FUNDAMENTALS OF COMPUTER VISION – 2nd PARTIAL EXAM JANUARY 20TH 2023 SOLVED

Name and Surname:

NIA

1.- What is a feature? Which types of features do we have? Can you put examples of each of the types? (1 point)

Piece of information that is useful to solve a given task

Interesting part of the image

Types of features:

Global: global properties of the whole image

Mean grey level, mean colour, main colours, histogram

Local: properties of a part of the image with their own entity

2.- What is a corner? How do we detect them? Please use the term "eigenvalues" in your explanation. (1 point)

Change in property (intensity) in both horizontal and vertical directions

Harris detector: search for these changes in a neighborhood around each pixel. Definition of the structural tensor with derivatives in x and y axis.

Analysis of eigenvalues→ corner: lambda_1 and lambda_2 high and similar in order

3.- Which is the main difference between Prewitt, Sobel and Roberts edge detectors? How can we reduce the high number of edges provided by these methods by default? (1 point)

Differences are related to the convolution kernel.

$$Prewitt(im) = \left(im * \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, im * \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \right)$$

$$Sobel(im) = \left(im * \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, im * \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \right)$$

$$Roberts(im) = \left(im * \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, im * \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right)$$

Reduction of number of edges:

- Non maximum suppression (after)
- Smoothing (before)
- Parameter tuning
- Morphological gradient (before)

4.- What is a blob? How can we detect them (think in a general way) (1 point)

- Regions in a digital image that differ in properties, such as brightness or color, compared to surrounding regions.
- All the points in a blob can be considered in some sense to be similar to each other

Detection: convolution of the image with a filter at multiple scales, look for extrema of filter response in the resulting scale space. Example of filters: Laplacian of Gaussian.

Other examples: Blob Detection in Binary Images, MSER

5.- Shape Descriptors: which are the groups and subgroups they are divided in? Please put an example of each of them (2 points)

Two groups:

Contour-based: we extract features only from the contour of the shape.

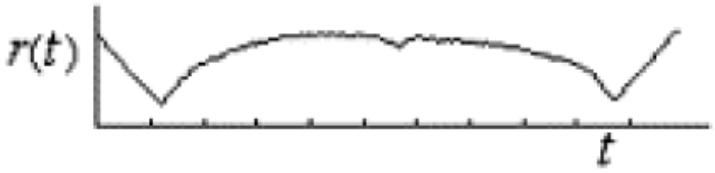
Region-based: we use the whole shape region.

Each group is also subdivided into structural approaches (representing the shape by segments/sections) or global approaches (represent the shape as a whole).

Descriptor	Type	Rotation invariant	Translation invariant	Scale invariant
Shape signature	Contour-based Global	Not direct	After normalization	After normalization
Convex Hull	Contour-based Global	Yes	Yes	Yes
Wavelet Transform	Contour-based Global	No	No	Yes
Minimum Boundary Circle	Contour-based Global	Yes	Yes	Yes
Chain Code	Contour-based Structural	Possible	No	Yes
Polygons Decomposition	Contour-based Structural	No	No	No
Shape Context	Contour-based Structural	Possible	Possible	Possible
Chamfer	Contour-based Structural	No	Yes	Yes

Descriptor	Type	Rotation invariant	Translation invariant	Scale invariant
Zernike Moments	Region-based Global	Yes	No	No
Shape Matrix	Region-based Global	Yes	Yes	Yes
ARP	Region-based Global	No	No	Yes
Grid-based	Region-based Global	Possible	No	No
Skeleton	Region-based Structural	Yes	No	Yes

6.- Shape Signature. One of its implementations creates as output a feature vector in which each component represents the distance of a given boundary pixel and the object center. Is it rotation, translation or scale invariant? In case it is not for any of them please provide a solution on how to fix it. (1.5 points)



Translation invariant: yes by normalization

Rotation invariant: no. How to do it? Try to set a point that appears in all images of the same type and align. Case specific, shift matching

Scale invariant: yes by normalizaton

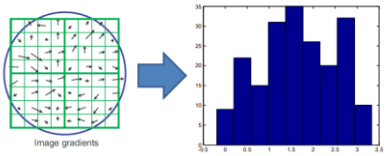
7.- SIFT descriptor. How do we calculate the Histogram of Oriented Gradients? Which is the main difference with respect to ORB? (1 point)

Compute the gradient magnitudes and orientations in a small window around the keypoint at its appropriate scale

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1))/(L(x+1, y) - L(x-1, y)))$$



Histogram of gradient orientation – the bin-counts are weighted by gradient magnitudes and a Gaussian weighting function. Usually, 36 bins are chosen for the orientation.

Main difference:

- SIFT is patented
- Circular regions, 17 spatial bins with 16 orientation bins (272 dimensions)

8.- Practical problem (1 point) How do we create the vocabulary words in Bag of Words method? Which parameter determines the final number of words? What are the problems associated to having very small o very large dictionaries?

Vocabulary creation:

- Clusterization of the feature descriptors extracted from training database images

Parameter:

- Number of clusters

Very small:

- Words do not represent cues from specific categories

Very large:

- Words are too specific, difficult to generalize

9.- Practical problem (1.5 points) Polyp detection in colonoscopy. Given this input image, what would you do to be able to generate the best descriptor possible for the polyp region? You can combine multiple feature extractors and descriptors if you want.

Free answer