NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 1 EXAMINATION 2024-2025

EE6222 – MACHINE VISION

November/December 2024   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 filter with impulse response $h(x,y) = 2\delta(x,y) - \delta(x, y-1) - \delta(x-1, y)$ is applied to an image $f(x,y)$ plotted in Figure 1.

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	1	1	1	1	1	-1	-1
-1	-1	1	1	1	1	1	-1	-1
-1	-1	1	1	1	1	1	-1	-1
-1	-1	1	1	1	1	1	-1	-1
-1	-1	1	1	1	1	1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

Figure 1. $f(x,y)$

Suppose the image $f(x,y)$ out of the Figure 1 is constant -1.

(a) Show the filter mask.

(5 Marks)

(b)   Express the output image $g(x, y)$ in terms of the input image $f(x, y)$.

(5 Marks)

(c)   Plot the output image $g(x, y)$ of size 9×9 in the same way as Figure 1.

(10 Marks)

(d)   The semantic meaning of the input image $f(x, y)$ is a square object in a constant background. What is the semantic meaning of the output image $g(x, y)$? How to map image $g(x, y)$ to another image $g_1(x, y)$ to make its semantic meaning clearer?

(5 Marks)

2.   Given a database of c classes, each class has n samples and each sample has m features. Each element of the database is denotted by x_{kij}, $1 \leq i \leq m$, $1 \leq j \leq n$, $1 \leq k \leq c$.

(a)   Compute the class-conditional mean μ_{ki} and the element σ_{kpq} of the covariance matrix.

(6 Marks)

(b)   Construct a column feature vector $\mathbf{x}_{kj}$ of each sample from the given x_{kij} and then compute the class-conditional mean vector $\boldsymbol{\mu}_k$ and the covariance matrix $\boldsymbol{\Sigma}_k$ using $\mathbf{x}_{kj}$.

(6 Marks)

(c)   Construct the class-conditional data matrix $\mathbf{X}_k$ using $\mathbf{x}_{kj}$ and $\boldsymbol{\mu}_k$ computed in (b) such that the mean feature vector of $\mathbf{X}_k$ is zero and compute $\boldsymbol{\Sigma}_k$ using $\mathbf{X}_k$.

(5 Marks)

(d)   We want to extract a one-dimensional feature. The unit length column vector that spans this dimension in the feature space is $\mathbf{a}$. Derive the class-conditional variance v_k of this one-dimensional feature from the definition of the variance using $\mathbf{a}$, $\mathbf{x}_{kj}$ and $\boldsymbol{\mu}_k$ or $\mathbf{X}_k$ to the final result in terms of $\mathbf{a}$ and $\boldsymbol{\Sigma}_k$.

(8 Marks)

3. A two-layer fully connected feed-forward neural network (also called multilayer perceptron) has two inputs, x_1 and x_2, two hidden neurons and one output y.

(a) Draw a diagram of this network and show all network parameters in your diagram.

(5 Marks)

(b) The above network has a nonlinear activation function $f(s) = \dfrac{1}{1 + \exp(-s)}$ for the hidden neurons and linear activation function $g(s) = s$ for the output neuron. Express the output y of the network as a function of the inputs x_1 and x_2.

(5 Marks)

(c) Construct the input vector $\mathbf{x}$, network parameter matrics $\mathbf{W}$ and $\mathbf{V}$ using inputs x_1 and x_2 and parameters you used in (a) and (b) so that the output of the network is expressed in a simple vector/matrix form of $y = g(\mathbf{V}^T f(\mathbf{W}^T \mathbf{x}))$.

(5 Marks)

(d) In a step-by-step manner, derive a learning rule that adjusts the network parameters of the output neuron to make the square error between the network output y and the desired output t smaller.

(10 Marks)

4. Deep learning has been used for video processing and stereo matching.

   (a) What is the difference between ReID and JDE in object detection and tracking?

   (6 Marks)

   (b) Describe the main pipeline of stereo matching.

   (6 Marks)

   (c) In a deep learning stereo matching algorithm such as PSMnet, the cost volume would be up-sampled to size $H \times W \times D$ via bilinear interpolation before the final step of regression to calculate the disparity map with size $H \times W$ where H and W denote the height and width of the input images, respectively. Answer the following questions.

      (i) What does D stand for?

      (ii) How does the value of D affect the stereo matching results?

      (iii) What is the physical meaning of a voxel in the cost volume C at voxel (h, w, d)?

      (13 Marks)

END OF PAPER

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.