XEE4C081

UNIVERSITY OF DUBLIN

TRINITY COLLEGE

FACULTY OF ENGINEERING, MATHEMATICS & SCIENCE

SCHOOL OF ENGINEERING

Electronic and Electrical Engineering

Senior Sophister **Trinity Term, 2010**

Engineering

Annual Examinations

DIGITAL MEDIA PROCESSING (4C8)

Date: 11th May 2010 **Venue: GOLDHALL** **Time: 14:00-16:00**

Prof. A. C. Kokaram

Answer question 1 and any THREE (3) of the remaining FOUR (4) questions. Please answer questions from each section in separate answer books.

Permitted Materials:

- **Calculator**
- **Drawing Instruments**
- **Mathematical Tables**
- **Graph Paper**

COMPULSORY

1. The image $f[h,k]$, shown in figure 1 (top left) is to be processed with an enhancement operation designed to improve the sharpness of the high frequency components of the image.

(a) The Matlab code below shows the design of a 1-D low pass filter using a Gaussian window. `x = -50:50;var = 4.5*4.5; g = exp(-x.*x/(2*var)); g =g/sum(g);` Why is it necessary to make the d.c. gain of this filter equal to unity and in which line is this achieved in the code above? **[5 marks]**

(b) Figure 1 shows on the top right, the frequency reponse of two 1-D low-pass filters, both having 101 taps. The Gaussian filter $g[k]$, has a frequency response shown by the solid line, while the near-ideal low pass filter $w[k]$ has the frequency response shown by the dashed line. Both of these filters are used to create the respective 2D seperable filter $g'[h,k]$ and $w'[h,k]$. For instance $g'[h,k]$ is created using $g[k]$. The image $f[h,k]$ is then processed with these two seperable filters to yield two output images $f_g[h,k]$ and $f_w[h,k]$ respectively. The output images are shown in Figure 1 in no particular order.

Match the output images with the correct filter that produced them, giving your reasons for your selection. **[10 marks]**

(c) Figure 2 shows an image edge enhancement system that boosts the high frequency content of an image by a factor of 4. The input to the system is $f[h,k]$ shown previously. The system diagram (Fig 2) shows a number of output signals labelled as A, B. A is an intermediate output while B is the final image output. The key component of the system is the low pass filter indicated by **LPF**. Figure 3 shows a number of possible output images when $g'[h,k]$ or $w'[h,k]$ are used as the **LPF** stage. Match A, B to the correct image from this set when i) $g'[h,k]$ is used as the **LPF** and ii) $w'[h,k]$ is used as the **LPF**. Explain your choices ensuring to relate what you see to the frequency frequency reponse of the respective filters. **[10 marks]**

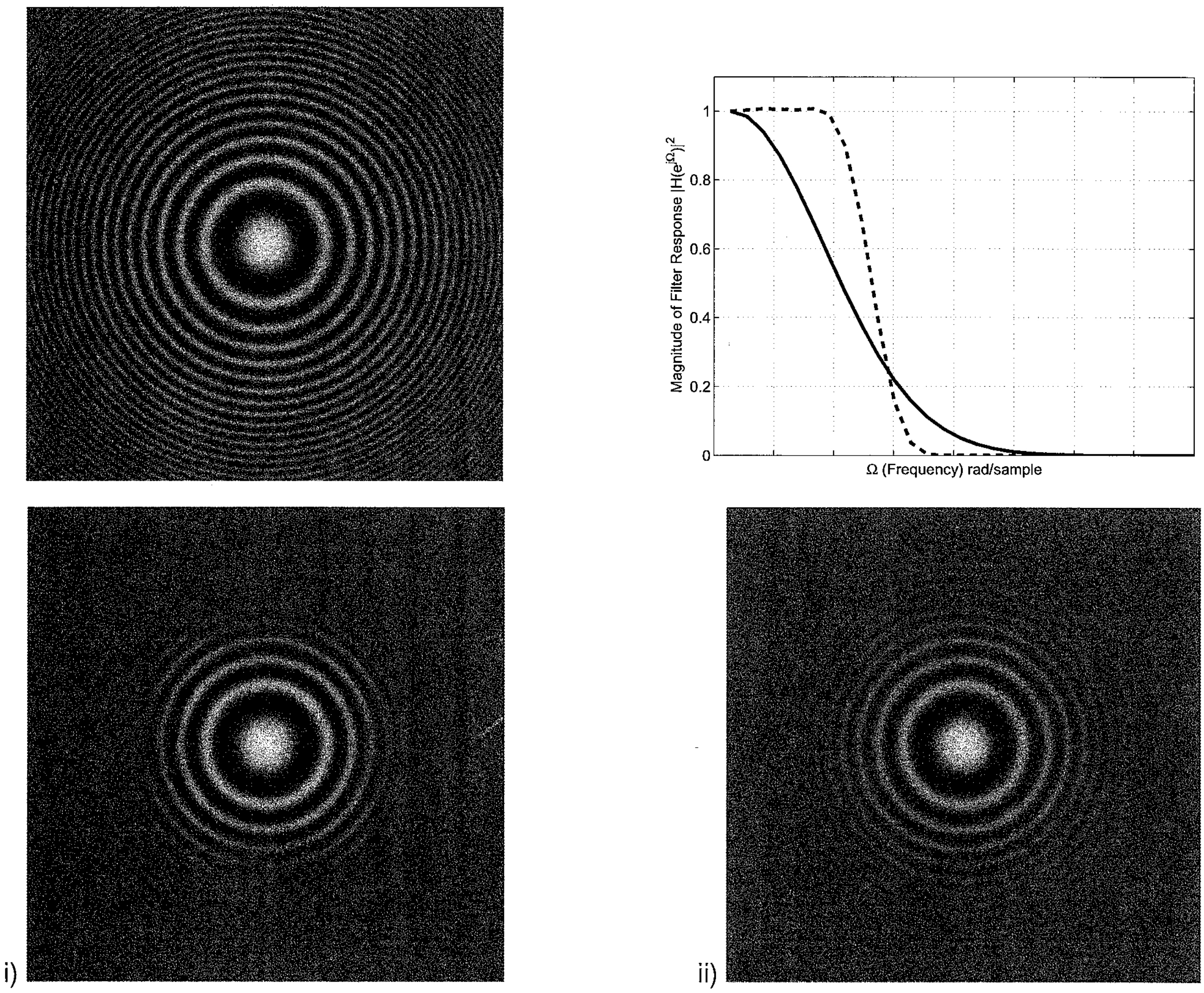

Figure 1: *Top Left: Original Zoneplate Image (mid-gray = 0 intensity), Top Right: Frequency response of 1-D filters. Bottom Row: Low pass filter outputs (mid-gray = 0 intensity).*

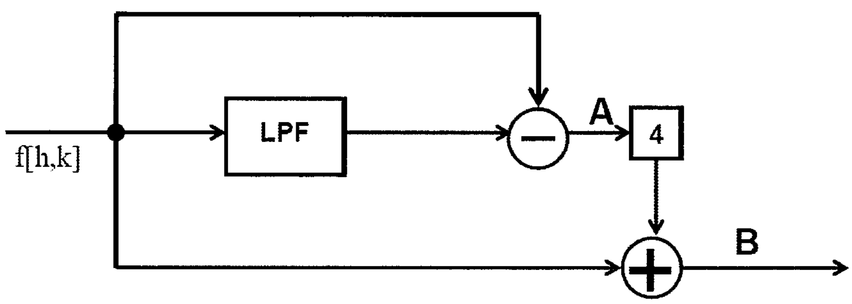

Figure 2: *Enhancement System*

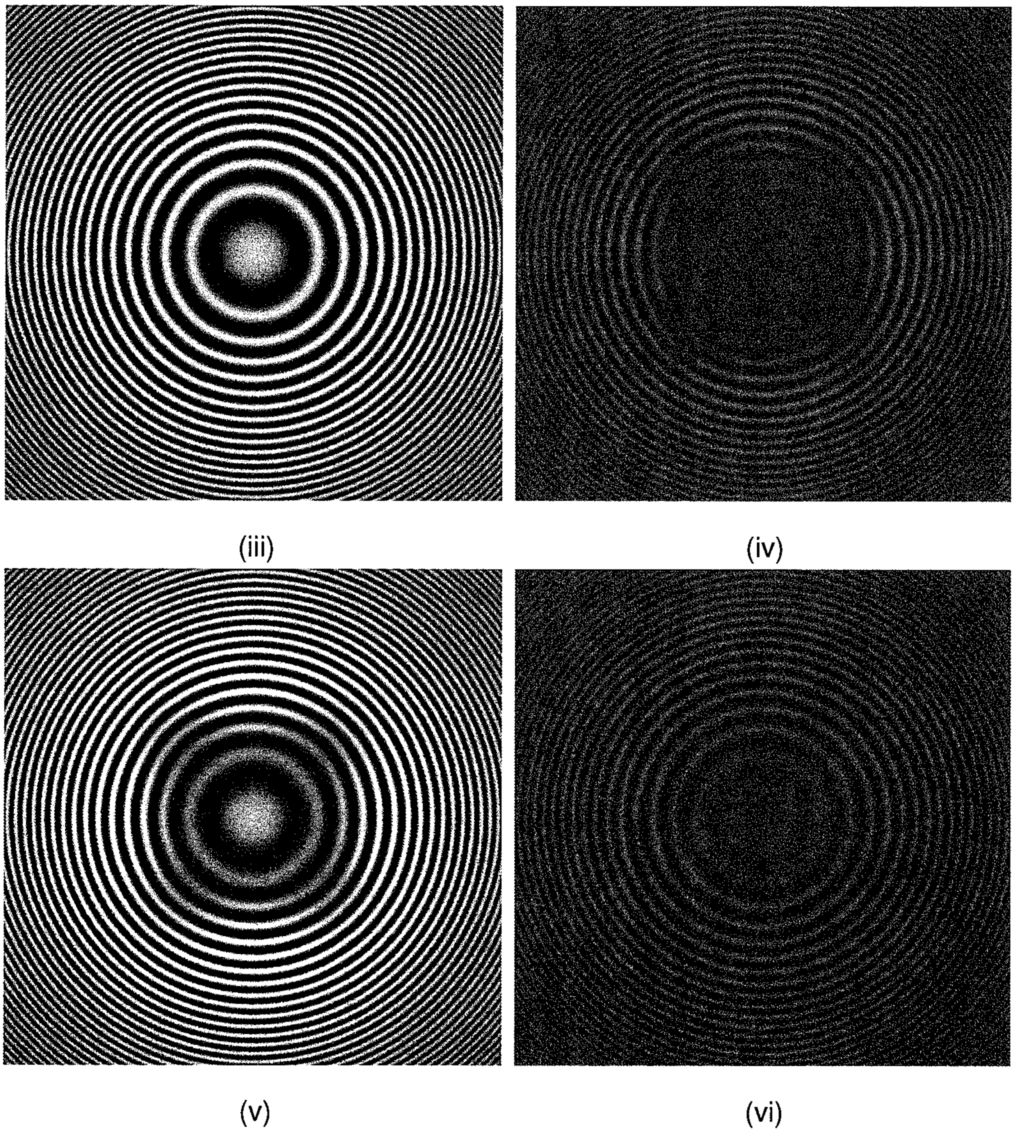

Figure 3: *Intermediate and Final outputs from the enhancement system. (mid-gray = 0 intensity)*

ANSWER THREE (3) QUESTIONS OUT OF THE REMAINING FOUR (4).

2. (a) Write down the image sequence model that is assumed implictly in the process of Full Search Block Matching for motion estimation. What do the terms *block size*, and *search width* refer to in that process? **[5 marks]**

(b) A pixel based motion detector is used to detect motion in a real video signal, $G_n(\mathbf{x})$. It thresholds the absolute value of the inter-frame pixel difference $\Delta_n(\mathbf{x}) = G_n(\mathbf{x}) - G_{n-1}(\mathbf{x})$ as follows

$$b_n(\mathbf{x}) = \begin{cases} 1 & \text{If } |\Delta_n(\mathbf{x})| > T_\Delta \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

where $b_n(\mathbf{x})$ is the motion detector output and is set to a value of 1 (one) at each pixel that motion is detected. In a real video signal, it is possible to model the p.d.f. of $|\Delta_n(\mathbf{x})|$ as a mixture of a Gaussian p.d.f., where the picture is not moving, and a Laplacian p.d.f where it is moving.

Assume that the probability distribution of the pixel difference $\Delta_n(\mathbf{x})$ in moving areas is Laplacian with $x_0 = 6.5$ as follows.

$$p(\Delta_n(\mathbf{x})) = \frac{1}{2x_0} \exp\left(\frac{-|\Delta_n(\mathbf{x})|}{x_0}\right) \tag{2}$$

Therefore the probability that a pixel is incorrectly classified as stationary when it is moving (i.e. $b_n(\mathbf{x}) = 0$) in these areas is $p_m(T_\Delta) = 1 - \exp(-T_\Delta/x_0)$.

i. Assume that the probability distribution of the pixel difference $\Delta_n(\mathbf{x})$ in stationary areas is Gaussian with variance $\sigma_v^2 = 100$ and zero mean. Show that the probability that a pixel is misclassified as moving (i.e. $b_n(\mathbf{x}) = 1$) in stationary areas is $p_n(T_\Delta) = \text{erfc}(T_\Delta/\sqrt{2\sigma_v^2})$. See Table 1 for the required integral. **[5 marks]**

ii. Over the range $T_\Delta = [3:9]$, plot on the same graph, curves of $p_n(T_\Delta)$, $p_m(T_\Delta)$ and hence choose an optimum threshold for correct classification of moving regions.
Explain your choice. **[10 marks]**

iii. A 1D $[1,\ 1,\ 1]/3$ filter is used in a separable implementation to create a 2D low pas filter H. This low pass filter is used to pre-process the pixel difference $\Delta_n(\mathbf{x})$. Calculate the reduction in the variance of Δ_n in stationary areas and, assuming there is no significant effect in moving areas, estimate the new optimal threshold that should be used.

[5 marks]

3. Figure 4 (below) shows the filter structure used in the analysis and reconstruction units of a 1D Perfect Reconstruction filterbank.

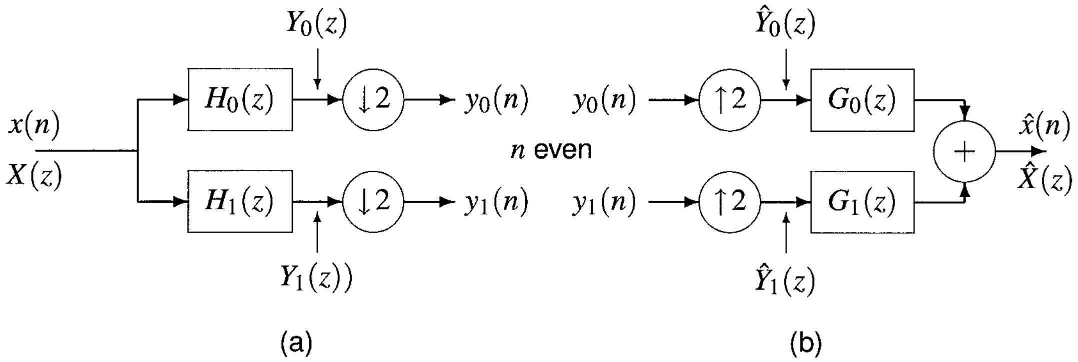

Figure 4: *Two-band filter banks for analysis (a) and reconstruction (b).*

(a) Using the Haar Wavelet, the analysis filters in the system above are as follows.

$$H_0(z) = \frac{1}{\sqrt{2}}\left(z^{-1}+1\right)$$
$$H_1(z) = \frac{1}{\sqrt{2}}\left(z^{-1}-1\right)$$

A signal $x[k] = 1,\ 1,\ 0,\ 0,\ 1,\ 1,\ 0,\ 0$ (for $k = 0:7$, 0 otherwsie) is input to this Haar filter bank. Calculate the output signals $y_0[n]$ and $y_1[n]$. **[4 marks]**

(b) The Haar filterbank in the configuration above creates a fully decimated transform and is *not* shift invariant. Using your signal output from above, explain the meaning of the terms *fully decimated* and *shift invariant.* Also explain why shift invariance is important in image manipulation? **[10 marks]**.

(c) Draw a block diagram showing how the analysis stage above can be applied to create a 2D analysis filter bank resulting in 2 levels of decomposition. **[6 marks]**

(d) Figure 5 shows a single grey scale image. Figure 6 shows a series of bandpass images which correspond to the 2D DWT of the image to the left of Figure 5 using the Haar wavelet at level 1 (the low pass sub-band is omitted). The brightness of the images has been adjusted for easier viewing, the mid-gray scale level corresponds to a value of 0 (zero). Identify the corresponding Hi-Lo, Lo-Hi and Hi-Hi sub-bands at Level 1 of the 2D DWT from the bandpass image set shown in Figure 5. Use the typical convention shown for arrangement of 2D DWT sub-bands as indicated in Figure 5. Explain your selections.

[5 marks]

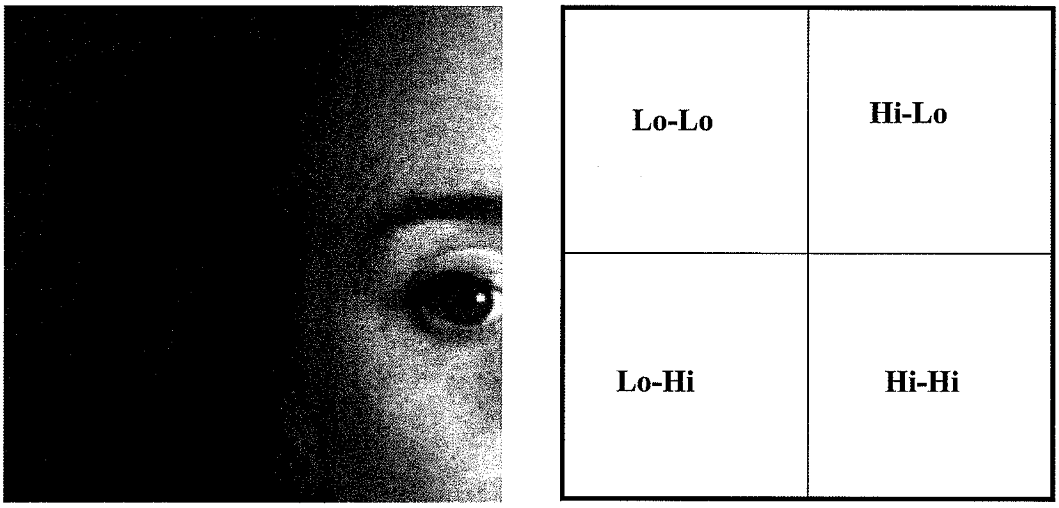

Figure 5: *Left : A Grey Scale Image. Right : The convention used for ordering the Level 1 sub-bands of the 2D DWT*

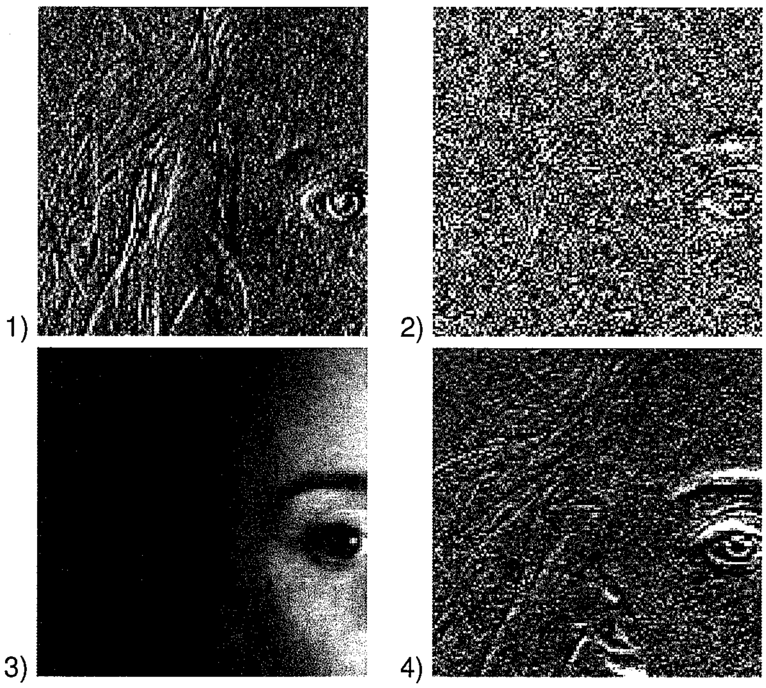

Figure 6: *Level 1 Subbands of the 2D DWT using the Haar Transform.*

4. (a) The 2-point, one-dimensional Haar transform of $[x(1),\ x(2)]$ is given by the following relationships

$$\begin{bmatrix} y(1) \\ y(2) \end{bmatrix} = \mathbf{T} \begin{bmatrix} x(1) \\ x(2) \end{bmatrix} \qquad \text{Where } \mathbf{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3}$$

Derive an expression for the 4-point transformation matrix that is equivalent to 2 levels of the Haar Transform. **[5 marks]**

(b) The DCT is used for almost all the core compression techniques in media transmission today. The DCT coefficient matrix for transforming a single 8 pixel row into 8 coefficients of the DCT is shown below. Since it is such an important transformation, it is typically required to be as efficient an implemementation as possible in hardware or software. Show how the number of multiply/add operations required to perform the 8 point DCT can be reduced by exploiting the structure in the matrix below. Estimate the reduction in operation count that you achieve. **[10 marks]**

$$\begin{bmatrix}
0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 \\
0.4904 & 0.4157 & 0.2778 & 0.0975 & -0.0975 & -0.2778 & -0.4157 & -0.4904 \\
0.4619 & 0.1913 & -0.1913 & -0.4619 & -0.4619 & -0.1913 & 0.1913 & 0.4619 \\
0.4157 & -0.0975 & -0.4904 & -0.2778 & 0.2778 & 0.4904 & 0.0975 & -0.4157 \\
0.3536 & -0.3536 & -0.3536 & 0.3536 & 0.3536 & -0.3536 & -0.3536 & 0.3536 \\
0.2778 & -0.4904 & 0.0975 & 0.4157 & -0.4157 & -0.0975 & 0.4904 & -0.2778 \\
0.1913 & -0.4619 & 0.4619 & -0.1913 & -0.1913 & 0.4619 & -0.4619 & 0.1913 \\
0.0975 & -0.2778 & 0.4157 & -0.4904 & 0.4904 & -0.4157 & 0.2778 & -0.0975
\end{bmatrix}$$

(c) A mobile phone manufacturer has a choice of either the Haar Transform or the DCT in its propreitary software codec to compress images of size 480×640 streamed at 20 frames per second i.e. as individual compressed frames.

i. Estimate the number of operations per second needed to perform each transform and hence recommend the best filter to use in terms of least computatinal load. You may assume that 1) 8×8 blocks are used in the compression process 2) The optimal operation count for the Haar transform is best calculated by considering the Transform as a filterbank, and just 2 levels of decomposition are used and 3) The optimal operation count for the DCT of a row vector is as you have calculated previously. **[10 marks]**

5. The MPEG compression standard has been in place for over ten years and is the backbone of the Digital Television industry. A key component in the compression process is the step of predicting one frame from another, allowing for motion.

(a) In creating frame n from frame $n-1$, MPEG schemes employ motion compensated prediction. There are two ways of achieving this. i) Frame n can be built by copying pixels from respective motion compensated locations in rame $n-1$ or ii) pixels from frame $n-1$ can be copied into their respective motion compensated locations in frame n. In i) a pixel at site $[h,k]$ in frame n is given a value at site $[h+d_1, k+d_2]$ from frame $n-1$ (where d_1, d_2 are the two components of motion between the frames). In ii) pixels at site $[h,k]$ in frame $n-1$ are copied into location $[h-d_1, k-d_2]$ in frame n. Although these two prediction equations are mathematically equivalent, MPEG schemes choose to PULL ie. use (i) for motion compensation rather than PUSH i.e. (ii). Explain why. **[5 marks]**

(b) Explain the meaning of the terms I-frames, B-frames, P-frames and GOP, and explain their purpose in the MPEG codec. **[5 marks]**

(c) Use a simple example to show and explain why the decode order of the frames in a GOP is different from the display order. **[4 marks]**

(d) The bit rate of compressed movies tends to spike dramatically at shot cuts. Why is this and how can an MPEG codec be modified to reduce this effect? **[3 marks]**

(e) The design of real time codecs requires that attention be paid to buffer sizes for motion compensated prediction, and different broadcasters select different encoding strategies depending on their real time requirements. YouTube uses H264 encoding with 90 frame GOPs having no B frames. Digital TV broadcasters typicaly use 12 frame GOPS with all frame types (typically only two consecutive B frames). State the likely frame sequence that is used to create GOPs in these two cases and explain why these particular GOP choices are appropriate for the needs of the respective broadcaster. **[8 marks]**

x	erfc(x)	x	erfc(x)
0.10	0.8875	1.30	0.0660
0.20	0.7773	1.40	0.0477
0.30	0.6714	1.50	0.0339
0.40	0.5716	1.60	0.0237
0.50	0.4795	1.70	0.0162
0.60	0.3961	1.80	0.0109
0.70	0.3222	1.90	0.0072
0.80	0.2579	2.00	0.0047
0.90	0.2031	2.10	0.0030
1.00	0.1573	2.20	0.0019
1.10	0.1198	2.30	0.0011
1.20	0.0897	2.40	0.0007

Table 1: Values for the erfc(x) function. erfc(x) $= \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-u^2)du$