Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M1. Introduction to human and computer vision** **Final exam**
Date: December 2nd, 2019 **Time: 2h30**
Teachers: Ramon Morros, Javier Ruiz, Philippe Salembier, Javier Vázquez, 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 Javier Vázquez (2 points)

1. Which are the cells responding to light in the retina? Which of them are the ones focusing on color? Explain in half a page their main characteristics.

2. What is colour constancy? Explain the Von Kries Law.

3. List the main steps in the camera colour processing pipeline, giving a short -1 line- explanation for each of the steps.

4. Explain unsharp masking. Why is it needed in the camera processing pipeline?

Problem II Philippe Salembier (2 points)

1. In this problem, our goal is to evaluate the performances of two image classification systems, which have been designed to differentiate between urban images and non-urban images. In order to perform an objective assessment a small database of images has been created. It is composed of the following 10 images.

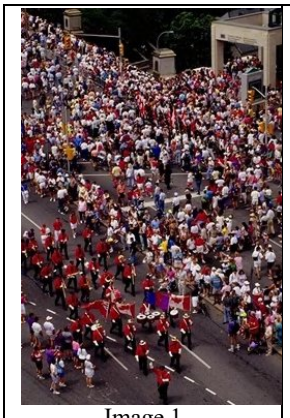


Image 1


Image 2


Image 3

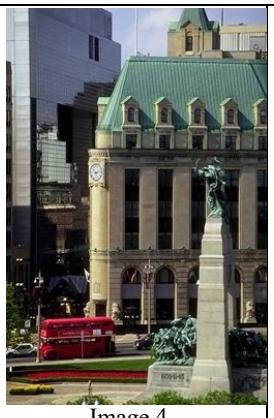


Image 4


Image 5

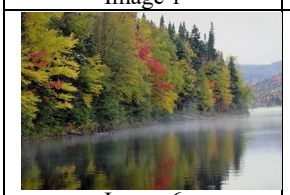


Image 6

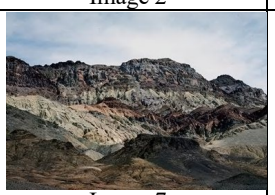


Image 7

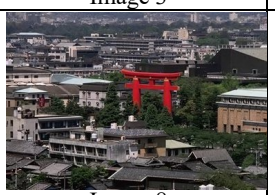


Image 8

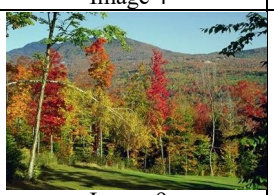


Image 9

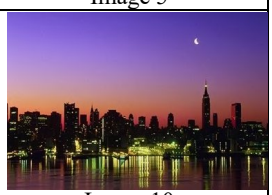


Image 10

It has been decided to assess the performances of the systems with the F measure which is the geometric mean of Precision, P, and Recall, R; that is $F = 2.P.R/(P + R)$

The two systems are run on the database and give the following classification results:

System	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9	Image 10
A	Urban	Non-urb	Urban	Urban	Urban	Non-urb	Urban	Urban	Non-urb	Urban
B	Non-urb	Non-urb	Urban	Urban	Urban	Non-urb	Non-urb	Urban	Non-urb	Urban

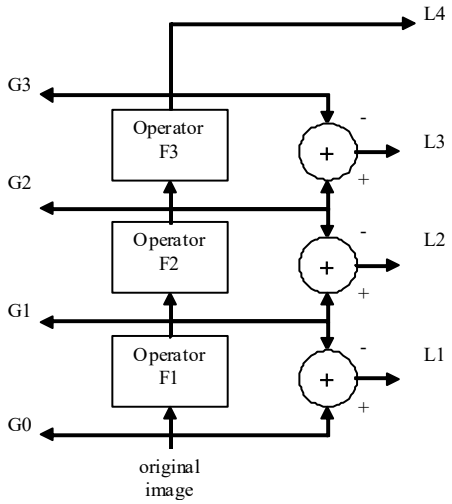
Note that both systems make two incorrect classifications. Both classify *Image 5* as an urban image, which is not the case. Moreover, System A is based on color information and misclassifies *Image 7* assuming that we see a lot of areas made of concrete. On its turn, System B, which is based on texture, estimates that *Image 1* is from a non-urban scene.

Assuming that the class of **"True" samples is the class of Non-urban images**, compute the precision, recall and F values of both systems and decide which one has the best performances.

2. Consider the following range transform: $s = T(r) = 4(\mathrm{r} - \frac{1}{2})^2$, for $0 \leq r < 1$
   Justify whether this transform increases, maintains or reduces the entropy of discrete images.

3. State the three important algebraic properties of a closing and define mathematically each property.

4. Multiscale image decomposition can be performed with the scheme of the following figure. In the sequel, various schemes are created with several operators Fi. In all cases, you can ignore the image border effects (that is you can consider that the image is of infinite size).



   a) Demonstrate that, for any operators Fi, the sum of the Li images is always equal to the original image.
   b) Assume that F1 is an opening, F2 a closing and F3 a closing of opening (closing(opening(.)). All operators use the same flat structuring element of size 5x5. Describe the content of the Li images.

   c) Assume that F1 is an erosion with a horizontal flat structuring element 1x3, F2 an erosion with a vertical flat structuring element 3x1 and F3 a dilation with a square structuring element of size 3x3. Is there any order relationship between L4 and G0?

   d) Assume that F1 is an opening with a flat structuring element of size 5x5, F2 an opening with a flat structuring element of size 3x3 and F3 a closing with a flat structuring element of size 5x5. Describe the content of the Li images.

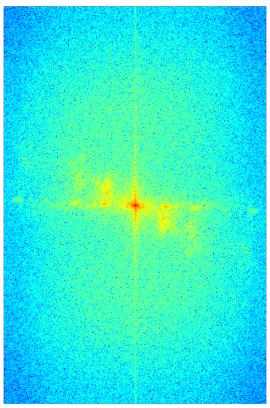
Problem III Javier Ruiz (3 points)

1. Consider the following LSI filter with impulse response (kernel) of size 3x3: $h[m,n] = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$ and the image $x[m,n] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ of size 4x4.

   a. What is the size of the filtered image $y[m,n] = x[m,n] * h[m,n]$?
   b. Compute the values of the filtered image $y[m,n] = x[m,n] * h[m,n]$ (If necessary zero-padding may be assumed).
   c. Does the filter correspond to a low-pass or high-pass filter? In which (horizontal or vertical) component?
   d. Justify if the filter detects any vertical or horizontal contours.

2. We want to analyse the frequency content of an image of 330x500 pixels (Figure a) with its DFT-2D of 330x500 samples (Figure b):

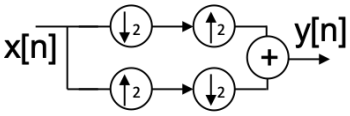
(a) Original image



(b) DFT-2D magnitude

   a. What discrete frequencies (F1 horizontal and F2 vertical) correspond to the DFT sample located at column 202 / row 251? column 130 / row 251? column 202 / row 271 and column 130 / row 231? (remember that Figure b shows the centered representation of the DFT)
   b. What region of the image corresponds to those frequencies?

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



4. Given a decomposition of an image X into a Laplacian pyramid elements with 3 levels: L_1, L_2 and G_3.

   a. Comment briefly the differences between Gaussian and Laplacian pyramids.
   b. Compute the number of samples (pixels) of the Laplacian representation for an image of size NxN.
   c. Is the Laplacian pyramid a complete or overcomplete representation of the image X?

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

6. Enumerate two main ways to reduce the resolution of images/features at each layer of convolutional neural networks?

Problem IV Verónica Vilaplana (1 point)

1. Compare and contrast Harris and Difference of Gaussians (DoG) detectors.

2. Assuming feature points have been previously detected using the SIFT feature detector, briefly describe the main steps of creating the SIFT feature descriptor at a given feature point.

Problem V Ramón Morros (2 points)

1. Describe how the gradient can be used to improve the Hough algorithm for detecting shapes in images. Explain the benefits of using the gradient in Hough transform.

2. In the RANSAC algorithm, what is the relationship between the number of outliers and the necessary number of iterations?

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

4. Explain briefly the difference between region merging and region growing methods in segmentation.