PAPER C315

COMPUTER VISION

Friday 20 March 2020, 10:00
Duration: 120 minutes
Post-processing time: 30 minutes
Answer THREE questions

Paper contains 4 questions

1 Image filtering

a Given a 5x5 image as shown below, perform Prewitt filtering and calculate a 3x3 output image (without considering the boundaries).

1	1	1	1	1
1	2	2	2	1
1	3	3	3	1
1	4	4	4	1
1	5	5	5	1

Image

$* h =$

Filter Output

i) Write down the 3x3 horizontal Prewitt filter h_x and vertical Prewitt filter h_y.

ii) Perform convolution between the image and the filters. Write down the output.

iii) Explain how the Prewitt filter h_x can be separated as two filters.

iv) In certain computer vision tasks (e.g. Harris corner detection), to calculate the image gradient, Gaussian filtering is applied prior to Prewitt filtering. Explain the motivation for performing Gaussian filtering.

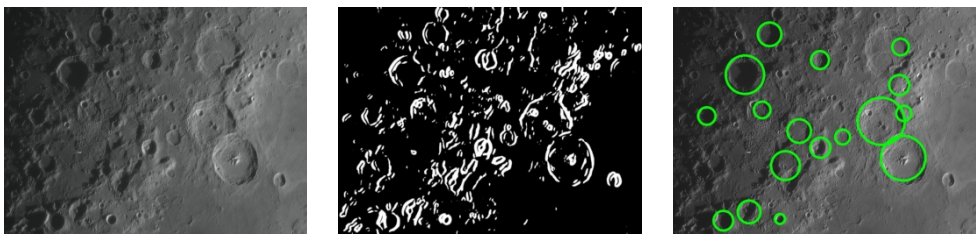
b The 2D Gaussian filter kernel is described by the following equation,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

i) To implement the kernel, how would you design the kernel size K?

ii) Suppose the Gaussian kernel size is $K \times K$ and the input image size is $N \times N$, evaluate the computational complexity using the big O notation for two implementations respectively: direct 2D Gaussian filtering and separable filtering.

c The image below was taken from the moon, where Apollo 16 landed. A binary edge map has been calculated after Prewitt filtering. We plan to detect the moon craters from the edge map using Hough transform.



(a) Moon image (b) Edge map (c) Detected craters

i) Explain the basic idea of Hough transform.

ii) We assume that the moon craters have circular shapes. To represent a circle, we need three parameters a, b, r, where (a, b) denotes its centre and r denotes its radius. The points on a circle is described with the following parametric model,

$$\begin{cases} x = a + r\cos\theta \\ y = b + r\sin\theta \end{cases}$$

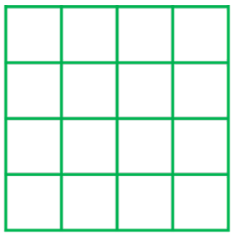
where (x, y) denotes a point on the circle and θ denotes the orientation of the point with regard to the centre. Let H be an accumulator, with 3 dimensions that each contains the bins for a, b and r. Describe the procedure for circle detection using Hough transform.

The three parts carry, respectively, 50%, 20%, 30% of the marks.

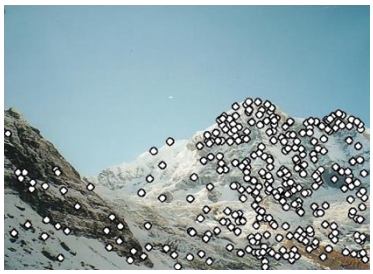
2 Feature detection and description

a Scale-invariant feature transform (SIFT) is an algorithm for detecting interest points in images and describing their features. Describe the procedure of interest point detection in SIFT.

b After the interest points are detected, SIFT describes the feature for each interest point. Suppose we use 4x4 subregions centred at the interest point (shown on the right) and within each subregion there are 4x4 pixels. The SIFT feature is a 128-dimensional vector. Explain what this feature vector stands for.



c The interest points detected by SIFT can be used for image matching. For the following two images, suppose SIFT detects M interest points in image A and N interest points in image B.



(a) Image A

(b) Image B

i) For each of the M interest points in image A, explain how we can find its matching point in image B.

ii) Evaluate the computational complexity of the point matching procedure using the big O notation. Could you suggest anyway to accelerate the matching?

d SIFT uses the difference of Gaussians (DoG) for interest point detection, which approximates the scale-normalised LoG. The Gaussian filter is formulated as,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

i) Prove the following equation,

$$\frac{\delta G}{\delta \sigma} = \sigma \nabla^2 G$$

where ∇^2 is the Laplacian operator.

ii) Let us define the difference of Gaussians as $DoG(x, y, \sigma) = G(k\sigma) - G(\sigma)$. Given that we can approximate the derivative $\frac{\delta G}{\delta \sigma}$ using the finite difference,

$$\frac{\delta G}{\delta \sigma} \approx \frac{G(k\sigma) - G(\sigma)}{k\sigma - \sigma}$$

prove that $DoG(x, y, \sigma)$ approximates the scale-normalised LoG, i.e. $\sigma^2 \nabla^2 G$.

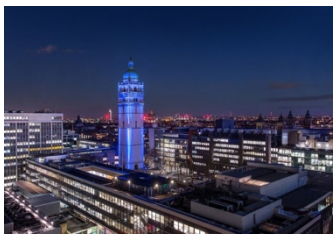
The four parts carry, respectively, 20%, 15%, 25%, 40% of the marks.

3 Image classification and object detection

a Given a dataset that consists of images of the Queen's Tower and some other buildings, we plan to build an object detection method that can tell whether a new image contains the Queen's Tower or not. Some of the sample images are shown below. We plan to combine histogram of oriented gradients (HOG) features and a linear SVM model to perform the task.


(a) Queen's Tower


(b) Queen's Tower


(c) Queen's Tower


(d) Big Ben


(e) The Shard


(f) The Gerkin

The first step is pre-process the images so that it will be easier to extract features and to train the model. What pre-processing steps would you perform?

b Explain the motivation of using histogram of oriented gradients (HOG) as features for image classification, instead of using simply pixel intensities. Is it rotation-invariant?

c What does SVM stand for? Briefly describe how a linear SVM model performs classification.

d Suppose the SVM is trained using the HOG features. At test time, a test image may contain the Queen's Tower but the size of the tower in terms of pixels is unknown. How would you apply the SVM to detect the Queen's Tower in a scale-robust way?

e On a test image, a number of overlapping regions are detected for the same object. Explain how to handle the overlapping detection results.

f Finally, we need to evaluate the detector in terms of classification accuracy (whether an image contains Queens' Tower or not) and localisation accuracy (how accurate is the detected bounding box). Describe how these two metrics can be evaluated.

The six parts carry, respectively, 10%, 15%, 30%, 15%, 10%, 20% of the marks.

4 Motion estimation

a The optic flow method is a technique to estimate motion between two images, for example, two time frames in a video (shown below).

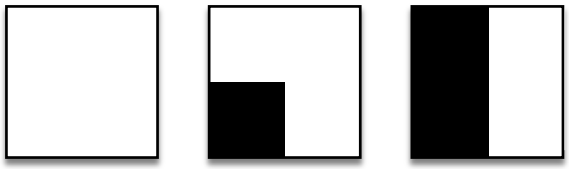
(a) time frame t

(b) time frame $t + 1$

Explain the basic assumptions in the optic flow method and derive the optic flow constraint equation.

b In the optic flow constraint equation, (u, v) denotes the unknown displacement vector at a pixel. This means that there are two unknowns to solve in a single equation. It is an under-determined system. How does the Lucas-Kanade method convert this under-determined system into an over-determined system?

c Optic flow can essentially track the motion for all the pixels and regions in an image. However, some regions may be more reliably tracked than other regions. In the following examples, which region can be more reliably tracked? Explain why.



(a) Flat region. (b) Corner. (c) Edge.

d Smartphones and handheld cameras are now prevalent, which enable us to capture special moments in our life with more ease. However, such casual videos may be influenced by shaky motion due to the instability of the camera or our hands. Camera shake causes visible frame-to-frame jitter in the recorded videos. Could you design a video post-processing method based on optic flow that can reduce the motion artefacts and stablise the videos please? Explain your idea.

The four parts carry, respectively, 30%, 20%, 25%, 25% of the marks.