DATA COMPRESSION ON DOMAIN SPECIFIC DATAWAREHOUSE

Supriya*

ABSTRACT

Domain Specific Compression Algorithms are useful when there is specific data/image to be compressed, that should be lossless. Compressing and storing an image is a difficult task as it would be applied to some specific and important areas such as Medical Diagnosis, where the image is used for treatment, Automated Fingerprints Identification System, where the image is used for criminal jurisdiction, skull measurement, etc. The domain specific compression algorithms require exactness and high compression ratio. In this paper comparisons are made between several currently used compression techniques like PPM, BPM, JPEG, GIF, TIF and JPEG2000 by considering numerous medical images of X-rays, Ultrasound etc, and a conclusion is made that JPEG200 is lossless and highly compressed technique which can be used in domain specific data like medical images where requirement of these two features is most important. To fulfill the above requirement a compressor and decompressor for JPEG2000 is implemented using JAVA. It has been observed that it is the best technique among several compressing techniques which are used currently on the basis of high compression ratio and lossless.

Keywords: *JPEG2000, JPEG, BPM, PPM, compression ratio.*

*Assistant Professor, College of Engineering Roorkee, Roorkee

INTRODUCTION

When an image or a sequence of images are to be transmitted over a network or a telephone line to a number of diagnosis centers around the globe to guarantee the proper diagnosis the Image compression play an important role. The problem, however, is that while high compression rates are desired, the usefulness of the image depends on preserving significant characteristics of the original image after the decompression has been done .Medical image compression should be lossless compression where all the relevant image information is preserved and effective diagnosis is possible than lossy compression where the information transmission and storage are very good but there is no guarantee that the information was preserved for processing and diagnosis [8],[6],[3].

MEDICAL DIAGNOSIS SYSTEM ARCHITECTURE

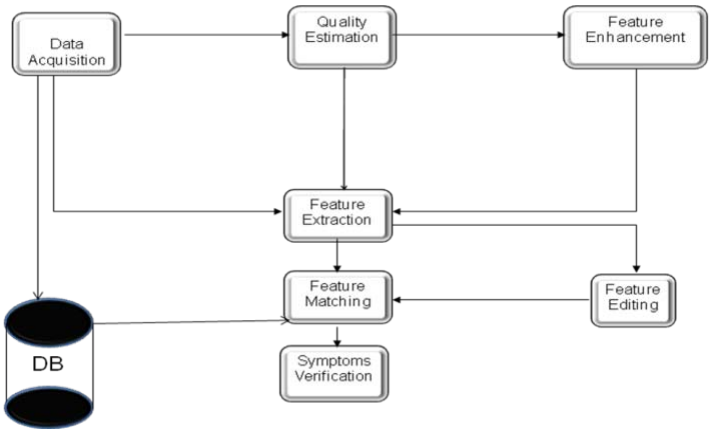

Figure1 Medical Diagnosis System Architecture

Motivation behind Compression

IMAGE TYPE	SIZE	BITS/ PIXEL	UNCOMPRESSED SIZE	TRANSMISSION BANDWIDTH	TRANSMISSSION TIME(28.8K MODEM)
GRAYSCALE	512×512	8	262 Kbytes	2.1 Mbit/image	1 min 13 sec
COLOR	512×512	24	786 Kbytes	6.29 Mbit/image	3 min 39 sec
MEDICAL	2048×1680	12	5.16 Kbytes	41.3 Mbit/image	23 min 54 sec
SHD	2048×2048	24	12.56 Kbytes	100 Mbit/image	58 min 15 sec

Table 1 .Motivation behind Compression

The motivation for the compression of images is illustrated through the use of Table1. This table shows the storage size, transmission bandwidth, and transmission time needed for various types of uncompressed images. It is clear from these values, that images require much storage space, large transmission bandwidths, and long transmission times. With the present state of technology, the only solution is to compress images before their storage and transmission. Then, at the receiver end, the compressed images can be decompressed [7],[2].Some of the advantages of Image Compression are Reduction in Storage, Rate (Data Transfer), Data Security, Backup and Recovery, Performance. Disadvantages are Overhead, Disruption, Portability, Output length, Reliability, Storage [9],[1].

JPEG2000 ARCHITECTURE

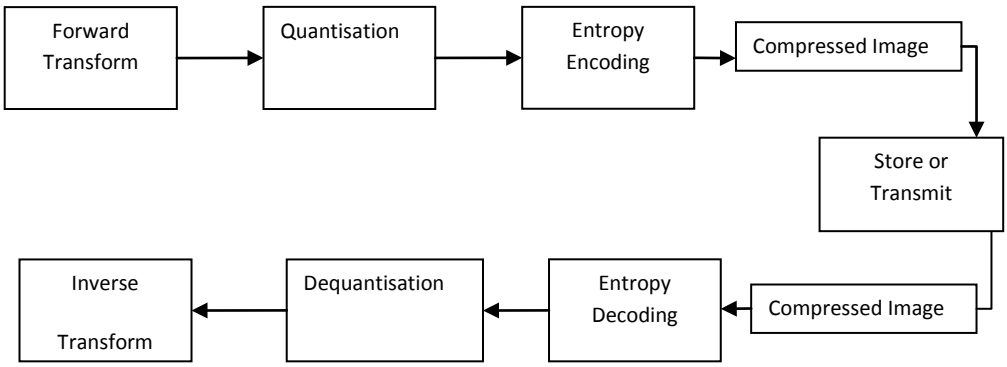

Figure.2 Block diagram of the JPEG2000 process

The encoding procedure is as follows:

1. The source image is decomposed into components. The image components are (optionally) decomposed into rectangular tiles. The tile-component is the basic unit of the original or reconstructed image.
2. A wavelet transform is applied on each tile. The tile is decomposed into different resolution levels.
3. The decomposition levels are made up of sub-bands of coefficients that describe the frequency characteristics of local areas of the tile components, rather than across the entire image component.
4. The sub-bands of coefficients are quantized and collected into rectangular arrays of "code blocks."
5. The bit planes of the coefficients in a code block (i.e., the bits of equal significance across the coefficients in a code block) are entropy coded.

6. The decomposition levels are made up of sub-bands of coefficients that describe the frequency characteristics of local areas of the tile components, rather than across the entire image component.
7. The sub-bands of coefficients are quantized and collected into rectangular arrays of "code blocks."
8. The bit planes of the coefficients in a code block (i.e., the bits of equal significance across the coefficients in a code block) are entropy coded.
9. The encoding can be done in such a way that certain regions of interest can be coded at a higher quality than the background.
10. Markers are added to the bit stream to allow for error resilience.
11. The code stream has a main header at the beginning that describes the original image and the various decomposition and coding styles that are used to locate, extract, decode and reconstruct the image with the desired resolution, fidelity, region of interest or other characteristics [3,4,5].

COMPARITIVE STUDY

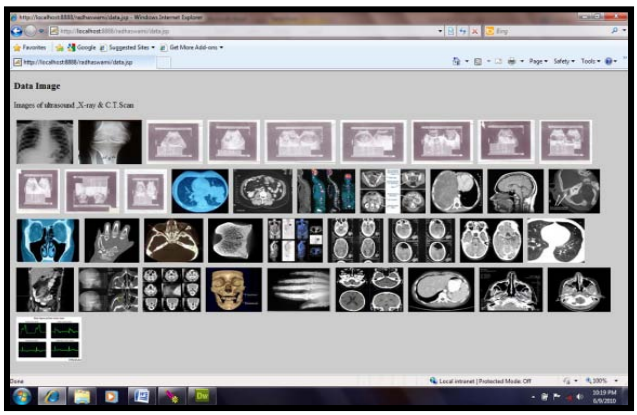

Figure 3 Real Medical Images.

A number of real medical images of X-rays, ultrasound, etc as shown in figure 3 are collected and Comparative analysis is presented in following tables, on the basis of compression ratio and lossless property. These are important features for domain specific areas.

As the table 2 reflects the image of original image retain after decompressed, which represent that this compression is lossless. This is most preferred feature of compression for medical diagnosis. In the above table 30 real medical images is consider and it has been observed that not a single bit of data is loss during the process of compression and decompression.

Table 2. Compare Result after Decompression

	IMG1(KB)	IMG2(KB)	IMG3(KB)	IMG3(KB)	IMG4(KB)	IMG6(KB)	IMG7(KB)	IMG8(KB)
ORIGINAL IMAGE	126.33	1436.23	3603.38	3361.12	1424.76	1424.76	1424.2	1628.16
COMPRESSED	15	17	38	36	16	16	16	17
DEOMPRESSED	126.33	1436.23	3603.38	3361.12	1424.76	1424.76	1424.2	1628.16
	IMG9(KB)	IMG10(KB)	IMG11(KB)	IMG12(KB)	IMG13(KB)	IMG13(KB)	IMG14(KB)	IMG16 (KB)
ORIGINAL IMAGE	1424.4	1492	1424	1802.23	47	31.3	37.2	30.3
COMPRESSED	16	17	12	14	1	1	1	1
DEOMPRESSED	1424.4	1492	1424	1802.23	47	31.3	37.2	30.3
	IMG17(KB)	IMG18(KB)	IMG19(KB)	IMG20(KB)	IMG21(KB)	IMG22(KB)	IMG23(KB)	IMG23(KB)
ORIGINAL IMAGE	37.3	39.3	36.4	34.1	26.8	33.2	39.4	33.6
COMPRESSED	1	1	1	1	1	1	1	1
DEOMPRESSED	37.3	39.3	36.4	34.1	26.8	33.2	39.4	33.6
	IMG24(KB)	IMG26(KB)	IMG27(KB)	IMG28(KB)	IMG29(KB)	IMG30(KB)		
ORIGINAL IMAGE	233	39.6	39.6	32.3	37.3	39.3		
COMPRESSED	1	1	1	1	1	1		
DEOMPRESSED	233	39.6	39.6	32.3	37.3	39.3		

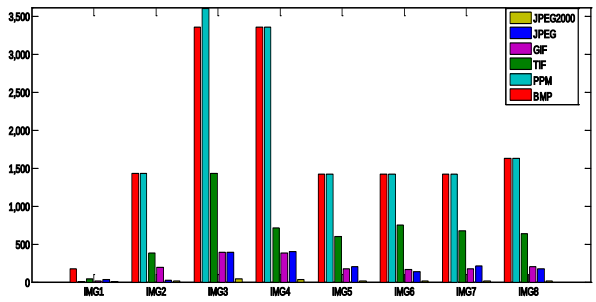

Figure4. Comparision Of Compression Rate Of Difference File Formate

In table 3 different currently used image compression techniques like TIF, GIF, PPM, BPM, JPEG are compared with JPEG2000 through compressing these real medical images and it has been observed that the compression ratio of JPEG200 higher than all the others compression techniques.

Table 3. Compare Result of Various Formats with JPEG2000

	IMG1(KB)	IMG2(KB)	IMG3(KB)	IMG3(KB)	IMG4(KB)	IMG6(KB)	IMG7(KB)	IMG8(KB)
JPEG	33.3	24.2	392	403	199	140	213	173
PPM	1.33	1436.23	3603.38	3361.12	1424.76	1424.76	1424.2	1628.16
TIF	38.3	380	1436.2318	710	600	747	680	634
GIF	13.3	193	389	380	173	166	172	200
BMP	179	1436.23	3361.76	3360	1424.76	1424.7	1424.2	1628.16
JPEG2000	**2**	**17**	**38**	**36**	**16**	**16**	**16**	**17**
	IMG9(KB)	**IMG10(KB)**	**IMG11(KB)**	**IMG12(KB**	**IMG13(KB)**	**IMG13(KB)**	**IMG14(KB)**	**IMG16(KB)**
JPEG	163	166	164	176	3.69	3.36	3.37	3.31
PPM	1424.4	1492	1424	1802.23	47	31.3	37.2	30.3
TIF	633	636	692	38.4	18 .7	31.3	18.1	14.1
GIF	179	172	183	199	8.89	4.72	7.7	4.31
BMP	1424.4	1492	1424	1802.23	47	31.3	37.3	30.73
JPEG2000	**16**	**17**	**12**	**14**	**1**	**1**	**1**	**1**
	IMG17(KB)	**IMG18(KB)**	**IMG19(KB)**	**IMG20(KB)**	**IMG21(KB)**	**IMG22(KB)**	**IMG23(KB)**	**IMG23(KB)**
JPEG	2.47	3.16	2.31	3.09	2.87	1.98	3.49	3.18
PPM	37.3	39.3	36.4	34.1	26.8	33.2	39.4	33.6
TIF	17.3	13.6	20.3	28.3	10.8	30.7	26.9	26.9
GIF	3.68	4.33	3.47	4.33	4.29	3.24	7.19	6.43
BMP	37.4	39.3	36.6	34.3	27.1	33.3	33.8	224
JPEG2000	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	IMG24(KB)	**IMG26(KB)**	**IMG27(KB)**	**IMG28(KB)**	**IMG29(KB)**	**IMG30(KB)**		
JPEG	23.9	3.81	3.77	3.96	3.39	2.86		
PPM	233	39.6	39.6	32.3	37.3	39.3		
TIF	128	23.3	23.4	32.6	16.8	27.3		
GIF	34.6	7.47	7.44	7.64	3.72	4.43		
BMP	39.9	39.9	32.7	37.3	39.3	30		
JPEG2000	**1**	**1**	**1**	**1**	**1**	**1**		

Similar pictorial results are shown by the figure4 given below. This graph clearly reflects the effectiveness of JPEG2000 in terms of compression ratio on large size images. For example in image IMG3 the size of BMP and PPM format is in the range of 3,000-3,500Kb, but for PEG2000 format it only 2 Kb.

DISCUSSION

By considering above 30 images of medical diagnosis, it is clearly observed that JPEG2000 reduces the size of image too much smaller than any other formats and it retain its size and quality after decompression. Causes reduction of storage, transmission time and bandwidth

for transmission, which are crucial resources for transmission of large images in medical field.

RESULTS AND CONCLUSION

JPEG2000 is the new standard for still image compression in Medical Field that are large and high quality imaging, required in diagnostically important regions. It provides a new framework and an integrated toolbox to better address increasing needs for compression and functionalities for still image applications, like medical imagery , Internet, color facsimile, printing, scanning, digital photography, remote sensing, mobile applications, digital library and E-commerce. Lossless and lossy coding embedded lossy to lossless, progressive by resolution and quality, high compression efficiency, error resilience and lossless color transformations are some of its extended features. Comparative results show that JPEG2000 is indeed superior to existing still image compression standards. It can form an effective system (high compression rate) and accurate system (lossless) in medical field .It is expected that JPEG2000 standard will become the one of most widely used medical image compression standard in future.

REFERENCES

[1] Abramson, N. (1963).” *Information Theory and Coding”*. McGraw-Hill, New York

[2] Bijay Shrestha, Dr. Charles G. O’Hara, Dr. Nicolas H. Younan (2004), “JPEG2000: IMAGE QUALITY METRICSASPRS 2004 Annual Conference,’Geospatial Goes Global: From Your Neighborhood to the Whole Planet’Baltimore, Maryland .

[3] Christophe, Pierre (2002),” JPEG 2000, an adapted compression method for ultrasound images, A comparative study”, Laboratoire Vision et Robotique.

[4] Christopoulos, A. Skodras, and T. Ebrahimi (2000), ‘The JPEG2000 Still Image Coding System: An Overview’, *IEEE Trans. Cons. Elect.*, Vol. 36., No. 3, pp. 1103 – 1127.

[5] David Taubman1, Erik Ordentlich2, Marcelo Weinberger, Gadiel Seroussi (2001)”Embedded Block Coding in JPEG2000”, Information Theory Research Group, HP laboratories Palo Alto, HPL-2001-34.

[6] Givon Zirkind (2003),”AFIS Data Compression: An Example of How Domain Specific Compression Algorithms Can Produce Very High Compression ratios,”ACMSIGGRAPH, Publications, Computer Graphics Quarterly, Volume 31, number 3.

[7] Rountree, J. C., Webb, B. N., & Marcellin, M. W. 2003, in ASP Conf. Ser., Vol. 314 Astronomical Data Analysis Software and Systems XIII, eds. F. Ochsenbein, M. Allen, & D. Egret (San Francisco: ASP), 554

[8] S. E. Ghrare, M. A. M. Ali, M. Ismail (2008), "The Effect of Image Data Compression on the Clinical Information Quality of Compressed Computed" Tomog European Journal of Scientific Research, ISSN 1340-216X Vol.23 No.1, pp.6-12.

[9] Taubman and M. Marcellin, (2002), "JPEG2000: Standard for Interactive Imaging," *Proc. of the IEEE*, Vol. 90, No. 8, pp. 1336 – 1347.