MULTIMEDIA COMMUNICATION

INSTITUTO SUPERIOR TÉCNICO

Academic Year 2019/2020 – 2nd Semester, Responsible: Prof. Fernando Pereira

1st Exam – 19th June 2020 (Friday), 11.30am

The marks should be out before **22th June (Monday), 6pm** at the CMul Web page. If you want to check the exam scoring, please contact the course responsible by email to schedule a conversation.

The exam is **2.5 hours long**. Answer all the questions in a detailed way, including all the computations performed and justifying well your answers.

Don't get 'trapped' by any question; move forward to another question and return later. ***Good luck !***

I (1 + 0.5 + 0.5 + 1.0 + 1.0 = 4.0 val.)

a) Why is it a great idea to use motion vectors to efficiently code video data ? (R: exploit the temporal redundancy notably in the presence of motion, targeting reducing the prediction error and increasing the compression efficiency)

b) Where is motion estimation performed ? At encoder, decoder or both ? Why ? (R: encoder)

c) Where is motion compensation performed ? At encoder, decoder or both ? Why ? (R: both)

d) Which process is more complex between motion estimation and motion compensation ? Why ? (R: motion estimation)

e) What is the maximum number of motion vectors per macroblock for the H.261, MPEG-1 Video, MPEG-2 Video and H.264/AVC video coding standards ? Explain your numbers. (R: 1, 2, 2, 32)

II (1.0 + 1.0 + 0.5 + 0.5 + 1.0 val. = 4.0 val.)

Consider the JPEG standard to code photographic images with a 288×352 luminance resolution and 8 bit/sample.

a) Determine the 'price' in bits per pixel if 4:4:4, 4:2:2 and 4:2:0 color subsampling is used and no compression is used. (R: 24, 16, 12 bpp)

b) Determine the average 'price' in bits per pixel if 4:4:4, 4:2:2 and 4:2:0 color subsampling is used and compression is applied with a luminance compression factor of 10 and a chrominances compression factor of 15. (R: 1.866, 1.333, 1.066 bpp)

c) Determine the average 'price' in bits for one single luminance sample if compression is applied with a luminance compression factor of 15 and a chrominances compression factor of 20. What is the 'price' variation if a chrominance sample is coded instead ? (R: 0.533, -0.133 bpp)

d) What is the minimum number of DCT coefficients that would you have to code for the luminance component of an image with 576×720 pixels in order **no** 8×8 luminance block of the image is flat ? (R: 12960)

e) Determine the maximum number of DCT coefficients that may be coded for each luminance block of an image with 576×720 pixels, 4:2:2, if each coefficient costs, on average, 5 bits for the luminance and 4 bits for the chrominance and a maximum total number of bits of 400000 is desired; consider that luminance blocks always use 2 coefficients more that each chrominance block and additionally consider that the EOB (End of Block) word costs 2 bits. (R: 7 for luminance)

III (0.5 + 1.0 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 val. = 4.0 val.)

Consider a videotelephony communication using Recommendation ITU-T H.261. The video sequence is coded with a CIF spatial resolution and a frame rate of 10 Hz at a channel bitrate of 256 kbit/s. The video content to code is horizontally divided into two equal parts; however, while the top part is fixed, the bottom part is moving.

Since the encoder processes sequentially the macroblocks, it is observed that all bits are uniformly generated in the second half of the time interval that the encoder usually dedicates to encode each image. At the encoder, the bits wait for transmission in an output buffer.

Knowing that the first image has used 38400 bits, the second image 51200 bit, and the third image 12800 bits, determine:

a) The time instants at which the sender finishes transmitting all bits for the first, second and third images assuming t=0 is when the first image is acquired. (R: 150, 350, 400 ms)

b) The minimum size of the encoder output buffer in order all bits above are transmitted without problems. (R: 38400 bit)

c) The initial visualization delay associated to the system defined in b). (R: 250 ms)

d) In which time period does the buffer fullness grow at faster speed? (R: between 150 and 200 ms)

e) The maximum number of bits that the 5th image may spent assuming that the 4^{th} image produces 12800 bits (still assuming that images only spend bits in the bottom half). (R: 51200 bit)

f) What will happen to the user quality of experience (space and time) if the channel rate is reduced from 256 kbit/s to 64 kbit/s ?

g) What will happen to the user quality of experience (space and time) if the frame rate is reduced from 10 Hz to 5 Hz ?

IV (1.0 + 1.0 + 1.0 + 0.5 + 0.5 = 4.0 val.)

Assume that you are contacted by a company to design a digital storage system for video clips. The company requires some editing flexibility and needs to store the largest number of 4 minutes clips in a disk. The maximum access speed to the disk is 80 Mbit/s. The clips have 4K resolution with the following characteristics: 3840 x 2160 (Y), 4:2:2, 10 bit/sample at 25 Hz.

a) Assuming that you have at your disposal, providing the required video quality, a JPEG coding solution with average compression factors of 40 and 45 for the luminance and chrominances, respectively, determine the maximum access time for an image knowing that the compression factors for critical frames are 20% lower than average. (R: 61.2 ms)

b) Assuming now that you have at your disposal, providing the required video quality, a MPEG-2 Video coding solution with N=12 and M=4 with the following average compression factors:
   - I frames: 30 and 35 for the luminance and chrominances, respectively
   - P frames: 40 and 50 for the luminance and chrominances, respectively
   - B frames: 50 and 60 for the luminance and chrominances, respectively

   Determine the maximum access time for an image knowing that the compression factors for critical frames are 25% lower than average. (R: 346.26 ms)

c) Determine, justifying, which coding solution would you propose to your client if the target is only to maximize the number of clips stored in the disk. (R: MPEG-2 with a lower rate of 83.27 Mbit/s)

d) Determine, justifying, which coding solution would you propose to your client if a maximum random access requirement of 100 ms is put forward together with the requirement of maximizing the number of clips stored in the disk. (R: JPEG as MPEG-2 has a MAT>100ms)

e) How many full video clips would you be able to store in the disk for the JPEG solution if the disk has a capacity of 10 TByte (10^{12}). (R: 3404)

V (0.5 + 1.0 + 0.5 + 0.5 + 0.5 + 1.0 = 4.0 val.)

Consider the audio coding solutions as adopted in the MPEG-1 Audio standard.

a) What is the basic idea behind perceptual audio coding ?

b) What does the so-called *threshold of hearing (in quiet)* express in terms of human auditory perception. Is this precise threshold very important in practice, e.g. when coding music ? Why ?

c) Explain how does the frequency masking effect help in reducing the rate for a similar quality.

d) Explain how does the temporal masking effect help in reducing the rate for a similar quality.

e) Explain how does the use of the so-called *scale factors* help in reducing the rate for a similar quality.

f) Identify and explain the 2 main ways to perform frequency driven audio coding as adopted in the MPEG-1 Audio standard. Which specific audio coding solution does simultaneously exploit these two ways ?