First Name : Last Name :

Exam – Master VICO – UE Multimedia Communication

September 2018 – duration 1h30 – answer **directly on the exam subject**

Authorized documents: handwritten and personal notes.
Laptop and phone are not authorized.

Indicative scale (/ 20): Exercise part on 14 points, Questions part on 6 points.

Exercise (on Motion estimation and compensation):

We are interested by motion estimation / compensation approaches with a sequence of monochrome frames where a gray level is coded on 8 bits. Motion estimation is based on Block Matching.

First, we consider a particular step of the motion estimation algorithm between 2 successive frames (called *frame(t-1)* and *frame(t)* respectively) of the sequence:

- Figure 1 represents *b*, the (4x4) block of the *frame(t)*, *b* has to be matched:
  - the 1st row and the 1st column of the table are the position indices of the *b* 's pixels;
  - in the following, we will indicate the block pixel position by: *(index of line, index of column)*.
- Figure 2 represents *B*, the (16x16) search window of the *frame(t-1)* used for the motion estimation:
  - the first row and the first column of the table are the *B*'s pixels position indices;
  - the position of a (4x4) bloc within *B* window, will be noted $b_B[u,v]$ where *[u, v]* indicates the position of its top left pixel. For example, $b_B[7,7]$ is the block shown in Figure 2 in the center of *B*.
  - The initial position of the *b* block, namely its position in *frame(t)*, corresponds to the center of the *B* window in *frame(t-1)*.

The metric used to compute the distance between *b* and a block of *B* is an MSE (Mean Squared Error):

$$D(b,b_B[u,v]) = MSE(b,b_B[u,v]) = \frac{1}{16} \times \sum_{(i,j)=(1,1)}^{(4,4)} \left(b(i,j) - b_B(u+i-1, v+j-1)\right)^2$$

The motion vector between 2 blocks, will be represented by an arrow connecting their respective upper left corners. This motion vector will be characterized by: *(displacement according to the rows, displacement according to the columns)*.

Questions :

1. Is this a "forward" or "backward" motion estimation, explain?

2. What is the maximal possible amplitude for the estimated motion between 2 successive frames? [it is, therefore, the maximal possible distance for a pixel, this distance will be given in absolute value and in pixels].

   $|horizontal\ maximal\ amplitude| =$

   $|vertical\ maximal\ amplitude| =$

First Name : Last Name :

3. Detail the calculation of the distance between *b* block in Figure 1, and the block in the center of *B* in Figure 2. (Give the literal formula)

$$D(b, b_B[7,7]) =$$

4. The motion estimation is performed by using a "Full Search" approach (noted FS approach).
   - Plot in red on Figure 2, the motion vector obtained with this FS method, and indicate in red the block found (frame it).
   - What is the value of this motion vector?

$$\overrightarrow{V^{FS}_{t\to(t-1)}} = (\quad , \quad)$$

   - Give the computation cost involved in order to match *b* block with its nearest block in *B*:
     - Number of subtractions & additions (explain, detail):
     - Number of multiplications & divisions (explain, detail):
     - Number of search for the minimum (explain, detail):

5. In order to reduce the computational complexity, a "one-at-a-time" approach is used (noted OT approach, it means “one direction at a time”).
   - Plot in green on Figure 2 the motion vector obtained by using this method (you can exactly represent the two vectors obtained successively: first the horizontal vector, and next the vertical one), and frame in green the block found.
   - What is the value of this motion vector?

$$\overrightarrow{V^{OT}_{t\to(t-1)}} = (\quad , \quad)$$

- Give the computation cost involved in order to match *b* block with its nearest block in *B* when using this method.
  - Number of subtractions & additions (explain, detail):
  - Number of multiplications & divisions (explain, detail):
  - Number of search for the minimum (explain, detail):

6. The motion vector obtained by the full search method is used to compensate the *b* block in order to predict it from the *frame(t - 1)*.
   1. Is this a "forward" or "backward" motion compensation, explain?
   2. If the *b* bloc position in *frame(t)* is $b_{frame(t)}[100, 100]$, what is the position of its prediction in *frame(t-1)* ? (complete below)

$$\widetilde{b_{frame(t-1)}}[100 \pm \quad , 100 \pm \quad]$$

	1	**2**	**3**	**4**
1	2	2	120	120
2	2	2	120	120
3	2	2	120	120
4	2	2	120	120

Figure 1: The (4x4) block of *frame(t)*.

First Name : Last Name :

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	100	100	100	40	40	40	60	60	60	60	60	60	60	60	60	60
2	100	100	100	40	40	40	40	10	10	2	2	2	2	120	120	130
3	100	100	100	40	40	40	40	10	10	2	2	2	2	120	120	130
4	100	100	100	40	40	40	40	10	10	2	2	2	2	120	120	130
5	100	100	100	40	40	40	40	10	10	2	2	2	2	120	120	130
6	100	100	100	40	40	40	40	40	40	10	4	4	4	130	130	140
7	100	100	100	40	40	40	40	40	40	10	4	4	130	130	130	140
8	100	100	100	40	40	40	40	40	40	10	4	4	130	130	130	140
9	130	130	130	70	70	70	40	40	40	10	4	4	130	130	130	140
10	130	130	130	70	70	70	40	40	40	10	4	4	130	130	130	140
11	130	130	130	70	70	70	40	40	40	10	4	4	130	130	130	140
12	130	130	130	70	70	70	40	40	40	30	3	3	120	120	160	160
13	130	130	130	70	70	70	40	40	40	30	3	3	120	120	160	160
14	130	130	130	70	70	70	40	40	40	30	3	3	120	120	160	160
15	140	140	140	80	80	80	50	50	50	40	3	3	120	120	170	170
16	140	140	140	80	80	80	50	50	50	40	40	40	40	170	170	170

Figure 2: The (16x16) search window in f*rame(t-1)*.

Questions:

1. How much information (in *bits / second*) is there in a progressive HD video (1080x1920 resolution, 60 *fps*, 4: 4: 4 format)? (give the literal formula)

2. What is the advantage of the 4: 2: 0 or 4: 2: 2 video formats for transmission purpose? (explain)

3. What is a « Macro Block » for H.261?

First Name : Last Name :

4. With H.261, the inter-frame prediction proceeds by motion estimation-compensation. For a block coded in inter mode, what are the 2 types of information that have to be transmitted in order to permit its decoding? (Think about errors and motion)

5. What new prediction mode is introduced in the GOP by MPEG1 in comparison with H.261?

6. Let's be the following GOP used to code a frame sequence by MPEG1:
   $I_1, B_1, B_2, B_3, P_1, B_4, B_5, B_6, P_2, B_7, B_8, B_9, I_2, \ldots$
   Which frames have be decoded, and in which rank, in order to display B_6 ?

7. MPEG2 takes into account of the interleaved frames for motion estimation/compensation. What phenomenon, when an object is moving, does explain this? (Consider the spatial vicinity and temporal gap existing between even and odd fields)

8. With MPEG2, the rate distribution is not uniform between the I, P, B frames of a GOP. Rank these types of frames, from the most to the least expensive, according to their typical costs (in *bits*).

9. We consider two video sequences (or video shots) with very different contents respectively: one video is with complex textures and moving objects, on the contrary, the other video gathers still objects with simple textures.
   These 2 videos are compressed by using an MPEG2 encoder. Will the bit rates, needed to achieve an equivalent objective quality of the decoded videos, be the same for these 2 videos? Explain.

First Name : Last Name :

10. Give an example of one intra prediction mode used by H.264/AVC to predict a (4x4) block.

11. H.264/AVC uses different block partitions for its inter prediction mode. What does it mean?

12. CABAC is a complex entropic coding used as last step by H.264/AVC. Does this step will introduce reconstruction errors in the decoded video? What is its goal then? Explain.

13. Compared to MPEG2, H.264/AVC can encode SD videos with 2 times less bits for an equivalent objective quality of the decoded content. Is this statement correct?

14. The decorrelation of the (8x8) bloc content is one reason why the DCT transform is used by JPEG. Give another reason (think about visual perception).

15. In a MICD encoder, why the predictor is fed only by reconstructed values?

16. What is the transform at the core of JPEG2000?