北京郵電大學

Queen Mary University of London

EBU7240 A

Joint Programme Examinations 2020/21

EBU7240 Computer Vision

Paper A

Time allowed 2 hours

Answer ALL questions

Use Answer book to answer the questions.

NOT allowed: electronic calculators and electronic dictionaries.

INSTRUCTIONS

1. **You MUST use the supplied Answer book to answer the questions.**
2. You must NOT take **question papers or answer books**, used or unused, from the examination room.
3. Write only with a black or blue pen **and in English**.
4. Do all rough work in the answer book – **do not tear out any pages**.
5. If you use Supplementary Answer Books, tie them to the end of this book.
6. Write clearly and legibly.
7. **Read the instructions on the inside cover.**

Examiners

Dr Changjae Oh, Dr Marie-Luce Bourguet

Filename: 2021_EBU7240_A

Instructions

Before the start of the examination

1) Place your BUPT and QM student cards on the corner of your desk so that your picture is visible.

2) Put all bags, coats and other belongings at the back/front of the room. All small items in your pockets, including wallets, mobile phones and other electronic devices must be **placed in your bag in advance**. **Possession of mobile phones, electronic devices and unauthorised materials is an offence.**

3) Please ensure your mobile phone is switched off and that no alarm will sound during the exam. **A mobile phone causing a disruption is also an assessment offence.**

4) Do not turn over your question paper or begin writing until told to do.

During the examination

1) You must not communicate with or copy from another student.

2) If you require any assistance or wish to leave the examination room for any reason, please raise your hand to attract the attention of the invigilator.

3) If you finish the examination early you may leave, but not in the first 30 minutes or the last 10 minutes.

4) For 2 hour examinations you may **not** leave temporarily.

5) For examinations longer than 2 hours you **may** leave temporarily but not in the first 2 hours or the last 30 minutes.

At the end of the examination

1) You must stop writing immediately – **if you continue writing after being told to stop, that is an assessment offence.**

2) Remain in your seat until you are told you may leave.

Question 1

a) This question is about **camera model**.

[7 marks]

i) Given the optical centre, O, at the origin, and the focal length f, and the image plane parallel to xy-plane, **1)** draw the pinhole projection model, including the 3D point $P = (x_1, y_1, z_1)$ and its projected 2D image point $P' = (x'_1, y'_1)$. Also, **2)** represent the coordinate $P' = (x'_1, y'_1)$. in terms of x_1, y_1, and z_1.

(5 marks)

ii) Describe how depth of field is affected if aperture size becomes smaller.

(2 marks)

b) This question is about **image filtering**.

[8 marks]

i) Given a 3 × 3 image, compute the output value of a centre pixel in grey by applying two different filters: **1)** Uniform mean filtering with the 3 × 3 filter kernel, **2)** Median filtering with the 3 × 3 filter kernel. (Show your calculations)

(2 marks)

0	2	0
6	200	3
6	6	2

ii) Explain why the average mean filter is good at removing zero-mean additive white gaussian noise (AWGN) N that has the following probability density function of a Gaussian random variable z.

$$P(z = N) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right)$$

(2 marks)

iii) Explain the bilateral filtering including **1)** its mathematical definition and **2)** the advantages over Gaussian filtering in image denoising.

(4 marks)

c) This question is about **feature detection and matching**.

[10 marks]

i) By using Harris corner detector with 3×3 window of equal weighting, the empirical constant k = 0.05, and differentiation kernel below (*d/dx* and *d/dy*), **1)** find the Harris matrix, and **2)** the corner response for the centre of the following image I_1, and **3)** determine whether the point is flat, edge, or corner.

(3 marks)

$$d/dx = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad d/dy = [0 \quad -1 \quad 1], \quad I_1 = \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 \\ 0 & 0 & 1 & 6 & 6 \\ 6 & 6 & 6 & 6 & 6 \\ 6 & 6 & 6 & 6 & 6 \end{bmatrix}$$

ii) Describe how *key point descriptor construction* works in Scale Invariant Feature Transform.

(4 marks)

iii) We have two sets of features $\{f_i | i = 1, \dots, N\}$, from a reference image I_1, and $\{g_j | j = 1, \dots, M\}$, from a target image I_2. Given a reference feature, f_1, describe how nearest neighbour matching works on f_1.

(3 marks)

Question 2

a) This question is about **fitting**.

[8 marks]

i) The figure below shows the pseudo code for estimating an affine transformation with RANSAC. Fill out the blanks in the code.

(5 marks)

```
Input: A set of N matched points MP = {(p_0,p'_0),(p_1,p'_1) …,(p_{N-1},p'_{N-1})}
Output: Affine transform T_F

S: the number of trials
count_mat: S × 1 vector
IN: a set of inliers

Initialize count_mat to 0
for i = 0 ~ S-1
   Randomly select k matched points from MP (Usually, k=3,4)
   Estimate an affine transformation T_i with [ (1) ]
   for j = 0 ~ N-1
      if [ (2) ]
         count_mat[i]++

Choose the best affine transformation T ← T_K where [ (3) ]

IN = NULL
for j = 0 ~ N-1
   if [ (4) ]
    IN ← IN ∪ {(p_j,p'_j)}

Re-estimate an affine transformation T_F with IN
```

ii) Explain the *Hough transform algorithm* with illustrations.

(3 marks)

b) This question is about **grouping**.

[8 marks]

i) The figure below shows the pseudo code for *K-means algorithm*. Fill out the blanks.

(3 marks)

1. Randomly initialize the cluster centers, $c_1, \ldots, c_K$
2. Given cluster centers, determine points in each cluster
   - For each point p, find the closest c_i. Put (1)
3. Given points in each cluster, solve for c_i
   - Set c_i to be the (2)
4. If c_i have changed, repeat (3)

ii) State three advantages and two drawbacks of *Mean-shift algorithm*.

(5 marks)

c) This question is about **calibration**.

[3 marks]

State **1)** the definition of camera calibration with **2)** the illustration of image, camera, and world coordinates.

(3 marks)

d) This question is about **stereo matching**.

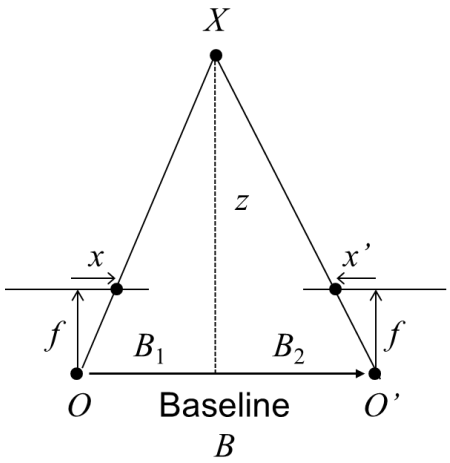
[6 marks]

i) Explain the uniqueness constraint used for improving local stereo matching performance. Provide your description with an illustration.

(3 marks)

ii) In the figure below, derive the relationship between disparity, $x - x'$ and depth z.

(3 marks)



Question 3

a) This question is about **Tracking.**

[4 marks]

Kanade-Lukas-Tomasi (KLT) Tracker solves the objective function below:

$$\min_{\boldsymbol{p}} \sum_{\boldsymbol{x}} \left[I\left(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p})\right) - T(\boldsymbol{x}) \right]^2$$

where $\boldsymbol{p}$ is warp parameters, I is image intensity, $\boldsymbol{W}$ is a warping function, T is template image intensity, and $\boldsymbol{x}$ is pixel coordinate. The figure below shows the procedure of KLT Tracker. Fill out the blanks with the proper notations.

(4 marks)

1. Warp image [$I\left(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p})\right)$]
2. Compute error image [$I\left(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p})\right) - T(\boldsymbol{x})$]
3. Compute gradient [$I(\boldsymbol{x}')$ ($\boldsymbol{x}'$: coordinates of the warped image)]
4. Evaluate Jacobian [ ]
5. Compute Hessian [ ]
6. Compute $\Delta\boldsymbol{p}$ = [ ]
7. Update parameters [ ]

b) This question is about **Recognition**.

[13 marks]

i) Describe *semantic segmentation, instance segmentation* and their difference.

(4 marks)

ii) Describe the training and testing phases in *statistical learning framework* with illustration.

(5 marks)

iii) State the **1)** two advantages and **2)** two drawbacks of *nearest neighbour classifier*.

(4 marks)

c) This question is about **detection**.

[8 marks]

i) In Viola/Jones face detector, integral images are used for fast feature evaluation during testing. Find **1)** the integral image of the figure below and compute **2)** the sum of pixels in the grey region based on the integral image. (Show your calculations)

(4 marks)

5	4	3	8	3
3	1	1	2	6
1	6	0	5	7
1	3	6	5	9
1	2	2	1	3

ii) Viola/Jones face detector employs boosting for feature selection. Describe the process, with illustration, of *boosting round 1*.

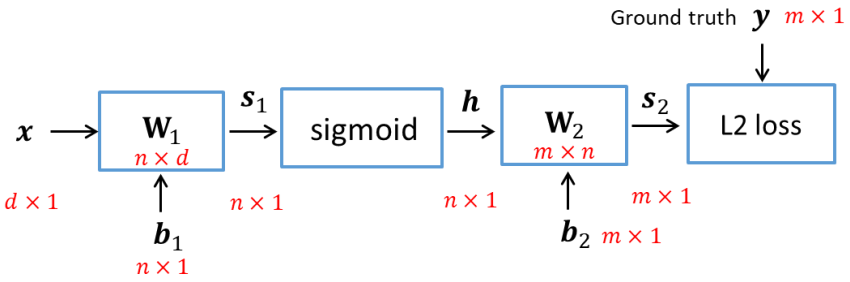
(4 marks)

Question 4

a) This question is about **Backpropagation in Deep Learning**.

[10 marks]

The figure below is 2-layer neural network with batch size 1:



where $\boldsymbol{s}_1 = \mathbf{W}_1\boldsymbol{x} + \boldsymbol{b}_1$, $\boldsymbol{s}_2 = \mathbf{W}_2\boldsymbol{h} + \boldsymbol{b}_2$, and the loss function $L = (\boldsymbol{s}_2 - \boldsymbol{y})^2$.

And the sigmoid function is defined as below:

$$\boldsymbol{h} = \sigma(\boldsymbol{s}_1) = \frac{1}{1 + e^{-\boldsymbol{s}_1}}$$

i) Derive the following derivative, $\frac{\partial L}{\partial \boldsymbol{s}_2}$.

(2 marks)

ii) Derive the following derivatives as a function of $\frac{\partial L}{\partial \boldsymbol{s}_2}$:

$$1)\ \frac{\partial L}{\partial \boldsymbol{b}_2}, \quad 2)\ \frac{\partial L}{\partial \boldsymbol{h}}$$

(4 marks)

iii) In the above 2-layers neural net, the batch size is 1. Re-draw this figure when the batch size is 16. State the dimension of all variables clearly.

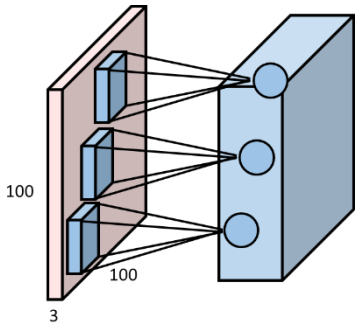
(4 marks)

b) This question is about **Feedforward process in deep learning**.

[7 marks]

The feed-forward process of the convolution is performed as below. Assume that matrix and vector multiplication is used for an efficient implementation on GPU. Note that the figure is drawn with BS = 1.

```
Input: [100x100x3]
Conv filter: 32 filters of [3x3x3]
Stride: 1
Padding: 1
Batch size: BS
```



i) Compute the size of output activation map when $BS = 1$. Use the formula for computing the size of activation map.

(2 marks)

ii) Explain the feed-forward process of the convolutional layer in a form of $\mathrm{O} = \mathrm{WX} + \mathrm{B}$. Assume that the batch size $BS = 1$. Hint) Explain this by considering O as 32×10000 matrix. O should be reshaped into the output activation map.

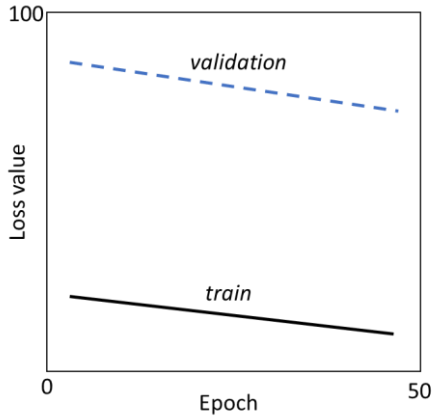
(5 marks)

c) This question is about **Training a neural network.**

[8 marks]

i) Based on the loss curve in the figure below, state the problem and suggest the solution.

(4 marks)



ii) State two advantages and two drawbacks of neural networks.

(4 marks)

END OF PAPER