Version JL/2

EGT3
ENGINEERING TRIPOS PART IIB: SOLUTIONS

Monday 24 April 2017 2 to 3:30

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.*

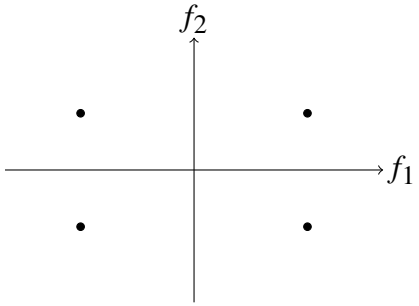
STATIONERY REQUIREMENTS
Single-sided script paper

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

10 minutes reading time is allowed for this paper.

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

1 (a) (i) Since $g(u_1,u_2) = \sin(\Omega_1 u_1)\sin(\Omega_2 u_2)$, and $\Omega_1 = \pi/8$, $\Omega_2 = \pi/16$, we have $f_1 = \Omega_1/(2\pi) = 1/16$Hz and $f_2 = \Omega_2/(2\pi) = 1/32/$,Hz. Therefore we have a 2D sinewave with these frequencies, ie peaks in the frequency domain at $(\pm 1/16, \pm 1/32)$. Sketching this:

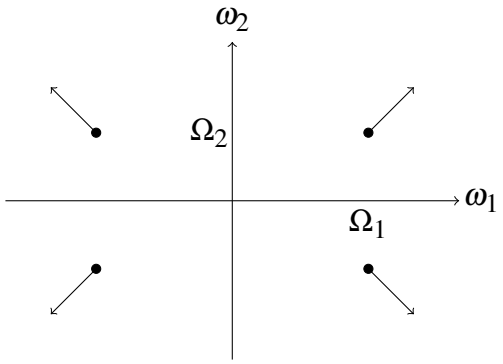


[10%]

(ii) Now suppose we instead have $g(u_1,u_2) = \sin(\phi_1(u_1))\sin(\phi_2(u_2))$. The instantaneous frequency is given by $\frac{d\phi}{du}$ [if $\phi = \Omega u$ then $\frac{d\phi}{du} = \Omega$, as expected]. Therefore:

$$\frac{d\phi_1}{du_1} = \Omega_1 + \frac{2\pi u_1}{16} \text{ and } \frac{d\phi_2}{du_2} = \Omega_2 + \frac{2\pi u_2}{32}$$

This therefore tells us that we have a 2D 'chirp' with frequency increasing as u_1 and u_2 increase. This is illustrated in the 2D frequency plane in the figure below.



[15%]

(iii) 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. The eyes are sensitive to phase while the ears are sensitive to amplitude.

[10%]

(cont.

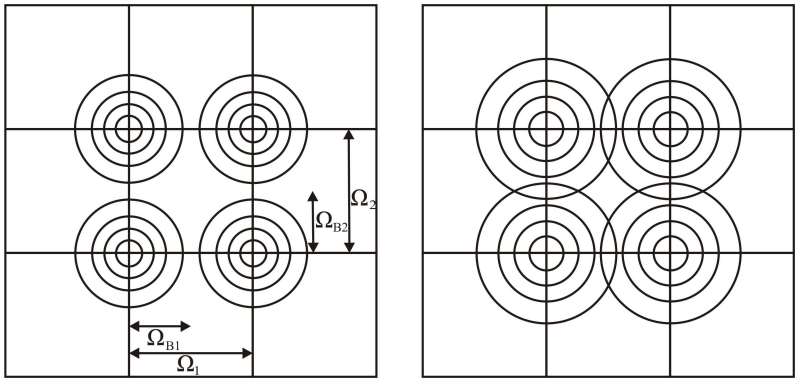
(b) (i)

$$G_s(\omega_1,\omega_2) = \frac{1}{\Delta_1\Delta_2}\sum_{p_1=-\infty}^{\infty}\sum_{p_2=-\infty}^{\infty} G(\omega_1 - p_1\Omega_1, \omega_2 - p_2\Omega_2)$$

(ii) From the equation above, it can be seen that the Fourier transform or spectrum of the sampled 2d signal is the periodic repetition of the spectrum of the unsampled 2d signal (centred on the 'grid' points in frequency space) – precisely analogous to the 1d case. This periodic repetition of the spectrum is called *aliasing*. If the image is spatially bandlimited to Ω_{B1} and Ω_{B2} then the original continuous image can be recovered from the sampled image by ideal low-pass filtering at Ω_{B1}, Ω_{B2} if the samples are taken such that $\Omega_{B1} < \frac{1}{2}\Omega_1$ and $\Omega_{B2} < \frac{1}{2}\Omega_2$ so that the periodic repeats of the spectrum do not overlap – this can also be written as:

$$\frac{2\pi}{\Delta_1} > 2\Omega_{B1} \quad \frac{2\pi}{\Delta_2} > 2\Omega_{B2}$$

$2\Omega_{B1}$ and $2\Omega_{B2}$ are known as the *2d Nyquist frequencies*. Thus the *2d sampling theorem* states that a bandlimited image sampled at or above its u_1 and u_2 Nyquist rates can be recovered without error by low-pass filtering the sampled spectrum. This is illustrated in the figure below.



[10%]

(iii) If $g(u_1,u_2) = \cos(\alpha u_1 + \beta u_2)$, the FT is

$$G(\omega_1,\omega_2) = \int\int \cos(\alpha u_1 + \beta u_2)\mathrm{e}^{-j(\omega_1 u_1 + \omega_2 u_2)}du_1 du_2$$

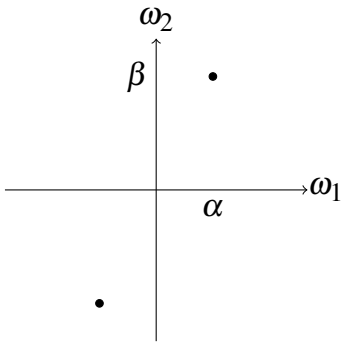
Now expand the cosine term to give $\cos\alpha u_1 \cos\beta u_2 - \sin\alpha u_1 \sin\beta u_2$, which then gives:

$$G = \int \cos\alpha u_1 \mathrm{e}^{-j\omega_1 u_1}du_1 \int \cos\beta u_2 \mathrm{e}^{-j\omega_2 u_2}du_2 - \int \sin\alpha u_1 \mathrm{e}^{-j\omega_1 u_1}du_1 \int \sin\beta u_2 \mathrm{e}^{-j\omega_2 u_2}du_2$$

which can be simplified as

$$G = \pi^2 [(\delta(\omega_1+\alpha)+\delta(\omega_1+\alpha))(\delta(\omega_2+\beta)+\delta(\omega_2+\beta))]+$$
$$\pi^2 [(\delta(\omega_1-\alpha)-\delta(\omega_1+\alpha))(\delta(\omega_2-\beta)-\delta(\omega_2+\beta))]$$
$$= 2\pi^2 [\delta(\omega_1-\alpha)\delta(\omega_2-\beta)+\delta(\omega_1+\alpha)\delta(\omega_2+\beta)]$$

Clearly g is bandlimited.



[20%]

(iv) Nyquist frequencies are simply twice the highest (vertical and horizontal) frequencies in the image, ie

$$\Omega_{n1} = 2\alpha \quad \Omega_{n2} = 2\beta$$

[10%]

(v) Sampling frequencies are given by

$$\Omega_{s1} = \frac{2\pi}{\Delta_1} = \frac{2\pi}{0.4\pi} = 5$$

$$\Omega_{s2} = \frac{2\pi}{\Delta_2} = \frac{2\pi}{0.2\pi} = 10$$

So, $5 \geq 2\alpha$ and $10 \geq 2\beta$, so we therefore take

$$\alpha = 5/2 \quad \beta = 5$$

[15%]

This was the least popular question. Parts a)(i)(iii) of this question were done well by almost all candidates. Part a)(ii) caused most difficulty – it was clear that many candidates did not know how to get instantaneous frequencies from $\sin\phi(t)$ *by differentiating* ϕ.

Part b) was generally well done. Almost everyone who attempted this question got full marks for the bookwork parts. Most marks were lost on part (iii) due to an inability to integrate.

2 (a) (i) 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%]

(ii) Taking the inverse FT of an ideal zero-Phase 2D frequency response, H, will normally create an impulse response with infinite support. Windowing is therefore necessary to produce a filter with finite support. The effect of the window is to smooth H_d, since multiplying by the window function w in the spatial domain leads to convolving the H with the FT of w, say W,in the frequency domain – 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.

[10%]

(iii) Two methods of windowing are via the *product* and *rotation* methods.
The Product Method for obtaining a 2-D window from 1-D windows is to simply take the product of two 1-D windows:

$$w(n_1, n_2) = w_1(n_1)\, w_2(n_2)$$

The Rotation Method of forming a 2-D window from 1-D windows is to obtain a 2-D *continuous* window function $w(u_1, u_2)$ by rotating a 1-D continuous window $w_1(u)$.

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

The continuous 2-D window is then sampled to produce a discrete 2-D window $w(n_1, n_2)$:

$$\boxed{w(n_1,n_2) = w(u_1,u_2)|_{u_1=n_1\Delta_1,\, u_2=n_2\Delta_2}}$$

The actual filter frequency response $H(\omega_1,\omega_2)$ is given by the **convolution** of the desired frequency response $H_d(\omega_1,\omega_2)$ with the window function spectrum $W(\omega_1,\omega_2)$.

[10%]

(b) (i) If we neglect the noise in the equation $\mathbf{y} = h * x + n$, we are left with

$$y(n_1,n_2) = \sum_{m_1}\sum_{m_2} h(m_1,m_2)x(n_1-m_1,n_2-m_2)$$

Since the relationship between x and y is a 2-D convolution, a straightforward approach to the problem of reconstruction is to take the Fourier transform of each side of the above to give:

$$Y(\omega_1,\omega_2) = H(\omega_1,\omega_2)X(\omega_1,\omega_2)$$

where: $H(\omega_1,\omega_2) = \sum_{n_2=-\infty}^{\infty}\sum_{n_1=-\infty}^{\infty} h(n_1,n_2)e^{-j(\omega_1 n_1+\omega_2 n_2)}$

$$\therefore X(\omega_1,\omega_2) = \frac{Y(\omega_1,\omega_2)}{H(\omega_1,\omega_2)} \text{ and } x(n_1,n_2) = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} X(\omega_1,\omega_2)e^{j(\omega_1 n_1+\omega_2 n_2)}d\omega_1 d\omega_2$$

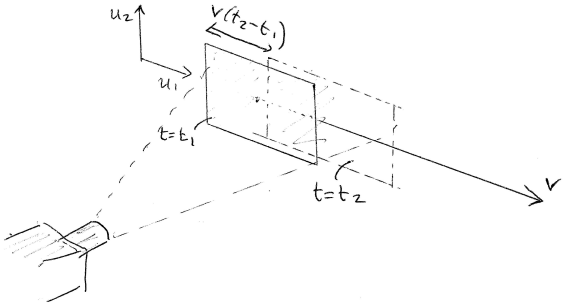
Thus, if we neglect noise and know the psf, h, we can estimate our true image by a process known as *inverse filtering*, which, as we see above, involves dividing the fourier transform of the observed image by the fourier transform of h – the inverse filter is therefore $1/H$.

[10%]

(ii) At time t the object will have moved a distance vt in the u_1 direction. If the shutter is open for T seconds, the camera will effectively 'sum' all of these images to produce a blurred image given by

$$y(u_1,u_2) = \int_0^T x(u_1 - vt, u_2)dt$$

(cont.



[15%]

(iii) Let $P(u_1)$ be the pulse given by

$$P(u_1) = \begin{cases} 1/v & 0 < u_1 < vT \\ 0 & \text{otherwise} \end{cases}$$

Now, if we let $\tau = vt$ in the above convolution, we get:

$$y(u_1,u_2) = \int_0^{vT} x(u_1-\tau,u_2)\frac{1}{v}d\tau$$

which we can write as

$$y(u_1,u_2) = \int_{-\infty}^{\infty} x(u_1-\tau,u_2)P(\tau)d\tau = P(\tau) * x$$

Thus the filter h makes no changes in the u_2 direction but is equal to P in the u_1 direction.

[20%]

(iv) For this part of the question a number of answers were acceptable. If we neglect noise, then since we have a convolution, the simplest *deblurring filter* would be the *inverse filter* as discussed in part (i). ie the simplest deblurring filter would be $1/P$. One could then add more robustness by discussing the *generalised inverse filter*. [15%]

(v) All inverse filters perform poorly in the presence of significant noise – improved performance can be achieved via the *Weiner filter*, g, whose spectrum is given by

$$G(\omega) = \frac{P_{xy}(\omega)}{P_{yy}(\omega)}$$

G is the optimal filter amongst the class of *linear filters* – the derivation minimises the sum of residuals. However, if we allow *non-linear filters*, meaning we generally need to solve via iterative methods, we can often get far superior results. Two common non-linear methods are the *Maximum Entropy* and *Pixon* methods.

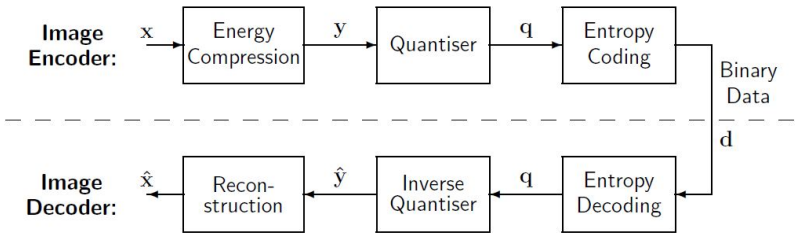
This question had the lowest average on the paper. Part (a) was mostly bookwork and almost all answers were very good. Part (b) was mostly straightforward, though parts (iii) and (iv) required some thinking outside the lecture notes, so drew the majority of the errors.

3 (a) Energy compression: compresses most of the input energy into a small proportion of the coefficients.

Quantiser: quantised the coefficients so that many of them become zero, because of energy compression.

Entropy coding: converts the spares integers into a binary data stream, using entropy coding to acheive the lowest mean bit rate consistent with lossless transmission of the quantised integers.

The 3 decoder blocks invert these processes and reconstruct the decoded image. Information is lost only in the quantiser.



[20%]

(b) If T is orthonormal, then the inverse of T is its transpose, ie $T^T T = TT^T = I$. Also the energy normalisation of each row of T means that the energy of an input vector $\mathbf{x}$, is preserved in the transformed vector $\mathbf{y} = T\mathbf{x}$, and vice versa. This means that quantising noise added to the coefficient $\mathbf{y}$ is equivalent to adding the same energy of noise to $\mathbf{x}$.

To transform the columns of a square matrix X, we calculate TX, and then to transform the rows of the result, we calculate $(TX)T^T$. Hence for a 2D transform we have $Y = TXT^T$.

To reconstruct X we compute

$$T^T YT = T^T TXT^T T = IXI = X$$

[20%]

(c) To show that T is orthonormal, we must show that $\mathbf{t}_j \mathbf{t}_k^T = 0$ if $j \neq k$, where $\mathbf{t}_j$ is the jth row of T. We must also show that $\mathbf{t}_j \mathbf{t}_j^T = 1$ for all j.

$$\mathbf{t}_1 \mathbf{t}_2^T = ab + ac - ac - ab = 0$$
$$\mathbf{t}_1 \mathbf{t}_3^T = a^2 - a^2 - a^2 + a^2 = 0$$

$$\mathbf{t}_1\mathbf{t}_4^T = ac - ab + ab - ac = 0$$
$$\mathbf{t}_2\mathbf{t}_3^T = ba - ca + ca - ba = 0$$
$$\mathbf{t}_2\mathbf{t}_4^T = bc - cb - cb + bc = 0$$
$$\mathbf{t}_3\mathbf{t}_4^T = ac + ab - ab - ac = 0$$

$$\mathbf{t}_1\mathbf{t}_1^T = 4a^2 = 1$$
$$\mathbf{t}_2\mathbf{t}_2^T = 2b^2 + 2c^2 = 2\frac{1}{2}(\cos^2(\pi/8) + \sin^2(\pi/8) = 1$$
$$\mathbf{t}_3\mathbf{t}_3^T = 4a^2 = 1$$
$$\mathbf{t}_4\mathbf{t}_4^T = 2b^2 + 2c^2 = 1$$

Hence T is orthonormal (unitary).

The inverse 2D DCT is $X = T^T Y T$. The basis function for $y_{1,1}$ is the X that results from $y_{1,1} = 1$ and all other elements of Y being zero. This is therefore $\mathbf{t}_1^T\mathbf{t}_1$ as $y_{1,1}$ selects the first column of T^T and the first row of T, similarly, $y_{i,j}$ selects $\mathbf{t}_j^T\mathbf{t}_i$.

Thus:

$$\mathbf{t}_1^T\mathbf{t}_1 = [a,a,a,a]^T[a,a,a,a]$$

ie all elements are a^2

$$\mathbf{t}_2^T\mathbf{t}_1 = [b,c,-c,-b]^T[a,a,a,a]$$

ie all columns are the same

$$\mathbf{t}_1^T\mathbf{t}_2 = [\mathbf{t}_2^T\mathbf{t}_1]^T$$

ie all rows are the same

$$\mathbf{t}_2^T\mathbf{t}_2 = [b,c,-c,-b]^T[b,c,-c,b]$$

[20%]

(d) Since a,b,c are all positive and $b \gg c$, we see by inspection that the rows of T are gradually increasing in frequency from top to bottom row. The form of the remaining basis functions are products of increasing vertical frequency from right to left and increasing horizontal frequency from top to bottom of the form $\mathbf{t}_j^T\mathbf{t}_i$.

Images of the real world have strong correlations between nearby pixels, and so they tend to have much more energy at low frequencies than at high frequencies.

The basis functions for $y_{i,j}$ contain frequencies proportional to i horizontally and j vertically, so the energy (and entropy) tend to decrease roughly as a function of $(i+j)$.

[20%]

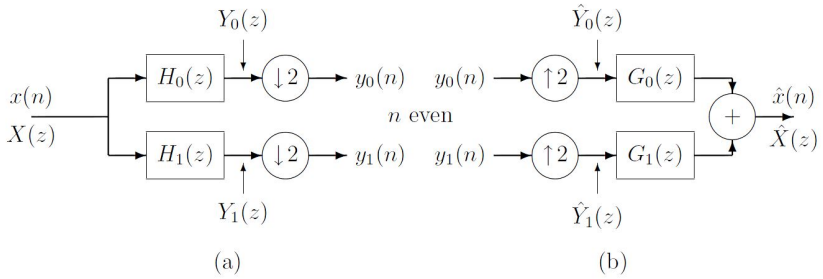
(cont.

(e) The contrast sensitivity of the human visual system tends to decrease as the spatial frequency of the input image increases. Therefore we can use coarser quantisation of coefficients $y_{i,j}$ as the frequency content of the (i, j) basis function increases. Hence the quantiser step size may be increased as (i, j) increases. This tends to reduce the entropy of the coded data for the higher frequency basis functions and gives a a lower overall data rate for the encoded data, for a given level of perceived image distortion.

[20%]

This was the second most popular question, which was well done. The question was predominantly bookwork, though some candidates nevertheless struggled with doing all the necessary calculations in part (c).

4 (a) See figure below, where the left hand side is the Analysis filter bank and the right hand side is the Reconstruction filter bank.



$H_0(z)$ and $G_0(z)$ are lowpass filters and H_1 and G_1 are highpass filters.

If the filter outputs Y_0 adn Y_1 were not downsampled, there would be a 2:1 redundancy introduced by the two filters H_0 and H_1. Since H_0 and H_1 each reduce the bandwidth of X by approx $1/2$, we can afford to downsample y_0 and Y_1 by $1/2$ so that the total sample rate of y_0 and Y_1 remains the same as X. Redundancy is not good when we are trying to achieve data compression. [20%]

(b) Consider the data samples y_n with z-transform

$$Y(z) = \sum_{n=-\infty}^{\infty} y_n z^{-n}$$

If y_n is downsampled by 2 and then upsampled by 2 to give $\hat{y}_n$, then the z-transform of $\hat{y}_n$ will be:

$$\begin{aligned} Y_0(z) &= \sum_{evenn} y_n z^{-n} \\ &= \sum_{alln} \frac{1}{2}\left[y_n z^{-n} + y_n(-z)^{-n}\right] \\ &= \frac{1}{2}\sum_n y_n z^{-n} + \frac{1}{2}\sum_n y_n(-z)^{-n} \\ &= \frac{1}{2}Y(z) + \frac{1}{2}Y(-z) \\ &= \frac{1}{2}\left[Y(z) + Y(-z)\right] \end{aligned}$$

.

[15%]

 (cont.

(c) Applying the result in (b) to the filterbank of (a) 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

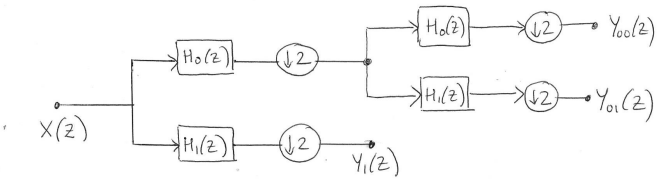
$$G_0(z)H_0(-z) + G_1(z)H_1(-z) = 0$$

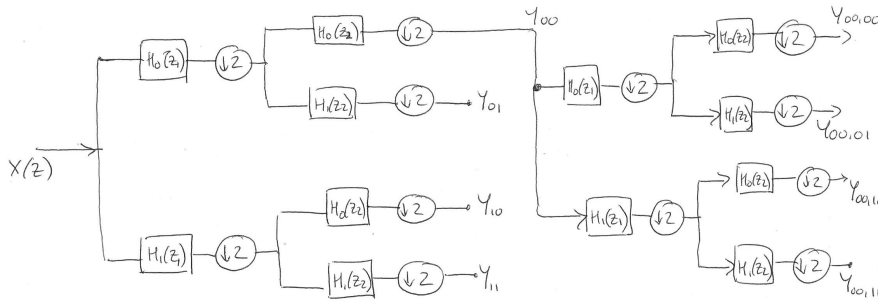
For perfect reconstruction, the $X(z)$ term must be multiplied by unity, so we require that

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

[20%]

(d) the figures below show the 2-level wavelet transforms for 1D signals and 2D signals;





At scale 1, y_{00} is the result of lowpass filtering in both directions (row and column), and it picks out the mean brightness over roughly a 2×2 region. y_{01} is the result of lowpass filtering along rows and highpass filtering down columns, so it picks out near-horizontal edges. Similarly, y_{10} picks our near-vertical edges. y_{11} is the result of highpass filtering in both directions and picks out corners, small dots and high frequency textures (eg grass). We get similar features picked out at larger scale by the four scale 2 outputs, $y_{00,00}$ to $y_{00,11}$.

[25%]

(e) The outputs of the above 2D wavelet transform may be inverted in order to reconstruct the image x by a mirror image of the above diagram based on reconstruction filter banks in which the H filters are replaced by equivalent G filters and the downsamplers by upsamplers.

If the G and H filters form a perfect reconstruction (PR) set by satisfying the conditions of part (c), then each pair of outputs above may be used to reconstruct the input to that pair of H filters, and so the whole process of the 2D wavelet transform may be easily inverted, to create a 2D inverse wavelet transform.

Compression artefacts from a wavelet-based image coding system tend to be less visible if the reconstruction lowpass filter (G_0) is as smooth as possible. The H_0 filter is less critical in this respect, although it should also be relatively smooth. Hence we should allocate factors of the product filter $H_0(z)G_0(z)$ so that $G_0(z)$ is as smooth as possible while still preserving the PR property of the system.

[20%]

This was the most popular question – and the candidates chose wisely, as it was also the question with the highest average. All parts were essentially bookwork and were done well by most. The majority of errors were as a result of not reading the question properly.

END OF PAPER