**COMUNICAÇÃO DE ÁUDIO E VÍDEO**

**INSTITUTO SUPERIOR TÉCNICO**

Year 2019/2020 – 1st Semester, Responsible: Prof. Fernando Pereira

**2nd Exam – 28th January 2020, 8am (Tuesday)**

The grading should be out before **28th January (Tuesday), 5pm** at the CAV Web page and the exam checking session will be on the **28th January (Tuesday), 6pm** in room LT4.

The exam is **3 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 (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 1.0 val. = 3.5 val.)

a) Indicate two fundamental differences between the luminance signal of an analogue and a digital image.

b) Why do humans see colors less well during the night ?

c) What is a masking effect in images ?

d) What fundamentally defines the number of frames per second needed for a video signal ?

e) How does the number of pixels in an image vary when the chrominance subsampling factor changes from 4:4:4 to 4:2:2 and 4:2:0 ?

f) Assuming you have to show images under multiple different conditions, e.g. display distance to viewer and display size, would you prefer images with increased spatial resolution or increased number of bits per sample ? Why ?

II (0.6 + 0.6 + 0.6 + 0.7 + 1.0 val. = 3.5 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: 0.4 bps, -0.133 bps)

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)

III (0.5 + 1.0 + 0.5 + 0.5 + 0.5 val. = 3.0 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. (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: 150 to 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)

IV (1.0 + 0.5 + 0.5 + 0.5 + 0.5 = 3.0 val.) Consider the MPEG-1 Audio standard.

a) Consider a mono, audio signal with 44 kHz sampling rate with the usual number of bit/sample, coded with Layer 2. What would the percentage variation of the rate if: i) the signal becomes stereo; ii) the codec changes to Layer 3; iii) the sampling rate is reduced to 40 kHz. (R: 100%, -33.33%, -9%)

b) Is perceptual audio coding more based on redundancy or irrelevance exploitation ? Why ? (R: Irrelevance)

c) What is the basic reason for audio signals to have a larger bandwidth than speech signals ?

d) Why it is reasonable to say that it is rather difficult to compress audio signals without accumulating algorithmic delay ?

e) What is the Layer 3 coding tool that makes this codec more 'chirurgical' in terms of frequency processing ? Why ?

V (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 Video)

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)

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 clips)

VI (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 3.0 val.) Consider a virtual reality system based on 360 degrees video.

a) What is the main difference between virtual reality and augmented reality experiences ?

b) What is the main purpose of performing stitching in this context ?

c) Why is the spherical visual data projected into a rectangle as when using the equirectangular projection ? d) What is the so-called *viewport* ?

e) What is precisely the so-called *motion to photon delay* ?

f) What is the main visualization ‘parameter’ that the user can (indirectly) control when navigating the visual data, which is critical to select the visual data to be displayed ?