SGN-12007 Introduction to Image and Video Processing

Sample Exam (Shared with Students)

Instructor: Professor Moncef Gabbouj

Last Name (print in bloc letters): ____________________
First Name (print in bloc letters): ____________________
Student Id. Number: ____________________

Instructions: This is a closed-notes and closed book exam**. Exam papers must be returned.** Please do not talk to anyone during the exam. If you get caught cheating, your exam paper will be confiscated, you will fail the course and your action will be dealt with according to TUT Code of Honor. **Answer the questions in the space provided and use the back of the page if needed. Mark all correct answers, there might be more than one correct answer**. **Additional papers WILL NOT be graded. Exam proctor: Collect all the exam papers! Do not accept additional pages!**

Problem 1 (4 marks):

1.1) Many digital images do not fully utilize the dynamic range ($0, L-1$) and this can be verified by looking at their histogram and observing that some of the bins near 0 and near $L-1$ are totally empty. To fix this problem, one can apply histogram modification to the original image f. Which of these mathematical expressions g can spread the intensity values in the original image so that the lowest intensity value is 0 and the highest intensity value is $L-1$. Mark all correct answers, there might be more than one correct answer.

(a) $g = f - f_{min}$ (f_{min} is the minimum intensity value of image f)
(b) g = $(L-1)/f_{max}$ (f_{max} is the maximum intensity value of image f)
(c) g = $(L-1)(f - f_{min}) / max(f - f_{min})$ (where the *max* is computed over all values of f).
(d) g = $(L-1)(f - f_{min})/(f_{max} - f_{min})$

Answer: (c) and (d)

Let f denote the original image. First subtract the minimum value of f denoted $f_{\min}$ from f to yield a function whose minimum value is 0:

$$g_1 = f - f_{\min}$$

Next divide g_1 by its maximum value to yield a function in the range $[0,1]$ and multiply the result by $L-1$ to yield a function with values in the range $[0, L-1]$

$$\begin{aligned} g &= \frac{L-1}{\max(g1)} g1 \\ &= \frac{L-1}{\max(f - f_{\min})}(f - f_{\min}) \end{aligned}$$

Keep in mind that $f_{\min}$ is a scalar and f is an image.

Accept also (d) : g = $(L-1)(f - f_{min})/(f_{max} - f_{min})$

1.2) Given an image *f*. Image *g* is the histogram equalized version of *f*. Let image *h* represent the equalized version of image g (round-off errors are neglected). Which of the following is the correct answer?

(a) $h = g$
(b) $h = f$
(c) $h = g \,.\, f$
(d) $h \neq g$

Answer (a)

Let $n = MN$ be the total number of pixels and let n_{r_j} be the number of pixels in the input image with intensity value r_j. Then, the histogram equalization transformation is

$$s_k = T(r_k) = \sum_{j=0}^{k} n_{r_j} / n = \frac{1}{n} \sum_{j=0}^{k} n_{r_j}$$

Because every pixel (and no others) with value r_k is mapped to value s_k, it follows that $n_{s_k} = n_{r_k}$. A second pass of histogram equalization would produce values v_k according to the transformation

$$v_k = T(s_k) = \frac{1}{n} \sum_{j=0}^{k} n_{s_j}$$

But $n_{s_j} = n_{r_j}$, so

$$v_k = T(s_k) = \frac{1}{n} \sum_{j=0}^{k} n_{r_j} = s_k$$

which shows that a second pass of histogram equalization would yield the same result as the first pass. We have assumed negligible round-off errors.

Problem 2 (4 marks):

Given

$$p(x,y) = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

2.1) When $p(x,y)$ is used as a filer mask, write the expression of the filtered image $g(x,y)$ in the spatial domain.

Answer:

The Laplacian kernel from Eq. (5-90) yields a filtered image in the spatial domain based on the expression

$$g(x,y) = 4f(x,y) - [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)]$$

2.2) When $p(x,y)$ is used as a mask for filtering an image, what kind of operation this mask represents?

Answer: This is a Laplacian kernel

2.3) Show that the Fourier transform of $p(x,y)$ is given by

$$P(u,v) = 4 - 2\cos\frac{2\pi u}{M} - 2\cos\frac{2\pi v}{N}$$

(hint: the translation property of the DFT might be handy here)

$$f(x,y)e^{j2\pi(\frac{u_0 x}{M}+\frac{v_0 y}{N})} \leftrightarrow F(u-u_0, v-v_0)$$

$$f(x-x_0, y-y_0) \leftrightarrow F(u,v)e^{-j2\pi(\frac{u_0 x}{M}+\frac{v_0 y}{N})}$$

Answer:

The Laplacian kernel from Eq. (5-90) yields a filtered image in the spatial domain based on the expression

$$g(x,y) = 4f(x,y) - [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)]$$

From property 3 in Table 4.4,

$$\begin{aligned} G(u,v) &= \left[4 - \left\{e^{j2\pi u/M} + e^{-j2\pi u/M} + e^{j2\pi v/N} + e^{-j2\pi v/N}\right\}\right]F(u,v) \\ &= H(u,v)F(u,v) \end{aligned}$$

where

$$\begin{aligned} H(u,v) = P(u,v) &= 2\left[2 - \cos(2\pi u/M) - \cos(2\pi v/N)\right] \\ &= 4 - 2\cos(2\pi u/M) - 2\cos(2\pi v/N) \end{aligned}$$

Consider the behavior of this function in the interval $[-M/2, M/2]$. The function is zero at the origin (eliminating the dc term) and passes frequencies on either side of it. Thus, it behaves as a highpass filter transfer function, as expected.

Problem 3 (3 marks):

Consider the following filter which acts on an image g_r in a local window of size *m* by *n* (assume that both *m* and *n* are odd numbers). The filter removes the lowest *d/2* and the highest *d/2* samples from the local input pixels (to obtain a local set of pixels S_{xy}) and averages the remaining samples. The output image is *f*:

$$f(x,y) = \frac{1}{mn-d} \sum_{(s,t)\in S_{xy}} g_r(s,t)$$

3.1) What is the name of the filter when *d=0*?

Answer: the filter becomes the arithmetic mean filter.

3.2) What is the name of the filter when *d/2=(mn-1)/2*?

Answer: the filter becomes the median filter.

3.3) How does the filter behave for other values of *d* (other values than *0* and *mn-1*) and what is the filter commonly called?

Answer: the filter combines order statistics and averaging filters and it is called Alpha-trimmed mean filter.

Problem 4 (6 marks):
4.1) What is the blind spot in the human eye?

Answer:
The blind spot in the human eye is an area on the retina where there are NO receptors (no rods and no cones).

4.2) What are rods, what is the shape of a rod, how many rods are there in a human eye, and what type of vision they produce?

Answer:
Rods are light receptors in the retina of the eye. Rods are long and slender receptors (counting 75-150 Million). Rods are responsible for scotopic vision.

4.3) What are cones, what is the shape of a cone, how many cones are there in a human eye, and what type of vision they produce?

Answer:
Cones are light receptors in the retina of the eye. Cones are shorter and thicker (counting 6-7 million). Cones are responsible for photopic vision.

4.4) Define subjective brightness.

Answer: Subjective brightness is the perceived intensity which is a logarithmic function of the light intensity incident on the eye.

4.5) Give an example of a color space that separates intensity from color information.

Answer: HSI and YUV (or YCbCr) where H and Y represent the intensity value and the rest represent the color information.

4.6) How does the following pixel with normalized RGB color values (½, 1, ½) appear in a color image?

(a) a pure green with a boost in intensity
(b) a dark green
(c) almost white since it combines all three colors
(d) a combination of red and blue color

Answer:
(a) The pixel appears like a pure green with a boost in intensity due to the additive grey component:
(½, 1, ½) = (½, ½, ½) + (0, ½, 0) = achromatic mid-level grey + mid-level green color.
Alternatively, can compute the hue component (H) of this point using RGB to HSI conversion formula and get H=120° (corresponding to Green color).

Problem 5 (7 marks):

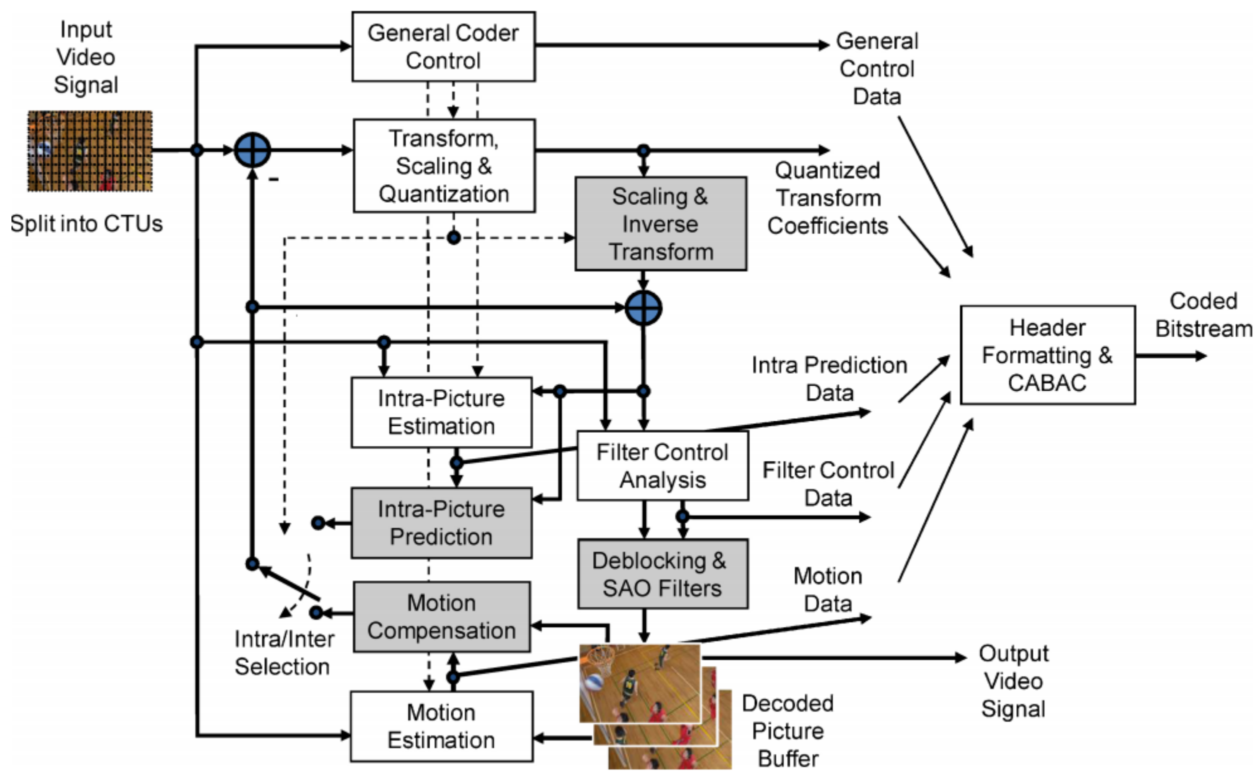
5.1) List three major benefits of video when it is represented in digital format.

Answer:
Exactness: Exact reproduction without degradation and accurate duplication of processing result;
Convenient & powerful computer-aided processing, which can perform rather sophisticated processing through hardware or software; and
Easy storage and transmission: DVD can store a three-hour movie, and transmission of high quality video through network in reasonable time

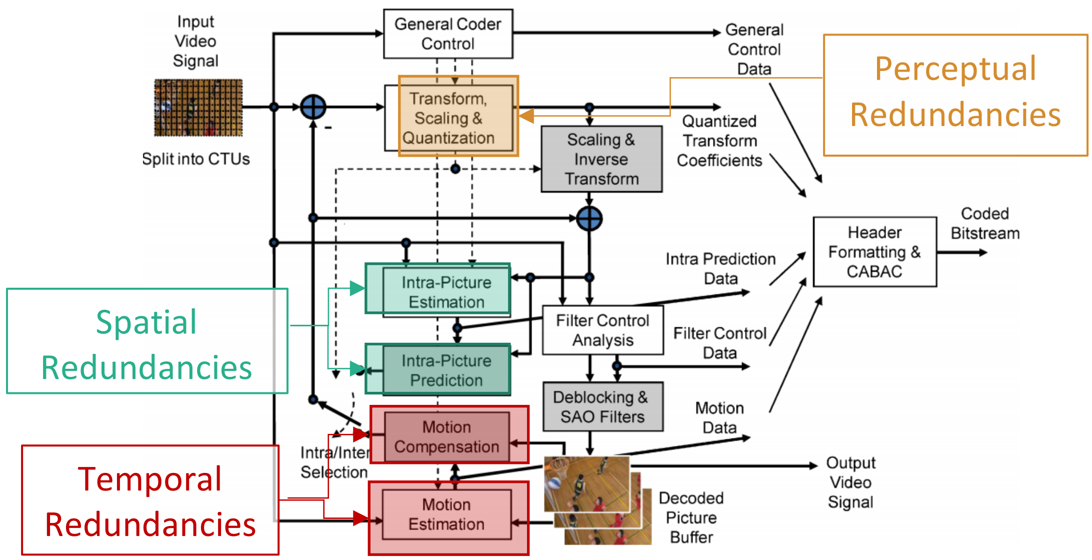
5.2) Most modern video encoders are called "hybrid video codecs" because they remove three types of redundancies. What are the three types of redundancies removed by hybrid encoders?

Answer:
The three types of redundancies are called Spatial, Perceptual, and Temporal redundancy.

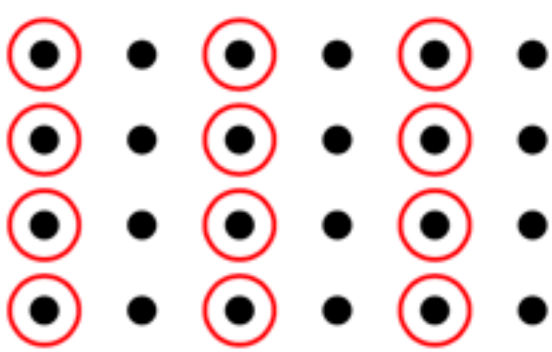
5.3) Specify on the diagram below which bloc or blocs remove which redundancy.

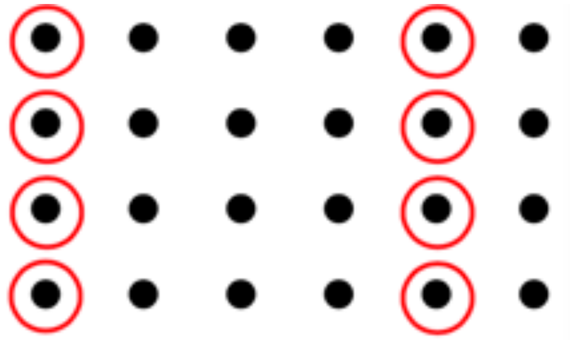


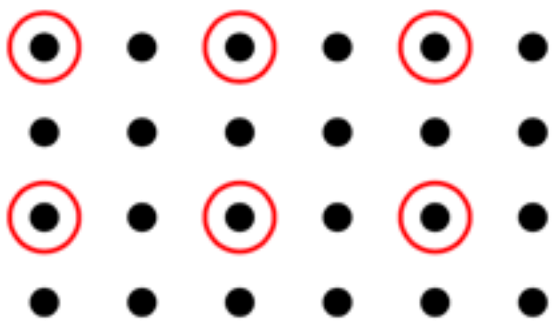
Answer:
(a) perceptual redundancy is removed in the Transform and quantization bloc, (b) intra-frame/picture estimation/prediction removes spatial redundancy; while (c) inter-frame compression through Motion compensation/estimation blocs removes temporal redundancy.



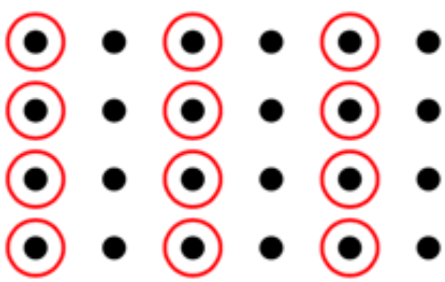
Problem 6 (4 marks):
6.1) Label the chrominance subsampling formats shown below with the representation (J:a:b) and tell what is the down-sampling factor used for the color component (the large solid black dots are Y samples and red circles are color samples, even if the circles are printed in black and white!)





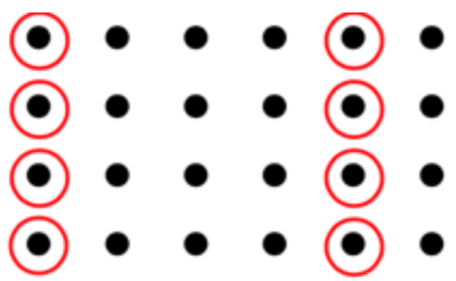


Answer:



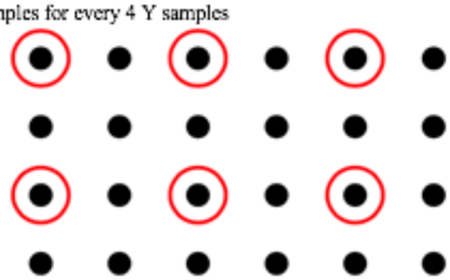
4:2:2

he color information is down-sampled by a factor of 2
orizontally from the full resolution intensity image
e. there are two Cb and two Cr samples for every 4 Y samples



4:1:1

4:1:1 sampling yields 1 Cb and 1 Cr sample for every
4 horizontal Y samples, notice the asymmetric resolution



4:2:0

The color information is down-sampled by a factor of 2 horizontally and vertically from the full resolution intensity image,
again 1 Cb and 1 Cr for every 4 Y samples, notice the more symmetric resolution

6.2) Draw the sampling structure in which the color components of the signal are sampled at the same rate as the luminance signal and provide its (J:a:b) representation.

Answer:
It is 4:4:4 and it is the same as 4:2:2 with no down-sampling (no skipping, all Y samples are red-circled)