Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module:	**M1. Introduction to human and computer vision**	**Final exam**
Date:	November 29th, 2021	**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. Explain Color constancy. Explain the Von Kries law.

   Answer: Color constancy is our ability to perceive objects as having a constant color despite changes in the color of the illuminant. In 1905, Johannes von Kries formulated an explanation for color constancy that is known as von Kries' coefficient law and which is still used to this day in digital cameras to perform white balance. This law states that the neural response of each type of cone is attenuated by a gain factor that depends on the ambient light [73]. In practice, von Kries' law is applied by dividing each element of the tristimulus value by a constant depending on the scene conditions but not on the stimulus: typically, each element is divided by the corresponding element of the tristimulus value of the scene illuminant. Regardless of the original chromaticity of the illuminant, after applying the von Kries' rule the chromaticity coordinates of the illuminant become $x = y = z = 1/3$, which correspond to achromatic, white light. In other words, von Kries' coefficient law is a very simple way to modify the chromaticity coordinates so that, in many situations, they correspond more closely to the perception of color.

2. Explain the Color formation equation. Which two corollaries follow from this equation?

   Answer: The color formation teaches us that color depends on three physical properties: The incident illuminant, the surface reflectance, and the photoreceptors response. The pointwise multiplications of the surface reflectance and the incident illuminant is called the color signal -this colour signal is the information reaching our eyes or the camera-. This color signal is then point-wise multiplied by the cones, and integrated over the visual spectra. There are two corollaries from this equation:
   i) The metamerism problem: Two different reflectances under a particular illuminant can have exactly the same tristimulus values.
   ii) Thricromacy property: we can generate any color by mixing three given colors –called primaries-, just by adjusting the amount of each one:

3. Explain gamma correction. How has been gamma correction modified for HDR images?

   Answer: Originally the CRT monitors had a non-linear relation between the voltage applied and the intensity shown. Also, our perception of intensity is not linear (Weber's law). Originally for both of these reasons –compensating the display non-linearity and improving our perception of it-, and currently just for the second reason, the camera applies a non-linear transformation to the output image in the form of a power-law (gamma-correction).
   For High-Dynamic range images, gamma-correction has shown not to be good enough –in HDR images it may wash-out important details on the bright areas, and create artifacts in them-. For this reason, in HDR images a logarithmic correction is used.

4. Explain unsharp masking.

   Answer: Unsharp masking is a basic edge enhancement method, i.e. it aims at enhancing the edges present in the image to give a "sharper" look. Unsharp masking is linear, and it works in the following way.

From an image I, we compute an edge map E, by the difference of the original image and the result of convolving the original image with a gaussian E = I - g * I. Then, this edge map is added to the original image with a scaling factor k I'=I+K·E

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

1: Consider the following image which is quantized with 3 bits.

```
1  4  4  4
1  4  4  5
1  5  4  5
4  5  4  1
```

Compute the image after histogram equalization.

Solution: The image after histogram equalization is:

```
2 5 5 5
2 5 5 7
2 7 5 7
5 7 5 2
```

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

$$\begin{matrix} -\infty & -\infty & -\infty & -\infty & -\infty \\ -\infty & -\infty & 0 & -\infty & -\infty \\ -\infty & 0 & \underline{-\infty} & 0 & -\infty \\ -\infty & -\infty & 0 & -\infty & -\infty \\ -\infty & -\infty & -\infty & -\infty & -\infty \end{matrix}$$

Consider the operator, $\boldsymbol{\psi}$, that consists in dilating twice the image with the structuring element SE: $\boldsymbol{\psi}(.) = \boldsymbol{\delta_{SE}}(\boldsymbol{\delta_{SE}}(.))$.
Is this operator ψ increasing, idempotent and extensive? (Precisely justify your answers)

Solution: The operator is equivalent to a dilation with the following structuring element SE2:

$$\begin{matrix} -\infty & -\infty & 0 & -\infty & -\infty \\ -\infty & 0 & -\infty & 0 & -\infty \\ 0 & -\infty & \underline{0} & -\infty & 0 \\ -\infty & 0 & -\infty & 0 & -\infty \\ -\infty & -\infty & 0 & -\infty & -\infty \end{matrix}$$

The operator is therefore increasing (as all dilation), non-idempotent (dilating once or twice is different as shown by the difference between structuring elements SE and SE2. So, if we iterate the results will change) and extensive (as the space origin belongs to SE2).

3: We construct a family of structuring element $\{SE_k\}$ based on SE defined in the previous question, with:

$$SE_k(m,n) = \begin{cases} 0 \ \ if \ \ m = \pm k, n = 0 \\ 0 \ \ if \ \ n = \pm k, m = 0 \\ \ \ -\infty \ \ otherwise \end{cases}$$

Can we compute a granulometry with the openings based on these structuring elements: $\boldsymbol{\gamma_{SE_k}}(.)$?
If yes, define the block diagram of the granulometry.

Solution : We cannot define a granulometry because the structuring element is not convex. Therefore increasing the size of the structuring element does not increase the strength of the opening or closing.

4: Consider the following operators: $\boldsymbol{\psi_1}(.) = \boldsymbol{\varphi}(\boldsymbol{\gamma}(.))$ and $\boldsymbol{\psi_2}(.) = \boldsymbol{\gamma}(\boldsymbol{\varphi}(.))$ where $\boldsymbol{\gamma}$ is an opening and $\boldsymbol{\varphi}$ a closing with a square structuring element of size 3x3.
Is the operator $\boldsymbol{\psi_2}(\boldsymbol{\psi_1}(.))$ increasing, idempotent and anti-extensive or extensive?

Solution: $\psi_2(\psi_1(.)) = \gamma\varphi\varphi\gamma(.) = \gamma\varphi\gamma(.)$ because the closing ($\boldsymbol{\varphi}$) is idempotent. We know that the combination $\gamma\varphi\gamma(.)$ is a **morphological filter** that is: increasing and idempotent. It is not anti-extensive nor extensive as many counterexamples can be found, for example a white image with one black point. The resulting filtered image will be all white. The filter is not extensive either as a black image with a white point will be transformed into a black image.

Problem III	**Javier Ruiz**	**(3 points)**

1. Consider the following 5x5 image $x[m,n] = \begin{bmatrix} \underline{0} & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ filtered by the Laplacian filter $h[m,n] = \begin{bmatrix} 0 & 1 & 0 \\ 1 & \underline{-4} & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Answer the following questions:

a) Compute the resulting image $y[m,n]$ of filtering $x[m,n]$ with the Laplacian filter.

b) Is the Laplacian filter robust against noise? How can you increase the robustness when filtering?
c) How can you detect contours on the resulting image $y[m,n]$?

2. Consider the following 1D filter defined by the impulse response $h[m] = [1, \underline{0}, 1]$
   d) Compute the frequency response, $H(F)$, (note: leave the expression in trigonometric form)
   e) Represent the modulus, $|H(F)|$, of the frequency response.

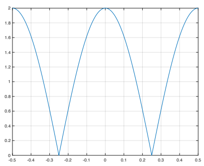
Now we use the same filter to also process the rows of an image, therefore the 2D filter is defined as $h[m] * h[n]$
   f) Calculate the values of the 2D impulse response of the filter as a 3x3 matrix.
   g) How would you categorize this 2D filter (low-pass, high-pass, band-stop, band-pass)?

3. We would like to compute the discrete wavelet transform of a 2D image using the Haar wavelet: $h_0[n] = \frac{1}{\sqrt{2}}\{\underline{1}, 1\}$, $h_1[n] = \frac{1}{\sqrt{2}}\{-\underline{1}, 1\}$, $g_0[n] = \frac{1}{\sqrt{2}}\{\underline{1}, 1\}$ and $g_1[n] = \frac{1}{\sqrt{2}}\{1, -1\}$.
   a) Draw the filter bank decomposition scheme that allows computing one level (scale) of the 2D discrete wavelet decomposition of images.
   b) Is the Haar wavelet an orthogonal wavelet? Does the Haar wavelet have linear phase? (justify)
   c) If the original image has NxN pixels, where can we find the diagonal wavelet coefficients of scale 1 in the decomposition? Precisely define the area corners.

SOLUTIONS
QUESTION 1:

a)
```
y =  0   0   0   0   0   0   0
     0   0   1   1   1   0   0
     0   1  -2  -1  -2   1   0
     0   1  -1   0  -1   1   0
     0   1  -2  -1  -2   1   0
     0   0   1   1   1   0   0
     0   0   0   0   0   0   0
```
b) No. You can low pass filter the image before filtering.
c) Detecting zero-crossings.

QUESTION 2:
a) $H(F) = 2\cos(2\pi F)$

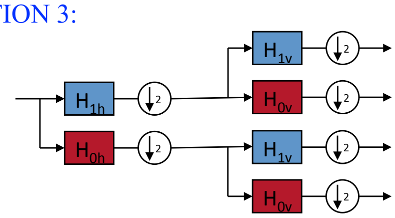


b)

c) $h[m,n] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & \underline{0} & 0 \\ 1 & 0 & 1 \end{bmatrix}$

d) The filter can be considered as a band-stop filter as it preserves both low frequencies (around Fx=Fy=0) and high frequencies (around Fx=Fy=1/2) but removes middle ones (around Fx=Fy=1/4)

QUESTION 3:



a)
b) Yes, it is orthogonal as $h_0[n] = g_0[-n]$ and $h_1[n] = g_1[-n]$ (without taking into account delays). Yes, it is has linear phase as impulse responses have symmetry.
c) They will be at the lower-right corner so (N/2,N/2), (N/2,N-1), (N-1,N/2),(N-1,N-1)

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

1. What is the difference between the derivative of a Gaussian filter and the difference of Gaussians filters? Mention one application of each.

Answer
The derivative of a Gaussian is used to compute the derivative of the image while at the same time smoothing to remove noise, hence it can be used to detect edges.
The difference of a Gaussian, on the other hand, is used to approximate the Laplacian of the image, hence it can be used to detect blobs. For example in the SIFT detector.

2. How do you make a patch descriptor rotationally invariant? How is it done in SIFT?

Answer
One can make descriptors rotationally invariant by assigning orientations to the key points and then rotating the patch to a canonical orientation. In SIFT this is done by constructing Histograms of Gradients in a neighborhood around the feature point, and assigning the largest bin as the corresponding orientation of the keypoint. Later, all detected features are rotated so that the corresponding orientations are vertically aligned. There may be more orientations assigned if there are peaks above 80% of the highest peak. In this case, additional keypoints are created having the same spatial location and scale as the original keypoint for each additional orientation.

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

1. After detecting the contours in an image, to find straight lines, the Least Square method proceeds by minimizing the energy functional $E = ||\mathbf{Ap}-\mathbf{y}||^2$.
    1. Explain how are **A** and **y** formed.
    2. Solve the minimization problem to obtain **p**.

2. For the RANSAC method, the number of iterations can be computed using:

$$k = \frac{\log(1-p)}{\log(1-w^n)}$$

a. Explain the different terms (k, p, w, n) in the equation.
b. Explain how to compute the number of iterations of the above formula if the percentage of outliers is not know *a priori*. Use pseudocode for the explanation.

3. In region growing, describe the most common similarity criterion (the one explained in class)

4. Explain the concept of Region Adjacency Graph (RAG) used in region merging. For the given gray level image and the corresponding partition, construct the RAG using the Approximation of Mean Squared Error as similarity criterion

3	2	1	5
11	13	12	6
19	20	21	7

Gray level image

1	1	1	2
3	3	3	2
4	4	4	2

Partition

ANSWERS

1. Least Squares
    1. **A** is a two columns matrix. The first column is created using the x coordinates of the contour points. The second column is all ones. **y** is a vector created using the y coordinates of the contour points.
    2. See the slides MCV_M1_L09_-_Grouping_segmentation_Classification_I, page 8

- **Minimization:** $\mathbf{p} = \underset{p \in \Re^2}{\operatorname{argmin}} \{E\}$ $\quad \mathbf{p} = [a\ b]$

Energy

$$E = \sum_{i=0}^{N} (y_i - ax_i - b)^2$$

coordinates of contour points

In matrix form:

$$E = \sum_{i=0}^{N} \left([x_i\ 1] \begin{bmatrix} -a \\ -b \end{bmatrix} + y_i \right)^2 = \left\| \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} -a \\ -b \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \right\|^2 = \|\mathbf{A}p - \mathbf{y}\|^2$$

$$= \mathbf{y}^T\mathbf{y} - 2(\mathbf{Ap})^T\mathbf{y} + (\mathbf{Ap})^T(\mathbf{Ap})$$

Minimization:

$$\frac{dE}{d\mathbf{p}} = -2\mathbf{A}^T\mathbf{y} + 2\mathbf{A}^T\mathbf{A}\mathbf{p} = 0 \qquad \mathbf{p} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} = \mathbf{A}^{-1}\mathbf{y}$$

Python:
p = np.linalg.pinv(A)*y

Source: S. Lazebnik, J. Hays

2. RANSAC

a. Number of model parameters (*n*), number of iterations (k), percentage of inliers (w), probability that RANSAC in some iteration selects only inliers (p)

b. If we do not know the percentage of inliers: compute an estimate at each iteration:

```
if ninliers > bestscore     % Largest set of inliers so far...
    bestscore = ninliers;   % Record data for this model

    % Update estimate of N, the number of trials to ensure we pic
    % with probability p, a data set with no outliers.
    fracinliers =  ninliers/npts;
    k = log(1 - p)/log(1 -  fracinliers^n);
end
```

3. The similarity has two terms: one describing how similar the pixel is to the safe region and the other that expresses the variation of the contour length and that tries to impose smooth contours.

For each potential merging between p_0 and R_i, compute:

$\Delta C = \alpha\,(p_0 - M_{R_i})^2 + (1-\alpha)\,\Delta contour$

Pixel gray level | Region mean | Variation of the contour length

4. In a RAG, the nodes are the regions and the edges express the neighborhood relationship. There are edges between neighbor regions, and they are labeled with the similarity between the regions.