Questions 1**

How to go from analogue to digital?

The main technique is **sampling**. Discretization in time. Take a signal and every so often take a sample of it. This is the basis of digital signal processing.

Sampling theorem

$$f_s > 2f_{max}$$

That means that the sampling frequency must be greater than twice the maximum frequency of the signal.

What is a problem of subsampling images?

Strange effect like Moiré patterns. This is because the sampling frequency is too low.

What is quantization?

Discretization in amplitude. The value of the sample is rounded to the nearest value in a finite set of values.
This is done to reduce the number of bits needed to represent the signal.

What is multimedia processing?

Multimedia processing is a discipline that takes advantages on all types of media, that can be written, recorder, smelled, etc, to derive solutions.

How humans perceive images?

Humans perceive images by the light that is reflected from the objects. The light is captured by the eyes and the brain processes the information.

What is retinotopy

Retinal images on the brain

What does cochlea do?

The cochlea is a spiral-shaped cavity in the inner ear that produces nerve impulses in response to sound vibrations. Converts time dependent sound information to its frequency spectrum like fourier transform.

What is HVS

Human Visual System

What are the 2 types of visual sensors in humans?

- Rods: Sensitive to black and white information. Around 75-150 millions and they have nice night vision.
- Cones: Sensitive to color information. Around 6-7 millions. Care about photopic vision, that is the vision in the daylight.

How are cones and rods distributed over the area of the eye?

Rods are distributed over the retina, cones are more int he back of the eye.

What is the blind spot?

The blind spot is the point where the optic nerve leaves the eye. There are no rods or cones in this area, and so there is no information detected in this area.

What is the fovea

Fovea provides nice high resolution vision and occupies a small area of the retina.

What is brightness adaption?

HVS can adapt to a large range of brightness levels. To do so, it changes the sensitivity of the rods and cones to the light.

What is intensity discrimination?

The ability to distinguish between different intensities of light. For example, the ability to distinguish between different shades of grey.
Weber ratio is the minimum difference in intensity that can be detected.
The lower the ratio, the better the discrimination.
To increase the ratio, for humans is good to have a good illumination in the background.

Simultaneous contrast

A region perceived brighter when surrounded by a darker region doesn't depend only on the intensity of the region, but also on the intensity of the surrounding region.

How to calculate the size of an object in the retina?

We have a formula that takes in account:

- distance of the object *d*
- Real object height *H*
- focal length of the eye *f*

$$h = f\frac{H}{d}$$

What is the motivation of Vision Problems?

Images are not convering on the retina.

WHat is Camera Focal Length f?

The distance between the lens and the image sensor when the subject is in focus, usually stated in millimeters (e.g., 28 mm, 50 mm, or 100 mm).

$$\frac{h}{f} = \frac{H}{d} \iff H = d\frac{h}{f}$$

What does camera aperture do?

It controls the amount of light that reach the sensor.

$$F - stop = \frac{focal\_length}{diameter}$$

What's the composition of a camera?

1. Lens: Focus the light
2. Visual system to view the image
3. Aperture: Controls the amount of light
4. Shutter: Controls the time of exposure of the sensor to the light
5. Sensor: Captures the light

What are noise and saturation problems?

When we have a digital image, we have a finite number of bits to represent the intensity of the light.

- Saturation problem: The image is too bright and we cannot distinguish anymore the different intensities of light.
- Noise problem: The image is too dark and we cannot distinguish anymore the details.

What are the two copmonents for image acquisition?

1. Illumination: Amount of source illumination that is incidented on the scene
2. Reflectance: The amount of light that is reflected from the object.

The final formula for each pixel is:

$$f(x,y) = i(x,y) * r(x,y)$$

And we want this to be normalized between 0 and 1 or 0 and 255.

What is spatial resolution?

The spatial resolution is the *level of quality* of the image. It is the number of pixels that we have in the image.

We have a *MxN* image of M rows and N columns and each pixel should be represented by a $L = 2^k$ (k bit/pixel)

What is the total number of bit to represent a black and white image?

$$b = M \times N \times k$$

What is the image is colored?

We have 3 channels: Red, Green and Blue. Each channel has a bit depth of k.

$$b = M \times N \times 3 \times k$$

What happens if we have an image NxN and we lower the value of N?

The image will be more pixelated, because the amount of pixels is lower for the same area.

What happens if we lower the value of L?

We will have less representation of the intensity of the light, so the image will have less shades of grey (or color if colored).

Which was one of the first instrument for 3D visualization?

The stereoscope.

What is the range of human hearing?

20 Hz to 20 kHz

How does the microhpone work?

It transform a sound wave into an electrical signal.

How does a speaker work?

It transforms an electrical signal into a sound wave.

Usual values for voice and music sampling

Voice: 8000 samples/s x 8 bits/sample = 64 kbps
Music: 44100 samples/s x 16 bits/sample = 705 kbps

How to achieve compact representation in multimedia?

1. Irrelevancy: Remove the irrelevant information that are not perceived by the human.
2. Redundancy: Remove the redundancy in the information or the predictable information.

How does frequency representation for image compression work in JPEG?

The image is transformed into the frequency domain using the Discrete Cosine Transform (DCT). The DCT is a lossy transformation that transforms the image into a set of coefficients that represent the frequency of the image. Now *quantization*, that is the process of rounding the values of the coefficients to the nearest value in a finite set of values, is applied to reduce the number of bits needed to represent the image. At the end, data is compressed.

How does MPEG explores temporal correlation?

MPEG exploit the temporal correlation by using some frames as reference frames and then *predicts* the next frames based on the reference frames. The less qualitative frames and part of the frames are the ones in which the motion is higher.

Questions of the test 2021

1

The human visual system (HVS) allows to see both with very low, as well as with high, illumination intensities.

a) What is the type of photoreceptors supporting night vision?

The rods

b) Which visual process explains the fact that a period of time is needed, e.g., when entering an illuminated room from the dark outside during the night, for the HVS to adapt to the new illumination conditions?

The brightness adaptation, that is the ability of the HVS to adapt to a large range of brightness levels.

2

Consider an original scene (left) and acquired image (right). What can explain the acquired image characteristics? How should the acquisition system be modified to solve the problem?

The image has a very low spatial resolution. A possible solution is to increase the pixel density of the camera.

3

Consider a CCD sensor with a resolution of 1000x1000 pixels, where each picture element occupies an area of 10□m x 10□m. If the size of the object to be imaged is 1m x 1m, and a lens with focal distance of 10 mm is used, how far away from the object should the acquisition system be placed so that the object image fully occupies the sensor? (Start by drawing a figure illustrating the image formation process)

So let's write the elements we have here:

- f = 10 mm = 0.01 m
- h = 10 um = 0.00001 m
- H = 1 m

We are searching the distance of the object.
We know that

$$\frac{h}{f} = \frac{H}{d} \iff d = f\frac{H}{h}$$

$$d = 0.01\frac{1}{1000 \times 10^{-6}} = 1m$$

We are doing *1000* x *10^-6* because we have 1000 pixels and each pixel is 10 micrometers.

4

Consider an application that uses and audio signal with maximum frequency of 8 kHz. This signal is digitized using 8 bits to represent the value of each sample. The application needs to transmit the resulting digital signal using a channel with maximum bitrate of 50 kbit/s. Can it be done? How? Briefly discuss your proposed solution.

So, we have max freq = 8000 Hz and 8 bits/sample. We need to calculate the bitrate.

$$b = 8000 \times 2samples/s \times 8bit/sample = 128kbps$$

Note that we are doing 2 samples/s because we need to sample the signal at least twice the maximum frequency.

So we can't transmit the signal using the channel with 50 kbps. We need to compress the signal.

5

Explain why, in MPEG video coding, the order of image/frame encoding and visualization is not the same.

The key here is the temporal redundancy. Since a lot of frames are similar, we can use some frames as reference frames and then predict the next frames based on the reference frames. This is why the order of encoding and visualization is not the same.

Questions of the test 2022

1

For telephonic communications an acoustic speech signal can be converted into an analogue electrical signal using a microphone. The resulting analogue speech signal to be transmitted can be limited to the frequency range from 20 Hz to 4000Hz.

a) What needs to be done so that this speech signal can be transmitted using a digital communications system? (specify the main operations involved)

The main operation is the sampling of the signal. We need to sample the signal at least twice the maximum frequency. So we need to sample the signal at 8000 Hz. Then we need to quantize the signal, that is, round the values of the samples to the nearest value in a finite set of values. This is done to reduce the number of bits needed to represent the signal.

b) What is the typical bitrate (bit/s) resulting from the conversion from analogue to digital speech? (justify all the values you consider)

We know that we can use 8 bits to represent the value of each sample. So we have 8000 samples/s x 8 bits/sample = 64 kbps.

2

The human visual system captures the light coming from a scene which reaches the retina.

a) What are the two types of light “sensor” cells present in the retina, and what are the main characteristics of each one?

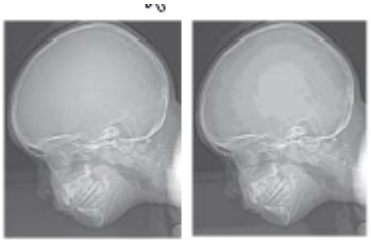
We are talking about rods and cones. The first are good for low light condition and the seconds are good for daylight conditions. and color vision.

b) Are there areas in the retina that are favoured in terms of human vision? Or areas of the retina that do not contribute to human vision? (briefly explain why)

There is a particular part of the retina called fovea that provides high resolution vision.
However, there is even a blind spot in the retina where the optic nerve leaves the eye. There are no rods or cones in this area, and so there is no information detected in this area.

3

Consider an original scene (left) and an acquired image (right). What can explain the acquired image characteristics? How should the acquisition system be modified to solve the problem?



We can see that the number of bits used to represent the intensity of the light is too low. We need to increase the number of bits to represent the intensity of the light.

4

Consider a camera with a small CCD sensor, with a resolution of 200x200 pixels, where each pixel occupies an area of 5□m x 5□m. The camera is used to take a picture of a 10cm x 10cm object, using a lens with focal length of 1 cm. How far away from the object should the camera be placed so that the object image completely occupies the camera sensor area? (Start by drawing a figure illustrating the image formation process; 1 □m = 1 x 10-6 m)

It's the same as before:

- f = 1 cm = 0.01 m
- resolution = 200 x 200 pixels
- H = 10 cm = 0.1 m
- h = 5 um = 5 x 10^-6 m

Now we need to calculate the distance of the object:

$$d = f\frac{H}{h}$$

$$d = 0.01\frac{0.1}{200 \times 5 \times 10^{-6}} = 1$$

We are doing *200* x *5* x *10^-6* because we have 200 pixels and each pixel is 5 micrometers.

5

Consider a digital image of size 1000 x 1000 pixels, each pixel represented using 8 bit/pixel. Would it be possible to store this image using at most 80 kbit? What type of operations can be considered for this purpose (if any)?

We have 1000 x 1000 pixels and each pixel is represented using 8 bits. So we have:

$$b = 1000x1000x8 = 8Mbits$$

This is if the image is black and white. If the image is colored, we have 3 channels and we need to multiply by 3:

$$b = 1000x1000x8x3 = 24Mbits$$

So, in order to solve this problem, we need to compress the image. We can exploit redundancy and irrelevancy to compress the image.