THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22- Issue-12-November-2019

Compact Improvement in Proficiency of Huffman Coding

Dr. Vijay Kumar Garg
Associate Professor, CSE
Lovely Professional University
Jalandhar, Punjab
vijay.garg@lpu.co.in

Abstract- Compaction helps to remove duplicity in the representation of data which further reduces the need of storage in it. The task of compaction is to free up valuable space on a web server typically by encoding techniques. A Huffman code is a specific type of best prefix code that is often used to compress lossless data. In this paper, many other techniques adopted by various researchers are described which help in compacting redundant data and improve proficiency on Huffman coding. Some of the suggested techniques like arithmetic coding, minutia score matching; Hybrid quantized technique and JPEG image coding etc. provide higher compaction than earlier Huffman coding.

Keywords- Huffman Algorithm, Compaction, Redundant, Discrepancy

I. INTRODUCTION

Huffman coding is a lossless data compression algorithm. Information is encoded with bits-1's and 0's in computer science. Bit strings encode the information that will allow the computer to execute the command. In this, variable-length codes are allocated to store characters and the length of assigned codes are depends upon probabilities of similar characters. The most recurring characters get the shortest code and less recurring characters get highest code. Suppose we want to allocate 26 different codes to English alphabets and want to allocate an English book (only letters) then with the help of the Huffman code, we will need less memory if we assign small length codes to most recurring characters.

1.1 Background

In 1951, David Huffman and classmates were given the option of a writing assignment or a final study in information theory at the MIT. The subject of finding the most effective binary code was given by Professor Robert M. Fano. Huffman, unable to prove that any code was the most capable, was about to give up and start the final analysis when he came across the idea of using a frequency based binary tree and quickly proved to be the most capable of this process. In doing

ISSN:0971-1260
Vol-22- Issue-12-November-2019

so, the student exceeds his instructor, who collaborated with Claude Shannon, the founder of information theory, to create an equivalent code. The succeeding is expected from this article:- ***Condense and Compact storage space of 0's and 1's.***

1.2 The construction of algorithm

We might have noticed that postal code and STD codes for crucial cities are generally smaller. This is very elementary concept in information theory. If the probability table is somewhat invalid, the Huffman algorithm will still provide a valid encoding, but the encoding value would be higher than it might have been if you had used a right probability table. But this is generally not an issue, because we create probability table based on actual text that is to encoded, so probability table will be "best" for text that we are used to encode.

1.3 Huffman Code Table

The following is the example of Huffman code table which is to be found according to the below demonstrated Fig. 1 in which fixed length and variable length code has been mentioned. It is also clearly seen from the Table 1 that character A has maximum variable length code than other characters.

Table 1
HUFFMAN CODE TABLE AS PER FIG. 1

Character	Probability	Fixed length code	Variable length code
A	42	000	10111
B	11	001	0111
C	10	010	011
D	8	011	01
E	9	100	1

ISSN:0971-1260
Vol-22- Issue-12-November-2019

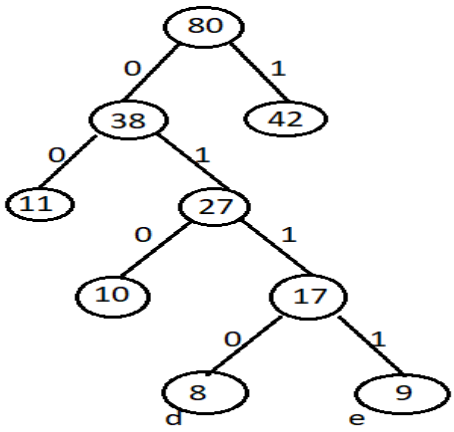

Fig. 1 Huffman Tree

1.4 Huffman Tree

Although Huffman coding with a given distribution of input probability is optimal for symbol-by-symbol coding, its optimality can sometimes be extensive accidentally. For example, compression is generally more capable of arithmetic coding and LZW coding. Each of these methodologies can merge an arbitrary value of signs with more competent programming and generally comply with the real input statistical data, both of which are useful if input probabilities are not known for sure or vary widely throughout the flow. Upgrade generally results from the interconnection of input symbols.

1.5 Discrepancy in Huffman coding

There is a great deal of disagreement with Huffman coding, some using a Huffman-like algorithm, and others preferring optimal prefix codes. A full list of Huffman articles coding the difference is given by "Code and Parse Trees for Lossless Source Encoding".

1.5.1 Lempel Ziv-Welch coding

Because of its simplicity, it is the most effective technique for general data compression and adaptability. It is dictionary based technique that is extensively used in the GIF image format.

1.5.2 FGK algorithm

In this algorithm, both encoder and decoder maintain effectively changing Huffman code trees.

ISSN:0971-1260
Vol-22- Issue-12-November-2019

The encoder sends the code word in the current tree for that symbol for each symbol and then updates the tree.

1.5.3 AAC audio coding

It is a technique used for compressing and encoding scheme digitised audio files. The sound quality of AAC audio is very effective.

1.5.4 ABC (artificial bee colony) algorithm

It is a maximization *t*echnique that affects the foraging behaviour of honey bees, and has been successfully applied to various practical problems.

1.5.5 DNA coding

DNA coding sequences are distributed through long regions of DNA called introns that do not have any function. Sometimes, recognized as extrons also in coding of DNA.

1.5.6 Minutia score matching

It is widely used for fingerprint recognition and can be categorized as ridge ending and ridge division.

1.5.7 Arithmetic coding

It is used in both lossy and lossless data compaction. It encodes entire word into a single number. It considers the current information as a range, prescribed by two numbers.

1.5.8 DWT (discrete wavelet transform) coding

It is used to reduce duplicity in video data without effecting visual quality. It captures both frequency and location information.

1.6 Applications

Although, due to its easiness, large pace and absence for patent barriers, many other methodologies offer good compression appearance than Huffman coding. Due to this reason, it is still widely used.

ISSN:0971-1260
Vol-22- Issue-12-November-2019

To some other compression technique, Huffman coding is often used as a "back-end" today. Devaluate and digital media instructions such as JPEG and MP3 get front-end model and quantization of Huffman coding.

The remaining part of this article is classified in the following sections. In Section 2, there is an impressive table that gives quick summary about source topic. Total 16 research papers have been included in this article. In Section 3, there is a pseudo-code of Huffman coding that gives informal representation of what algorithm must do. In Section 4, the article includes vital facts from background research or conclusion to explain results as needed. In Section 5, there must be references that specify the use of a source of information in order to ascertain source topic.

II. LITERATURE REVIEW

This article provides a brief review about papers that includes a concrete findings, as well as theoretical and methodological contributions to a specific topic. The following Table 2 gathers the required information from 16 useful research papers. It provides the author name, the techniques used by different authors, various limitations and disadvantages and complexity of a given researched papers. It makes comparison among research papers. This table relates, encapsulates, assess and simplify the literature. And according to situation, the literature review may figure out the sources and advise the reader on the most significant.

TABLE 2

COMPARISON OF VARIOUS HUFFMAN ALGORITHMS

Sr. No.	Authors	Techniques	Limitations	Advantages	Complexity
1.	Abdurazzag Ali Aburas et al. [1]	FGK algorithm (Faller, Gallagher, Knuth)	-	Easy error detection and error correction. Memory space is higher than binary tree.	**Time complexity-** O(n * log|E|) **Space complexity-** O(|E|)

ISSN:0971-1260
Vol-22- Issue-12-November-2019

2.	Reza Hashemian et al. [2]	CHT(condensed Huffman table)	End to end delay and time varying channel. So, network is a challenging task.	Reduces cost of Huffman code table. Space requirement is significantly reduced.	**Time complexity-** O(nlogn)
3.	Karl Spritting et al. [3]	DCT (discrete cosine transform) and uniform quantization.	Does not work well for continuous variables. Data can be overflowed.	Simple to solve. Powerful as iteration method	**Time complexity-** O(n)power 2
4.	Menachem Ailenberg et al. [4]	DNA coding (Data storage in DNA)	Retrieval time is very slow. Cost is high	Huge amount of data can be stored. No electricity is required.	-
5.	Abdurazzag Ali Aburas et al. [5]	Minutia score matching	-	Provides reliability, simplicity. Requires only Image for matching fingerprint patterns	**Time complexity and space complexity-** O(n)
6.	Aiswarya S, Dr. S. Veni et al. [6]	EZW(embedded Zero tree Wavelet encoder) technique	Files that do not contain any repetitive data at all cannot be compressed	It is very simple to implement. It is loss less compression technique.	**Time complexity-** O(n)

ISSN:0971-1260
Vol-22- Issue-12-November-2019

7.	Kshitija Pol et al. [7]	DWT(discrete wavelet transform) technique	Provides redundancy And less efficient	Provides time and frequency information	**Time complexity-** O(4N^2LOG2N)
8.	Samuel L. Braunstein et al. [8]	Storage of Quantum signals	Hard to encoding and decoding without disrupting the signals of the quantum.	Encoding and decoding can be done Code word of variable length are used to find out the efficiency	-
9.	Natalia TATAR Stefan HOLBAN et al. [9]	ABC (artificial bee colony) algorithm	Sometimes accuracy of optimal value cannot meet the requirement	Few control parameters Simplicity and robustness	**Time complexity-** T(n)=c*n
10.	Cung Nguyen et al. [10]	JPEG image coding	It is lossy compression	Improves performance in terms of error and image perception	**Time complexity-** O(n)z
11.	Asadollah Shahbahrami et al. [11]	Arithmetic coding	It is paid algorithm Statistical technique	Lossless compression. Reduce file size dramatically.	**Time complexity-** **O(|E|+n)**
12.	Milan R. Dincic et al. [12]	Hybrid quantizer technique	Frequency response of the system cannot be drawn or estimated	It gives optimum results for larger size of images.	**Time complexity-** O(R- +(n)\ Ax
13.	Kussy Nugamesh	Discrete Wavelet	Provides redundancy	Provides time and frequency	**Time complexity-**

ISSN:0971-1260
Vol-22- Issue-12-November-2019

	Mutter et al. [13]	Transform technique	And less efficient	information	O(4N^2LOG2N)
14.	Bayadir Abbas AL-Himyari et al. [14]	Run length encoding technique	Compression Ratio is less than other algorithms.	Easy implementation Faster execution	**Time complexity-** O(m*n)
15.	C.Saravanan, M.Surender et al. [15]	Lempel Ziv coding technique	The amount of storage is infinite	Dictionary based technique	**Time complexity-** O(n)
16.	Jie Zhu,Rang-Ding Wang et al. [16]	AAC audio coding technique	-	Provides good imperceptibility	**Complexity-** It is an extension of low complexity AAC optimized for low-bitrate applications

III. PSEUDO-CODE

The following are the certain steps that to be followed to create a Huffman tree and the same Huffman tree can be used to generate the Huffman table as shown in Table 1 and Fig 1.

Input: Characters and their Occurrences

(c1, O[c1]), c2, O [c2])… c1, c2… O [cn])

Output: returns the Huffman Tree

Make priority queue Q using c1, c2… cn;

n=|C|

Q=C

If z. left is equal to x

Then EXTRACT-LEAST (Q)

Else

If z. right is equal to y

ISSN:0971-1260
Vol-22- Issue-12-November-2019

Then EXTRACT-LEAST (Q)

Print z.Occur= x.Occur + y.Occur

Insert (Q, z)

Return

EXTRACT-LEAST (Q)

IV. CONCLUSION

To conclude, the whole paper encapsulates to provide the conclusion that tells What Huffman coding is used for? What are the limitations and advantages of Huffman coding? What are the findings for Huffman coding? What are the requirements to provide discrepancy in Huffman coding? What are the alternative solutions to provide better proficiency?

The various Huffman algorithms were discussed in this paper on the basis of the complexity, advantages and limitations achieved by the adopted method. It was found from the literary table that DNA coding and run-length coding algorithm is very slow and moreover it has the high cost to implement in terms of the complexity. DWT algorithm is also less efficient as it has very high time complexity while LZW, Minutia score matching, EZW algorithms are best in terms of their complexities.

V. REFERENCES

[1] Abdurazzag Ali Aburas and Salem MA Rehiel, "Fingerprint patterns recognition system using Huffman coding", IJCSNS,January,2009.

[2] Reza Hashemian,"Condensed table of Huffman coding, a new approach to efficient decoding", IEEE transactions on communication, March, 2004.

[3] Karl Spritting, John Husoy and Suen Ole Aase," Improved Huffman coding using Recursive Splitting" , Department of electrical and computer engineering,January,1999.

[4] Menachem Ailenberg,Ori D. Rotstein "An Improved Huffman coding method for achieving text, images and characters in DNA", Department of Swegery, university of Toronto,Li Ka Shing knowledge institute,September,2009.

[5] Abdurazzag Ali Aburas,Salem MA Rehiel" Fingerprint patterns recognition system using Huffman coding", Proceedings of the world congress of engineering, vol 3, july,2008.

ISSN:0971-1260
Vol-22- Issue-12-November-2019

[6] Aiswarya S, Dr. S. Veni, "Hardware implementation of EZW based image compression using Huffman coding" , international journal of engineering science and innovative technology, Volume 2,issue 5, September,2013.

[7] Kshitija Pol," Image steganography based on DWT using Huffman encoding with arithmetic coding",Computer Science and Engineering Department,Gurgaon,Volume III,October,2014.

[8] Samuel L. Braunstein, Christopher A. Fuchs, Daniel Gottesman, "A Quantum Analog of Huffman Coding ",IEEE transactions of Information Theory, Volume- 46 ,4 july,2000.

[9] Natalia TATAR Stefan HOLBAN, "A bio inspired alternative to Huffman coding", 11th international conference on DEVELOPMENT AND APPLICATION SYSTEMS, Romania, May 17-19, 2012.

[10] Cung Nguyen, "Fault tolerant Huffman coding for JPEG image coding system", Department of Computer Science and Engineering", University of California, CA 95616.

[11] Asadollah Shahbahrami,Ramin Bahrampour,Mobin Sabbaghi Rostami, Ostafa Ayoubi Mobarhan, "Evaluation of Huffman and arithmetic algorithms for multimedia compression standards", Department of Engineering", University of Guilan,Rasht,Iran, Oct 25,2016.

[12] Milan R. Dincic, Peric Zoran "Design of Quantizers with Huffman coding for Laplacian Source", ISSN 1392-1215 Electronic and Electrical Engineering", 2010,No.10(106).

[13] Kussy Nugamesh Mutter, Zubir Mat Jafri, Azlan Bin Abdul Aziz, "Hybrid Hopfiel neural network, discrete wavelet transform and Huffman coding for image recognition",IJCSNS International journal of computer science and network security, VOL 9,No.6,June,2009.

[14] Bayadir Abbas AL-Himyari, "Role of run length encoding on increasing Huffman effect in text compression", Journal of Kerbala University,Vol.6 No. 2 Scientific.2008.

[15] Bayadir Abbas AL-Himyari, "Enhancing efficiency of Huffman coding using Lempel Ziv Coding for Image Compression", International journal of soft computing and engineering(IJSCE)ISSN:2231-2307",Volume-2,Issue-January,2013.

[16] Jie Zhu,Rang-Ding Wang,Juan Li and Di-Qun Yan, "A Huffman coding section-based Steganography for AAC Audio", Information technology", Volume 10 (10):1983-1988,2011.