COMUNICAÇÃO DE ÁUDIO E VÍDEO

INSTITUTO SUPERIOR TÉCNICO

Year 2020/2021 – 1st Semester, Responsible: Prof. Fernando Pereira

1st Exam – 18th January 2021, 8am (Monday)

The exam grading should be out before **24th January (Sunday), 7pm** at the CAV Web page. The exam checking session time will be announced in due time, and will be live or online depending on the pandemic evolution.

The exam is **2.5 hours (150 minutes) 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 + 1.0 + 0.5 + 1.0 + 0.5 val. = 4.0 val.)

Consider the JPEG standard to code photographic images with a 1080×1920 luminance resolution, 4:2:0 color subsampling and 8 bit/sample.

a) How many total pixels, samples and blocks exist in this type of image. (R: 2073600 pixels: 3110400 samples; 48600 blocks)

b) Determine the average price in bits (considering both the luminance and the chrominances) to code 1000 pixels in this type of image if a codec with a luminance compression factor of 20 and a chrominances compression factor of 25 is used. (R: 560 bit)

c) What is the average price of a luminance sample for the codec in b) ? And for a chrominance sample ? (R: 0.4 and 0.32 bit)

d) Determine the total number of bits that have to be spent to code only the luminance component of an image if an average number of 3 DCT coefficients are coded per block and each coefficient costs, on average, 4 bits; additionally consider that the EOB (End of Block) word costs 2 bits. (R: 453600 bit)

e) What does it mean saying that the entropy encoder exploits the signal statistical redundancy ?

II (0.5 + 0.5 + 0.5 + 1.0 + 0.5 + 1.0 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 (constant) channel bitrate of 64 kbit/s. The bits for each frame are uniformly generated in the time interval that the encoder usually dedicates to encode each image. At the encoder, the bits wait for transmission in an output buffer.

Answer the following INDEPENDENT questions …

a) Assuming that the buffer size is 18000 bits, what is the maximum number of bits that the first frame may produce ? (R: 24400 bit)

b) Assuming that the first frame produces 15000 bits, what is the minimum size of the buffer ? (R: 8600 bit)

c) Assuming that the first frame produces 15000 bits and the buffer size is 18000 bits, what is the maximum number of bits that the second frame may produce ? (R: 15800 bit)

d) Assuming that the first frame produces 20000 bits and the buffer size is 15000 bits, what is the maximum number of bits that the third frame may produce ? (R: 14200 bit)

e) Assuming that the buffer size is 12800, what is maximum number of bits that frame 467 may produce ? (R: 19200 bit)

f) Assuming that the buffer size is 12800, what is latest time the full set of bits for the second frame may be received at the decoder ? (R: 400 ms)

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

Consider the MP3 audio coding standard.

a) What is the first key consequence in terms of user experience if longer audio frames are coded with MP3 ? What type of positive impact could this have ? (Idea: more delay; less rate as more temporal redundancy to exploit)

b) What is the major impact in terms of coding parameters when audio masking increases the hearing threshold at a specific bandwidth range ? What would be an important practical impact at service level ? (Idea: quantization step increase; less rate for same quality)

c) Is lossy perceptual audio coding compatible with 'transparent quality coding' ? How ? (R: Yes, by removing only perceptually irrelevant information.)

d) What would happen if a MPEG-1 Audio Layer 1 decoder is asked to decode a MP3 stream ? And a MPEG-1 Audio Layer 2 decoder? (Idea: No decoding would be impossible as the later codecs have less coding tools and thus a simpler coding syntax)

IV (1.0 + 2.0 + 1.0 = 4.0 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 video 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 chrominances, 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 chrominances, 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: 550.3 kbit/s; 103 ms)

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 as a function of the acquisition-visualization delay requirement defined by the client ? (Idea: MPEG-2 Video if delay requirement > 400 ms; H.261 if delay requirement < 400 ms and > 103 ms; no solution if delay requirement < 103 ms)

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

Consider the H.264/AVC video coding standard.

a) What is the key feature of H.264/AVC hierarchical 'B-frames' that is not available in the MPEG-1/2 Video coding standards? (Idea: B frames may be used as reference frames for prediction)

b) This standard considers two lists (List 0 and List 1) of decoded frames from which (two) reference frames may be selected for prediction when coding a B-slice macroblock (one reference frame from each list). What is the intent behind including the same frames in the two lists ?

c) Why are the same frames in the two lists above ordered in a different way ?

d) This standard allows to leave a certain frame for an undetermined time in List 0. What type of practical situation may benefit from this feature ?