An Implementation on Pattern Creation and Fixed Length Huffman's Compression Techniques for Medical Images

Trupti Baraskar
S.G.B.A. University
Amravati, India

Vijay R. Mankar
S.G.B.A. University
Amravati, India

ABSTRACT

Today more and more digital medical images are used by physicians in their clinical diagnosis. DICOM format images are used by physicians in their clinical diagnosis. DICOM images are too large in size. They require much space for storage and bandwidth for transmission. Thus, medical images are to be compressed due to their large size and repeated usage for diagnostic purpose. Three pattern Huffman compression algorithm uses the concept of pattern creation. The operation of pattern creation is to create patterns using fixed length coding. Using these patterns, compression of image is carried out. Three best patterns are created in such a way, which yields better compression ratios. Three pattern Huffman compression algorithm uses lossless compression technique and can be applied to all types of medical images like CT scans, MRIs, PET, etc. without compromise in quality. The core concept of the algorithm is based on building up a collection of n-length patterns in the image. The basic model of new compression algorithm is similar to that of the Huffman encoder except for the pattern finder. The operation of the pattern finder is to find the best pattern, which is the most frequent occurring pattern. Therefore the best pattern will also be an input to the encoder. The output of the encoder will be the code along with footer information.

Keywords

DICOM, Three Pattern Compression, Pattern Finder, Haar Transform

1. INTRODUCTION

Medical images are at the heart of the healthcare diagnostic procedures. They have provided to view cross-sections of internal organs, tissues, bone and other features of patients but also a mean for physicians to evaluate the patient's diagnosis and monitor the effects of the treatment, and for investigators to conduct research of the mention disease. Medical images come from a board spectrum of imaging sources such as computed axial tomography (CT), magnetic resonance imaging (MRI), digital mammography and positron emission tomography (PET), and they generate a huge amount of image data and important medical information. Data compression is one of the most important need of Hospital Management System. Data compression refers to the process of reducing the amount of data requires to represent a given quantity of information. Data Compression can be defined as reducing of the amount of storage space required to store a given amount of data. The objective of image compression is to reduce the data side by removing the redundancies within the image and without any loss of image quality. Users are also looking for a storage solution that provides faster response times and increased data availability [1]. Lossless compression is important to diagnostic quality images since compressed reconstructed image matches the original, and

lossy compression is generally avoided. An ideal medical image compressor should also be able to compress images up to 16-bit grayscale and 48-bit RGB images, at a relatively high compression ratio, and with a high speed compression and decompression rate [2].

2. RELATED WORK

This related work is served for comparison of Huffman's compression technique such as Static Huffman's and Adaptive Huffman's coding, pre computed Huffman's codes, Lossless Grey-scale Image Compression using Source Symbols Reduction, Haar wavelet with Huffman coding, an n-pattern Huffman compression method, Global thresholding and Huffman encoding, Contourlet Transform with modified Zerotree Wavelet and Huffman coding. These methods are analyzed on the basis of compression ratio, PSNR and computational simpler code. There are many loss-less compression techniques such as Arithmetic coding, Run Length Encoding, Huffman Coding and some famous Dictionary based algorithms like Lempel-Ziv- Welch (LZW) coding, though Huffman Coding forms the basis of many compression Algorithms like modifying Huffman's technique and optimizing it, yields a more effective compression algorithm that increases the compression ratio on medical images. Huffman's procedure creates the optimal code for a set of symbols and probabilities subject to the constraint that the symbols be coded one at a time. The Huffman code is designed by merging the lowest probable symbols and this process is repeated until only two probabilities of two compound symbols are left and thus a code tree is generated and Huffman codes are obtained from labeling of the code tree[3][4][5].

2.1 Static Huffman Encoding

The idea behind Huffman static coding is, representing frequently occurring pixel in shorter code word instead of representing fixed length code word based on frequency of each pixel. The code words are stored in Huffman-Tree. Based on this tree, pixels are presented with shortest code word as applicable. The simplest forms of statistical modeling use a static table of probabilities. Huffman encoding/decoding trees were then built and stored. Compression programs had access to this static model and would compress data using it [5][6].
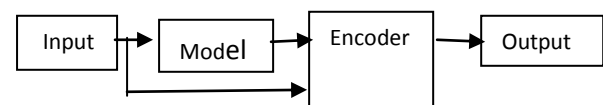

Fig.1 Static Huffman Encoder

2.2 Adaptive Huffman Encoding

This is dynamic version of Huffman encoding technique and avoids preparation of frequency table to know the occurrence of each pixel in an image at the beginning [9][10]. It calculates as it reads after each pixel. The frequency of each symbol is called weight of that symbol and this is done at stage. One problem with adaptive models is that they start knowing essentially nothing about the data. So when the program first starts, it doesn't do a very good job of compression. Most adaptive algorithms tend to adjust quickly to the data stream and will begin turning in respectable compression ratios after only a few thousand bytes. One advantage that adaptive models have over static models is the ability to adapt to local conditions when compressing executable files data may change drastically as the program file changes from binary program code to binary data.
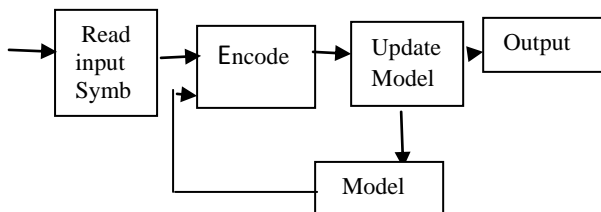

Fig.2 Adaptive Huffman Encoder

2.3 Source symbols reduction and Huffman technique

In this compression method the compression ratio is achieved by reducing number of source symbols. The number of source symbols are reduced by combining together two source symbols. There for the number of generated Huffman code are also reduced. This severed method achieves 10% more compression ratio then regular Huffman coding [11].
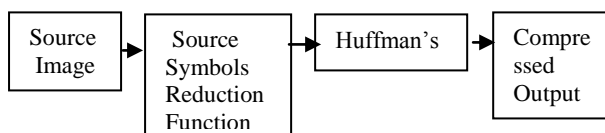

Fig.3 Source Symbols reduction Method

2.4 Lossless Huffman's coding using Binary tree

In this method image are converted in to an array using Delphi image control tool. This control tool will read image and convert them in to a text file. This text file sort based on frequency count of each character in file. Build Huffman's tree based on priorities list and generates canonical Huffman's codes [12].

2.5 Haar and Fast Haar Wavelet with Huffman coding

The Haar transform is one of the simplest and basic transformations from the space/time domain to a local frequency domain, which reveals the space/time-variant spectrum. The attracting features of the Haar transform, includes speed for implementation and ability to analyze the local features, such as signal and image compression. A HT decomposes each signal into two components, one is called average (approximation) or trend and the other is known as difference (detail) or fluctuation. A precise formula for the values of first average sub signal, a = (a1, a2... aN / 2), at one level for a signal of length N. Four parameter calculated those are A, H, V and D respectively. The approximation

component contains most of the energy, after Haar transform has been applied. Hence, it is clear that excluding information from the approximation area will result in largest distortion to the compressed image and excluding information from the diagonal area will result in least distortion to the compressed image [13]. The Fast Haar Wavelet Transform is the enhanced algorithm which can reduce the calculation work in HT [14].
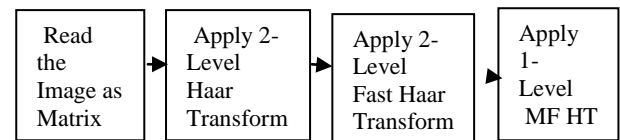

Fig.4 Image Compression Using Modified Fast Haar Wavelet

2.6 Effective Compression Algorithm for Medical Image

The new compression algorithm is similar to the Huffman except for the pattern finder. The pattern finder is use to find best pattern which is the most frequency occurring pattern. There are four basic assumption to implements the effective compression method[15].

1. The pattern is restricted to length 3.
2. The first and last character in the pattern must not same.
3. The second character to be replace by footer information.
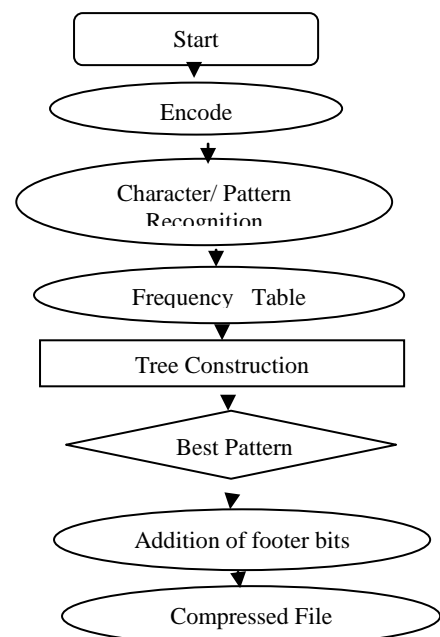4. Each pattern must trace once.

Fig.5 Flow chart of effective compression algorithm

2.7 Basic Huffman compression techniques

Huffman coding is a statistical coding technique that forms the basis for many compression techniques. Huffman compression techniques consist of techniques guaranteed to generate an exact duplicate of input data stream after an expand cycle. This type of compression is used when storing database records, spreadsheets, or word processing file and image formats. In these applications, the loss of even a single

bit could be catastrophic. Huffman algorithm can be used in the case of medical image compression where there should not be any loss of information during compression that will affect proper diagnosis. Huffman coding finds an optimal way to take advantage of varying character frequencies in a particular file.

The idea behind the coding is to give less frequent characters and groups of characters, longer codes. Further, the coding is constructed in such a way that no two constructed codes are prefixes of each other. This property about the code is crucial for deciphering the code easily. Huffman's algorithms proceed to build a code. At the beginning, all nodes are roots of their own degenerate tree of only one leaf. The algorithm merges the trees with the least probability first and repeats this procedure until only one tree is left. The resulting code is {a → 00, b → 01, c → 10, d → 110, e → 111}, for an average length of 2.17 (while the entropy is 2.11 bits/symbol, a discrepancy of 0.06 bits/symbol)[16].

Table 1. Building a Huffman Tree

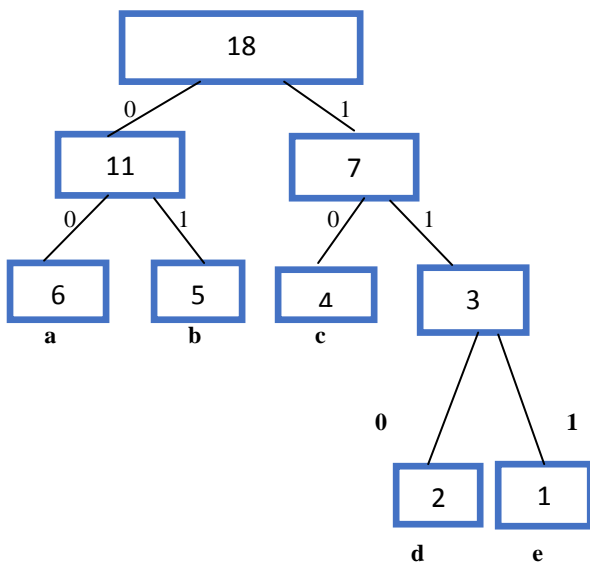| Character	Frequency	Code
a	6	00
b	5	01
c	4	10
d	2	110
e	1	111

Fig. 6: Huffman's Algorithm

The algorithm will proceed iteratively. The process is illustrated in Fig. 6. At the start, all symbols are given a tree node that is the root of its own sub tree. Besides the symbol and its probability, the node contains pointers to a right and a left child. They are initialized to null, symbolized here by ⊥.All the roots are put in a list L. Requirement (c) asks for the two symbols with lowest probability to have codes of the same length. To remove the two roots having the smallest probabilities from L, let them be a and b. We create a new root c having probability P (a) + P (b) and having children a and b. We add c to L. This will cause a and b to share a common prefix, the code for c. We find our self with one less tree in L. If we repeat this until only one tree is left in L, we will have completed the tree-structured code satisfying all of the requirements. The algorithm in pseudocode looks like this:

L = {(a1, P(a1),⊥,⊥), (a2, P(a2),⊥,⊥), . . . , (an, P(an),⊥,⊥)} While |L| > 1
{a = minP L
L = L − {a} b = minP L
L = L − {b} c = (⊥, P (a) + P(b), b, a)
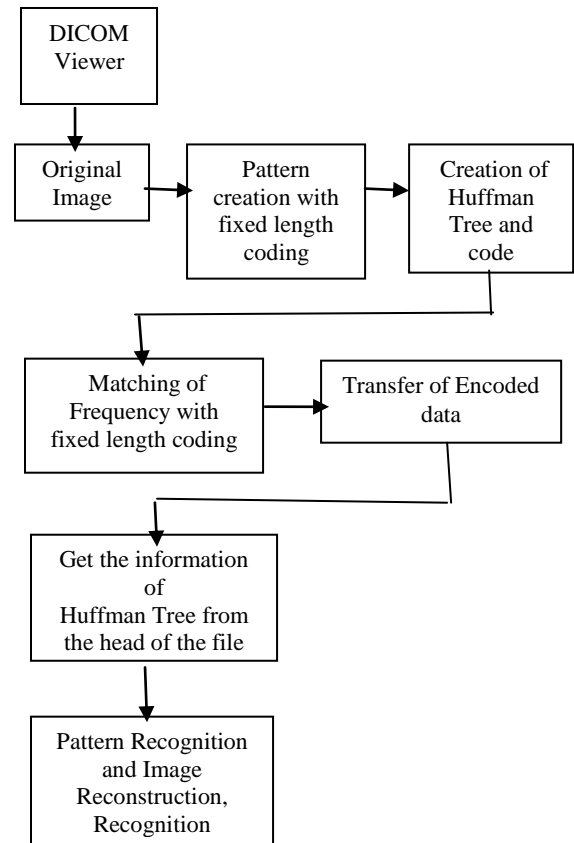L = L ∪ {c}

3. PROPOSED METHOD

Fig.7: A pattern creation and fixed length Huffman's compression techniques

3.1 DICOM Viewer
DICOM Viewer is a tool used for viewing DICOM format medical images as this kind of images requires different image viewer application. This part of the architecture diagram consists of 3 sub-parts which includes parsing of .dcm file, reading of the tags in order to separate image tags from the patient information tags and the third part consists of image and patient details in separated format.

3.2 Encoder
In this part of the system architecture diagram, compression of the image is performed. For this purpose, patterns are created with the help of our fixed length coding. Once the patterns are created we apply Huffman algorithm in order to create a tree and then matching of frequency with the fixed length codes is done. Before creating the Huffman codes we recreate the Huffman tree. Using these Huffman codes we reduce the size of image and this compressed file is stored with .huff extension.

3.3 Decoder
At the decoder side, to perform the decompression operation we read the file header first in order to obtain information about compressed file. Reading of header provides all the

information required to perform further operations. Using the information obtained Huffman tree is reconstructed.

4. IMPLEMENTATION

4.1 DICOM Viewer
Algorithm:

1. Take input as a .dcm DICOM file.

2. Parse the .dcm file to read in built tags in order to separate the image and the patient data.

3. Store the patient data in a text file for further reference, similarly store the pixel (i.e. image) data in .bmp image format to provide as an input for next module.

4.2 Encoder
Algorithm:

1. Take the .bmp image from DICOM viewer module as an input to encoder module.

2. Pattern Finder is used to find patterns using fixed length coding.

3. A Huffman tree is generated based on the corresponding patterns.

4. A matching function matches the frequency with the fixed length coding.

5. Recursive Huffman tree is created using which Huffman code is generated.

4.3 Decoder
Algorithm:

1. The encoded bits generated as an output of encoder module are an input to decoder module.

2. Huffman tree information is extracted from the header of the file.

3. Pattern is recognized and the image is reconstructed using the header bits.

4. The decompressed image can then be viewed for analysis.

5. PERFORMANCE EVALUATION
The proposed algorithm was tested on 20 medical images (256x256). Compression parameters that are used and evaluated are: Compression Ratio (CR), Compression Factor (CF), and Peak Signal to Noise Ratio (PSNR). It is observed that, with 3-pattern Huffman Technique, the average compression ratio is 30.84%. The basic Huffman technique, it is around 10-12%. Figure below shows the original images and their corresponding decompressed images that are used in this implementation.
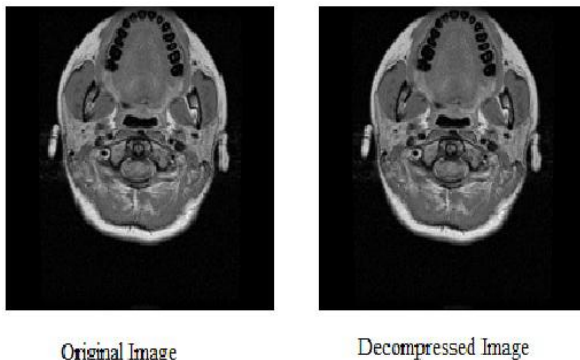

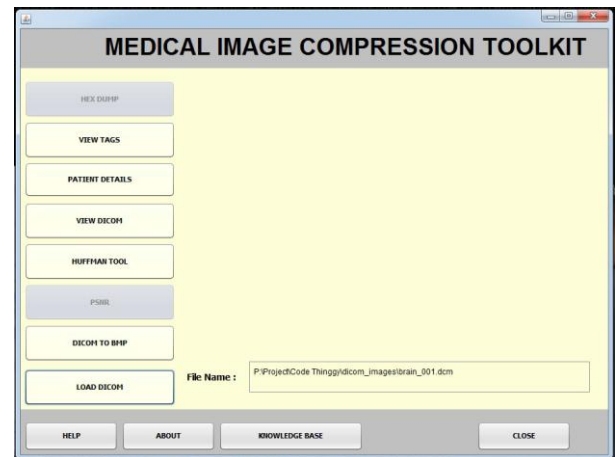**Fig.8: Original Image1 and Corresponding Decompressed Image**
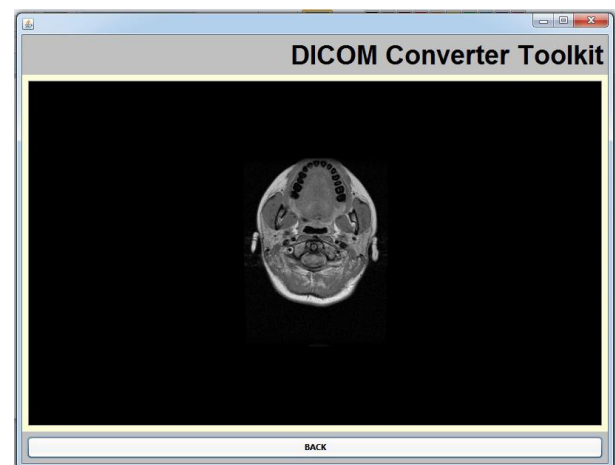

Fig.9: Screen Shot: Load DICOM Image
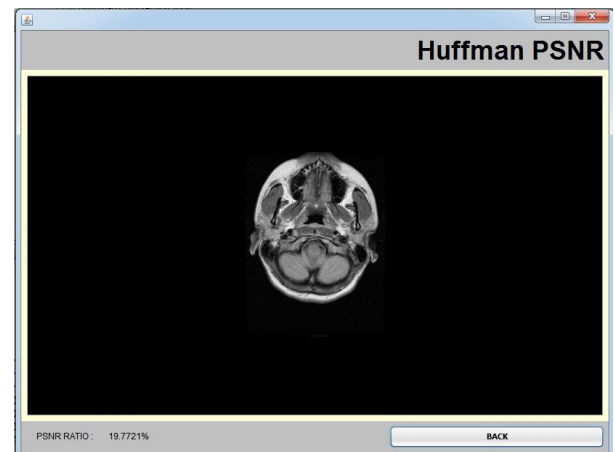

Fig.10: Screen Shot: View DICOM Image

Fig.11: Screen Shot: PSNR value of an Image

6. CONCLUSION
In today's world, JPEG 2000 compression technique is widely used for compression of images. But, the major drawback of this technique is, this method is a lossy compression technique existing Huffman. The proposed method made certain improvements on the existing Huffman technique to determine the best component and the pattern in the image to be processed. The proposed technique is purely lossless method in which the experimental results reveal improved compression percentage, than the existing method. This kind of lossless compression is a basic need for any kind of

medical image. The proposed work is restricted to search for patterns of length 3. This project is mainly done to make small attempt of bringing more compression. There can be numerous improvements compression is a basic need for any kind of medical image. The proposed work is restricted to search for patterns of length 3. This project is mainly done to make small attempt of bringing more compression. There can be numerous improvements made in this new work. When the work gets extended further, there will be an improved compression ratio than basic Huffman.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] J. Janet and T.R. Netesan, "effective compression algorithm for Medical Images as an Aid to Telemedicine", Asian Journal of Information Technology 4(12); 1180 – 1186, 2005.

[2] Neil F. Johnson, "In Search of the Right Image: Recognition and Tracking of Images in Image Databases, Collections, and the Internet." Center for Secure Information Systems, George Mason University, 1999

[3] E.R. Fiala and D.H. Greene, "Data compression with finite windows,"Comm. of the ACM, vol. 32, pp. 490–505, April 1989.

[4] R. G. Gallager, "Variations on a theme by Huffman," IEEE trans. Inf. Theory, vol. 24, pp. 668–674, Nov. 1978.

[5] M. Buro, "On the maximum length of Huffman codes," Information Processing Letters, vol. 45, pp. 219–223, 1993.

[6] Ida Mengyi Pu, Fundamental Data Compression, 1st ed., Butterworth-Heinemann publications, 2006.

[7] Jagadish H Pujar, "A New Lossless Method of Image Compression And Decompression Using Huffman Coding Techniques", Journal of Theoretical and Applied Information Technology, vol. 15, no. 1, pp. 18-23, 2010.

[8] Nelson, M. and J.L. Gailly, 1995, "The Data Compression Book, 2nd Edition, M and T Book, New York.

[9] Li-yangchun, "Image Compression Based on P-tree", in 2nd International Conference Information Science and Engineering (ICISE), pp. 4550-4553, Dec. 4-6, 2010.

[10] Jeeffrey Scott Vitter, "Dynamic Huffman Coding", ACM Transactions on Mathematical Software, vol. 15, no. 2, pp. 158-167, June 1989.

[11] C. Saravanan & R. Ponalagusamy, "Lossless Grey-scale Image Compression using Source Symbols Reduction and Huffman Coding", International Journal of Image Processing (IJIP), Volume (3) : Issue (5)

[12] Mridul Kumar Mathur, Seema Loonker, Dr. Dheeraj Saxena, "Lossless Huffman Coding Technique for Image Compression and Reconstruction using B trees, IJCTA, Jan-Feb 2012

[13] D Mohandas, "A Novel Huffman Coding Mechanism For Medical Image Compression In Telemedicine", Chapter3, shodhganga.in_bnet.ac.in, 2013

[14] Anuj Bhardwaj and Rashid Ali "Image Compression Using Modified Fast Haar Wavelet Transform", World Applied Sciences Journal 7 (5): 647-653, 2009, ISSN 1818-4952 © IDOSI Publications, 2009

[15] J. Janet and T.R. Netesan, "effective compression algorithm for Medical Images as an Aid to Telemedicine", Asian Journal of Information Technology 4(12); 1180 – 1186, 2005.

[16] D Mohandas, "A Novel Huffman Coding Mechanism For Medical Image Compression In Telemedicine", Chapter3, shodhganga.in_bnet.ac.in, 2013