Version JL/2

EGT3
ENGINEERING TRIPOS PART IIB: SOLUTIONS

Monday 23 April 2018   2 to 3:40

Module 4F8

IMAGE PROCESSING AND IMAGE CODING

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

Answers to questions in each section should be tied together and handed in separately.

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 (a) Perception of images is very much concerned with lines and edges. It can be shown that if we discard the amplitude information present in the 2D FT of an image, we can still reconstruct a recognisable image due to the fact that edge information is retained in the phases of the FT.

If a filter phase response is non-linear, then the various frequency components which contribute to an edge in an image will be phase-shifted with respect to each other in such a way that they no longer add up to produce a sharp edge – i.e. dispersion takes place. It is often simplest to enforce the *zero-phase* condition, i.e. insisting that the frequency response is purely real, so that

$$H(\omega_1, \omega_2) = H^*(\omega_1, \omega_2)$$

Thus, ensuring that our filters are zero-phase will ensure that we preserve edges – crucial for image recognition. [10%]

(b) The standard result for the impulse response of this simple ideal lowpass filter is (see notes, or work it out, or use databook)

$$h_1(n_1\Delta_1, n_2\Delta_2) = \frac{\Delta_1\Delta_2\Omega_c^2}{(\pi)^2}\operatorname{sinc}(\Omega_c n_2\Delta_2)\operatorname{sinc}(\Omega_c n_1\Delta_1)$$

In order to avoid aliasing, we require $\Omega_c < \pi/\Delta_1$ and $\Omega_c < \pi/\Delta_2$, ie sampling rates are greater or equal to twice the largest frequency present. [10%]

(c) We know that the ideal impulse response, $h_2(n_1\Delta_1, n_2\Delta_2)$ is given by

$$h_2(n_1\Delta_1, n_2\Delta_2) = \frac{\Delta_1\Delta_2}{(2\pi)^2}\int_{-\pi/\Delta_2}^{\pi/\Delta_2}\int_{-\pi/\Delta_1}^{\pi/\Delta_1} H_2(\omega_1, \omega_2)\mathrm{e}^{j(\omega_1 n_1\Delta_1 + \omega_2 n_2\Delta_2)}d\omega_1 d\omega_2$$

$$= \frac{\Delta_1\Delta_2}{(2\pi)^2}\int\int_R \mathrm{e}^{j(\omega_1 n_1\Delta_1 + \omega_2 n_2\Delta_2)}d\omega_1 d\omega_2$$

where R is the region where $H_2 = 1$. Noting that the lines of the rotated square have equations $\omega_2 = \omega_1 \pm \Omega_s$ and $\omega_2 = -\omega_1 \pm \Omega_s$, it seems sensible to change to the following variables

$$\omega_1{}' = \omega_1 + \omega_2 \quad \text{and} \quad \omega_2{}' = \omega_1 - \omega_2$$

so that the diagonal lines in R become lines of constant ω_1' or ω_2'. Thus the integral becomes

$$\frac{\Delta_1\Delta_2}{(2\pi)^2}\int_{-\Omega_s}^{\Omega_s}\int_{-\Omega_s}^{\Omega_s} \mathrm{e}^{j\frac{\omega_1'}{2}(n_1\Delta_1+n_2\Delta_2)}\mathrm{e}^{j\frac{\omega_2'}{2}(n_1\Delta_1-n_2\Delta_2)}|J|d\omega_1'd\omega_2'$$

where $|J|$ is the magnitude of the jacobian of the transformation from (ω_1,ω_2) to (ω_1',ω_2'); $|J| = \left|\frac{\partial\omega_1}{\partial\omega_1'}\frac{\partial\omega_2}{\partial\omega_2'} - \frac{\partial\omega_1}{\partial\omega_2'}\frac{\partial\omega_2}{\partial\omega_1'}\right| = \frac{1}{2}$.
Thus, the integral can be evaluated as

$$\frac{\Delta_1\Delta_2}{2(2\pi)^2}\left[\frac{\mathrm{e}^{j\frac{\omega_1'}{2}(n_1\Delta_1+n_2\Delta_2)}}{\frac{j}{2}(n_1\Delta_1+n_2\Delta_2)}\right]_{-\Omega_s}^{\Omega_s}\left[\frac{\mathrm{e}^{j\frac{\omega_2'}{2}(n_1\Delta_1-n_2\Delta_2)}}{\frac{j}{2}(n_1\Delta_1-n_2\Delta_2)}\right]_{-\Omega_s}^{\Omega_s}$$

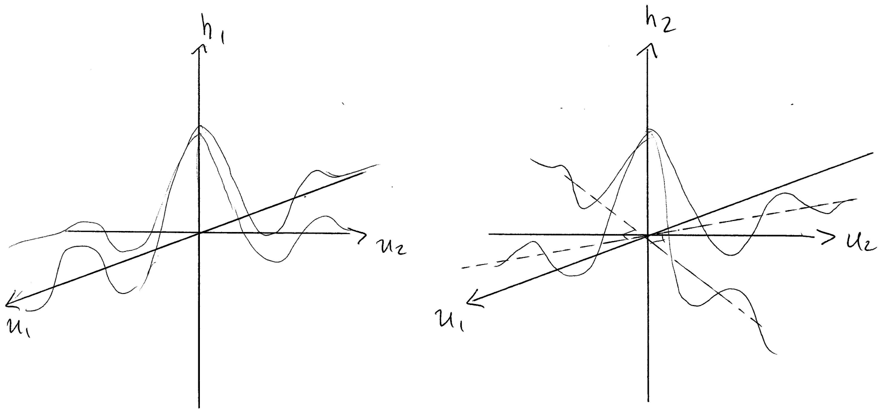
Giving

$$\frac{\Delta_1\Delta_2}{2(\pi)^2}{\Omega_s}^2\ \mathrm{sinc}(n_1\Delta_1+n_2\Delta_2)\frac{\Omega_s}{2}\ \mathrm{sinc}(n_1\Delta_1-n_2\Delta_2)\frac{\Omega_s}{2}$$

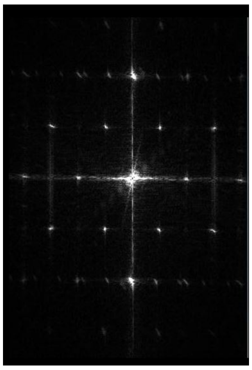
Compare this expression with that for h_1 in part (b): we see that the two expressions are very similar – ie can we simply obtain h_2 by considering some form of h_1 in a rotated coordinate system? Clearly, if we look at H_1 in coordinates which are $\omega_1' = \omega_1+\omega_2$ and $\omega_2' = \omega_1-\omega_2$, we will be looking at a diamond shape as in H_2 with $\Omega_s = \sqrt{2}\Omega_c$ and sampling intervals which are scaled by a factor of $\sqrt{2}$. Careful substitutions (noting also that the sampling moves from a rectangular grid to a diamond grid) should then enable us to obtain h_2 from h_1.

[30%]

(d) Sketches of h_1 and h_2 are given below.



If we were to take a 2D Fourier transform of the brick image in figure 2, we obtain a number of frequencies as illustrated below (there are many frequencies due to patterns on a number of scales in such an image)



Filtering with H_2 will retain more high frequencies along the axes than filtering with H_1, but will retain fewer regions of the frequency plane which have high frequencies in both directions. Thus, we might expect H_2 to produce an image which does not pick out diagonal frequencies/patterns as well as H_1. [20%]

(e) Taking the inverse FT of the ideal frequency response will give an impulse response which does not have finite support – to remedy this we multiply by a *window function* which forces the impulse response coefficients to zero for (n_1, n_2) outside R_h, the desired support region. The actual filter frequency response $H(\omega_1, \omega_2)$ is then given by the **convolution** of the desired frequency response $H_d(\omega_1, \omega_2)$ with the window function spectrum $W(\omega_1, \omega_2)$.

Thus the effect of the window is to smooth H_d – clearly we would prefer to have the mainlobe width of $W(\omega_1, \omega_2)$ small so that H_d is changed as little as possible. We also want sidebands of small amplitude so that the ripples in the (ω_1, ω_2) plane outside the region of interest are kept small.

The two most popular methods of forming 2d windows from 1d windows are

(i) Taking the product of 1d windows:

$$w(u_1, u_2) = w_1(u_1)\, w_2(u_2)$$

(ii) Rotating a 1d window:

$$w(u_1, u_2) = w_1(u)|_{u=\sqrt{(u_1^2+u_2^2)}}$$

[10%]

(f) We know that the 2D spectrum is the product of the 1D spectra.

$$W_i(\omega_i) = \int_{-U_i}^{U_i} \cos^2\left(\frac{\pi u_i}{U_i}\right) \mathrm{e}^{-j\omega_i u_i} du_i = \frac{1}{4}\int_{-U_i}^{U_i} (2 + \mathrm{e}^{j2\pi u_i/U_i} + \mathrm{e}^{-j2\pi u_i/U_i})\mathrm{e}^{-j\omega_i u_i} du_i$$

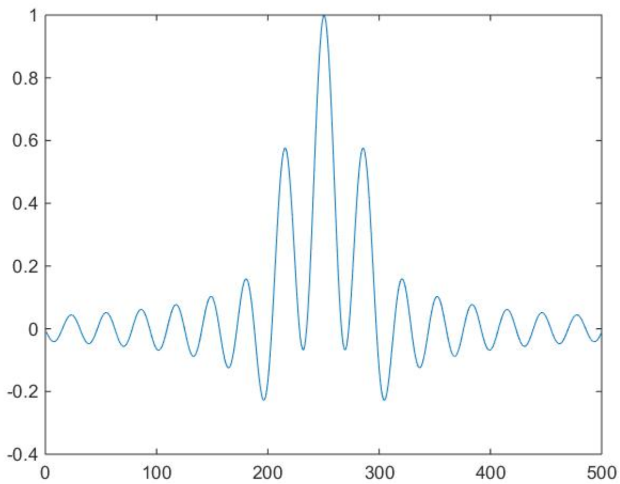
using $\cos^2\theta = (1+\cos 2\theta)/2$.

This can then be easily integrated to give:

$$W_i(\omega_i) = \frac{U_i}{2}\left(2\ \mathrm{sinc}\omega_i U_i + \ \mathrm{sinc}\left[U_i\left(\frac{2\pi}{U_i} - \omega_i\right)\right] + \ \mathrm{sinc}\left[U_i\left(\frac{2\pi}{U_i} + \omega_i\right)\right]\right)$$

and $W(\omega_1, \omega_2) = W_1(\omega_1)W_2(\omega_2)$.

Sketch of the spectrum along the ω_1 axis is given below ($U_1 = 1$) – the window shows reasonable behaviour: it has a mainlobe which is not too wide and sidelobes which show decent behaviour, though the sidelobes next to the main lobe are not attenuated. Note that no tuning is possible here so the amplitudes of the sidelobes cannot be reduced at all.



[20%]

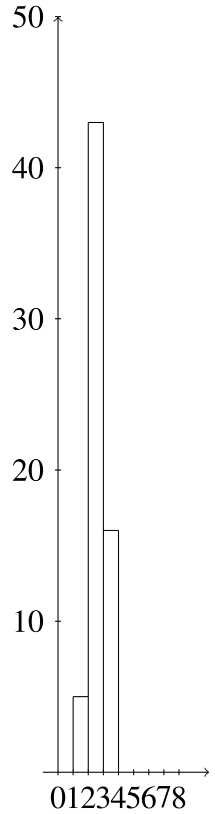
This was the least popular question. Parts a) and b) of this question were done well by almost all candidates. Part c) caused most difficulty – many people tried to do the change of variables in an ad-hoc fashion, using the given result to guess what they had to do. Sketching in part d) was also poorly done.

Part e) was bookwork on windowing and was well done by all candidates. Part f) was less well done, but mostly this appeared to be because time was running out.

2 (a) (i) A histogram plot of frequency of occurrence of grey levels in an image against grey level will tell us how much the available grey levels are used. An intuitively appealing idea would be to apply a transformation or mapping to the image pixels in such a way that the probability of occurrence of the various grey levels should be constant, i.e, all grey levels are equiprobable, which would correspond to a constant amplitude histogram. This process is called *histogram equalisation*.

Histogram equalisation is often useful in bringing out detail in images which make poor use of the available grey levels – this may occur due to poor illumination of the scene, or non-linearity in the imaging system. [10%]

(ii) The histogram of the image in Figure 3 (of the paper) is shown below. We can see that the grey levels used are concentrated around the bottom end of the range, i.e. 1-3, with mid and high levels unused. [10%]



[10%]

(iii) It often helps to draw up a table when performing histogram equalisation: below let $H(i)$ be the frequency values and $C(i)$ be the cumulative frequency values

i	1	2	3	4	5	6	7	8
$H(i)$	5	43	16	0	0	0	0	0
$C(i)$	5	48	64	64	64	64	64	64

The transformed levels are given by

$$y_k = \sum_{i=1}^{k} L \frac{N_i}{NM}, \quad k = 1 \ldots 8$$

where $N \times M$ are the dimensions of the image, N_i is the number of pixels in grey level i (equivalent to $H(i)$ above) and L is the range in grey level space. Therefore, $L = 8$, $NM = 64$ and

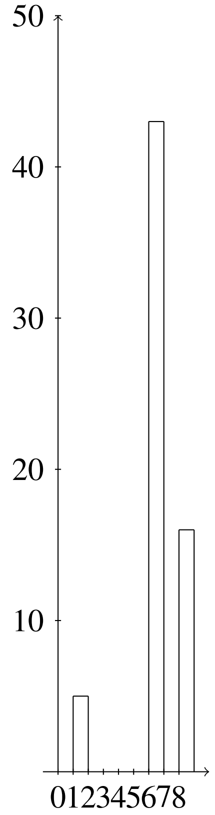
$$y_k = \frac{L}{NM} \sum_{i=1}^{k} N_i = \frac{1}{8} \sum_{i=1}^{k} N_i = \frac{1}{8} C(k), \quad k = 1 \ldots 8$$

We can now add an extra line to our table to show the transformed values:

i	1	2	3	4	5	6	7	8
$H(i)$	5	43	16	0	0	0	0	0
$C(i)$	5	48	64	64	64	64	64	64
$y(i)$	0.635	6	8	8	8	8	8	8

From this table it is now easy to draw the new image and sketch the new histogram

6	6	6	6	6	6	6	6
6	6	6	8	8	8	6	6
6	6	8	8	1	8	8	6
6	6	8	1	1	1	8	6
6	6	8	8	1	8	8	6
6	6	6	8	8	8	6	6
6	6	6	6	6	6	6	6
6	6	6	6	6	6	6	6



We can see from the new histogram that the process has succeeded in spreading out the grey levels more evenly across the scale but that the distribution is far from being uniform. The discreteness of the problem means that the equalisation process tries to do the best job it can according to the rules prescribed. The spread of greylevels can then be improved by interpolation after the histogram equalisation process, if required. [25%]

(b) (i) For *spatially stationary processes* $x(\mathbf{n})$, $y(\mathbf{n})$, the cross correlation function is defined as follows and is 'translationally invariant', i.e.

$$R_{xy}(\mathbf{0},\mathbf{n}) \equiv R_{xy}(\mathbf{n}) = E[x(\mathbf{k})y^*(\mathbf{k}-\mathbf{n})] \;\; \forall \mathbf{k}$$

i.e. the cross-correlation between the origin in the x image and the point $\mathbf{n}$ in the y image is independent of where the origin is taken.
The *cross-power spectrum* of two jointly stationary processes $x(\mathbf{n})$ and $y(\mathbf{n})$ is written as $P_{xy}(\omega)$ and is given by the FT of the cross-correlation function

$$P_{xy}(\omega) = FT(R_{xy}(\mathbf{n}))$$

Note here that the following convention for the cross correlation (denote as R^+_{xy}) is also fine

$$R^+_{xy}(\mathbf{n}) \equiv R_{xy}(-\mathbf{n}) = E[x(\mathbf{k})y^*(\mathbf{k}+\mathbf{n})] \;\; \forall \mathbf{k}$$

[10%]

(ii) We choose the filter impulse response $g(\mathbf{q})$ such that the expectation of the squared error is minimised.

$$\text{Minimise } Q = E\{[x(\mathbf{n}) - \hat{x}(\mathbf{n})]^2\}$$

$$= E\left\{\left[x(\mathbf{n}) - \sum_{\mathbf{q}\in\mathbf{Z^2}} g(\mathbf{q})\, y(\mathbf{n}-\mathbf{q})\right]^2\right\}$$

Differentiate this objective function with respect to $g(\mathbf{p})$

$$\frac{\partial Q}{\partial g(\mathbf{p})} = E\left\{2\left[x(\mathbf{n}) - \sum_{\mathbf{q}\in\mathbf{Z^2}} g(\mathbf{q})\, y(\mathbf{n}-\mathbf{q})\right][-y(\mathbf{n}-\mathbf{p})]\right\} = 0 \quad \forall \mathbf{p} \in \mathbf{Z^2}$$

$$\therefore\; E\{x(\mathbf{n})\, y(\mathbf{n}-\mathbf{p})\} = \sum_{\mathbf{q}} g(\mathbf{q})\, E\{y(\mathbf{n}-\mathbf{q})\, y(\mathbf{n}-\mathbf{p})\}$$

If the images are spatially stationary (with x and y real), we can then write:

$$E\{x(\mathbf{n})\, y(\mathbf{n}-\mathbf{p})\} = R_{xy}(\mathbf{p})$$

If we rewrite $y(\mathbf{n}-\mathbf{p})$ as $y(\mathbf{n}-\mathbf{q}+\mathbf{q}-\mathbf{p})$, the expectation in the RHS of the earlier equation becomes

$$E\{y(\mathbf{n}-\mathbf{q})\, y(\mathbf{n}-\mathbf{p})\} = E\{y(\mathbf{n}-\mathbf{q})\, y(\mathbf{n}-\mathbf{q}+\mathbf{q}-\mathbf{p})\}$$

$$= E\{y(\mathbf{k})\, y(\mathbf{k}+\mathbf{q}-\mathbf{p})\} = R_{yy}(\mathbf{p}-\mathbf{q})$$

$$\therefore R_{xy}(\mathbf{p}) = \sum_{\mathbf{q}} g(\mathbf{q})\, R_{yy}(\mathbf{p}-\mathbf{q}) \quad \forall \mathbf{p} \in \mathbf{Z^2}$$

Taking the Fourier transform of the above equation will lead (after some manipulation) to the form of the Wiener filter (frequency domain) given. [30%]

(iii) Following derivation not necessary, but if candidates don't remember the final result then they will have to go through it: [Derivation:

$$R_{yy}(\mathbf{p}) = E\{y(\mathbf{n})\, y(\mathbf{n}-\mathbf{p})\} \;\text{ where }\; y(\mathbf{n}) = \sum_{\mathbf{m}} h(\mathbf{m})\, x(\mathbf{n}-\mathbf{m}) + d(\mathbf{n})$$

So, if signal and noise are uncorrelated (so that the cross terms are zero) and noise is zero mean:

$$R_{yy}(\mathbf{p}) = E\left\{\sum_{\mathbf{m}}\sum_{\mathbf{q}} h(\mathbf{m})\, x(\mathbf{n}-\mathbf{m})\, h(\mathbf{q})\, x(\mathbf{n}-\mathbf{p}-\mathbf{q})\right\} + E\{d(\mathbf{n})\, d(\mathbf{n}-\mathbf{p})\}$$

$$= \sum_{\mathbf{m}} \sum_{\mathbf{q}} h(\mathbf{m})\, h(\mathbf{q}) E\left\{x(\mathbf{n}-\mathbf{m})\, x(\mathbf{n}-\mathbf{p}-\mathbf{q})\right\} + R_{dd}(\mathbf{p})$$

$$\therefore R_{yy}(\mathbf{p}) = \sum_{\mathbf{m}} \sum_{\mathbf{q}} h(\mathbf{m})\, h(\mathbf{q})\, R_{xx}(\mathbf{p}+\mathbf{q}-\mathbf{m}) + R_{dd}(\mathbf{p})$$

as $E\left\{x(\mathbf{n}-\mathbf{m})x(\mathbf{n}-\mathbf{p}-\mathbf{q})\right\} = E\left\{x(\mathbf{n}-\mathbf{m})x(\mathbf{n}-\mathbf{m}-(\mathbf{p}+\mathbf{q}-\mathbf{m}))\right\}$

Now take the Fourier transform of each side to give:

$$P_{yy}(\omega) = \sum_{\mathbf{p}} \left\{ \sum_{\mathbf{m}} \sum_{\mathbf{q}} h(\mathbf{m})\, h(\mathbf{q})\, R_{xx}(\mathbf{p}+\mathbf{q}-\mathbf{m}) \right\} e^{-j\omega^T \mathbf{p}} + P_{dd}(\omega)$$

where P_{dd} is the FT of the autocorrelation function of the noise. Interchange order;

$$P_{yy}(\omega) = \sum_{\mathbf{m}} \sum_{\mathbf{q}} h(\mathbf{m})\, h(\mathbf{q}) \sum_{\mathbf{p}} R_{xx}(\mathbf{p}+\mathbf{q}-\mathbf{m})\, e^{-j\omega^T \mathbf{p}} + P_{dd}(\omega)$$

Let $\mathbf{k} = (\mathbf{p}+\mathbf{q}-\mathbf{m})$, then:

$$P_{yy}(\omega) = \sum_{\mathbf{m}} \sum_{\mathbf{q}} h(\mathbf{m})\, h(\mathbf{q}) \sum_{\mathbf{k}} R_{xx}(\mathbf{k})\, e^{-j\omega^T (\mathbf{k}-\mathbf{q}+\mathbf{m})} + P_{dd}(\omega)$$

$$\therefore P_{yy}(\omega) = \left\{ \sum_{\mathbf{m}} h(\mathbf{m})\, e^{-j\omega^T \mathbf{m}} \right\} \left\{ \sum_{\mathbf{q}} h(\mathbf{q})\, e^{j\omega^T \mathbf{q}} \right\} \left\{ \sum_{\mathbf{k}} R_{xx}(\mathbf{k})\, e^{-j\omega^T \mathbf{k}} \right\} + P_{dd}(\omega)$$

end derivation]

$$\therefore P_{yy}(\omega) = |H(\omega)|^2\, P_{xx}(\omega) + P_{dd}(\omega)$$

as h is real. Reiterate we are assuming that the signal and noise are uncorrelated and that the noise is zero mean. [15%]

This was the most popular question and was done by almost all candidates. Part a) on histogram equalisation in images was uniformly well done – clearly everyone had revised this. Part b) was less well done, but as it was mainly bookwork, many people had thoroughly revised the ideas behind Wiener filtering and everyone managed some marks. Sub-part (ii) caused most trouble – of those who got through the derivation, many forgot to state the assumptions they made.

3 (a) The key characteristics of vision that are exploited in image compression are:

(i) The human visual system (HVS) is much more sensitive to overall intensity (luminance) changes than to colour changes. Usually most of the information about a scene is contained in its luminance rather than its colour (chrominance).

(ii) The bandwidth of the HVS for luminance components is much wider than for chrominance (typically about 5 times as wide).

(iii) The contrast sensitivity of the HVS for luminance is also around 3 times better than red-green sensitivity and around 6 times better than blue-yellow.

(iv) The luminance sensitivity also drops off at low spatial frequencies, but only if there is no temporal fluctuation (flicker).

(v) Activity masking occurs, such that in the presence of high image activity (eg a strong texture) it is much more difficult to notice coding distortions than in smooth areas of low activity.

These characteristics are exploited by performing a transformation from RGB to YUV colour space, and by using a lower sampling rate and coarser quantisation for the colour channels. Also the quantisation step size may be designed to adapt to local image activity levels to take advantage of activity masking. [20%]

(b) If T is an orthonormal matrix, $T\,T^T = I$:

$$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = I$$

T is therefore orthonormal.

The 2-D Haar transform of X is given by

$$Y = T\,X\,T^T = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} a+b & a-b \\ c+d & c-d \end{bmatrix} = \frac{1}{2}\begin{bmatrix} a+b+c+d & a-b+c-d \\ a+b-c-d & a-b-c+d \end{bmatrix}$$

Since $Y = T\,X\,T^T$, we see that $T^T Y T = (T^T T)X(T^T T) = X$, as T is orthonormal; X can therefore easily be obtained from the transform Y using T.

We have seen that the Haar matrix, T, is orthonormal; therefore the inner product of any two different columns or rows is zero, and the inner product (energy) of each row/column with itself is unity. It is well known that multiplying a vector or matrix by an orthonormal (or unitary) matrix preserves the energy of the input vector or matrix. We see this by considering the energy of $\mathbf{y} = T\mathbf{x}$

$$\mathbf{y}^T\mathbf{y} = \mathbf{x}^T T^T T\mathbf{x} = \mathbf{x}^T I\mathbf{x} = \mathbf{x}^T\mathbf{x}$$

since $T^T T = I$. [20%]

(c) To apply the Haar transform to a complete image, we group the pixels into 2×2 blocks and apply the transform to each block – we then group all top left components into a subimage, all top right components into a subimage etc; giving us 4 subimages.

- Consider the top left coefficients from each 2×2 image block. These represent half the sum of the four pixels in each block, which represents an averaging or lowpass filtering of each block. Hence the top left subband is a slightly blurred version of the original image, scaled up in amplitude by a factor 2, but reduced in size by 2:1 horizontally and vertically.
- The top right coefficients can be written $\frac{a+c}{2} - \frac{b+d}{2}$, which measures the mean horizontal gradient of each 2×2 block. Hence this subband picks out vertical edges in the image, where pixels a and c are significantly different from pixels b and d.
- The lower left coefficients can be written $\frac{a+b}{2} - \frac{c+d}{2}$, which measures the mean vertical gradient of each 2×2 block. Hence this subband picks out horizontal edges in the image, where pixels a and b are significantly different from pixels c and d.
- The lower right coefficients can be written $\frac{a+d}{2} - \frac{b+c}{2}$, which measures the diagonal curvature of each 2×2 block. Hence this subband picks out corners and textures in the image, where pixels a and b are significantly different from pixels c and d.

Most of the energy (generally over 90%) is contained in the lo-lo (top left components) subimage and similarly, lo-lo typically contains over 50% of the entropy.

Since this top left subimage is the result of lowpass filtering in both directions, it has similar characteristics to the original image (except that it is smaller). Image compression techniques can therefore usefully be applied to this lo-lo subimage. [20%]

(d) Suppose we quantise both the original image, X, and the transformed image, Y, with a given step size Q_{step}. We know that the Haar transform preserves energy. Quantising errors can be modelled as independent random processes with variance (energy) $= Q_{step}^2/12$ [recall IB Comms] and the total squared quantising error (distortion) will tend to the sum of variances over all pixels. This applies whether the energies are summed before or after the inverse transform (reconstruction) in the decoder. Hence equal quantiser step sizes before and after an energy-preserving transform should generate equivalent quantising distortions.

We saw in part (c), that the top left subimage has similar characteristics to the original image but is smaller. We may then apply the Haar transform again to this subimage, splitting it into four more bands. The lowpass result of this second stage may be further decomposed by a third Haar transform, and so on for as many levels as required. The energy compaction properties of the transform are improved by each of the first few levels (typically four), but beyond that there is virtually no additional gain. These multi-level transforms are a key feature of most image compression algorithms: we can compress our data by compressing the subimages which contain most of the energy. [20%]

(e) When subbands are quantised severely, the coefficients from the finer levels are often set to zero, so the reconstructed image is largely made up from basis functions of the lowpass subbands. At level 1 the Haar lowpass basis functions are just 2×2 blocks of equal pixels, and at coarser levels the block size increases to 4×4, 8×8 etc. Hence the image is made up from different sized block patterns, giving it a somewhat 'blocky' appearance, with larger blocks in areas of greater smoothness in the original image.

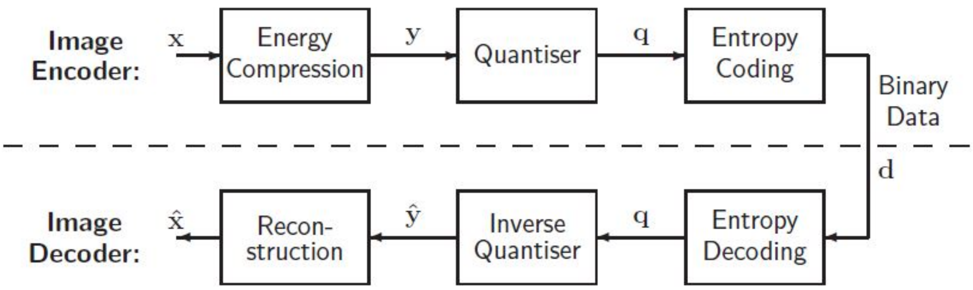
The sum and difference operations of the Haar transform may be regarded as simple types of FIR filters, applied to the rows and columns of the image. To avoid the 'blocky' artifacts, we must modify these filters so that they have smoother responses. Wavelet concepts allow us to do this. Using the concept of a two-band filter bank with perfect reconstruction inverse filters, it is possible to design filters which have much better smoothness than the Haar filters by forcing them to have multiple zeros at $z = -1$ in the z-plane. This makes the filters more complex but gives quantisation artifacts which are less visible because the lowpass basis functions have smooth boundaries, rather than the sharp edges of the Haar functions. [20%]

This was the second most popular question, and was largely well done. The question was predominantly bookwork. Strangely, part (a), which should have been the easiest part, was the part with the lowest average, with most people simply giving one or two characteristics of the human visual system.

4 Note: this question discusses the 1D cases – since when applying to an image, the rows and columns are done separately, so it is effectively two 1D transforms.

The figure below sketches the main components of an encoder-decoder system.

(a)



The *compression* and *reconstruction* blocks will be transforms that concentrate a high proportion of the image energy into as few coefficients as possible, while preserving energy.

The *quantiser* represents coefficients to a given level of accuracy and should be set so as to control the trade-off between distortion and bit rate. The *inverse quantiser* gives the best estimate of the image transform given the decoded data.

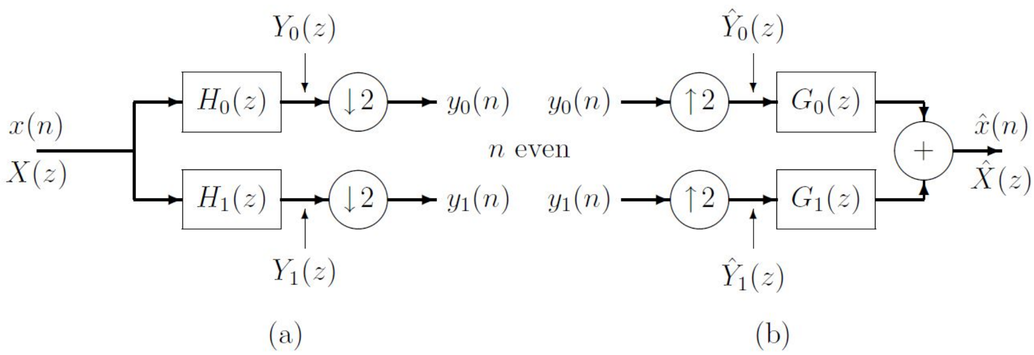
The *coder* encodes the output of the quantiser into a bit stream – this should attempt to minimise the total number of bits based on the statistics of various classes of samples.

The compression/reconstruction and the coding/decoding should all be *lossless*. It is only the quantiser that introduces loss and distortion.

Perfect Reconstruction in such a system refers to the *transform* and *inverse transform* blocks – this means they are lossless and the input can be recovered **exactly** from the output. In this case, it is indeed only the quantiser step that introduces loss and distortion.

[20%]

(b) A sketch of a two-band analysis/reconstruction filter bank system is shown in the figure below((a) shows analysis and (b) shows reconstruction).



The downsamplers by 2 omit all samples $y(n)$ when n is odd.

The upsamplers by 2 insert zeros in place of the missing odd-numbered samples.
The downsamplers and upsamplers are used to avoid redundancy in the two systems. [20%]

(c) We are given that $\hat{y}(n) = y(n)$ when n is even and $\hat{y}(n) = 0$ when n is odd. We know that

$$\begin{aligned} Y(z) &= \sum_{-\infty}^{\infty} y(n)z^{-n} \equiv \sum_{n \text{ even}} y(n)z^{-n} \\ &= \sum_{\text{all } n} \frac{1}{2}\left[y(n)z^{-n} + y(n)(-z)^{-n}\right] \\ &= \frac{1}{2}\sum_{\text{all } n} y(n)z^{-n} + \frac{1}{2}\sum_{\text{all } n} y(n)(-z)^{-n} \\ &= \frac{1}{2}\left[Y(z) + Y(-z)\right] \end{aligned}$$

[15%]

(d) For the filter banks in part (b) we have

$$Y_0(z) = H_0(z)X(z) \quad \text{and} \quad Y_1(z) = H_1(z)X(z)$$
$$\hat{Y}_0(z) = \frac{1}{2}\left[Y_0(z) + Y_0(-z)\right] \quad \text{and} \quad \hat{Y}_1(z) = \frac{1}{2}\left[Y_1(z) + Y_1(-z)\right]$$

and

$$\hat{X}(z) = G_0(z)\hat{Y}_0(z) + G_1(z)\hat{Y}_1(z)$$

Combining these expressions we have:

$$\begin{aligned} \hat{X}(z) &= \frac{1}{2}G_0(z)\left[H_0(z)X(z) + H_0(-z)X(-z)\right] + \frac{1}{2}G_1(z)\left[H_1(z)X(z) + H_1(-z)X(-z)\right] \\ &= \frac{1}{2}X(z)\left[G_0(z)H_0(z) + G_1(z)H_1(z)\right] + \frac{1}{2}X(-z)\left[G_0(z)H_0(-z) + G_1(z)H_1(-z)\right] \end{aligned}$$

For antialiasing, the $X(-z)$ term must be zero and so we require that

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

For perfect reconstruction, the $X(z)$ term must be multiplied by unity, so we require that

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

[25%]

(e) To satisfy the antialiasing condition and make $G_1(z)$ and $H_1(z)$ highpass when $G_0(z)$ and $H_0(z)$ are lowpass, we let

$$G_1(z) = z^k H_0(-z) \quad \text{and} \quad H_1(z) = z^{-k} G_0(-z)$$

with k being an odd integer (usually ± 1).

Then the perfect reconstruction (PR) condition becomes

$$G_0(z)H_0(z) + G_0(-z)H_0(-z) = 2$$

or

$$P(z) + P(-z) = 2$$

since $P = G_0 H_0$.

For symmetric left/right and up/down filter behaviour in images, we usually assume linear-phase filters, so that $p_{-n} = p_n$.

The PR condition causes all odd coefficients p_n in $P(z)$ to be cancelled when it is added to $P(-z)$, so it only constrains the even coefficients.

Hence $p_0 = 1$ and $p_2, p_4, p_6, ...$ are all zero. Thus $P(z)$, with symmetry, will be of the form

$$P(z) = \ldots + p_5 z^5 + p_3 z^3 + p_1 z^1 + 1 + p_1 z^{-1} + p_3 z^{-3} + p_5 z^{-5} + \ldots$$

The design process is to find a good set of coefficients, $\{p_1, p_3, p_5, \ldots\}$, such that $P(z)$ is a well-shaped lowpass filter that can be factorised into two lowpass filters $H_0(z)$ and $G_0(z)$.

[20%]

All parts were essentially bookwork and were done well. Even part (d), which was reasonably involved (deriving the anti-aliasing and perfect reconstruction conditions for filter banks), was done perfectly by most.

END OF PAPER