Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M1. Introduction to human and computer vision** **Final exam**
Date: December 1st, 2014 **Time: 2h30**
Teachers: Marcelo Bertalmío, Ramon Morros, Nava Rubin, Javier Ruiz, Philippe Salembier, Verónica Vilaplana

- Books, lecture notes, calculators, phones, etc. are not allowed.
- All sheets of paper should have your name.
- Answer each part in a separate sheet of paper.
- All results should be demonstrated or justified.

1: Why was the XYZ colorspace introduced alongside RGB? How is the function that transforms an RGB triplet into XYZ, linear or non-linear? Explain why, for any trichromatic display device, there are always colors that we can perceive but that the display is not able to reproduce.

Answer in section III of the course notes.

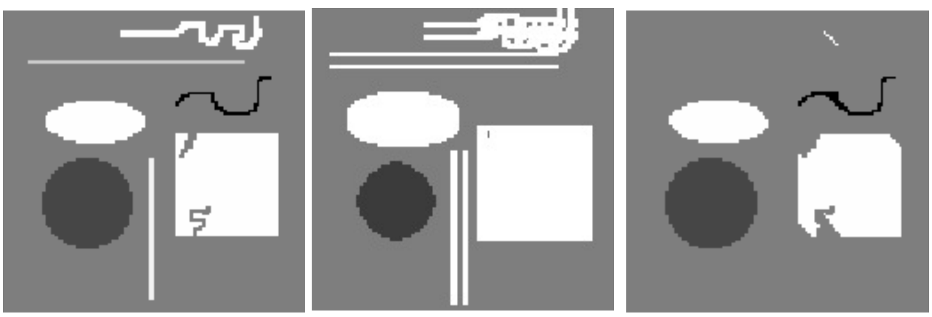
2: What is the color constancy property of the human visual system? How do cameras emulate it?

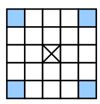
Answer in section X of the course notes.

3: Histogram equalization redistributes the gray levels of images. Is it possible to increase the entropy of a digital image through histogram equalization? Why?
Answers: No, because in the case of digital images, the gray levels of lower probability will be merged and information will be lost.

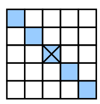
4: State the three important algebraic properties of an opening and define mathematically each property.
Answers:
Increasing: if $x \leq y \implies \gamma(x) \leq \gamma(y)$
Anti-extensive: $\forall x, \gamma(x) \leq x$
Idempotent: $\forall x, \gamma(\gamma(x)) = \gamma(x)$

5: An original test image has been processed with two morphological operators T_1 and T_2. The resulting images are shown in the following Figure. The operators can be either an erosion, a dilation, a morphological opening or a morphological closing with one of the structuring element shown on the right.





SEa



SEb

Original image | Image 1 (processed with T_1) | Image 2 (processed with T_2) | Structuring element

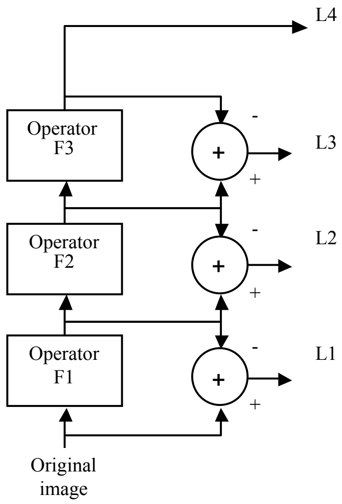
1. Define the T_1 operator and the structuring element used to create Image 1. Justify your response.
2. Define the T_2 operator and the structuring element used to create Image 2. Justify your response.

Answers:
T_1 is a dilation with SE_a. In the processed image, maximas have been duplicated on the basis of the four points of the structuring element. Thin minimas have been removed.
T_2 is an opening with SE_b. In the processed image, maximas that are smaller than the structuring element have been removed. Minimas have not been modified.

6: Consider the image analysis scheme given on the right side of the page.

In the sequel, various schemes are created with several operators Fi. For each scheme, you are asked to classify images **L1, L2 and L3** in one of the following class:

A: Zero image,
B: Positive image,
C: Negative image,
D: None of the previous cases.



a) System 1:
F1: Dilation with a flat square structuring element of size 5x5.
F2: Erosion with the same structuring element as F1.
F3: Closing with a flat square structuring element of size 3x3.

b) System 2:
F1: Opening with a flat square structuring element of size 7x7.
F2: Closing(Opening(.)) with the same structuring element as F1.
F3: Closing(Opening(.)) with the same structuring element as F1.

Answers:
System 1: L1: Negative image (the dilation is extensive as the space origin belongs to the structuring element), L2: Positive image (the erosion is anti-extensive as the space origin belongs to the structuring element), L3: Zero image (the concatenation of F1 and F2 is a closing with a 5x5 structuring element. F3 is a closing with a 3x3 structuring element. So it has no effect (see the section on granulometry)).
System 2: L1: Positive image (L1 is a Top-Hat), L2: Negative image (The opening has no effect as it is idempotent and F1 was already the same opening. So the operator acts like a pure closing which is extensive), L3: Zero image (the concatenation of F1 and F2 is an alternating filter, which is a morphological filter. So it is idempotent. So the second Closing(Opening(.)) has no effect).

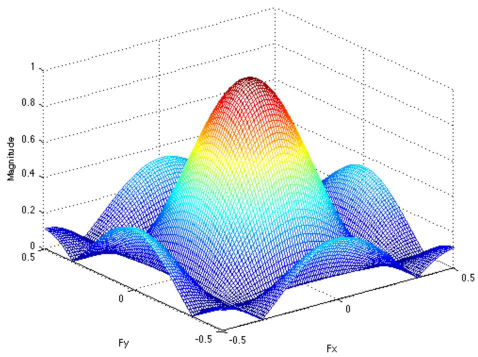
7: Compute the Discrete Fourier Transform of NxN samples of the image defined by $\boldsymbol{x[m,n] = \delta[m]}$ with $\boldsymbol{\delta[m] = \begin{cases} 1 \; if \; m = 0 \\ 0 \; otherwise \end{cases}}$

Answer: $DFT_{NxN}\{\delta[m]\} = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} \delta[m]\, e^{-j2\pi\frac{k}{N}m} e^{-j2\pi\frac{l}{N}n} = N\delta[l]$

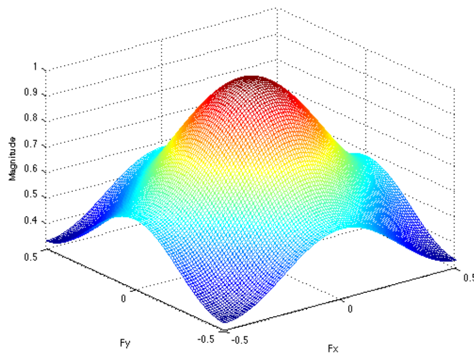
8: Consider the image x[m,n] of NxN pixels with values between [0,1], X[k,l] its Discrete Fourier Transform (DFT) of NxN samples and the image y[m,n] = 1 – x[m,n]. Compute Y[k,l], the DFT of y[m,n], in terms of X[k,l] and the inverse DFT of X[k,l]+Y[k,l].

Answer: $Y[k,l] = N^2\delta[k,l] - X[k,l]$ and $DFT^{-1}\{Y[k,l] + X[k,l]\} = DFT^{-1}\{N^2\delta[k,l]\} = 1$

9: Consider the following frequency responses of two low pass filters (Average and Gaussian) of 3x3 samples. Justify which frequency response corresponds to each filter.



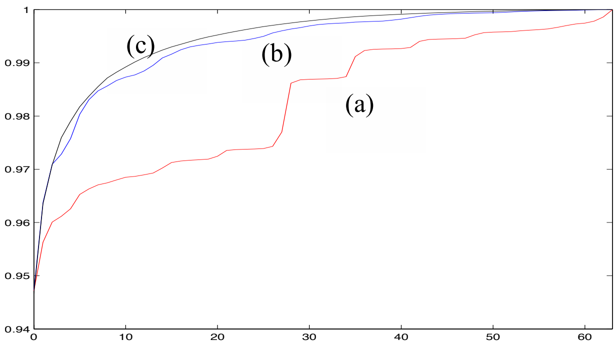
(a)



(b)

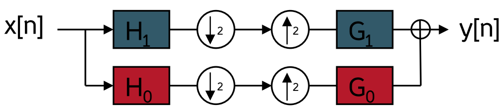
Answer: Filter (a) is average while filter (b) is Gaussian. Average filter corresponds to a 2D sinc function in frequency so it has zeros in the modulus at multiples of 1/N =1/3 while the Gaussian filter does not.

10: The following figure shows the normalized cumulative energy for the zigzag scanned coefficients of the Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and the Karhunen-Loeve Transform (KLT) using blocks of 8x8 pixels of an image. Justify which curve corresponds to which transformation.



Answer: Curve (c) corresponds to KLT as it is optimal in the energy compactness sense, curve (b) is DCT as it also has very high compactness properties. Finally curve (a) is DFT.

11: Consider the following filter bank decomposition of a 1D signal x[n]. Justify the condition on the filters $h_0[n]$, $h_1[n]$, $g_0[n]$ and $g_1[n]$ to be <u>orthogonal</u> filters.



Why are orthogonal filters usually <u>not</u> used in image processing as filters in the wavelet decomposition?

Answer: In order to achieve perfect reconstruction: $\boldsymbol{G_0(F)H_0(F) + G_1(F)H_1(F) = 1}$ and $\boldsymbol{G_0(F)H_0(F - {}^{1}\!/_{2}) + G_1(F)H_1(F - {}^{1}\!/_{2}) = 0}$. Orthogonal filters also comply with $h_0[n] = g_0[-n]$ and $h_1[n] = g_1[-n]$. They are not used in image processing as they cannot have linear phase (except Haar wavelet).

12: Discuss briefly the differences between Gaussian and Laplacian pyramids.

Answer: The Gaussian pyramid consists of a series of images that are iteratively filtered using a Gaussian filter and scaled down. The Laplacian pyramid is similar to the Gaussian but it computes the difference between up-sampled Gaussian pyramid level and the Gaussian pyramid level effectively. It represents a band pass filter (except for the first and last level).

13: List the sequence of steps involved in Canny edge detection, including image preprocessing. Describe the function of each step in one or two sentences.

Answer :

Step 1: Smoothing: Use a Gaussian smoothing filter to reduce the local noise in the image. Edge detectors are strongly affected by noise, yielding many false positives. This is especially true for the small filters that we are using here.

Step 2: Compute gradients: Use a convolution filter to compute the vertical and horizontal components of the gradient throughout the image. This information is used to estimate the strength and orientation of edges.

Step 3: Non-maxima suppression: Whichever pixel has a lower edge strength than either of its two neighbors in the direction of the edge gradient (perpendicular to the edge orientation) is set to zero. This way, detected edges are thinned to a width of one pixel.

Step 4: Linking and Hysteresis Thresholding: Track high magnitude contours, removing weak small segments. Use a high threshold to start edge curves and a low threshold to continue them.

14: The Harris corner detector finds corners in an image. Conceptually, corner detection can be thought of as an auto-correlation of an image patch. Consider a window which slides over an image patch

a) Describe intensity changes in the window when (i) the image patch is constant or 'flat', (ii) there is an edge in the patch or (iii) there is a corner in the patch.
b) The change in intensity for a shift (u,v) in a neighborhood of a point (x,y) is approximated by a quadratic form:

$$E(u,v) \approx [u \ \ v] \ M \begin{bmatrix} u \\ v \end{bmatrix}$$

Give the expression of M and explain how M is related to the behavior described in (a)
c) Discuss the invariance or covariance of the Harris corner detector with respect to (i) intensity changes and (ii) rotation

Answer :

a) If the image patch is constant or "flat", then there will be little to no intensity changes in the window. If the image patch has an edge, then there will be no intensity changes in the window along the direction of the edge. If a corner is present, however, then there will be a strong intensity changes in the window regardless of the direction.

b) A local quadratic approximation of E(u,v) in the neighborhood of a point is given by the second-order Taylor expansion

$$E(u,v) \approx [u \ \ v] \ M \begin{bmatrix} u \\ v \end{bmatrix}$$

where

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Image points can be classified using the eigenvalues of M. If the image patch is flat, λ_1 and λ_2 are small; E is almost constant in all directions. If there is an edge, one of the eigenvalues is much larger than the other, $\lambda_1 >> \lambda_2$ or $\lambda_2 >> \lambda_1$. At a corner, both eigenvalues λ_1 and λ_2 are large, E increases in all directions

c) Harris corner detection is partially invariant to affine intensity change. It is invariant to intensity shifts, but not to intensity scaling.
Harris corner detection is covariant to rotation: second moment ellipse rotates but its shape (i.e. eigenvalues) remains the same.

15: Provide the pseudo-code for the implementation of the Hough transform to detect circumferences in an image. NOTE: The equation describing a circle with center *(a,b)* and radius *r* can be written as:

$$a = x - r\cos(t)$$
$$b = y - r\sin(t)$$

Answer:

1. Quantize the parameter space
   int P[r_{max}][a_{max}] [b_{max}]; // accumulators
2. Loop for all edge points

```
for each edge point (x, y) {
        // Compute parameters
        for (t = 0; t < 360; t = t+Δt) {
        for (r = 0; r < r_max; r = r+Δr) {
                a = x - r·cos(t);
                a' = quantize(a);
                b = y - r·sin(t);
                b' = quantize(b);
                (P[t][a'][b'])++;
        }
        }
}
```

3. Find the peaks in P[r][a'][b']

16: Describe how to combine RANSAC and Least Squares in order to find instances of a given model in an image.

Answer: First, the parameters describing the model are obtained using RANSAC. Obtain the inlier points. Refine the result by using LS computed on the inliers.

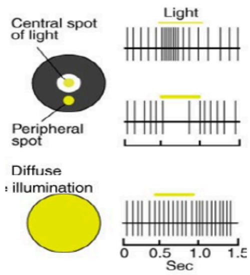
17: Explain briefly the difference between region merging and region growing methods in segmentation.
Answer: Region merging: we start from a partition (all pixels belong to a connected region with unique label) of the image. Neighboring regions are iteratively merged using a region similarity criterion until a given stop criterion is reached.

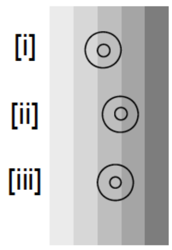
Region growing: we starting from a partial segmentation, defined by some markers ('safe' areas defining the interior of the final regions) and an uncertainty zone (unlabeled pixels, not belonging to any region). The pixels in the uncertainty zone connected to the markers are iteratively assigned to the regions using a similarity criterion, until all pixels in the uncertainty zone are labeled (assigned to a region).

18: Describe the main problems when using a segmentation technique based on gradient binarization.

Answer: It is difficult to define the binarization threshold. The contours may not be closed. Noise in the image may lead to false contours. Contours may be thick, more than one pixel wide.

19:





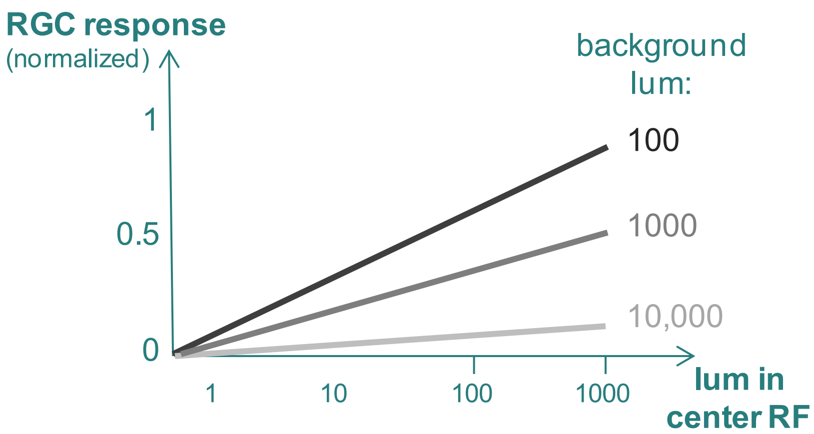
The left figure (taken from the slides shown in class) shows how a typical retinal ganglion cell (RGC), with a center-surround receptive field of type "ON-center", responds in lab experiments. Three cases are shown: when light is projected only on its center (top panel), only in the surround region (middle) or diffusely all over its receptive field (bottom panel).

For the image shown on the right (vertical stripes of different gray levels), consider three such RGCs with receptive field locations as marked (denoted by [i]-[iii]). For each, indicate whether it will produce <u>more</u>, <u>less</u> or the <u>same</u> number of spikes/sec (on average), compared with its response to a homogeneous gray image.

[i] ________ [ii] ________ [iii] ________

Answers:
[i] <u>more</u> [ii] <u>same</u> [iii] <u>less</u>

20:



Experiments have shown that RGCs responses can be affected also by changes in the amounts of light falling *outside* their receptive field (RF) center- and surround- regions. Specifically, as shown in the schematic diagram above (taken from the slides shown in class), the average level of luminance in the “background” (a few degrees outside the RF of each RGC) will affect the gain of the cell’s response to light within its RF center.

A. What is the purpose of this phenomenon? (ie, what is its computational advantage?)
B. How is it called?
C. What mechanism serves a similar purpose in a film-based (non-digital) camera?
D. What is the main advantage of the RGC mechanism, compared to that of the camera?
E. (Optional – extra points) How is this mechanism achieved (implemented) by RGCs?

Answers:

A. The purpose is to match (or ‘adapt’) the sensitivity of an RGC so that its response is relative to the locally prevalent luminance, rather than to any pre-set range, since luminance levels can vary over an extremely wide range (eg, sunlight vs a dimly-lit room). This allows to use the dynamic range of RGC responses more effectively.

B. “Luminance gain control” and/or “light adaptation”.

C. Exposure control (with a shutter)

D. The sensitivity (gain) of each RGC is determined locally. Since images often contain, simultaneously, regions of high- and low-levels of light, this allows the retina to represent only the relevant light gradients within each region. This means it, effectively, represent more light gradients overall than any single RGC is capable of (ie, maximize overall “digital depth” without increasing the digital depth of the individual elements). In contrast, a shutter provides only ‘global’ control. This is is equivalent to adjusting the gain of all RGCs by a single factor (determined by the mean luminance of the entire image).

E. The neural circuitry of RGC achieves luminance gain control by (roughly) dividing the within-RF luminance by the mean luminance in the “background” zone.