Data Provided: None

DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING

Spring Semester 2013-14 (2.0 hours)

EEE6081 Visual Information Engineering -Solutions

1.

a. For orthogonality:

$a^2 + b^2 = 1$ (1)

$c^2 + d^2 = 1$ (2)

$ac + bd = 0$ (3)

For regularity:

$a + b = \sqrt{2}$ (4)

from (4) and (1)

$(\sqrt{2}-b)^2 + b^2 = 1$

$1 - 2\sqrt{2}b + 2b^2 = 0$

$(1 - \sqrt{2}b)^2 = 0$

$b = 1/\sqrt{2}$

from (4) $a = 1/\sqrt{2}$

from (3) $c + d = 0$ therefore, $d = -c$

from(2) $2c^2 = 1$ $c = \pm 1/\sqrt{2}$

choose $c = 1/\sqrt{2}$ then $d = -1/\sqrt{2}$

(5)

b. For the input $[x_1\ x_2\ x_3\ x_4]$,

The transform matrix is

$$\begin{bmatrix} a & b & 0 & 0 \\ 0 & 0 & a & b \\ c & d & 0 & 0 \\ 0 & 0 & c & d \end{bmatrix}$$

(2)

c. Four basis functions

F1=(a b 0 0)

F2=(0 0 a b)

F3=(c d 0 0)

F4=(0 0 c d) **(4)**

Taking the dot products

F1.F1=F2.F2= $a^2 + b^2 = 1$

F3.F3= F4.F4= $c^2 + d^2 = 1$

F1.F2=F1.F4=F3.F4=F3.F2=0

F1.F3=F2.F4= $ac + bd = 0$

Therefore the system is orthogonal

Therefore the inverse is the transpose

$$\begin{bmatrix} a & 0 & c & 0 \\ b & 0 & d & 0 \\ 0 & a & 0 & c \\ 0 & b & 0 & d \end{bmatrix}$$

d. y_0 consists of low pass filtered half resolution representation of the signal. It provides an approximation of the input signal.

If the input signal is highly correlated, the most of the energy and the entropy of the signal are compacted in this channel. This can be further decomposed using as the input to the wavelet transform. For applications, the lowpass signal can be used with low quantization in compression and low complexity operations for low resolutions.

y_1 consists of high pass filtered half resolution representation of the signal.

These are used in compression by eliminating or highly quantizing. The high pass signal are also used in denoising (by eliminating the values lower than athreshold), edge detection or feature extraction.

(4)

e.

2D single level decomposition:

Apply the wavelet transform on the rows

Then apply on the columns

This will give four subands: LL, LH, HL and HH

Then use the LL subband repeat the single level 2D decomoposition

Repeat this process iteratively on LL subband successively for at least 4 levels of decomposition.

For high pass samples identify the noise coefficients:

Use a threshold T. Replace the coefficients whose magnitudes are lower than T, with zero.

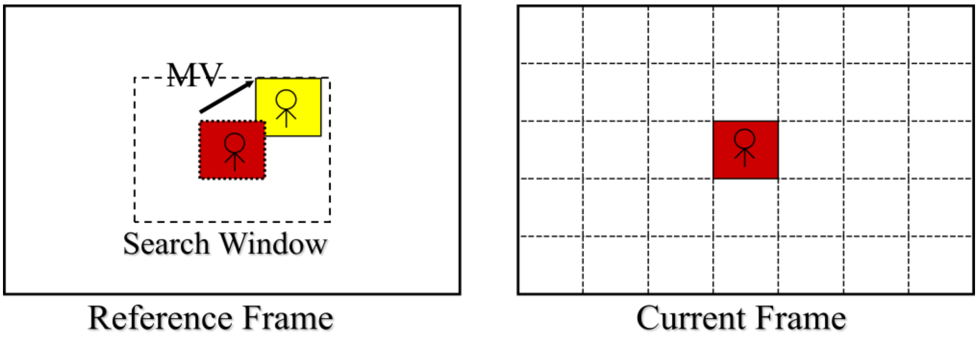
Then do the inverse 2D wavelet transform, applying the 2D single level wavelet transform on the smallest resolution decomposition level iteratively to build up **(5)** the original resolution level.

2D single level reconstruction:

Apply the wavelet transform on the columns

Then apply on the rows.

2. **a.**



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:

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

(4)

b. A frame of pixels N x M

Blocks of BxB

Number of Blocks = $NM/(B^2)$

Motion vector range -w to w

Search window is (2w+1) x (2w+1)

Number of operations per search point in SAD computation is B^2

For the full search per window per block: $B^2(2w+1)^2$

For the whole frame $= B^2(2w+1)^2NM/(B^2) = NM(2w+1)^2$

This is independent of the block size.

Linearly proportional to frame dimension and square proportional to the motion vector range.

(4)

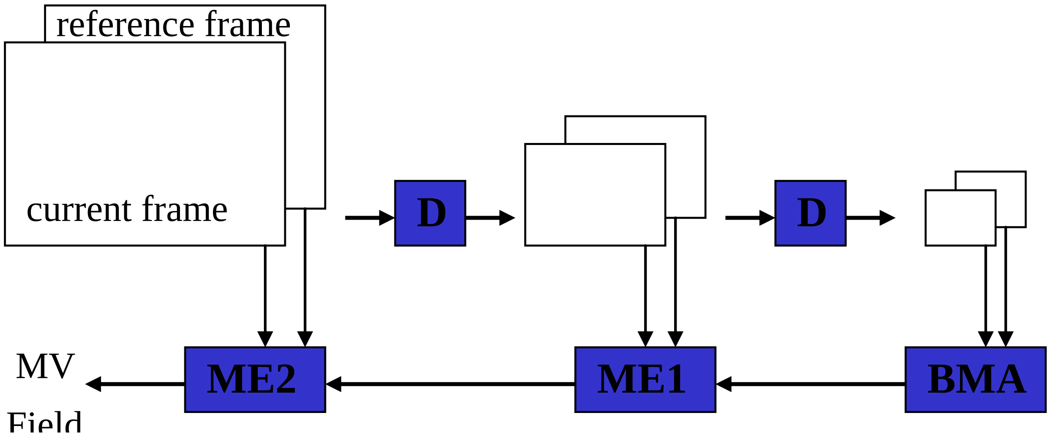
c. compression efficiency: this depends on the cost of motion vectors and the accuracy of the motion prediction leading to smaller prediction residuals. A content adaptive variable block size is good for this case.

prediction accuracy: smaller block size is better

cost of motion vectors: bigger block sizes result in fewer number of vectors, thus with low cost

error propagation: smaller block sizes result in smaller areas are being predicted. If motion vectors are corrupted due to transmission errors, the effect is slower if the block sizes are smaller. **(4)**

d.



D is any multi-resolution decomposition like wavelet transform.

BMA is applied in the ¼ resolution frames with a new search area w/4. The new complexity is (NM/16) $(w/2+1)^2$

This perovides a coarse estimation. Then ME1 and ME2 refines the motion for higher resolutions. Only extra 8 searches per block is required at each stage.

(4)

e.

a) Motion Activity Descriptors– can be determined using the density and the magnitude of motion vectors

b) Camera Motion Descriptors – BMA only considers translational motion – so only possible to extract camera pan information. Not possible to get zooming in out information

c) Motion Trajectory Descriptors – motion active regions can be segmented by merging the blocks with motion vectors with specific magnitude and directions

d) Parametric Motion Descriptors - only translational motion parameters can be estimated

(4)

3. **a.** 8 bits can represent 256 values. 0-255

4 bits can represent 8 different values 0 -15

Therefore 0-255 has to be mapped to 0-15

All pixel values must be re-quantized by dividing by (256/16) 16.

(3)

b. low spatial frequency regions - due to quantization artificial contours will appear

high frequency regions- for low quantization, such artificial contours won't appear. But for high quantization, some high frequency details might be lost

(3)

c. CIF resolution: 1080 x 1920

Sampling factor for 4:2:0 is 1.5

Data rate 1080 x 1920 x 8 x 1.5 x 50 = 1244160000 bits/sec

=1.16 G bits /sec

(4)

d. (i) Quality scalable image coding – Encoded bit streams can be decoded at any data rate lower than the original data rate to decode an image with lower quality compared to that of the original. The bit streams are organised with the progressive quality.

(ii) Quality scalability – Using dyadic wavelet decomposition and encoding them from bit plane by bit plane, sp that with every bit plane the quantization bin sizes are refined to half of the bin sizes in the previous bit plane.

(iii) Progressive downloading and displaying of images. With first view a low quality image is displayed. The quality will be improved as more bits are downloaded and decoded.

(5)

e. Forward transform

$$y_1 = x_1 + \{ax_0\}$$
$$y_0 = x_0 + \{by_1\}$$

where {} represents rounding to the nearest integer

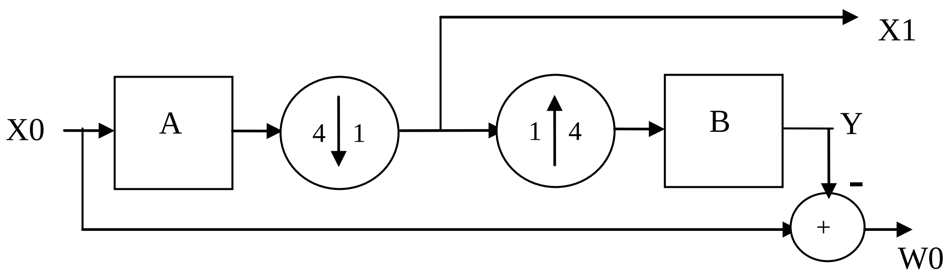
inverse transform

$$x_0 = y_0 - \{by_1\}$$
$$x_1 = y_1 - \{ax_0\}$$

where {} represents rounding to the nearest integer

(5)

4. a.



X0 is the original image

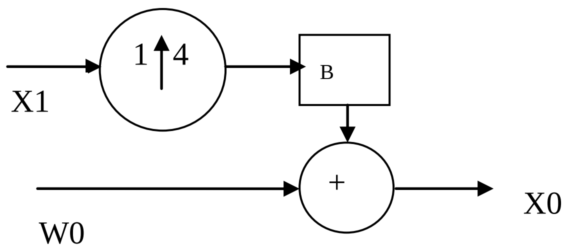
X1 is the half resolution (quarter) approximation, which is the image after the downsampling operation.

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

X1 and W0 form a pyramidal representation of X0 with X1 being the approximated down sample image 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.

Reconstruction



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

(5)

b. The sampling redundancy factor for a 1 level decomposition

1 + 1/4

The sampling redundancy factor for a 2 level decomposition

(1 + 1/4) + 1/16

The sampling redundancy factor for a 3 level decomposition

((1 + 1/4) + 1/16)+ 1/64

(5)

This leads to a geometric series with a=1 and r=1/4

Therefore the highest value of the redundancy factor when the number of levels are increased to a large number is

a /(1-r) = 1/ (1-1/4) = 4/3.

c. Transform the image into a 5 level pyramid transform of the image of NxM resolution.

Leading to 5 subbands of high frequency components (d0, d1, d2, d3, d4 with resolutions, NxM, N/2xM/2, N/4xM/4, N/8xM/8, N/16xM/16 respectively and the 5th level approximation band a5 with resolution N/32xM/32.

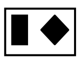
Use the quantisation parameters Q0>Q1>Q2>Q3>Q4>Q5 for the six subbands to quantize and entropy encode.

The spatial resolution scalability can be achieved by combined decoding as follows

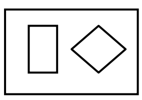
1) a0 only to get N/32xM/32 image
2) a0 and d4 to get N/16xM/16 image
3) a0, d4 and d3 to get N/8xM/8 image
4) a0, d4, d3 and d2 to get N/4xM/4 image
5) a0, d4, d3 d2 and d1 to get N/2xM/2 image
6) a0, d4, d3 d2 d1 and d0 to get N/2xM/2 image

(5)

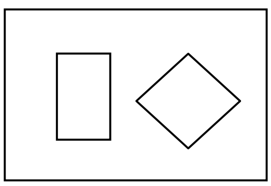
d.



quarter resolution low pass (M/4 x N/4 pixels)



Half resolution detail (M/2 x N/2 pixels)



Full resolution detail sub band (M x N pixels)

(5)

GCKA