A REVIEW ON THE TECHNIQUES DEPLOYED IN DATA COMPRESSION

Sagargouda S. Patil

Department of Computer Science& Engg,
KLE Dr.M.S.Sheshagiri CET, Udyambag, Belgaum, (India)

ABSTRACT

Data compression is an art of representing input data or original data into another form of data which is in a compact form. This technology plays a vital role in many applications. There are a large number of data compression algorithms available for compressing files of different formats. This paper provides a survey of different basic techniques to compress larger data into smaller one which results in reduction in time and increased efficiency.

Keywords: Data Compression, Huffman Coding, JPEG.

I INTRODUCTION

Data compression is useful because it helps toreduce resources usage, such as data storage space or transmission capacity [1]. Data compression is also widely used in backup utilities, spreadsheet applications, and database management systems. Certain types of data, such as bit-mapped graphics, can be compressed to a small fraction of their normal size [2].

Data compression can be classified into two techniques-one being the lossy compression, andthe other, lossless compression.In lossy compression some information is lost during the process, where the image data is stored into important and unimportant data.

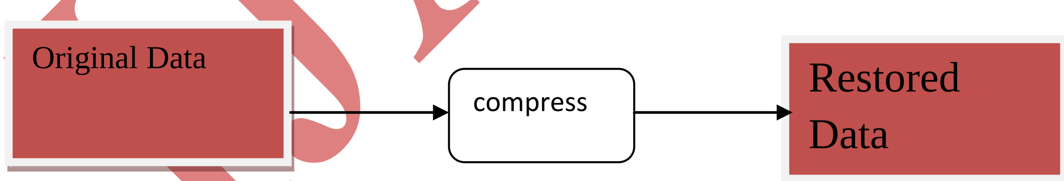

Fig. 1 Block Diagram of Lossless Compression.

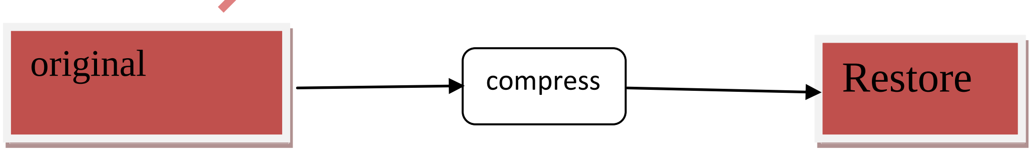

Fig. 2. Block Diagram of Lossy Compression

The system then discards the unimportant data. It provides much higher compression rates, but there will be some loss of information compared to the original source file.In lossless compression [11] no information is lost during the compression and the decompression process [3, 12]. Here the reconstructed image is mathematically and visually identical to the original one [13]. Fig. 1, 2 shows the pictorial representation of lossless and lossy compression techniques, respectively.

The rest of the paper is organized as follows. Effective data compression techniques have been briefed in Section 2. Section 3 deals with their applications. The paper is concluded in Section 4.

II COMPRESSION TECHNIQUES

2.1 Huffman coding

It is an entropy based encoding algorithm [10], designed by David A. Huffman [6]. The running time of Huffman's method is O(log n)[5].This is also similar to Shannon Fano method, but the difference between these two were in Shannon Fano method[8,9],it constructs its code from top to bottom. In Huffman coding method it constructs its code from bottom to top.

Two major parts in Huffman Coding:-

1. We can build a Huffman tree from a set of input characters.
2. Traversing the Huffman Tree and assign codes to each character [14].

Fig 2. gives an example for Huffman coding problem.

a1	0.35	0.35	0.35	**0.40**	**0.6**
a2	0.25	0.25	0.25	0.35	**0.4**
a3	0.20	0.2	0.2	0.25	
a4	0.10	0.1	**0.2**		
a5	0.05	**0.1**			

Fig 2. Example for Huffman coding

2.2 JPEG

JPEG (Joint Photographic Expert Group) is used to compress images effectively and video data [7].This method is of lossy compression for digital photography [4] (Fig. 3.). JPEG achieves 10:1 compression. The three main goals of the jpeg compression method are as follows:-

1. Higher Compression Ratio

2. Highly Sophisticated

3. This makes good results in pixel aspect ratios, colour spaces, other image features.

JPEG image coding formats in JPEG2000[17,18] and JPEG were

i) JPEG 2000 **ii) JPEG**

Fig 3. Examples for JPEG 2000 and JPEG.

2.3. LZ77 and LZ78

These two techniques are the types of lossless data compression methods, devised by Abraham Lempel and Jacob Ziv in 1977 [19] and 1978 [20]. They are well-known as LZ1 and LZ2, respectively. They are responsible for many variants like LZW, LZSS, LZMA, etc. They were used exclusively for GIF and DEFLATE algorithm used in PNG. They are called as dictionary techniques. LZ77 comprises of a sliding window during compression, which was then showed to be equivalent to the explicit dictionary constructed by LZ78. However, they are equivalent only when the whole of the data is to be decompressed.

III APPLICATIONS

3.1 Image Compression

This can be either lossless or lossy. Lossy technique used at usually lower bit rates, introduce compression artifacts. Lossy method is usually appropriate for natural images like photographs in applications where slight loss of fidelity is acceptable. However, lossless compression is often recommended for technical drawings, medical images, clip art, face recognition applications, etc [21-23]. There are many algorithms available to employ lossless image compression, popular among them being the following:

1) Predictive Coding and DPCM

2) Area image compression

3) Run-length encoding

4) Entropy encoding

5) Deflation

6) Adaptive dictionary algorithms like LZW

7) Chain codes

3.2 Video Compression

The technique mainly focuses on partial or total elimination of redundancy in video data [24-26]. They primarily combine temporal motion compensation and spatial image compression. It is a practical implementation of source coding. They widely rely on lossy compression techniques. MPEG-4 achieves a compression factor between 20 and 200. There are many challenges ingrained in video compression: highly compressed video may lead to distracting or visible artifacts. Deploying lossy schemes may cause a trade-off in video quality. The cost of processing the compression and decompression and system requirements is considerably high.

3.3 Audio Compression

It is a type of lossless or lossy compression in which the total amount of data in the audio waveform is reduced to differing extents for storage or transmission [27-29]. Lossy audio compression bestows higher compression and is used in wide number of audio applications. They mainly use psychoacoustics to eliminate less meaningful or less audible sounds. The following are a set of lossless audio compression formats:

1) Shorten [30]
2) Apple Lossless (ALAC)
3) Free Lossless Audio Codec (FLAC)
4) WavPack
5) Windows Media Audio 9 Lossless (WMA Lossless) by Microsoft
6) Monkey's Audio
7) MPEG-4 ALS

3.4 File Compression

The UNIX compress command is one of the earliest applications of LZW. The dictionary size is adaptive.

3.5 Portable Network Graphics (PNG)

It is based on the implementation of LZ77[16].This allows match length between 3 and 258.

3.6 Graphics Interchange Format (GIF)

It is also another implementation of LZW[15]. Initial size of the dictionary is 2b+1.Compressed image is stored with 1st byte.

IV CONCLUSION

The various algorithms which have caused acceleration in data compression are discussed in the paper. But since, data compression is still a developing field, many enhancements to the available techniques is crucial so as to boost up their performance.

ACKNOWLEDGEMENTS

The author is immensely grateful for the valuable guidance and support provided by Dr.V.S. Malemath ,Senior Lecturer, Department of Computer Science& Engg,KLE Dr.M.S.Sheshagiri CET, Udyambag, Belgaum;and colleague Rahul Dashrath Gavas.

REFERENCES

[1] J.Ziv and A.Lempel, "A universal algorithm for data compression", IEEE Transactions on Information Theory, IT-23(3):337-343, May 1977.

[2] J.Ziv and A.Lempel, "Compression of individual sequences via variable-rate coding", IEEE Transactions on Information Theory, IT-24(5):530-536, September 1978.

[3] CCSDS Secretariat, "Lossless Data Compression", The Consultative Committee for Space Data Systems.

[4] TOURADJ EBRAHIMI, MURAT KUNT, "Visual Data Compression for Multimedia Applications", PROCEEDINGS OF THE IEEE, VOL. 86, NO. 6, JUNE 1998.

[5]. Mridul Kumar Mathur, Seema Loonker, Dr. Dheeraj Saxena, " LOSSLESS HUFFMAN CODING TECHNIQUE FOR IMAGE COMPRESSION AND RECONSTRUCTION USING BINARY TREES",Int.J.Comp.Tech.Appl,Vol 3 (1), 76-79.

[6] Huffman D.A., "A method for the construction of minimum redundancy codes", *Proceedings of the Institute of Radio Engineers, 40 (9), pp. 1098–1101, September 1952.*

[7] TOURADJ EBRAHIMI, MURAT KUNT, "Visual Data Compression for Multimedia Applications", PROCEEDINGS OF THE IEEE, VOL. 86, NO. 6, JUNE 1998.

[8] Fano R.M., "The Transmission of Information", Technical ReportNo. 65, Research Laboratory of Electronics, M.I.T., Cambridge,Mass.; 1949.

[9]Shannon C.E., "A mathematical theory of communication," BellSys. Tech. Jour., vol. 27, pp. 398-403; July, 1948.

[10] Huffman D.A., "A method for the construction of minimumredundancycodes", *Proceedings of the Institute of RadioEngineers*, 40 (9), pp. 1098–1101, September 1952

[11] R. Sankaralingam, R. Orugani, and N. Touba, "Static compactiontechniques to control scan vector power dissipation," in Proceedings ofthe IEEE VLSI Test Symposium (VTS '00), pp. 35–40, Montreal,Canada, May 2000.

[12] Manber, U.; Myers, G. Suffix arrays: A new method for on-line stringsearches. SIAM J. Comput.1993, 22, 935–948. 2. Larsson, N.J.;Sadakane, K. Faster suffix sorting. Theoret. Comput. Sci. 2007, 317, 258–272.

[13] Manzini, G.; Ferragina, P. Engineering a lightweight suffix arrayconstruction algorithm.Algorithmca 2004, 40, 33–50.

[14] Abouelhoda, M.I.; Kurtz, S.; Ohlebusch, E. Replacing suffix trees withenhanced suffix arrays. J. Discrete Algorithms 2004, 2, 53–86.

[15] Welch T.A., "A technique for high-performance datacompression", *IEEE Computer*, 17, pp. 8–19, 1984.

[16] Ziv. J and Lempel A., “A Universal Algorithm for Sequential DataCompression”, IEEE *Transactions on Information Theory* 23 (3),pp. 337–342, May 1977.

[17] “New work item: JPEG 2000 image coding system,” ISO/IECJTC1/SC29WG1 N390R, JPEG2000, 1997.

[18] “Call for contributions for JPEG 2000 image coding system,”ISO/IEC JTC1/SC29WG1 N505, JPEG2000, 1997.

[19] Ziv, Jacob; Lempel, Abraham (May 1977). "A Universal Algorithm for Sequential Data Compression". *IEEE Transactions on Information Theory* **23** (3): 337–343.doi:10.1109/TIT.1977.1055714

[20]Ziv, Jacob; Lempel, Abraham (September 1978). "Compression of Individual Sequences via Variable-Rate Coding". *IEEE Transactions on Information Theory* **24** (5): 530–536.
doi:10.1109/TIT.1978.1055934

[21] Taubman, David. "High performance scalable image compression with EBCOT."*Image Processing, IEEE transactions on* 9, no. 7 (2000): 1158-1170.

[22] Marcellin, Michael W. *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards, and Practice*. Vol. 1. springer, 2002.

[23] Lewis, Adrian S., and G. Knowles. "Image compression using the 2-D wavelet transform." *Image Processing, IEEE Transactions on* 1, no. 2 (1992): 244-250.

[24] Le Gall, Didier. "MPEG: A video compression standard for multimedia applications." *Communications of the ACM* 34, no. 4 (1991): 46-58.

[25] Sullivan, Gary J., and Thomas Wiegand. "Video compression-from concepts to the H. 264/AVC standard." *Proceedings of the IEEE* 93, no. 1 (2005): 18-31.

[26] Bhaskaran, Vasudev, and Konstantinos Konstantinides. *Image and video compression standards: algorithms and architectures*. Springer, 1997.

[27] Pan, Davis. "A tutorial on MPEG/audio compression." *IEEE multimedia* 2, no. 2 (1995): 60-74.

[28] Gersho, Allen. "Advances in speech and audio compression." *Proceedings of the IEEE* 82, no. 6 (1994): 900-918.

[29] Hans, Mat, and Ronald W. Schafer. "Lossless compression of digital audio."*Signal Processing Magazine, IEEE* 18, no. 4 (2001): 21-32.

[30] Robinson, Tony. "SHORTEN: Simple lossless and near-lossless waveform compression." (1994).