Paper Number(s): **E4.08**
SO14
ISE4.33

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE
UNIVERSITY OF LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2002

MSc and EEE/ISE PART IV: M.Eng. and ACGI

DIGITAL IMAGE PROCESSING

Friday, 26 April 10:00 am

There are FOUR questions on this paper.

Answer THREE questions.

Corrected Copy

Time allowed: 3:00 hours

Examiners responsible:

First Marker(s): Stathaki,T.

Second Marker(s): Clarke,T.J.W.

Special instructions for invigilators: None

Information for candidates: None

1. (a) (i) Explain why the Fourier transform amplitude of an image alone often does not capture most of the intelligibility of the image. [1]

(ii) Explain why it is common to work only with unitary transforms. [1]

(iii) In a specific experiment it is observed that the amplitude of the Fourier transform of an image exhibits low values within a significantly large area around the origin and takes high values within the rest of the two-dimensional frequency plane. State the implications of this observation as far as the original image is concerned. [2]

(b) Let $f(x,y)$ denote an $M \times N$-point two-dimensional (2-D) sequence that is zero outside $0 \le x \le M-1,\ 0 \le y \le N-1$. In implementing the 2-D Discrete Fourier Transform (DFT) of $f(x,y)$, we relate $f(x,y)$ to a new $M \times N$-point sequence $F(u,v)$.

(i) Define the sequence $F(u,v)$ in terms of $f(x,y)$. Show that the 2-D DFT is separable and symmetric. [2]

(ii) Explain how it is possible to calculate the 2-D DFT using only the one-dimensional DFT. [2]

(iii) Comment on the energy compaction property of the Discrete Fourier Transform. [2]

(iv) Consider an image that is corrupted by random noise. Propose a method to reduce the noise using the Discrete Fourier Transform. [2]

(c) Consider the population of random vectors $\underline{f}$ of the form

$$\underline{f} = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix}.$$

Each component $f_i(x,y)$ represents an image. The population arises from their formation across the entire collection of pixels.

Consider now a population of random vectors of the form

$$\underline{g} = \begin{bmatrix} g_1(x,y) \\ g_2(x,y) \end{bmatrix}$$

where the vectors $\underline{g}$ are the Karhunen-Loeve transforms of the vectors $\underline{f}$.

(i) Explain what the elements of the covariance matrix of $\underline{f}$ represent. [2]

(ii) Write down the relationship between $\underline{g}$ and $\underline{f}$. [2]

(iii) Suppose that the covariance matrix of $\underline{f}$ is $\begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}$. It is proposed to reconstruct approximations to the two original images by using in each case only one principal component image associated with the largest eigenvalues. What would be the mean square error incurred in doing so? [2]

(iv) Repeat (iii) in the case where the covariance matrix of $\underline{f}$ is $\begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix}$. [2]

2. (a) (i) Describe the technique of spatially adaptive histogram equalisation. [3]

(ii) An image has the gray level probability density function $p_r(r) = 1,\ 0 \le r \le 1$. It is desired to transform the gray levels of this image so that they will have the specified probability density function $p_z(z) = 3z^2,\ 0 \le z \le 1$. Assume continuous quantities and find the transformation (in terms of r and z) that will accomplish this. [3]

(b) Suppose that an image is corrupted by random noise. One of the properties of the human vision is that the noise is much less visible in the edge regions than in the uniform background regions.

(i) Give examples of spatial masks that are used for noise reduction. Compare the results between smaller and larger masks. [2]

(ii) Propose a method to calculate the local signal-to-noise ratio (SNR) of the image. [2]

(iii) Using the information regarding the local SNR, propose a method that uses variable size spatial filters to reduce background noise without blurring the image significantly. [2]

(c) Give examples of 3×3 Prewitt and Sobel spatial masks. Compare the results arising from their use. Justify your answer. [4]

(d) Give the 3×3 Laplacian spatial mask. Prove that this mask approximates a local second derivative operator. Discuss a possible drawback of the Laplacian mask. Justify your answer. [4]

3. We are given the degraded version g of an image f such that in lexicographic ordering

$$g = Hf + n$$

where H is the degradation matrix which is assumed to be block-circulant, and n is the noise term which is assumed to be zero mean, independent and white.

(a) (i) Consider the Wiener filtering image restoration technique. Write down without proving them the expressions for both the Wiener filter estimator and the restored image both in the spatial domain and the frequency domain and explain all symbols used. [3]

(ii) Discuss the weaknesses of Wiener filtering. [4]

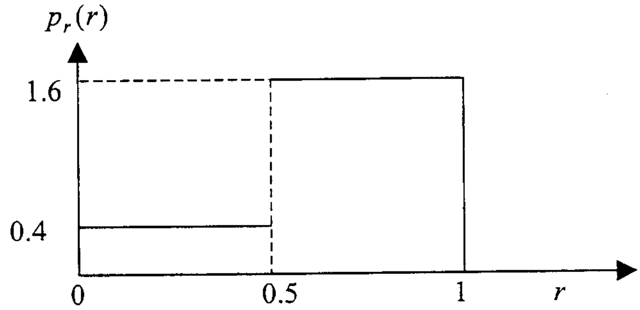
(iii) Discuss a possible formulation of the Wiener filtering in an iterative form. [3]

(b) (i) Describe the image restoration technique of constrained least squares (CLS). Prove the expressions for both the constrained least squares (CLS) filter estimator and the restored image both in the spatial domain and the frequency domain and explain all symbols used. [3]

(ii) Comment on the choice of the regularization parameter in the case of restoration of an image that contains mainly low frequencies and in the case of an image that contains mainly high frequencies. [4]

(iii) Propose a technique to restore an image using a spatially adaptive constrained least squares (CLS) filter. What is the advantage of this technique? [3]

4. (a) Consider an image with intensity $f(x,y)$ that can be modelled as a sample obtained from the probability density function sketched below:



(i) Suppose two reconstruction levels are assigned to quantize the intensity $f(x,y)$. Determine these reconstruction levels using a uniform quantizer. [1]

(ii) Suppose the two intensity levels found in (i) are to be transmitted using extended by two Huffman coding. Find the Huffman codewords. For your codeword assignment, determine the average number of bits required to represent r. [4]

(iii) Determine the entropy, the redundancy and the coding efficiency of the Huffman code for this example. [1]

(b) Consider a Discrete Memoryless Source (DMS) that consists of two symbols s_1, s_2 with probabilities p_1, p_2 respectively with $p_1 + p_2 = 1$. Suppose that differential coding is used to transmit the symbols. Find the new set of symbols with the respective probabilities. [7]

(c) (i) Explain the concepts of category and residual in the lossless JPEG standard for compression of images. [3]

(ii) State how and why we transmit both the category information and the residual information. [4]

E4.08
SO14
ISE 4.33

SOLUTIONS-DIGITAL IMAGE PROCESSING 2002

QUESTION 1

(a)

(i) In viewing a picture, some of the most important visual information is contained in the edges and regions of high contrast. Intuitively, regions of maximum and minimum intensity in a picture are places at which complex exponentials at different frequencies are in phase. On the other hand, regions of flat intensity in a picture are places at which complex exponentials at different frequencies have random phases.
Therefore, it seems plausible to expect the phase and not the amplitude of the Fourier transform of a picture to contain much of the information in the picture, and in particular, the phase should capture the information about the edges.

(ii) Because the inverse of the transformation matrix is obtained very easily.

(iii) We should expect the original image to have many details and very dark backgrounds.

(b)

(i) If $f(x,y)$ is an $M \times N$ array, such as that obtained by sampling a continuous function of two dimensions at dimensions M and N on a rectangular grid, then its two dimensional Discrete Fourier transform (DFT) is the array given by

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

$u = 0,\ldots,M-1$, $v = 0,\ldots,N-1$

It is straightforward to prove that the two dimensional Discrete Fourier Transform is separable and symmetric.

(ii) Use the property of separability. Calculate the DFT row by row and then column by column.

(iii) The low order coefficients are significantly larger due to the low pass nature of a real image.

(iv) DFT/Low pass filtering/Inverse DFT.

(c)

(i) The diagonal elements represent variances of the individual images. The off diagonal elements represent covariances between image.

(ii) The mean vector of the population is defined as

$$\underline{m}_f = E\{\underline{f}\} \Rightarrow \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} E\{f_1\} \\ E\{f_2\} \\ \vdots \\ E\{f_n\} \end{bmatrix}$$

The covariance matrix of the population is defined as

$$\underline{C}_f = E\{(\underline{f} - \underline{m}_f)(\underline{f} - \underline{m}_f)^T\}$$

For M vectors from a random population, where M is large enough, the mean vector and covariance matrix can be approximately calculated by summations

$$\underline{m}_f = \frac{1}{M} \sum_{k=1}^{M} \underline{f}_k \,,\quad \underline{C}_f = \frac{1}{M} \sum_{k=1}^{M} \underline{f}_k \underline{f}_k^{\,T} - \underline{m}_f \underline{m}_f^{\,T}$$

Very easily it can be seen that $\underline{C}_f$ is real and symmetric. In that case a set of n orthonormal eigenvectors always exists.

Let $\underline{A}$ be a matrix whose rows are formed from the eigenvectors of $\underline{C}_f$, ordered so that the first row of $\underline{A}$ is the eigenvector corresponding to the largest eigenvalue, and the last row the eigenvector corresponding to the smallest eigenvalue.

The Karhunen-Loeve transform maps the vectors $\underline{f}$'s into vectors $\underline{g}$'s with the relationship

$$\underline{g} = \underline{A}(\underline{f} - \underline{m}_f)$$

(iii) $\lambda_1 = 7, \lambda_2 = 2$. Mean square error is 2.

(iv) $\lambda_1 = 6,\ \lambda_2 = 3$. Mean square error is 3.

QUESTION 2

(a)

(i) The histogram equalisation is obtained through the expression $s = T(r) = \int_0^r p_r(w)dw$. In spatially adaptive histogram equalisation we define a window around each pixel and calculate the above transformation. We use this mapping for that pixel of interest.

(ii) In the specific example $p_r(r) = 1,\ 0 \le r \le 1$ and $p_z(z) = 3z^2,\ 0 \le z \le 1$.

$$r = G(z) = \int_0^z p_z(w)dw = z^3$$

$$z = G^{-1}(r) = r^{\frac{1}{3}}.$$

(b)

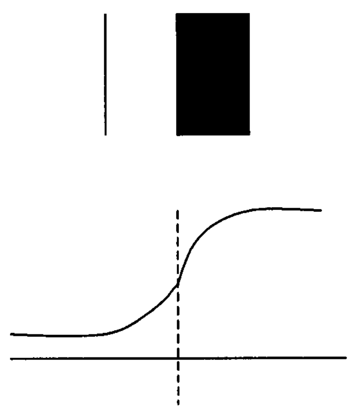
(i) Could be masks of the form $\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. A large mask destroys edges and reduces more the noise a small mask the other way round.

(ii) For each pixel define a neighborhood and calculate SNR.

(iii) Large local SNR would require small mask small local SNR would require large mask.

(c)

An edge is the boundary between two regions with relatively distinct grey level properties (see Figures below). The magnitude of the first derivative can be used to detect the presence of an edge in an image.



An approximation of the local first derivative using a 3×3 mask is the **Prewitt** operator

$$\nabla f \cong \left|(z_7 + z_8 + z_9) - (z_1 + z_2 + z_3)\right| + \left|(z_3 + z_6 + z_9) - (z_1 + z_4 + z_7)\right|$$

and the **Sobel** operator

$$\nabla f \cong \left|(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)\right| + \left|(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)\right|$$

The Sobel operator has the advantage of providing both a differencing a smoothing effect. Lets consider the first Sobel mask that results in a differentiation along the vertical direction. Along the horizontal direction we have the effect of a filtering process through the filter $\{h(-1), h(0), h(1)\} = \{1,2,1\}$. In z − domain we have $z^{-1} + 2 + z$ and in frequency domain $2\cos\theta + 2 = 2(\cos\theta + 1)$. This function has a type of a smoothing effect. The corresponding analysis for the Prewitt mask would give $2\cos\theta + 1$. This is a function that goes to 0 at $\theta = \frac{2\pi}{3}$ and the starts increasing again.

(d)
The areas where the second derivative has zero crossings can be also used to detect the presence of an edge in an image
The **Laplacian** of a 2-D function $f(x,y)$ is a second order derivative defined as

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

In practice it can be also implemented using a 3x3 mask as follows

$$\nabla^2 f = 4z_5 - (z_2 + z_4 + z_6 + z_8).$$

Suppose the first derivative along the horizontal direction is $d_n^{(1)}(m,n) = x(m,n) - x(m,n+1)$ and the second derivative along the horizontal direction is $d_n^{(2)}(m,n) = d_n^{(1)}(m,n) - d_n^{(1)}(m,n-1)$, then we have $d_n^{(2)}(m,n) = 2x(m,n) - x(m,n-1) - x(m,n+1)$. If we follow the same procedure for the vertical direction we obtain the Laplacian mask.
The main disadvantage of the Laplacian operator is that it produces double edges.
If we represent an edge as the change from intensity r_1 to intensity r_2 as follows

$$\begin{matrix} r_1 & r_1 & r_2 & r_2 \\ r_1 & r_1 & r_2 & r_2 \\ r_1 & r_1 & r_2 & r_2 \\ r_1 & r_1 & r_2 & r_2 \end{matrix}$$

Then the response of the Laplacian mask will be

$$\begin{matrix} 0 & r_1 - r_2 & r_2 - r_1 & 0 \\ 0 & r_1 - r_2 & r_2 - r_1 & 0 \\ 0 & r_1 - r_2 & r_2 - r_1 & 0 \\ 0 & r_1 - r_2 & r_2 - r_1 & 0 \end{matrix}$$

QUESTION 3

(a)

(i)

The Wiener filter is

$$\mathbf{W} = \mathbf{R}_{\mathbf{fy}}\mathbf{R}_{\mathbf{yy}}^{-1} = \mathbf{R}_{\mathbf{ff}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{R}_{\mathbf{ff}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{\mathbf{nn}})^{-1}$$

and the estimate for the original image is

$$\hat{\mathbf{f}} = \mathbf{R}_{\mathbf{ff}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{R}_{\mathbf{ff}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{\mathbf{nn}})^{-1}\mathbf{y}$$

Note that knowledge of $\mathbf{R}_{\mathbf{ff}}$ and $\mathbf{R}_{\mathbf{nn}}$ is assumed.

In frequency domain

$$W(u,v) = \frac{S_{ff}(u,v)H^{*}(u,v)}{S_{ff}(u,v)|H(u,v)|^{2} + S_{nn}(u,v)}$$

$$\hat{F}(u,v) = \frac{S_{ff}(u,v)H^{*}(u,v)}{S_{ff}(u,v)|H(u,v)|^{2} + S_{nn}(u,v)}Y(u,v)$$

(ii)

Computational issues

The noise variance has to be known, otherwise it is estimated from a flat region of the observed image.

In practical cases where a single copy of the degraded image is available, it is quite common to use $S_{yy}(u,v)$ as an estimate of $S_{ff}(u,v)$. **This is very often a poor estimate !**

(iii)

Step 0: Initial estimate of $\mathbf{R}_{\mathbf{ff}}$

$$\mathbf{R}_{\mathbf{ff}}(0) = \mathbf{R}_{\mathbf{yy}} = E\{\mathbf{y}\mathbf{y}^{\mathrm{T}}\}$$

Step 1: Construct the i^{th} restoration filter

$$\mathbf{W}(i+1) = \mathbf{R}_{\mathbf{ff}}(i)\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{R}_{\mathbf{ff}}(i)\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{\mathbf{nn}})^{-1}$$

Step 2: Obtain the $(i+1)^{\text{th}}$ estimate of the restored image

$$\hat{\mathbf{f}}(i+1) = \mathbf{W}(i+1)\mathbf{y}$$

Step 3: Use $\hat{\mathbf{f}}(i+1)$ to compute an improved estimate of $\mathbf{R}_{\mathbf{ff}}$ given by

$$\mathbf{R}_{\mathbf{ff}}(i+1) = E\{\hat{\mathbf{f}}(i+1)\hat{\mathbf{f}}^{\mathrm{T}}(i+1)\}$$

Step 4: Increase i and repeat steps 1,2,3,4.

The block-circulant assumption enables us to work with DFT's.

(b)

(i)

It refers to a very large number of restoration algorithms.

The problem can be formulated as follows.

minimize

$$J(\mathbf{f}) = \|\mathbf{n}(\mathbf{f})\|^{2} = \|\mathbf{y} - \mathbf{H}\mathbf{f}\|^{2}$$

subject to

$$\|\mathbf{C}\mathbf{f}\|^{2} < \varepsilon$$

where

$\mathbf{Cf}$ is a high pass filtered version of the image.

The idea behind the above constraint is that the highpass version of the image contains a considerably large amount of noise!

Algorithms of the above type can be handled using optimization techniques.

Constrained least squares (CLS) restoration can be formulated by choosing an $\mathbf{f}$ to minimize the Lagrangian

$$\min\left(\|\mathbf{y}-\mathbf{Hf}\|^2+\alpha\|\mathbf{Cf}\|^2\right)$$

α represents either a Lagrange multiplier or a fixed parameter known as regularisation parameter.

α controls the relative contribution between the term $\|\mathbf{y}-\mathbf{Hf}\|^2$ and the term $\|\mathbf{Cf}\|^2$.

The minimization of the above leads to the following estimate for the original image

$$\mathbf{f}=\left(\mathbf{H}^{\mathbf{T}}\mathbf{H}+\alpha\mathbf{C}^{\mathbf{T}}\mathbf{C}\right)^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{y}$$

(ii)

For an image with low frequencies α should be large. For an image with high frequencies the other way round.

(iii)

The functional to be minimized takes the form

$$M(\mathbf{f},\alpha)=\|\mathbf{y}-\mathbf{Hf}\|^2{}_{\mathbf{W}_1}+\alpha\|\mathbf{Cf}\|^2{}_{\mathbf{W}_2}$$

where

$\|\mathbf{y}-\mathbf{Hf}\|^2{}_{\mathbf{W}_1}=(\mathbf{y}-\mathbf{Hf})^{\mathbf{T}}\mathbf{W}_1(\mathbf{y}-\mathbf{Hf})$

$\|\mathbf{Cf}\|^2{}_{\mathbf{W}_2}=(\mathbf{Cf})^{\mathbf{T}}\mathbf{W}_2(\mathbf{Cf})$

$\mathbf{W}_1,\mathbf{W}_2$ are diagonal matrices, the choice of which can be justified in various ways. The entries in both matrices are non-negative values and less than or equal to unity.

In that case

$\Phi(\mathbf{f})=\nabla_{\mathbf{f}}M(\mathbf{f},\alpha)=(\mathbf{H}^{\mathbf{T}}\mathbf{W}_1\mathbf{H}+\alpha\mathbf{C}^{\mathbf{T}}\mathbf{W}_2\mathbf{C})\mathbf{f}-\mathbf{H}^{\mathbf{T}}\mathbf{W}_1\mathbf{y}$

A more specific case is

$$M(\mathbf{f},\alpha)=\|\mathbf{y}-\mathbf{Hf}\|^2+\alpha\|\mathbf{Cf}\|^2{}_{\mathbf{W}}$$

where the weighting matrix is incorporated only in the regularization term. This method is known as **weighted regularised image restoration**. The entries in matrix $\mathbf{W}$ will be chosen so that the high-pass filter is only effective in the areas of low activity and a very little smoothing takes place in the edge areas.

QUESTION 4

(a)

(i)

Reconstruction levels are at $r_0 = (0+\frac{1}{2})\frac{1}{2} = \frac{1}{4}$, $r_1 = (\frac{1}{2}+\frac{2}{2})\frac{1}{2} = \frac{3}{4}$

(ii)

We have the symbols $r_0r_0, r_0r_1, r_1r_0, r_1r_1$ with probabilities $\frac{1}{16}, \frac{3}{16}, \frac{3}{16}, \frac{9}{16}$

The Huffman code is found below. Probabilities for each r_i are found by evaluating the integral of the PDF over the relevant decision region. The result is shown below.

Symbol	***Probability***
$s_0 = r_0r_0$	1/16
$s_1 = r_0r_1$	3/16
$s_2 = r_1r_0$	3/16
$s_3 = r_1r_1$	9/16

Step 1		**Step 2**		**Step 3**	
s_3	9/16	s_3	9/16	s_3	9/16
s_2	3/16	$\{s_1, s_0\}$	4/16	$\{s_2, \{s_1, s_0\}\}$	7/16
s_1	3/16	s_2	3/16		
s_0	1/16				

Symbol	***Codeword***
s_0	001
s_1	000
s_2	01
s_3	1

(iii)

Average number of bits to represent f

$$l_{avg} = 3\cdot\frac{1}{16} + 3\cdot\frac{3}{16} + 2\cdot\frac{3}{16} + 1\cdot\frac{9}{16} = \frac{27}{16} \text{ bits/word}$$

For the above example we have:

Entropy $H(s) = -\sum_{i=1}^{5} p_i \log_2(p_i)$ bits/symbol

Average length of code $l_{avg} = \sum_i l_i p_i = \frac{27}{16}$ bits/symbol

Redundancy $l_{avg} - H(s)$ bits/symbol

Coding efficiency $H(s)/l_{avg}$

(b)

Symbol	*Probability*
$r_0 = s_0 - s_0$	p_0^2
$r_1 = s_0 - s_1$	$p_0 p_1$
$r_2 = s_1 - s_0$	$p_0 p_1$
$r_3 = s_1 - s_1$	p_1^2

(c)
(i)
The lossless compression method within JPEG is fully independent from transform-based coding. It uses differential coding to form prediction residuals that are then coded with either a Huffman coder or an arithmetic coder. The prediction residuals usually have a lower entropy; thus, they are more amenable to compression than the original image pixels.
In lossless JPEG, one forms a prediction residual using previously encoded pixels in the current line and/or the previous line. The prediction residual for pixel x in Figure is defined as $r = y - x$ where y can be any of the following functions:

$y = 0$
$y = a$
$y = b$
$y = c$
$y = a + b - c$
$y = a + (b - c)/2$
$y = b + (a - c)/2$
$y = (a + b)/2$

Note that, pixel values at pixel positions a, b, and c, are available to both the encoder and the decoder prior to processing x. The particular choice for the y function is defined in the scan header of the compressed stream so that both the encoder and the decoder use identical functions. Divisions by two are computed by performing a one-bit right shift.

	c	b	
	a	x	

Figure: Lossless JPEG prediction kernel

The prediction residual is computed modulo 2. This residual is not directly Huffman coded. Instead, it is expressed as a pair of symbols: the category and the magnitude. The first symbol represents the number of bits needed to encode the magnitude. This number is called category.

(ii)
Only category is Huffman coded. This is because small categories occur very frequently and large categories very infrequently. Residual cannot be Huffman coded because the number of different residuals is very high. The compressed representation for the prediction residual consists of the Huffman codeword of the category followed by the actual binary representation for the magnitude. In general, if the value of the residual is positive, then the code for the magnitude is its direct binary representation. If the residual is negative, then the code for the magnitude is the one's complement of its absolute value. Therefore, codewords for negative residual always start wish a zero bit.