Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M1. Introduction to human and computer vision** **Final exam**
Date: December 3rd, 2018 **Time: 2h30**
Teachers: Marcelo Bertalmío, 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 Marcelo Bertalmío (2 points)

1. Explain the trichromacy property and the experiments to derive the color matching functions.
   Answer in Section II of the notes.

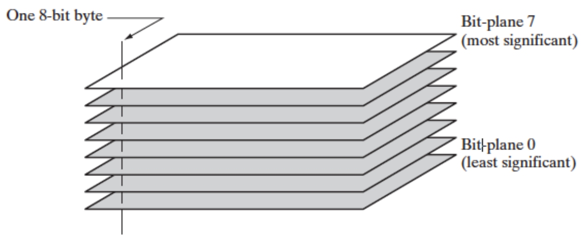
2. Define color constancy and state von Kries law.
   Answer in Section IV of the notes.

3. What were the reasons to propose the XYZ color space alongside RGB? How is the transform from RGB to XYZ, linear or non-linear? Why, for any given display, there are always colors that we can see but that the display is not able to reproduce?
   Answer in Section III of the notes.

4. Describe the two most popular schemes that cameras use to perform automatic white balance.
   Answer in Section X of the notes.

Problem II Philippe Salembier (2 points)

1. Assume a gray level image is quantized with 8 bits, we are interested in highlighting the contribution of specific bits to the total gray level appearance. Consider that the image is composed of eight "1-bit planes", ranging from bit-plane 0 for the least significant bit to bit-plane 7 for the most significant bit, as shown in the figure below:



   Separating a digital image into its bit-planes (bit-plane slicing) is useful for analyzing the relative importance played by each bit of the image.

   a. The images below correspond to bit-planes 3 and 7 of the original image on the left. Could you tell which image corresponds to which bit-plane and why?


Original image


bit-plane X

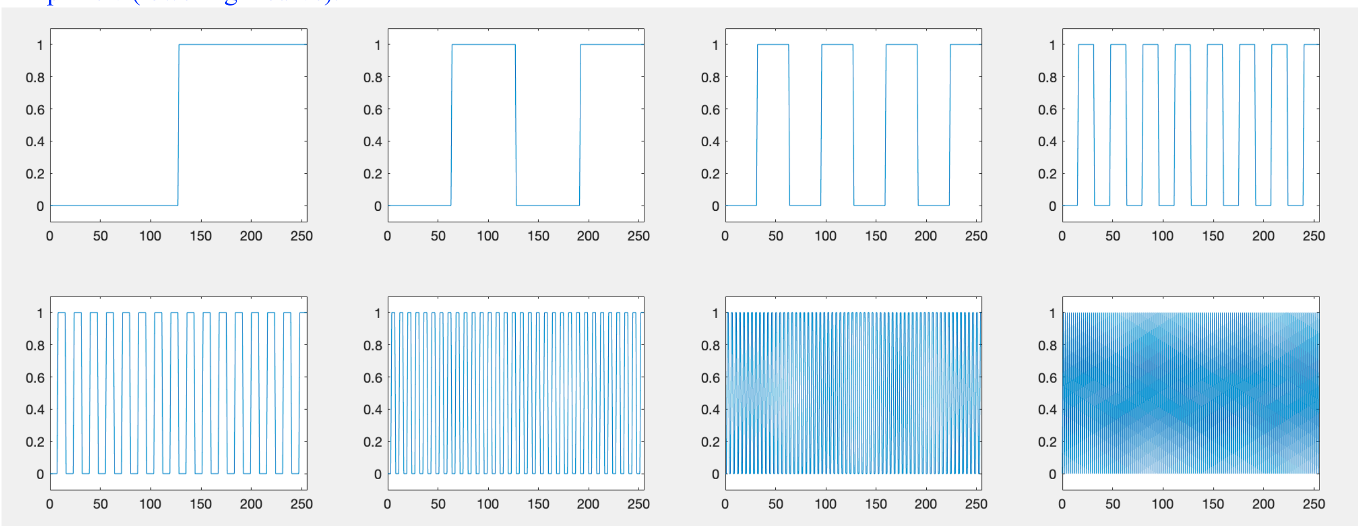

bit-plane Y

b. Propose a set of range transforms $s = T_i(r)$, $0 \le i < 8$ capable of producing all the individual bit-planes (i) of an 8-bit gray level image. Note: You may define the range mapping either through a graphic representation (which is probably easier) or through an analytical formula.

Solution:

a. Bit-plane X image represents small variation of the gray level values of the original image. It corresponds to bit-plane 3. By contrast, bit-plane Y image identifies the pixels of high gray level values of the original image. It corresponds to bit-plane 7.

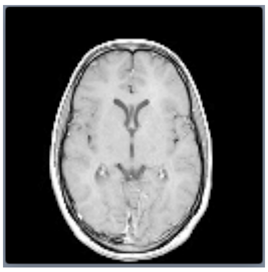
b. The following graph represents the 8 range transforms starting from bit plane 7 (upper left curve) going down to bit-plane 0 (lower right curve).



Each curve can be obtained with the following analytical formulation:

$$s = T_i(r) = Modulo(floor(\frac{r}{128/2^{7-i}}),2)$$

2. You have to segment a set of 10.000 Magnetic Resonance Images (MRI) similar to the one shown on the right side of the page. You want in particular to extract pixels corresponding to the cerebrospinal fluid. From this dataset, you also have access to a set of 50 images which have been manually segmented by experts in MRI analysis.



After analyzing a reasonable set of images, you conclude that cerebrospinal fluid areas are sufficiently contrasted and that you may use a simple thresholding technique to extract them.

Describe precisely how you would define the parameters of the algorithm so that optimal results are obtained.

Solution: The main parameter of the thresholding technique is of course the threshold. To define its value optimally, one could use the classical approach when groundtruth is available (here represented by the 50 images segmented by experts). On this set of 50 images, the comparison of the binary masks extracted by the algorithm with the groundtruth masks defined by the experts allows one to compute the average Precision and Recall values as well as the F-score (Harmonic mean of Precision and Recall) for each possible threshold. The threshold value could be the one optimizing the F-score.
If it is felt that some postprocessing of the masks is necessary, for example cleaning of the masks with morphological opening or closing, the size of the structuring element can also be optimized following the same procedure.

3. In image processing on a square grid, the Laplacian operator estimates the sum of second partial derivatives of the image with respect to the vertical and the horizontal coordinates.

a. This operator can be implemented through a linear translation invariant (LTI) filter. Define its impulse response and explain how this impulse response is related to LTI filters that estimate the first derivative of the image. Illustrate the behavior of the filters on a simple transition between a dark object to a bright object.

b. In mathematical morphology, there exist three operators that are related to the estimation of the first derivative of an image. Define these operators and illustrate their behavior on a simple transition between a dark object to a bright object

c. Relying on the tools discussed in the previous point, propose a morphological Laplacian operator that would also estimate the sum of second partial derivatives of the function with respect to the vertical and the horizontal coordinates. Finally, illustrate the behavior of this operator on a simple transition between a dark object to a bright object.

Solution:

a. The impulse response of the linear Laplacian operator is given by: $h[m,n] = \begin{bmatrix} 0 & 1 & 0 \\ 1 & \underline{-4} & 1 \\ 0 & 1 & 0 \end{bmatrix}$. It is a combination of high pass filters estimating the second derivative in horizontal and in vertical. The impulse response of these filters is given by: $[1, \underline{-2}, 1]$ which can be viewed as a combination of "Simple differences" that are high-pass filters estimating the first derivative: Forward difference: [1, <u>-1</u>] and backward difference [<u>1</u>,-1].
On a transition such as 0,0,0,0,0,1,1,1,1,1,1, the filter response is an oscillation as 0,0,0,0,0,0,1,-1,0,0,0,0,0. The response is positive before the transition and negative afterwards.

b. In mathematical morphology, three gradients are defined: The gradient by erosion: x - Erosion(x); the gradient by dilation: Dilation(x) - x and the morphological gradient: Dilation(x) - Erosion(x). All of them with the smallest possible structuring element. On the same transition as in the previous point, the gradient by erosion will give a positive response of one pixel <u>after</u> the transition; the gradient by dilation will give a positive response of one pixel <u>before</u> the transition and the morphological gradient will produce of positive response of two pixels on both sides of the transition.

c. A morphological Laplacian operator can be defined as: Gradient by dilation – Gradient by erosion: That is:
Dilation(x) – x – (Erosion(x) - x) = Dilation(x) – Erosion(x) – 2x.
It will give the same response as the linear operator on simple transitions: 0,0,0,0,0,0,1,-1,0,0,0,0,0.

4. Define whether an erosion with an arbitrary structuring element is increasing, anti-extensive and idempotent. In case the property does NOT hold, find a counterexample illustrating the lack of the property.

Solution: An erosion is increasing, not idempotent and non extensive nor anti-extensive for arbitrary structuring elements. Consider for example a structuring element which is a square centered on the space origin <u>but not including</u> the space origin. Consider as input signal an image equal to 1 everywhere except one point where it is equal to 0. If the erosion is applied on this image, the output will be by the shape of the structuring element and will be neither smaller or larger than the input. This therefore shows that the operator is non extensive nor anti-extensive. If the operator is iterated twice, the image of the structuring element will be further eroded and expended. As a result, the operator is not idempotent.

Original image

1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	0	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1

First Erosion

1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	0	0	0	1	1
1	1	0	1	0	1	1
1	1	0	0	0	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1

Second Erosion

1	1	1	1	1	1	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	1	1	1	1	1	1

Problem III Javier Ruiz (3 points)

1. Consider three filters of size 3x3 whose impulse responses are:

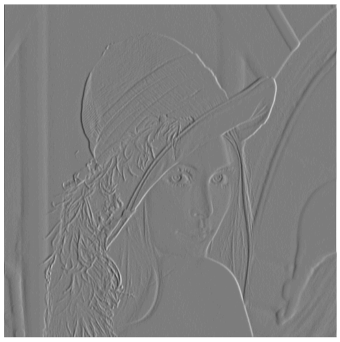
$$h_1[m,n] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad h_2[m,n] = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \qquad h_3[m,n] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Consider the input image x[m,n] shown on the top row of the following figure. **<u>Justify</u>** which images on the bottom row correspond to the outputs of the three filters above (please note that the output images have been transformed so grey levels match the input image dynamic range):

Input image $x[m,n]$

(a)

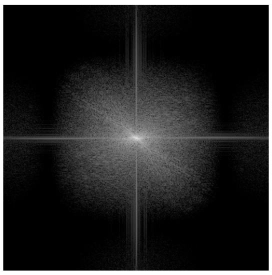
(b)



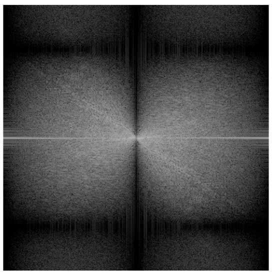
(c)

Solution: (a) → h3 high pass filter on the vertical frequencies (horizontal contours)
(b) → h1 low pass version of the input image
(c) → h2 high pass filter on the horizontal frequencies domain (vertical contours)

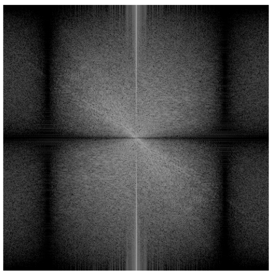
2. Using the same filters of the previous question **justify** to what images (input image, filtered with h_1, filtered with h_2 or filtered with h_3) correspond the following modulus of the Discrete Fourier Transform (centred representations):



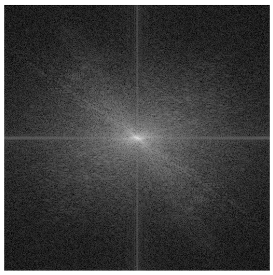
(a)



(b)



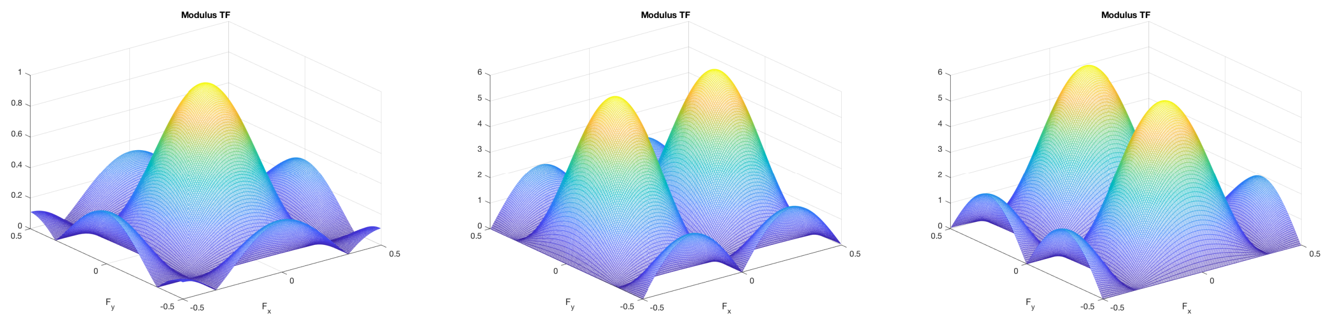
(c)



(d)

Solution: (a) → h1: low pass filtered in all directions
(b) → h2: high pass filtered in the horizontal frequencies
(c) → h3: high pass filtered on the vertical frequencies
(d) → Original image (all frequencies are preserved)

3. Using the same filters of the previous question justify which filter corresponds to which modulus of its Fourier Transform.



(a) (b) (c)

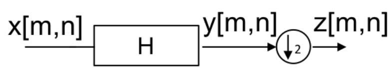
Solution: (a) → h1: low pass filter in all directions. Average filter is a sinc.
(b) → h2: high pass filter horizontal frequencies.
(c) → h3: high pass filter vertical frequencies.

4. Under which condition the Discrete Fourier Transform of MxN samples of an image, X[k,l], represents the sampled version of its Fourier Transform $X(F_x,F_y)$?

$$X[k,l] = X(F_x,F_y)\big|_{F_x=\frac{k}{M},F_y=\frac{l}{N}}$$

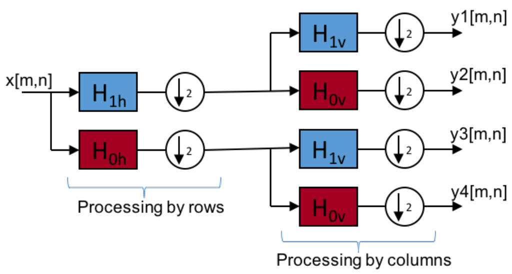
Solution: If MxN is greater than or equal to the size of the image.

5. Consider the following system using an anti-aliasing filter H and a down-sampling process by a factor of two in both the horizontal and vertical dimensions.
What would the ideal anti-aliasing filter be? What would its cut off frequency be?



Solution: Ideally, the anti-aliasing filter is an ideal loss pass filter with a cut off frequency of 1/4 in both horizontal and vertical directions.

6. 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

Problem IV	**Verónica Vilaplana**	**(1 point)**

1. The Canny edge detector uses three parameters, *sigma, low_threshold, high_threshold*. Explain what do these parameters control and what is the effect of increasing / decreasing their values.

Solution:

See Lecture8, slides 20 to 28.
Sigma: controls the width of Gaussian filter. Large sigma detects large scale edges, small sigma detects fine features.
Low-threshold and high-threshold control the hysteresis thresholding. Edges start at strong edge pixels (values higher than the high *hreshold) and continue through weak edge pixels (values higher than the low*_threshold). Decreasing/increases the high threshold tends to produce more (may be spurious) /less contours, decreasing / increasing the low_threshold produces longer (without gaps) /shorter contours (gaps between contours)

2. Explain the basic idea of the Harris corner detection algorithm. Use the notion of eigenvalues and thresholds on these. How can an edge be defined in terms of the eigenvalues defined above?

Solution:

See Lecture8, slides 40 to 50.

In the region around a corner, the image gradient has two dominant directions

The Harris corner detection algorithm computes a 2×2 matrix M at each pixel in terms of derivatives at that point and then computes the two eigenvalues of the matrix, λ1 and λ2,

$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$, the elements are the products of the first derivatives I_x and I_y.

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 $R < T1 < 0$ (or, alternatively, $\lambda 1 \approx 0$ (corresponding to the direction of the edge) and λ2 is large (corresponding to the normal direction at the edge)) (here we assume $\lambda 1 \leq \lambda 2$.)

3) labels a pixel C if $R > T2 > 0$ (or, alternatively, λ1 and λ2 are both large).

Problem V	**Ramón Morros**	**(2 points)**

1. Segmentation with region merging

The similarity criterion for a region merging segmentation algorithm is given by:

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

where:

$$C_{color}(R_1, R_2) = N_{R1} \| M_{R1} - M_{R1 \cup R2} \|_2^2 + N_{R2} \| M_{R2} - M_{R1 \cup R2} \|$$
$$C_{contour}(R_1, R_2) = -Length\ of\ common\ contour$$

a) Explain the operation of the region merging algorithm.

b) Describe the effects of the two components C_{color} and $C_{contour}$. Particularize the explanation for the two particular cases where $\alpha = 0$ and $\alpha = 1$.

Solution: Segmentation with region merging

a) See slide 66 in MCV_M1_L10_-_Grouping_segmentation_Classification_II.pdf

b) C_{color}: Defines the similarity between regions in terms of the MSE between the mean of the pixels in each region and the mean of the merging of the two regions. Penalizes mergings between regions with different gray levels.
$C_{contour}$: A factor that penalizes mergings that increment the common contour length. When Alpha = 1, the effect of the contours is not considered. The segmentation may have ‘noisy’ contours. When Alpha = 0, gray level similarity is not considered. The resulting regions have very smooth contours but do not reflect precisely the true contours of the objects.

2. RANSAC algorithm

a) A key parameter in the RANSAC algorithm is the threshold used to determine if a given point is an inlier or an outlier. Explain the effect of using a too small or too large value for this parameter when using RANSAC to obtain the parameters of a line given a set of contour points in an image.

b) Let’s assume that, for each point, we have a quality score that measures the likelihood of the point belonging to the model. Explain how these scores can be used to improve the operation of the RANSAC algorithm (This is known as the PROMEDS algorithm)

Solution: RANSAC

a) Threshold too small: small misalignments will be considered as outliers. This may prevent the line being detected because of too few inliers. Threshold too large: non-collinear points may be considered as inliers, giving a false estimation of the line parameters.

b) Points are ordered from best to worse quality. The algorithm starts picking subsets of points among the points having the largest quality. If a suitable solution having a minimum number of inliers is not found, the algorithm selects matches with lower quality. Usually, solutions are found with a few iterations. Much faster than RANSAC, with similar quality.