Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M1. Introduction to human and computer vision** **Final exam**
Date: November 27th, 2017 **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. After much hard work you develop a novel denoising algorithm that you believe will outperform existing methods. The aim of the algorithm is to reduce the visibility of noise in an image and the algorithm uses a state-of-the-art model of the human visual system to predict the visibility of noise in an image. The algorithm works by attempting to minimize the estimated visibility of noise across the whole image.

   To evaluate the model against other methods you can either run a subjective test to see which denoising method subjects prefer. Alternatively, a number of papers use a state-of-the-art image quality metric (IQM) to evaluate the denoising algorithms instead. The IQM has been developed by an external research group and provides a local estimate of the visibility of noise across an image.

   What are the advantages and disadvantages of using an IQM?

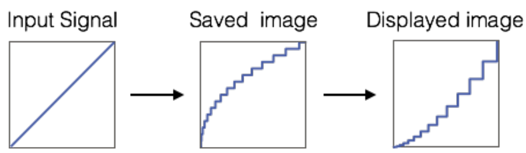
The answer may receive a maximum of two points; the point reward for each correct answer is shown in the brackets.

Solution:

Advantages
- IQM is quick.
- The results of an IQM are independent of the experimenter.

Disadvantages
- An IQM may only have limited predictive power and thus may incorrectly predict the best metric.
- It may be impossible to evaluate the statistical significance of the IQM results.
- Both the IQM and your denoising algorithm attempt to model the visibility of noise in an image. This is the circularity problem. There is no way to know which model is better and thus the use of the IQM has no scientific value.
- The results of the IQM may not be valid for image sets, viewing conditions, research questions and noise types that are different from the originally ones used to design and evaluate the IQM.

2. A simplified schematic of the image-processing pipeline is shown below. Please (a) describe the encoding and decoding non-linearities (b) explain how this pipeline helps reduce the visibility of quantization artifacts in the displayed image and (c) explain why the pipeline may have to be re-evaluated for high dynamic range displays or different viewing conditions such as a cinema or home computer.



Solution:
- The encoding nonlinearity is approximately the inverse of the decoding nonlinearity.
- Because quantization is applied before the decoding nonlinearity quantization in non-linear.
- This non-linearity quantization is close to the nonlinear processing of the HVS.

- As perception varies with dynamic range and the viewing condition, the shape of the encoding and decoding nonlinearities must be reevaluated.

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

1. What is the color constancy property of the human visual system? How do cameras emulate it?
   Answer in section X of the course notes.

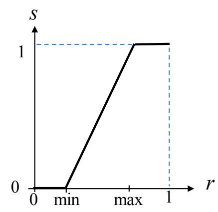
2. List and briefly explain all the color processing operations that are performed in-camera.
   Answer: sections VII-G, X, XI of lecture notes.

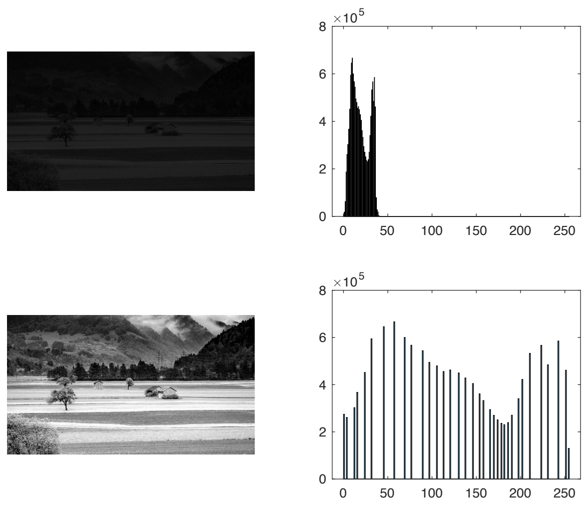
Problem III	**Philippe Salembier**	**(2 points)**

1. We want to increase the contrast of an image. To this goal, we can use either a range mapping as the one described in the figure on the right or an histogram equalization.

   The figure below shows on the first row an original image and its histogram. After contrast enhancement, we obtain the image and the corresponding histogram shown in the second row of the figure.

   Which contrast enhancement technique has been used? Justify your response.





Solution: Histogram equalization. We see that the spacing between bins is not always the same. We also see that some gray level values of low probability have been merged.

2. Define the notions of Precision, Recall and F-score.
Solution: The precision is the relation between True Positive (TP) and Positive detections (P), that is TP/P. The Recall is the relation between True Positive (TP) and True samples (T), that is TP/T. The F-score is the harmonic mean of Precision and Recall: 2 P R / (P+R)

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

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

   Compute the dilation with this structuring element of the following image.

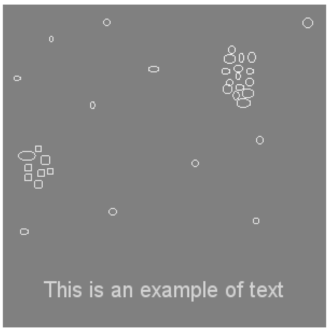
$$\begin{array}{cccccc} 100 & 100 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 & 100 \\ 100 & 200 & 200 & 200 & 200 & 100 \\ 100 & 100 & 100 & 100 & 100 & 100 \\ 100 & 50 & 50 & 50 & 50 & 100 \\ 100 & 100 & 100 & 100 & 100 & 100 \end{array}$$

   Note: If necessary zero-padding may be assumed.

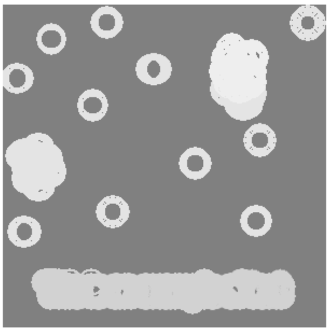
   Solution:

100	100	100	100	100	0
200	200	200	200	100	100
100	100	100	100	100	100
100	200	200	200	200	100
100	100	100	100	100	100
100	50	50	50	50	100

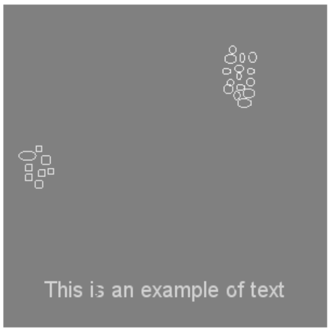
4. Consider the operator $\psi(f) = \delta_\circ(f) \wedge f$, where f is the original image, $\delta_\circ(.)$ a dilation with an annular structuring element (not a disc) centered on the space origin (without including it).
   a) Analyze de algebraic properties of $\delta_\circ(.)$ to define whether it is extensive, increasing and idempotent.
   b) Assuming that $\psi(.)$ is idempotent, show that $\psi(.)$ is an opening.
   c) The opening $\psi(.)$ is known as the annular opening. The following figure shows a processing example. What is the practical interest of this opening?



Original image: f



Dilation: $\delta_\circ(f)$



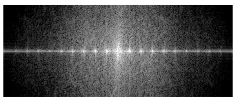
Annular opening: $\psi(f)$

Solution:
   a) $\delta_\circ(.)$ is a dilation. So, as any dilation, it is increasing and not idempotent. Moreover, as the space origin is not included in the structuring element, the dilation is not extensive (nor anti-extensive).
   b) To show that $\psi(.)$ is an opening we need to demonstrate that it is increasing and antiextensive.
      a. Increasing: if $f \leq g$, then $\delta_\circ(f) \leq \delta_\circ(g)$ (as the dilation is increasing). By taking the infimum of both inequalities, we get: $\delta_\circ(f) \wedge f \leq \delta_\circ(g) \wedge g \Rightarrow \psi(f) \leq \psi(g)$. This demonstrates that $\psi(.)$ is increasing.
      b. Antiextensive: the computation of $\psi(.)$ involves taking the minimum of the dilation with the input signal. So, the output signal cannot be larger than the input. The operator is antiextensive.
   c) The annular opening allows to remove maxima that are isolated. Groups of maxima that are close to each other (in comparison with the diameter of the ring used as structuring element) are preserved.

Problem IV Javier Ruiz (3 points)

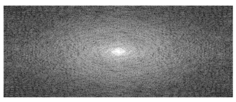
1. Justify which modulus of the Discrete Fourier Transformation (centered representation on the right) corresponds to which image (on the left).

(A)



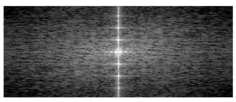
(a)

(B)



(b)

(C)



(c)

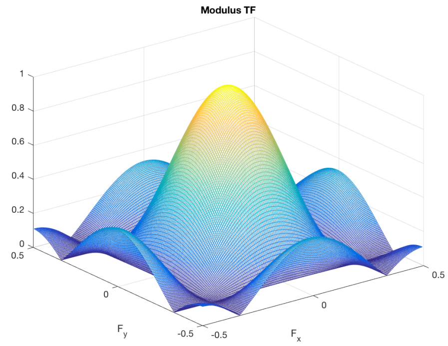
Solution: (A) → (b) High frequencies with no specific direction
(B) → (a) Vertical contours that correspond to horizontal frequencies

(C) → (c) Horizontal contours that correspond to vertical frequencies

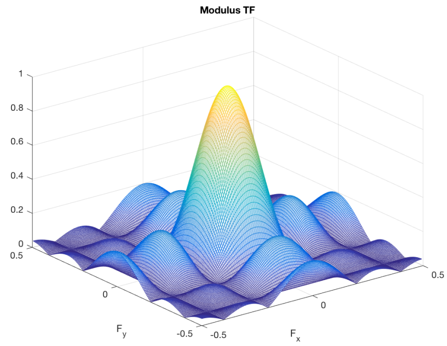
2. Using the modulation property of the Discrete Fourier Transform (DFT), express the DFT of MxN samples of the image $\tilde{x}[m-1,n]$ (periodic version of x[m,n]) in terms of $X[k,l]$, the DFT of MxN samples of x[m,n]

Solution: $X[k,l] = DFT_{MxN}\{\tilde{x}[m-1,n]\} = X[k,l] \cdot e^{-j2\pi\frac{k}{M}}$

3. Consider two average filters of size 3x3 and 5x5. Justify which filter size corresponds to which modulus of its Fourier Transform.



(a)



(b)

Solution: (a) 3x3 cut-off frequency is around 1/3 for both horizontal and vertical frequencies.
(b) 5x5 cut-off frequency is around 1/5 for both horizontal and vertical frequencies.

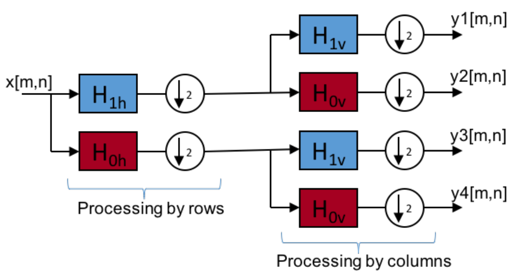
4. Enumerate the three impulse responses that can be used to approximate the vertical gradient $\frac{\partial f(x,y)}{\partial y}$ of an image $f(x,y)$. How can they be used to detect contours in an image?
Solution: Forward difference: $h_f[n] = \{1, \underline{-1}\}$; Backward difference: $h_b[n] = \{\underline{1}, -1\}$; Symmetric difference: $h_s[n] = \{1, \underline{0}, -1\}$. Contours can be detected by finding maxima and minima on the output.

5. Given a decomposition of an image X into a laplacian pyramid elements with 3 levels: L_1, L_2 and X_3. Give the steps and equation to re-construct image X from L_1, L_2 and X_3 (use F_i to denote the blur-and-upsample operator at level i of the pyramid).
Solution:

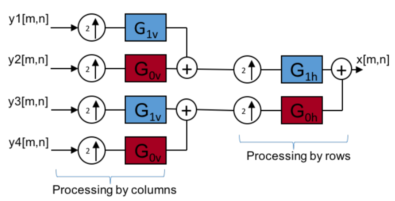
$X_2 = F_2X_3 + L_2$

$X = F_1X_2 + L_1 = F_1(F_2X_3 + L_2) + L_1$

6. Given the following wavelet decomposition of an image x[m,n], draw the corresponding filter-bank to reconstruct the image from the decomposition y_1[m,n], y_2[m,n], y_3[m,n] and y_4[m,n] (consider H_0, H_1, G_0 and G_1 bi-orthogonal filters).



Solution:



1. **Canny's edge detection:**

   a. Calculating the gradient magnitude and angle is part of the Canny's method. Below you can see a part of an image $f(x,y)$

f(x,y)

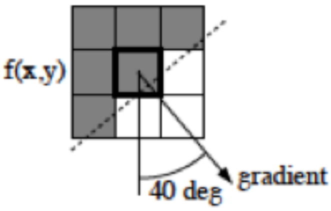
2	3	3	3	3	3
1	2	2	3	3	3
0	0	$\boxed{2}$	2	2	3
0	0	0	0	1	2
0	0	0	0	0	1

Use these two masks

$$h_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} / 8 \quad , \quad h_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} / 8$$

to calculate the gradient magnitude $M(x,y)$ and angle at the marked pixel. Note: As you are not allowed to use calculator for this exam, you may define the gradient angle with a closed-form expression involving classic trigonometric functions.

   b. Non-maxima suppression is part of the Canny's method. In the surrounding image $f(x,y)$ illustrated below, the gradient direction in the central pixel is given. Also, the corresponding gradient magnitude $M(x,y)$ is given.



M(x,y)

3	3	3
3	$\boxed{4}$	5
3	5	5

Will the central pixel in $M(x,y)$ be suppressed or not? Describe carefully how you make the decision.

Solution:

a. At the marked point

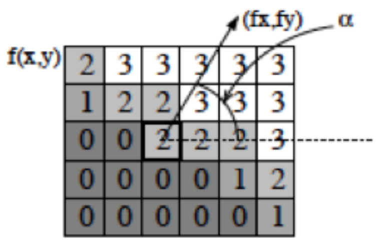
$$f_x = \frac{\partial f(x,y)}{\partial x} \approx \frac{1.3 + 2.2 + (-1).2}{8} = \frac{5}{8}$$

$$f_y = \frac{\partial f(x,y)}{\partial y} \approx \frac{1.2 + 2.2 + 1.3}{8} = \frac{9}{8}$$

Therefore

$$M(x,y) = \sqrt{f_x^2 + f_y^2} = \frac{\sqrt{106}}{8}$$

$$\alpha = \arctan\left(\frac{9/8}{5/8}\right) .180^{\circ}/\pi \approx 61^{\circ}$$



b. For the actual gradient direction, the pixel pair in $M(x,y)$ marked with circles should be considered:

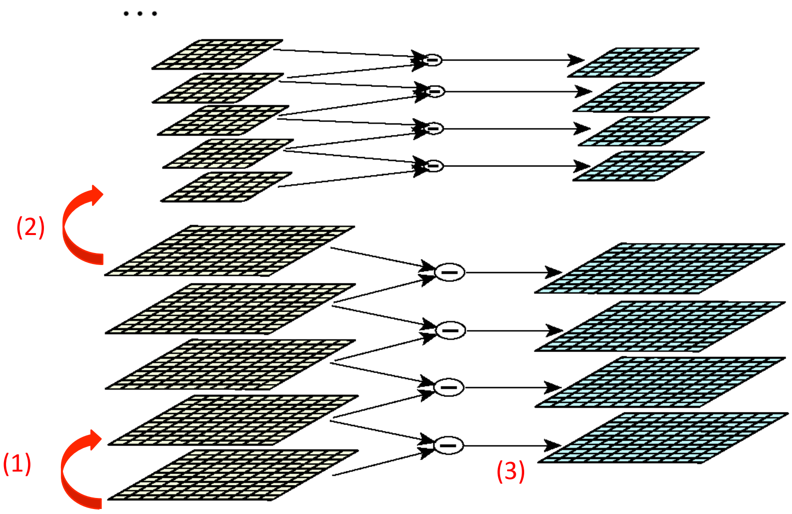
M(x,y)

3	3	3
3	$\boxed{4}$	5
3	5	5

Since for one of the pixels $M(x,y) = 5 > 4$, the central pixel in $M(x,y)$ will be suppressed.

2. Question 2: Sift detector

The SIFT (Lowe) detector finds keypoints using the image pyramid shown in the following figure



a) Explain which operations are performed in steps (1), (2) and (3).
b) Explain how keypoints are found using this pyramid.

Solution

a) (1) Convolution with a Gaussian filter at scale kσ (the original image is convolved with a Gaussian filter at scale σ, the next image is obtained by convolving the previous result with a Gaussian filter at scale kσ)
(2) Downsampling by a factor of 2 in each dimension
(3) Difference of two Gaussian filtered images.

b) We compare each pixel in the DOG images to its neighbors in space and scale: 8 neighbors at the same scale and 9 neighbors in each of the adjacent (upper and lower) scales. A pixel is selected as candidate keypoint if its value is the maximum or minimum among the compared pixels.

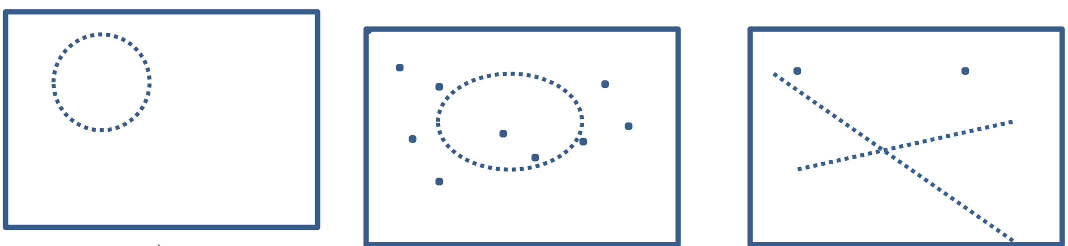
Problem VI	Ramón Morros	(2 points)

1. Let us suppose we are using the Hough transform to find ellipses in an image of WxH pixels. 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 an ellipse with center (h,k) and semi-axes (a,b) can be written as:

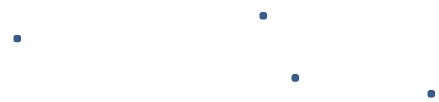
$$h = x - \mathrm{a} \cdot \cos(t)$$
$$k = y - b \cdot \sin(t)$$

Solution:
We quantize the semi-axes a and b into R levels each and the parameter t into T levels. Then, for each contour point (x,y), the parameters (h,k) are computed and stored for each combination of a, b and t. This is, $N \cdot R^2 \cdot T \cdot$ (2add + 2mult + 2 trig) operations. At the end of this step, we have to find the M maxima of the resulting $W \cdot H \cdot R^2$ matrix (a,b,h,k). This can be done using $M \cdot W \cdot H \cdot R^2$ comparison operations.

2. Given the cases represented in the following figure, discuss the advantages and drawbacks of Least Square (LS) vs. Hough Transform vs. RANSAC to find instances of the given shape(s) in an image.
   a. Single object (circumference), no noise
   b. Single object (ellipse), noise can be present
   c. More than one object (lines), noise can be present



a)



b) c)

Solution:
a) The three methods can detect the shape. However, the best option is LS because it will be much faster and has no parameters.
b) LS is not a robust method so it is not a good option in presence of noise (same for c). Both RANSAC and Hough transform will be able to solve the problem but the high number of parameters will result in a high number of operations for the Hough transform.
c) In this case, the Hough transform is the best method because it is the only of the three methods that can handle multiple instances.

3. Segmentation using Mean-Shift (MS):
   a. Explain the steps of the mean-shift algorithm for image segmentation in order to obtain an image partition.
   b. In remote sensing, satellite images are often captured with multispectral sensors that capture data at several wavelengths. For instance, Landsat 8 satellite collects data from 9 spectral bands and there are sensors with a larger number of bands. Discuss the application of the MS algorithm in the case of high dimensionality feature spaces.

Solution:
a) For each point in the feature space, run the MS algorithm (slides, page 48). This assigns a label for each point in the feature space. These labels are translated to each pixel of the image, resulting in a classification of the pixels. To obtain a segmentation, a connected component labelling step has to be applied.
b) In case of a high dimensional feature space (FS), methods based on FS segmentation suffer the curse of dimensionality problem: the number of feature points is given by the number of pixels in the image and is much smaller than the number of points in the feature space. This results in very sparse feature representations. It is difficult to estimate the maxima of the pdf in these conditions.

4. In region growing segmentation, describe the concept of markers.

Solution:
Markers are initial seeds for the objects to segment. They are connected regions that mark the interior of the objects to segment. They form a partial segmentation of the image. The pixels in the uncertainty zone (non-labelled pixels, the ones that do not belong to any of the markers) are progressively added to these initial seeds until the final partition is obtained.