Data Provided: None

DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING

Spring Semester 2010-2011 (2 hours)

EEE6081/EEE421 Visual Information Engineering 6

Solutions

1.

a. For the orthogonality condition

If the inner product $\langle f_n, f_m \rangle = 1$ when n = m and
$= 0$ when $n \neq m$.

In other words $\sum_{i=0}^{3} f_{ni} f_{nm} = \delta_{nm}$

<f0,f0> =<f1,f1> = (hxh)+(hxh) = $2h^2=1$
h = +/- 1/sqrt(2)

(2)

b. <f0,f0> =<f1,f1> =1

<f0,f1> = ((hxh)+(-hxh)) = 0

Satisfies $\langle f_n, f_m \rangle = 1$ when n = m and
$= 0$ when $n \neq m$.

(2)

c. y_0 contains the downsampled output of low pass filter. Half resolution low passed signal containing the same features as the original signal.

y_1 contains the downsampled output of high pass filter. Half resolution high passed signal contains the singularities in the single.

(3)

d.

$$T^{-} = \begin{bmatrix} h & h & 0 & 0 \\ 0 & 0 & h & h \\ h & -h & 0 & 0 \\ 0 & 0 & h & -h \end{bmatrix}$$

(3)

e. F is orthogonal. Therefore, the inverse matrix is the transpose.

$$T^{-1} = \begin{bmatrix} h & 0 & h & 0 \\ h & 0 & -h & 0 \\ 0 & h & 0 & h \\ 0 & h & 0 & -h \end{bmatrix}$$

(2)

f.

$$T = \begin{bmatrix} h & h & 0 & 0 \\ h & -h & 0 & 0 \\ 0 & 0 & h & h \\ 0 & 0 & h & -h \end{bmatrix}$$

(3)

g. Transform matrix for 2 data points **(5)**

$$T = \begin{bmatrix} h & h \\ h & -h \end{bmatrix}$$

Create low pass transform matrix L (64x128) and H (64x128)

$$L = \begin{bmatrix} h & h & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & h & h & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & h & h \end{bmatrix}$$

$$H = \begin{bmatrix} h & -h & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & h & -h & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & h & -h \end{bmatrix}$$

$$T = \begin{bmatrix} L \\ H \end{bmatrix}$$

Apply T on columns first, and then on rows to get 4 sub bands LL, HL, LH and HH.

Now create the second level transform matrices matrix L2 (32x64) and H2 (32x64)

$$L2 = \begin{bmatrix} h & h & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & h & h & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & h & h \end{bmatrix}$$

$$H2 = \begin{bmatrix} h & -h & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & h & -h & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & h & -h \end{bmatrix}$$

$$T2 = \begin{bmatrix} L2 \\ H2 \end{bmatrix}$$

Apply T2 as separable processes on each of the four sub bands.

2.

a. Any three of the following:

- fast
- Easy to construct new wavelet families (No need to use spectral factorization)
- in place computations
- easy to construct adaptive and non-linear transforms
- reversible integer transforms **(3)**

b. The lifting steps in the matrix form:

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

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

The top row represents the low pass filter and the second row represents the high pass filter. **(3)**

c.

$$x_i = p_{2i}$$
$$y_i = p_{2i+1}$$

(2)

d.

$$y_i^{'} = y_i^f + \left(\sqrt{2}-1\right)x_i^f$$
$$x_i = x_i^f - \frac{1}{\sqrt{2}}y_i^{'}$$
$$y_i = y_i^{'} + \left(\sqrt{2}-1\right)x_i$$

(3)

e.

$$y_i^{'} = y_i - \left\{\left(\sqrt{2}-1\right)x_i\right\}$$
$$x_i^f = x_i + \left\{\frac{1}{\sqrt{2}}y_i^{'}\right\}$$
$$y_i^f = y_i^{'} - \left\{\left(\sqrt{2}-1\right)x_i^f\right\}$$

where { } represents rounding to the nearest integer. **(2)**

f.

$$y_i^{'} = y_i - \frac{\left(\sqrt{2}-1\right)}{2}\left(x_i + x_{i+1}\right)$$
$$x_i^f = x_i + \frac{1}{2\sqrt{2}}\left(y_i^{'} + y_{i-1}^{'}\right)$$
$$y_i^f = y_i^{'} - \frac{\left(\sqrt{2}-1\right)}{2}\left(x_i^f + x_{i+1}^f\right)$$

(3)

g. This is usually obtained by performing the temporal decomposition of a group of pictures first, followed by the 2D separable decomposition of the temporally decomposed frames (t+2D).

The temporal decomposition is done by using the motion compensated temporal lifting (MCTL). The prediction and update step are performed by considering the motion compensation. The first level of temporal decomposition results in a half a GOP each of low pass and high pass frames. Low pass frames are repeatedly decomposed using the MCTLto get a dyadic temporal decomposition. **(4)**

3.

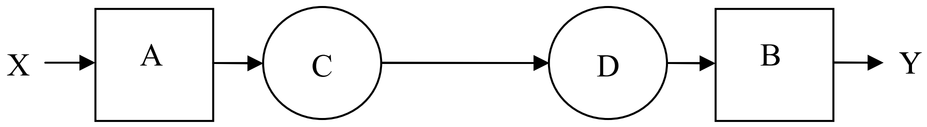
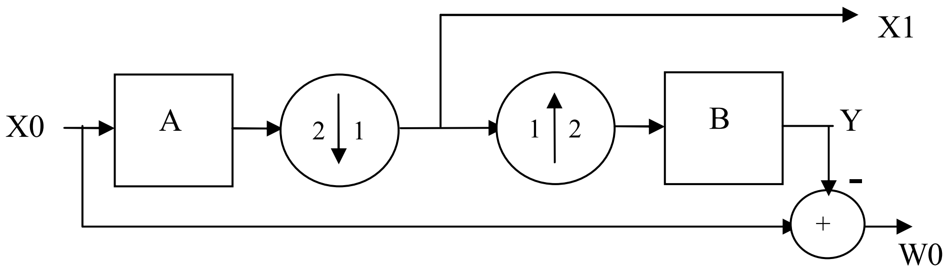


Figure 1

Figure 1 shows a block diagram of a system for creating an approximated version (Y) of a one-dimensional signal (X). A and B represent low pass filters. C and D represent down-sampling and interpolating (i.e., inserting zero –valued samples) by a factor of 2, respectively.

a.



X0 is the original signal

X1 is the half resolution approximation, which is the signal after the downsampling operation.

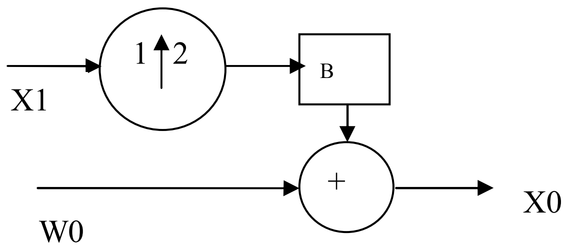
W0 is the details of the higher resolution, which is obtained using the difference between the original and the approximated signal.

X1 and W0 form a pyramidal representation of X0 with X1 being the approximated down sample signal and W0 being the details at the original representation.

X1 can be further decomposed into two components by using the same system as cascaded operations.

(3)

b.



At each level, the image from the previous level (the scaled down image) is interpolated and added to the corresponding details in that level **(3)**

c. A and B 2-Dimensional low pass filters

C is a 4: 1 decimator and D is a 1:4 interpolator. That means for each pixel, right, bottom and right-bottom neighbours are filled in with zero-valued samples. **(2)**

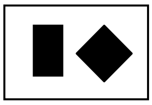
d. The sampling redundancy factor for a 3 level decomposition

$$((1 + 1/4) + 1/16) + 1/64 = 1 + 21/64$$

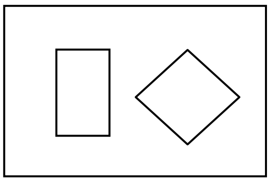
(2)

e. The discrete wavelet transform is a non-redundant representation. Therefore, the redundancy factor is 1. But the redundancy factor of a pyramid scheme for a 1D scheme approaches to 2 and for 2D scheme approaches to 4/3. Both are greater than 1. Less memory space is required for the wavelet transform coefficient storage. **(2)**

f.



Half resolution low pass



detail sub band

(3)

g. A 3 level 2-dimensional pyramid-based multi-resolution representation results in 4 subbands:

H1 – Very high frequency detailed sub band from the level 1 (full resolution)

H2 - high frequency detailed sub band from the level 2 (1/2 resolution)

H3 - high frequency detailed sub band from the level 3 (1/4 resolution)

L4- low frequency sub band from the level 3 (1/8 resolution)

Medium magnitude coefficients in mid frequency sub bands (H2, and H3). The coefficients should correspond to textured area so that the modifications are not visible.

For minimum effect on visual quality low magnitude coefficients need to be chosen. Lower magnitude leads to lower distortion. But for robustness high frequency, high magnitude coefficients are better. So to compromise the two complementary requirements, medium magnitude coefficients in mid frequency sub bands can be chosen. **(5)**

4. **a.** The answer should contain the following:

The DWT decorrelates the image.

Compacts the energy of the image into a fewer number of coefficients.

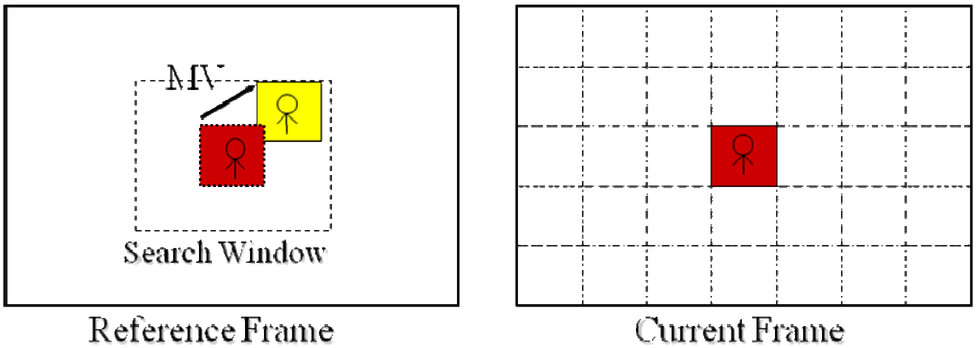
Provides a multi resolution framework.

By encoding the coefficients according to their importance on the contribution to the total image energy provides quality scalability. This is usually done by bit plane by bit plane coding resulting in hierarchical embedded quantizers.

Resolution scalability is achieved by coding/deciding coefficients from the highest level of decomposition to the lowest decomposition.

(5)

b.



The current frame (C) is partitioned into non-overlapping blocks.

For each block, within a search window in the reference frame (R), find the motion vector (displacement) that minimizes a pre-defined mismatch error (e.g., sum of absolute difference (SAD)), using a full search, where all possible MV candidates within the search range are investigated.

SAD for a block at (x,y) location (top-left hand coordinates), for a specific displacement (dx,dy) is computed as follows:

$$SAD(dx,dy) = \sum_{i=0}^{b-1}\sum_{j=0}^{b-1}\left|C(x+i,y+j) - R(x+i+dx,y+j+dy)\right|$$

To get the accurate motion models, in modern video coding standards,

i. fractional pixel motion vectors

ii. hierarchical variable block sizes fields to account for the motion of variable size objects

are used.

(5)

c. Advantages: Higher coding gain due to inter-frame prediction

Disadvantages: High computational complexity, Error propagations into P and B frames, no real-time decoding (requires a buffer) **(3)**

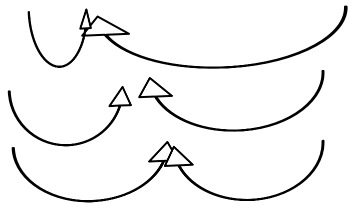
d. Only one I frame, followed by sequential predictions. An error in any frame is propagated to subsequent frames.

Include an I frame at regular intervals. (e.g., every 6 -8 frames). **(2)**

e.

I B B B B B I

1 2 3 4 5 6 7



Coding/decoding order 1, 7, 2, 3, 4, 5 , 6 **(2)**

f. Coding/decoding order for method 4: 1, 7, 2, 3,4, 5 , 6

Coding/decoding order for method 3: 1, 3, 2, 5 , 4, 7, 6

Method 4 has a higher delay compared to the method 3.

The complexity of method 4 is higher as more B frames are involved (two sets of motion vector fields as opposed to one in P frames) **(3)**

GCKA