Faculty of Computer & Information Sciences
Ain Shams University
Course Code: SCO 421
Course Name: Computer Vision
Exam: (Final) 22/1/2023
Year: (1st term) 3rd & 4th Levels (SC-CSys)

Instructors: Dr. Dina Khattab
Offering Dept: Scientific Computing
Academic year: 2022-2023
Duration: 2 Hours
Total Grade: 50 Marks

Question (1) — Marks: 10

State whether each of the following statements is <u>True</u> of <u>False</u>.

1. When filtering an image with a Gaussian filter, the filter size is the most important because it defines the smoothing scale.
2. Averaging is considered an intuitive solution for image smoothing as it assumes the independence of the noise added to the image.
3. If you are unsure of how many clusters you have in your data, the best method to do clustering would be K-means.
4. Linear filtering complexity is dependent on both image size and filter size.
5. Sobel filters are typically used as edge detectors rather than sharpening tools.
6. Canny edge detector is based on the 1st derivative of a Gaussian.
7. Given the following filter, both convolution and correlation operations will be similar.
$$\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$$
8. The Median filter will perform better in removing noise from the following image.

9. Good feature descriptors should be invariant to both photometric and geometric transformations.
10. It is important to convolve an image with a Gaussian filter before convolving with a Laplacian filter.
11. For finding horizontal edges in an image, we use the following filter $\begin{matrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{matrix}$
12. RANSAC is used to get rid of outliers in situations when there are wrong point correspondences.
13. In canny edge detector, reducing both the high and low thresholds of the threshold hysteresis step, will get more discontinuous edges.
14. The basic building block for object detection in the Mask R-CNN architecture is the Fast R-CNN.
15. The main parameter of the Mean-shift algorithm is the initial window position.
16. The number of visual words is the main parameter for classifying images using BOW.
17. The task of image recognition is considered a sub-task of image verification.
18. Classifying an object incorrectly is considered a false negative error in object detection.

19. The YOLO model for object detection is more likely to produce multiple object responses in the image.

20. The U-Net architecture for semantic segmentation applies both convolution and up-convolutions processes in the up-sampling sub-network.

Question (2) **Marks: 20**

A) **[2 Marks] Design** a 3x3 filter that by being convolved with an image ***f(x,y)*** with gray levels ranging from 5 through 20 will change it to an image ***g(x,y)*** with gray levels ranging from 10 through 40. Suppose that after receiving the image ***g(x,y)***, you have subtracted from it the result of convolving ***f(x,y)*** with a uniform averaging filter, to receive a final image ***h(x,y)***. **What** operation would ***h(x,y)*** represent relative to the image ***f(x,y)***?

B) **[7 Marks]** The following very useful operator $\boldsymbol{\nabla^2 h_\sigma(x;y)}$ is often applied to an image ***I(x; y)*** in computer vision algorithms, to generate a related image ***g(x; y),*** where $\boldsymbol{h_\sigma(x;y)}$ is the Gaussian operator with parameter **σ**, and $\boldsymbol{\nabla^2} = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$.

- **i.** **Give** the general name for this mathematical operator, and the purpose that it is used for in computer vision.
- **ii.** **What** image properties should correspond to the zeroes of the equation, i.e. those points **(x; y)** in the image ***I(x; y)*** where the above result ***g(x; y)*** = 0?
- **iii.** What is the significance of the parameter **σ**? If you increased its value, would there be more or fewer points ***(x; y)*** at which ***g(x; y)*** = 0? And **why**?
- **iv.** **Explain** which of the following two operations is preferable to apply and **why**? where * signifies 2D convolution operation:

a. $\nabla^2[h_\sigma(x;y) * I(x;y)]$ **b.** $\nabla^2 h_\sigma(x;y) * I(x;y)$

C) **[7 Marks]** Given the following pairs of images, we need to apply image matching between images of each pair. The first step is to apply feature detection and use these features for the matching process.

- **i.** **Which** pair do you think that Harris detector will be appropriate for this mission? **Explain** why.

Pair 1

Pair 2

- **ii.** The Harris algorithm works by computing the derivatives at each image pixel, then compute the second Moment matrix M in a window around the pixel. Consider the following derivative matrices Ix and Iy computed for some image pixel in a 3x3 window W, **Compute** the Harris M matrix using the equation:

$$M = \sum_{x,y \in W} \begin{bmatrix} Ix(x,y)^2 & Ix(x,y) * Iy(x,y) \\ Ix(x,y) * Iy(x,y) & Iy(x,y)^2 \end{bmatrix}$$

$Ix =$

1	2	2
3	2	1
2	1	1

$Iy =$

4	6	8
8	6	7
6	6	4

Do you think that this pixel corresponds to a corner? **Explain** why.

D) **[4 Marks]** Consider the following 4 matrices are the output of applying the Difference of Gaussians (DoG) on images of the scale space in the SIFT algorithm. There is a single SIFT keypoint that exists in this scale. Your job is to find it. Apply the Keypoint localization step on these matrices to detect the keypoint.

i. **Report** its location in terms of x, y and its scale (the one it exists in).

ii. **Explain** why is this a SIFT keypoint.

0	2	0	1
0	-2	0	1
1	2	0	1
2	2	1	1

Scale =1

1	2	0	2
0	-2	0	1
1	1	0	2
1	1	1	1

Scale=2

2	3	0	2
1	-3	1	1
2	0	1	1
1	1	2	1

Scale=3

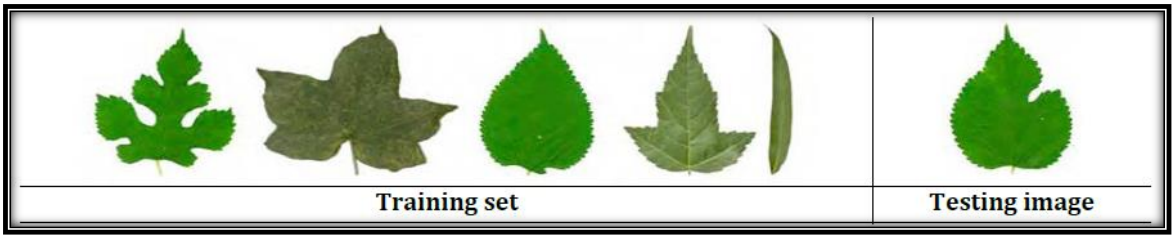
2	5	0	4
4	-2	8	6
0	8	-2	3
-1	1	2	2

Scale =4

Difference of Gaussian (DoG) images

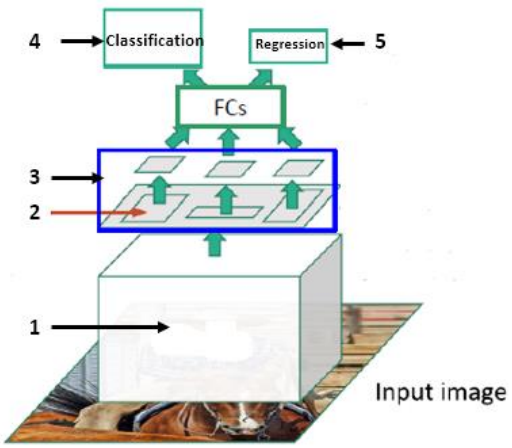
Question (3) — Marks: 20

A) **[6 Marks] Design** a complete algorithm using **handcrafted** features to automatically determine which of the 5 training images is most similar to the testing image. Your algorithm should generate a numerical similarity score, or a distance score, between a pair of leaf images, `[e.g., similarity (training image #1, testing image) = 0.8, … etc.]`



B) **[8 Marks]** Given the following image of the Fast R-CNN region-based object detector, **answer** the following questions guided by the legends on the figure:

1. **What** is the name of this block? **What** is its function?
2. **Mention** the process that is used to generate these region proposals.
3. **What** is the title of this operation? **What** is it used for?
4. **Explain** the output of this classification process. **Which** classifier is used for it?
5. **Explain** the output of this regression process.



C) **[6 Marks] Draw** a general architecture of the Fully Convolutional Network (FCN) for semantic segmentation, showing the input, core and output components. **Explain** the importance of using skip connections.

With Best Regards,

Dr. Dina Khattab