Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

EE4C08

FACULTY OF ENGINEERING, MATHEMATICS & SCIENCE

SCHOOL OF ENGINEERING

Electronic & Electrical Engineering

Engineering

Senior Sophister

Annual Examinations

Digital Media Processing (EE4C08)

Sample Exam Solutions

Dr. F. Pitié

Instructions to candidates:

Answer FOUR (4) out of the FIVE (5) questions.

Please answer questions from each section in separate answer books.

Materials Permitted for this Examination:

New Formulae & Statistic Tables

Graph Paper

Non-programmable calculators

Question 1

1. The Haar Transform is a simple 2-point transform specified by the following transformation matrix:

$$T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Explain how the 1D Haar Transform is extended to 2D image matrices.

[5 marks]

In 2-dimensions the pixels used in the transformation are encoded in 2×2 matrices. Say the input 2×2 block is denoted as matrix X. We may transform first the columns of X, by pre-multiplying by $T^\top$, and then the rows of the result by post-multiplying by T. The output coefficients are encoded in a 2×2 image as follows: $Y = T^\top X T$. With Haar we have $T^\top = T$.

2. Given an input 2x2 image with intensity values as follows:

$$\begin{bmatrix} 10 & 19 \\ 18 & 20 \end{bmatrix}$$

compute the Haar Transform coefficients for the 2×2 image and identify each of the subbands.

[5 marks]

LoLo: $(10 + 18 + 19 + 20)/2 = 33.5$

LoHi: $(10 - 18 + 19 - 20)/2 = -4.5$

HiLo: $(10 + 18 - 19 - 20)/2 = -5.5$

LoLo: $(10 - 18 - 19 + 20)/2 = -3.5$

3. The quantisation step matrix given to the encoder is as follows:

$$Q = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}$$

compute the quantised coefficients and compute the corresponding inverse image.

[5 marks]

Compute the MSE between the original and quantised images.

Quantised LoLo: round$(33.5/2) * 2 = 34$

Quantised LoHi: round$(-4.5/4) * 4 = -4$

Quantised HiLo: round$(-5.5/4) * 4 = -4$

Quantised HiHi: round$(-3.5/8) * 8 = 0$

The inverse transform:

$Y_1 1 = (34 - 4 - 4 + 0)/2 = 13$

$Y_1 2 = (34 - 4 - (-4 + 0))/2 = 17$

$Y_2 1 = (34 - (-4) - 4 - (+0))/2 = 17$

$Y_2 2 = (34 - (-4) - (-4) + 0)/2 = 21$

The MSE between two images I and G is given by

$$MSE = \frac{1}{N} \sum_{\mathsf{x}} (I(\mathsf{x}) - G(\mathsf{x}))^2$$

Thus here we have $MSE = ((13-10)^2+(18-17)^2+(19-17)^2+(20-21)^2)/4 = 15/4 = 3.75$

4. Write down the code for the Matlab function that computes the entropy in bits/pixel of an 2d image with intensity range of 0:255.

[5 marks]

Say we have a grayscale image `Y` in the `0:255` range.

We first need to compute the histogram:

```
h = hist(Y(:), 0:255) ;
```

and normalise the histogram:

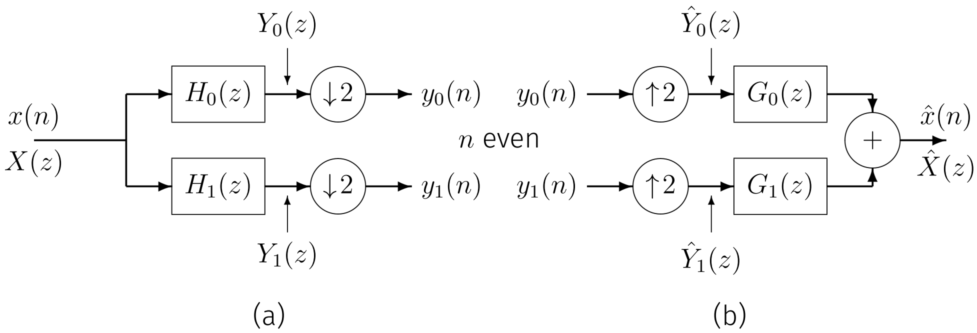
```
h = h/sum(h);
```

Then the entropy is given by

```
H = -sum(h(h>0).*log2(h(h>0)))
```

Question 2

Consider the two-band filter banks as follows:



We choose the LeGall filters 3/5 for the analysis and reconstruction filters:

$$H_0(z) = \frac{1}{8}\left(-1 + 2z^{-1} + 6z^{-2} + 2z^{-3} - z^{-4}\right)$$

$$H_1(z) = \frac{1}{2}\left(1 - 2z^{-1} + z^{-2}\right)$$

$$G_0(z) = H_1(-z) = \frac{1}{2}\left(1 + 2z^{-1} + z^{-2}\right)$$

$$G_1(z) = -H_0(-z) = -\frac{1}{8}\left(-1 - 2z^{-1} + 6z^{-2} - 2z^{-3} - z^{-4}\right)$$

1. LeGall filters are used in JPEG2000 image compression standard. Are they used for lossy or lossless compression?

   [3 marks]

   The LeGall 3/5 are used for lossless compression as they allow perfect reconstruction with integer operations. The CDF 9/7 offer better compression performance but use filters with floating point coefficients that cannot be easily inverted with any machine precision. The CDF 9/7 are thus better suited to lossy compression.

2. What is Perfect Reconstruction and what are the conditions for H_0, H_1, G_0, G_1? Show that the LeGall filters meet the anti-aliasing condition.

   [12 marks]

Perfect Reconstruction is achieved when the reconstructed signal is identical to the original input signal. That is given by:

$$\hat{X}(z) \equiv X(z)$$

This leads to the following equations:

$$H_0(z)G_0(z) + H_1(z)G_1(z) = 2$$

$$H_0(z)G_0(-z) + H_1(z)G_1(-z) = 0$$

The first equation is known as the no-distortion condition and the second condition is known as the anti-aliasing condition.

$$\begin{aligned} H_0(z)G_0(-z) + H_1(z)G_1(-z) &= H_0(z)H_1(-(-z)) + H_1(z)(-H_0(-(-z))) \\ &= H_0(z)H_1(z) - H_1(z)H_0(z) \\ &= 0 \end{aligned}$$

3. The filterbank is fully decimated and not shift invariant. Explain what decimated and shift invariance mean. Explain the importance of shift invariance in image processing. [10 marks]

Decimated means that downsampling at each sub band occurs, so that we have As many coefficients as input samples.
Shift invariance is the equivalent of time invariance and it means that if $y[n]$ is the response of the system to $x[n]$, then $y[n-k]$ is the response of the system to $x[n-k]$. Shift invariance is key to analysis. We expect the shape of the output to be preserved by image translation. If the system is not shift invariant, the analysis will be harder to perform as we can expect a small shift to greatly perturbs the coefficients at discontinuities in the input signal.

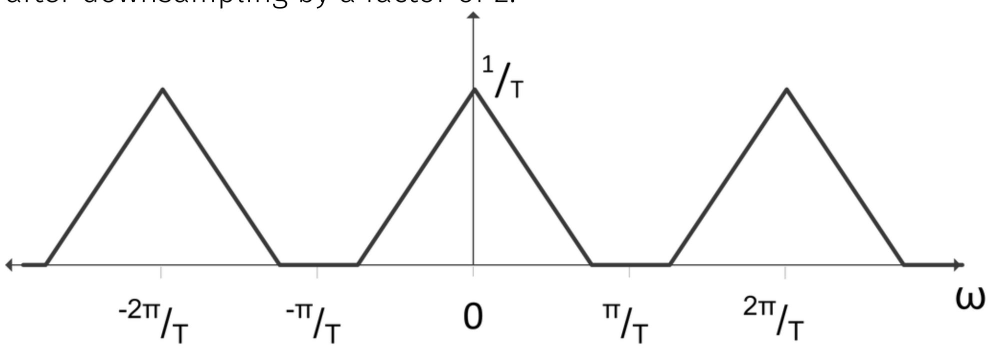
Question 3

1. What is the necessary condition for the aliasing-free sampling of a 2D analogue signal?

[4 marks]

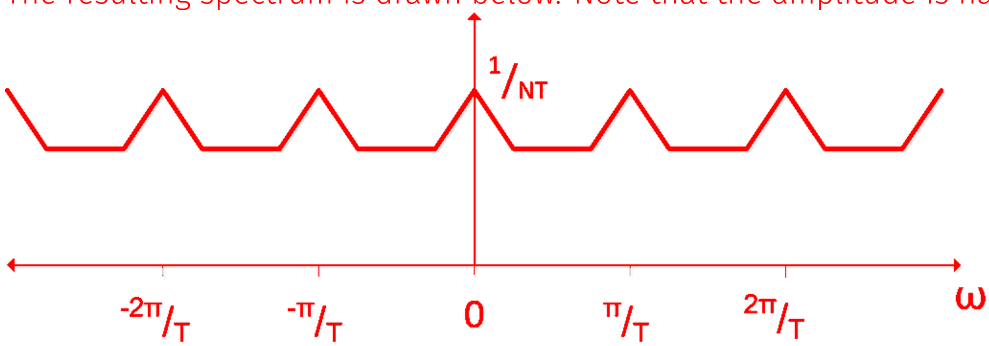
This is the sampling theorem. To avoid an overlap in the spectrum, we need to make sure that the sampling frequency is at least twice of that of the highest analogue signal frequency in both x and y directions.

2. Below is the spectrum of a 1D digital signal, sampled at frequency $2\pi/T$: Draw the signal after downsampling by a factor of 2.



[5 marks]

The resulting spectrum is drawn below. Note that the amplitude is halved.



3. The *sinc* filter can be used as an anti-aliasing filter before downsampling. What is effect of the *sinc* filter in the frequency domain? Explain why it is necessary to filter before resampling.

[10 marks]

The sinc filter corresponds to an ideal low-pass filter. In 2D it has the following form:

$$H(\omega_x, \omega_y) \propto \begin{cases} 1 & \text{if } \omega_x < \Omega_x, \omega_y < \Omega_y \\ 0 & otherwise \end{cases}$$

It is necessary to apply the sinc filter before resampling as we need to cut-off frequencies that will be aliased by the resampling step. Using the example of previous subquestion, a downsampling by 2 effectively halves the available frequency bandwidth and frequencies that are higher than the new Nyquist frequency must be cut-off.

4. What are the practical considerations to consider when implementing an FIR filter approximation of the sinc filter?

[6 marks]

The sinc filter is an IIR filter, thus its impulse response must be truncated if we want to use a FIR filter approximation. [2 marks]
The truncation of the sinc filter can cause ripple artefacts in the output image. [2 marks]
A good way to avoid these ringing artefacts is to apply a windowing mask to smooth out the discontinuity caused by the truncation. [2 marks]

Question 4

1. Explain in which ways the JPEG standard exploits the characteristics of the human visual system to colour and spatial contrast to improve the perceived quality of the compressed images.

[9 marks]

JPEG is based on the YCbCr colour space.
Also JPEG allows to subsample the chrominance space (eg. 4:2:2 , 4:20 schemes) as we are less sensitive to chrominance resolution.
JPEG uses perceptually optimized quantisation matrices for both the luminance and chrominance channels. The chrominance quantisation is stronger for the chrominance channels. Also, the quantisation is roughly increasing with spatial frequency (expect for a small dip at mid frequencies for the luminance channel). There is an asymmetry in the quantisation steps between the vertical and horizontal frequencies.

2. What is the computational complexity of the DCT

[2 marks]

$$\mathcal{O}(n \log(n))$$

3. How is the DCT computed on 2D images?

[2 marks]

We can apply the 1D DCT vertically to the columns and then horizontally to the resultant vertical DCT. Swapping the rows/columns order also works.

4. Explain differential coding why and where it is used in JPEG.

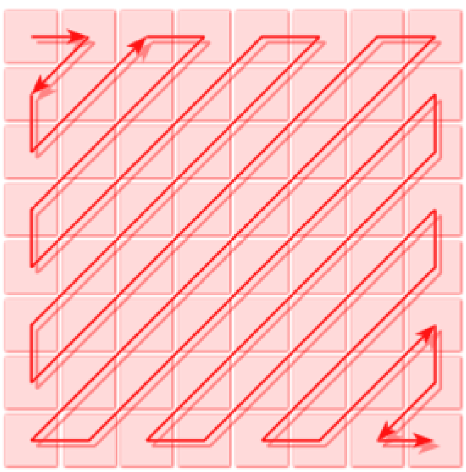
[6 marks]

Differential coding is idea of encoding the difference between two consecutive values instead of the values themselves. The idea is that the distribution of the difference

signal is more compact than distribution of the original signal, hence the entropy is lower after transformation. [3 marks]
Differential coding is used for the DC coefficients. The DC values form a downsampled image which is hard to encode (high entropy). The differential coding offers a simple way to reduce the entropy. [3 marks]

5. Explain the order in which the AC coefficients are read and why such an order is used. [6 marks]

The AC coefficients are read in a zigzag manner, starting from the lowest frequencies and progressing to the highest frequency values (see figure below)



The idea is to maximize the run length of consecutive 0's after quantisation. It is expected that higher frequencies have higher probability of being quantised to 0 (because of the structure of the quantisation matrix). Hence by making sure that lower frequencies are packed at the beginning of the scan, we minimise the number of non-zeros entries at the end of the scan and hence we maximise the number of consecutive zeros. In particular, this maximizes the length of the last run of zeros, thus shortening the length of the block to encode.

Question 5

1. Briefly explain the process of Full Search Block Matching and introduce block size, search width.

   [5 marks]

   Full search block matching is a motion estimation technique. Say we have two frames n and $n+1$. Frame n image is typically subdivided in macroblocks of size 16×16 (block size). For each block, we find the associated motion vector by finding the location of the best matching block in frame $n+1$, within a radius of $\pm$ search width pixels.

2. Estimate the number of operations required to do full search block matching on a 4K resolution footage (3840×2160), with a Block Size of 16×16 pixels and a search size of ± 64 pixels.

   [5 marks]

   Denote N the block size and w the search width. The cost per block is about: $3N^2(2w+1)^2$ per block. In our case we have $N = 16$ and $w = 64$, thus the cost per block is $3x16^2(2x64+1)^2 = 1.28e+07$ there are $3840/16 = 240x2160/16 = 135$ blocks in the image, hence we have $240x135x1.28e+07 = 4.14e+11$ operations.

3. Explain the meaning of I-frame, B-frame, P-frame and GOP and explain their purpose in MPEG2.

   [5 marks]

   I-frames are 'intra-coded pictures' and are in effect a fully encoded picture (like JPEG). P-frames are 'predicted pictures' and encode only the changes between the current frame and the motion compensated previously decoded frame.
   B-frames are 'bi-directive pictures' and encode the difference between the current frame and the motion compensated preceding of succeeding frames. GOP is a 'group of pictures'.
   The GOP structure specifies the order in which the I,P and B frames are decoded. The GOP starts with an I-frame and then only contains P or B frames.

The GOP size refers to the distance between two I-frames.

4. Consider video surveillance application of a static scene. We are only interested when there is activity in the video. Devise a simple and fast motion detection system that operates in the compressed stream.

[6 marks]

Different options are possible.
It is expected that P-frames will be very efficient when there is no activity, hence a possible solution would be to set a threshold on the size of the P-frames.

5. Explain how we could optimize the bandwidth by adapting the GOP structure to the presence or not of motion.

[5 marks]

Theoretically the bandwidth could be optimized by setting a very long GOP when there is no activity. (eg. I-frame every 90 frames) and choose only P frames in the GOP structure. Then, when motion is detected, set the next GOP to be smaller (say around 12 to 15 frames) as in typical MPEG2.