CITY UNIVERSITY OF HONG KONG

Course code & title : CS5187 Vision and Image

Session : Semester B 2018/19

Time allowed : Two hours

This paper has FOUR pages (including this cover page).

1. This paper consists of 6 questions.
2. Answer ALL questions.

*This is an **open-note** examination (allowed to have five A4 papers as reference notes)*

Students are allowed to use the following materials/aids:

Approved Calculator

Five A4 papers (10 pages) as reference notes

Materials/aids other than those stated above are not permitted. Candidates will be subject to disciplinary action if any unauthorized materials or aids are found on them.

Question 1: Image Processing [15%]

(a) Name two applications of non-maximum suppression (NMS) in image processing. Describe how non-maximum suppression works in one of the applications. [5%]

(b) Give two advantages of performing pooling operation in CNN (convolutional neural network)? If a convolutional filter has size 3×3 and stride 1, and pooling layer has pool size of 4×4 , what is the size of receptive field for a pixel after pooling operation? [5%]

(c) Give an example of second moment matrix that has rank 1. What kind of image patches can produce such matrix? [5%]

Question 2: Feature Extraction [15%]

(a) What is the difference between the 1st derivative of Gaussian filter and Laplacian of Gaussian (LoG) filter? [4%]

(b) Explain how local interest point operators such as SIFT can achieve invariances in (i) illumination, (ii) scale, (iii) rotation. [6%]

(c) What is "feature map" in CNN (convolutional neural network)? The last convolution layer of AlexNet is a tensor with size $13 \times 13 \times 256$, where 13×13 is image resolution and 256 is number of feature maps. Describe how to convert the tensor into a vector of low dimensional space for applications such as image retrieval. [5%]

Question 3: Image Transformation [20%]

(a) Show that the perspective transformation of an infinity points is not at infinity. [4%]

(b) Given a 3×3 homography H, show that the projective transformation H of a line l is $H^{-T} l$.
[6%]

(c) To compute homography H between two images, a set of 100 corner points between two images are matched using SIFT. It is estimated that up to 40% of matched points are outliers. Calculate what is the minimum number of iterations required by RANSAC to estimate correct answer with 99% probability? Show your steps. [6%]

(d) Comment how possible a mobile phone with one camera can measure the height of a 2D poster being vertically hanged. You can assume that the intrinsic parameters of the camera is known, and the orientation of camera is known when capturing picture. [4%]

Question 4: Epipolar Geometry [15%]

(a) Explain why stereo image rectification is usually performed for the reconstruction of disparity map. [4%]

(b) Show the vector presentation for epipole after image rectification. [2%]

(c) Show the fundamental matrix for a rectified image pair. [2%]

(d) In image rectification, two homographies, H and H', are required for projecting the left and right image planes onto a common plane. Derive an equation to show the relationship between the fundamental matrix before rectification and the two homographies. [7%]

Question 5: Optical Flow [15%]

(a) What is the requirement made by Lucas-Kanade Tracker about the spatial-temporal property for robust estimation of optical flow? [2%]

(b) What is the minimum number of pixels required to estimate an optical flow? Explain your answer. [3%]

(c) If the motion is not translation but affine transformation, show how should brightness constancy equation be modified to estimate optical flow? What is the minimum number of pixels required for estimation? Show your steps for the estimation of affine parameters. Brightness constancy equation for translation is given as: $I_x u + I_y v + I_t \approx 0$.
[10%]

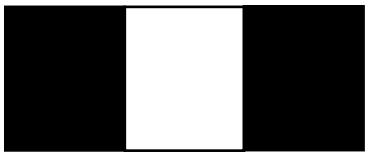
Question 6: Object Recognition [20%]

(a) Show the integral image of the following image patch I . Given a rectangle filter R of the same size as I , calculate the feature value when convolving R with I . Show your steps.
[5 marks]

4×6 image patch I

1	1	1	1	1	1
1	1	2	2	1	1
1	1	2	2	1	1
1	1	1	1	1	1

Rectangle filter R
The size of a black or white region is 4×2 .



(b) In Viola-Jones face detection, a chain of classifiers (or attentional cascade) is used to speed up detection. What is the requirement for each of the classifiers in the cascade? Explain the reason.
[5 marks]

(c) Why faster R-CNN (regional convolutional neural network) is much faster than R-CNN? If an application is to detect objects of both very large size and very small size, which network is likely to give better performance? Explain your reason. [5 marks]

(d) What are the two major steps used by FCN (fully convolutional network) for image upsampling? How does FCN generate image segmentation result with higher accuracy?
[5 marks]

- END -