Analysis of File Formats and Lossless Compression Techniques for Medical Images

V. Punitha
Department of Computer Science and Applications
The Gandhigram Rural Institute (Deemed to be University)
Dindigul, Tamilnadu, India
punithacnp@gmail.com

P. Kalavathi
Department of Computer Science and Applications
The Gandhigram Rural Institute (Deemed to be University)
Dindigul, Tamilnadu, India
pkalavathi.gri@gmail.com

***Abstract*— The rapid development of Medical Images in the medical field helps the doctor with integrated and remote healthcare services for accurate diagnostic, treatment and surgical planning. With the growing advancement in medical imaging technologies, a huge volume of medical images are being acquired and processed every day leading to huge storage requirements and increased bandwidth for transmission. To fulfill the above requirements, Medical Image compression techniques are developed to compress the Medical Images acquired from various imaging modalities with different image format. This paper discusses the various Medical Imaging formats used in clinical practices and also performance analysis of some of the existing lossless image compression techniques for the Medical Images obtained from different modalities.**

Keywords — *Image Compression, Medical Image Compression, Medical Image File Formats, Lossless Compression.*

I. Introduction

In recent years, the utilization of medical imaging has become an indispensable tool in clinical practices. Medical imaging is a techniques and process used to examine the internal organs of the human body for diagnostic and treatment process. Medical imaging used to exam and decline in mortality, fewer hospital admissions, shorter hospital stays, longer life expectancy, and reduced need for exploratory surgery. There are different medical imaging modalities are used for imaging a human body such as X-Ray, Magnetic Resonance Imaging (MRI), Ultra Sound Imaging (US), Computed Tomography Scanning (CT), Positron Emitted Tomography (PET) etc. Nowadays, development and need of medical imaging modality tremendously increased, and the need for producing, transferring and sharing of medical images have also been amplified. It leads to insufficient bandwidth of network and storage of memory device. Medical images contains more details or information than simple text or document files so it required extra demand of bandwidth to travels through different types of networks [1]. Therefore we have to reduce the size of image before it is going to be stored or transmitted. Here comes the need for compression. Compression defines reduces the size of data required to represent given amount of information. Therefore, efficient compression techniques are essential in the field of telemedicine and its applications.

Rest of this paper is organized as follows: In Section 2 we discuss about Types of Medical images and Types of compression, In Section 3 we provide a brief discussion about lossless compression techniques for Medical Images, In Section 4, introduces material and metrics used for compression, Finally, Section 5 provides a Results and Discussions.

II. Types of Medical Image Formats and Compressions Methods

A. *Types of Medical Image File Formats*

Since the medical images are stored in different format by different modalities, the retrieval, processing and transmitting is a challenging task. This paper describes four major file formats currently used in medical imaging. They are Neuroimaging Informatics Technology Initiative (Nifti), Analyze, Minc and Digital Imaging and Communications in Medicine (DICOM). Medical Image file formats are categorized into two types. First is the format wished-for to standardize images generated by medical imaging modality like DICOM. Second is the formats aim to facilitate and strengthen the post processing analysis like Analyze, Minc, and Nifti.

Analyze is one of the de-facto standard for Medical image post processing techniques. It has been designed for multidimensional data volume. This volume consists two types of binary files: image file with extension of “.img” that contains the voxel raw data and a header file with extension “.hdr” containing the meta data such as number of pixels in the x, y and z directions, voxel size and data type. This format is old but still widely used by many software packages [3].

Nifti is a file format developed by National Institutes of Health for neuroimaging to overcome the limitation of Analyze format so it is referred as revise Analyze format. Nifti also require storing the header and image data in separate files. In this format images are typically saved as a single file with “.nii” extension where both files are merged together. Analyze format is rapidly replaced by Nifti format for neuroimaging research and also adopted as default format for some of the mostly used software domain packages such as FSL, SPM (Statistical Parametric mapping) and AFNI (Analysis of Functional NeuroImages) Image viewer and analysis software like 3D slicer, ImageJ and OsiriX are also supported with this format.

The Minc file format was developed at the Montreal Neurological Institute (MNI). It provides a flexible data format for medical imaging. This format mainly used by a viewer and software processing library called MNI Brain Centre. The first version of minc based on standard Network common data format (NetCDF). Next version of minc2 is based on the standard Hierarchical Data Format version 5 (HDF5). This format uses the ".mnc" as an image extension.[3]

The DICOM is an international standard used for medical imaging in the area of medical informatics. It was established by the American college of Radiology and the National Electric Manufacturers Association at the end of 1990. DICOM is not only a file format but also a network communicational protocol. It assists manufacturers and users to interconnect devices with medical imaging equipment on networks and provide access to several types of medical images [4]. This format is more user friendly in terms of access, exchange and usability of diagnostic medical records. DICOM file containing Metadata, pixel data merging together as a single file and header file contains the complete description about imaging procedure with patient information such as name, gender, age, weight and height. This format images are typically stored with .dcm extension [3].

Table 1: Summary of the Medical Image File formats

Format	Header	Extension	Software Used
Analyze	Fixed length : 348 byte binary format	.img & .hdr	SFL, SPM, MRIcro
Nifti	Fixed length : 352 byte binary format	.nii	FSL, AFNI, SPM, 3D Slicer, OsiriX
Minc	Extensible binary Format	.mnc	MNI Brain Imaging Centre
DICOM	Variable length binary format	.dcm	PACS

B. *Types of Compression Methods*

Commonly image compression methods are classified into two types: Lossy and Lossless compression. In lossy compression the details of the image or quality of the image are lost where as in lossless compression, it produces the image without any loss of information as like as original image with reduced storage space.[5][8] Often medical image compression requires lossless compression because medical images have high valuable information and details, it is necessary to preserve those details while compression. Nowadays, a near lossless compression technique are becoming popular, it is irreversible and does not affect the data much after decompression.

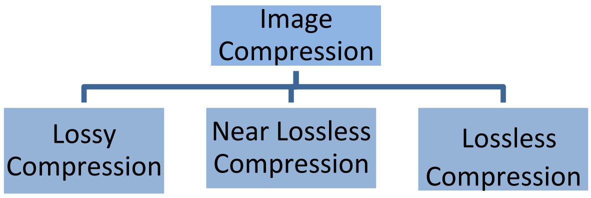


Fig. 1. Types of Medical Image Compression

III. Analysis of Lossless Compression Techniques for Medical Images

In Lossless compression the original and the decompressed images must be identical. Since the Medical image contains high valuable information and thus the decompressed data should be identical as original. Various kinds of compression techniques are developed for Lossless compression. In this paper we analyzed the performance of the following three popular lossless image compression techniques for medical images.

Methods used for Lossless compression:

1. Run length Encoding
2. LZW Encoding
3. Huffman Coding

A. Run Length Coding

Run length coding is a form of lossless data compression technique used to compress the repeated pixel information stored in an image. Here the data values of same data are stored as a pair of single value and its count rather than retain a long original data [12]. The run length code for gray scale image is represented as $\{V_i, C_i\}$ where V_i is the value of intensity and C_i refers to the number of pixels with the same intensity value V_i and is shown in Figure 2 [2][13].

25	25	25	32	32	32	32	48	48

$\{25, 3\}$	$\{32, 4\}$	$\{48, 2\}$

Fig. 2. Run Length Encoding

The main disadvantage of this RLE is in worst case the size of compressed data may be twice as the original data when the image contains less number of redundant data. Later on, to overcome this problem Huffman algorithm is developed [7].

B. LZW Compression

Abraham Lempel, Jacob Ziv, and Terry Welch developed LZW compression technique in the year 1984. It is a kind of dictionary based algorithm. In this, the original file is compressed into a smaller file using table based look up algorithm. It reads each sequence of bits or symbols, grouping them as a string and converts those strings to codes of a given length and creates an entry into a table. The Pointer size selected as 12 bits and it creates 4096 dictionary entries [4]. LZW performs well when image file contains more number of redundant data. This algorithm mainly used in GIF and often used in TIFF and PDF. It is easy to implement and provides a high throughput [5][10].

C. Huffman Encoding

Huffman encoding algorithm is developed by David Huffman in the year of 1950 [2]. It is a lossless compression algorithms based on the frequency of occurrence of a symbol in the original file that is going to be compressed. The main idea behind this algorithm is fixed length codes are replaced by variable length codes. The most probably occurring symbols are represented by shorter bit codes. This is implemented by creating a Huffman tree where each node represents a symbol in the alphabet. After constructing the Huffman tree, it creates a prefix code for each symbol from the alphabet. The tree traversing from root to the node corresponds to the symbol. It assigns the smaller bit 0 for left branch and 1 for right branch like a binary tree. These symbols are stored in the header of compressed file that is being used at the decoding process.[8]

IV. METRICS AND DATASETS USED

Time and Space efficiency are the two important factors in compression. Performance of the compression largely depends on the redundancy of data in the image [9][11]. Performance measures used in this paper is as follows:

Compression Ratio: The ratio between number of bits needed for original image and compressed details is called as compression ratio (CR). The following expression is used to calculate the Compression Ratio.

$$CR = \frac{No.of\ bits\ in\ original\ Image}{No.of\ bits\ for\ compressed\ file}$$

Bit Per Pixel: Number of bits required to store a pixel is called as Bits per Pixel (BPP)/Bit Rate (BR).

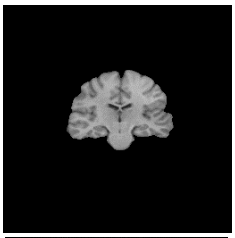
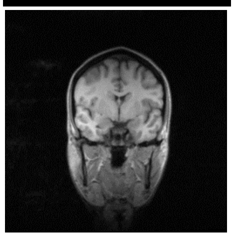
$$Bit\ Per\ Pixel = \frac{Size\ of\ compressed\ file}{No.of\ samples}$$

Datasets Used:

For the performance analysis, we have taken images from IBSR (Internet Brain Segmentation Repository) for MRI images and RadioPedia for other medical image modalities.

V. RESULTS AND DISCUSSIONS

The computed results for selected lossless compression techniques are represented in the following Table 2, Table 3, and Table 4 for RLE, LZW, and Huffman respectively for the 5 test images from different modalities (Fig 3). These tables show the Size of original image, compressed image in pixel, Compression Ratio and Bit Per Pixel. Table 2 computed, compression percentage using RLE, LZW and Huffman method for the test images shown in Figure 3. The performance comparison of Compression Ratio (CR) between the selected three algorithms RLE, LZW, and Huffman for the images shown in Figure 3 are depicted in the graph Figure 4. Similarly the performance comparison of BPP obtained by the three lossless compressions techniques such as RLE, LZW and Huffman on the tested images (Figure 3) are shown in Figure 5. From Figure 4 and Figure 5, it is evident that among the three selected lossless compression methods Huffman method perform well on the test images by producing BPP and higher Compression Ratio value. Thus, among these selected three lossless methods Huffman method perform well on compressing medical images. Working in the system of IEEE 802.11, administration contrast entiation in the MAC convention can be acquired by shifting the backoff clock dispersion, the concede time (DIFS), and the size of the parcels [1]. Expecting that bundle lengths can't be constrained by the MAC layer for continuous traffic, we concentrate on the initial two parameters and next present our proposed instrument for circulated need booking and versatile backoff for IEEE 802.11.

Image No	Image Name	Original Image
1.	MRIBrain	
2.	MRI	

3.	CT-Spine	
4.	Ultra sound	
5.	XRay	

Image No	Image Name	Original image size in "pixels"	Compressed image size in "pixels"	Compression ratio	Bit Per Pixel	Comp_Per (%)
1	MRI-Brain1	524288	16224	32.31	4.03	96.90
2	MRI	524288	38624	13.57	1.69	86.41
3	CT-Spine	524288	60320	08.69	1.08	99.91
4	Ultra Sound	524288	59120	08.86	1.10	88.71
5	X Ray1	524288	52432	09.99	1.24	89.98

Table 2:Computed Compression Measures for RLE Method

Table 3: Computed Compression Measures for LZW Coding

Image No	Image Name	Original image size in "pixels"	Compressed image size in "pixels"	Compression Ratio	Bit Per Pixel	Comp_Per (%)
1	MRI-Brain1	524288	8017	8.17	1	87.76
2	MRI	524288	16224	5.16	1.54	80.62
3	CT-Spine	524288	14651	4.47	1.78	77.62
4	Ultra Sound	524288	14904	4.39	1.81	77.22
5	X Ray1	524288	16928	3.87	2.06	74.16

Table 4: Computed Compression Measures for Huffman Coding

Image No	Image Name	Original image size in "pixels"	Compressed image size in "pixels"	Compression ratio	Bit Per Pixel	Comp_Per (%)
1	MRI-Brain1	524288	115479	4.54	1.76	43.18
2	MRI	524288	343679	1.52	5.24	34.21
3	CT-Spine	524288	284113	1.84	4.33	45.65
4	Ultra Sound	524288	307820	1.70	4.69	41.17
5	X Ray1	524288	479728	1.09	7.32	08.25

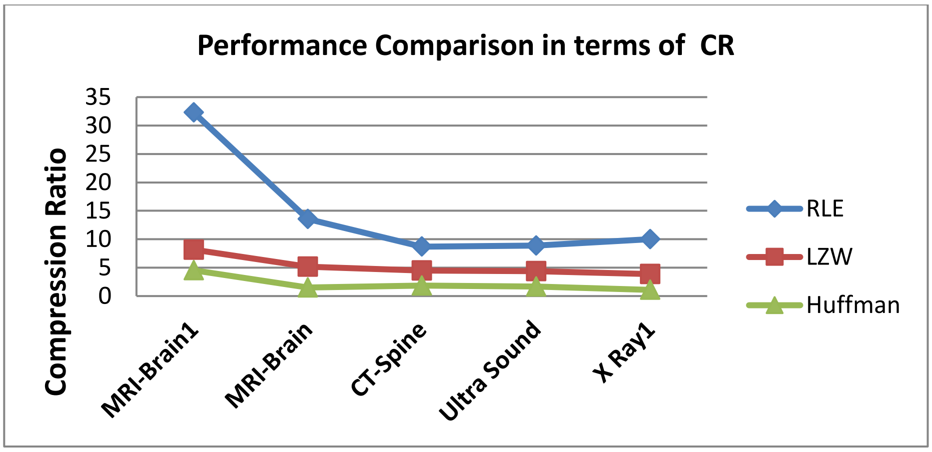


Fig. 3. Performance Comparison in terms of CR

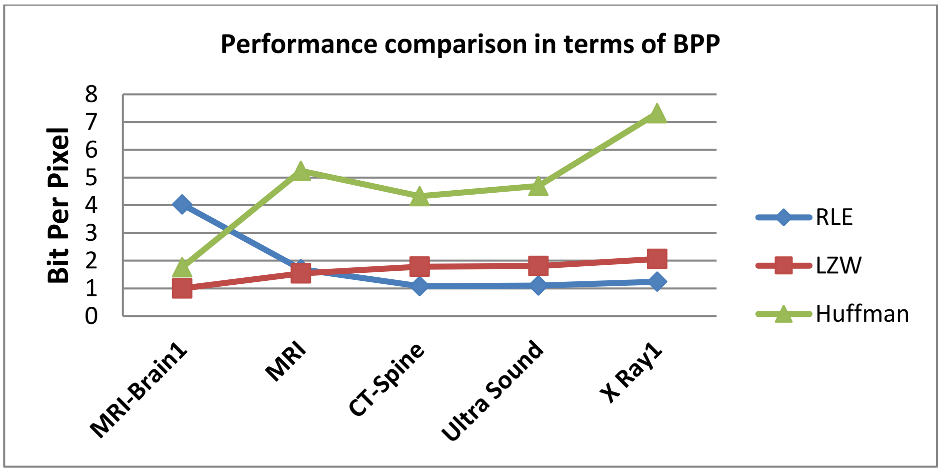


Fig. 4. Performance Comparison in terms of BPP

VI. CONCLUSION

In this paper, we have compared three lossless compression methods RLE, LZW, Huffman for medical images obtained from different modalities and we evaluated the methods with different performance metrics such as Compression Ratio, Bit Per Pixel, and Compression Percentage. We found that Huffman coding provides a better result for medical images. In future, we could modify this method to yield good compression result for 3D medical images.

Acknowledgement

This work was supported by Council of Scientific & Industrial Research (CSIR), File Number: 25(0304)/19/EMR-II/, Human Resource Development Group, and Government of India.

References

[1] M. Hasan and K. Nur, "A Lossless Image Compression Technique using Location Based Approach," vol. 1, no. 2, 2012.

[2] Walaa Z. Wahba, Ashraf Y. A. Maghari, "Lossless Image compression Techniques Comparative Study", International Research Journal of Engineering and Technology, Volume: 03 Issue: 02, 2016.

[3] Michele Larobina, Loredana Murino, "Medical Image File Formats", Journal of Digital Imaging volume 27, pages200–206, 2014.

[4] Liting Yan, Yahui Yu, Hongliang Ma, "DICOM standard and Its Application in PACS system", Medical Imaging Process and Technology, Enpress, 2018.

[5] Bhumika Gupta, "Study of Various Lossless Image Compression Techniques", International Journal of Emerging Trends & Technology in Computer Science. Volume 2, Issue 4-2013.

[6] Michele Larobina, Loredana Murino, "Medical Image File Formats", Journal of Digital Imaging volume 27, pages200–206, 2014.

[7] Hardik Jani, Jeegar Trivedi, " A Survey on Different Compression Techniques Algorithm for Data Compression", International Journal of Advanced Research in Computer Science & Technology, Vol.2,Issue 3, 2014.

[8] S.Boopathiraja and P.Kalavathi, "A Near Lossless Three-Dimensional Medical Image Compression Technique using 3D-Discrete Wavelet Transform", International Journal of Biomedical Engineering and Technology, 2019.

[9] S.Boopathiraja, P.Kalavathi and and C.Dhanalakshmi, "Significance of Image Compression and its Upshots – A survey", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2019

[10] S.Boopathiraja, P.Kalavathi and S.Chokkalingam "A Hybrid Lossless Encoding Method for Compressing Multispectral Images using LZW and Arithmetic Coding", International Journal of Computer Sciences and Engineering, May 2018.

[11] S.Boopathiraja and P.Kalavathi, "A Near Lossless Multispectral Image Compression using 3D-DWT with application to LANDSAT Images", International Journal of Computer Sciences and Engineering, ISSN: 2347-2693, Vol.06, No.04, pp.332-336, May 2018.

[12] P.Kalavathi and S.Boopathiraja, "A Medical Image Compression Technique using 2D-DWT with Run Length Encoding ", Global Journal of Pure and Applied Mathematics. ISSN: 0973-9750. Vol.13, no.5, pp. 87-96. September 2017.

[13] P.Kalavathi and S.Boopathiraja, "A Wavelet Based Image Compression with RLC Encoder", Proceedings of the National Conference on Computational Methods, Communication Techniques and Informatics (NCCVNI), pp.68-74. January 2017.