Data Compression using a Generic Approach

Shreyas Kalrao
shreyaskalrao@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

Someshwar Gaikwad
gaikwadsomeshwar6@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

Sushant Sontakke
smsontakke111@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

Yash Rathi
yashrathi878@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

ABSTRACT

The amount of data being generated in the great age of Big Data is enormous, which in turn has led to a large demand for storage infrastructures. With this growing demand the storage infrastructures that depend on natural resources like silicon for hardware will become expensive. Hence there is a need to use techniques that minimize the size required by the data and efficiently store the data on these infrastructures. Data compression is a technique that reduces thesize of data. This project's goal is to reduce the size of the original data without losing the quality of the data. We propose a generalized system with which data ofany type or more commonly multimedia can be reduced in size without losing the quality of the data.

Keywords: Data Compression, Multimedia, Quality

1. INTRODUCTION

Storage device capacity and speed have been steadily increasing since the 1980s. Today, new types of memory that are less expensive and more efficient are being developed. Nowadays the documents gotten too huge, this issue is exacerbated by today's rapidly increasing file sizes and the amount of multimedia storage used per user. One reason is that the quality of the images and audio that comprise the users' data has grown exponentially, and another is that the growing use of multimedia currently is further exacerbating the storage requirements. It would be beneficial if we could somehowEncode them to a lower size and then restore them when they are required later. As a result, we will be able to use more of our Diskettes, hard discs, CDs, and USB flash drives, tapes, etc.

The key challenge is to save this large amount of data in very small memory spaces without losing the quality perceived. To overcome this problem an efficient data compression approach is highly recommended. Data compression is a way to improve energy efficiency, improved memory use, and efficient transmission of data. Data compression works by various techniques of compression and uses various algorithms for data compression to reduce the size of user data. Such techniques are important today when it comes to the digital storage of information on computer discs and transmission over communication networks.

2. LITERATURE SURVEY

Mohammad Sheraj, Ashish Chopra [1] discovered an algorithm which aims at achieving highest data compression ratio in a lossy scenario while still being able to retainoriginal image's resolution and audio file's bitrate. For this, feature extraction is used on chunks of data and stored in a database with a hash as a key. This hash is stored in the fileand full data is later reconstructed from the database. The size of the hash comes out as a very small fraction as compared to the size of the whole chunk leading to a big amount of compression. This algorithm is independent of the input file as it works on raw data. The algorithm is found to provide significant advantages over conventional compression formats, both lossy and lossless.

R.Mahalakshmi, S.K.Mahendran [2] highlighted the importance of data compression and various compression algorithms that are currently being used by different multimedia (text, image, audio, video) files. They talk about the importance of lossy and lossless compression and highlight the point that lossless compression is suited for important text files such as bank records and text articles. They have categorized the different algorithmsfalling under lossy and lossless compression and have given a brief overview of all such algorithms. The purpose of their paper was to highlight a very important fact that differenttypes of files require different algorithm. A simple binary data can be compressed using algorithms like Run-Length Encoding and Huffman Coding, but one can only determine a suitable algorithm if they know the contents of the file.

Nor Asilah Khairi, Asral Bahari Jambek [3] proposed that for transmitting and storing data lots of energy and memory space is required so data compression is a solution for better performance regarding energy-saving, better memory usage, and data transmission efficiency. In this article, they discussed about several data compression algorithms. A tabular data has been gathered and summarised in the explanation of compression algorithms in the paper. Due to its great potential for satisfying their project objectives,the study identified MAS compression algorithm as a project prototype. The paper also contains comparison between different types of data compression technique and explainedMAS compression algorithm thoroughly.

K.Muthuchamy [4] elaborated different data compression types and and also proposed different techniques to implement those types. Data compression can also speed up a transmission of data from one person to another. In performing a compression requires a method of data compression that can be used, the method can then be used to compress a data. The paper also explains how data compression works through several compressing techniques and software solutions that utilise data compression algorithms to reduce the data size.

Luluk Anjar Fitriya, Tito Waluyo Purboyo, Anggunmeka Luhur Prasasti [5] present a review different types of data compression techniques. Data compression is widely used by the community because through a compression we can save storage. This paper also explains how a method works in doing a compression and explains which method is well used in doing a data compression in the form of text. The output generated in doing canbe known through the compression file size that becomes smaller than the original file.

3. CONCLUSION

Data compression is the art of finding short descriptions for long strings. Every compression algorithm can be decomposed into zero or more transforms, a model, and a coder. The advantages of compression are a reduction in storage hardware, data transmission time and communication bandwidth and the resulting cost savings.

The compressed file also requires less time for transfer, and it consumes less network bandwidth than an uncompressed file. Various compression algorithm techniques have advantages and disadvantages in doing a compression. The algorithm is based on a database which can be varied in size according to the nature of the deployment, whether it is on device or server based.

4. FUTURE SCOPE

The further gains in lossless compression will be incremental at best, but there arehuge opportunities in lossy compression. Future scope of project are updates in existing project with newer and faster techniques.

And lossy compression hasn't slowed down. Each decade sees about a 50% reduction in bitrate required for a given quality experience. Combined with faster speeds andcheaper storage.

As a comparison point, a DVD used MPEG-2 up to about 9.2 Mbps. We can now deliver 4K HDR with an average bitrate of 9.2 Mbps. That's 24x more pixels, and much better pixels and image compression is a lot better. HEIF can reduce bitrate versus JPEG or PNG from 50–98% depending on content. That's a combination of big improvements. Similar upgradation of technologies will be the future scope.

5. REFERENCES

[1] Mohammad Sheraj, Ashish Chopra, "Data Compression Algorithm for Audio and Image using FeatureExtraction," 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP).

[2] R.Mahalakshmi, S.K.Mahendran, "Data Compression in Multimedia (Text, Image, Audio and Video),"International Journal of Engineering Sciences and Research Technology ISSN 2227-9655.

[3] Nor Asilah Khairi, Asral Bahari Jambek, "Study on data compression algorithm and its implementation in portable electronic device for Internet of Things applications," EPJ Web of Conferences 162, 01073 (2017).

[4] K.Muthuchamy, "A study on various data compression types and techniques," International Journal ofResearch and Analytical Reviews vol. 5 iss. 3, ISSN 2349-5138, August 2018.

[5] Luluk Anjar Fitriya, Tito Waluyo Purboyo, Anggunmeka Luhur Prasasti , "A Review of Data Compression Techniques," International Journal of Applied Engineering Research ISSN 0973-4562 vol. 12, no. 19 (2017) pp. 8956-8963.