Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M1. Introduction to human and computer vision** **Final exam**
Date: November 30th, 2020 **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.
- All results should be demonstrated or justified.

At the beginning of your exam, please write the following statement and sign below:

I hereby certify that I am doing this exam without using any books, lecture notes, personal notes, video related to the course content and that I have no communication with anyone beside the teachers.

PB 1 (0.5 point): Explain the idea of color opponency.
Solution: Color opponency states that in our Visual System, the retina first obtains the signals for the Long, Medium, and Short Cones, and then these signals are recombined into three opponent channels. The opponent channels are two chromatic and one achromatic, being the chromatic ones red versus green, and blue versus yellow, and the achromatic one white versus black. The idea of opponency can be easily understood by thinking that we cannot imagine colors such as reddish green or blueish yellow because they are opponent between them.

PB 2 (0.5 point): Explain the CIE Lab color space. Which is its main advantage over the CIE XYZ space?
Solution: CIELab color space is an opponent-based color space, and therefore it presents 1 channel for luminance and two channels for chrominance. Furthermore, the two channels fro chrominance represent the red versus green opponency and the blue versus yellow opponency. The main advantage of CIELab over CIEXYZ is that CIELAB is a perceptually-based space, meaning that distances in this space correlate with the distances a human observer will perceive. This is not that case in the CIEXYZ color space, as in that case a small distance in the blue part of the color space is equivalent to a much larger distance in the green part of the color space, as can be visualized by looking at the MacAdam ellipses.

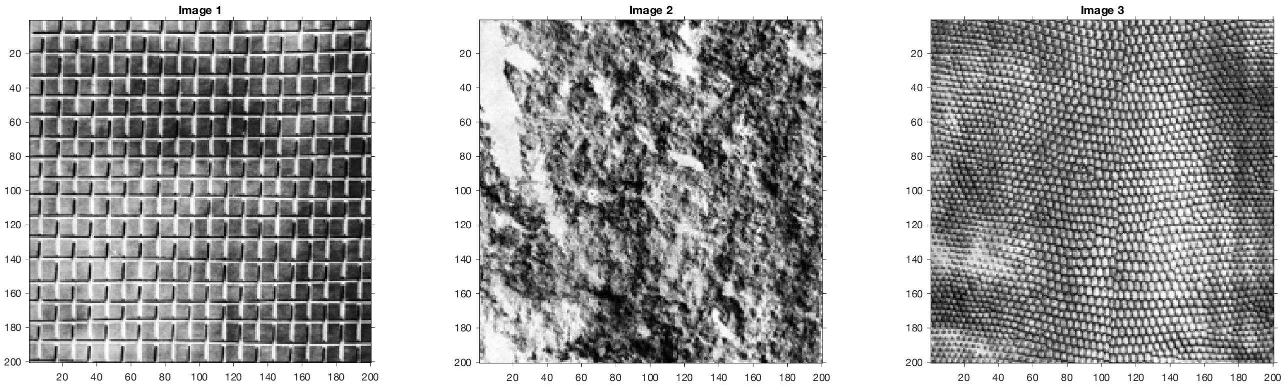
PB 3 (0.5 point): Explain the Grey-World and the MaxRGB methods for white balance
Solution: White balance methods aim at discounting the illuminant presented in the scene. To this end, both Grey-World and MaxRGB assume some conditions on the image captured.
Grey-World: The Grey-World method assumes that the World is Grey in average, and therefore any image captured under the white illuminant will have a grey average. So, if we are given an image (with channels R,G,B) that has a per channel average: R_{average}, G_{average}, and B_{average}, the Grey-World hypothesis will modify it in the following way:
R_new=R*(127/R_{average})
G_new=G*(127/G_{average})
B_new=B*(127/B_{average})

MaxRGB: The MaxRGB method assumes that there are in the image either a White object, or Red, Green, and Blue objects (i.e. objects aiming to maximize the response in each of the channels). From here, MaxRGB also assumes that an image captured under the white illuminant will have the maximum response of 1 in each of the channels, modifying the input image (R,G, B) as follows:
R_new=R*(255/R_{max})
G_new=G*(255/G_{max})
B_new=B*(255/B_{max})

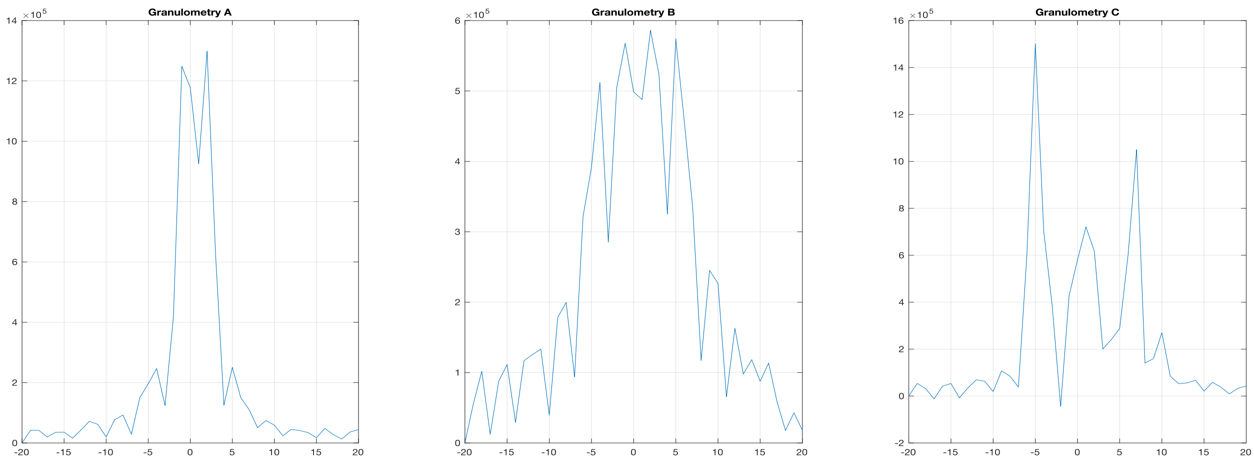
PB 4 (0.5 point): Explain the basic auto-exposure method that cameras use.
Solution: Most auto-exposure algorithms work as follows:
1) Take a (temporary, not to be recorded) picture with a predetermined exposure value EV_{pre}.

2) Compute a single brightness value B_{pre} from that picture.
3) Select an optimal brightness B_{opt}
Then, they compute the optimal exposure as:
EV_{opt}= EV_{pre}+ log2(B_{pre}/B_{opt})
Finally, they modify the aperture and shutter speed (and gain) to take the exposure-corrected image.

PB 5 (0.5 point): Considering the following three images: image 1, 2 and 3.



We have computed their granulometry with circular structuring element. The three pattern spectrums: granulometry A, B and C are shown below.



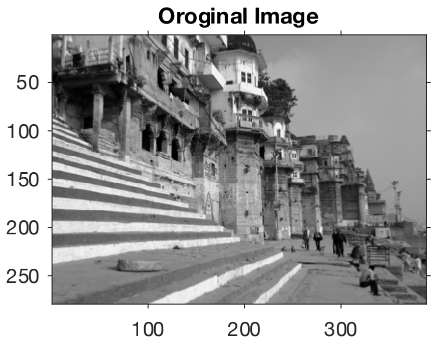
Define the correspondence between the granulometric curves (A,B,C) and the images (1,2,3). **Justify precisely your reasoning.**
Solution:

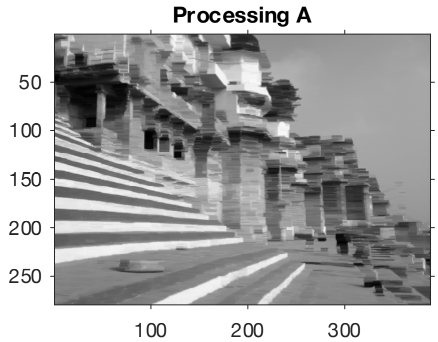
Granulometry A = image 3 (the image involves essentially small minima and maxima)
Granulometry B = image 2 (the texture is noisy and involve minima and maxima of quite different sizes)
Granulometry C = image 1 (large minima can be observed and small macximas)

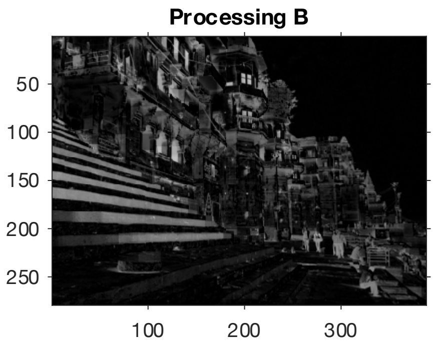
PB 6 (0.5 point): The original image of size 388x279 shown in the upper left part of the following figure has been processed by three different operators:

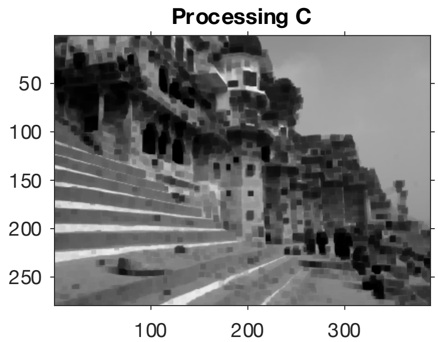
- ❑ An erosion with a horizontal structuring element on length 5 followed by an erosion with a vertical structuring element on length 5.
- ❑ A closing with a horizontal structuring element on length 7 followed by a closing with a horizontal structuring element on length 7.
- ❑ A dual tophat involving a closing with a structuring element of size 13x13.

The results of are shown in Processing A, B and C. Assign an operator to each of these results. Justify your answer taking into account the various operator properties and the observed visual effects.









Solution:

- Processing A = A closing with a horizontal structuring element on length 7 followed by a closing with a horizontal structuring element on length 7. This is equivalent to a closing with a horizontal structuring element on length 7.
- Processing B = A dual tophat involving a closing with a structuring element of size 13x13. Small minima of the original image appear as bright elements.
- Processing C = An erosion with a horizontal structuring element on length 5 followed by an erosion with a vertical structuring element on length 5. This is equivalent to an erosion with a 5x5 square structuring element.

PB 7 (0.5 point): Consider the image I below (left side) and its negative (right side). All images are quantized with 4 bits.

Image I

0	4	4	12	4
0	4	4	12	0
0	0	4	15	12
4	0	4	12	15
4	15	4	15	15

Negative of I

15	11	11	3	11
15	11	11	3	15
15	15	11	0	3
11	15	11	3	0
11	0	11	0	0

Compute in both cases the image after histogram equalization. Are the resulting images negative of each other?

Solution: The images after histogram equalization are:

4	10	10	12	10
4	10	10	12	4
4	4	10	15	12
10	4	10	12	15
10	15	10	15	15

15	11	11	5	11
15	11	11	5	15
15	15	11	3	5
11	15	11	5	3
11	3	11	3	3

They are not negative of each other.

PB 8 (0.5 point): We would like to construct two sets of structuring elements of increasing sizes in order to use them to compute granulometries. The first set is based on SE_1^A and the second one on SE_1^B (show below). In each case, the locations included in the structuring element is shown with "×".

$SE_1^A[m,n] = \begin{bmatrix} . & \times & . \\ \times & . & \times \\ . & \times & . \end{bmatrix}$ and $SE_1^B[m,n] = \begin{bmatrix} \times & \times & . \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix}$

We consider these structuring elements are of size one (represented by the sub-index 1) and in order to create structuring elements of larger size we simply dilate them by themselves. As examples, we have:

$SE_2^A[m,n] = SE_1^A[m,n] \oplus SE_1^A[m,n]$, $SE_3^A[m,n] = SE_1^A[m,n] \oplus SE_2^A[m,n]$, etc.

$SE_2^B[m,n] = SE_1^B[m,n] \oplus SE_1^B[m,n]$, $SE_3^B[m,n] = SE_1^B[m,n] \oplus SE_2^B[m,n]$, etc.

Compute the structuring elements $SE_2^A[m,n]$ and $SE_2^B[m,n]$.
Can we use the set of SE_k^A or of SE_k^B structuring elements to compute granulometries?

Solution:
$SE_2^A[m,n]$

```
. . . . . . .
. . . x . . .
. . x . x . .
. x . x . x .
. . x . x . .
. . . x . . .
. . . . . . .
```

$SE_2^B[m,n]$

```
. . . . . . .
. x x x . . .
. x x x x . .
. x x x x x .
. x x x x x .
. x x x x x .
. . . . . . .
```

SE_k^A cannot be used to compute a granulometry because they are not convex.
SE_k^B can be used to compute a granulometry because they are convex.

PB 9 (1.0 point): Consider the image x[m,n] with values between [0,1], X[k,l] its Discrete Fourier Transform (DFT) of NxN samples (X[k,l] =DFT_{NxN}{x[m,n]}) and X(F_x,F_y) the Fourier Transform of x[m,n]

a. Justify the maximum size of the input image so 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): $X[k,l] = X(F_x, F_y)\big|_{F_x=\frac{k}{N}, F_y=\frac{l}{N}}$

Consider the size of the input image NxN and the image y[m,n] = 1 + x[m,n].

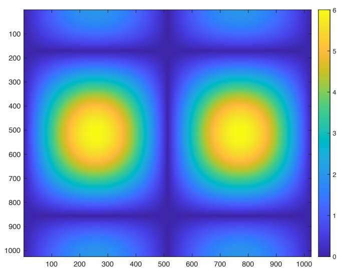
b. Compute Y[k,l]=DFT_{NxN}{y[m,n]}, in terms of X[k,l]
c. Compute the inverse DFT of Y[k,l] - X[k,l]

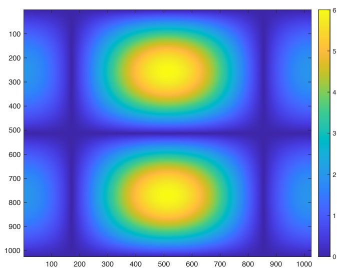
Solution:
a) Size of the image should be less or equal to NxN
b) Y[k,l]= N^2 δ[k,l] + X[k,l]
c) DFT^{-1}{Y[k,l] - X[k,l]} = DFT^{-1}{N^2 δ[k,l]} =1

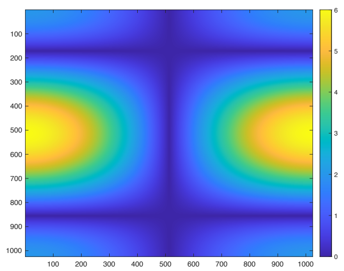
PB 10 (1.0 point): Consider the following filter h[m,n] with impulse response (kernel) of size 3x3 and the image x[m,n] of size 4x4:

$$h[m,n] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \qquad x[m,n] = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

a. Compute the filtered image $y[m,n] = x[m,n] * h[m,n]$ (If necessary zero-padding may be assumed).
b. Does the filter correspond to a low-pass or high-pass filter? In which (horizontal or vertical) component?
c. Justify if the filter detects any vertical or horizontal contours.
d. Justify which on the following 3 images represents the modulus of the DFT of the filter h[m,n]


A)


B)


C)

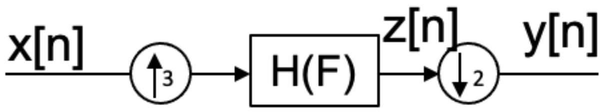
Solution:

a)

1	2	4	5	4	2
1	2	4	5	4	2
-1	-2	-4	-5	-4	-2
-1	-2	-4	-5	-4	-2
0	0	0	0	0	0
0	0	0	0	0	0

b) The filter corresponds to a high pass filter on the vertical component.
c) Horizontal contours relate to high frequency in the vertical component.
d) B) as the filter is high pass in vertical and low pass in horizontal.

PB 11 (1.0 point): Consider the following decomposition using down-sampling and up-sampling processes (without filtering).



1. Assume we are using an ideal filter H(F). Would you use a low-pass or a high pass filter?
2. Using again an ideal filter H(F), what would the cut off frequency be? (use normalized discrete frequency values)
3. Are there any aliasing in the signal z[n]?
4. Express the Fourier Transform Z(F)=FT{z[n]} as a function of X(F)=FT{x[n]} and H(F).
5. Express the Fourier Transform Y(F)=FT{y[n]} as a function of Z(F)=FT{z[n]}.

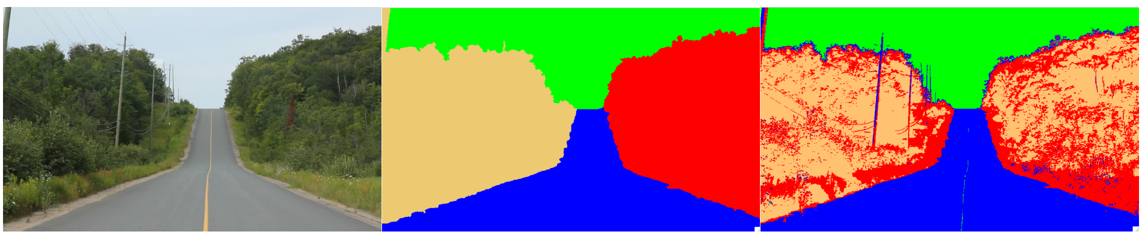
Solution:
a) Low-pass
b) 1/(2*3) = 1/6
c) No, there is no aliasing in the up-sampling process.
d) Z(F) = X(3F)H(F)
e) Y(F)=1/2Z(F/2) + 1/2Z(F/2-1/2)

PB 12 (0.5 point): Describe the Laplacian of Gaussian blob detector. In particular, explain the motivation and the idea of the scale selection.
Solutions:
1. Slides 58 – 63

PB 13 (0.5 point): Describe the SIFT descriptor. That is, describe how the pixel neighborhood around a detected keypoint is converted to a 128-dimensional feature vector. Mention at least two computer vision tasks or applications where keypoint detectors/descriptors like SIFT are commonly used. Explain what is the benefit of using SIFT in the applications (e.g. compared to methods which are not scale invariant).

Solutions:
2. Slides 82 – 87

PB 14 (0.5 point): We have applied two segmentation algorithms, k-means and watershed, to image (1). For k-means, clustering has been performed in the RGB space using k=4. For watershed, 4 markers have been defined manually by drawing 4 strokes over the image and using the strokes as markers. The results are shown in images (2) and (3). Explain which image corresponds to k-means and which to watershed (justify your answer). In each case, explain if an extra step is needed to obtain a partition.



(1) (2) (3)

Solution: (2) has been obtained with watershed and (3) with k-means. The image label (3) shows regions that are not connected (it is a classification, not a partition. This is typical of the methods performing segmentation in the feature space. Watershed provides connected regions (partition), as in image (2).
To obtain a partition for image (3), a re-labeling step is necessary. Image (2) is already a partition so no further steps are necessary.

PB 15 (0.5 point): Hough Transform:

1. Write the pseudocode to detect circumferences in a contour image using the Hough transform.
2. Assuming that we know the gradient at each contour position, explain how this information can be used to improve the previous algorithm, Provide the pseudo-code of the modified algorithm.

   Equation of a circumference: $\begin{aligned} a &= x - r\cos(t) \\ b &= y - r\sin(t) \end{aligned}$

Solution: Pseudocode:

```
int P[amax][bmax][rmax]; // accumulators
for each edge point (x, y) {
    // Compute parameters
    for (r = rmin; r <= rmax; r = r+Δr) {
        for (t = 0; t <= tmax; t = t+Δt) {
                    a = x - r·cos(t);
                    b = y - r·sint(t);
                    P[a][b][r]++;
        }
    }
}
```

The gradient direction provides the orientation of the edge, so there is no need to loop over all possible angles. This results in faster computation.

```
int P[amax][bmax][rmax]; // accumulators
for each edge point (x, y) {
    // Compute parameters
    for (r = rmin; r <= rmax; r = r+Δr) {
                        t = gradient_orientation(x,y)
                a = x - r·cos(t);
                b = y - r·sint(t);
                P[a][b][r]++;
    }
}
```

PB 16 (0.5 point): RANSAC: Explain the LMedS algorithm and its advantages over RANSAC

Solution: (MCV_M1_L09_-_Grouping_segmentation_Classification_I slides, pp 43)

Similar to MSAC: Minimize the median of the residuals. The threshold is determined automatically as a value proportional to the median of the residuals produced for each subset. Typically produces better results than RANSAC/MSAC when the inliers contain some noise because in such cases RANSAC or MSAC might have thresholds set too large.

PB 17 (0.5 point): Segmentation using region merging: Give two examples of similarity measures between two regions.

Solution

1. Approximation of Mean Squared Error:
2. $C_{color}(R_1,R_2) = N_{R1}\left\|M_{R1} - M_{R1\cup R2}\right\|_2^2 + N_{R2}\left\|M_{R2} - M_{R1\cup R2}\right\|_2^2$
   Contour length variation:

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

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