International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.07, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

A REVIEW OF DATA COMPRESSION TECHNIQUES

Ibrahim M. Omar[1], S.N. Prasad[2]

[1,2] *School of Electronics and Communication Engineering, REVA University, (India)*

ABSTRACT

In this paper a review kind of data compression techniques is presented. Data compression is used widely by various sectors because using compression can save some storage. Data compression can also speed up the transmission time of data from a source to a destination. Performing compression requires to a method of data compression that can be used to compress the data. Data cannot be only text data but it can be video or images. Image Compression is a solution of the problems associated with transmission of digital images and the storage of large amount of information for digital Image and compression is the only way to deal with this problem. Data compression techniques are divided into two named lossless compression and lossy compression. Some of lossless compression techniques are Run-Length Encoding, Huffman, Arithmetic coding, and Lempel Ziv welch. Each method has the ability to perform a different kind of compression. The paper explains which method is well used in doing data compression, the output generated can be known as the compression file which its size becomes smaller than the original file.

Keywords: *Data Compression, compression techniques, lossless compression, image compression, run-length encoding, Huffman, Shannon Fano.*

1-INTRODUCTION

The proliferation of computer and other electronic systems/devices into various sectors are continuously generating data at unprecedented rate. This results in critical management and storage of data. The unprecedented rate of data generation calls for a better, low-cost memory solutions to store. Since, a large amount of data gets generated at every moment, the memory management and storage formats must be simple and fast. The more information being dealt with means the more it costs in term of storage and transmission costs. The capacity problem of storage media could be resolved to an extent by adopting to compression technology that would result in better use of the memory space.

To store terabytes of data, especially of the type human-readable text, it is beneficial to compress the data to gain significant savings. There is a large size that can hinder data transmission quickly and save on existing storage in the computer. To overcome the problem of information or data to be transmitted or transmitted can be done quickly than required a compression that can save storage and transmission of data to be done.

Compression is the process of converting a data set into a code to save the need for storage and transmission of data making it easier to transmit a data. Compression techniques are being rapidly developed to compress large data files such as images, there are some algorithms that perform the compression in different ways; some are lossless and some are lossy. Lossless keep the same information as the original image and in lossy some information loss during the compression.

2- NEED OF COMPRESSION

In general uncompressed data occupies large amount of memory in storage and needs more bandwidth from the transmission media to transmit, and it takes more time to transfer from a source to a destination. So if we want to transfer or store large amount of data then we have to compress it first for fast speed of transmission and to store in a less space. Hence compression is very essential for modern multimedia application.

3- Lossless Compression Techniques

In lossless compression algorithms, the original data can be recovered exactly from the compressed data. It is used for discrete data such as computer generated data, text and certain kinds of image and video information. It can achieve only a modest amount of compression of the data and hence it is not useful for sufficiently high compression ratios.

The advantage of this kind of algorithms is they maintain quality but the main disadvantage is they don't reduce the file size as much as lossy compression does.

There are some techniques of lossless compression:

- Run length encoding
- Huffman coding
- LZW coding
- Arithmetic coding

3.1- Run length encoding: Run length encoding is one of the simplest data compression algorithms. This compression technique is useful in case of repetitive data, when we have sequence of same intensity pixel or symbols then this sequence is replaced by shorter symbols and it is represented by a sequence. Runs is sequences in which the same data value occurs in many consecutive data elements are stored as a single data value and their count, rather than as the original run. The run length code for a gray scale image is represented by a sequence {Vi, Ri} where Vi is the intensity of pixel and Ri refers to the number of consecutive pixels with the intensity Vi.

10 10 10 10 12 12 12 12 12 15 15 15 20 20
(10,4)(12,5)(15,3)(20,2)

Figure: Run length coding

3.2- Huffman coding: Huffman coding is an entropy encoding algorithm used for lossless data compression.The method encodes symbols or characters with the help of a binary tree by combining the two

smallest frequencies to form a code tree. Huffman codes are based on the number of character frequencies that often appear. The larger the Huffman code frequency the less number of bits produced. Conversely, the fewer character appearances the more number of bits produced during a compression. The Huffman's algorithm is generating minimum redundancy codes compared to other compressing algorithms.

3.3- Lempel Ziv Welch: Lempel ZivWelch algorithm is a lossless compression algorithm and uses a dictionary. LZW compression will form a dictionary during the process of the compression takes place. Dictionary based coding can be static or dynamic. In the static dictionary coding, the dictionary is fixed when the encoding and decoding processes. In dynamic dictionary coding, the dictionary is updated on fly. In encoding process the algorithm goes over the stream of information, then coding it; if a string is never smaller than the longest word in the dictionary then it transmits. In decoding process, the algorithm rebuilds the dictionary in the opposite direction way, it thus does not need to be stored.

3.4- Arithmetic coding: The aim of Arithmetic Coding is to define a method that provides code words with an ideal length. Like for every other entropy coder, it is required to know the probability for the appearance of individual symbols. Arithmetic Coding is the most efficient method to code symbols according to the probability of their occurrence. The average code length is very close to the possible minimum given by information theory. Arithmetic Coding assigns an interval to each symbol whose size reflects the probability for the appearance of this symbol. The code word of a symbol is an arbitrary rational number belonging to the corresponding interval.

4- Lossly Compression Techniques

Lossy compression algorithms as the name imply causes loss of some information after compression. The compressed data is similar to the original uncompressed data but not just like the previous as in the process of compression some information concerning to the original data has been lost. Lossy compression algorithms provide a higher compression ratio than lossless compression.

5- LITERATURE SURVEY

In [1] the paper discusses Arithmetic coding which provides an effective mechanism for removing redundancy in the encoding of data. The paper shows how arithmetic coding works and describes an efficient implementation that uses table lookup as a fast alternative to arithmetic operations. The reduced-precision arithmetichas a provably negligible effect on the amount of compression achieved. We can speed up the implementation further by the use of parallel processing. In this paper the role of probability models is discussed and how they provide probabilityinformation to the arithmetic coder. We conclude with perspectives on the comparative advantages and disadvantages of arithmetic coding.

In [2] the paper presents a modified scheme for Run length encoding (RLE). Run length encoding algorithm performs compression of input data based on sequences of identical values. RLE is having some limitations and

they have been highlighted and discussed in this paper. In RLE largest number of sequences may increase the number of bits to represents the length of each run, which may increase the size of memory stack which may results in performance degradation. Bit stuffing has been suggested in this paper. A new bit different from the original sequence is added in between reduces the repeat length, thereby with the same stack we can represent length as well. The technique is described using VHDL and is implemented on Saprtan3 FPGA.

In [3] in this paper, an improved dynamic bit reduction algorithm is developed to compress and decompress the text data based on lossless data compression approach. Various experiments have been conducted on different datasets such as Random, Alphanumeric, Numeral and Special Characters dataset. The results obtained by the proposed system are compared with the existing data compression techniques Bit Reduction and Huffman Coding using parameters such us Compression Ratio and saving percentage. From the results analysis, it is concluded that the proposed system shows very good compression results in terms of Compression Ratio and Saving Percentage as compared to the existing techniques for all the datasets that have been considered.

In [4] the paper discusses compression and transmission on a compound image. This journal uses the decomposition technique and a modified image is then compressed using the LZW algorithm. The performance measures the PSNR value and the MSE value (Mean Square Error) is calculated. When compared with using embedded compression the algorithm has higher PSNR value using LZW algorithm.

In [5] the paper presents vector quantization based image compression technique [5]. In this paper encoding of the difference map between the original image and compressed image is adjusted and after that it is restored in VQ compressed version. Result of this experiment shows that although this scheme needs to provide extra data, it can improve the quality of Vector quantized compressed images, and further to be adjusted according to the difference map between the lossy compression and the lossless compression.

In [6] the paper described a new test data compression technique based on Reversed Leading Bits Coding and Huffman Coding (RLBC-HC). RLBC-HC divides the test data into codeword segments with proper size, and achieves a high Huffinan coding efficiency consequently. The decoder hardware is very simply, and easy to be implemented in hardware. The performance of the proposed pattern generator has been demonstrated by experimental results and comparisons against other test data compression techniques.

In [7] the paper presents the approach of the lossless image compression using the novel concept of image folding [14]. This proposed method uses property of adjacent neighbor redundancy for the prediction. The method applies column folding followed by row folding iteratively on the image until the image size reduces to a smaller pre-defined value. Then the method is compared with the existing lossless image compression algorithms and the obtained result shows a comparative performance of various methods. Data folding method

is a simple technique for compression of images which provides good efficiency and offer lower computational complexity as compared to the SPIHT technique of lossless compression.

In [8] the paper presents a novel method for compressing the image named as five module method (FMM). In the method they convert each pixel value in 8x8 blocks into a multiple of 5 for each of RGB array [8], and then the value is divided by 5 to obtain new values which are known as bit length for each pixel and uses less storage space than the original values which is 8 bits. The paper shows the potential of the FMM based image compression techniques. The advantage of this method is, it provides high PSNR although it haslow CR [compression ratio]. The method is good for bi-level like black and white medical images where the pixel of the images is presented using one byte [8 bit].

In [9] the paper presents a technique that encryption and compression of the image is done first, the encryption is used for stream cipher to encrypt the image then a compression technique named SPIHT [10] is used for compressing the image. In this paper to provide good encryption stream cipher encryption is used. SPIHT compression results better compression ratio as well as the size of the larger images can be selected and can be decompressed with the least or no loss in the original image. Therefore confidential and high encryption and the best compression rate has been energized to provide better security.

In [10] the paper presents a technique that utilizes the reference points coding with threshold values for image compression. The paper gives the idea of an image compression method which can be used to perform bothlossy and lossless compression. A threshold value is associated in the compression process, by varying the threshold values, different compression ratios can be achieved and by setting the threshold value to zero then lossless compression can be performed. The quality of the decompressed image can be calculated during the process of the compression, when the threshold value of a parameter assumes positive values, then Lossy compression can be achieved.

In [11] the paper presents a technique for image compression that uses different embedded Wavelet based image coding with Huffman-encoder for further compression. In the paper they implemented the EZW and SPIHT algorithms with Huffman encoding [15] which uses different wavelet families for compression and then comparison for the PSNRs and bit rates of these families are made. These algorithms were performed on various images, and it is shown that the results have good quality and it also provides high compression ratio as compared to the previous existing lossless image compression techniques.

In [12] the paper presents a novel hybrid image compression technique. This technique inherits the properties of localizing the global spatial and frequency correlation from wavelets and classification and function approximation tasks from modified forward-only counter propagation neural network (MFOCPN) for image compression. In this scheme several tests are used to investigate the usefulness of the proposed scheme. In this

paper, they explore the use of MFO-CPN [12] networks to predict wavelet coefficients for image compression. In this method, they combined the classical wavelet based method with MFO-CPN. The performance of the proposed network is tested for three discrete wavelet transform functions. In this they analysis that Haar wavelet results in higher compression ratio but the quality of the reconstructed image is not good. On the other hand db6 with the same number of wavelet coefficients leads to higher compression ratio with good quality. Overall they found that the application of db6 wavelet in image compression out performs other two.

In [13] the paper presents a new modified international data encryption algorithm to encrypt the full image in an efficient secure manner, and encryption after the original file will be segmented and converted to other image file. By using Huffman algorithm the segmented image files are merged and they merge the entire segmented image to compress into a single image. Finally they retrieve a fully decrypted image. Next they find an efficient way to transfer the encrypted images to multipath routing techniques. The above compressed image has been sent to the single pathway and now they enhanced with the multipath routing algorithm, finally they get an efficient transmission and reliable, efficient image.

6- CONCLUSION

Using the compression technique can reduce the number of file sizes.This paper presents various types of data compression techniques. There are basically two types of compression techniques. One is lossless compression and the other is lossy compression technique. Comparing the performance of compression technique is difficult unless identical data sets and performance measures are used, lossy compression techniques provides high compression ratio than lossless compression scheme. Lossy compression is used for more compression ratio and Lossless compression is used when the original image and reconstructed image are to be identical.

This review paper also gives the idea about various image types and performance parameter of image compression. Based on review of different types of images and its Compression algorithms we conclude that the compression algorithm are useful in their related areas and basically depends on the three factors i.e. quality of image, amount of compression and speed of compression.These are still a challenging task for the researchers and academicians. Some of these techniques are obtained good for certain applications like security technologies. After study of all these techniques it is found that lossless image compression techniques are most effective over the lossy compression techniques.

REFERENCES

[1] Paul G. Howard, Jeffrey Scott Vitter, “Arithmetic Coding for Data Compression”, Proceedings of the IEEE, Vol 82, No.6 June 1994.

[2]S. Sarika, S. Srilali, “Improved Run Length Encoding Scheme For Efficient Compression Data Rate”, Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013.

[3] Amandeep Singh Sidhu, MeenakshiGarg, "Research Paper on Text Data Compression Algorithm using Hybrid Approach" International Journal of Computer Science and Mobile Computing, Vol.3 Issue.12, December- 2014.

[4] G.R.Gnana King, C.SeldevChristoper, And N.Albert Singh, " Coumpound Image Compression Using Parallel Lempel Ziv-Welch Algorithm", IET Chennati Fourth International Conference on Sustainable Energy and Intelligent System,Chennati,December 12-14, 2013

[5] Jau-JiShen and Hsiu-Chuan Huang, An Adaptive Image Compression Method Based on Vector Quantization, IEEE, pp. 377381, 2010.

[6] HaoqiRen, " A data Comprssion Technique based on Resersed Leading Bits Coding and Huffman Coding", International Conference on Communication and Networking, China, 2015.

[7] Suresh Yerva, Smita Nair and Krishnan Kutty, Lossless Image Compression based on Data Folding,IEEE, pp. 999-1004, 2011.

[8] Firas A. Jassim and Hind E. Qassim, "Five Modulus Method for Image Compression" SIPIJ Vol.3, No.5, pp. 19-28, 2012.

[9] C. Rengarajaswamy and S. Imaculate Rosaline, SPIHT Compression of Encrypted Images, IEEE, pp. 336-341,2013.

[10] Yi-Fei Tan and Wooi-Nee Tan, Image Compression Technique Utilizing Reference Points Coding with Threshold Values,IEEE, pp. 74-77, 2012.

[11] S. Srikanth and SukadevMeher, Compression Efficiency for Combining Different Embedded Image Compression Techniques with Huffman Encoding, IEEE, pp. 816-820, 2013.

[12] AshutoshDwivedi, N Subhash Chandra Bose, AshiwaniKumar,‖ A Novel Hybrid Image Compression Technique: Wavelet-MFOCPN‖ pp. 492-495, 2012.

[13] S. Dharanidharan , S. B. Manoojkumaar and D. Senthilkumar,‖ Modified International Data Encryption Algorithm using in Image Compression Techniques,‖ IJESIT , pp. 186-191,2013.

[14] S. A. Mohamed; Dept. of Electr. Eng., Queen" S Univ., Kingston, Ont., Canada ; M. M. Fahmy, Image compression using VQBTC, IEEE Transactions on Communications.

[15] A.S. Ragab, Abdalla S.A. Mohmed, M.S. Hamid, "Efficiency of Analytical Transforms for Image Compression" 15th National Radio Science Conference, Feb.24-26, 1998.

[16] AkhandPratap Singh, Dr. Anjali Potnis, Abhineet Kumar, "A Review On Latest Techniques Of Image Compression" International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 07 | July-2016.

[17] LulukAnjarFitriya, Tito WaluyoPurboyo, AnggunmekaLuhurPrasasti, "A Review of Data Compression Techniques" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 19 2017.

[18] Ritu, Puneet Sharma, "A Review of Data Compression Technique" International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 4, Jul-Aug 2014.

[19] GauravVijayvargiya, Dr. Sanjay Silakari, Dr.RajeevPandey, "A Survey: Various Techniques of Image Compression" International Journal of Computer Science and Information Security, Vol. 11, No. 10, October 2013.

[20] PriyankaRaichand, Rinkle Rani Aggarwal, "A Short Survey Of Data Compression Techniques For Column Oriented Database" Journal of Global Research in Computer Science, Volume 4, No. 7, July 2013.