

# A Study on Various Encoding and Compression Techniques

<sup>1</sup>Devi. A, <sup>2</sup>Abhay Kumar Srivastav

<sup>1</sup>Asst.Professor, <sup>2</sup>Asst.professor

School of Computer Science and Applications, REVA University, Bangalore

## Article Info

Volume 83

Page Number: 3477-3479

Publication Issue:

May-June 2020

## Article History

Article Received: 19 August 2019

Revised: 27 November 2019

Accepted: 29 January 2019

Publication: 12 May 2020

## Abstract

This paper exhibits survey of various data compression and image compression techniques. It is dealing about existed encoding and compression techniques combined together for literature survey. It is also been discussed about the transmission speed and phasing the various issues of using different compression techniques to reduce the size of plain text. This survey discusses about several compression techniques with different survey research paper and it is been provided knowledge of compression and encoding techniques for future work.

**Keyword:** Arithmetic encoding, Huffman Algorithm, BWT-mtf, Lempel-Ziv algorithm.

## 1. Introduction

Encoding is used to change the original form of plaintext into some other form. The bit rate reduction algorithm to minimize the bits which are used to compare the original information and also used to transfer minimum bits over the network to save the storage capacity and the bandwidth. The transmission of data is so easy and efficient. Lossy and lossless are the two different types of compression techniques involved in compression of data, namely, Shannon-Fano, run-length, Huffman coding, Arithmetic coding, Lempel-Ziv algorithm.

Compression of data is a technique to change the original form of message as input into the different cipher text in reduced size. Compression of Lossless would never be removed any important content in the original message. But in case of lossy compression, it may remove some information from the original data. Kumar et al experimented a method called "Lossy Data Compression Logarithm". From this, the result has been proved that the method can be generated the ratio of compression up to 1:60 in many cases. In 2013, Kavinder [3] proved that the ratio of compression is increased due to combine the Huffman coding and discrete cosine of transform by using vector quantization.

In Huffman coding, the length of code is depending on the concurrence of characters. Smallest code represents often continuous characters and less continuous characters getting the biggest code. Arithmetic coding is nothing but entropy encoding and it is varied from other entropy

encoding, namely, Huffman coding. In 1970, the adaptive Huffman coding for data compression took its advanced form. It is almost same as the Huffman coding but not exactly the same.

Due to overcome the disadvantages in adaptive Huffman coding, the Arithmetic coding is being developed to modify and to replace each bit with a codeword. Huffman coding provides the codeword, particularly to each symbol but in the arithmetic coding codeword is provided to the whole string. Shannon-Fano coding is an important compression technique, which is the first method for compression, developed by Claude Shannon and RM Fano at MIT & Bell Lab. It is based on the probabilities of symbol in the message. Whenever the probability of symbol is larger than the codes they vary in length.

## 2. Literature Survey

The technique of compression of data based on dictionaries may not be encoded single symbol into different length of bit strings. It can encode variable-length symbol strings as single tokens [3], the LZW algorithm is also used in Dictionary-based data compression. Static and dynamic are two types of dictionary-based algorithm. Fixed size of dictionary is used in static for encoding and decoding process. The different size of dictionary is used for decoding and encoding process.

Vector quantization is a lossy data compression due to some number of bits are loss after decoding the

message and it is based on principle of block coding [2]. It is mainly designed for multi-dimensional data. R.S. Brar, B. Singh, et al [3], proposed "A study of various compression techniques and the algorithm for bit reduction to compress text data". This paper surveyed different techniques in lossy and lossless compression. Based on these methods, the authors have proposed a bit reduction algorithm for data compression using the number theory method and which is free of time complexity.

S.Y. Chaudhary, J. Joshi and M. Porwal. Jain, suggested "Methods of data compression for lossless data and algorithm compression". It has been proposed that the various lossless data compression techniques are compared with arithmetic coding based on their performance. The author has been shown to have a higher compression ratio in arithmetic coding than the Huffman encoding, and has also shown to have decreased channel bandwidth. Iwan

Handoyo Putro, Petrus Santoso [4] have been implemented a technique based on arithmetic encoding in data compression. The issues of length of message and sending more than 160 characters in the message have been discussed.

Asmita A. Bardekar et al [5], showed in this paper, the lossy and lossless data compression method but focuses only on the vector quantization technique for image compression. In vector quantization the training set of input as the codeword and set of codeword's are the codebook. In this paper, image as the input vector called codeword and block of image is the codebook. Three steps that follow vector quantization, first one is codebook design second one is image encoding and last step is decoding the image.

### 3. Text Compression

Anita Singhrova, Dr. Nupur Prakash [6] have proposed the analysis of various security protocols in mobile in the field of mobile communication and the different encryption, authentication techniques required while transmitting data over mobile. Stephan Rein, Clemens Guhmann [7] have been experimented a technique compressed message is transferred through Mobiles and compressed message is too less complexity of arithmetic coding.

Abu Shamim Mohammad Arif [8] gave a "Bengali compressed messages" technique using the "masking and dictionary" principle and also, a static data compression on short messages. The technique implemented especially for the short Bengali messages. Vimal Kishore et al [5] relates speech compression to multimedia communication. It is a very important part of compression because today, the multimedia communication is very popular. This compression method converts the human speech to compress encoded format.

LZW [9], coding is implemented for decreasing the consumption of energy and increasing the lifetime of network. The networks are one-third of

initial file size. A "DC" lossless technique called "Sequential-Lossless Entropy Compression" is proposed to provide robustness in "WSN" [10]. A lossless data compression algorithm [11], has been proposed for compressing the data from pulse digitizing electronics and used for compressing the original message from "time projection chamber", "A large Ion Collider experiment" and also the three steps namely "delta calculation" with "vector quantization and Huffman encoding".

Sundararajan et al. [12] propose an instant coding method that uses LWCs with "codeword slimming" scheme. This system transmits words rather than the traditional "word-serial bit-parallel" model, in the "bit-serial word-parallel" model. In this method, Words are transmitted simultaneously over bus W-lines and thus each line can be considered independently for reducing bits. It uses a displacement-coding method where coding is performed on the "transitions between consecutive bits" sent over the bus line instead of on the original bits of the data. So certain bits of displacement are coded using LWCs to reduce bits.

### 4. Data Compression

Behnam et al. [13] proposed a method that the best in both binary and Desc encoding [20]. This method dynamically adjusts the encoding between Desc or binary encoding to boost the energy-delay product based on application behaviour of the device. This method calculates (1) the overall Desc encoding latency and (2) the average delay of each cache block that is transferred.

Therefore, (1) if a group of cache blocks' average latency is lower than the average binary encoding latency, therefore Desc encoding is used for all transfers as Desc will fulfil the latency deadline and energy budget. (2) If the "average latency" of Desc encoding is lower than the "full latency," the binary encoding scheme shall be used for all transfers, since the high latency of Desc cannot be tolerated. (3) otherwise such blocks are transferred using Desc and others use binary encoding based on considerations of latency / energy. This method can take decisions after a process (i.e., similar number of successive requests) or a burst (i.e., high-bandwidth request stream).

Additionally, either LS or TS may be used for binary encoding. This method increases performance and energy efficiency while at the same time incurring little overhead area. The drawback of the Behnam et al. method [13] is that the approach significantly increases the complexity of the design.

Wang present a technique of encoding that works by grouping related data into clusters. They use the algorithm k-majority which is a modified version of the clustering algorithm k-means. The similarity between bit-strings is determined in the k-majority algorithm using Hamming (and not Euclidean) distance and cluster centres are recalculated using majority vote instead of arithmetic mean.

The Centres are randomly initialized and they concluded that value-similarity-based techniques are primarily useful for situations where adjacent values have the same or similar values, such as array address pointers, picture neighbouring pixels, nearby integer values, and so on.

Due to the peculiarity of floating-point representation, however, performing XOR between nearby data values cannot reduce the count of values '1'. Many technologies such as machine-learning and image processing are inherently error-tolerant. Additionally, ambient noise and error during the conversion between analog and digital environments contribute to inherent sensor data errors. Consequently, if the errors during transmission are negligible compared to other sources of error or can be accepted by the application, a loss encoding scheme may be used to conserve energy with low quality losses.

## 5. Conclusion

It is concluded that from the existing literature, an LZW coding is combined with Arithmetic coding of double compression with cryptosystem. So that, a Hybrid compression is being proposed for future research of encoding of original message to create an intermediate message and in performing encryption, RSA encryption or Elgamal algorithm to enhance the security.

## References

- [1] Mark Nelson, Sean-Loup Gailly, "The data compression book", second edition, "IDG books Worldwide, Inc."
- [2] <http://www.datacompression.com/vq.html>
- [3] R.S. Brar and B. Singh, "A survey on different compression techniques and bit reduction algorithm for compression of text data.", "IJARCSE", Vol 3, Issue 3, March 2013.
- [4] Iwan Handoyo Putro<sup>1</sup>, Petrus Santoso, Maya Basoeki, "A Short Text Compression Scheme based on Arithmetic Coding", 2007.
- [5] Asmita A. Bardekar et al., "A Review on LBG Algorithm for Image Compression", *International Journal of Computer Science and Information Technologies*, Vol2, 2011.
- [6] Anita Singhrova, Dr. Nupur Prakash, "Performance Analysis of Mobile Security Protocols: Encryption and Authentication",
- [7] Stephan Rein, Clemens Guhmann, Frank H.P. Fitzek, *Low-Complexity Compression of Short Messages*, 2006.
- [8] Abu Shamim Mohammad Arif, Asif Mahamud, Rashedul, "An Enhanced Static Data Compression Scheme of Bengali Short Message", *IJCSIS*, Vol. 4, No.1&2, August 2009.
- [9] B. Ruxanayamin, Dr. B. Ananda Krishna, "Minimization of power consumption in MANETs", *IJERT*, PP.31-35, 2013.
- [10] Liang Y, Li Y, "An efficient and robust data compression algorithm in wireless sensor networks". *IEEE Commun. Lett.* 18, 439-442.
- [11] Patauner C, Marchioro A, Bonacini S, Rehman A.U, Pribyl W, "A lossless data compression system for a real-time application in HEP data acquisition". *IEEE Trans. Nucl. Sci.* 58, 1738-1744, 2011.
- [12] V. Sundararajan and K. K. Parhi, "Reducing bus transition activity by limited weight coding with codeword slimming," in *Great Lakes symposium on VLSI*, pp. 13-16. 2000.
- [13] P. Behnam, N. Sedaghati, and M. N. Bojnordi, "Adaptive Time-based Encoding for Energy-Efficient Large Cache Architectures," *IWEES, ACM*, 2017.