CITY UNIVERSITY OF HONG KONG

I declare that I do not access to Internet during the exam period and the exam paper is completed by me alone. I sign my student number below as a vow of trust.

Student Number: _____

Course code & title : CS5187 Vision and Image

Session : Semester B 2019/20

Time allowed : Two hours

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

---

1. This paper consists of 14 questions in 2 sections.
2. Answer ALL questions in Section A and Section B.
3. **Give your answers directly on this file.** Use more pages if necessary.
4. You are **NOT allowed** to assess Internet during the exam

---

*This is an **open-book** examination*

Students are allowed to use the following materials/aids:

Approved Calculator

Books, 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.

Section A: Short Question (50%)
Five marks (5%) for each question.

Q1
What are the parameters required to set when generating a pyramid of Gaussian images for a picture? Why pyramid of Gaussian is applied in various image processing algorithms?

Q2
Calculate the **normalized co-occurrence matrix** for the given image patch below using the vector $d=(1,2)$ which specifies the spatial relation between two pixels. What is the property of this image patch? Note: $d=(dr,dc)$ is the relation in row (*dr*) and column (*dr*).

2	0	0	0	0
1	0	2	0	0
1	0	1	0	2
0	0	1	0	1
0	0	0	0	1

Q3
The Harris operator detects corner by computing the two eigenvalues λ_1 and λ_2 of the Harris matrix, where $\lambda_1 > \lambda_2$. A threshold, T , is then set to determine whether a corner is detected. Which of the following criterion is more appropriate to detect corner? Justify your answer.
(i) $\lambda_1 > T$
(ii) $\lambda_1 \lambda_2 > T$

Q4

A set of 100 pixels are matched between the two images captured by a pair of uncalibrated cameras. It is estimated that at least 70% of matchings are correct. To estimate the disparity map for the two images, RANSAC algorithm is applied. What is the minimum number of iterations required by RANSAC to estimate correct answer with 99% probability? Show the steps of calculation.

Q5

A photographer takes 12 pictures using a perspective projection camera mounted on a tripod. The tripod is rotated by 30 degree in angle each time before taking a picture. Can the scene depth be recovered using the pictures taken in 360 degree? Explain your answer.

Q6

Consider a set of parallel lines on the X-Z plane, where Z is the optical axis of a camera. Why these lines will intersect when being projected on the image plane (X-Y plane)? What is the property of the intersection point on the image?

Q7

Economic beehoon is a popular breakfast in South East Asia. There are around 30 different categories of food items (e.g., rice noddle, fried egg, sausage) for choice. Suppose you want to classify, segment and count the food items in the breakfast (see the sample image). Which one of the following neural networks you will use: faster R-CNN, fully convolutional network (FCN), Mask R-CNN? Explain your choice.

Q8

Why structure-from-motion is considered as a chicken-and-egg problem? Suppose the inputs include 20 images and 2000 scene points, how many unknown variables are required to be solved for structure-from-motion?

Q9

Why stereo vision and optical flow estimation are possible to be solved by the same algorithm? Briefly describe how convolutional neural network can be modified for both problems of stereo vision and optical flow estimation.

Q10

What does “receptive field” mean in video processing? Describe how does receptive field is being used in 3D convolutional neural network for activity classification.

Section B: Long Question (50%)

Q11 Projection Matrix [15%]

The followings are the notations for two cameras C_1 and C_2:
I: Identity matrix of size 3×3
0: Zero vector of size 3×1
K_1: Calibration matrix of C_1
K_2: Calibration matrix of C_2
R: Rotation matrix of size 3×3
T: a vector of 3×1

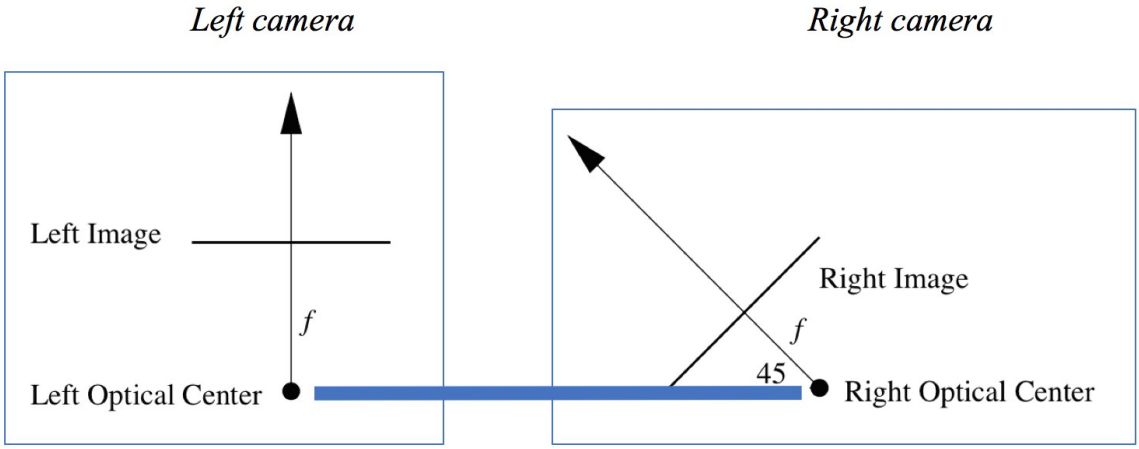
(a) If the optical axes of C_1 and C_2 are parallel, and C_2 is at the position $\begin{bmatrix} T_x & 0 & 0 \end{bmatrix}$ in the coordinate system of C_1.

(i) What are the rotation matrices of both cameras? [1%]
(ii) What is the essential matrix that relates both cameras? [2%]

(b) If the optical axes are not parallel, and the projection matrices of cameras are given as followings:
Projection matrix of C_1: $M_1 = K_1 \begin{bmatrix} I & O \end{bmatrix}$
Projection matrix of C_2: $M_2 = K_2 \begin{bmatrix} R & T \end{bmatrix}$

(i) Where is the camera C_2 located in the world coordinate? [6%]
(ii) What is the essential matrix that relates both camera? [6%]

Q12 Stereo Vision [15%]



(a) Where are the epipoles of the left and right images? [2%]

(b) Draw at least five epipolar lines on each of the left image and right image below. [4%]

Left image

Right image

(c) Describe how to rectify the image pair above. [5%]

(d) Suppose the intrinsic and extrinsic parameters of both cameras are known. Describe how to estimate the scene depth given the rectified image pairs. [4%]

Q13 Image Transformation [10%]

A poster hanged on a building is captured by a camera. Assuming the top-left corner of the poster correspond to the world coordinate. As the poster is a 2D plane, the scene point on the poster is denoted as $(X, Y, 0)$. Its corresponding pixel coordinate in the image is denoted as (u, v) .

(a) What is the geometric transformation between $(X, Y, 0)$ and (u, v) ? Express the answer in terms of matrix representation. [2%]

(b) How many unknown parameters in the transformation matrix? [2%]

(c) To solve the transformation matrix, what is the minimum number of correspondence points required between the poster and image? [2%]

(d) What happen if the transformation matrix is solved by more number of correspondence points than the answer in (c)? [2%]

(e) Given two invariant properties of the transformation. [2%]

Q14 Optical Flow [10%]

Let $I(x, y, t)$ be an image sequence. Assume that there is an affine change in the image intensities between two adjacent images, as following:

$$I(x+u, y+v, t+1) = aI(x, y, t) + b$$

where (u, v) is optical flow, $a(x, y)$ and $b(x, y)$ are photometric parameters that change the pixel intensity at location (x, y). Show the linear system of equations for the estimation of four unknown parameters (u, v, a, b). What is the minimum window size for estimation?

- END -