Solutions: 4F8 2011

ENGINEERING TRIPOS PART IIB

Thursday 12 May 2011 2.30 to 4.00

Module 4F8

IMAGE PROCESSING AND IMAGE CODING

1 (a) (i) 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 $W(\omega_1, \omega_2)$.

This is exactly as we should expect since we multiply in the spatial domain and must therefore convolve in the frequency domain.

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

A. Taking the product of 1d windows:

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

B. Rotating a 1d window:

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

[15%]

(ii) You can do this either by direct Fourier transforming, or by using the fact that the triangular pulse is the convolution of two rectangular pulses.

If doing it directly: first find the FT of w_1

$$W_1(\omega_1) = \int_{-U_1}^{U_1} \left(1 - \frac{|u_1|}{U_1}\right) \mathrm{e}^{-j\omega_1 u_1} du_1$$

$$= \int_{-U_1}^{0} \left(1 + \frac{u_1}{U_1}\right) \mathrm{e}^{-j\omega_1 u_1} du_1 + \int_{0}^{U_1} \left(1 - \frac{u_1}{U_1}\right) \mathrm{e}^{-j\omega_1 u_1} du_1$$

$$= \int_{-U_1}^{U_1} \mathrm{e}^{-j\omega_1 u_1} du_1 + \int_{0}^{U_1} -\frac{u_1}{U_1} \left[\mathrm{e}^{-j\omega_1 u_1} + \mathrm{e}^{j\omega_1 u_1}\right] du_1$$

Which can be rewritten as

(cont.

$$\left[\frac{e^{-j\omega_1 u_1}}{-j\omega_1}\right]_{-U_1}^{U_1} - \int_0^{U_1} \frac{u_1}{U_1} 2\cos(\omega_1 u_1) du_1$$

$$= 2U_1 \text{sinc}(\omega_1 U_1) - \frac{2}{U_1}\int_0^{U_1} u_1 \cos(\omega_1 u_1) du_1$$

Integrating the second term by parts then gives

$$\frac{2}{U_1}\int_0^{U_1} u_1 \cos(\omega_1 u_1) du_1 = \frac{2}{U_1}\left(\left[\frac{u_1 \sin(\omega_1 u_1)}{\omega_1}\right]_0^{U_1} - \int_0^{U_1} \frac{\sin(\omega_1 u_1)}{\omega_1} du_1\right)$$

$$= 2U_1 \text{sinc}(\omega_1 U_1) + \frac{2}{\omega_1{}^2 U_1}\left[\cos(\omega_1 U_1) - 1\right]$$

The whole integral is therefore

$$2U_1 \text{sinc}(\omega_1 U_1) - 2U_1 \text{sinc}(\omega_1 U_1) - \frac{2}{\omega_1{}^2 U_1}\left[\cos(\omega_1 U_1) - 1\right]$$

$$= \frac{2}{\omega_1{}^2 U_1} 2\sin^2(\omega_1 U_1/2) = U_1 \text{sinc}^2 \frac{\omega_1 U_1}{2}$$

.

Thus the total 2-D window function will be

$$W(\omega_1, \omega_2) = W(\omega_1) W(\omega_2) = U_1 U_2 \text{sinc}^2 \frac{\omega_1 U_1}{2} \text{sinc}^2 \frac{\omega_2 U_2}{2}$$

Note: that one can get the above result fairly easily by taking the standard result for the FT of a rectangular pulse (a sinc) and noting that since the triangular pulse is the convolution of two rectangular pulses, the FT of the triangular pulse must be the multiplication of the FT , ie a sinc^2 – a bit of care must be taken to get the correct factors. [30%]

(iii) Note that along the axes, the sinc^2 function above has its first zeros at $\frac{\omega_1 U_1}{2} = \pm\pi$ and $\frac{\omega_2 U_2}{2} = \pm\pi$, ie at $\omega_k = \pm 2\pi/U_k$, and subsequent zeros at multiples of these values. The mainlobe is therefore wider than a simple rectangular window function, but the sidelobes decay at a much more rapid rate.

Thus for a given ideal frequency response we know that the effect of windowing is to convolve with the FT of the window function – thus the

(TURN OVER for continuation of SOLUTION 1

freq response will be spread out more than for a rectangular window, but the sidelobes will be much better. This will therefore give a relatively wide transition band but a flat stop-band. [15%]

(b) (i) To find H we can use two approaches – either directly use the IFT or use a combination of standard results (probably the most common choice). Both approaches are given below:

Direct IFT

Can first of all do this via straightforward FTs. Firstly we can write the frequency response as

$$H(\omega_1, \omega_2) = H_0 - H_1 H_2$$

where

$$H_0(\omega_1, \omega_2) = \begin{cases} 1 & \text{if } |\omega_1| < \Omega_U \text{ and } |\omega_2| < \Omega_U \\ 0 & \text{otherwise} \end{cases}$$

$$H_1(\omega_1, \omega_2) = \begin{cases} 1 & \text{if } \Omega_L < |\omega_1| < \Omega_U \\ 0 & \text{otherwise} \end{cases}$$

$$H_2(\omega_1, \omega_2) = \begin{cases} 1 & \text{if } \Omega_L < |\omega_2| < \Omega_U \\ 0 & \text{otherwise} \end{cases}$$

Taking the IFT of $H(\omega_1, \omega_2)$ gives us

$$h(n_1, n_2) = \frac{\Delta_1 \Delta_2}{(2\pi)^2} \int_{-\pi/\Delta_2}^{\pi/\Delta_2} \int_{-\pi/\Delta_1}^{\pi/\Delta_1} [H_0 - H_1 H_2] e^{j(\omega_1 n_1 \Delta_1 + \omega_2 n_2 \Delta_2)} \, d\omega_1 \, d\omega_2$$

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

$$- \frac{\Delta_1 \Delta_2}{(2\pi)^2} \left[\int_{-\Omega_U}^{-\Omega_L} e^{j\omega_1 n_1 \Delta_1} d\omega_1 + \int_{\Omega_L}^{\Omega_U} e^{j\omega_1 n_1 \Delta_1} d\omega_1 \right] \times$$

$$\left[\int_{-\Omega_U}^{-\Omega_L} e^{j\omega_2 n_2 \Delta_2} d\omega_2 + \int_{\Omega_L}^{\Omega_U} e^{j\omega_2 n_2 \Delta_2} d\omega_2 \right]$$

Evaluating these integrals gives

(cont.

$$\frac{\Delta_1\Delta_2}{(2\pi)^2}\left\{\left[\frac{e^{j\omega_1 n_1\Delta_1}}{jn_1\Delta_1}\right]_{-\Omega_U}^{\Omega_U}\left[\frac{e^{j\omega_2 n_2\Delta_2}}{jn_2\Delta_2}\right]_{-\Omega_U}^{\Omega_U}\right\}$$

$$-\frac{\Delta_1\Delta_2}{(2\pi)^2}\left\{\left[\left[\frac{e^{j\omega_1 n_1\Delta_1}}{jn_1\Delta_1}\right]_{-\Omega_U}^{-\Omega_L}+\left[\frac{e^{j\omega_1 n_1\Delta_1}}{jn_1\Delta_1}\right]_{\Omega_L}^{\Omega_U}\right]\left[\left[\frac{e^{j\omega_2 n_2\Delta_2}}{jn_2\Delta_2}\right]_{-\Omega_U}^{-\Omega_L}+\left[\frac{e^{j\omega_2 n_2\Delta_2}}{jn_2\Delta_2}\right]_{\Omega_L}^{\Omega_U}\right]\right\}$$

$$=\frac{\Delta_1\Delta_2}{(2\pi)^2}\{2\Omega_U 2\Omega_U \text{sinc}(n_1\Delta_1\Omega_U)\text{sinc}(n_2\Delta_2\Omega_U)\}$$

$$-\frac{\Delta_1\Delta_2}{(2\pi)^2}[2\Omega_U\text{sinc}(n_1\Delta_1\Omega_U)-2\Omega_{L1}\text{sinc}(n_1\Delta_1\Omega_L)]\times$$

$$[2\Omega_U\text{sinc}(n_2\Delta_2\Omega_U)-2\Omega_L\text{sinc}(n_2\Delta_2\Omega_L)]$$

$$=\frac{\Delta_1\Delta_2}{(\pi)^2}\{\Omega_U\Omega_L\text{sinc}(n_1\Delta_1\Omega_U)\text{sinc}(n_2\Delta_2\Omega_L)+$$

$$\Omega_L\Omega_U\text{sinc}(n_1\Delta_1\Omega_L)\text{sinc}(n_2\Delta_2\Omega_U)-\Omega_L\Omega_L\text{sinc}(n_1\Delta_1\Omega_L)\text{sinc}(n_2\Delta_2\Omega_L)\}$$

Using combinations of standard results

It is also possible to arrive at the above by using the standard results for a rectangular lowpass and bandpass filters.

Standard result for a lowpass filter (H_0) is:

$$h(n_1\Delta_1,n_2\Delta_2)=\frac{\Delta_1\Delta_2}{\pi^2}[\Omega_U{}^2\,\text{sinc}(\Omega_U n_2\Delta_2)\,\text{sinc}(\Omega_U n_1\Delta_1)]$$

Standard result for a separable bandpass filter (H_1H_2) is

$$h(n_1\Delta_1,n_2\Delta_2)=$$

$$\frac{\Delta_1\Delta_2}{\pi^2}[\Omega_U\text{sinc}(\Omega_U n_1\Delta_1)-\Omega_L\text{sinc}(\Omega_L n_1\Delta_1)]\,[\Omega_U\text{sinc}(\Omega_U n_2\Delta_2)-\Omega_L\text{sinc}(\Omega_L n_2\Delta_2)]$$

Thus, as our filter can be formed from $H_0 - H_1H_2$, our impulse response is:

$$=\frac{\Delta_1\Delta_2}{\pi^2}\Big[\Omega_U{}^2\,\text{sinc}(\Omega_U n_2\Delta_2)\,\text{sinc}(\Omega_U n_1\Delta_1)-[\Omega_U\text{sinc}(\Omega_U n_1\Delta_1)-\Omega_L\text{sinc}(\Omega_L n_1\Delta_1)]\times$$

$$[\Omega_{U2}\text{sinc}(\Omega_U n_2\Delta_2)-\Omega_{L2}\text{sinc}(\Omega_L n_2\Delta_2)]]$$

which simplifies to give the same expression as above.

(TURN OVER for continuation of SOLUTION 1

As well as taking $(H_0 - H_1 H_2)$ we can also treat the shaded region as the sum of lowpass filters ($|\omega_1| < \Omega_U$ and $|\omega_2| < \Omega_L$), ($|\omega_1| < \Omega_L$ and $|\omega_2| < \Omega_U$) minus another lowpass filter ($|\omega_1| < \Omega_L$ and $|\omega_2| < \Omega_L$) etc. (giving the same form as the direct IFT).

From the above results we see that if $\Omega_U = \Omega_L$ our expression for h reduces to

$$\frac{\Delta_1 \Delta_2}{\pi^2}\left[{\Omega_U}^2 \mathrm{sinc}(\Omega_U n_1 \Delta_1)\mathrm{sinc}(\Omega_U n_2 \Delta_2)\right]$$

which is indeed the h of a square lowpass filter with side Ω_U.

[30%]

(ii) If we consider the value of h on the u_1 axis, where $n_2 = 0$, we see that the expression reduces to:

$$\frac{\Delta_1 \Delta_2}{\pi^2}\left[\Omega_L \Omega_U \mathrm{sinc}(\Omega_U n_1 \Delta_1) + \Omega_L(\Omega_U - \Omega_L)\mathrm{sinc}(\Omega_L n_1 \Delta_1)\right]$$

Similarly, along the u_2 axis we have:

$$\frac{\Delta_1 \Delta_2}{\pi^2}\left[\Omega_L \Omega_U \mathrm{sinc}(\Omega_U n_2 \Delta_2) + \Omega_L(\Omega_U - \Omega_L)\mathrm{sinc}(\Omega_L n_2 \Delta_2)\right]$$

Thus, along the axes we will get the sum of two sincs, which will give sinc-like behaviour.

However, if we look at what happens on the diagonals ($u_1 = u_2$), we see that we will get sinc^2 behaviour, so that the sidelobes are smaller and decay more rapidly. The sketch below indicates the behaviour of the impulse response.

[10%]

2 (a) Our observed image, $\mathbf{y}$, is modelled as a linear distortion, L, of the true image, $\mathbf{x}$, plus additive noise, $\mathbf{d}$, i.e. $\mathbf{y} = L\mathbf{x} + \mathbf{d}$.

(i) If we have access to the imaging system and to a range of sources to image, we can image something resembling (as much as possible) a point source. The resulting image can then be taken as our estimate of L (also known as the *point spread function* – this of course neglects the noise, but nevertheless is used. In microscopy, special *point source beads* are used to estimate the point spread function (L). [10%]

(cont.

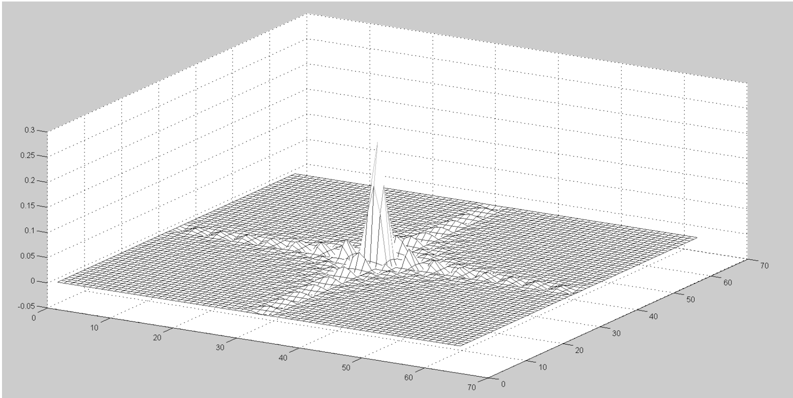


Fig. 1

(ii) If we neglect noise we can write $\mathbf{y} = L\mathbf{x} + \mathbf{d}$ in discrete form as

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

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:

$$Y(\omega_1, \omega_2) = \mathscr{L}(\omega_1, \omega_2)X(\omega_1, \omega_2)$$

where:

$$\mathscr{L}(\omega_1, \omega_2) = \sum_{n_2=-\infty}^{\infty}\sum_{n_1=-\infty}^{\infty} L(n_1, n_2)e^{-j(\omega_1 n_1 + \omega_2 n_2)}$$

$$\therefore X(\omega_1, \omega_2) = \frac{Y(\omega_1, \omega_2)}{\mathscr{L}(\omega_1, \omega_2)}$$

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, L, 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 L – the inverse filter is therefore $1/\mathscr{L}$.

(TURN OVER for continuation of SOLUTION 2

If the transfer function $\mathscr{L}(\omega_1,\omega_2)$ has zeros then the inverse filter, $1/\mathscr{L}$, will have infinite gain. i.e. when $\mathscr{L}(\omega_1,\omega_2)$ is very small, $1/\mathscr{L}(\omega_1,\omega_2)$ is very large (or indeed infinite if there are zeros) and therefore, small noise in the regions of the frequency plane where $1/\mathscr{L}(\omega_1,\omega_2)$ is very large, can be hugely amplified. In practice a method of lessening this sensitivity to noise is to threshold the frequency response, leading to the so-called, pseudo-inverse or generalised inverse filter $\mathscr{L}_g(\omega_1,\omega_2)$. This is given by

$$\mathscr{L}_g(\omega_1,\omega_2) = \begin{cases} \frac{1}{\mathscr{L}(\omega_1,\omega_2)} & \frac{1}{|\mathscr{L}(\omega_1,\omega_2|} < \gamma \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

or

$$\mathscr{L}_g(\omega_1,\omega_2) = \begin{cases} \frac{1}{\mathscr{L}(\omega_1,\omega_2)} & \frac{1}{|\mathscr{L}(\omega_1,\omega_2|} < \gamma \\ \gamma\frac{|\mathscr{L}(\omega_1,\omega_2|}{\mathscr{L}(\omega_1,\omega_2)} & \text{otherwise} \end{cases} \tag{2}$$

Clearly for $\frac{1}{|\mathscr{L}(\omega_1,\omega_2|} \geq \gamma$ in equation 2, the modulus of the filter is set as γ, whereas in equation 1 it is set as 0. [20%]

(iii) In the Bayesian derivation of the Wiener filter we assume, for simplicity, that $E[\mathbf{x}] = 0$ and $E[\mathbf{d}] = 0$, i.e. that both the signal and the noise are zero mean. To find an estimate of $\mathbf{x}$, we maximise $P(\mathbf{x}|\mathbf{y})$, i.e. the *probability* of the original image *given* the observed data. We form this posterior vis Bayes which tells us that $P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$, with

$$P(\mathbf{y}|\mathbf{x}) \propto \mathrm{e}^{-\frac{1}{2}\mathbf{d}^T N^{-1}\mathbf{d}} = \mathrm{e}^{-\frac{1}{2}(\mathbf{y}-L\mathbf{x})^T N^{-1}(\mathbf{y}-L\mathbf{x})}$$

Here we have assumed that the noise is gaussian distributed with covariance matrix $N = E[\mathbf{d}\mathbf{d}^T]$ so that the $\mathbf{d}^T N^{-1}\mathbf{d}$ term is the vector equivalent of the $\frac{1}{\sigma^2}$ term in the 1d gaussian – if N is diagonal then N^{-1} will be diagonal with elements $\frac{1}{\sigma_i{}^2}$.

The *prior* probability $P(\mathbf{x})$ incorporates any *prior* knowledge we may have about the distribution of the data and we assume an ideal world in which $\mathbf{x}$ is a gaussian random variable, described by a *known* covariance matrix $C = E[\mathbf{x}\mathbf{x}^T]$ (including all cross-correlations etc.) so that

$$P(\mathbf{x}) \propto \mathrm{e}^{-\frac{1}{2}\mathbf{x}^T C^{-1}\mathbf{x}}$$

(cont.

An alternative image deconvolution algorithm is known as *Maximum Entropy* deconvolution, again assuming Gaussian noise, the likelihood would be as for the Wiener filter above but the prior is an *entropic* prior which takes the form

$$P(\mathbf{x}) \propto \mathrm{e}^{\alpha S}$$

where one version of the *entropy* S (sometimes known as the *cross entropy*) of the image is given by

$$S(\mathbf{x}, \mathbf{m}) = \sum_i \left[x_i - m_i - x_i \ln\left(\frac{x_i}{m_i}\right)\right]$$

where $\mathbf{m}$ is the *measure* on an image space (*the model*) to which the image $\mathbf{x}$ defaults in the absence of data. (Can see global maximum of S occurs at $\mathbf{x} = \mathbf{m}$.)

Or – another alternative prior is the Pixon prior. This is harder to describe as no real detail was given in the notes! However, marks will indeed be given if anybody has read up on this and has managed to adequately describe the prior used (in terms of assuming a distribution for the sizes of the pixons).

[30%]

(b) Suppose the original image has size a_1 in the u_1 direction and size a_2 in the u_2 direction. The sampling intervals in the u_1 and u_2 directions are therefore:

$$\Delta_1 = \frac{a_1}{m} \quad \Delta_2 = \frac{a_2}{n}$$

We know that the FT of a sampled image is the FT of the continuous image repeated at intervals (in 2-D space) of the sampling frequencies. So, for our $m \times n$ sampled image, the spectrum is repeated at intervals of

$$\Omega_{1s} = \frac{2\pi}{\Delta_1} = \frac{2\pi m}{a_1} \quad \text{and} \quad \Omega_{2s} = \frac{2\pi}{\Delta_2} = \frac{2\pi n}{a_2}$$

Now, suppose we downsample by a factor of d_1 in the u_1 direction and by a factor of d_2 in the u_2 direction, the new sampling intervals are then

$$\Delta_1' = \frac{a_1}{(m/d_1)} \quad \Delta_2' = \frac{a_2}{(n/d_2)}$$

so that the FT of this newly sampled image is repeated at sampling intervals of

(TURN OVER for continuation of SOLUTION 2

$$\Omega'_{1s} = \frac{2\pi m}{a_1 d_1} = \frac{2\pi}{\Delta_1 d_1} \quad \text{and} \quad \Omega'_{2s} = \frac{2\pi n}{a_2 d_2} = \frac{2\pi}{\Delta_2 d_2}$$

Thus, for no aliasing in the resampled image we need

$$\Omega'_{1s} > 2\Omega_1 \quad \text{and} \quad \Omega'_{2s} > 2\Omega_2$$

Which means that

$$d_1 < \frac{\pi m}{a_1 \Omega_1} \equiv \frac{\pi}{\Delta_1 \Omega_1} \quad \text{and} \quad d_2 < \frac{\pi n}{a_2 \Omega_2} \equiv \frac{\pi}{\Delta_2 \Omega_2}$$

So, we obtain the minimum size of the image (to avoid aliasing) as

$$x_1 = m/d_1 = \frac{a_1 \Omega_1}{\pi} \quad \text{and} \quad x_2 = n/d_2 = \frac{a_2 \Omega_2}{\pi}$$

Since we need integer values, (m_{min}, n_{min}) are given by rounding up x_1 and x_2. [30%]

If we do not sample sufficiently frequently, we will have aliasing which will occur via frequencies from the repeated spectrum falling into the 'main' spectrum. In many cases we will get aliased frequencies which will then lie very close to each other in the frequency domain. With two close frequencies, the effect will be to produce artefacts at the sum and difference of the two frequencies – it is generally the difference frequency which will then manifest itself as ringing/beating (ie moire fringe effects) artefacts in the aliased image.

[10%]

3 (a) Consider an $(n \times n)$ block of image pixels denoted by the matrix X. If we multiply X on the left by the DCT matrix T, we produce another $(n \times n)$ matrix, Y

$$Y = TX$$

It is easy to see that the ijth element of Y, Y_{ij} is given by

$$Y_{ij} = \mathbf{t}_i \cdot \mathbf{x}_j$$

where $\mathbf{t}_i$ is the ith **row** of T and $\mathbf{x}_j$ is the jth **column** of X. We can see therefore that if $\mathbf{y}_i$ is the ith column of Y then

(cont.

$$T\mathbf{x}_i = \mathbf{y}_i$$

is the 1D DCT of $\mathbf{x}_i$ $(\mathbf{y}_i)_j = \mathbf{t}_j \cdot \mathbf{x}_i$).

Thus, the columns of Y give the 1D DCTs of the columns of X.

Similarly, we can see that if we take $W = XT^T$, then the ijth element of W is $\mathbf{x'}_i \cdot \mathbf{t}_j$ where $\mathbf{x'}_i$ is the ith **row** of X. Thus, the rows of W give the 1-D DCT of the rows of X.

It is therefore clear that TXT^T will first form the 1-D DCT of the columns of X, and then take the 1-D DCT of the rows of TX – thus performing a 2-D DCT transform (operation is separable). [20%]

END OF SOLUTIONS

3(b) (cont 2)

For $n = 8$:

No. of multiplies for $\underline{y} = T\underline{x}$

is $8 \times 8 = 64$

and no. of adds is $8 \times 7 = 56$

Whereas ~~using the~~ splitting this into two $\frac{n}{2}$ transforms gives:

No of multiplies $= 2 \times 4 \times 4 = 32$

No of ~~a~~adds $= 2 \times 4 \times 3 + \underline{2 \times 4} = 32$

↑ to form $\underline{u}$ & $\underline{v}$

If we do the same trick on one of the ~~$\frac{n}{2}$~~ 4×4 matrices above one of the 4×4 multiplies is replaced by $2 \times 2 \times 2 = 8$

and one of the 4×3 adds is replaced by $2 \times 2 \times 1 + 2 \times 2 = 8$

~~8 & 4 a~~ saving $16 - 8 = 8$ mults and $12 - 8 = 4$ adds.

~~Again~~ Hence the totals are $32 - 8 =$ <u>24 mults</u>

and $32 - 4 =$ <u>28 adds</u>

(2 further mults can be saved as the 2×2 DCT needs only 2 mults).

For an 8×8 2D DCT we need to transforms 8 separate columns of X first and then 8 rows of the result, so the totals are: $(8+8) \times 24 = 384$ mults

and $(8+8) \times 28 = 448$ adds

This compares with $(8+8) \times 64 = 1024$ mults & $(8+8) \times 56 = 896$ adds for the direct method.

3 (c) No. of bits to code subband i,j

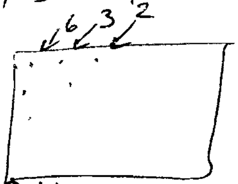
$$= H_{i,j} \cdot (128 \times 96) = H_{i,j} \cdot 12K \qquad (1K = 1024)$$

There is 1 band with entropy $H_{1,1} = \frac{6}{2-1} = 6$ bits/coef

2 bands " " $H_{1,2} = \frac{6}{3-1} = 3$ bits/ "

3 " " " $H_{1,3} = \frac{6}{4-1} = 2$ bits/ "

4 " " " $H_{1,4} = \frac{6}{5-1} = 1.5$ bits/ "



etc.

Hence the total bits for the image $= \left(\sum_{i=1}^{8} \sum_{j=1}^{8} H_{i,j} \right) 12K$

$$= 12K \Big(6.1 + 3.2 + 2.3 + 1.5.4 + \tfrac{6}{5}.5 + \tfrac{6}{6}.6 + \tfrac{6}{7}.7 + \tfrac{6}{8}.8 + \tfrac{6}{9}7 + \tfrac{6}{10}.6 + \tfrac{6}{11}.5 + \tfrac{6}{12}.4 + \tfrac{6}{13}.3 + \tfrac{6}{14}.2 + \tfrac{6}{15}.1 \Big)$$

$$= 12K \left(6 \times 8 + 6 \left(\tfrac{7}{9} + \tfrac{6}{10} + \tfrac{5}{11} + \tfrac{4}{12} + \tfrac{3}{13} + \tfrac{2}{14} + \tfrac{1}{15} \right) \right)$$

$$= 12K \, (48 + 6 \, . \, 2.6059) = 12K \, . \, 63.6357$$

$$= \underline{\underline{781{,}956 \text{ bits}}}$$

In practice it would be a little larger than this, say about 800,000 bits.

3(d) The case of $n=2$ is the simple Haar transform, which is known to ~~be~~ give much worse compression than the $n=8$ DCT, especially if only used at one level (ie not in multi-level form like a wavelet transform). However $n=2$ is very simple to compute.

The case of $n=4$ gives better compression than $n=2$, but not as good as $n=8$.

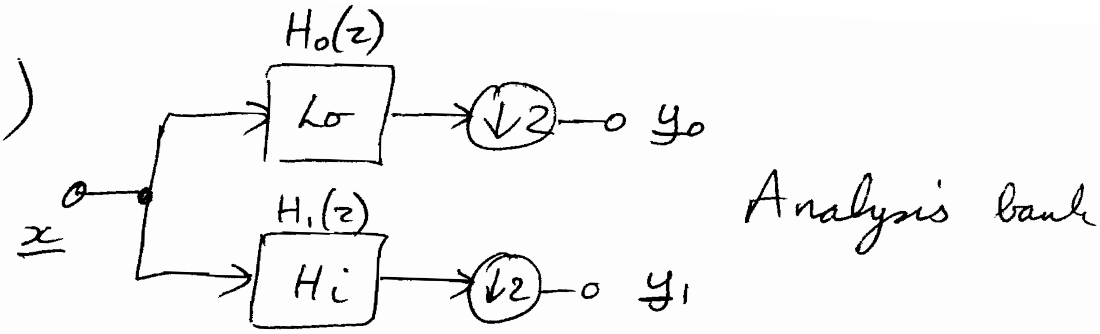
The case of $n=16$ ~~to~~ gives slightly lower entropy than $n=8$, <u>but</u> the artifacts are bigger ~~so the~~ and therefore more visible.

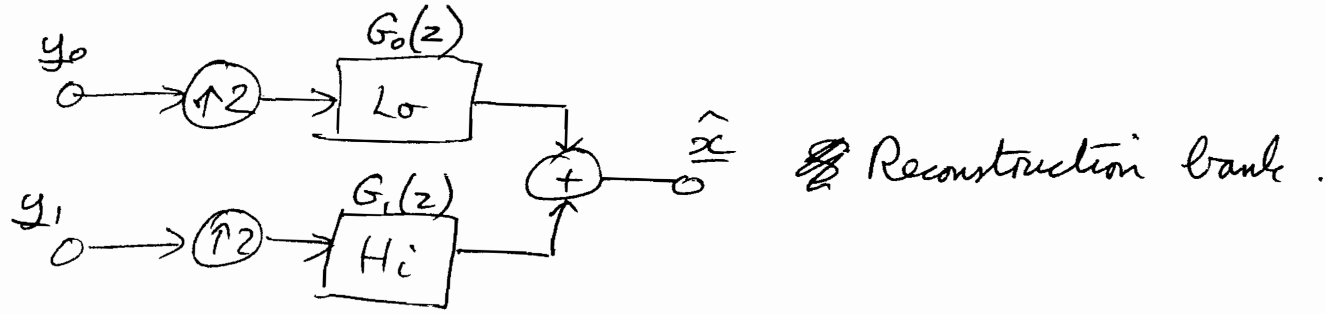
Hence $n=8$ turns out to be the optimum DCT size for most purposes and is the size used in JPEG.

Computation goes up with size n, but not dramatically so because of the efficiency of the factorization strategy in part (b), and the fact that the number of blocks $\propto \frac{1}{n^2}$

This question was attempted by 8 IIB candidates and 1 graduate. Parts a) and d) were answered well by almost all candidates (mostly bookwork). Parts b) and c) were the source of most lost marks – candidates generally knew what was required but some were unable to recall the details.

4. (a)

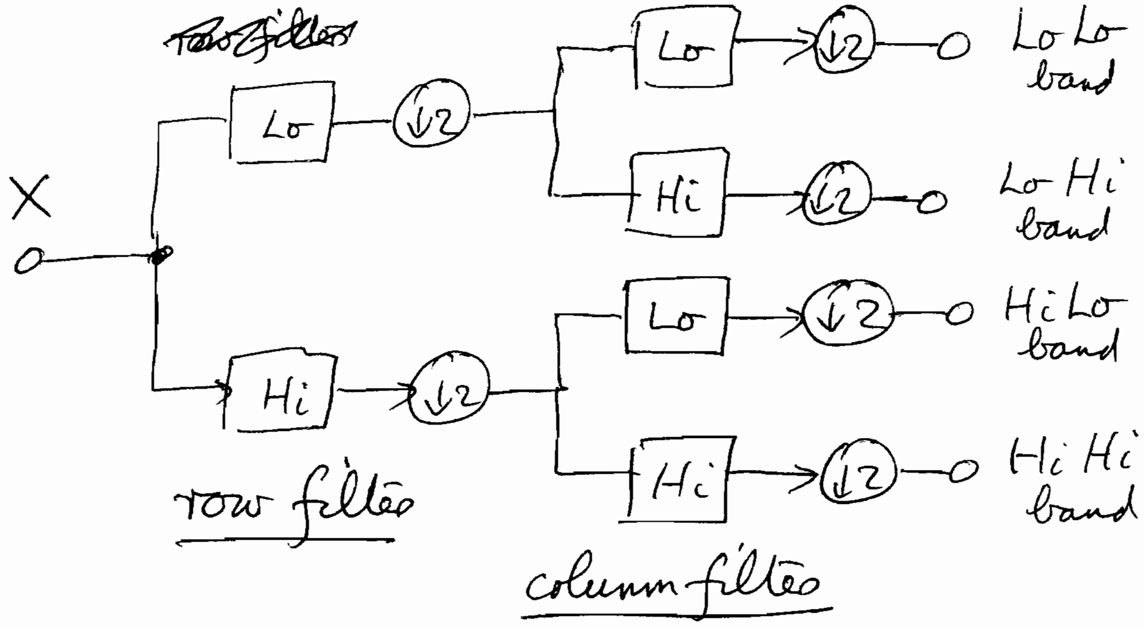




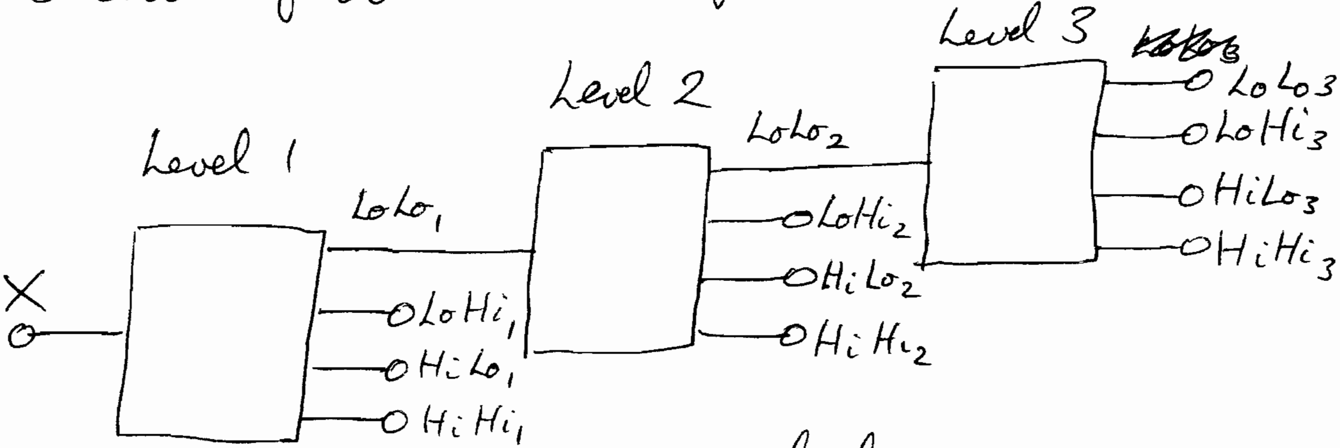
The perfect reconstruction requirement is that the reconstructed vector $\hat{\underline{x}}$ should be identical to the input vector $\underline{x}$, if y_0 & y_1 are fed unaltered from the analysis bank to the reconstruction bank.

This is important in an image coding system, so that the only source of distortion is due to the quantisation of transform coefficients.

4(b) One level of 2-D filterbank:

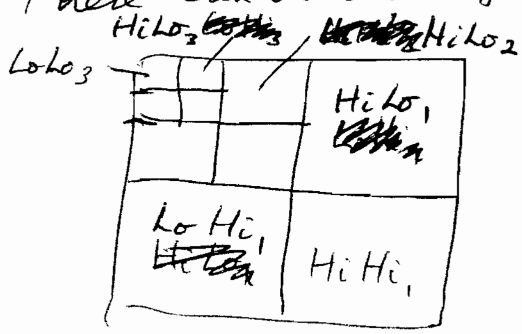


3 levels of the above 2-D filterbank would be

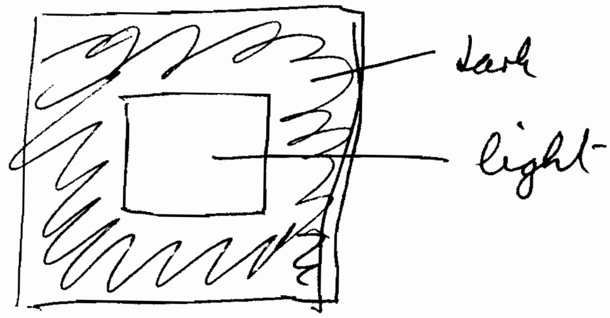


There are 3 subbands at each of levels 1 and 2 and 4 at level 3 giving 10 subbands in all.

These can be arranged in a big matrix as follows



4 (c) Image:

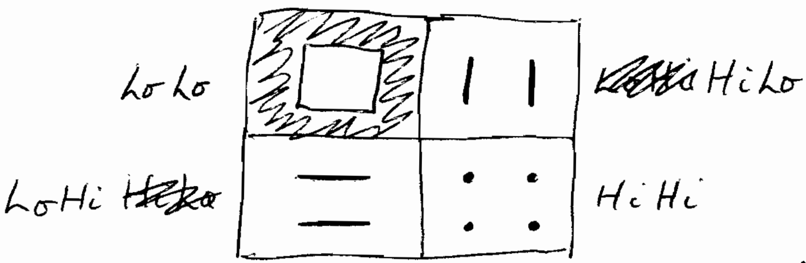


Highpass filters tend to detect edges & ignore smooth regions of the data.

Lowpass filters tend to ignore fine detail and respond to smooth regions.

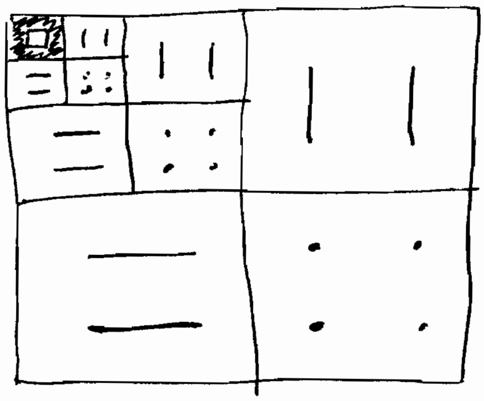
Hence a highpass filter along the rows will tend to detect vertical edge structures, and a highpass down the columns will detect horizontal edge structures.

A highpass filter in both directions will only respond to corners of the square. Hence the ~~3~~4 subbands at <u>level 1</u> will look like



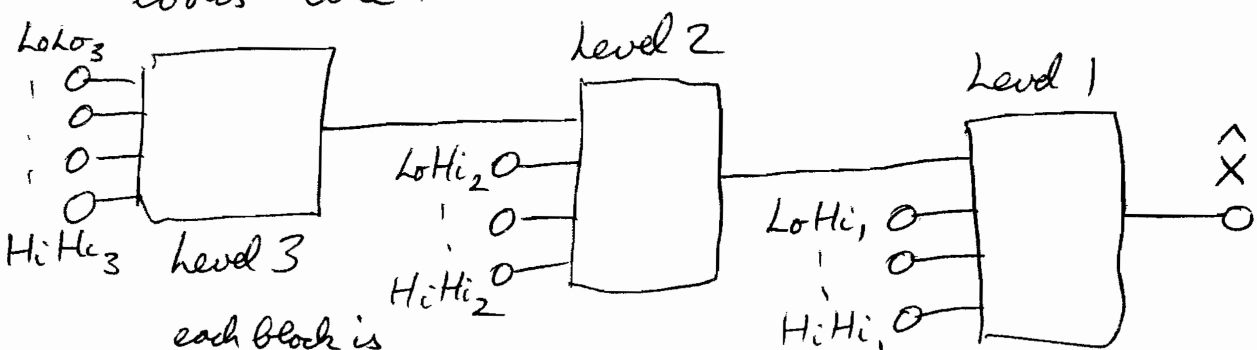
The LoLo band image is just a smaller version of the original, so the same effects will occur in the subband at levels 2 & 3 and the set of 10 subbands will be:

4(c) (cont.)

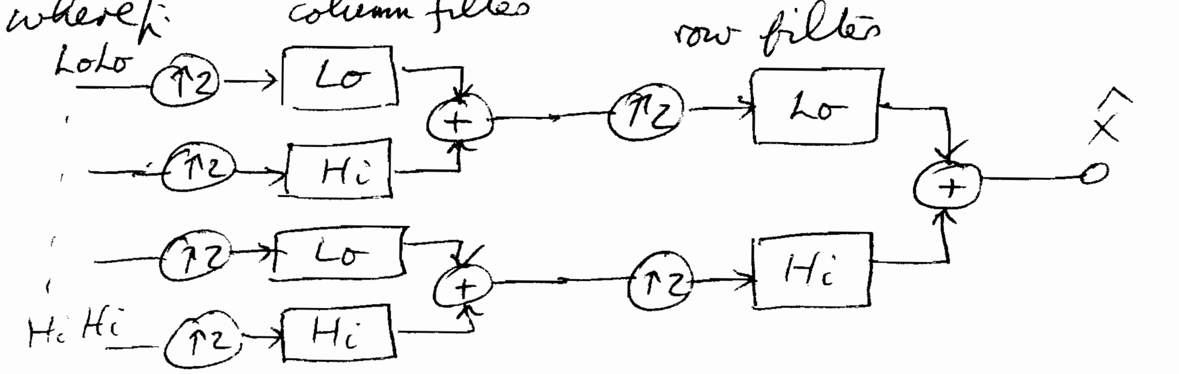


Note that here we plot the ~~mag~~ magnitudes of the wavelet coefs, such that dark lines/points indicate large amplitude coefs, while white paper indicates low amplitude coefs.

4(d) The reconstruction filter bank is just the reverse of the analysis bank. Hence it looks like:



each block is where:



4 (d) (cont.)

Since the reconstruction filters in part (a) can be designed to perfectly reconstruct $\hat{\underline{x}}$ to be identical to $\underline{x}$, the above 2-D reconstruction block will invert the operations of the 2-D analysis block in part (b) at each level. Hence the whole 3-level filter bank above will invert the 10 bands of the 3-level analysis bank in part (b).

Data compression tends to suppress or heavily quantise the higher-frequency wavelet ~~co~~ coefficients, so it is very important that the image that is reconstructed just from the lower frequency coefficients is relatively smooth and free of high frequency or blocking artifacts. Hence the impulse response of the lowpass <u>reconstruction</u> filter, $G_0(z)$ in part (a), should be as smooth as possible, while still preserving the perfect reconstruction of the group of filters H_0, H_1, G_0 and G_1.

This question was attempted by 8 IIB candidates and 1 graduate. It was well done by all who attempted it and the high average mark reflects the fact that it was mainly bookwork. All parts were answered competently, with perhaps (c) causing a bit of confusion with some candidates. This question was the easiest on the paper, but the fact that only 8 IIB candidates did it reflects that one needed to have revised well to answer it.