NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 1 EXAMINATION 2023-2024

EE6222 – MACHINE VISION

November / December 2023 Time Allowed: 3 hours

INSTRUCTIONS

1. This paper contains 4 questions and comprises 4 pages.
2. Answer all 4 questions.
3. All questions carry equal marks.
4. This is a closed-book examination.
5. Unless specifically stated, all symbols have their usual meanings.

---

1. A digital greyscale image has the image histogram $P_{of}(f)$ given by

$$P_{of}(f) = 0.2\delta[f-0.7] + 0.1\delta[f] + 0.3\delta[f-0.8] + c\delta[f-1] + 0.2\delta[f-0.9],$$

where f is the grey level, and c is a constant.

(a) Determine the minimal and the maximal grey levels, f_{min} and f_{max}, of the image. How many different grey levels do the image have?

(6 Marks)

(b) Determine the value of c and plot the histogram $P_{of}(f)$.

(6 Marks)

(c) A gamma correction $g = f^2$ is applied to the image. Compute and plot the histogram of the gamma-corrected image $P_{\gamma g}(g)$.

(6 Marks)

(d) Histogram equalization is applied to correct the image f. Compute and plot the histogram of the histogram-equalized image $P_{heg}(g)$. It is not necessary for you to normalize the output image into some range such as [0, 255].

(7 Marks)

2. Let $\boldsymbol{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$ be the feature vector to be classified, which has known class-conditional Gaussian PDFs for all classes. A minimum weighted distance classifier is described by the following discriminant function of class ω_i:

$$g_i(\boldsymbol{x}) = -\sum_{j=1}^{n} (x_j - \mu_{ij})^2/\sigma_{ij}^2 + c_i$$

where μ_{ij} and σ_{ij}^2 are the class-conditional mean and variance of x_j for class ω_i. The most frequently applied linear classifier can be formulated by the following discriminant function of class ω_i:

$$g_i(\boldsymbol{x}) = \boldsymbol{w}_i{}^T\boldsymbol{x} + b_i.$$

(a) Derive the conditions of $\boldsymbol{x}$ under which the minimum weighted distance classifier minimizes the probability of the misclassification or the error rate.

(10 Marks)

(b) Derive the conditions of $\boldsymbol{x}$ and determine the weighting vector $\boldsymbol{w}_i$, with which the linear classifier minimizes the probability of the misclassification or the error rate. (**Hint:** You may directly use some intermediate results you derived in part (a).)

(8 Marks)

(c) Compare the complexity of the above two classifiers in the training phase and the classification phase.

(7 Marks)

Hint: The general expression of Gaussian PDF is given by:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right].$$

3. The inputs and outputs of a typical fully-connected layer of neural network are denoted by $\boldsymbol{x} = [x_1, x_2, \ldots, x_{100}]^T$ and $\boldsymbol{y} = [y_1, y_2, \ldots, y_{98}]^T$. A linear activation function is applied in this layer and the network parameters to connect the inputs and outputs are given by a matrix $\boldsymbol{W} = [w_{ij}]$ of size 98×100 and a 98-dimensional vector $\boldsymbol{\theta} = [\theta_1, \theta_1, \ldots, \theta_{98}]^T$.

(a) (i) Express the outputs in term of inputs in vector-matrix form and scalar form.

(ii) Compute the numbers of trainable parameters, multiplications and summations required in this layer to compute the outputs from the inputs.

(iii) What is the ratio of the number of outputs to the number of trainable parameters?

(12 Marks)

Note: Question No. 3 continues on page 3.

(b) Replace this layer by a convolutional neural network layer that has 20 learnable filters of size 3 with trainable parameters, $\boldsymbol{w}^k = [w_{-1}^k, w_0^k, w_1^k]^T$ and θ^k, $1 \leq k \leq 20$. This generates 20 output feature maps, denoted by $\boldsymbol{y}^k = [y_2^k, y_3^k, \ldots, y_{99}^k]^T$, $1 \leq k \leq 20$. A linear activation function is applied in this layer.

(i) Express the outputs in term of inputs in scalar form.

(ii) Compute the number of trainable parameters, multiplications and summations required in this layer to compute the outputs from the inputs.

(iii) What is the ratio of the number of outputs to the number of trainable parameters?

(13 Marks)

4. (a) MEI and MHI are two important motion encoders that represent motions from video frames in intuitive manners. Given a frame at time t represented as $I(x, y, t)$ and a frame located in i frames previously represented as $I(x, y, t - i)$, formulate the frame difference between these two frames. Next, formulate the **binary** MEI representation E_τ and MHI representation H_τ. State the difference between the MHI and MEI representations. Include necessary explanations over the notations used in the two representations. Briefly discuss the disadvantages (at least 3 of them) of leveraging MEI and MHI for motion encoding.

(10 Marks)

(b) There are three key assumptions that enable the estimation of optical flow, one of which is the assumption of brightness constancy. Describe the brightness constancy in brief. Suppose that the brightness of a pixel located at (x, y) at the timestamp t is represented as $I(x, y, t)$ and the same pixel is displaced and relocated to $(x + u, y + v)$ at time $t + 1$. Derive the brightness constancy constraint equation using the Taylor expansion with the first-order Taylor approximation. State the key shortcoming of leveraging the brightness constancy constraint for motion identification and how such shortcomings are formulated **mathematically**. To overcome the aperture problem, some algorithms such as Lucas-Kanade estimation leverage on other key assumption(s) for optical flow estimation. What assumption is used for the Lucas-Kanade algorithm?

(5 Marks)

Note: Question No. 4 continues on page 4.

(c) Given four collinear points A, B, C, D, which are projected on the image plane as points a, b, c, d as shown in Figure 1. We further define the cross-ratio as $[A, B, C, D] = \frac{\|AC\|}{\|BC\|} / \frac{\|AD\|}{\|BD\|}$. Prove that the cross-ratio is projective invariant. (*Hint*: use the law of sine).

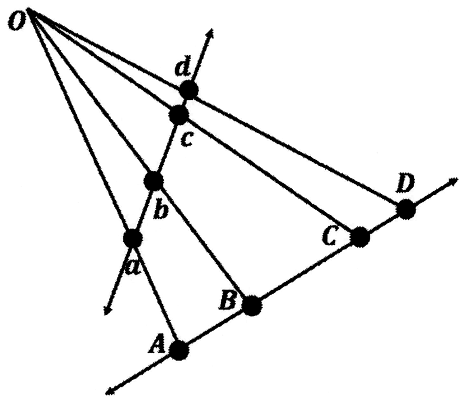

Figure 1

(4 Marks)

(d) Suppose point $\boldsymbol{P}$ is projected onto two image planes at points $\boldsymbol{p}$ and $\boldsymbol{p}'$. Suppose that the intrinsic matrices of the two cameras are both $\boldsymbol{K}$ while the extrinsic matrix of Camera 2 with respect to Camera 1 is defined as $[\boldsymbol{R}, \boldsymbol{t}]$. What is the essential matrix $\boldsymbol{E}$? Derive in detail the epipolar constraint with the essential matrix. Further suppose that the intrinsic matrix of the two cameras is **unknown**. Derive in detail how the epipolar constraint would be re-formulated. (*Hint*: derive and formulate the epipolar constraint with the fundamental matrix)

(6 Marks)

END OF PAPER

EE6222 MACHINE VISION

Please read the following instructions carefully:

1. **Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**

2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.

3. Please write your Matriculation Number on the front of the answer book.

4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.