EE4C08 - Tutorial 1
Image and Video Processing
Prof. Anil Kokaram

1. (a) Derive the difference equations for the following linear filters with transfer functions as shown. You can assume that z_1 and z_2 correspond to the horizontal and vertical 2D z-transform variables respectively.

$$
\begin{aligned}
H_1(z_1, z_2) &= \frac{1}{1 - 0.99z_1^{-1} - 0.9z_2^{-1} + 0.891z_1^{-1}z_2^{-1}} \\
H_2(z_1, z_2) &= 0.64z_1^{-1}z_2^{-1} + 0.8z_1^{-1} + 0.64z_1^{-1}z_2 + 0.8z_2^{-1} + 1.0+ \\
&\quad 0.8z_2 + 0.64z_1z_2^{-1} + 0.8z_1^{1} + 0.64z_1z_2 \\
H_3(z_1, z_2) &= 1 - z_1^{-1} \\
H_4(z_1, z_2) &= -z_2^{-1}z_1^{-1} + z_1^{-1} + -z_2^{-1} + 1 \\
H_5(z_1, z_2) &= \frac{1}{1 - 0.99z_1^{-1} - 1.0z_2^{-1} + 1.98z_1^{-1}z_2^{-1}}
\end{aligned}
$$

   (b) For each of the filters above, classify them as IIR or FIR. State in addition their causality i.e. whether they are causal, non-causal or semi-causal.

   (c) Classify the filters as separable and non-separable. In the case of separable filters, find the row and column filters.

   (d) A video processing system requires a vertical edge strength measurement of an input image. Which of the above filters would you use for this?

2. A signal $f_l(h,k)$ is processed with a variety of filters with convolution masks as shown below. The results of processing are shown in figure 1, *but not in the right order.* The images have also been modified so that the mid-gray value represents the 0 value, darker pixels indicatwe negative values and brighter values indicate positive values. By observing the filter masks, and using your knowledge of what filters do to images, identify the corresponding filter and output image. Explain your selection.

$$p_1(h,k) \equiv [1 \; -1]$$

$$p_2(h,k) \equiv \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$p_3(h,k) \equiv \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$p_4(h,k) \equiv \begin{bmatrix} 0.0013 & 0.0086 & 0.0159 & 0.0086 & 0.0013 \\ 0.0086 & 0.0581 & 0.1076 & 0.0581 & 0.0086 \\ 0.0159 & 0.1076 & 0.1993 & 0.1076 & 0.0159 \\ 0.0086 & 0.0581 & 0.1076 & 0.0581 & 0.0086 \\ 0.0013 & 0.0086 & 0.0159 & 0.0086 & 0.0013 \end{bmatrix}$$

$$p_5(h,k) \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

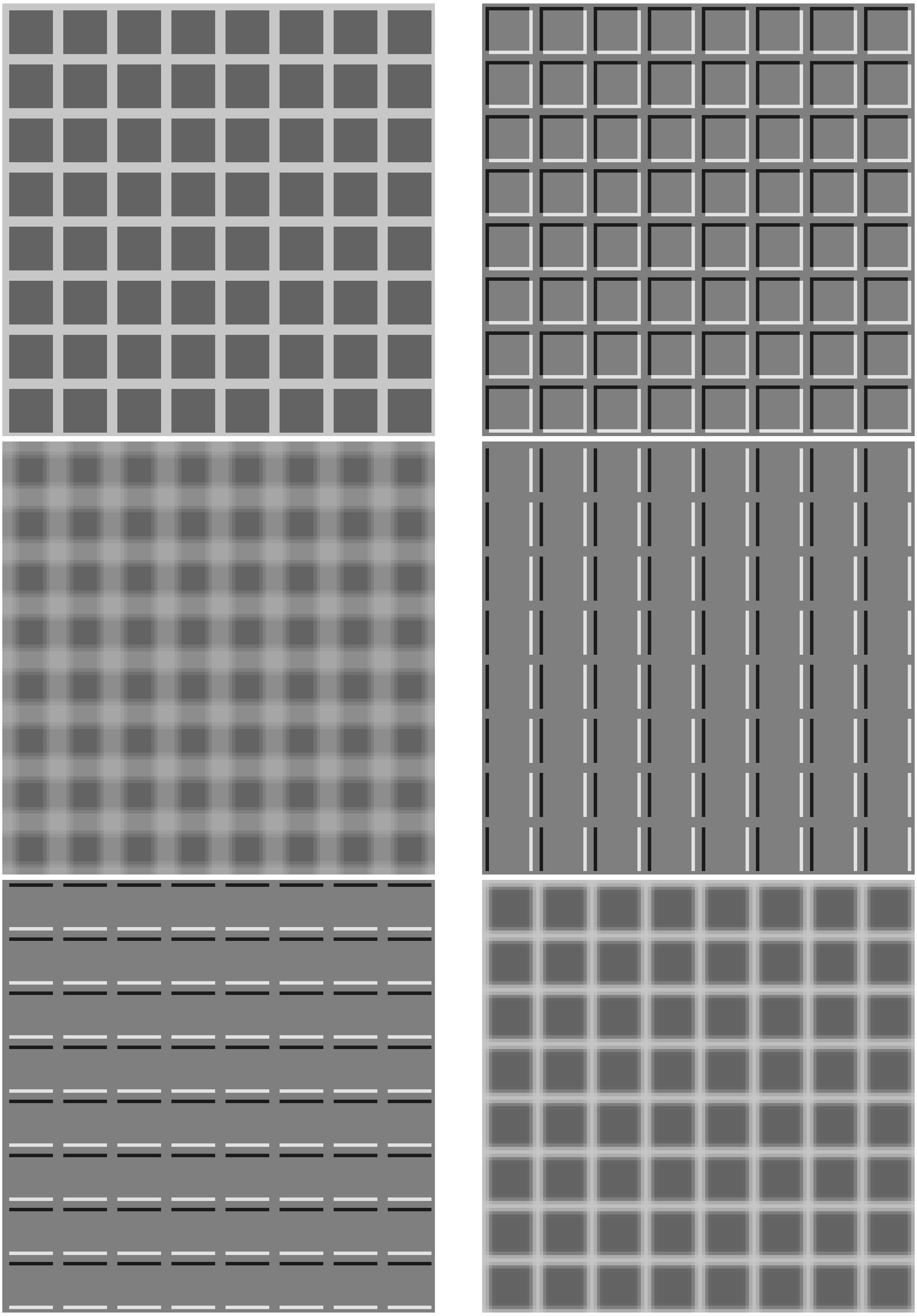

Figure 1: Top Left: $f(h,k)$, Clockwise from top right: Output A, B, C, D, E

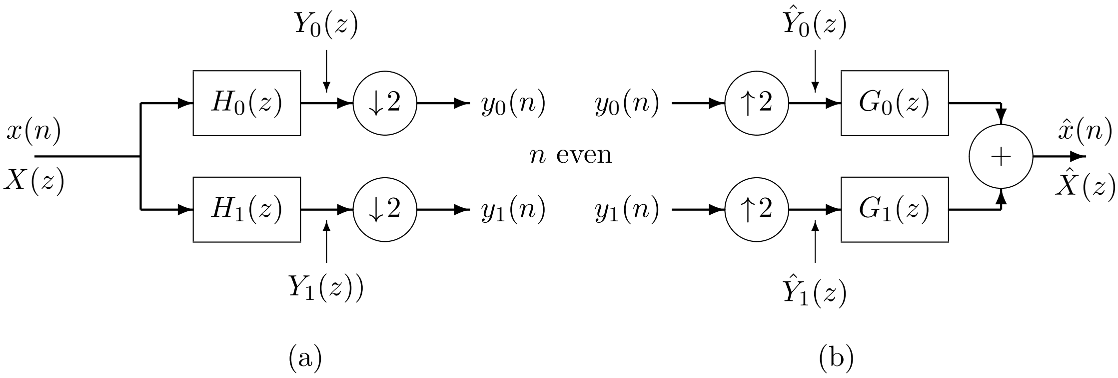

Figure 2: Two-band filter banks for analysis (a) and reconstruction (b).

3. The system shown in figure 2 shows the filters used in the basic unit of a Perfect Reconstruction filterbank. In this course you know that the following expressions must hold for Perfect Reconstruction.

$$G_0(z)H_0(z) + G_1(z)H_1(z) \equiv 2 \tag{1}$$

and

$$G_0(z)H_0(-z) + G_1(z)H_1(-z) \equiv 0 \tag{2}$$

Show that the LeGall 3-5 tap filters satisfy the Perfect Reconstruction requirement. The LeGall filters are as below.

$$\begin{aligned}
H_0(z) &= \frac{1}{2}(z + 2 + z^{-1}) \\
G_0(z) &= \frac{1}{8}(-z^2 + 2z + 6 + 2z^{-1} - z^{-2}) \\
G_1(z) &= \frac{1}{2}z(-z + 2 - z^{-1}) \\
H_1(z) &= \frac{1}{8}z^{-1}(-z^2 - 2z + 6 - 2z^{-1} - z^{-2})
\end{aligned} \tag{3}$$

4. The images on the left of figure 3 show a series of images of differing content and on the right (but not in order) are the 20× Log magnitudes of their 2D FT's. Find the correct signal and transform picture pairs, explaining your choices.

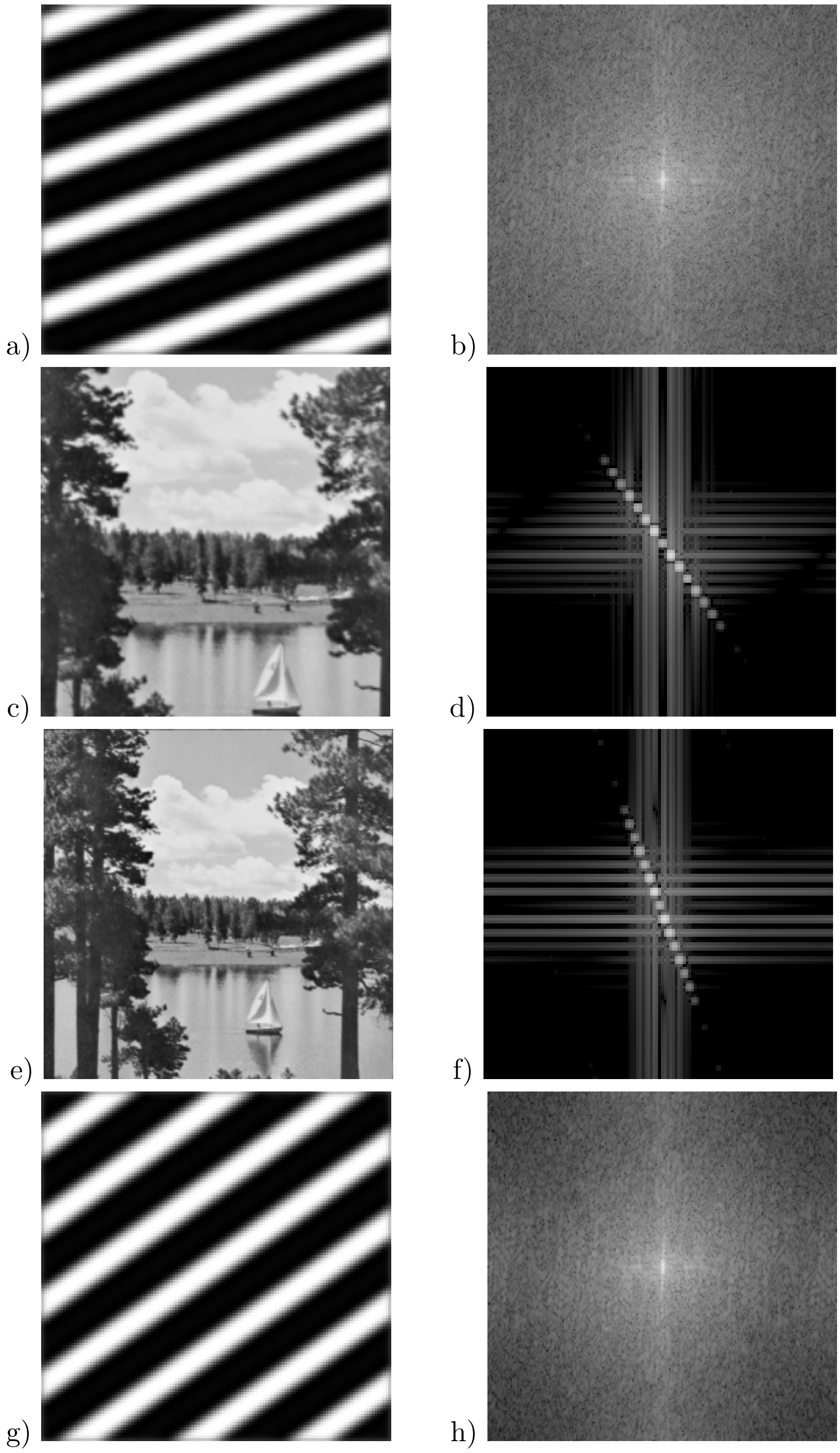

Figure 3: *A mixture of pictures and their 2D Log Power Spectra estimated using the 2D DFT*

5. An 8×8 block of pixels in a greyscale image gives the following set of symbols at the output of a standard implementation of RLC in a JPEG encoder.

```
(4) 1111 (0,3) 110 (0,3) 000 (0,1) 1 (0,2) 10 (0,1) 1 (2,1) 0 (1,1) 1
(7,1) 1 (6,2) 00 (0,1) 1 (3,1) 1 (4,1) 0 (4,1) 1 EOB
```

(a) Given that the dc coefficient of the previous block is -6, calculate the 64 DCT coefficients of the block arranged in the standard 8×8 square matrix format.

(b) Given that each coefficient of your answer represents the quantisation level for each band of the DCT, indicate how you would calculate the grayscale intensities of the reconstructed image blocks.

The following questions are based on questions from the 2010 and 2011 exams.

6. A pixel based motion detector is used to detect motion in a real video signal, $G_n(\vec{x})$, before FSBM. It measures the absolute value of the pixel difference at a site, $|\Delta_n(\vec{x})|$, and assumes motion when $|\Delta_n(\vec{x})| > T_\Delta$, where T_Δ is some threshold.

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

$$p(\Delta_n(\vec{x})) = \frac{1}{2x_0} \exp\left(\frac{-|\Delta_n(\vec{x})|}{x_0}\right) \tag{4}$$

   (a) Show that the probability that a pixel is detected as moving is $p_m(T_\Delta) = \exp(-T_\Delta/x_0)$.

   (b) A block based motion detector is built using this pixel based detector, with $T_\Delta = 1$, by detecting motion in a block when more than 50% of the pixels in a block have been flagged as moving by the pixel based motion detector. Given a block size of 8×8 pixels, calculate the probability that a block is detected as moving.

   (c) Hence or otherwise estimate the computational savings C_1/C_2 created by the motion detector, where C_1 is the number of operations required to perform FSBM (with a search radius of 4 and an 8×8 block) after motion detection on a frame of size 1920×1080, and C_2 is as defined above. Note that your estimate of C_1 should also include the operation count for performing the motion detection stage.

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

$$b_n(\vec{x}) = \begin{cases} 1 & \text{if } |\Delta_n(\vec{x})| > T_\Delta \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

   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.

   - $\Delta_n(\mathbf{x})$ in moving areas is Laplacian with $x_0 = 6.5$ as follows.

$$p(\Delta_n(\mathbf{x})) = \frac{1}{2x_0} \exp\left(\frac{-|\Delta_n(\mathbf{x})|}{x_0}\right) \tag{6}$$

     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 $p_m(T_\Delta) = 1 - \exp(-T_\Delta/x_0)$.

   - In stationary areas $\Delta_n(\mathbf{x})$ is Gaussian with variance $\sigma_v^2 = 100$ and zero mean.
   - The proportion of pixels actually undergoing motion in the image is 0.1.

   (a) Show that the probability that a pixel is misclassified as moving (i.e. $b_n(\mathbf{x}) = 1$) in stationary areas is $p_n(T_\Delta) = \mathrm{erfc}(T_\Delta/\sqrt{2\sigma_v^2})$. See Table 1 for the required integral.

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. $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-u^2) du$

(b) 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. The optimum threshold is the value of T_Δ that minimises the total probability of error.

(c) 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 on the probability distribution in moving areas, estimate the new optimal threshold that should be used.