EEE6218 2014-15
Solutions

1.

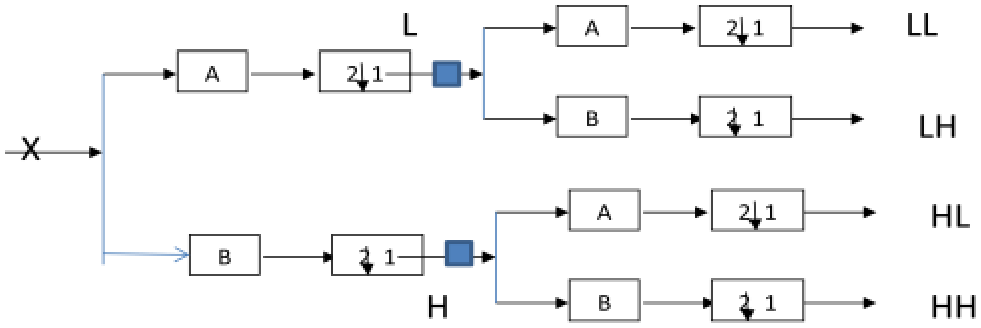
a. X_L consists of low passed half resolution representation of the signal. It provides an approximation of the input signal.

(1 mark)
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.

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

(3)

b.



Row operations column operations

The blue box represents transpose operator.
(1 mark)

The input image, X, is first transformed using the 1D transform on rows to get 2 subbands, L (Low frequency) and H (High frequency).

Then each subband transposed and apply the same 1D transform again on the new rows (effectively on columns of the initial orientation) to get the 4 new subbands, LL (Low-Low frequencies in both directions), LH (Low frequency on rows anf high frequency in columns) ,HL (High frequency in rows and Low frequency in columns) and HH (High-High frequency in both directions).

(2 marks)

(3)

c. Separabale filtering and non-separable filtering are two different approaches for operating filters and transforms on multidimensional data.

In separable filtering, a one dimensional filter is applied on each of the dimension separately. For example, separable filtering for an image involves applying the filter first on rows followed by applying the same filter in columns. The order of operation does not matter if the filters are linear time invariant.

In non-separable filtering, the corresponding multidimensional filter is designed and applied on the multidimensional data considering all dimensions at the same time. For example, for an image now the filter will become a 2D matrix.
(3 marks)

Two advantages: (any two of the following)
1. a fewer number of multiplications – so less complex
2. ability to design non-linear filters
3. ability to design filters to capturespecific features in multidimensional domains

(1 mark)

(4)

d. Two filters
A= {a,b,c,d} and
B={p,q,r,s}

To get LL -apply A on rows and A on columns
Non- separable filter AA = Croneckor product (A, A')
= Croneckor product ({a,b,c,d}, {a,b,c,d}')

$$[a \quad b \quad c \quad d] \otimes \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \equiv \begin{bmatrix} aa & ba & ca & da \\ ab & bb & cb & db \\ ac & bc & cc & dc \\ ad & bd & cd & dd \end{bmatrix}$$

(1.5 marks)

To get LH -apply A on rows and B on columns
Non- separable filter AB = Croneckor product (A, B')
= Croneckor product ({a,b,c,d}, {p,q,r,s}')

$$[a \quad b \quad c \quad d] \otimes \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix} \equiv \begin{bmatrix} ap & bp & cp & dp \\ aq & bq & cq & dq \\ ar & br & cr & dr \\ as & bs & cs & ds \end{bmatrix}$$

(1.5 marks)

(6)

To get HL -apply B on rows and A on columns

Non- separable filter BA = Croneckor product (B, A')
= Croneckor product ({p,q,r,s}, {a,b,c,d}')

$$[p \quad q \quad r \quad s] \otimes \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \equiv \begin{bmatrix} pa & qa & ra & sa \\ pb & qb & rb & sb \\ pc & qc & rc & sc \\ pd & qd & rd & sd \end{bmatrix}$$

(1.5 marks)

To get HH -apply B on rows and B on columns
Non- separable filter BA = Croneckor product (B, B')
= Croneckor product ({p,q,r,s}, {p,q,r,s}')

$$[p \quad q \quad r \quad s] \otimes \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix} \equiv \begin{bmatrix} pp & qp & rp & sp \\ pq & qq & rq & sq \\ pr & qr & rr & sr \\ ps & qs & rs & ss \end{bmatrix}$$

(1.5 marks)

e. The filterbank can be first applied for temporal decomposition on time domain. This means applying the filterbank on each of the corresponding pixel in all the frames. Affter that each frame is considered individually and performed the 2D transform as a separable application of the 1D transform on rows and columns. This would make a t+2D decomposition. Since it is separabale one could also apply 2D transform first, followed by the temporal transform, leading to a 2D+t decomposition.

(3 marks)

The main problem in this type of transform is that on temporal domain high pass would result in large magnitudes of high frequency content due to object motion. The temporal low pass would not represent the true approximation of the signal as the motion is not considered in the transform

(1 mark)

(4)

2. **a.** In R-G-B format, each colour plane contains the same amount of data and would have the same data rate and bits per sample. But in Y Cb Cr, we can exploit the human eye's tolerance to reduced resolution in chrominance layers. Therefore they can de down sampled and the data rates can be reduced.

Y Cb Cr format also useful in transition from back and white tv transmissions to colour tv transmissions. **(3)**

b. low spatial frequency regions - due to quantization artificial contours will appear **(1 mark)**

high frequency regions- for low quantization, such artificial contours won't appear. But for high quantization, some high frequency details might be lost **(2 marks)**

(3)

C. **i.** Specify the required resolution of the display in terms of number of pixels.
Vert. Angle subtended = arctan(1/12) = 4.76
Horiz. Angle subtended = arctan(4/3 x 1/12) = 6.34
(1 mark)

Vert cycles = 4.76 x 150 = 714
Horiz cycles = 6.34 x 150 = 951
(1 mark)

Nyquist criterion would require doubling of these figures to give 1428 x 1902 pixels on screen. Assume BW is limited and use Kell Factor so use 1020 x 1359 pixels
(1 mark)

(3)

ii. Can sense 38k hues – i.e, 15.2 bits – call it 16 bits.
(1 mark)

In colour + luminance representation --- 200 luminance i.e., 8 bits for luminance
(1 mark)

Memory per picture = 1020x 1359 x 3 bytes = 3.96 Mbytes
(1 mark) **(3)**

iii. Can perceive flicker up to 80 Hz – so Nyquist would suggest 160 Hz frame rate but use interlacing.
(1.5 marks)

Bit rate (bandwidth) = 3.96 x 80 = 316.8 Mbytes/s = 2.47 Gbps.
(1.5 marks)

(3)

iv. Currently use 4:4:4 sampling for Luminance and chrominance channels
Can reduce sampling of chrominance –
in 4:2:0 video – which gives H/2xW/2 resolution for chrominances. **(3)**

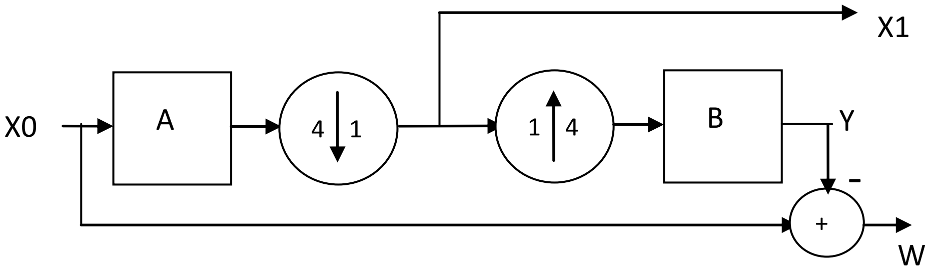
(1 mark)

In 4:4:4 total pixels = 3 HW **(1 mark)**

In 4:2:0 total pixels = (HxW) + (H/2xW/2) +(H/2xW/2) =1.5HW **(1 mark)**

v. Any standard that can be useful for digital video broadcasting (DVB): MPEG-2 or H.264 or AVC (Advanced Video Codedc) **(2)**

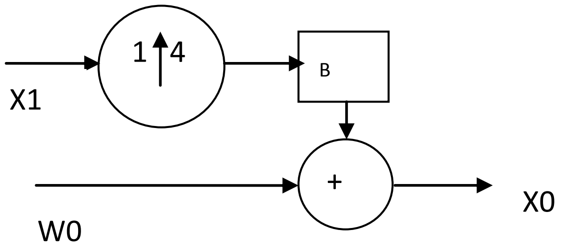
3. a.



(1 mark)

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.
(1.5 marks)

Reconstruction



(1 mark)

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

(4)

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

(2 marks)

(4)

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

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.
(1 mark)

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.

(1.5 marks)

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

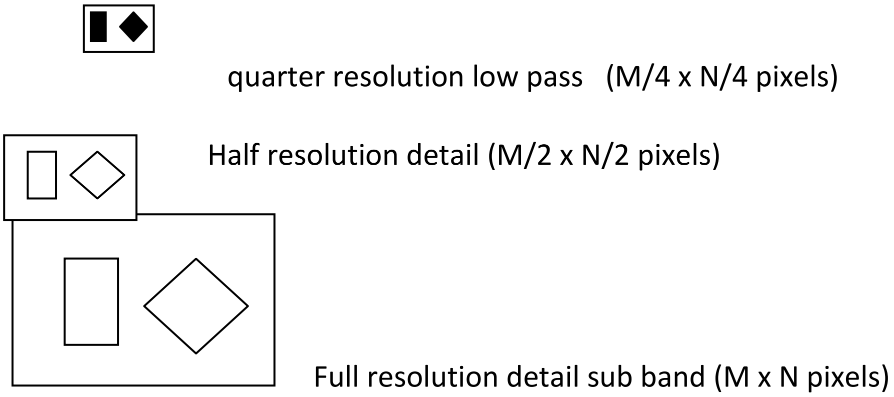
The spatial resolution scalability can be achieved by combined decoding as follows

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

(1.5 marks)

(4)

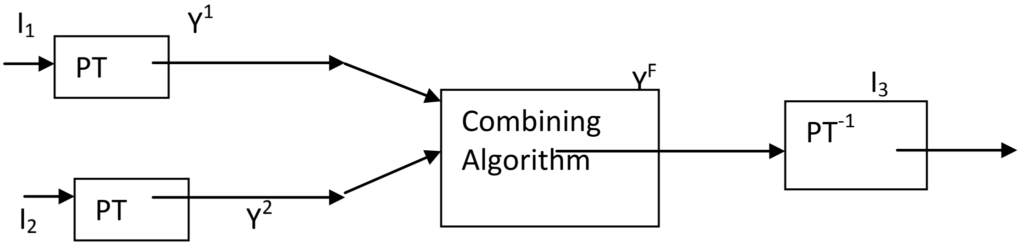
d.


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)

(4)

Marks deistribution:
Correct dimensions - **(1 mark)**
Correct description - **(1 mark)**
Correct approximation features **(1 mark)**
Correct detail features **(1 mark)**

e.



(1 mark)

Images 1 and 2 are decomposed into sub bands using the Pyramid transform (PT).

(0.5 marks)

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)

(2 marks)

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

(0.5 marks)

(4)

4. 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. **(1 mark)**
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)

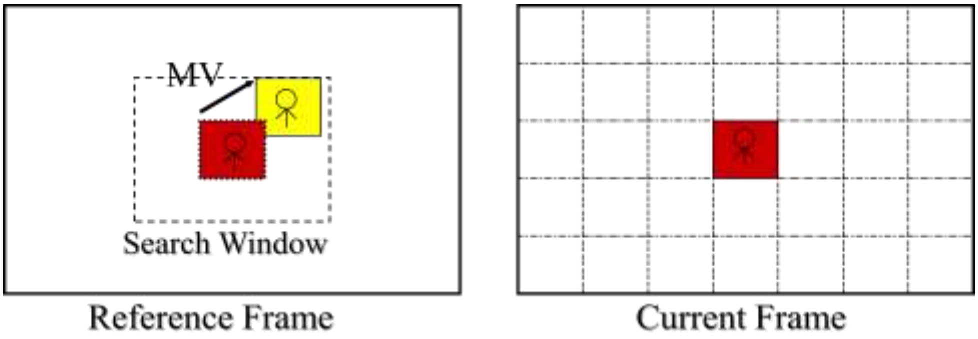
4. Coding Redundancy: This is present in images when the probability of occurrence of symbols has not been taken into account when assigning binary codes. Entropy coding – (VLC, Huffman coding , run length coding , arithmetic coding) is used
(1 mark)

(4)

b. The answer should contain the following:
The DWT decorrelates the image.
(1 mark)
Compacts the energy of the image into a fewer number of coefficients.
Provides a multi resolution framework.
(1 mark)
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.
(1 mark)
Resolution scalability is achieved by coding/deciding coefficients from the highest level of decomposition to the lowest decomposition.
(1 mark)

(4)

c.



(4)

(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.

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

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

(1.5 marks)

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.

(1.5 marks)

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

(1 mark)

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

(1 mark)

(2)

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

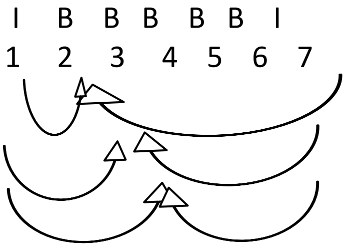
(1 mark)

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

(1 mark)

(2)

f.



(1 mark)

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

(1 mark)

(2)

g. 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 **(2)**

Method 4 has a higher delay compared to the method 3.
(1 mark)
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)
(1 mark)