**PROJECT REPORT**

**TOPIC-BLACK AND WHITE IMAGE COMPRESSION AND TEXT COMPRESSION THROUGH HUFFMAN CODING**

SUBMITTED BY-

MANANYA KOHLI (2022275)

PANDILLAPELLY HARSHVARDHINI (2022345)

ADITYA KUMAR SINHA (2022034)

**ABSTRACT**

Our project focuses on black and white image compression as well as text compression in the light of Huffman coding. The project provides detailed insights of how Huffman coding works, from frequency analysis to generation of Huffman codes and subsequent compression and decompression methods. Moreover, the mathematical proof establishes that Huffman coding adheres to the prefix property, ensuring uniquely decodable encoded sequences.

It covers the algorithm and implementation of the text and image compression. At last, it gives the limitations and conclusion of the huffman coding method.

**INTRODUCTION**

Huffman coding is a variable-length prefix coding algorithm used for data compression. It was developed by David A. It is a greedy technique and we use this to encode data and compress the data. Here encoding means we have to digitize it, and write it in the form of 0 and 1.

It is widely used in various compression algorithms, including those used for image, audio, and video compression, as well as in file compression formats like ZIP.

The principle on which huffman encoding operates is variable length encoding , in this suppose if frequency of some alphabet is less then it will be assigned a larger binary code in comparison to the alphabet which has a high frequency , and this is the very reason why it is so good in compression.

Huffman encoding operates by making a huffman tree where each leaf node represents a symbol and its associated frequency.This tree is built from the bottom , and each symbol behaves as a node here. and the nodes with the lowest frequencies are present at the bottom of the tree, and the summation of their frequency give rise to a new node, this process continues until a single root node is formed.

We are using black and white images for simplicity in the compression process through huffman coding.

**EXISTING LITERATURE**

In the vast landscape of image compression techniques, researchers have explored several methods to effectively reduce image file sizes.

Three notable methods are LZW (Lempel-Ziv-Welch), Run-Length Encoding, and Huffman coding. Among these, Huffman coding stands out as an approach based on information theory.This coding method assigns different codes to different parts of the image based on the probability of occurrence.

Adaptive versions such as Vitter's algorithm, this act a bit smarter than the huffman algorithm. Instead of generating all the short codes at start, they change and generate as they read.

Thus, if a new word or pattern occurs frequently, the algorithm can adjust its codes to make them even shorter. This is more concise. Here is the method that we have studied:

**THE TOPIC OF STUDY**

This study focuses on the use of Huffman coding for grayscale image compression,which aims to optimize the pixel intensity representation that reduces the file size without sacrificing image quality.

The first step is to compute a probability distribution *P( X = xi )*, where X represents pixel intensity values, and *xi* denotes a particular intensity level, where *i* is an index representing a specific intensity value.

This probability distribution basically indicates the probability of encountering different pixel intensities in a grayscale image. In this, shorter codes are assigned to high-efficiency pixels enhancing compression efficiency.

We tried to implement something new which incorporates the concept of image smoothing with techniques such as Gaussian filtering, which helps to optimize the overall pixel representation.

This compression scheme is designed not only to reduce storage requirements but also to speed up the transmission and retrieval of grayscale images, making Huffman coding a valuable tool in visualization and data transmission as studies progress deepen the fascinating speed of this application. Huffman tree construction, and the resulting binary representation, eventually contributes to a broader understanding of Huffman coding in the context of grayscale graphics..

**CONCLUSION**

This study highlights the effectiveness of Huffman coding as a powerful tool for compressing grayscale image data, without compromising the ability to reconstruct images.

The compression ratio (CR), a key metric calculated by dividing the original image size by the size of the encoded image.The higher the compression ratio, the more just the image data is compressed. As we know, one of the limitations of the huffman algorithm is “If the distribution changes, the optimal code may change, requiring a new tree.”

So, looking ahead, there exists future exploration of ​​developing Huffman variable coding. Adaptive Huffman coding, with the ability to dynamically update the codebook during the encoding process, presents opportunities for even greater compression efficiency.