IMPERIAL COLLEGE LONDON

TIMED REMOTE ASSESSMENTS 2020-2021

BEng Honours Degree in Computing Part III
BEng Honours Degree in Electronic and Information Engineering Part III
MEng Honours Degree in Electronic and Information Engineering Part III
MEng Honours Degree in Electronic and Information Engineering Part IV
BEng Honours Degree in Mathematics and Computer Science Part III
MEng Honours Degree in Mathematics and Computer Science Part III
MEng Honours Degrees in Computing Part III
MSc Computing
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant assessments for the Associateship of the City and Guilds of London Institute

PAPER COMP60006=COMP96046=COMP96047

COMPUTER VISION (TERM2)

Friday 26 March 2021, 10:00
Duration: 140 minutes
Includes 20 minutes for access and submission

Answer ALL FOUR questions
Open book assessment

Paper contains 4 questions

1 Image filtering

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

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

Image

$* h =$

Filter

Output

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

ii) Perform convolution between the image and the filters. Briefly explain the procedure and then write down the output.

b Gaussian filtering is commonly used in computer vision. The 2D Gaussian filter kernel is described by the following equation,

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

i) Gaussian filtering is combined with the Harris detector for interest point detection. Describe the motivation for introducing Gaussian filtering here.

ii) Show that the 2D Gaussian filter is a separate filter.

iii) Suppose the 2D Gaussian kernel size is $K \times K$, the input image size is $N \times N$ and Gaussian filtering is performed. Each multiplication is defined as one operation. Each addition is also defined as one operation, the same as multiplication. If we treat boundary pixels in the same way as inner pixels, how many operations do we need for direct 2D Gaussian filtering and separable filtering respectively?

c Median filtering is often used for image denoising. Suppose a uniform region of an image is corrupted by salt and pepper noise (salt: random white pixels; pepper: random black pixels). Show that provided less than half of the pixels in a neighbourhood are corrupted, a 3x3 median filter will almost perfectly restore the uniform region.

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

2 Feature detection and description

a Interest point detection

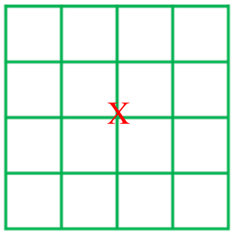
i) Let $I[x, y]$ denotes an 2D image, $G[x, y]$ denotes a 2D Gaussian filter and $*$ denotes convolution. Show that

$$\frac{\partial^2(I * G)}{\partial x^2} + \frac{\partial^2(I * G)}{\partial y^2} = I * (\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2})$$

ii) Explain what the difference of Gaussian (DoG) filter is and why it is used for interest point detection.

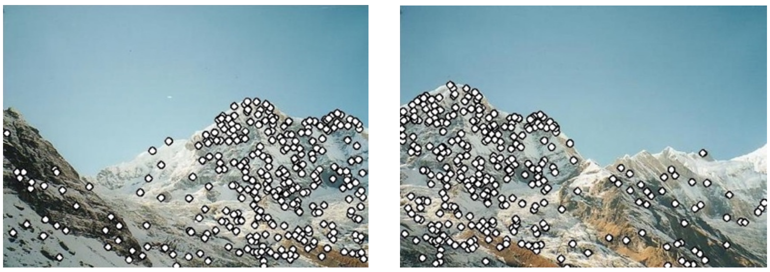
b Feature description

After interest points are detected, we use SIFT to describe the feature for each interest point. Suppose 4x4 subregions are used, centred at the interest point (shown on the right). Explain how the SIFT descriptor is calculated for this interest point (we already know its location and scale).



c Image matching

i) After feature description, we match interest points between Images A and B (shown below). Suppose we know the two images are related by translation and isotropic scaling only. Let (x, y) denote an interest point in A and (u, v) denote an interest point in B. Write down the equation that maps (x, y) to (u, v). How many pairs of points are needed to uniquely determine the mapping?



(a) Image A (b) Image B

ii) We have obtained a number of point pairs between Images A and B. However, since some of these pairs are outliers, we decide to use RANSAC to fit the transformation model. Suppose 50% of the point pairs are outliers and in RANSAC each time we sample 3 pairs. How many times do we need to sample to assure that with 95% probability, we have at least once that all 3 point pairs are inliers? You can keep the logarithm in the answer.

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

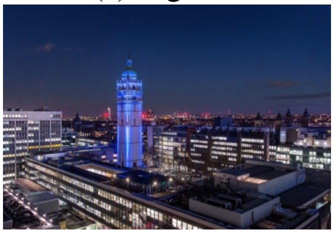
3 Image classification, detection and segmentation

a You are given a dataset that consists of the Big Ben and other landmark buildings in London (some examples shown below).


(a) Big Ben


(b) Big Ben


(c) Big Ben


(d) Queen's Tower


(e) The Shard


(f) The Gerkin

i) Your first task is to develop a classifier to detect the Big Ben in new images. You downsample the images to reduce dimensionality and then train a classifier on these images to differentiate images that contain Big Ben versus those not. However, you find the classifier performs poorly on test images. Explain potential reasons why the classifier is not working well.

ii) Your second task is to develop an image retrieval system using the dataset of London buildings. Given a new image of some building, the system will retrieve similar images. Describe how you would design such a system.

iii) Now the image retrieval system has been developed and you are to evaluate its performance. You provide it with a query image and the system responds with a ranked list of images in the order of decreasing probability. The relevances of the returned images are [+1, +1, -1, +1, +1, -1], where +1 means that the image is indeed relevant to query and -1 means not relevant. Calculate the precision and recall values needed for plotting the precision-recall curve.

b The virtual background function is commonly used for video conferencing (e.g. Zoom or Microsoft Teams). To this end, we need to separate the foreground from background for an image and blend it with a new background according to,

$$I_i = \alpha_i \cdot F_i + (1 - \alpha_i) \cdot B_i$$

where I_i denotes the blended image intensity or colour at pixel i, F_i denotes the foreground, B_i denotes the new background and α_i denote the probability of foreground, $0 \leq \alpha \leq 1$, with 1 to be pure foreground and 0 to be pure background.

i) Suppose the user in video conferencing uses a green screen, how would you estimate the α image?

ii) For general cases without a green screen, how would you estimate the α image?

The two parts carry, respectively, 70%, 30% of the marks.

4 Motion estimation and tracking

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 $t + 1$

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

ii) The optic flow constraint equation at a pixel location is an under-determined system. How does the Lucas-Kanade method convert it to an over-determined system to solve the flow field?

iii) Estimation of large displacements is a challenge in optic flow methods. How does the Lucas-Kanade method address this challenge?

iv) If the Lucas-Kanade method is applied to a colour video, how would you estimate the flow field?

b Given a video dataset, you are developing an algorithm for car tracking on the street (figure below). However, a challenge is that some cars may be occluded by other objects in certain time frames and then appear again later. How would you design the algorithm so that it keeps tracking of the same car through occlusions? Explain your idea.

(a) Car (yellow arrow) (b) The same car (c) Occlusion (d) Re-appearance

The yellow arrow points to a car that we track, which is occluded at certain frame.

The two parts carry, respectively, 70%, 30% of the marks.