

# AN ALGORITHM FOR IMAGE COMPRESSION USING HUFFMAN CODING TECHNIQUES

**Sanjay Kumar Gupta**

*Dronacharya College of Engineering, Gurgaon, (India)*

## ABSTRACT

*Compression is used about everywhere. Images are very important documents nowadays; to work with them in some applications they need to be compressed, more or less depending on the purpose of the application. The need for an efficient technique for compression of Images ever increasing because the raw images need large amounts of disk space seems to be a big disadvantage during transmission & storage. There are various algorithms that performs this compression in different ways; some are lossless and keep the same information as the original image, some others loss information when compressing the image. In this paper we proposed the Lossless method of image compression and decompression using a simple coding technique called Huffman coding. This technique is simple in implementation and utilizes less memory.*

**Keywords:** *Huffman Encoding, Compression, Lossless*

## I INTRODUCTION

**Compression** refers to reducing the quantity of data used to represent a file, image or video content without excessively reducing the quality of the original data. Image compression is the application of data compression on digital images. In effect, the objective is to reduce redundancy of the image data in order to be able to store or transmit data in an efficient form [1]. Data compression has become requirement for most applications in different areas such as computer science, Information technology, communications, medicine etc. In computer science, Data compression is defined as the science or the art of representing information in a compact form [2]. It also reduces the number of bits required to store and/or transmit digital media. To compress something means that you have a piece of data and you decrease its size. There are different techniques and they all have their own advantages and disadvantages. Huffman coding is a lossless data compression technique. Huffman coding is based on the frequency of occurrence of a data item i.e. pixel in images. The technique is to use a lower number of bits to encode the data in to binary codes that occurs more frequently. It is used in JPEG files.

## II FUNDAMENTALS FOR COMPRESSION

Compression can be divided into two categories, as Lossy and Lossless compression.

**Lossy compression** means that some data is lost when it is decompressed. Lossy compression bases on the assumption that the current data files save more information than human beings can "perceive". Thus the irrelevant data can be removed.

**Lossless compression** means that when the data is decompressed, the result is a bit-for-bit perfect match with the original one. The name lossless means "no data is lost", the data is only saved more efficiently in its compressed state, but nothing of it is removed.

### 2.1 Compression Principle

A common characteristic of most of the images is that the neighboring pixels are correlated and therefore contain redundant information. The foremost task then is to find less correlated representation of the image. Two fundamental components of compression are redundancy and irrelevancy reduction.

- a. Redundancy reduction aims at removing duplication forms of the signal (Image/Text).
- b. Irrelevancy reduction omits parts of the signal that will not be noticed by signal receiver, namely the Human visual system.

In an image there are three types of redundancies in order to compress file size. They are:

- a) **Coding redundancy:** Fewer bits to represent frequently occurring symbols.
- b) **Interpixel redundancy:** Neighboring pixels have almost same value.
- c) **Psycho visual redundancy:** Human visual system cannot simultaneously distinguish all colours.

### 2.2 Huffman coding

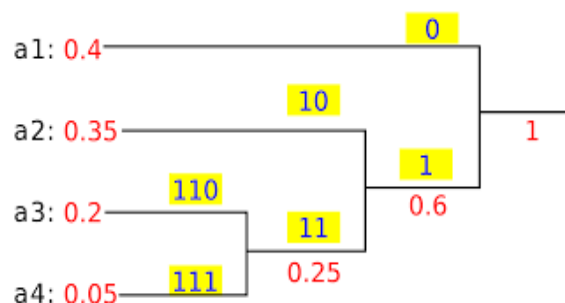
Huffman coding is classical data compression techniques invented by David Huffman. It is optimal prefix code generated from set of probabilities and has been used in various compression applications. These codes are of variable code length using integral number of bits. This idea causes a reduction in the average code length and thus overall size of compressed data is smaller than the original.

Huffman code procedure is based on the two observations [3].

- a) More frequently occurred symbols will have shorter code words than symbol that occur less frequently.
- b) The two symbols that occur least frequently will have the same length.

The Huffman code is designed by merging the lowest

probable symbols and this process is repeated until only two probabilities of two compound symbols are left and thus a code tree is generated and Huffman codes are obtained from labeling of the code tree.



**Fig: Huffman coding**

At the far left of the table I the symbols are listed and corresponding symbol probabilities are arranged in Decreasing order and now the least probabilities are Merged as here 0.05 and 0.2 are merged, this gives a Compound symbol with probability 0.25 and the compound symbol probability is placed such that again the probabilities should be in decreasing order. So this process is continued until only two probabilities are left at the far right shown in the above table as 0.6 and 0.4.

The average length of the code is given by the average of the product of probability of the symbol and number of bits used to encode it.

## Entropy (H)

The entropy of a symbol is defined as the negative logarithm of its probability. To determine the Information content of a message in bits we express the entropy.

$$H = -\sum_{i=1}^N (P_i \log_2 P_i)$$

Here  $P_i$  is the occurrence probability of symbol  $S_i$ . In compression, entropy determines how many bits of information are actually present in a message.

## Compression Ratio (CR)

Compression ratio (CR) is defined as number of bits to represent the size of original image to the number of bit to represent the size of compressed image. Compression ratio shows that how much time the image has been compressed.

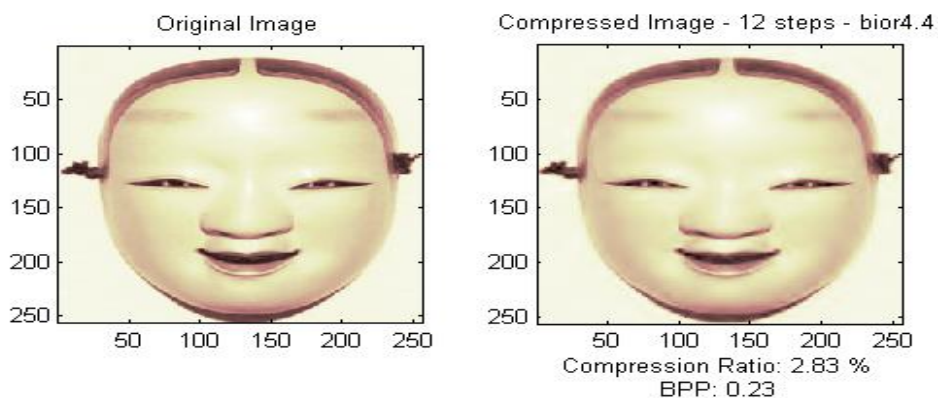
## III DEVELOPMENT OF HUFFMAN CODING AND DECODING ALGORITHM

- 1) Read the image on the workspace of mat lab
- 2) Call a function which will find the symbols (i.e. pixel value which is not repeated).
- 3) Call a function which will compute the probability of each symbol.
- 4) Probabilities of symbols are arranged in decreasing order and lower probabilities are merged. This step is continued until only two probabilities are left and codes are assigned according to rule that, highest probable symbol will have a shorter length code.
- 5) Further Huffman encoding is performed i.e. mapping of code words to the corresponding symbols will in compressed data
- 6) The original image is reconstructed i.e. decompression is done using Huffman decoding.
- 7) Match the code words with code dictionary to get the reconstructed image.
- 8) Generate a tree equivalent to the encoding tree.
- 9) Read input character wise until last element is reached.
- 10) Output the character encode in the leaf and return to the root, and continue the step9 until all the Codes of corresponding symbols are known.

## IV RESULT

We have developed Huffman algorithm for image and analyzed performance parameter like compression ratio and bits per pixel (bpp).

Simulation results show the lossless image compression scheme.



The final compression ratio (2.8%) and Bit-Per-Pixel ratio (0.23) are very satisfactory. Recall that the CR means that the compressed image is stored using only 2.8% of the initial storage size.

## V ADVANTAGES AND DISADVANTAGES

### 5.1 Advantages of Data Compression:

- Less disk space required
- Faster writing and reading
- Faster file transfer
- Variable dynamic range
- Byte order independent

### 5.2 Disadvantages of Data Compression:

- Added complication
- Effect of errors in transmission
- Slower for sophisticated methods (but simple methods can be faster for writing to disk.)
- ``Unknown" byte / pixel relationship
- Need to decompress all previous data

## VI CONCLUSION

Starting from a given image, the goal of true compression is to minimize the number of bits needed to represent it, while storing information of acceptable quality.

The final compression ratio (2.8%) and Bit-Per-Pixel ratio (0.23) are very satisfactory.

### Current Research

A wide research is being done in the Image processing technique.

1. Cancer Imaging – Different tools such as PET, MRI, and Computer aided Detection helps to diagnose and be aware of the tumor.
2. Brain Imaging – Focuses on the normal and abnormal development of brain, brain ageing and common disease states.

3. Image processing – This research incorporates structural and functional MRI in neurology, analysis of bone shape and structure, development of functional imaging tools in oncology, and PET image processing software development.

4. Imaging Technology – Development in image technology have formed the requirement to establish whether new technologies are effective and cost beneficial. This technology works under the following areas:

- Magnetic resonance imaging of the knee
- Computer aided detection in mammography
- Endoscopic ultrasound in staging the esophageal cancer
- Magnetic resonance imaging in low back pain

## V FUTURE

We all are in midst of revolution ignited by fast development in computer technology and imaging. Against common belief, computers are not able to match humans in calculation related to image processing and analysis. But with increasing sophistication and power of the modern computing, computation will go beyond conventional, Von Neumann sequential architecture and would contemplate the optical execution too. Parallel and distributed computing paradigms are anticipated to improve responses for the image processing results.

## REFERENCES

- [1] Mamta Sharma, “Compression Using Huffman Coding”, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.5, May 2010, 133-141.
- [2] Sachin Dhawan,”A Review of Image Compression and Comparison of its Algorithms”, Dept. of ECE, UIET, Kurukshetra University, Haryana, India. IJCET VOL.2, Issue 1, March 2011.
- [3] JAGADISH H. PUJAR , “A NEW LOSSLESS METHOD OF IMAGE COMPRESSION AND DECOMPRESSION USING HUFFMAN CODING TECHNIQUES”,  
Journal of Theoretical and Applied Information Technology
- [4] Dalvir Kaur, “Huffman Based LZW Lossless Image Compression Using Retinex Algorithm”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 8, August 2013
- [5] Sandhya Sharma, “ Image Compression using hybrid of DWT,DCT and Huffman Coding”, “International Journal for Science and Emerging Technologies with Latest Trends” 5(1): 19-23 (2013)
- [6] Rafael C. Gonzalez & Richard E Woods, “digital image processing”
- [7] Rudra Pratap, “getting started with matlab”
- [8] David Solomon ,”Huffman coding”

- [9] Darrel Hankersson, Greg A. Harris, and Peter D. Johnson Jr. Introduction to Information Theory and Data Compression. CRC Press, 1997
- [10] Introduction to Data Compression, Khalid Sayood, Ed Fox (Editor), March 2000
- [11] [http://en.wikipedia.org/wiki/Adaptive\\_Huffman\\_coding](http://en.wikipedia.org/wiki/Adaptive_Huffman_coding)
- [12] [http://en.wikipedia.org/wiki/Huffman\\_coding](http://en.wikipedia.org/wiki/Huffman_coding)
- [13] Gilbert Held and Thomas R. Marshall. Data and Image Compression: Tools and Techniques