LaTeX Huffman Encoding for SBE201 Project Report

Ahmed Hossam Mohammed
ahmed.mohamed16@eng-st.cu.edu.eg

Esraa Mohammed Saeed
esraa.mohamed99@eng-st.cu.edu.eg

Alaa Tarek Samir
Alaa.mohamed99@eng-st.cu.edu.eg

Amira Gamal Mohammed
amira.omar99@eng-st.cu.edu.eg

Sohaila Mohammed Maher
Sohila.mohamed99@eng-st.cu.edu.eg

1. Introduction

David Huffman developed the Huffman algorithm in 1951. This algorithm is mainly used for data compression (which is based on encoding data). Here is how it works, It depends on the frequency of the input data in the file. The characters, which is most generated or repeated, will take the small binary code, and the characters, which is least generated or repeated, will take the large binary code. The method of representing data is called the Huffman binary tree of nodes. It is built by a priority queue and the node with the lowest frequency is the highest priority. Here we divided the code into three main parts, the first one is to read and write PGM files, the second one is to encode the files, build the Huffman tree, and generate the compressed files, and finally the third part is to decode the new generated files to get the original PGM files.

2. Motivation

We knew about Huffman algorithm and its applications in various fields. So the topic was quite interesting for us to search for. We also thought that it is going to help us in the data base course next year. Another thing is that compressed files help greatly with space and memory problems especially, in case of huge amount of data or files like in hospitals for example.

3. Resources

1. $include < opencv2/opencv.hpp >$ This is the library we use to read/write PGM files
2. Using namespace cv We use this line to avoid writing cv:: before imread/imwrite.
3. $include < vector >$ we use vector to write PGM file
4. $include < fstream >$ This library helps us to deal with an external file open/close or write in it.

4. Challenges and Problems

1. We faced a problem that the picture "NORMAL2 -IM-1431-000.1.pgm" is the only picture, which includes comment: , but our code read the file line by line and expects to have a comment in the second line, width and height in the Third line. However, the other seven pictures don't have a comment in the second line so it took the width and height as a comment.

Figure 1. Problem (1)

2. But we faced another problem, the pixel values that we put in the text file is slightly different than pixel values we get from the MATLAB.

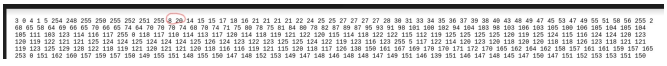

Figure 2. Problem (2)

3. At the beginning of the encoding part we faced the problem of reading the .txt files that has pixel values as the files contained taps and end lines after each row that took too much space and made it difficult to read the files.

4. The decoding part the images were not the same as the original images as there. We later discovered that it was because of the R letter in the X-ray images it was pure white that is represented by -1 and the encoding part only read positive values and the zero .

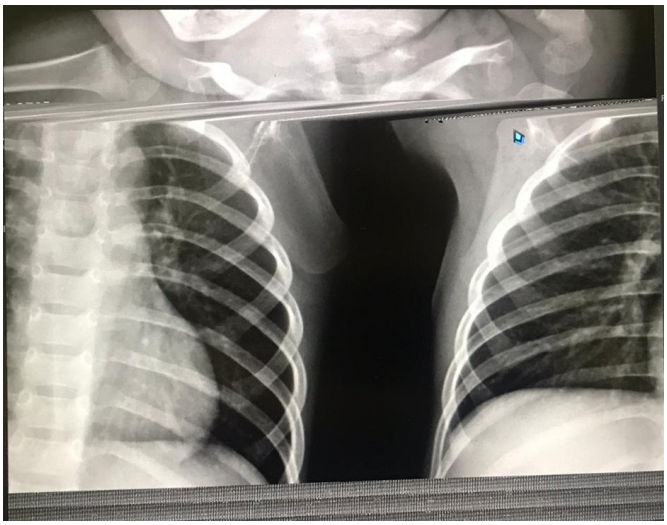

Figure 3. Problem (4)

5. User manual for the system

1. To run the test.cpp file (read the Images) open the terminal and write the following command: "$g++test.cpp - ooutput`pkg - config - -cflags - -libsopencv`$" then "$./a.out$".
2. To run the encoding file open the terminal and write the following command: "$g++EncodingAA.cpp$" then "$./a.out$".
3. To run the Decoding file: open the terminal and write the following command: "$g++DecodingH.cpp$" and then "$./a.out$".

6. Results

Screenshots for your running programs

7. Contributions

1. Esraa Moahmmed Saeed : Responsible for the first part (reading the PGM files) and improvement 2.1
2. Esraa Mohammed Saeed and Ahmed Hossam Mohammed: responisble for writing the report the parts related to the openCV library and the read PGM files.
3. Amira Gamal Mohammed and Alaa Tarek Samir : Responsible for the Encoding and compression part, writing the report the parts related to the encoding and Improvement 2.2
4. Ahmed Hossam Mohammed: responsible for the decoding part.
5. Sohaila Mohammed Maher: responsible for writing the report the parts related to the Decoding and Improvement 2.1

References