International Journal of Computer Science and Engineering Communications,
Vol.4, Issue.1, Page.1277-1280, (2016)
www.scientistlink.com

A Survey on Lossless and Lossy Data Compression Methods

K.A. Ramya[1], M.Pushpa[2]

[1] Research Scholar, [2] Assistant Professor,
Quaid-E-Millath College for Women (Autonomous), Chennai. India
[1]ramyanndr@gmail.com, [2] push_surya@yahoo.co.in

Abstract

Compression is built into a broad range of technologies like storage systems, databases, operating systems and software applications. It refers to the process of reducing the quantity of data used to represent the content without excessively reducing the quality of the original data. Their main purpose is to reduce the number of bits required to store and/or transmit digital media in a cost effective manner. There are number of data compression techniques used and they can be categorized as Lossless and Lossy compression methods. In this paper, we made an attempted to discuss about some of the general concepts of compression algorithm using Lossless and Lossy methods of compression.

Keywords: *Data compression, Lossless Compression, Lossy compression, RLE, LZW, Huffman coding and Transform coding, DCT and DWT.*

I INTRODUCTION

Now-a-days Databases are mandatory everywhere. Even when doing the shopping for dressing everyone is using internet. Database touches all aspects of our lives. Database that stores information about the users, personnel, official and contact information etc. And like anything else, storing information will require a space to be stored. When the database store space becomes as big, the server refuses to add new data to the database [1]. So the compression techniques are used. The data compression helps to compress the data inside a database, and it can help reduce the size of the database[2]. Lossless and Lossy Compression are two types of data compression. Lossless compression is used in text file, database tables and in medical image because of the law of regulations. Some of the main techniques are Run length Encoding, Arithmetic Encoding, Shannon-Fano, Lempel-Ziv-Welch, Huffman coding. Lossy compression algorithm is usually with the original data is not necessary after decompression. Some of the methods of Lossy Data Compression methods are Transform Coding, Discrete Cosine Transform, Discrete Wavelet Transform, Fractal Compression. In this paper we concentrate RLE, LZW, Huffman Coding and Transform Coding, DCT and DWT. In this paper, we made an attempt to discuss about these lossless and lossy algorithm.

II DATA COMPRESSION

Data Compression can also be made possible because most of the real world data is very redundant. It is basically defined as a technique that decrease the size of data by applying different methods that can be either be Lossless or Lossy[3]. It is an important applications in the areas of data transmission and data storage despite of the large capacity storage devices are available these days. Hence we need an efficient way to store and transmit different types of data such as text, image, audio and video to decrease execution time and memory size [4]. There are two types of data compression techniques:

- Lossless Data Compression
- Lossy Data Compression

A) LOSSLESS DATA COMPRESSION

Lossless compression means when the data is decompressed, the result is a bit-for-bit perfect match with the original one.

The name of lossless means no data is lost, the data is only saved more efficiently in its compressed state, but nothing of it is removed[5]. Lossless data compression methods may be categorized according to the type of data they are designed to compress. Compression algorithm algorithms are basically used for compression of text, images and sound.

Most of the Lossless compression programs use two different kinds of algorithms: one which generates a statistical model for the data of input and another which maps the input data to bit strings using this model in such a way that frequently encountered data will produce shorter output than less frequent data[6].

B) LOSSY DATA COMPRESSION

Lossy compression means that some of the data is lost when it is decompressed. Lossy compression bases on the assumption that the recent data files save more information than human beings can "perceive". Thus the irrelevant data can be removed [5]. Lossy image compression can be used in digital cameras, to increase storage capacities with minimal degradation of picture quality. Similarly, Digital Versatile Disks use the lossy MPEG-2 video codec for video compression. In this technique compression, methods of psychoacoustics are used to remove non-audible components of the signal. Compression of human speech is often performed with even more efficient techniques, so that "speech compression" or "voice coding" is sometimes differentiated as a separate discipline from "audiocompression"[3].

Table 1. Comparison between Lossy and Lossless compression

FACTORS	DATA COMPRESSION	
	LOSSLESS COMPRESSION	LOSSY COMPRESSION
Definition	Lossless compression is a class of data compression algorithms that allow the original data to be perfectly reconstructed from the compressed data[7].	Lossy compression is the class of data encoding methods that uses inexact approximations to represent the content. These techniques are used to reduce the data size for storage, handling, and transmitting content[8].
Algorithm	RLW, LZW, Arithmetic encoding, Huffman coding, Shannon-Fano coding	Transform coding, DCT,DWT, Fractal compression, Rectangle Segmentation and Sparse Matrix Storage (RSSMS).
Uses	Text or programs, images and sound	Images, audio and video.
Images	RAW, BMP, and PNG are all Lossless formats.	JPEG and GUI are lossy image formats.
Audio	WAV, FLAC, and ALAC are all Lossless formats.	MP3, MP4, and OGG are lossy audio formats.
Video	Few lossless video formats are in common consumer use, they would result in video files taking up a huge amount of space.	Common formats like H-264, MKV, and WMV are all lossy. H-264 can provides smaller files with higher qualities than previous generation of video codec because it has a "smaller" algorithm that's better at choosing the data to throw out.
Advantages	It maintains quality. Conversion in any other format possible without loss of audio information.	It can make a multimedia file much smaller than its original size.It can reduce file sizes much more than lossless compression.
Disadvantages	It doesn't reduce the file size as much as lossy compression. Lossless encoding technique cannot achieve high levels of compression.	Conversion to another format only with loss of audio information.It cannot be used in all types of files because it works by removing data. Text and data cannot be compressed because they do not have redundant information.

III LOSSLESS COMPRESSION TECHNIQUES

A) Run Length Encoding(RLE)

RLE is a simple data compression algorithm which is supported by bitmap file formats such as BMP. RLE basically compresses the data by minimize the physical size of a repeating string of characters. This repeating string is called a run which is typically encoded into two byte represents the total number of characters is the run and is called the run count and replaces runs of two or more of the same character with a number which represents the length of the run which will be followed by the real character and single characters are coded as runs of 1. RLE is useful where redundancy of data is high and it can also be used in combination with other compression techniques also [9].

B) Lempel-Ziv-Welch(LZW)

The LZW algorithm was created in 1984 by Terry Welch. LZW is a general compression algorithm capable of working on almost any type of data with references into the table of strings commonly occurring in the data being compressed, and replaces the actual data with references to the table. The table is formed during compression at the same time at which the data is encoded and the same time as the data is decoded [10]. LZW is a popular method. It has been applied for data compression. The important steps for this technique are given below; Firstly it will read the file and given a code to all character. If the same characters are found in a file then it will not assign the new code and then use existing code from a dictionary. The process is continuous until the characters in a file are null [3].

C) Huffman Coding

The Huffman coding algorithm is named after its inventor, David Huffman, who developed the method as a student in a class on information theory at MIT in 1950[11]. Huffman coding deals with data compression of ASCII characters. It is used in compression of many type of data such as text, image, audio, and video. This method is based on building a full binary tree for the different symbols that are in the original file after calculating the probability for each symbol and put them in descending order.

Table 2. Comparison between Some of the Lossless compression algorithm

FACTORS	Lossless Compression Techniques		
	RLE	**LZW**	**HUFFMAN CODING**
Advantages	It is easy to implement. It is a good alternative for a complex compression algorithm.	It is simple and good compression. Dynamic code word table built for each file. Decompression creates the code word table so it does not need to be passed.	It is easy to implement. Produce a lossless compression of images.
Speed	Fast to execute.	Fast compression	Fast to execute.
Application	TIFF, PDF, BMP	TIFF, GIF, PDF	ZIP, ARJ, JPEG, MPEG
Drawback	It cannot achieve the high compression ratios as compared to another advanced compression methods.	Management of string table is difficult. Only works for English text. Everyone needs a dictionary.	Relatively slow. Depends upon statistical model of data. Decoding is difficult due to different code lengths. Overhead due to Huffman tree.

After that, we derive the code words for each symbol from the binary tree, giving short code words for symbols with large probabilities[4].

IV LOSSY COMPRESSION METHODS

i. Transform coding

Transform coding is a type of data compression for natural data like audio signals or photographic images. The transform is typically lossy, resulting in a lower quality copy of the of the original input. In transform coding, knowledge of the application is used to choose information to discard, thereby lowering its bandwidth. The remaining information can then be compressed via a variety of methods, when the output is decoded, the result may not be identical to the original input, but is expected to be close enough for the purpose of the application[12].

ii. Discrete Cosine Transform

A discrete expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies[13]. Discrete Cosine Transform is a lossy compression

technique which is widely used in area of image and audio compression. DCTs are used to convert data in the summation of series of cosine waves oscillating at different frequencies. There are very similar to Fourier Transforms, but DCT involves use of Cosine functions and real coefficients, Fourier Transforms use both sine and cosine functions are much more efficient as fewer functions are needed to approximate a signal. Both Fourier a spatial domain into a frequency domain and their respective functions converting thing back[14].

iii. Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules. In other words, this transform decomposes the signal into mutually orthogonal set of wavelets, which is the main difference from the continuous wavelet transform, or its implementation for the discrete time series sometimes called Discrete-time continuous wavelet Transform (DT-CWT)[15].

DWT is applied to image blocks generated by the pre-processor. Two-dimension DWT leads to a decomposition of estimation coefficients at level j in four components: the approximation at level j=1, and the details in three orientations i.e. horizontal, vertical, and diagonal[14].

Table 3. Comparison between some of the Lossy compression algorithm

FACTORS	Lossy Compression Techniques		
	Transform coding	**DCT**	**DWT**
Advantages	It produce very good image quality. Increased encoding with adaptive bit assignment (ATC), an encoding complexity comparable to that of fully adaptive predictive coding (APC).	It is real valued. Better energy compaction. Coefficients are nearly correlated. Experimentally observed to work well.	It offer a simultaneous location of time and frequencies. It can be used to decompose a signal into component wavelet. Very small wavelet can be used to isolate fine details in a signal, while very large wavelet can identify coarse details.
Computation	It is computational1y intensive.	It is a fast computational approach.	It is computationally very fast.
Application	JPEG, MPEG	JPEG, MPEG, MJPEG	JPEG, MPEG
Drawback	Transform matrix cannot be factored into sparse matrix. High computational complexity.	Truncation of higher spectral coefficients results in blurring of the images, especially wherever the details are high. Coarse quantization of some of the low spectral coefficient introduces graininess in the smooth portions of the images.	DWT is shift sensitive because input signal shifts generate unpredictable changes in DWT coefficients. It suffers from poor directionality because DWT coefficients reveal only three spatial orientations.

VI CONCLUSION

In this paper we have presented the various lossless and lossy compression techniques. Lossless compression doesn't reduce the file as much as lossy compression. Sound data cannot be compressed well with conventional text compression algorithms. In Lossy compression text and data cannot be compressed. In future, different lossless algorithms can be improved to enhance the performance of the compression ratio/factor for text.

REFERENCES

1. www.webopedia.com
2. www.technet.microsoft.com
3. R.S. Brar and B. Singh, A Survey on different compression techniques and bit reduction algorithm for compression of text data" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 3, March 2013.
4. Elabdalla, A.R. and Irshid, M.I., An efficient bitwise Huffman coding techniques based on source mapping, Computer and Electrical Engineering 27 (2001) 265 – 272
5. Mamta Sharma and S.L. Bawa, "Compression Using Huffman Coding", International Journal of Computer science and Network Security, Vol.10 No.5, May 2010.
6. Amarjt Kaur, Navdeep Singh Sethi, Harinderpal Singh, " A Review on data compression techniques", International Journal of Advanced Research in Computer Science and Software Engineering", Volume 5, Issue 1, January 2015.
7.en.wikipedia.org/wiki/Lossless¬_compression
8. https://en.wikipedia.org/wiki/Lossy_ compression
9. Neha Sharma, Jasmeet Kaur, Navmeet Kaur, "A Review on various Lossless Text Data Compression Techniques", International Journal of Engineering Sciences, Issue Dec 2014, Vol 2.
10. S. Kapoor and A. Chopra, "A Review of Lempel Ziv Compression Techniques" IJCST Vol.4, Issue 2, April-June 2013.
11. S. Shanmugasundaram and R. Lourdusamy, "A Comparative Study of Text Compression Algorithms" International Journal of Wisdom Based Computing, Vol. 1 (3), December 2011.
12. https://en.wikipedia.org/wiki/Transform_coding
13.https://en.wikipedia.org/wiki/Discrete_cosine transform
14. Kitty Arora, Manshi Shukla, "A Comprehensive Review of Image Compression Techniques", International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 1169-1172
15.www.gwyddion.net/documentation/user-guide-en/wavelet-transform.