Data Provided: None

The University Of Sheffield.

Spring Semester 2012-13 (2.0 hours)

EEE6081 Visual Information Engineering 6 - SOLUTIONS

1. a. A transform is needed for

- Decorrelating the signal
- Compact energy into a fewer number of coefficients
- To achieve a different representation so that specific features are prominent.

(3 marks)

In addition to these an orthogonal transform needed to make sure

That the thransform is reversible (invertible)

The total energy in the transform domain is the same the total energy in the pixel domain.

This is important in estimating the mean square error in the coefficient domain during the quantization

(2 marks)

(5)

b. Check whether the transform is orthogonal.

The basis functions are

F1= (64 64 64 64)/128

F2=(84 35 -35 -84)/128

F3=(64 -64 -64 64)/128

F4=(35 -84 84 -35))/128

Show

F1. F2 =0

F1.F3=0

F1.F4=0

F2.F3=0

F2.F4=0 **(5)**

F3.F4=0

F1.F1=1

F2.F2=1

F3.F3=1

F4.F4=1

(3 marks)

Therefore H is orthogonal.

Therefore inverse is H^t

(1 mark)

$$H^t = \frac{1}{128}\begin{bmatrix} 64 & 84 & 64 & 35 \\ 64 & 35 & -64 & -84 \\ 64 & -35 & -64 & 84 \\ 64 & -84 & 64 & -35 \end{bmatrix}$$

(1 mark)

c. Divide the image into 4x4 blocks

For each block,

Apply the transform on columns

Apply the transform on rows

Set the coefficients whose magnitude is less than a threshold value to zero

Apply the inverse transform on rows

Apply the inverse transform on columns

(5)

d.

$$X=\begin{bmatrix} 440 & -24 & -17 & -4 \\ 19 & 12 & 8 & -2 \\ 5 & 4 & 3 & -1 \\ -3 & -3 & 2 & -4 \end{bmatrix}$$

After quantisation and de-quantisation

$$Y=\begin{bmatrix} 440 & -20 & -20 & o \\ 20 & 10 & 10 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(2 marks)

The original total energy is = sum (x.*x)

= 440x440 +24x24 + 17x17 + 4x4 + 19x19 +12x12 +8x8+ **(5)**

2x2+5x5+4x4+3x3+1x1+3x3+3x3+2x2+4x4

=195143

(1 mark)

The new energy is =sum (y*y)

= 440x440 +20x20 + 20x20 + 20x20 +10x10 +10x10

=195100

(1 mark)

% loss = (195143-195100)/195143 = 0.02%

(1 mark)

2. **a.** Forward transform

$$x_0 = X_{2i}$$
$$x_1 = X_{2i+1}$$

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

(2 marks)

where {} represents rounding to the nearest integer

(1 mark)

Inverse transform

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

$$X_{2i} = x_0$$
$$X_{2i+1} = x_1$$

(2 marks) **(5)**

b. Polyphase matrix

$$\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

$$\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{Vmatrix} 1+ab & b \\ a & 1 \end{Vmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

(2 mark)

High pass polyphase components a+1=0

a=-1

(1 mark)

Low pass poly phase components

(1+ab) + b =1

(1 mark)

Symmetry

(1+ab)=b=1/2 therefore b=1/2

(1 mark) **(5)**

c. The low pass component

-half resolution smoothed signal

-Carries most of the energy of the image

- Therefore, can be used for applying the transform again on this component leading to a dyadic decomposition and compacting energy into fewer number of coefficients

- Half resolution constitutes the platform for multi resolution analysis (MRA)

- Therefore, good for MRA –based image analysis tasks and reducing the **(5)** complexity by processing on lower resolutions.

d. The transform matrix is

$$T = \begin{Vmatrix} 1+ab & b \\ a & 1 \end{Vmatrix}$$

$$= \begin{bmatrix} 0.5 & 0.5 \\ -1 & 1 \end{bmatrix}$$

(1 mark)

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

$$L = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0.5 & 0.5 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}$$

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

Apply P on columns first and then on rows

This will result in 4 subbands.

(2 marks)

For the second level

Create low pass transform matrix L (32x64) and H (32x64)

$$L = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0.5 & 0.5 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & & 0 & 0 \\ \vdots & \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}$$

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

Apply P on columns first and then on rows in each of the subband

This will result in 16 subbands.

(2 marks)

(5)

3. a.

1. Spectral Redundancy: The correlation among different spectral bands. For example, the redundancy in RGB bands. Removed by using the RGB to YCbCR conversion (1 mark)

2. Inter-pixel Redundancy: This is due to spatial and temporal correlations with neighbouring pixels. Motion compensated prediction to remove temporal redundancy and transforms, such as DCT and DWT to remove spatial redundancy are used. (2 marks)

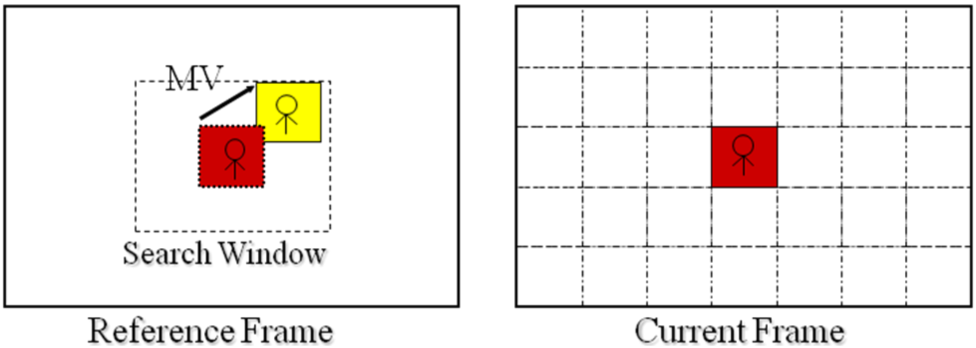
3. Psychovisual Redundancy: The eye does not response with equal sensitivity to all visual information. Certain information has less relative importance than other information in normal visual processing. Such information is said to be psychovisually redundant. Quantization of transform coefficients by choosing various quantisation parameters in various subbands according to their significance into the visual quality is the mai way to remove this redundancy (1 mark)

4. Coding Redundancy: This is present in images when the probability of occurrence of symbols has not been taken into account when assigning binary codes. Entropy coding – (VLC, Huffman coding , run length coding , arithmetic coding) is used

(1 mark)

(5)

b.



(1 mark)

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.

(1 mark)

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

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

(1 mark)

(5)

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.

(2 marks)

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

Disadvantages: High computational complexity, Error propagations into successive frames, buffer requirements

(2 marks)

(ii) Include an I frame at regular intervals. (e.g., every 6 -8 frames). Fast motion estimation algorithms

(1 mark)

(iii) Advantages: Higher coding gain as better predictions, the error propagation is limited compared to method 2

Disadvanatges. Higher computational complexity, coding decoding delays, more space for buffer needed as more reference frames are used

(2 marks)

(5)

d. Main MPEG-7 motion descriptors

a) Motion Activity Descriptors – describes the pace or intensity of a video segment (e.g., high or low activity. can be determined using the motion vectors (their density and magnitude)

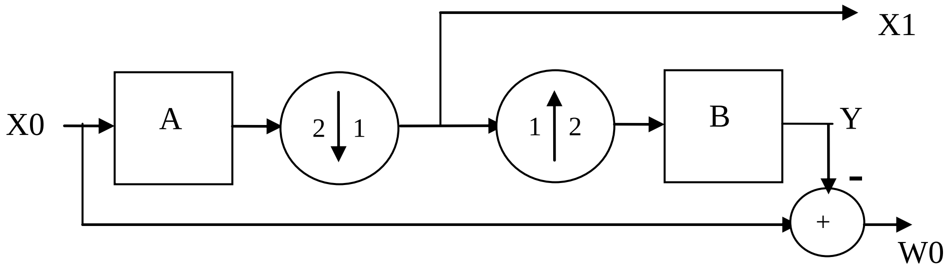
b) Camera Motion Descriptors - Camera track, Camera pan, tilt and roll motion modes for a video segment. BMA only considers translational motion – so only possible to extract camera pan information. Not possible to get zooming in out information

c) Motion Trajectory Descriptors – the moving path of an object or a region. A set of interpolation functions to describe the path . motion active regions can be segmented using the direction and magnitude of vectors.

d) Parametric Motion Descriptors – describes the motion evolution of objects or regions. Translation/rotation/scaling etc. only translational motion parameters are used in BMA. Therefore these information can be extracted

(1.25 marks each) **(5)**

4. a.



(1 mark)

X0 is the original signal

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

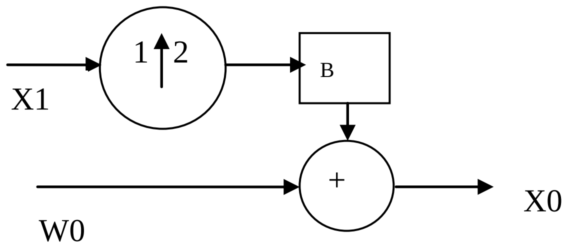
W0 is the details of the higher resolution, which is obtained using the difference between the original and the approximated signal.

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.

(2 marks)

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

(2 marks) **(5)**

b. The sampling redundancy factor for a 3 level decomposition

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

(3marks)

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

Therefore the highest value is $a/(1-r) = 4/3$.

(2 marks) **(5)**

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

(1 mark)

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 5^{th} level approximation band a5 with resolution N/32xM/32.

(1 mark)

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

(1 mark)

The quality 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

(2 marks) **(5)**

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

d0 – Very high frequency detailed sub band from the level 1 (full resolution)

d1 - high frequency detailed sub band from the level 2 (1/2 resolution)

d2 - high frequency detailed sub band from the level 3 (1/4 resolution)

a3- low frequency sub band from the level 3 (1/8 resolution)

(2 marks)

Medium magnitude coefficients in mid frequency sub bands (d1, and d2). The coefficients should correspond to textured area so that the modifications are not visible.

(1 mark)

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.

(2 marks) **(5)**

GCKA