STUDY OF TECHNIQUES FOR LOSSLESS COMPRESSION OF SINOGRAMS

Ishani Vajpai
Jodhpur Institute of Engineering and Technology
Jodhpur, Rajasthan.
ishani.vajpai@gmail.com

Aakriti Gupta
M.B.M. Engineering College
J. N. V. University
Jodhpur, Rajasthan

Abstract— The study of emerging techniques for Image Processing and analysis is essential for understanding the imaging based diagnosis used in the field of biomedical research. Apart from the requirement of a proper communication module between man and the machine, there is an ardent need for lossless compression techniques for the acquisition, transfer and archival medical images. The objective of this paper is to provide an overview of the sinograms and the lossless compression techniques that can be applied to compress them. This paper describes the most effective of the lossless techniques that include the well established conventional techniques as well as the latest emerging techniques. The paper also discusses the need to store sinograms in a compressed form rather than compressed CT (Computed Tomography) scan for better availability of the associated information for a longer period of time due to reduced storage requirements. It can be used for better diagnosis and treatment of the patients without the hazard of unduly exposing patients to radiation based measurements.

Keywords—sinogram, lossless compression techniques

1. **INTRODUCTION**

In order to diagnose or examine the internal parts of the human body, medical imaging is frequently used. Medical imaging often refers to the set of techniques that produce images of internal aspect of the body. It is estimated that till 2010, more than 5 billion medical image studies had been conducted worldwide [1]. X-Rays are used for producing the tomographic images of the required parts of body. As a result of this, there has been a continuous increase in the volume of the data captured. Due to large amount of data, their transmission without compression is prohibitive. JPEG2000 (Joint Photographic Expert Group) standard is being presently employed for compression of scan images. Alternatively, PACS (Picture Archiving and Communication System) systems, that are generally based on dedicated computer which can access data stored in digital image processors of different imaging modalities and transfer these data at high speed within internal or external location of hospital, are used [1]. Another solution is MIPAV(Medical Image Processing, Analysis, Visualization), which is a JAVA application that can run on any JAVA enabled platform such as Windows, Unix, Macintosh OS X [2]. As a result, this data can be transferred globally as per the requirement. On the other hand, these images also require sophisticated computerized quantification and visualization tools for analysis.

The following imaging techniques are used in Medical Diagnosis and Treatment-

1. MRI (Magnetic Resonance Imaging) -It is a powerful tool which uses strong magnetic field to produce images. For example it assesses the rate of hippocampal atrophy.
2. CT-A sensitive diagnosis tool. For example it can be used to diagnose tumors or brain after stroke.
3. PET/CT (Positron Emission Tomography) -This is mainly used in oncology. It can measure the brain’s metabolic activity by measuring regional glucose metabolism.
4. X-Ray-It is a diagnosis tool that uses small amount of ionising radiation. It can be used to

diagnose condition or disease in internal organs or bones.

5. Ultrasound-It is a versatile tool that generates images using high frequency sound waves. For example it can be used to access the stages of pregnancy.

6. Bone densitometry-It is used to calculate mineral matter per square centimetre of bones. For example it can be used to diagnose osteoporosis.

The raw data acquired, consisting of both the ROI (Region of Interest) and non-ROI region is known as sinogram. The sinograms are elaborated further in the next section.

2. SINOGRAMS

The raw projection data obtained in CT scans is termed as ‘sinogram’. This data is named as sinogram as it appears as sinusoidal bands and lines. X coordinate is determined by rotating X Ray tube generator whereas Y coordinate by the position of detector of rays that are transmitted by object. Compression ratio of sinogram is high as compared to that of reconstructed image formed from raw data. Raw file contains data of whole body of patient instead only ROI. This will give the consulting doctor an option to study not only the ROI demarcated by the Radiologist, but also the Non-ROI data if required, during diagnosis. It will obviate the need of performing CT again, if the ROI demarcated by the Radiologist does not match with the requirement of the consulting expert, thus making it not only convenient, but also economical. Further, even though the reconstructed image contains only ROI from sinogram, the whole body gets exposed to radiation. As sinogram contains both ROI and Non-ROI, the memory required by it is 3 to 4 times that of CT scan file. Hence, it would be more effective to use some data compression technique to reduce the memory requirement. It has been observed by researchers that due to their serial nature, sinogram files are compressible to a higher compression ratio, as compared to that obtained from the compressed CT scan images. It could be sent to CT scanner whenever reconstructed image is needed. This will improve the quality of image needed for diagnosis.

3. SINOGRAM DATA ACQUISITION

The data from the human body is acquired by the use of a special type of scanner. Figure 1 describes the basic layout of the gantry of this system. It consists of a fixed frame and a pivot-mounted rotating disk. An X-ray tube and a detector are mounted on the disk that gets powered over a slip ring. The X-ray tube casts X-rays through a measurement field in which the patient is located. The detector consists of an array or matrix of detector elements. The disk rotates with a speed of up to 3.5 rotations per second while the detector records in the order of 1000 projection images per rotation. These get processed and serialized for transfer over a capacitive or optical slip ring by the detector electronics and get stored in a memory buffer. This raw projection data is reconstructed to attenuation images or volumes on a reconstruction computer.[4] It is stored and then is used for diagnosis. Figure 2 shows the path of the captured signal in the data processing system, from point of acquisition to the display. This data is compressed and transferred on a data carrier or over a network. This data is decompressed back to the original form and these images are used for the further diagnosis.

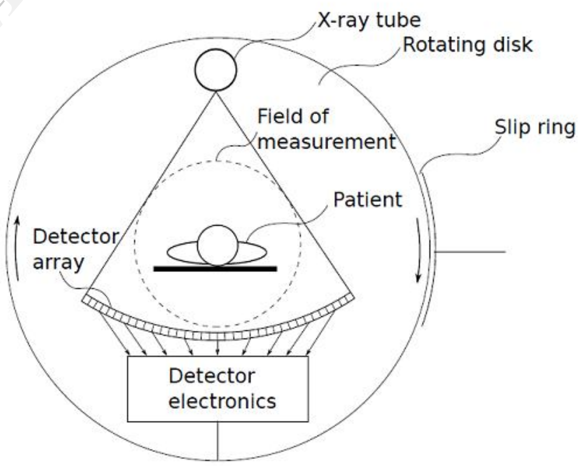

Figure 1 Layout of Gantry [4]

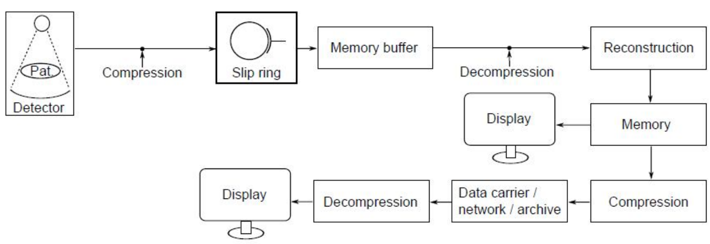

Figure 2 Data Processing Systems [4]

The process of image compression holds a place of very great significance in data processing. The various techniques for sinogram compression are explained in the next section.

4. COMPRESSION TECHNIQUES

The volume and quantity of data required for the analysis and diagnosis is very large and bulky. With the increase in the number of patients, this data is continuously increasing. In order to store and transmit this data, compression of data is carried out. There are two techniques used for compression-
1. Lossless compression- In this compression algorithms data of the image is exactly same bit by bit after decompression. But compression ratio is limited to a range of 1:2- 1:3.
2. Lossy compression- This technique permits loss of data. It has high compression ratio. Compression ratio up to 8:1 does not affect quality of image. For example if compression is done with compression ratio of 12:1, the file of 115MB compresses to 10 MB without affecting the quality of the image.
The lossless techniques are generally preferred over lossy techniques in the field of medical imaging. The major lossless techniques used in this field are as follows:

1. Shannon Entropy Based Compression
2. Entropy Coding Based Compression
3. Huffman Coding Based Compression
4. Deflate Coding Based Compression
5. Differential Algorithm and Entropy Based Compression
6. Spatial Indexing Based Compression

A brief account of these techniques is given below:

1. Shannon Entropy Based Compression:

The Shannon Entropy quantifies the expected value of the information contained in a message, usually in units such as bits [5]. It provides the best possible compression of data. It gives the method in which the data cannot compress more than one bit of entropy per bit of data. The data compressed in this manner is not shorter than its information content.

2. Entropy Coding Based Compression:

All the captured information consists of a lot of data which is redundant. This data can be represented in codes. Short codes are used for symbols with high probability and long codes for symbols with low probability. As a result, this data is compressed by eliminating the redundant data.

3. Huffman Coding Based Compression:

Huffman coding is an algorithm to construct a variable-length code where symbols with higher probability get shorter codes than symbols with lower probability [6]. The Huffman coding has a special concept of prefix codes. No two data can have the same prefix. The data is recognized on the basis of prefixes. The Huffman code is generated by using Binary tree. It is constructed by taking atleast two probable values and representing them by a new symbol that has the probability which is equal to the sum of two symbols.

4. Deflate Coding Based Compression:

It is a well known lossless data compression algorithm. It is known as .zip file format. It combines the Huffman coding and dictionary coding. It consists of 286 symbols. Symbols 0 to 255 represent bytes consisting of 8 bits for direct coding of input data. Symbol 256 denotes the end of a compressed data block and symbols 257 to 285 are the instructions to repeat a part of the data that was already decoded.

5. Differential Algorithm and Entropy Based Compression:

It focuses on the correlation. In this technique the differences between the consecutive sinuses are stored unlike separate codes for each symbol in Huffman coding. These are known as differential sinogram. As a result, redundancy is removed and they are compressed using Entropy based compression.

6. Spatial Indexed Based Compression:

In this technique, temporal information associated with a sinogram as a list of arrival times indexed by sinogram entries. It can be applied to differential sinograms also. The redundancies are removed by using Huffman coding.

The above techniques are the general lossless data compression. The last two, differential algorithm and entropy based compression and spatial indexed compression are new techniques. The compression ratio obtained is comparatively higher in these two techniques.

5. CONCLUSION

This study suggests that raw data associated with the CT images, i.e., sinogram, should be preserved for future use, instead of the images, as is the present practice. Since the sinogram data is usually in the serial form, it can be easily compressed by using lossless compression techniques. Hence, algorithms for implementing lossless compression of sinograms need to be developed. Generally, lossless techniques are preferable to the lossy techniques as complete data cannot be restored in the latter case. The study of lossless compression techniques that have been discussed in this paper reveals that the best two techniques for compression of sinograms are the differential algorithm and entropy based compression and spatial indexed compression. They have higher compression ratio and are hence, preferred over the others.

5. REFERENCES

1. H. U. Lemke, H. S. Stiehl, H. Scharnweber, and D. Jackél “Applications of Picture Processing, Image Analysis and Computer Graphics Techniques to Cranial CT Scans”, Journal of Digital Imaging, Vol. 16, No.1, pg 13–28, March, 2003, Published online April 30, 2003, doi: 10.1007/s10278-002-6030-9.

2. Medical Image Processing, Analysis, and Visualization Available at : http://mipav.cit.nih.gov

3 A. Wegener, “Compressing the Storage Demands for CT Imaging”, Medical Design Technology, Vol 11, Issue 9, pg 15, September,2007.

4. J. Wilhelmy “Lossless and Lossy Raw Data Compression in CT Imaging”. Inaugural-Dissertation Friedrich-Alexander-Universität Erlangen-Nürnberg, December 18,2012.

5. C. E. Shannon “A Mathematical Theory of Communication”, Bell System Technical Journal, 27:379–423, 623–656, 1948.

6. D. Huffman “A method for the construction of minimum-redundancy codes” In Proceedings of the I.R.E., pages 1098–1101, September,1952.