c)

7. *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.

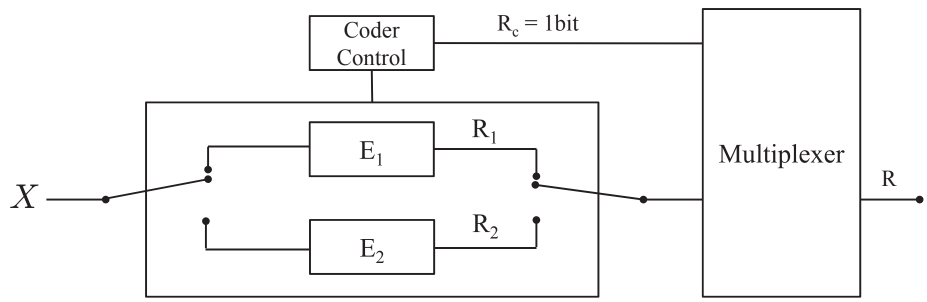


Figure 4: For each pixel of image X, the coder control chooses between encoder E_1 and E_2.

Let the distortion rate function of encoder E_1 be

$$D_1(R_1) = -2\frac{\sigma^2}{e^2}\left(R_1 - \frac{3}{2}\right) \quad \text{for} \quad 0 \leq R_1 \leq 1, \tag{12}$$

and that of encoder E_2 be

$$D_2(R_2) = \sigma^2 \exp(-2R_2) \quad \text{for} \quad R_2 \geq 0. \tag{13}$$

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.