A Survey on Lossless and Lossy Data Compression Methods

P.KAVITHA[1]
PG Scholar
SNS College of Technology
Coimbatore, Tamilnadu ,India.
pkavitha29@gmail.com

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 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.

Key words : Data compression ,Lossless compression and Lossy compression

I. Introduction

Now a days data bases are mandatory everywhere. 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 databases. 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, data base tables and in medical image because of the law of regulations, some of the main techniques are Run Length Encoding, Arthimatic Encoding, Shannon fano, Lempel Ziv Welch, Huffman coding,. Lossy compression algorithm is ususlly 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 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 a data transmission and a data storage despie 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 compression techniques

- Lossless Data Compression
- Lossy Data Compression

2.1 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 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].

2.2 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 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 "Video compression" is some times differented as a separate discipline from "audio compression"[3].

Table1.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.	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, RSSMS
USES	Text or programs, images and sound	Images, audio and video
IMAGES	RAW, BMP and PNG	JPEG and GUI are lossy image
Audio	WAV, FLAC AND ALAC	MP3, MP4 and OGG
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. 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.
Dis Advantages	It does not 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 bytes 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 character 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.

B) Lempel-Ziv-Welch (LZW)

The LZW is a general compression algorithm capable of working on almost any type of the 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 enoded and the at the same time data is decoded[10]. LZW is a popular method. It has been applied for the data compression. The important steps for this technique are given below; firstly it will read the file and given the code to all character. If the same character is found in the file then it will not assign the new code and then use the existing code from a dictionary. This process is continues until all the character in the file is null [3].

C) HUFFMAN CODING

The Huffman coding deals with data compression of ASCII characters. It is used in compression of many type of data such as text, audio, video and image. This method is based on the building a full binary tree for the different symbols that are in the original file after calculating the probability of each symbol and put in descending order.

Table 2. Compression between some of the Lossless Compression algorithm

Factors	Lossless Compression Techniques		
	RLE	LZW	Huffman coding
Advantages	It is easy to implement It is good alternative for 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 the Lossless compression of images
Speed	Fast to execute	Fast compression	Fast to execute
Application	TIFF, BMP, PDF	TIFF, GIF, PDF	Zip, ARG, JPEG, MPEG
Drawback	IT Cannot achieve the high compression ratios as compared to another advances compression methods.	Management of string table is difficult. Only works for English text. Every one needs dictionary.	Relatively slow Depends upon statistical model of data. Decoding is difficult due to different coding lengths. Overhead due to Huffman coding.

IV. LOSSY COMPRESSION METHODS

I Transform coding

Transform coding is a type of data compression for natural data like audio signals or images. The transform is typically lossy, resulting ina lower quality copy of the original input. In transform coding , knowledge of the application is used to choose information to discard, therby lowering its band width. The remaining information can be compressed via variety 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 (DCT)

A discrete expresses a finite sequence of data points in the terms of the sum of cosine functions oscillating at different frequencies. DCT 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 involues uses of cosine functions are much more efficient as fewer function are needed to approximate a signal.

III DISCRETE WAVELET TRANSFORM (DWT)

The DWT is an implementation of the wavelet transform using a discrete set of the values 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 differences from the continuous wavelet transform or its implementation fro the discrete time series sometimes called Discrete –time continuous wavelet transform (DT-CWT). DWT is applied to the image block generated by the pre processor .

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 compactations. Co efficient are nearly correlated. Experimentally observed to work well	It offer a simulataneous location of time and frequencies. It can be used to decompose a signal into component wavelets. Very small wavelet can identify coarse details.
Computation	It is computationally intensive.	It is fast computationally 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 co efficient 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 un predictable changes in DWT co efficient. It suffers from poor directionality because DWT coefficients reveal only three spatial orientations

CONCULSION

In this paper presented the various lossless and lossy compression techniques. Lossless compression does not 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 improves to enhance the performance of the compression ratio fro 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 Applied Research in Computer Science and Software Engineering (IJARCSSE) vol 3issue 3 march 2013.
[4] Ellabla. A.R and Irshid "An efficient bitwise Huffman coding techniques based on source mapping computer and electrical engineering "
[5] Mamta Sharma,"Compression using Huffman coding" international Journal of computer science and network security, vol 5 no 5 may 2012
[6] Amria raju, Navadeep singh " A Review on data compression techniques" International Journal of Applied Research in Computer Science and Software Engineering (IJARCSSE) vol5 jan 2015

[7] en.wikipedia.org/wiki/Lossless compression
[8] en.wikipedia.org/wiki/Lossy compression
[9] Neha Sharma, Jasmeet singh " a review on various lossless text data compression techniques" International journal of electrical science issue dec 2014
[10] S.Kapoor and A.Chopra, "A review of Lempel Ziv compression techniques" IJCST vol4 issue 2 april 2013
[11] https://en.wikipedia.org/wiki/transform_coding
[12] https://en.wikipedia.org/wiki/Discreate_cosine transform