# Huffman Encoding Algorithm and Its Application in Image Compression

Team Members:
Miten Shah (180001049)
Sundesh Gupta (180001057)

# Overview

As part of our Algorithm project, we have decided to work on “Application of Huffman Encoding Algorithm in Image Compression” and implement **An Efficient Encoding Algorithm Using Local Path on Huffman Encoding Algorithm for Compression, 2019** paper by E. Erdal et al. [1].

# Goals

1. Implement Huffman encoding and arithmetic coding algorithms.
2. Analyze Huffman encoding and arithmetic coding for image compression.
3. Implement and Analyze improved Huffman encoding for image compression.
4. Test the performance of the improved huffman encoding algorithm on different sets of pictures including grayscale and colored images and compare with existing algorithms.
5. Further optimize Huffman encoding for lesser time complexity and higher compression ratio, if possible.

# Specifications

We briefly explain the required topics for our project.

## Huffman Encoding

The Huffman encoding algorithm is one of the most commonly used data compression methods in the field of computer science. The algorithm developed by David Huffman is used to reduce coding redundancy without loss of data quality. The use of data frequency is the basic idea in the Huffman encoding algorithm. The approach in the algorithm assigns symbols from the alphabet with variable codewords according to their frequency. The symbol with a higher usage frequency is symbolized by shorter codes in order to obtain a higher compression result.

Huffman's method can be efficiently implemented in time linear **O(n)** to the number of input weights if these weights are sorted. However, Huffman encoding is not always optimal among other compression methods like Arithmetic Coding Algorithm and does not always achieve a higher data compression ratio.

## Arithmetic Coding Algorithm

Arithmetic coding is one of the most used data compression methods in the literature that decreases coding redundancy, like Huffman coding. It creates a code string that symbolizes fractional values between 0 and 1. The fundamental idea used in arithmetic coding is to divide the values from 0 to 1 according to the probability of the symbols in the message. When the symbol count is small and there are less differences in the probability of the characters, Huffman encoding loses its effectiveness, while arithmetic coding is more successful.

For Arithmetic coding, time complexity depends on a number of different input symbols and length of a message. So for length of message 'n', **n+n*|Σ|** is the computation cost , where |Σ| is a number of different input symbols. That is, at maximum **O(n*|Σ|)**.

## Image Compression

Image compression is a type of data compression applied to digital images, to reduce their cost for storage or transmission. A set of procedures used to compress an image is called an image compression system. Compressor and decompressor operations generate an image compression system. The compressor process consists of two stages: preprocessing and encoding. Unlike the compressor process,the decompressor operation consists of the postprocessing phase after the decoding process

The encoding phase is one of the most important steps in compression algorithms. Huffman encoding and Algorithmic coding algorithms are used in the encoding phase of image compression. Although the arithmetic coding algorithm has been developed to improve the efficiency of the Huffman encoding algorithm, it has deficiencies in mathematical processing and complexity. The improved Huffman encoding algorithm overcomes these limitations and is more efficient as well as offers more successful compression results.

## Improved Huffman Encoding Algorithm and Its Application in Image Compression

The Huffman encoding algorithm and arithmetic coding algorithm are frequently used coding algorithms in the kernel of compression algorithms. However, an improved encoding algorithm was developed that used the advantages of the Huffman encoding algorithm, and eliminated the gaps in the arithmetic coding algorithm. This improved encoding algorithm used a modified tree structure from the Huffman encoding algorithm.

The main objective of the improved Huffman Encoding algorithm, and all other compression algorithms, is to reduce the number of bits needed to symbolize the characters.Image compression measurements include:

1. Compression Percentage (CP):

$$Compression\ Percentage = 100\left(1 - \frac{Compressed\ file\ size\ (bytes)}{Original\ file\ size\ (bytes)}\right)$$

2. Number of bits per pixel (NoBPP):

$$Compression\ Percentage = 100\left(1 - \frac{Compressed\ file\ size\ (bytes)}{Original\ file\ size\ (bytes)}\right)$$

## Bibliography


1. An Efficient Encoding Algorithm Using Local Path on Huffman Encoding Algorithm for Compression, Erdal Erdal and Atilla Ergüzen (2019) [ref.]
2. Lossless Image Compression through Huffman Coding Technique and Its Application in Image Processing using MATLAB Vikash Kumar, Sanjay Sharma (2017) [ref.]
3. An Introduction to Arithmetic Coding, G. G. Langdon (1984) [ref.]
4. ttps://en.wikipedia.org/wiki/Huffman_coding
5. https://en.wikipedia.org/wiki/Image_compression