COMUNICAÇÃO DE ÁUDIO E VÍDEO

INSTITUTO SUPERIOR TÉCNICO

Year 2015/2016 – 1st Semester, Responsible: Prof. Fernando Pereira

1st Exam – 14th January 2016, 8am (Thursday)

The marks should be out before **17th January (Sunday), 8pm** at the CAV Web page and the exam checking session will be on the **18th January (Monday), 11am** 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 + 1.0 + 0.5 + 0.5 val. = 3.0 val.)

Consider the JPEG standard to code photographic images with a 576×720 luminance resolution, 4:2:0 color subsampling and 8 bit/sample.

a) How many total chrominance blocks exist in this type of image. (R: 3240 blocks)

b) Determine the average number of bits per pixel (considering both the luminance and the chrominances) that are spent to code this type of image with a luminance compression factor of 20 and a chrominances compression factor of 10. (R: 0.8 bit/pixel)

c) Determine the total number of bits that have to be spent to code an image if an average number of 8 DCT coefficients are coded per block and each coefficient costs, on average, 4 bits for the luminance and 3 bits for the chrominance; additionally consider that the EOB (End of Block) word costs 3 bits. (R: 314280 bits)

d) Why is it reasonable to say that the quantizer is the key controller of a JPEG encoder ?

e) Why is it reasonable to ask that an ideal transform for image compression provides uncorrelated coefficients ?

II (0.5 + 0.5 + 1.0 + 1.0 + 0.5 val. = 3.5 val.)

Consider a videotelephony communication using Recommendation ITU-T H.261. The video sequence is coded with a CIF spatial resolution, a frame rate of 10 Hz and a constant bitrate channel of 128 kbit/s. The bits for each coded image are uniformly generated in the time between the acquisitions of two images.

Knowing that the first image has used 19200 bits, the second image 32000 bit, and the third image 9600 bits, determine:

a) Considering that a constant bitrate channel is used, what architectural element allows the encoder to control the number of bits spent per macroblock ? (R: Quantizer)

b) The time instants at which the sender finishes to send all bits for the second and third images. (R: 400, 475 ms)

c) The minimum size of the encoder output buffer in order all bits above are transmitted without problems. (R: 25600 bit)

d) The initial visualization delay associated to the system defined in c) while justifying the formula used. (R: 300 ms)

e) The maximum number of bits that the 3rd image could have spent. (R: 12800 bit)

III (1.0 + 1.0 + 0.5 + 0.5 = 3.0 val.)

Consider the MPEG-1 and MPEG-2 Audio standards.

a) Determine the coding rate for mono audio content with a 22 kHz bandwidth and the usual number of bit/sample if coded with a Layer 3 codec to reach CD transparent quality. How many times would the rate increase if the sampling rate becomes 48 kHz and stereo audio is used. (R: 58,67 kbit/s; 2.18)

b) What are the 2 main ways that audio frequency masking contributes to reduce the bitrate when coding the audio signal ?

c) Why does the Layer 3 codec use the MDCT with an overlapping window ?

d) Why does the Layer 3 codec use a varying size window for the MDCT ?

IV (1.0 + 1.5 + 1.0 = 3.5 val.)

Consider that your company is contacted to design a videoconference system between the various main locations of a bank. The spatial resolution is CIF (352×288 luminance samples), 4:2:0, at 12.5 Hz, with the usual number of bits per sample. Assume that you have available, offering the target quality, two solutions:

1. H.261 based solution with average compression factors of 25 and 35 for the luminance and chrominance, respectively; the critical compression factors (for the images spending more bits) are 20 and 25 for the luminance and chrominance, respectively.
2. MPEG-2 Video based solution with N = M = 3 with average compression factors of 25 and 35 for the luminance and chrominance, respectively, for the I frames, and 30 and 45 for the luminance and chrominance, respectively, for the P and B frames. The critical compression factors are 75% of the average compression factors.

Assume that the transmission rate is always the same as the coding rate.

a) Determine the bitrate and acquisition-visualization delay for the H.261 based solution. (R: 550 kbit/s, 0.103 s)

b) Determine the bitrate and acquisition-visualization delay for the MPEG-2 Video based solution. (R: 483.8 kbit/s, 400 ms)

c) Assuming that your client always pretends to minimize the transmission rate, what solution from above would you select depending on the acquisition-visualization delay requirement ?

V (1.5 + 1.0 + 1.0 + 0.5 = 4 val.)

Consider the development of an application similar to YouTube.

a) What are the 3 main technologies required for the free access representation of the multimedia content involved ? For each of these technologies, define its main objective.

b) Considering the rate budget, provide reasonable estimates of the rate percentages would you allocate to the 3 main technologies referred in a) assuming video with standard definition and stereo audio ? Justify your answer. (R: e.g. 5% metadados, 80% vídeo, 15% áudio)

c) Considering the video media component, what codecs would you support on the uploading side and on the streaming side ? Consider both the cases of live broadcasting and storage for a posteriori streaming. Justify your choices.

d) What video coding standard would become more relevant to be selected if you were specifically addressing transmission environments with highly varying bandwidth ? Why ? (R: SVC)

VI (1.0 + 0.5 + 0.5 + 0.5 + 0.5 = 3.0 val.)

Consider the currently very popular H.264/AVC video coding standard.

a) What is an H.264/AVC access unit ? How many pictures may be decoded from an access unit ? (R: 1)

b) What is the purpose of H.264/AVC defining a lossless coding mode considering that it also defines a PCM coding mode ?

c) Why is it common to say that current video coding standards do not represent well non-translational motion ?

d) Why is it important to define good interpolation filters for the half and quarter sample positions ?

e) Why are the motion vectors coded using a prediction made based on the motion vectors for neighboring partitions?