Image Compression of medical images using VQ-Huffman Coding Technique

Kavinder[1], Vinay Chopra[2], Harsimranjeet Kaur[3]
[1]Student DAVIET Jalandhar
[2]AP DAVIET Jalandhar
[3]Sr. Lect. CEC, Landran

Abstract- *Digital Medical Imaging has grown very fast in recent years and hence plays a vital role in diagnosis, treatment, and research area. All the radiological modalities such as CT scanners, MRI, US, PET, X-Ray made by multiple vendors and located at one or many sites can communicate by means of DICOM across an network. Now days, hospitals need to store large volume of data about the patients that require huge hard disk space and high bandwidth. This would employ the need to compress DICOM images for efficient storage and transmission over the internet. In this paper, a new compression algorithm combining the features of both lossy (DCT) and lossless (Huffman Coding) compression techniques has been designed and implemented. The performance of proposed algorithm is then improved using Vector Quantization technique in the context of increasing Compression Ratio as well as preserving the quality of compressed images. Different quality metrics like MSE, PSNR and CR are computed on various medical test images. The experimental results show that proposed compression technique performs better than the existing techniques in terms of performance parameters.*

Keywords- *DICOM (Digital Imaging and Communication in medicine), DCT (Discrete Cosine Transform), Huffman Coding, Vector Quantization, PSNR (Peak Signal to Noise Ratio), MSE (Mean Square Error) and CR (Compression Ratio).*

1. INTRODUCTION TO DICOM

DICOM (Digital Imaging and Communication in Medicine) is a popular industry standard developed by ACR/NEMA to aid the distribution and viewing of medical images, such as CT-Scans, MRIs, X-Rays and Ultrasounds etc. The ACR is the American college of Radiology and NEMA is the National Electrical Manufacturers Association. DICOM is especially used for handling, storing, printing and transmitting information in medical imaging. This standard allows different manufacturers equipment to efficiently communicate medical information via computers as shown in Figure 1.1.

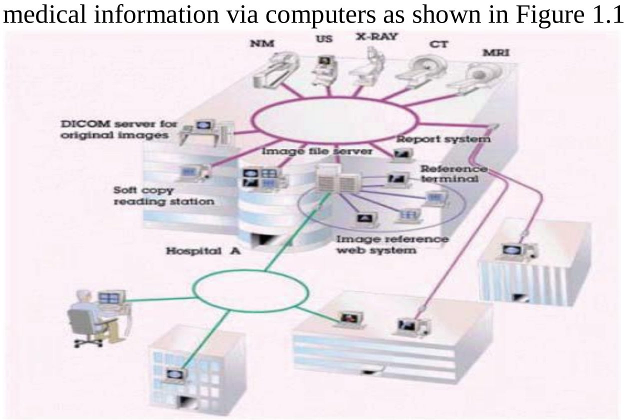

Figure 1.1 DICOM as standard of medical image format

A single DICOM file contains both a header (which stores information about the patient's name, the type of scan, image dimensions, etc) as well as all of the image data (which contains information in three dimensions) [24]. DICOM differs from other data formats in that it groups information together into a Data Set. DICOM specifies that information for medical images is defined in Information Object Definition (IOD) Model. IOD defines information object consisting of a number of attributes such as patient ID, patient's name, examination date etc. and one special attribute containing the image pixel data as shown in Figure 1.2. In the example, Tag (0010, 0020) is representing patient's ID. VR (Value Representation) specify the characteristics of information. VR field is represented as PN for patient name and DA for examination date. Length field displays the length of data in value field and Value field corresponds to actual data. One data element communicates one attribute and several data elements must be combined to make any person IOD instance [3].

DICOM increases the security and integrity of medical image archives retained for legal and compliance reasons. As size of medical images is very large, transmission bandwidth is not sufficient to send such a huge amount of image data so if we compress these medical image files before transfer then it becomes very easy to communicate such files through the network with lower bandwidth and high speed.

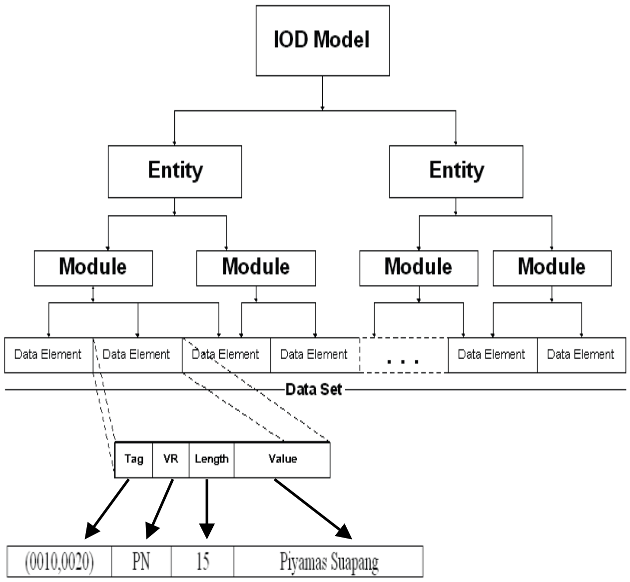

Figure 1.2 Information Object Definition (IOD) Model

2. IMAGE COMPRESSION

Image Compression deals with techniques for reducing the storage space required for saving an image and the bandwidth required for transmitting it. It is basically minimizing the size of a graphics file without degrading the quality of the image to an unacceptable level. The reduction in file size allows more medical images to be stored in a given amount of disk space. It also reduces the transmission time required for images to be sent over the Internet or downloaded from Webpage. Image data can be compressed using a variety of standards, including JPEG, JPEG2000, SPIHT, ROI, DCT and Huffman coding[1][2]. Various factors involved in the design of compression schemes are the amount of distortion introduced (if using a lossy compression scheme), including the degree of compression and the computational resources required to compress and decompress the data.

2.1 Classification of Image Compression Schemes

Image compression schemes are generally classified as lossless compression schemes and lossy compression schemes.

2.1.1 Lossless compression scheme

Lossless compression is an error free compression where the original data can be recovered after decompression. This scheme achieves low compression ratio but has several applications like compression of medical images where the loss of information is not acceptable [11].

2.1.2 Lossy compression scheme

In Lossy compression, some extend of the original data is lost during compression and only an approximation of original data is obtained after decompression. Lossy schemes achieves higher compression ratio and are used in applications, like compression of natural images where perfect reconstruction is not essential and we can afford the partial data loss as long as it is within tolerance [4].

2.2 Image Compression Techniques

2.2.1 Discrete Cosine Transform

DCT represents an image as a sum of sinusoids of varying magnitudes and frequencies. It works on the concept of separating image into parts of different frequencies. DCT cuts the image into blocks of 8×8 pixels, processes each block independently, shifting & simplifying the pixels so that there is less data to encode. Compression is achieved by quantizing large amount of high frequency components in the image [2].

2.2.2 Compressive Sensing

Compressive sensing also called Compressive Sampling or Sparse Sampling method is closely connected to transform coding technique in which sparse signals that contains coefficients close to or equal to zero are used. This technique is used as a part of quantization in which an average gray value that replaces the block of nearest neighboring gray values in the transformed image is found. In this method, compression is achieved by the elimination of sparse signal redundancy [7].

2.2.3 Vector Quantization

Vector quantization maps k-dimensional input vectors in the vector space to a finite set of output vectors. Each vector is called codeword and set of all codewords is called codebook. VQ encoder takes an input vector and outputs an index of the codeword that offers lowest distortion (closest codeword). At receiving side, VQ decoder replaces the received index with associated codeword and outputs the codeword [15].

2.2.4 Huffman Coding

It is also called as Variable length Coding in which coding redundancy has been reduced. It is based on the frequency of occurrence of pixel values in the image. This technique is to assign smaller codes to more probable (most frequently occurred) gray values than less probable ones that occur less frequently. Huffman codes can be properly decoded because they obey the prefix property which means that no code can be a prefix of another code, so the complete set of codes can be represented as a binary tree, known as a Huffman tree [6].

3. PROPOSED TECHNIQUE FOR DICOM IMAGE COMPRESSION

In the recent time, various lossy and lossless image compression techniques are emerged during the development in DICOM image compression field. The demand of higher compression ratio is ever increasing for the efficient storage and transmission of medical images due to increasing bandwidth requirement. As the compression ratio increases, the quality of the resulting image degrades. So, a tradeoff between compression ratio and the tolerance in the visual quality degradation need to be considered during image compression. This would imply the need for a compression scheme that would give high compression ratio as well as provide the best quality images [4][7].

A lot of work has already been done in the analysis of lossy and lossless compression techniques individually. Since their combination can give better results in terms of compression that further helps in storing and transmitting medical images on limited bandwidth [1]. So, implementation of a technique combining the benefits of both these image compression methods is important. Thus, main objective of the research is to design and implement new hybrid compression algorithm for compressing medical images which performs better than the existing techniques in terms of various performance parameters like Peak Signal to Noise Ratio, Mean Square Error and Compression Ratio.

Huffman Coding, a lossless technique to compress images is rather a very good technique to be used for compressing DICOM images as it gives high PSNR value, but it gives less compression ratio [6]. On the other hand, lossy techniques like DCT and Vector Quantization gives high compression ratio and low PSNR [7]. In this proposed work, a new compression algorithm which combines the features of both DCT (lossy) and Huffman Coding (lossless) techniques is designed and implemented and performance of the proposed technique is then improved using Vector Quantization. This results in achieving good PSNR and high CR value. As a result, good quality images are developed at higher compression rate [2][4]. The proposed technique attempts to deal with the problem of storing and transmitting medical images over the internet by combining the features of lossy and lossless image compression techniques to create an optimum solution.

1. Image loading in Memory: Image is get loaded into the MATLAB application through Image Acquisition toolbox. Using the GUI Interface for loading image in dicom format as in jpeg, bmp or in any other format.
2. DCT applied for basic compression – This lossy compression technique is used for basic level image compression, by using its inbuilt compression function. DCT cuts the image into blocks of 8x8 pixels. In this step, oscillating frequencies in the image are set to 0 according to predefined DCT level.
3. Vector Quantization applied for improvement: Compressive Sensing is first applied as a part of Vector Quantization compression that includes finding average gray level values for representing each block of nearest neighboring pixel values (excluding pixels having 0 density values) in the transformed image respectively. In this step of applying VQ, all pixel values in the block are replaced with quantized coefficients. This lossy technique is used to improve the performance of Huffman Coding technique by increasing the compression ratio.

The methodology adopted for the proposed algorithm is given below in the flowchart.

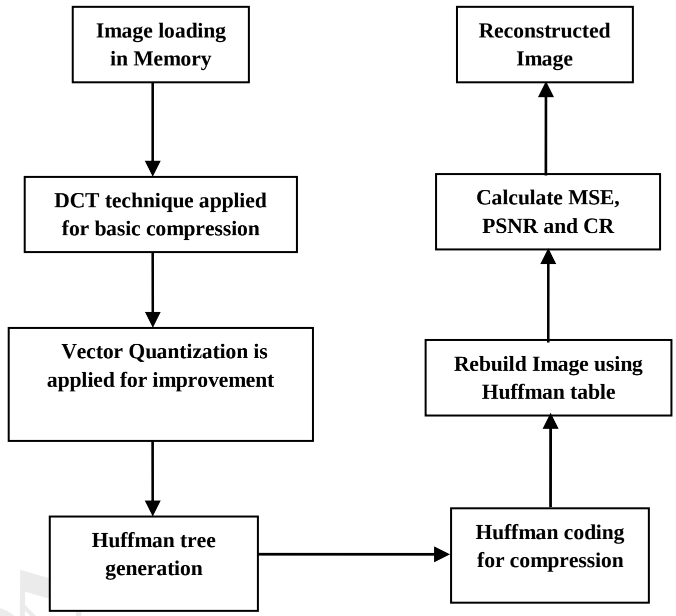

Figure 1.3 Block Diagram of proposed work

4. Huffman Tree Generation: This lossless compression technique varies the length of the encoded symbol in proportion to its information content. The more often a symbol or pixel is used, the shorter the binary string used to represent it in the compressed stream [5]. This step includes plotting of probability values and finding unique codes for individual probabilities whose sum is equal to 1. Complete set of codes can be represented as a binary tree, known as a Huffman tree.
5. Compression using Huffman Coding: This step includes implementing previously generated Huffman tree to the quantized image for compression, by replacing gray values in the image with the unique codes on the basis of their probabilities, both given in the Huffman probability table.
6. Rebuild image using Huffman Table: Image will be reconstructed through decompression process in which firstly inverse DCT is applied followed by vector dequantization and finally Huffman decoding process is applied using already generated Huffman table.
7. Calculate MSE, PSNR and CR: In the final step, various quality metrics like MSE (Mean Square Error), PSNR (Peak Signal to Noise Ratio) and CR (Compression Ratio) are calculated to measure the quality of the image and both subjective and objective results are obtained. The MSE is used to compute the ratio between the maximum possible power of a signal and the power of corrupting noise. The lower the value of MSE, the lower the error. The PSNR

computes the peak signal-to-noise ratio, in decibels, between two images. The higher the PSNR, the better the quality of the reconstructed image. The Compression Ratio (CR) is used to calculate the rate at which image is compressed. The higher the CR, the lower the bandwidth requirement.

$$Compression\ Ratio = N_1 / N_2$$

N_1 is No. of bits to represent compressed image and N_2 is No. of bits to represent original image

The block first calculates the Mean Square Error using the following equation:

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N}$$

Then the block computes the Peak Signal to Noise Ratio using the following equation:

$$PSNR = 10 \log_{10}\left(\frac{R^2}{MSE}\right)$$

4. RESULTS AND DISCUSSION

Both types of subjective and objective results are computed on various types of medical test images. Objective results of the images are evaluated and compared on the basis of being compressed with and without the application of Vector Quantization technique. Following are the subjective results obtained by compressing two medical test images such as X-Ray and CT-Scan images using DCT technique followed by Vector Quantization and then finally applied Huffman Coding compression technique. Vector Quantization is used to improve the performance of Huffman Coding technique in the context of improving compression ratio. Each medical test image is having three resultant figures i.e., Figure 4 to 6 for X-Ray image and Figure 7 to 9 for CT-Scan image, representing an original image, transformed image using DCT technique and finally compressed image using Huffman Coding with Vector quantization technique respectively and their corresponding histograms displayed along with the resulting images.

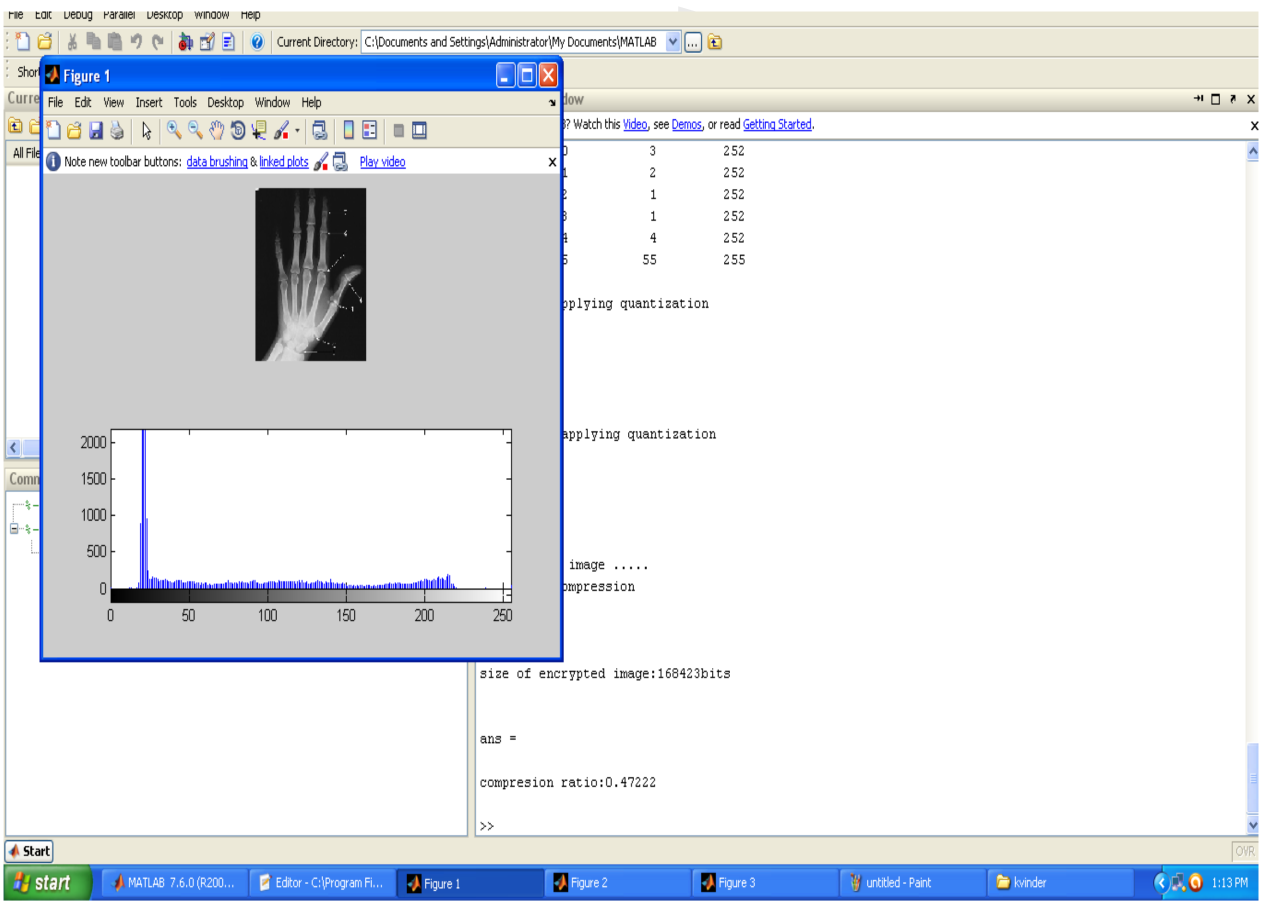

Figure 4: Results showing the original X-Ray image along with its Histogram.

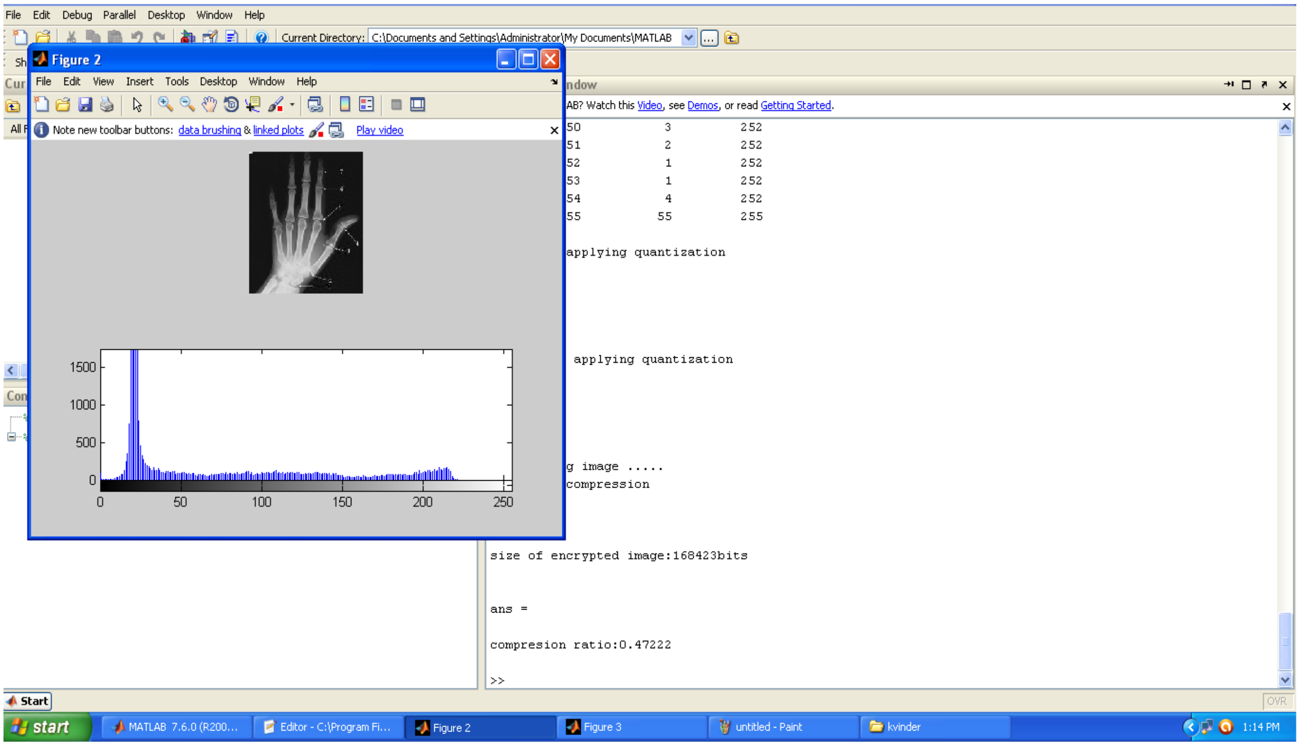

Figure 5: Results showing the X-Ray image after applying DCT along with its Histogram.

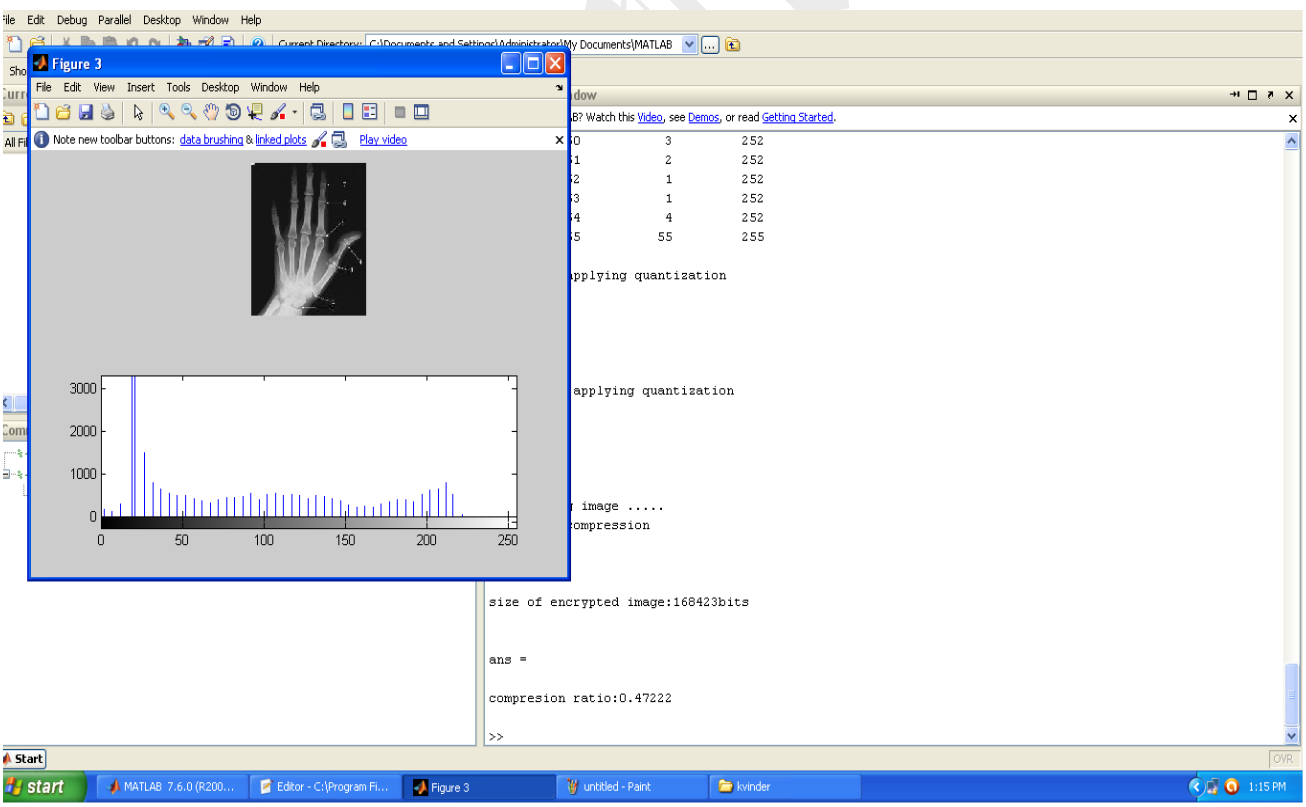

Figure 6: Results showing the X-Ray image compressed using Huffman Coding with Vector Quantization along with its Histogram.

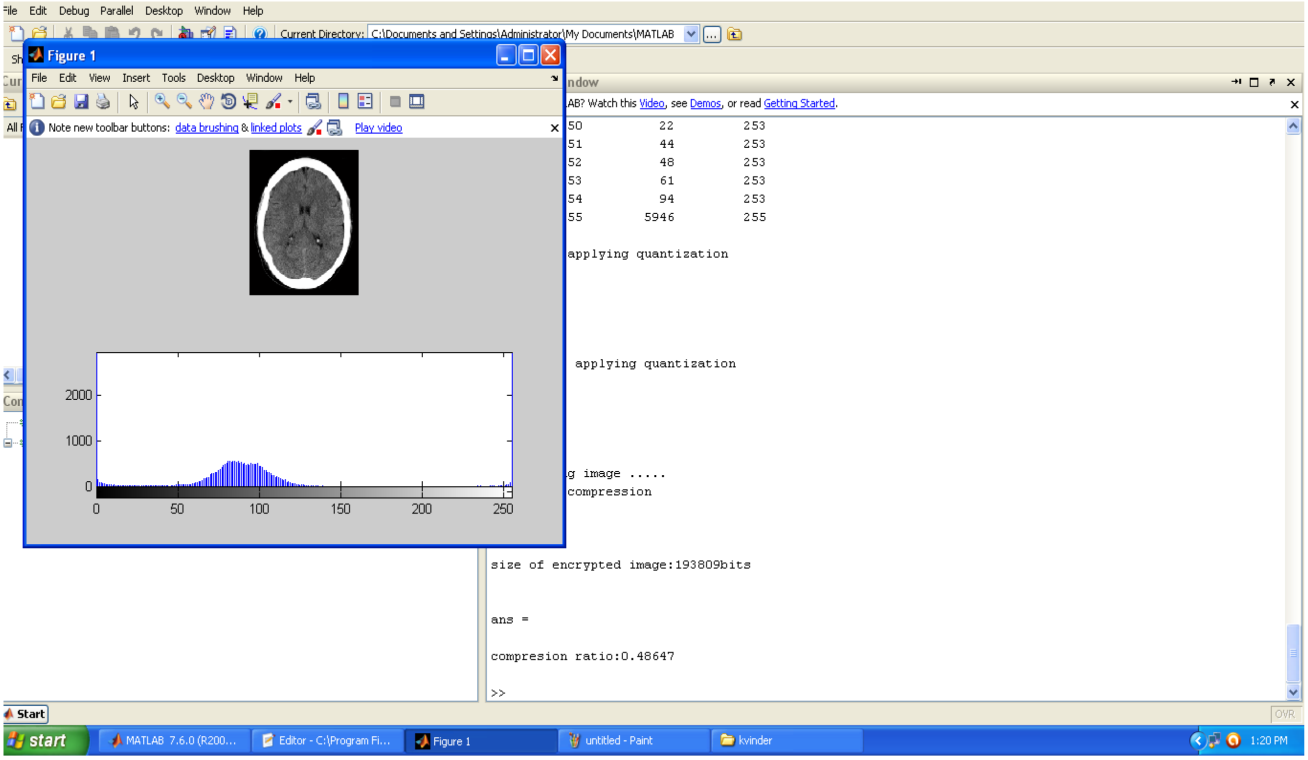

Figure 7: Results showing the original CT-Scan image along with its Histogram.

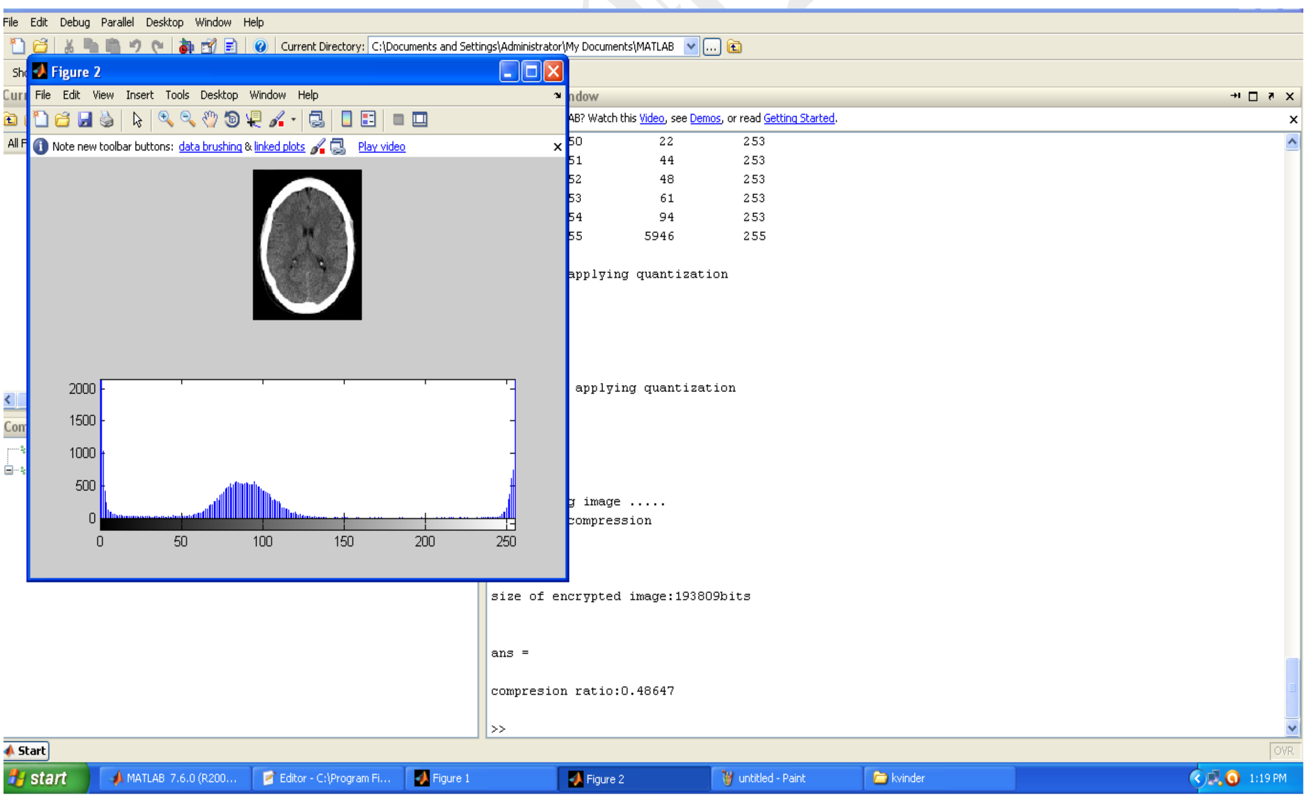

Figure 8: Results showing the CT-Scan image after applying DCT along with its Histogram.

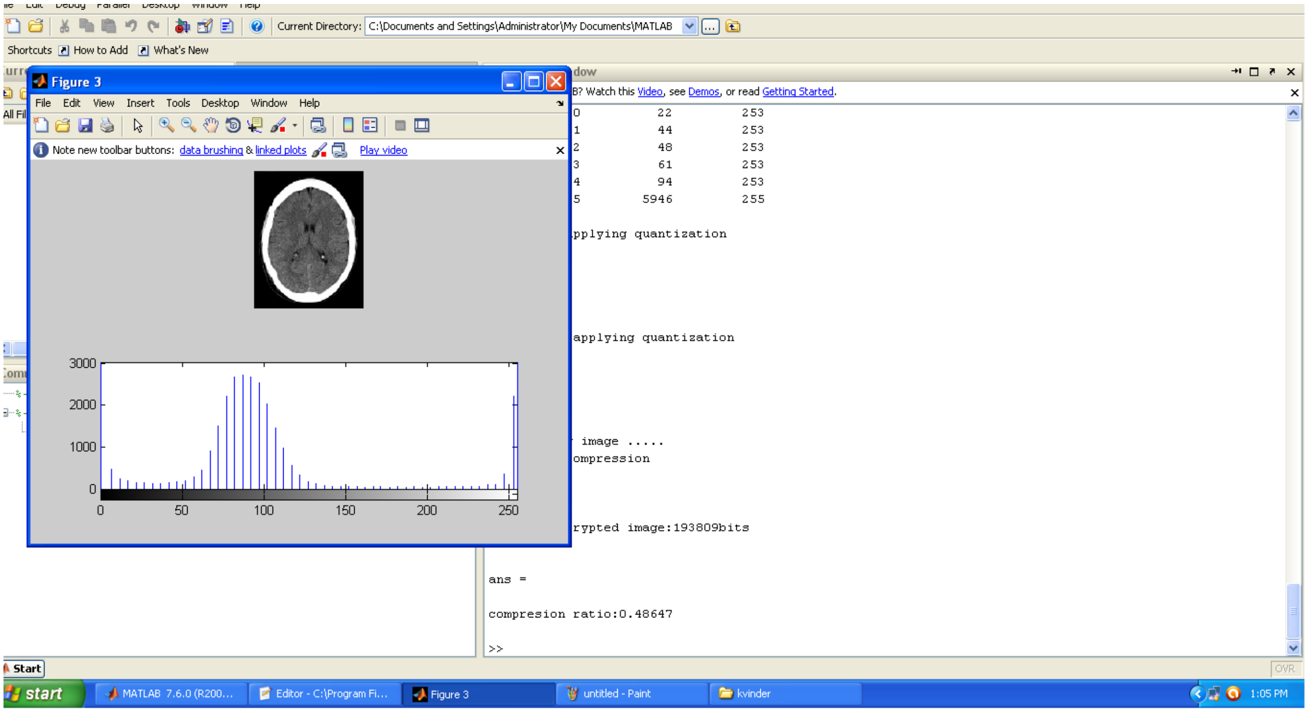

Figure 9: Results showing the CT-Scan image compressed using Huffman Coding with Vector Quantization along with its Histogram

Following tables present the objective results showing the values of different quality measure parameters such as Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR) and Compression Ratio (CR) calculated on various medical test images. Table 1 shows the results obtained after applying DCT and Huffman Coding technique only. Table 2 shows the results obtained after applying DCT followed by Vector Quantization and then finally applied Huffman Coding technique.

Table 1: Results showing the values of different quality metrics calculated on X-Ray and CT-Scan test images (without using VQ).

QUALITY METRICS	X-Ray Image	CT-Scan Image
MSE	0.0159	0
PSNR	66.1099	Inf
CR	.7397	.70486

Table 2: Results showing the values of different quality metrics calculated on X-Ray and CT-Scan test images (with using VQ).

QUALITY METRICS	X-Ray Image	CT-Scan Image
MSE	1.6491	1.4627
PSNR	45.9584	46.4792
CR	.47222	.48647

5. CONCLUSION AND FUTURE SCOPE

In the proposed technique, a new compression algorithm has been designed and implemented in which DCT technique has applied for basic level compression, followed by Vector Quantization technique used to improve the performance in the context of increasing compression ratio. Finally, Huffman Coding technique has been applied for compressing medical test images such as X-Ray and CT-Scan images. From the objective results presented in Table 1 and 2 respectively, it is concluded that compression ratio has been improved with the application of Vector Quantization technique before using Huffman Coding technique. As a result, good quality images are developed at higher compression rate. This achieves in efficient storage and transmission of medical images at limited memory space and bandwidth requirement. For the future scope, further research work could be done on other such combinations of lossy and lossless compression techniques. Some work could also be done on quality metric PSNR for more improvement, as it is slightly decreased when compression ratio increases but that is within tolerance of visual quality of medical images.

6. REFERENCES

[1] G. Murugan and Dr. Kannan, "*Low Ramification near Lossless Image Compression Technique*", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, pp: 447-452, April 2012.

[2] Nivedita, Pardeep Singh and Sonika Jindal, "*A Comparative Study of DCT and DWT-SPIHT"*, IJCEM International Journal of Computational Engineering and Management, Volume 15, Issue 2, pp: 26-32, March 2012.
[3] Piyamas Suapang, Kobchai Dejhan and Surapun Yimmun, "*A Web-based DICOM-Format Image Archive, Medical Image Compression and DICOM Viewer System for Teleradiology Application*", SICE Annual Conference 2010, pp: 3005-3011, August 2010.
[4] Yen-Yu Chen and Shen-Chuan Tai, "*Embedded Medical Image Compression Using DCT Based Subband Decomposition and Modified SPIHT Data Organization*", Proceedings of the Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE'04), pp:1-8, © 2004 IEEE.
[5] Mamta Sharma, "*Compression Using Huffman Coding* ", IJCSNS International Journal of Computer Science and Network Security, VOL.10 Issue No.5, pp: 133-141, May 2010.
[6] Mridul Kumar Mathur, Seema Loonker and Dr. Dheeraj Saxena, "*Lossless Huffman Coding Technique For Image Compression And Reconstruction Using Binary Trees",* IJCTA International Journal of Computer Technology and Applications, Volume 3 Issue No. 1, pp: 76-79, Jan-Feb 2012.
[7] S. Kadambe and J. Davis, "*Compressive Sensing and Vector Quantization based Image Compression*", Proceedings of the Forty Fourth IEEE Conference on Signals, Systems and Computers (ASILOMAR), pp: 2023-2027, Nov 2010.
[8] Armando Manduca and Amir Said, "*Wavelet Compression of Medical Images with Set Partitioning in Hierarchical Trees*", Proceedings of the International Conference on Engineering in Medicine and Biology Soc. (EMBS'2011), pp: 1-9, © 2011 IEEE.
[9] Yen-Yu Chen, "*Medical Images Compression for Remote Diagnosis Using Modified SPIHT Data Organization and Fidelity Enhancement Filter*", National Science Council of the Republic of China, Vol. 17, pp: 49–61, 2007.
[10] Dr. S. Shenbaga Devi and K. Vidhya, "*Development of Medical Image Compression Techniques",* Proceedings of the International Conference on Computational Intelligence and Multimedia Applications, pp: 97-101, © 2007 IEEE.
[11] Jagadish H. Pujar and Lohit M. Kadlaskar, "*A New Lossless Method Of Image Compression And Decompression Using Huffman Coding Techniques*", Journal of Theoretical and Applied Information Technology, pp: 18-23, JTAIT 2010.
[12] Puja Bharti, Dr. Savita Gupta and Ms. Rajkumari Bhatia, "*Comparative Analysis of Image Compression Techniques: A Case Study on Medical Images",* Proceedings of the International Conference on Advances in Recent Technologies in Communication and Computing, pp: 820-822, October 2009.
[13] Nivedita and Sonika Jindal, "*Performance Analysis of SVD and SPIHT Algorithm for Image Compression Application",* International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2 Issue No. 2, pp: 1-6, February 2012.
[14] K. Vidhya and Dr. S. Shenbagadevi, "*Performance Analysis of Medical Image Compression"*, Proceedings of the International Conference on Signal Processing Systems, pp: 979-983, ICSPS 2009 IEEE.
[15] Rupinder Kaur and Nisha Kaushal, "*Comparative Analysis of Various Compression Methods for Medical Images",* pp: 1-6, NITTTR 2008.
[16] Sungdae Cho, Dongyoun Kim and William A. Pearlman, "*Lossless Compression of Volumetric Medical Images with Improved 3-D SPIHT Algorithm",* Center for Image Processing Research, pp: 1-16, 2009.
[17] Armando Manduca, "*Medical Image Compression with Set Partitioning in Hierarchical Trees*", 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp: 1224-1225, 1996.
[18] M.A. Ansari and R.S. Anand, "*Recent Trends In Image Compression And Its Application In Telemedicine And Teleconsultation*", XXXII National Systems Conference, pp: 59-64, 17-19 December 2008.
[19] S.Bhavani and Dr. K.Thanushkodi, "*A Survey On Coding Algorithms In Medical Image Compression*", International Journal on Computer Science and Engineering (IJCSE), Vol. 02 Issue No. 05, pp: 1429-1434, 2010.
[20] T. Kesavamurthy and Subha Rani, "*Dicom Color Medical Image Compression using 3D-SPIHT for Pacs Application*", International Journal *of* Biomedical Science (IJBS), Vol. 4 Issue No. 2, pp: 113-119, June 2008.
[21] Matthew J. Zukoski, Terrance Boult and Tunç Iyriboz," *A novel approach to medical image compression*", International Journal of Bioinformatics Research and Applications, Vol. 2 Issue No. 1, pp: 89-103, 2006.
[22] Ashanta Ranjan Routray and Munesh Chandra Adhikary, "*Image Compression Based on Wavelet and Quantization with Optimal Huffman Code*", International Journal of Computer Applications,Volume 5 Issue No.2, pp: 6-9, August 2010.
[23] Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins, "*Digital Image Processing Using MATLAB",* Pearson Education, 2009