

# Data Compression using Huffman Coding Techniques & Comparison

Santosh Kumar Choudhary<sup>1</sup>, Dr. Deepak Sharma<sup>2</sup>

<sup>1</sup>Dept. Computer Science & Applications, Monad University, Hapur U.P

<sup>2</sup>Assoc. Prof., Dept. Computer Science & Applications, Monad University, Hapur U.P

**Abstract - Data Compression is a technique used now days everywhere. Basically, it is the technology which is used in reducing the size of data by eliminating the redundancy of data available in the original file. It is essential for reducing the size of data while storing data and reduces time access as well as easy to store large file by data compression technique.**

**Data compressions may be either Lossy or Lossless. The lossy technique may not regenerate the original form of the source while the Lossless technique regenerates the original form of the source during data compression. Basically, Lossy technique used to compress data for video & audio. Due to this technique, the number of bits required while storing data is also reduced.**

**There are various techniques available to reduce a large amount of data in which I have analyzed the Huffman coding algorithm and compare it with other algorithms available such as Arithmetic coding. Huffman Coding is easy while the implementation & utilization of reducing data storage.**

**Index Terms - Data compression, Huffman coding, Lossy compression, Lossless compression, Arithmetic coding.**

## 1.INTRODUCTION

Data compression is a method of representation of data in the form of the compact and reduced form so that size of the file can be stored in less memory as well as fast data transmission.

Basically, data compression is divided into two forms one is Lossy data compression, and the other one is Lossless data compression. In Lossy data compression data cannot recover in original form after data compressed that is this technique involves loss of data which is very important. Whereas in Lossless data compression technique involves an exact original form of data recovered after data compression. There is various use of this technique which involves medical use, even though the computer understands higher language, which means the executable file can be

definitely compressed by lossless data compression technique. Using lossy data compression method audio and video file can easily compressed. There is a high demand for data storage technique with maintaining privacy as well as better data transmission efficiency.

There is lots of lossless data compression technique involved such as Huffman algorithm, Arithmetic algorithm and Run length encoding techniques.

Compression is techniques which involve decreasing size of the data compared to the original size of data. Whereas in video and audio, the content of the quality of data should not be reduced to the quality of the original size of data. Also, the size of data is reduced in such a way that storage capacity is increased at least 6 times as compared with the original size of data.

Using Data Compression method, first of all, to find out the redundancy of data available in original file then eliminating the redundancy of data available in the source file. In Arithmetic algorithm technique data compression used for lossless data, compression was using variable-length encoding algorithm method such that whole original size of the message is converted into a single number and each message is indicated with the particular symbol which forms a code word. In Run-length encoding, algorithm technique is used for lossless data compression, comparatively less data compression used for image compression. It is a sequential form of same data occurred in a consecutive way are stored in single value and counted. This technique is not used for in which many runs are not available in the source file. In Huffman algorithm technique used for lossless data compression, this technique deals with the frequency of occurrence of data. As more frequently occurring symbol indicated with fewer bits compared with the less occurring symbol.

## 2. BACKGROUND

Data compressions the technique of decreasing data in the form of decoding, encoding, and converting as per required information so that data can be stored in comparatively less storage space. Data is used to indicate the same information, and which contain unnecessary/repeated information is called redundant data. Let x and y indicate no. of bits which represent the same information is called relative data redundancy (Re). It represents the percentage of redundant data.

$$Re = (1 - 1/Co)$$

Whereas Co represents compression ratio  $Co = x/y$

### 2.1 Various forms of redundancy

There are various forms of redundancy basically redundancies that can be identified in a two-dimensional representation of the image

- a. Coding redundancy: - A set of data is represented as Code in the form of symbols. Each piece of information or event is assigned a sequence of code symbol called Codewords. The various amount of symbol in each codeword has its length. In coding redundancy, the symbol which is used many times in image set a less number of symbols. In this way reducing the size of the image.
- b. Spatial redundancy-this types of redundancy called an inter-pixel redundancy. The neighbourhood pixels are correlated. So, the redundant data may be deleted to compress the image.
- c. Irrelevant Data: - The data ignored by the human visually is redundant data.

Data compression is attained when redundancies are eliminated. Huffman coding method uses a coding redundancy to compress the image.

### 3. DIFFERENT TYPES OF COMPRESSION

Lossless Data Compression technique in which data is uncompressed hence each bit is exactly matched with original data. It involves with no loss of information after uncompressed data, data stored with an efficient way in which not a single data is lost.

Lossy Data Compression technique which involves there is the loss of some information after uncompressed. Only important data is still available. By this method, data is reduced by eliminating redundancy data. Basically, this data compression

used for audio and video. Lossy data compression used in JPEG file.

Colour coding is the technique used in graph theory and detects the shortest path in the network as well as used to find out cycle in NP-complete problem.

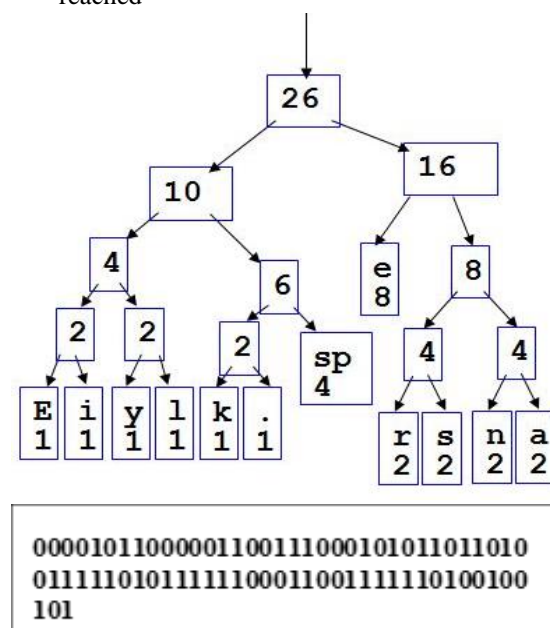
Digital coding is the technique through which letters, characters as well as symbols are represented in digital forms. The basic principle behind this technique is combining binary digits to depict in the form of a character. Data compression is done by using the depiction of the mean size of each symbol which is less than the standard form.

### 4. HUFFMAN CODING TECHNIQUE FOR CODING REDUNDANCY

Huffman coding method of compression uses a coding redundancy technique. In an image 2D matrix, the number of different symbol present is calculated. In the 8-bit representation, each pixel uses 8 bits to represent intensity in the image. With the probability method, the largest occurrence of the symbol will be calculated. This method reduces the number of bits for largely occurrence symbol. In this way, the storage capacity of the image may reduce.

Encoding the Data and Traverse Tree for Codes

- Traverse tree to discover new codewords
- Going left node indicated as 0 and going right node indicated as 1
- Codewords cannot achieve until a leaf node is reached



Encoding the File Results

- 73 bits to encode information
- ASCII would take  $8 * 26 = 208$  bits
- If modified code uses 4 bits per character if needed.
- Total bits equal to  $4 * 26 = 104$ .

Decoding the File Results

- Tree created for respectively data.
  - Consider occurrence for each data
  - Big hit on compression, especially for smaller files
- Tree determined
  - On basis of a statistical analysis of data
- Data transmissions is movement of bits
- Once receiver get a tree, then trace all arriving bit stream
- Going to left node: indicate 0
- Going to right node: indicate 1

Char	Code
E	0000
i	0001
y	0010
l	0011
k	0100
.	0101
space	011
e	10
r	1100
s	1101
n	1110
a	1111

Comparison of common algorithms with Huffman Coding

Algorithm	Huffman Coding	Run Length Coding	Arithmetic Coding
Advantage	1.Uniquely decodable code 2.Produce optimal compressed code 3.Lossless 4.Smallest average cod word size.	1.Easy to implement 2.Quick executable file 3.Lossless	1.More frequently occurring with fewer bits compared with less 2.Decrease file size 3. Lossless.

Disadvantage	1.Relatively slow. 2.Dependent on the statistical model 3.Decoding is complex due to various code size	1.The compression ratio is low as compared to other algorithms	1.paid algorithm 2.Statistically method.
Uses	Used in MPEG & JPEG.	Used in BMP and PCX files.	Used infrequently occurring sequences of pixels

5. CONCLUSION

Compression is a method in which the size of data is reduced while transmitting information through transmitting media as well as increase capacity of storing data by reducing compressed data. By this analysis, it helps to implement compressed software for compressing data with lossless data as well as reduce time complexity, space complexity and efficiently to implement by using the Huffman algorithm method.

REFERENCES

- [1] A.B. Watson, "Image compression using the DCT", Mathematical journal, 1995, pp 81-88.
- [2] <http://www.rz.go.dlr.de:8081/info/faqs/graphics/jpeg1.html>.
- [3] S. Annadurai and R. Shanmugalakshmi, "Fundamental of digital image processing", Pearson education, ISBN 81-7758-479-0.
- [4] J. H. Pujar and L. M. Kadlaskar, "A new Lossless method of image compression and decompression using Huffman coding technique", Journal of theoretical and applied information technology.
- [5] Klein, S.T., and Wiseman, Y.: 'Parallel Huffman Decoding with Applications to JPEG Files'. The Computer Journal, 46(5), British Computer Society,2003.
- [6] Kao, Ch., H, and Hwang, R. J.: 'Information Hiding in Lossy Compression Gray Scale Image', Tamkang Journal of Science and Engineering, Vol. 8, No 2, 2005, pp. 99- 108
- [7] Redmill, D. W. and Bull, D. R.: 'Error Resilient Arithmetic Coding of Still Images'. Image Communications Group, Centre for

Communications Research, University of Bristol,  
Bristol.

- [8] Xrysovalantis Kavousianos, Emmanouil Kalligeros, and Dimitris Nikolos” Optimal Selective Huffman Coding for Test-Data Compression “IEEE TRANSACTIONS ON COMPUTERS, VOL. 56, NO. 8, AUGUST 2007
- [9] Blelloch, G.: 'Introduction to Data Compression'. Carnegie Mellon University, September 2010.
- [10] K. Sayood, “Introduction to data compression”, 2nd and 3rd edition Academic press Morgain Kaufmann Publishers, 2000, 2006.