Master in Computer Vision Barcelona

UAB UOC UPC upf

Module: **M1. Introduction to human and computer vision** **Final exam**
Date: December 2nd, 2013
Teachers: Marcelo Bertalmío, Ramon Morros, Albert Oliveras, Nava Rubin, Javier Ruiz, Verónica Vilaplana **Time: 2h30**

- 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: Explain what is the shape of the color gamut of a display device, plotted in xy chromaticity coordinates, and what is the reason for this shape. Explain why, for any device, there are always colors that we can perceive but that the display is not able to reproduce.
Answer in section 2.6.3 of the course notes.

2: Explain the two most popular techniques for in-camera automatic white balance.
Answer in section 2.5 of the course notes.

3: The following Matlab program generates the granulometry ('pecstrum') shown in figure 1 of the 'circles' image of figure 2.

```
I = imread('circles.jpg');
imagesc(I);colormap('gray');
max_size = 16;
x = ((-max_size+1):max_size);
pecstrum = granulometry(I,'disk',max_size,'mav');
figure(1), plot(x, -derivate(pecstrum)), grid, title('Derivate Granulometry with a
''disk'' as SE')

function pecstrum = granulometry(A, type_SE, steps, measure)
for counter = 1:steps
    remain = imopen(A, strel(type_SE, counter-1));
    pecstrumA(counter) = sum(abs(remain(:)));
    remain = imclose(A, strel(type_SE, counter-1));
    pecstrumB(counter) = sum(abs(remain(:)));
end
pecstrum = cat(2, fliplr(pecstrumB), pecstrumA);
end
```

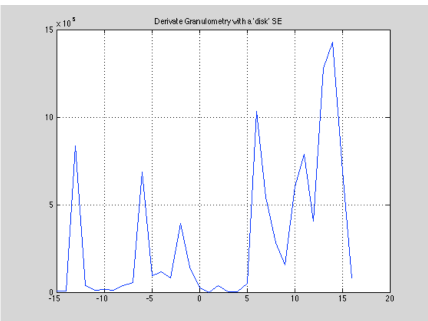


Figure 1: "pecstrum"

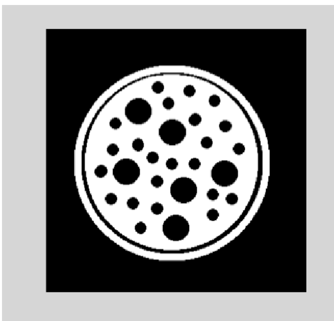


Figure 2: "circles"

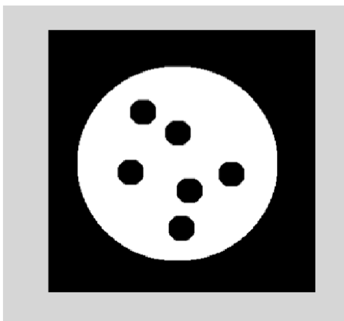


Figure 3

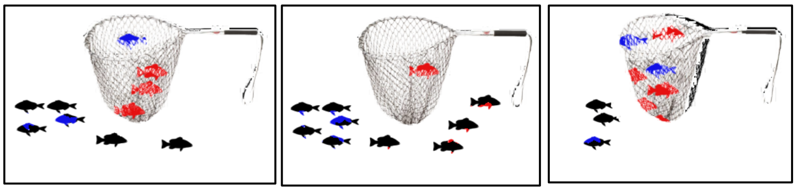
Write a Matlab program that generates the image of figure 3 from the image 'circles' and explain the purpose of each instruction.

Answer:

```
I   = imread('circles.jpg');
out = imclose(I, strel('disk', 13));
% This command corresponds to a closing of the image I, which removes dark circles
```

```
% of size lower or equal to 13. This size was selected because of the
% granulometry/pecstrum peak at -13 (minus means dark in this case)
Very similar results are obtained with a strel('disk', 7) to strel('disk', 13) in
the imclose instruction.
```

4: In the context of a fishing game for kids, where the objective is to fish all the red fishes, 3 players get the following results:



Player 1 Player 2 Player 3

As it is not a trivial situation to evaluate, we decided that the winner will be assessed on F measure (harmonic mean of precision and recall). To get an ordered list of the players, fill the following table:

Answer

		Player 1	Player 2	Player 3
TP	TruePositive	3	1	5
T	True	5	5	5
P	Positive	4	1	6
Precision	=TP/P	3/4	1	5/7
Recall	=TP/T	3/5	1/5	1
F measure	=2*(Prec*Recall)/(Prec+Recall)	2/3	1/3	5/6

5: Find and explain the correspondence between images and histograms.

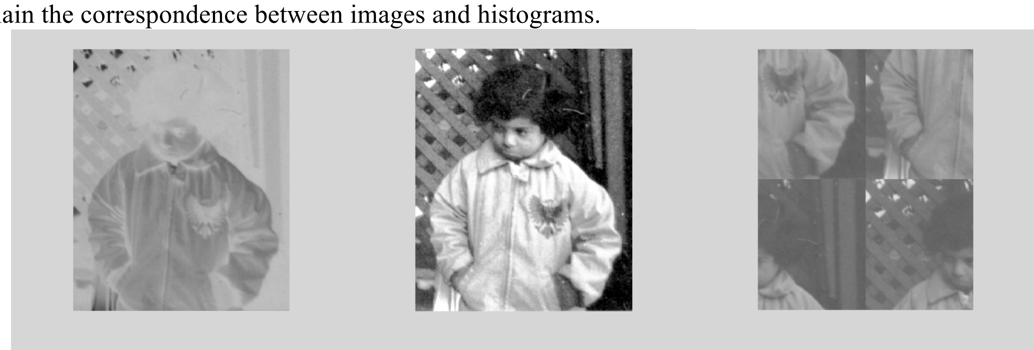
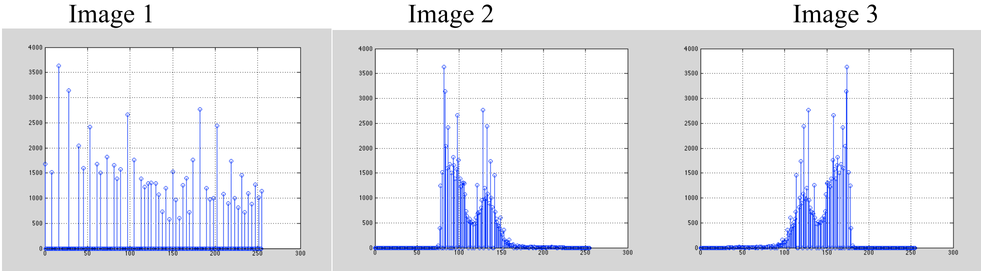


Image 1 Image 2 Image 3



Histogram 1 Histogram 2 Histogram 3

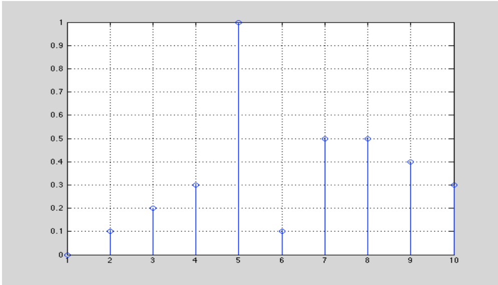
	Histogram Number
Image 1	3
Image 2	1
Image 3	2

Explanation:
Histogram 1 corresponds to an image that is more contrasted than the other two, that is image 2.
Histogram 2 corresponds to image 3 because has the same envelope as histogram 1 but less contrasted and the histogram is invariant to the shift of signal.
Histogram 3 corresponds to the unshifted "negative" of image 1 that is image 1.

6: Write the mathematical expressions of the following operations: Dilation, Erosion of a 1D signal x[n] with a Structuring Element b[n]: $\delta_b\{x[n]\} = x[n] \oplus b[n] =$

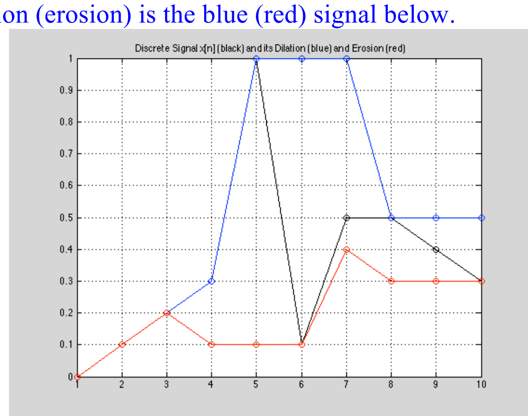
$$\varepsilon_b\{x[n]\} = x[n] \Theta b[n] =$$

According to the previous definitions, draw in the following graphic the corresponding dilation and erosion of the discrete signal x[n] by a flat structuring element $b[n] = \{-\infty, .., -\infty, \underline{0}, 0, 0, -\infty, .., -\infty\}$



Answer: $\delta_b\{x[n]\} = x[n] \oplus b[n] = \bigvee_{k=-\infty}^{\infty} (x[k] + b[n-k])$ and $\varepsilon_b\{x[n]\} = x[n] \Theta b[n] = \bigwedge_{k=-\infty}^{\infty} (x[k] - b[k-n])$

According to these definitions, the dilation (erosion) is the blue (red) signal below.



7: In Mean-Shift segmentation, which parameter has to be specified? Discuss the influence of this parameter in the final segmentation result.

Answer: The parameter is the window size. The size of the window is related to the number of modes of the pdf (number of clusters) in the feature space: Too small window sizes can result in an excessive number of local maxima (high number of classes) found. Too large window sizes may result in missing some classes.

8: K-Means vs. GMM/EM (Expectation Maximization) segmentation: which are the parameters needed in either case to represent the clusters during the iterative steps of the algorithms? Comment the pros and cons of each algorithm.

Parameters:

k-means: mean of each class.

GMM/EM: weight, mean and variance of each class.

Pros & cons: K-Means is faster (only computes class means and point-distances to the means) but is less robust as in each iterative step there is a hard assignment of the feature points to the classes. GMM/EM uses a probabilistic interpretation, where each point contributes to all classes and thus, is more robust at the cost of higher computational complexity (computation of Gaussian parameters and weights).

9: Let us suppose we are using the Hough transform to find circles in an image (using a naïve implementation). After contour detection, N contour points are found. Give an estimation of the computational complexity of the approach (number of basic operations: additions, multiplications, comparisons, trigonometric). 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: We quantize the radius *r* into R levels and the parameter *t* into T levels. Then, for each contour point, the parameters (a,b) are computed and stored for each combination of *r* and *t*. This is, N·R·T· (2add + 2mult + 2 trig) operations. At the end of this step, we have to find the M maxima of the resulting 3·N matrix (a,b,r). This can be done using 3·N·M comparison operations.

10: Discuss the advantages and drawbacks of LS vs. Hough Transform vs. RANSAC to find instances of a given shape in an image.

LS: fast, closed form, scales well with dimensionality. It is not robust to outliers. It can only detect one instance of the model.

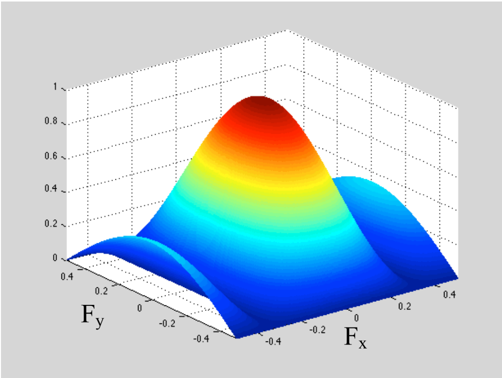
Hough: Can detect several instances of the shapes, robust. It does not scale well with the dimensionality of the problem in terms of computational complexity.

RANSAC: Scales well with dimensionality, robust to outliers. Usually, it can only detect one instance of the model. It uses an iterative process and computational complexity may be very high depending on the percentage of outliers.

11: Compute the Fourier transform, indicating the value in $F_x=0$ and $F_y=0$, of the image of MxM pixels defined by $\boldsymbol{x[m,n]} = \boldsymbol{\delta[m]}$ with $\boldsymbol{\delta[m]} = \begin{cases} \mathbf{1}\ \boldsymbol{if\ m = 0} \\ \mathbf{0}\ \boldsymbol{otherwise} \end{cases}$

Answer: $X(F_x, F_y) = TF\{x[m,n]\} = TF\{\delta[m]\} = e^{-j\pi F_y (M-1)} \frac{\sin(\pi F_y M)}{\sin(\pi F_y)}$ with $X(0,0) = M$

12: Consider the <u>average</u> filter h[m,n] with the following Fourier transform:



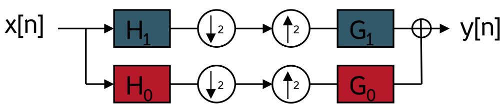
Justify the dimensions MxN (M columns and N rows) of the impulse response of the <u>average</u> filter h[m,n].

Answer: M=3 as it has zeros in $F_x=1/3=1/M$. Similarly, N=2 as it has zeros in $F_y=1/2=1/N$

13: Enumerate the laplacian pyramid elements and the steps needed to construct it from an image **X** for 3 levels of the pyramid (use G_i to represent the blur-and-downsample operator and F_i to denote the blur-and-upsample operator at level i of the pyramid).

Answer: Laplacian pyramid elements:
$\mathbf{L_1} = \mathbf{X} - F_1 G_1 \mathbf{X}$
$\mathbf{L_2} = X_2 - F_2 G_2 X_2$ (with $\mathbf{X_2} = G_1 \mathbf{X}$)
$\mathbf{L_3} = X_3 - F_3 G_3 X_3$ (with $\mathbf{X_3} = G_2 \mathbf{X_2}$)
$\mathbf{X_4} = G_3 \mathbf{X_3}$

14: Consider the following filter bank decomposition of a 1D signal x[n]. Justify the conditions on the filters H_0, H_1, G_0 and G_1 to achieve perfect reconstruction.



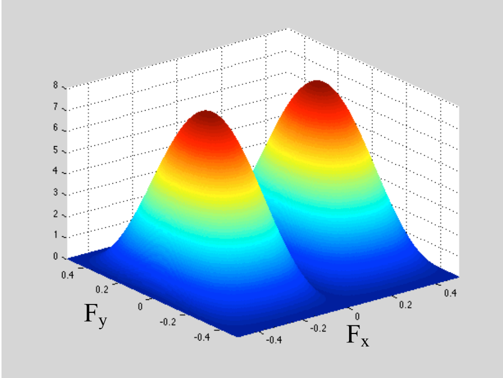
If $H_0=G_0$ correspond to ideal low pass filters with cut-off frequency $F_c=1/4$ and $H_1=G_1$ correspond to ideal high pass filters with $F_c=1/4$, justify that they are bi-orthogonal.

Answer: $G_0(F)H_0(F) + G_1(F)H_1(F) = 1$ and $G_0(F)H_0(F - {}^1\!/_2) + G_1(F)H_1(F - {}^1\!/_2) = 0$
From the equations above we see that if $H_0=G_0$ are ideal low pass filters and $H_1=G_1$ are ideal high pass filters then they comply with both equations and they correspond to bi-orthogonal filters.

15: Discuss the advantages and disadvantages of the Discrete Cosine Transform (DCT) versus the Karhunen-Loeve Transform (KLT).

Answer: The DCT is a complete, separable and orthogonal transform whose transformed coefficients are real and present good compaction characteristics (better than DFT).
The KLT is a data-dependent transform which is optimal in the sense of energy compactness and therefore allows space dimensionality reduction. The basis is data-dependent and it has to be computed for each data collection.

16: Justify which sobel filter ($\boldsymbol{h_1[m,n]} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{-1} & \mathbf{-2} & \mathbf{-1} \end{bmatrix}$ or $\boldsymbol{h_2[m,n]} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{-1} \\ \mathbf{2} & \mathbf{0} & \mathbf{-2} \\ \mathbf{1} & \mathbf{0} & \mathbf{-1} \end{bmatrix}$) corresponds to the frequency response below. What kind of contours does it detect?



Answer: The frequency response corresponds to filter $h_2[m,n]$ as it detects vertical contours that correspond to horizontal frequencies F_x (in the figure, the horizontal frequencies are band-pass filtered while vertical frequencies are not).

17: Edge Detection

Compare the Canny edge detector and the Laplacian-of-Gaussian (LoG) edge detector for each of the following questions.
(a) Which of these operators is/are isotropic and which is/are non-isotropic?
(b) Describe each operator in terms of the order of the derivatives that it computes.
(c) What parameters must be defined by the user for each operator?
(e) Which detector is more likely to produce long, thin contours? Briefly explain.

Answer:
(a) LoG is isotropic and Canny is non-isotropic.
(b) LoG is 2nd derivative and Canny is 1st derivative.
(c) LoG requires σ defining the scale of the Gaussian blurring. Canny requires σ and two thresholds for hysteresis.
(d) Canny because of non-maximum suppression which thins and hyteresis thresholding which can fill in weak edge gaps.

18: Corner detection

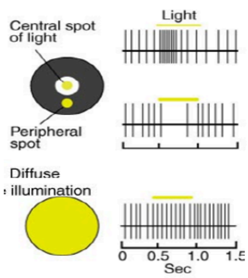
The Harris corner detection algorithm computes a 2×2 matrix at each pixel in terms of derivatives at that point and then computes the two eigenvalues of the matrix, λ_1 and λ_2, where $\lambda_1 \leq \lambda_2$.
(a) Which are the elements of the matrix?
(b) How can the two eigenvalues be used to label each pixel as either a locally smooth region (S), an edge point (E), or a corner point (C)? Give your answer by specifying "Label pixel S if …", "Label pixel E if …" and "Label pixel C if …"
Answer:

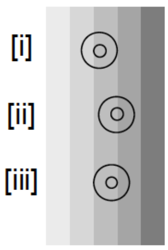
(a) $M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$

the elements are the products of the first derivatives I_x and I_y.

(b) The Harris corner detector first computes $R = \det(M) - \alpha \operatorname{trace}(M) = \lambda_1\lambda_2 - \alpha(\lambda_1 + \lambda_2)^2$ and then
1) labels a pixel S if $|R| \approx 0$ (or, alternatively, $\lambda_1 \approx \lambda_2 \approx 0$);
2) labels a pixel E if $R < T_1 < 0$ (or, alternatively, $\lambda_1 \approx 0$ (corresponding to the direction of the edge) and λ_2 is large (corresponding to the normal direction at the edge));
3) labels a pixel C if $R > T_2 > 0$ (or, alternatively, λ_1 and λ_2 are both large).

19:





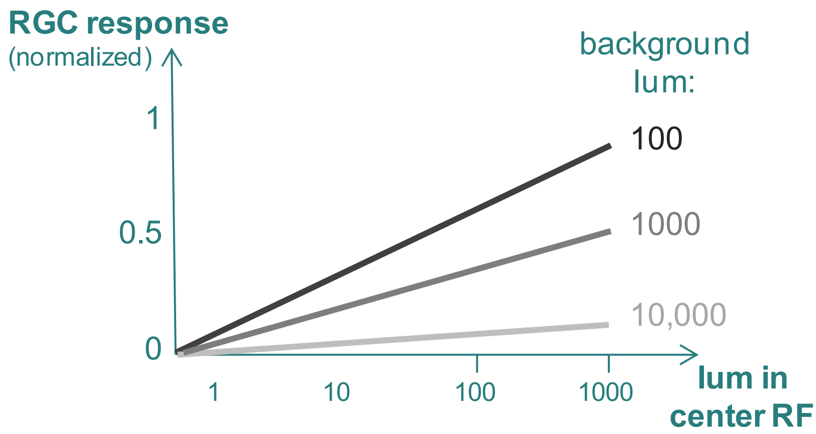
The left figure (taken from the slides shown in class) shows how a typical retinal ganglion cell (RGC), with a center-surround receptive field of type "ON-center", responds in lab experiments. Three cases are shown: when light is projected only on its center (top panel), only in the surround region (middle) 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 more, less or the same number of spikes/sec (on average), compared with its response to a homogeneous gray image.

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

Answers:
[i] __more__ [ii] __same__ [iii] __less__

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.