Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M1. Introduction to human and computer vision** **Final exam**
Date: November 30th, 2015 **Time: 2h30**
Teachers: Marcelo Bertalmío, David Kane, Ramon Morros, 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 problem in a separate sheet of paper.**
- All results should be demonstrated or justified.

Problem I **David Kane** **(1 point)**

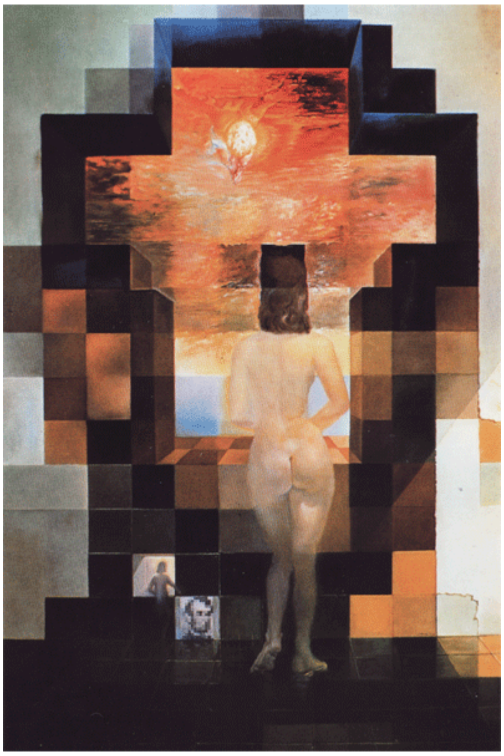
1: The human visual system (HVS) is capable of operating over a huge range of light levels, from 10^{-3} cdm^{-2} on a clear night, to 10^{7} cdm^{-2} to when looking directly into sunlight. However, the human visual system does not work the same at high and low light levels. Describe some of the differences between human vision at high and low light levels. Please include reference to the activity of (a) retinal rods and cones (b) the perception of color and (c) the perception of coarse and fine details.

Solution:

- Cones detect light at the different wavelengths, corresponding in part to the perception of the colors red, green and blue.
- The rods work in low light levels. The cones work in high light levels. The cones system has a higher acuity than the rod system. Thus vision is coarser at low luminance levels than high.
- Additional points for mentioning the Purkinje effect.
- Additional points for mentioning that colors appear more saturated at high light levels.
- Additional points for mentioning that the contrast sensitivity function shifts to higher spatial frequencies at higher light levels.

Condition	cd/m2		Cells
Clear night sky	0.001	Scotopic	Rods
Quarter moon	0.01	Scotopic	Rods
Full moon	0.1	Mesozoic	Rods & Cones
Late twilight	1	Mesozoic	Rods & Cones
Twilight	10	Photopic	Cones
Heavy overcast	100	Photopic	Cones
Overcast sky	1000	Photopic	Cones
Full daylight	10000	Photopic	Cones
Direct sunlight	100000	Photopic	Cones

2: The painting Gala Contemplating the Mediterranean Sea by Salvador Dalí appears to contain two overlapping images; One of the wife of Salvador Dalí looking out to sea and another of Abraham Lincoln. Describe with reference to the contrast sensitivity function (CSF) of the human visual system (HVS) why the image of Gala is more visible when the viewer stands close to the image and why the image of Abraham is more visible when the viewer stands further away from the image.



Far  view

Near view

Solution:

- The CSF describes the sensitivity of the HVS to different frequency components of the stimulus. The CSF is band pass.
- We are most sensitive to wavelengths at ~3 cycles per degree (cpd), less sensitive to coarse details that are less than 3 cpd and less sensitive to fine details that are greater than 3 cpd.
- When a subject moves towards or away from the painting the spatial frequency content of the stimulus changes. When we move away from the painting, fine details (high spatial frequencies) become less visible whilst coarse details (low spatial frequencies) become more visible and vice-versa.
- The image of Gala is a fine image; it has more energy in the high spatial frequencies. The image of Abraham is a coarse image; it has more energy in the low spatial frequencies. Thus at near distances the image of Gala is more visible and at far distances the image of Abraham is more visible.
- Additional points may be given by referring to the following
    - The Fourier transform.
    - Campbell and Robson sensitivity chart. Or any of the various charts used by optometrists (e.g. the Snellen acuity chart).
    - Fractal images.
    - A diagram of the CSF.
    - Psychophysics and the development of the CSF from contrast thresholds.

Problem II Philippe Salembier (2 points)

1: Consider the following image which is quantized with 4 bits.

1	4	4	4
1	4	4	5
1	5	4	5
4	5	4	1

Compute the image after histogram equalization.
What is the effect of histogram equalization on the entropy of this image.

Solution: The image after histogram equalization is:

4	11	11	4
4	11	11	15
4	15	11	15
11	15	11	4

As, for this particular image, the equalization simply redistributes the gray level values without merging any of them, the entropy will be preserved.

2: Consider the following flat structuring element SE (the underlined position indicates the m=n=0 point):

$$\begin{array}{ccccc} -\infty & -\infty & -\infty & -\infty & -\infty \\ -\infty & -\infty & 0 & -\infty & -\infty \\ -\infty & 0 & \underline{-\infty} & 0 & -\infty \\ -\infty & -\infty & 0 & -\infty & -\infty \\ -\infty & -\infty & -\infty & -\infty & -\infty \end{array}$$

Consider the operator, $\boldsymbol{\psi}$, that consists in dilating twice the image with the structuring element SE: $\boldsymbol{\psi}(.) = \boldsymbol{\delta_{SE}}(\boldsymbol{\delta_{SE}}(.))$.
Is this operator ψ increasing, idempotent and extensive? (Precisely justify your answers)

Solution: The operator is equivalent to a dilation with the following structuring element SE2:

$$\begin{array}{ccccc} -\infty & -\infty & 0 & -\infty & -\infty \\ -\infty & 0 & -\infty & 0 & -\infty \\ 0 & -\infty & \underline{0} & -\infty & 0 \\ -\infty & 0 & -\infty & 0 & -\infty \\ -\infty & -\infty & 0 & -\infty & -\infty \end{array}$$

The operator is therefore increasing (as all dilation), non idempotent (dilating once or twice is different as shown by the difference between structuring elements SE and SE2. So if we iterate the results will change) and extensive (as the space origin belongs to SE2).

3: We construct a family of structuring element $\{SE_k\}$ based on SE defined in the previous question, with:

$$SE_k(m,n) = \begin{cases} 0 \ \ if \ \ m = \pm k, n = 0 \\ 0 \ \ if \ \ n = \pm k, m = 0 \\ \ \ -\infty \ \ otherwise \end{cases}$$

Can we compute a granulometry with the openings based on these structuring elements: $\boldsymbol{\gamma_{SE_k}}(.)$?
If yes, define the block diagram of the granulometry.

Solution : We cannot define a granulometry because the structuring element is not convex. Therefore increasing the size of the structuring element does not increase the strength of the opening or closing.

4: Consider the following operators: $\boldsymbol{\psi_1}(.) = \boldsymbol{\varphi}\big(\boldsymbol{\gamma}(.)\big)$ and $\boldsymbol{\psi_2}(.) = \boldsymbol{\gamma}(\boldsymbol{\varphi}(.))$ where $\boldsymbol{\gamma}$ is an opening and $\boldsymbol{\varphi}$ a closing with a square structuring element of size 3x3.
Is the operator $\boldsymbol{\psi_2}(\boldsymbol{\psi_1}(.))$ increasing, idempotent and anti-extensive or extensive?

Solution: $\psi_2\big(\psi_1(.)\big) = \gamma\varphi\varphi\gamma(.) = \gamma\varphi\gamma(.)$ because the closing ($\boldsymbol{\varphi}$) is idempotent. We know that the combination $\gamma\varphi\gamma(.)$ is a **morphological filter** that is: increasing and idempotent. It is not anti-extensive nor extensive as many counterexamples can be found, for example a white image with one black point. The resulting filtered image will be all white. The filter is not extensive either as a black image with a white point will be transformed into a black image.

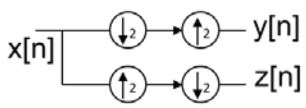
Problem III	**Marcelo Bertalmío**	**(1 point)**

1: Define trichromacy. Prove mathematically this property.
Answer: section II of lecture notes.

2. List and briefly explain all the color processing operations that are performed in-camera.
Answer: sections VII-G, X, XI of lecture notes.

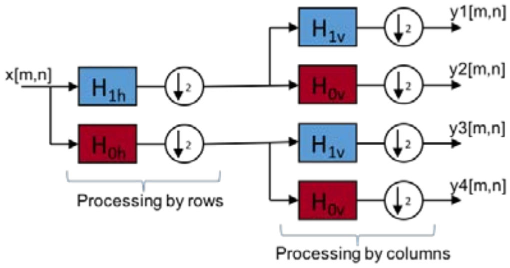
Problem IV	**Javier Ruiz**	**(3 points)**

1: Consider the following decomposition using down-sampling and up-sampling processes (without filtering). Express both Fourier Transforms Y(F)=FT{y[n]} and Z(F)=FT{z[n]} as a function of X(F)=FT{x[n]}.



Solution: $Y(F) = {}^1\!/_2\, X(F) + {}^1\!/_2\, X\left(F - {}^1\!/_2\right)$ and $Z(F) = X(F)$

2: Consider the following wavelet decomposition of an image where H0 and H1 correspond to 1D low-pass filter and high-pass respectively. Indicate which image (approximation, horizontal detail, vertical detail and diagonal detail) correspond to each output image (y1[m,n], y2[m,n], y3[m,n] and y4[m,n]) respectively.

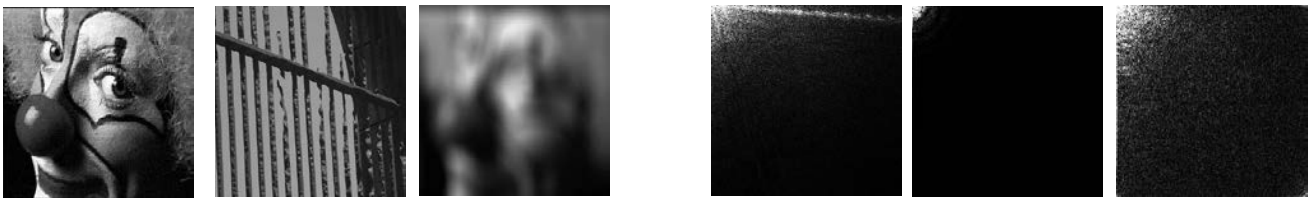


Solution: y1[m,n]: Diagonal detail; y2[m,n]: Vertical detail; y3[m,n]: Horizontal detail; y4[m,n]: Approximation

3: Enumerate the three discrete impulse responses that can be used to proximate the horizontal gradient $\frac{\partial f(x,y)}{\partial x}$ of an image $f(x,y)$.

Solution: Forward difference: $h_f[m] = \{1, \underline{-1}\}$; Backward difference: $h_b[m] = \{\underline{1}, -1\}$; Symmetric difference: $h_s[m] = \{1, \underline{0}, -1\}$

4: Justify which DCT transformation (on the right) corresponds to which image (on the left).



(a) (b) (c) (1) (2) (3)

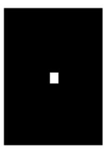
Solution:
(a) → (3) High energy at high frequencies without any specific direction
(b) → (1) High energy at high frequencies at a specific direction
(c) → (2) High energy only at low frequencies due to blurred image

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

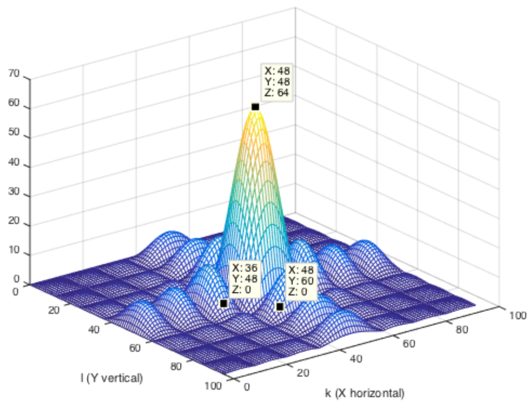
Solution:

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

6: Consider the image x[m,n] of 96x96 pixels composed by a black background of level 0 and a white square of level 1 of PxP pixels (Figure a). Figure b represents the magnitude of the Discrete Fourier Transform of 96x96 samples using the centered representation. From the DFT representation, obtain the size P of the white square of the image.



a) Image x[m,n]



b) Magnitude of the DFT with 96x96 samples

Solution:
Maxim value of 64 at X=48,Y=48 (centered representation) → k=0, l=0 → $P^2 = 64$ → P=8

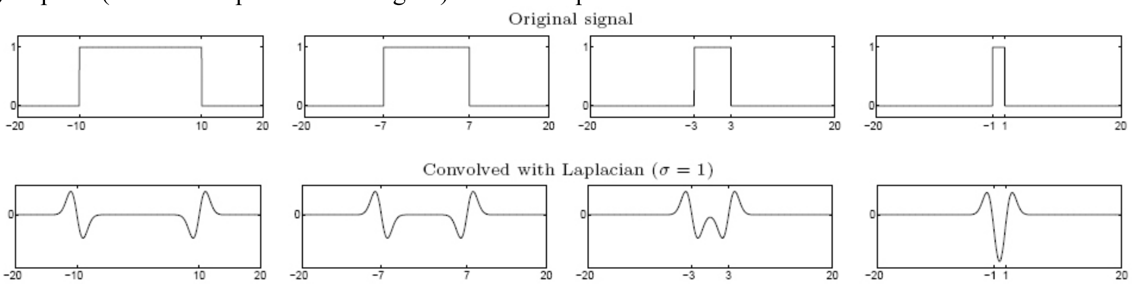
Alternatively: zero value at Y=60 (centered) → l = 60-48 = 12 (non centered) → 1/P = 12/96 → P=8

Problem V Verónica Vilaplana (1 point)

1: a) Define invariance and covariance in the context of key-point detection.
b) Is the Harris corner detector invariant or covariant to (i) affine intensity changes, (ii) image translation, (iii) image rotation, (iv) image scaling?
Why or why not?
c) Give an example of a computer vision application in which corners are more appropriate features to use than, for example, lines.

Solution:
a) **Invariance:** if the image is transformed and the feature locations do not change. **Covariance:** if we have two transformed versions of the same image, the features should be detected in the corresponding locations.
b) See Slides Lecture 8, pp. 56, 57
c) Detecting corners would be more appropriate when only a sparse set of points are needed, especially facilitating the detection of corresponding points in multiple images. This is used, for example, in stereo matching

2: a) Explain (with the help of the next figure) how the Laplacian of Gaussian can be used as a blob detector.



b) Why do we need to perform scale normalization?
c) What is the characteristic scale of a blob?

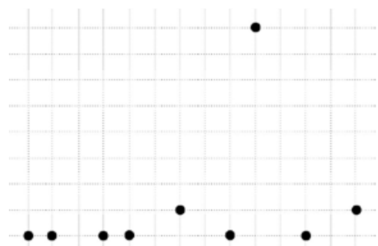
Solution:
a) The image is convolved with Laplacian of Gaussians at several scales. The magnitude of the response will achieve a maximum at the center of the blob, when the scale of the Laplacian matches the scale of the blob.
b) The Laplacian response decays as scale increases (the response of the derivative of the Gaussian filter to a step edge decreases as the scale sigma increases… see Slide 68). To keep the response scale-invariant it must be multiplied by scale^2.
c) The characteristic scale of a blob is the scale that produces a peak of the Laplacian response in the blob center.

Problem VI Ramón Morros (2 points)

1:

After a contour detection step, we obtain a set of contour points with the following coordinates:

$x_0 = (0, 1)$ $x_1 = (1, 1)$
$x_2 = (3, 1)$ $x_3 = (4, 1)$
$x_4 = (6, 2)$ $x_5 = (8, 1)$
$x_6 = (9, 9)$ $x_7 = (11, 1)$
$x_8 = (13, 2)$



We use RANSAC to estimate the parameters of the line ($y = ax + b$) that better fits the set of points. Three pairs of points are selected randomly to compute three RANSAC iterations: $r_1=[x_1,x_6]$, $r_2=[x_0,x_7]$, $r_3=[x_5,x_6]$.
a) Give the parameters of the line found after the three iterations of RANSAC if the threshold is $d = 1.1$. Explain concisely the steps performed.
b) Explain how can this result be improved in order to obtain a line that better fits the resulting inliers.
NOTE: the problem can be resolved graphically; hence a calculator is not needed.

Solution:
a) At each iteration, the line passing through the two selected points is drawn. The number of inliers (points where the distance to the line is <= d) is counted.

After all iterations, the line resulting in more inliers is selected
Iteration 1: 3 inliers
Iteration 2: 8 inliers
Iteration 3: 2 inliers
Best line corresponds to iteration 2: (a,b) = (0,1) → y = 0·x+1

b) Re-compute the model using only inliers using Least Squares

2: Segmentation using Mean-Shift:
a) Explain the mean-shift algorithm applied to segmentation in the feature space.
b) Discuss the application of the algorithm in the case of high dimensionality feature spaces.

Solution:
a) Slides Lecture 10, pp.47
b) High dimensionality feature spaces are very sparse with data points far away from each other. Only a very limited number of data points are available to infer local structure, which may lead to erroneous results. This phenomenon is known as the curse of dimensionality.

3: Let's assume we want to segment an image using a region-merging approach. Initially, the similarity between regions is modeled using a criterion based on the approximation of the Mean Squared Error, given by Eq. (1). When segmenting the image, we observe that this criterion leads to noisy contours. Propose a variation of the criterion to obtain regions with smoother contours.

$$C_{color}(R_1,R_2) = N_{R1}\left\|M_{R1} - M_{R1\cup R2}\right\|_2^2 + N_{R2}\left\|M_{R2} - M_{R1\cup R2}\right\|_2^2 \qquad (1)$$

Solution:
An additional criterion taking into account the length of the common contour can be added to the previous term:

$$C(R_1,R_2) = \alpha C_{color}(R_1,R_2) + (1-\alpha)C_{cont}(R_1,R_2)$$

$$C_{cont}(R_1,R_2) \approx -\text{Length of common contour}$$

4: Schematically explain the Max-Lloyd algorithm used to perform the k-means clustering.

Solution: Slides Lecture 10, pp.22