Data Provided: None

Spring Semester 2011-12 (2.0 hours)

EEE6081 Visual Information Engineering Solutions

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

3 bits can represent 8 different values 0 -7

Therefore 0-255 has to be mapped to 0-7

All pixel values must be re-quantized by dividing by (256/8) 32.

(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: 288 x 352

Sampling factor for 4:2:0 is 1.5

Data rate 288 x 352 x 8 x 1.5 x 50 = 60825600 bits/sec

=58 M bits /sec

(4)

d. Uncompressed data consists of a lot of bits as computed in part c compared to the available bandwidths in the networks and limited space in storage devices. Therefore, compression is necessary

Data compression is possible due to:

- Redundancy in source data (inter-pixel redundancy due to correlation and coding redundancy due to methods used in representation)
- Irrelevant data due to limitations of the human visual and hearing system

(4)

e. The main purpose is to remove inter-pixel redundancy.

The image is decorrelated, so that the probability distribution is Laplacian – narrow peak and long tails. This representation reduces the source entropy.

The energy of the image is compacted into a smaller number of coefficients. This enables applying quantization (to exploit HVS limitations) to improve compression

(3)

f. Two approaches:

a) The use of causal prediction schemes as 2D DPCM
b) Integer transforms. That is transforms that are perfectly reversible and can map integers in to integers without using rounding operations. **(3)**

2.

a. Filter A is length 2. Therefore, the coefficients are {a b}

Using orthogonality condition:

For k=0; $a^2 + b^2 = 1$

Using regularity condition:

a + b = sqrt(2)

Solving two equations a=b= 1/sqrt(2)

(4)

b. If the filter A is {a b}, filter B is {a –b}

The top half corresponds to the low pass filtering while the bottom half corresponds to the high pass filtering.

$$\begin{bmatrix} a & b & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a & b & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a & b & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & b \\ a & -b & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a & -b & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a & -b & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & -b \end{bmatrix}$$

(4)

c. X_L consists of low passed 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.

Compression is achieved by eliminating or highly quantizing the high pass coefficients and low quantizing of the low pass coefficients. **(4)**

d. Apply the wallet transform in multiple levels as a dyadic decomposition.

High pass sub bands capture the signal variations, which can be due to the edges in the signal and noise.

The variations due to edges are higher. While the variations due to the noise are smaller. So are the magnitudes of the high pass sub band coefficients. Therefore, by setting all high pass coefficients whose magnitudes are less than a threshold to zero and applying the inverse wavelet transform will remove noise.

. **(4)**

e. Transform matrix for 2 data points

$$T = \begin{bmatrix} a & b \\ a & -b \end{bmatrix}$$

Create low pass transform matrix L (64x128) and H (64x128) **(4)**

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

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

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

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) with the diagonal elements (a b) and (a –b) respectively for L2 and H2.

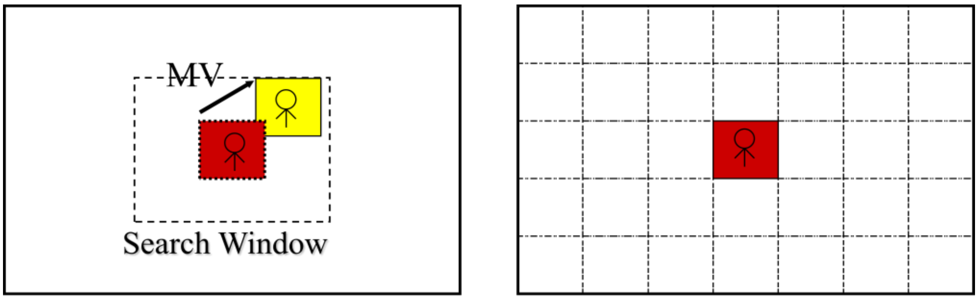
$$L2 = \begin{bmatrix} a & b & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & a & b & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & a & b \end{bmatrix}$$

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

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

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

3. a.



Reference Frame   Current Frame

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)

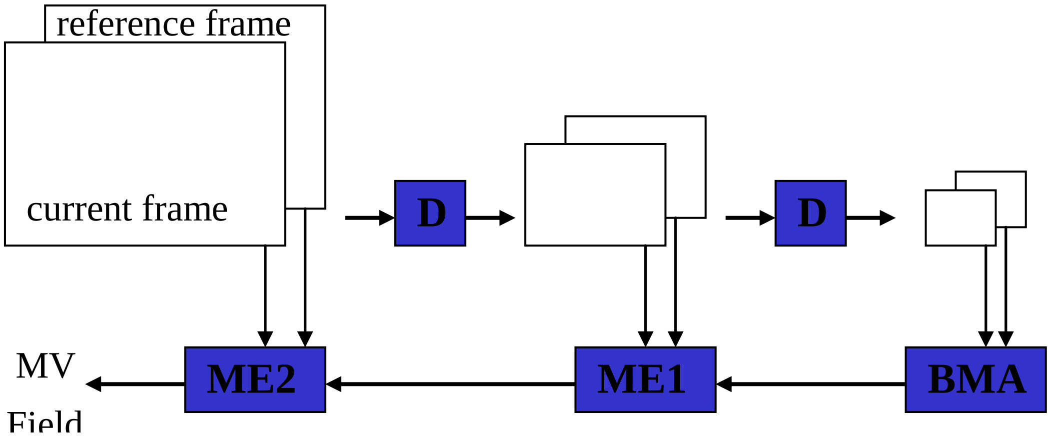
c. compression efficiency: this depends on the cost of motion vectors and the accuracy of the motion prediction leading to smaller prediction residuals. A content adaptive variable block size is good for this case.

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

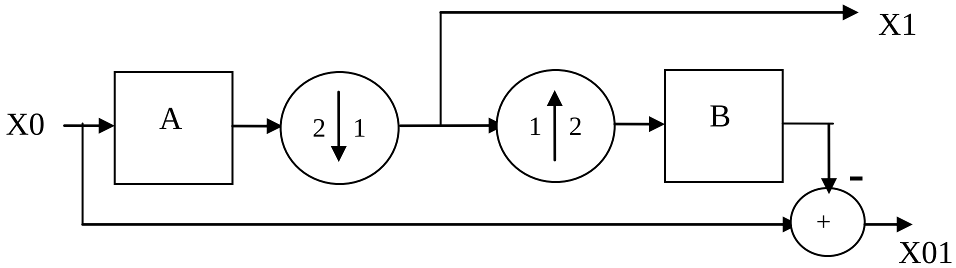
b) Camera Motion Descriptors – BMA only considers translational motion – so only possible to extract camera pan information. Not possible to get zooming in out information

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)**

4.

a.



X0 is the original signal

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

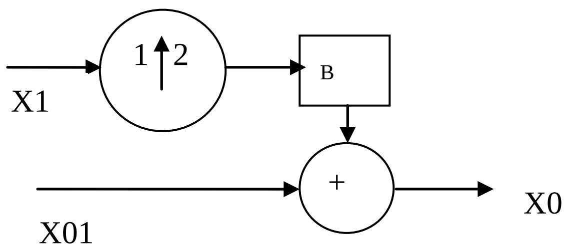
X01 is the details of the higher resolution, which is obtained using the difference between the original and the approximated signal.

X1 and X01 form a pyramidal representation of X0 with X1 being the approximated down sample signal and X01 being the details at the original representation.

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

(4)

b.



For a given level the signal reconstruction is as in the above figure.

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

If the details subbands after levels 1, 2 and 3 are X01, X12 and X23 respectively and the approximated subband at the level is X3,

- X2 is obtained using X3 and X23
- X1 is obtained using X2 and X12
- X0 is obtained using X1 and X01

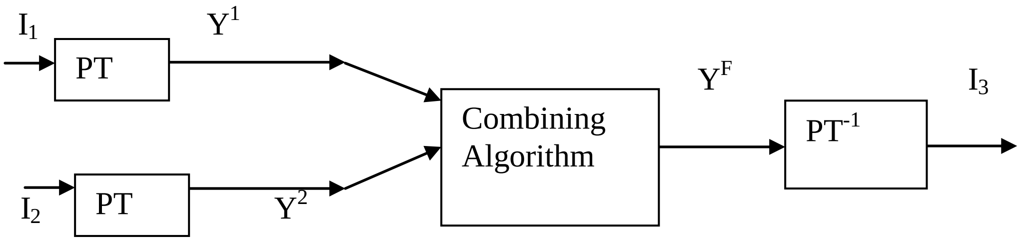
(4)

c. A and B become 2-Dimensional low pass filters. The decimator and the interpolator become 4: 1 and 1:4, respectively.

The sampling redundancy factor for a 3 level decomposition

$$((1 + 1/4) + 1/16)+ 1/64 = 1 + 21/64 = 85/64$$ **(4)**

d.



Images 1 and 2 are decomposed into sub bands using the Pyramid transform (PT). The combining algorithm computes the local activity level around individual coefficients and the energy levels. These values are used to one-to-one comparing in pyramid transform domain to select and combine coefficients from two transformed images to construct Y_F.

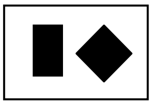
For example;

- $Y^F_{i,j} = (|Y^1_{i,j}| > |Y^2_{i,j}|) \ ? \ Y^1_{i,j} : Y^2_{i,j}$ (For high pass bands)
- $Y^F_{i,j} = aY^1_{i,j} + bY^2_{i,j}$ with a+b=1 (For the low pass band)

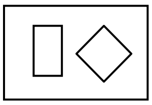
Then perform the inverse Pyramid transform to construct the fused image I3.

(4)

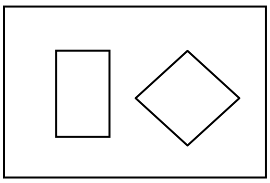
e.



Half resolution low pass



Half resolution detail



Full resolution detail sub band

(4)

GCKA