ISSN- 2394-5125 VOL 7, ISSUE 02, 2020

A HUFFMAN CODING: A WAY TO THE EFFICIENT UTILIZATION OF RESOURCES THROUGH DATA COMPRESSION

Naveen Bansal,
Assistant Professor, DEPARTMENT OF COMPUTER APPLICATIONS
CSB, Jhanjeri, Mohali, nbansal907@gmail.com

Preety Sharma,
Assistant Professor, DEPARTMENT OF COMPUTER APPLICATIONS
CSB, Jhanjeri, Mohali, preetysbrn12@gmail.com

Arshdeep,
Assistant Professor, UNIVERSITY INSTITUTE OF ENGINEERING
Chandigarh University, Gharaun, Mohali, arsh.grover1992@gmail.com

Dr. Jatinder Kaur
Assistant Professor, DEPARTMENT OF ELECTRONICS &
COMMUNICATION ENGINEERING CEC, Jhanjeri, Mohali,
jatinder_kor@yahoo.co.in

Abstract
Compression is a way of efficient utilization of bandwidth. Compression will efficiently reduce the total number of bits required to represent certain information. It also removes redundancy. In actual compression reduce the storage requirements. The various compression areas are data compression, image compression, video compression and audio compression. In this paper we will present the simulation of Huffman Coding in MATLAB. The first section of this paper presents the general introduction related to data compression and its various techniques. Then the next section includes the theory of Huffman coding. The subsequent section includes the problem statement. After that last section presents the simulation results.

Keywords: Compression, Data Compression, Huffman coding, Redundancy.

1. Introduction

Compression is a technique in which some bits of original data are eliminated but it doesn't modify the actual meaning of message and our visual system can't recognize this change. Data Compression consists of taking a stream of symbols as an input and converts those symbols into coded form as shown in figure 1.

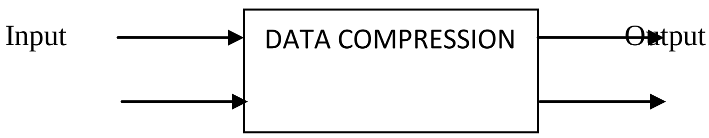


Figure 1. Basic Data Compression block diagram

The resulting stream of codes shows the effectiveness of compression. If the resulting stream is smaller than the original symbols it means the Compression is effective. Data Compression techniques helps to compress the data files which results in the form of improvement in efficiency and reduction in storage requirements. Data Compression is related with the concept of information theory because of its concern with redundancy. Data compression is a kind of lossless compression. Lossless compression is that in which the output is a duplicate copy of the input. In the lossless compression we just compress the data during transmission but the actual number of bits is not eliminated from it. Data compression is mainly used for file transfer etc. in which if we eliminate some information then the actual mean of input is changed. In contrast to lossless compression, lossy compression is the type of compression in which the output is differs from input. In this compression elimination of few bits is tolerable, since it can't change the actual mean of input. Image compression comes under lossy compression. In image compression if we eliminate few pixels of an image, our visual system can't identify this difference.

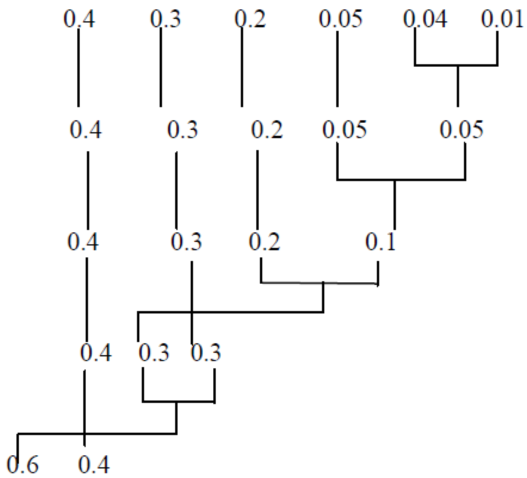
2. Theory of Huffman Coding

Huffman coding comes under Data Compression category. In the Huffman Coding the way to compression is providing the codeword with shorter wavelength to those symbols that are having higher probability of occurrence while symbols with lower probability of occurrence will have longer codeword lengths. Which results in the form of reduction in the average codeword length per symbol and this results to a smaller output data size. Average code length of Huffman code is greater than Shannon-fano coding. But it is more efficient than Shannon-fano coding because it will result in an optimum code when the code construction is efficient.

3. Problem formulation

In this paper we consider two cases in the first case we take six symbols and in the second case we consider ten symbols.

Case 1: Number of symbols=6



Case 2: Number of symbols=10

The ten symbols are sorted in descending order then follow the same procedure as shown for above case. We also may generate these symbols randomly by providing a specific range.

Table1. Samples and Code length table

No. of Samples	Code length
0.4	1
0.3	01
0.2	000
0.05	0011
0.04	00100
0.01	00101

L(X)= 0.4*1+0.3*2+0.2*3+0.05*4+0.04*5+0.01=2.05bits/sec.

ή(x) =

Table2. Formulas for calculating Code length, actual length and Entropy in Huffman coding

Code length	Actual length	Entropy
$L=\sum P_i*L_i$	$H= -\sum P_i*\log(P_i)$	$\eta=\frac{H}{L}$

4. Simulation model & Algorithm

- The first step related with Huffman coding algorithm is to arrange all the input symbols in the descending order on the basis of their probabilities of occurrence.

ISSN- 2394-5125 VOL 7, ISSUE 02, 2020

- Next step is based on the addition of two least probabilities symbols to generate a new symbol with probability of the sum of two combined symbols. This new symbol will have a probability that is sum of the probabilities of the two symbols. We will assign a '0' to the left of the new symbol and a '1' to the right of the symbol.
- Repeat step 1. If there are many symbols with the same probability then the newly created symbol will be placed as high as possible in the list. This procedure is continued until a binary tree containing all symbols has been generated.

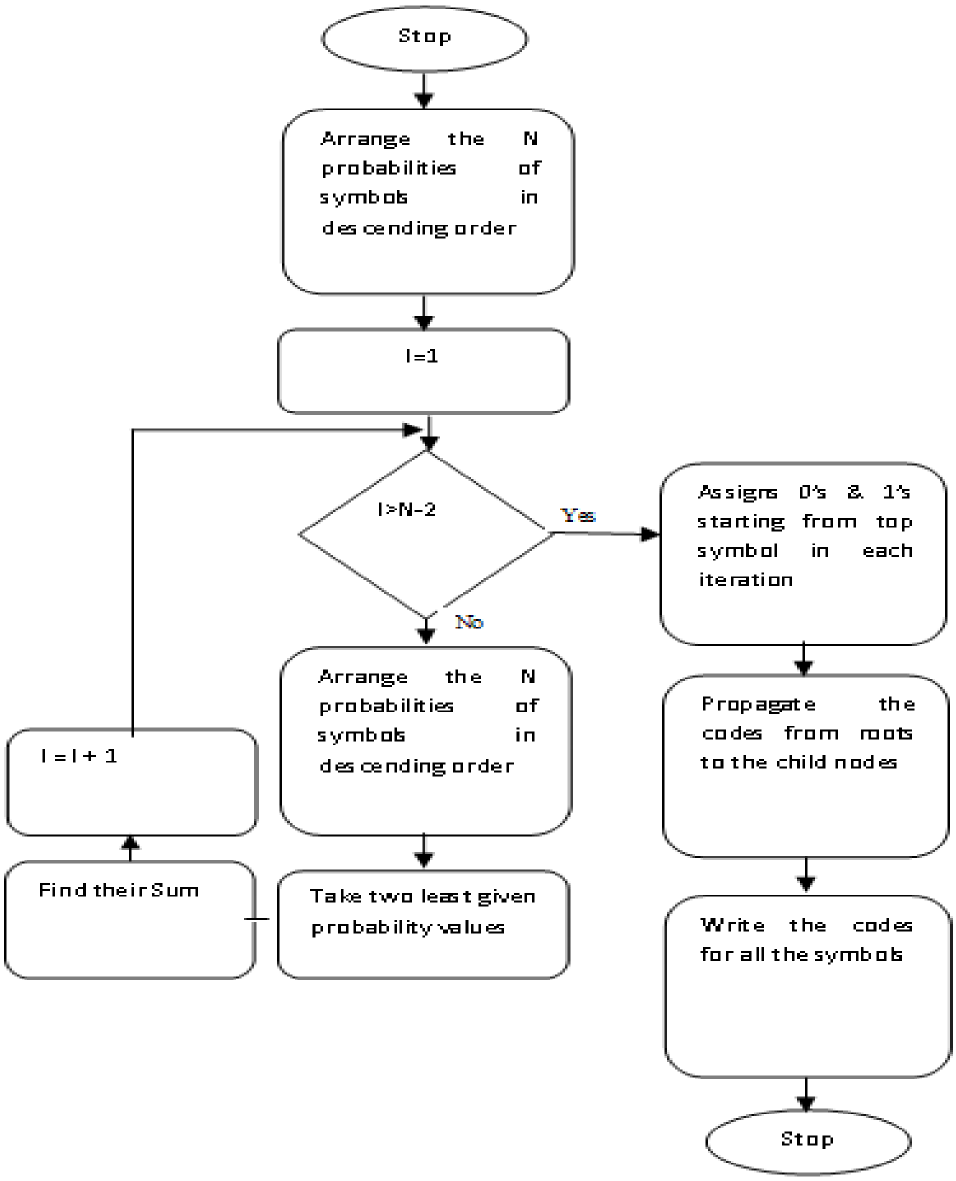
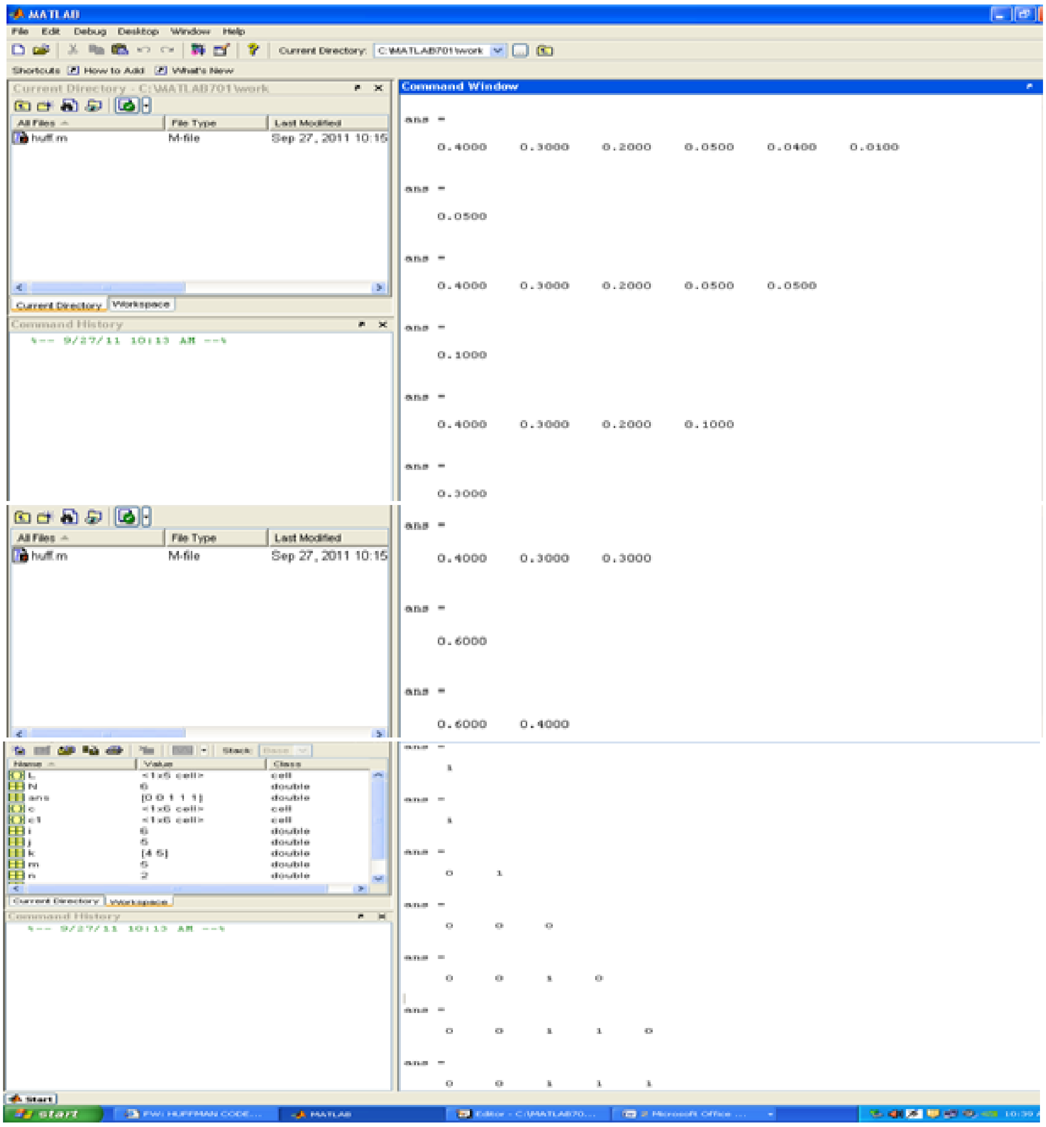


Figure2: Flow chart of Huffman coding

ISSN- 2394-5125 VOL 7, ISSUE 02, 2020

5. Simulation Results



6. Conclusion

In this paper we concluded that in the data compression by using Huffman coding, we require the less number of bits for the transmission of that symbols, having high probability of occurrence and vice versa. The code efficiency may be relatively poor in the region of the transition. If the probability at the transition is close to an optimum probability for the previous extension, the performance may deteriorate when moving to the higher extension.

References

[1] Dudgeon , D.E and R.M . Mersereau, Multidimensional digital signal processing, 1984, Englewood Cliffs, New jersey: Prentice –Hall.

[2] Russ, J.C., The image processing Handbook. Second ed. 1995, Boca Raton, Florida: CRC Press

[3] Castleman, K.R., Digital Image Processing, Second ed. 1996, Englewood Cliffs, New Jersey: Prentice-Hall.

[4] Giardina, C.R. And E.R.Dougherty, Morphological Methods in Image and Signal Processing, 1988, Englewood Cliffs, New Jersey: Prentice-Hall, pp. 321

[5] The Data Compression book, Second edition, Mark Nelson and Jean-Loup Gailly.

[6] www.real.com

[7] R.G. Gallager, "Variations on a theme by Huffman", IEEE Trans. Inform.. Theory, Vol. IT-24, no 6, pp 668-674, Nov. 1978.