COMUNICAÇÃO DE ÁUDIO E VÍDEO

INSTITUTO SUPERIOR TÉCNICO

Academic Year 2014/2015 – 1st Semester, Responsible: Prof. Fernando Pereira

1st Exam – 15th January 2015 (Thursday), 8am

MEEC: The marks should be out before **18th January (Sunday), 6pm** at the CAV Web page and the exam checking session will on the **19th January (Monday), 2pm** 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 (1.0 + 1.0 + 1.0 + 1.0 = 4.0 val.)

a) What does the human Contrast Sensitivity Function (CSF) express ? In which coding tool may the CSF have a larger impact ? Why ? (R: quantization)

b) How many components are typically used to represent an image ? Which are typically these components assuming that the images will have to be transmitted at some stage ? Why ? (R: 3; luminance and 2 chrominances)

c) What is the main advantage of using a non-uniform versus an uniform quantization function, e.g. for speech signals ? Does it make sense to use non-uniform quantization also for audio coding ? (R: Yes)

d) When we speak about transparent quality, are you speaking about lossy or lossless coding ? Why ? (R: Lossy)

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

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: 150-200 ms)

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

III (1.0 + 1.0 + 2.0 = 4 val.)

Suppose that your company has been contacted to design the video system feeding a giant screen to be installed at Copacabana beach for the Opening Ceremony of the next Olympics. The digital transmission will be in HD - 1920×1152 (Y) and 960×1152 (Cr, Cb) at 25 Hz (8 bit/sample). Assume that you have available providing the necessary quality for each frame, a MPEG-2 Video codec reaching the compression factors indicated in the table below. To guarantee adequate random access, at least one frame has to be coded in Intra mode every 300 ms. Finally, to reach the compression factors in the table below, no more than three B frames should be introduced consecutively. Assuming that the intention is to minimize the bitrate to reduce the transmission costs, determine:

Frame Type	Luminance Compression	Chrominance Compression

	Factor	Factor
I	10	15
P	15	20
B	20	30

a) The best M and N values characterizing the (regular) temporal coding structure of I, P and B frames to be adopted fulfilling the requirements above while minimizing the transmission bitrate. (R: N=6, M=3)

b) The average bitrate associated to the coding structured determined in a). (R: 45.4656 Mbit/s)

c) The initial visualization delay at the receiver assuming that the transmission is made at the bitrate determined in b), the N value is the one defined in a), M=N, and the critical compression factors for the I frames (for the 'more difficult' frames) are 10% lower than the average compression factors indicated in the table above (while they are the same for the other frame types). Assume also that the coding and decoding times are negligible. (R: 305 ms)

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

Make some quick computations …

a) Determine the average number of bits per pixel (considering both the luminance and the chrominances) that are spent when coding a 4:2:0 image with 8 bit/sample and a global compression factor (for the luminance and the chrominances) of 40. (R: 0.3 bi/pixel)

b) Determine the total bitrate to transmit a movie if the video has a spatial resolution of 720×576 luminance samples, 4:2:2 subsampling format, 25 Hz (using the typical number of bit/sample), and the audio (stereo) has a sampling frequency of 44 kHz (using the typical number of bit/sample), knowing that the audio compression factor is the typical MP3 compression factor (for transparent quality) and the video compression factor is three times the audio compression factors for all video components. (R: 4.73 Mbit/s)

c) Knowing that a DVB solution may 'insert' 20 Mbit/s of source rate in a 8 MHz bandwidth channel, what is the transmitted bitrate if all the system parameters stay the same with the exception of the channel coding ratio that goes from 2/3 to 2/4 and the modulation that goes from 4-PSK to 64-QAM. (R: 90 Mbit/s)

d) Knowing that a DVB solution may 'insert' 10 Mbit/s of source rate in an 8 MHz bandwidth channel, determine the channel coding ratio knowing that the transmitted rate is twice the source rate. (R: 1/2)

V (1.0 + 0.5 + 0.5 + 0.5 = 2.5 val.)

Consider the H.264/AVC standard.

a) Knowing that the temporal distance between I/P frames is 8 frames, how many layers of temporal scalability will result if hierarchical B frames are used ? Show it with a sketch.

b) Why can we say that the transform solution is a hierarchical transform ?

c) What is the meaning of cascading quantization ?

d) Why it may not be a good idea to simply select as the best coding mode the one associated to the lowest prediction error ?

VI (0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 2.5 val.)

Consider the HEVC standard.

a) What is the first entity used to divide and process each frame ? (R: CTU)

b) At what level/entity is performed the Inter versus Intra coding decision ? (R: CU)

c) Which HEVC partition entity resembles more to the macroblock entity used in the previous H.264/AVC standard ? Why ? (R: CU due to Inter/Intra coding mode decision)

d) What is biggest novelty in terms of prediction block partition regarding the previous standards ? (R: asymmetric splitting)

e) Why is the transform bypass coding mode particularly useful for graphics content ?