CISC3021 Multimedia Forensics and Security - Lab1

Huffman coding is a method used for lossless data compression, which builds an optimal prefix code based on the frequency of characters. Here are the basic construction rules for Huffman coding:

1. **Calculate Frequencies**: First, calculate the frequency of each character in the file to be compressed.
2. **Create a Priority Queue (Min Heap)**: Place each character along with its frequency as a node into a priority queue, where nodes with smaller frequencies have higher priority.
3. **Build the Huffman Tree**:
    - If there are more than two nodes in the priority queue, perform the following steps:
        - Remove the two nodes with the smallest frequencies from the priority queue, and merge them into a new node whose frequency is the sum of their frequencies.
        - Insert the new node back into the priority queue.
    - Repeat these steps until only one node remains in the priority queue; this node becomes the root of the Huffman tree.
4. **Assign Codes**:
    - Starting from the root node, agree that the left branch represents 0 and the right branch represents 1, and assign a unique Huffman code to each leaf node.
5. **Encode Output**: Using the constructed Huffman tree, encode each character according to its assigned Huffman code.

Please solve the following questions using Matlab or Python or your hands.

Exercise 1 - Basic Construction

Problem Description:

Given a list of characters and their frequencies, construct the corresponding Huffman tree and give the Huffman code for each character.

Example:

Assume we have the following characters and their frequencies:

- A: 40
- B: 12
- C: 20
- D: 13
- E: 15

Requirements:

- Construct a Huffman tree.
- Give the Huffman code for each character.

Exercise 2 - Encoding Efficiency

Problem Description:

Compare the compression efficiency of Huffman coding with other encoding methods (such as fixed-length coding).

Example:

Assume we have a sequence of characters with the following frequencies:

- A: 50
- B: 25
- C: 15
- D: 10

If using fixed-length coding, each character takes up 2 bits. If using Huffman coding, calculate the total number of bits for both encoding methods and compare their compression efficiencies.

Exercise 3 - Decoding Process

Problem Description:

Given a Huffman tree and a Huffman encoded string, decode the original character sequence.

A： 0　B:10　C:110 D:111

Example:

The Huffman encoded string is 100111110110, decode the corresponding character sequence.

Exercise 4 - Practical Application

Problem Description:

Design a simple text file compression program using Huffman coding to compress text and correctly decompress it.

Requirements:

- Write a program that reads a text file and counts the frequency of each character.
- Use Huffman coding to compress the text.
- Store the compressed data in another file.
- Write a decompression program that reads the compressed file and restores it to the original text file.

To be, or not to be, that is the question: Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles And by opposing end them. To die—to sleep, No more; and by a sleep to say we end The heart-ache and the thousand natural shocks That flesh is heir to: 'tis a consummation Devoutly to be wish'd. To die, to sleep; To sleep, perchance to dream—ay, there's the rub: For in that sleep of death what dreams may come, When we have shuffled off this mortal coil, Must give us pause—there's the respect That makes calamity of so long life. For who would bear the whips and scorns of time, Th'oppressor's wrong, the proud man's contumely, The pangs of dispriz'd love, the law's delay, The insolence of office, and the spurns That patient merit of th'unworthy takes, When he himself might his quietus make With a bare bodkin? Who would fardels bear, To grunt and sweat under a weary life, But that the dread of something after death, The undiscovere'd country, from whose bourn No traveller returns, puzzles the will, And makes us rather bear those ills we have Than fly to others that we know not of? Thus conscience doth make cowards of us all, And thus the native hue of resolution Is sicklied o'er with the pale cast of thought, And enterprises of great pith and moment With this regard their currents turn awry And lose the name of action.