NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 1 EXAMINATION 2022-2023

EE6427 – VIDEO SIGNAL PROCESSING

November / December 2022 Time Allowed: 3 hours

INSTRUCTIONS

1. This paper contains 5 questions and comprises 5 pages.
2. Answer all 5 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. The two-dimensional Discrete Cosine Transform (2-D DCT) of an $N \times N$ data matrix is given by:

$$S_{uv} = \alpha(u)\alpha(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} s_{ij} \cos\frac{(2i+1)u\pi}{2N}\cos\frac{(2j+1)v\pi}{2N} \qquad u, v = 0, \ldots, N-1$$

where

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } k = 0 \\ \sqrt{\frac{2}{N}} & \text{for } k = 1, 2, \ldots, N-1 \end{cases}.$$

(a) Calculate the 2-D DCT of the following 4×4 pixel block **A**.

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 10 & 10 & 0 \\ 0 & 10 & 10 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(7 Marks)

Note: Question No. 1 continues on page 2.

(b) Based on your result in part (a), calculate the 2-D DCT of the following pixel block **B**.

$$\mathbf{B} = \begin{bmatrix} 20 & 20 & 20 & 20 \\ 20 & 15 & 15 & 20 \\ 20 & 15 & 15 & 20 \\ 20 & 20 & 20 & 20 \end{bmatrix}$$

(4 Marks)

(c) DCT is used in the baseline JPEG compression. Draw a simple block diagram of JPEG encoder. Clearly label all the key components in the diagram.

(5 Marks)

(d) A student makes the following claim: "*DCT in the baseline JPEG compression is effective in compressing texture image patches as it can achieve excellent energy compaction that further provides effective entropy encoding for these patches*". State clearly whether you agree or disagree with the claim, and briefly justify your answer.

(4 Marks)

2. (a) A Vanilla Recurrent Neural Network (RNN) has the following settings.

Initial hidden state, $\mathbf{h}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,

Hidden state weight matrix, $\mathbf{W}_{hh} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix}$,

Input weight matrix, $\mathbf{W}_{xh} = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{bmatrix}$,

Output weight matrix, $\mathbf{W}_{hy} = \begin{bmatrix} 0.1 & 0.4 \end{bmatrix}$.

Assume no bias is used in the computation of the RNN.

A 2-timestep input is given by $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}$ where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$.

(i) Find the hidden state $\mathbf{h}_1$ at timestep $t = 1$.

(ii) Find the output y_1 at timestep $t = 1$.

(iii) Find the output y_2 at timestep $t = 2$.

Note: Question No. 2 continues on page 3.

(iv) A user would like to use a Vanilla RNN to predict the price of a share based on the following input features: the price, the trade volume, and the bid number of the share for the past 5 days. Suggest the change(s) that should be made to the RNN structure above to perform the share price prediction.

(15 Marks)

(b) A user would like to develop an image classification application using a model that can achieve good accuracy and uses attention mechanism when performing classification. He is considering the following 3 candidate models: (i) VGG, (ii) Vision Transformer (ViT), and (iii) Long Short-Term Memory (LSTM). State which model is most likely going to meet the user's need and briefly justify your answer.

(5 Marks)

3. (a) Briefly describe the main difference(s) between one-stage detectors and two-stage detectors for object detection in videos.

(4 Marks)

(b) With the aid of a diagram, briefly describe the key components in a two-stream network for human action recognition in videos.

(6 Marks)

(c) What are two typical approaches used for stereo matching? Discuss the respective advantages of these two approaches.

(4 Marks)

(d) The brightness constancy constraint is based on the assumption that for a small space-time step, the brightness of a point remains the same, i.e.,

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t).$$

Using the assumption, derive the brightness constancy equation.

(6 Marks)

4. (a) Draw a diagram of the MPEG encoder with the following blocks.

- Discrete Cosine Transform (DCT)
- Inverse Discrete Cosine Transform (IDCT)
- Quantization (Q)
- Inverse Quantization (IQ)
- Motion Estimation (ME)
- Variable Length Coding (VLC)
- Buffer

(4 Marks)

Note: Question No. 4 continues on page 4.

(b) Video compression algorithms such as MPEG use the slice structure when encoding a video frame.

(i) Discuss why the slice structure is used to encode a video frame.
(ii) Give an advantage and a disadvantage of using multiple slices per frame as compared with using one slice per frame in video coding.

(6 Marks)

(c) In the MPEG video coding, a Group of Pictures (GOP) structure may contain I-frame, B-frame(s) and P-frame(s). Briefly describe I-frame, B-frame and P-frame in a GOP.

(4 Marks)

(d) Huffman coding uses variable length coding in which a bitstream can be uniquely decodable into codewords. The symbols S = {a, b, c, d, e, f} have the corresponding codewords as shown in Table 1. State clearly whether the codewords can be Huffman codewords, and briefly justify your answer.

Table 1

S	a	b	c	d	e	f
Codeword	1100	1001	101	111	10	0

(6 Marks)

5. (a) A matrix **C** is given by

$$\mathbf{C}=\begin{bmatrix} 40 & 100 & 100 & 40 \\ 100 & 180 & 180 & 100 \\ 100 & 180 & 180 & 100 \\ 40 & 100 & 100 & 40 \end{bmatrix}.$$

Find the outputs of one-level Haar wavelet transform and two-level Haar wavelet transform of **C**.

(8 Marks)

(b) A matrix **A** is given by

$$\mathbf{A}=\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}.$$

The singular value decomposition of an $m \times n$ matrix **A** is a factorization of the form $\mathbf{A} = \mathbf{U\Sigma V}^{\mathrm{T}}$.

Note: Question No. 5 continues on page 5.

(i) Show that $\mathbf{\Sigma} = \begin{bmatrix} \sqrt{40} & 0 \\ 0 & \sqrt{10} \end{bmatrix}$.

(ii) Find the matrices **U** and **V**.

(iii) Given an observation matrix **W** with size $2F \times N$, where F is the number of frames and N is the number of feature points, briefly describe how to perform Tomasi–Kanade factorization in structure from motion application.

(12 Marks)

END OF PAPER

EE6427 VIDEO SIGNAL PROCESSING

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.