Medical Image Compression Using the Hybrid of Lempel ZIV Welch (LZW) and Huffman Algorithms

Ipeayeda Funmilola Wumi and Binuyo Adeyemi O.

Department of Computer Science and Department of Mathematical Sciences, Ajayi Crowther University, Oyo, Oyo State Nigeria

Abstract: - **In recent years the development and demand of multimedia products has grown increasingly fast, contributing to insufficient bandwidth of network and storage of memory device. Therefore, the theory of data compression becomes more and more significant for reducing the data redundancy to save more hardware space and transmission bandwidth.**

This paper explains the compression of medical images losslessly using the hybrid of two algorithms which are Lempel Ziv Welch (LZW) algorithm and Huffman algorithm. Compression in general is very useful because it helps reduce the consumption of expensive resources such as hard disk space and transmission bandwidth.

In this work, Matrix laboratory (MATLAB) which is both a programming language and also a programming environment was used to carry out the coding.

Conclusively, the lossless image compression is the best because there was no visible loss of quality in the image. It compresses brain images and computed tomography (CT) images losslessly. For further studies, it is recommended that this software should be imbibed into a patient management system in other to keep records of the images and utilize the fast transmission benefit.

Keywords: Image compression, Hybrid algorithm, Lossless and Lossy compression, Huffman algorithm, Lempel Ziv Welch (LZW) algorithm, DICOM images, MATLAB.

I. INTRODUCTION

Image compression reduces the irrelevance and redundancy of an image data in order to be able to store or transmit data in an efficient form. Where loss of any Information is not acceptable and data are critical then Lossless Compression method is applied. Compression of medical image is based on lossless compression method. For identification of disease and surgical planning, medical imaging is being used, and long-term storage is needed for profiling patient's data. To avoid loss of critical medical information, lossless compressions of the medical images are indispensable.

Medical images require large amounts of memory which is a concern for medical imaging users. On the accuracy of clinical diagnosis, it has become a crucial area for research to estimate the effect of image compression. The most usually used dimensions of image quality are peak signalnoise ratio and mean square error.

Medical images are increasingly displayed on a range of devices connected by distributed networks, which place bandwidth constraints on image transmission. As medical imaging has transitioned to digital formats such as DICOM and archives grow in size, optimal settings for image compression are needed to facilitate long term mass storage requirements.

Storage issues and bandwidth over networks have led to a need to optimally compress medical imaging files while leaving clinical image quality uncompromised.

One definition of optimal medical image compression is a degree of compression that decreases file size substantially but produces a degree of image distortion that is not clinically significant. A more conservative definition of optimal image compression would require a degree of image distortion that cannot be perceived by the viewer at all. Other methods that have been previously used to distinguish degrees of medical image compression include pixel analysis and blinded measurements of diagnostic accuracy.

II. STATEMENT OF THE PROBLEM

The researcher has identified the difficulty faced by medical practitioners when they want to send medical images such as x-rays and MRIs from one practitioner to another. This difficulty is due to the fact that medical images are very large and take up disk or memory space therefore making the bandwidth required to send the images over a network very high which is not economical and also the monotonous paper environment brought about by carrying films around needs to be eliminated since it does not encourage medical report management.

The process in existence involves the printing out of images on films. This process has the following problems:

I. Difficulty in maintenance of records
II. Time consuming
III. Damage coming upon the films
IV. High cost of transmission

III. LOSSLESS AND LOSSY IMAGE COMPRESSION

The lossless image compression has been a significant issue in recent years due to the increasing demand of storing huge amount of high quality multimedia data in a given small storage. With lossless coding, we restore every detail of the original data upon decoding. Obviously this is a necessity for numerical and financial documents.

Unlike lossless image compression, the lossy image compression is not advisable for numerical or medical images because although it reduces the size of the image it also reduces the quality of the image and causes loss of certain bits. This loss or reduction in quality makes the image not as clear and visible as its original image.

➢ *Lossy Image Compression*

The nature of the human eye's perception allows significant reduction of information in the image for JPEG2000. The transform coefficients are rounded and quantized causing partial loss of information. These algorithms are optimized for compression of photographic images, which are mostly used in computer industry. There are also transform-based algorithms optimized for different tasks such as Enhanced Compression Wavelets for the compression of aerial and satellite photos. DCT and especially wavelet -based algorithms present excellent compression efficiency in terms of compression vs. degradation tradeoff for the class of images to which they were optimized. In some applications, it is not necessary to transfer the whole image details into one continuous transmission. It is often more important to have a schematic thumbnail of the image faster than the whole image.

Moh'dAliMoustafaAlsayyh (2012) mentioned that lossy coding can only be applied to data such as images and audio for which humans will tolerate some loss of fidelity (faithfulness of our reproduction of an image after compression and decompression with the original image). He went further to explain that lossy compression is beneficial in applications such as television transmission, video conferencing, and facsimile transmission etc., in these circumstances certain amount of error is tolerable. Because we are no longer being held to the same requirements that underline the reproduction of financial or engineering data, we should be able to realize greater compression of the data as we increase the allowed loss of information. Lossy compression algorithms are capable of attaining high compression ratios.

Lossy compression algorithms obtain greater compression by allowing distortion of the image that will be recovered on decompression. In fact, lossy schemes provide much higher compression ratios than lossless schemes. In effect, these systems simplify the image by removing information from them. The more the degree of simplification, the less the recovered image will look like the original. Because the allowable distortion varies according to the purpose of the image, the amount of compression is typically set by the user of the compression system.

➢ *Lossless Methods For Image Compression*

Lossless data compression is a technique that allows exact reconstruction of the original image from its compressed form. This branch of image compression has achieved numerous inventions during last few decades. Some of the algorithms deal with an image in spartial domain while the others work in frequency domain. Lossless compression is used when the original and the decompressed image must have the same details. Thaneshwar Kumar(2015)states that lossless compression is necessary for many high performance applications such as geophysics, telemetry, nondestructive evaluation, and medical imaging.

DzulkifliMohamad (2012) reports that the lossless compression is particularly beneficial in image archiving and permits the image to be compressed and decompressed without dropping any data. . In lossless image compression algorithms, the reconstructed image is identical to the original image. Lossless algorithms, however, are limited by the low compression ratios they can attain. These methods include Run-length encoding – used as default method in PCX and as one of possible in BMP, TGA, TIFF, DPCM and Predictive Coding, Entropy encoding, Adaptive dictionary algorithms such as LZW – used in GIF and TIFF, Deflation – used in PNG, MNG, and TIFF and Chain codes.

Based on the study, it is possible to apply the lossless methods for image compression to DICOM images. This is because the system can help reduce bandwidth, storage required, patients waiting time and increase productivity. As a result, it is worth for investing on the system, whereby it can shorten the stress on both patients and doctors.

IV. HYBRID ALGORITHM

According to Wikipedia, Hybrid algorithm is an algorithm that combines two or more other algorithms that solve the same problem either choosing one (depending on the data), or switching between them over the course of algorithm. This is generally done to combine desired features of each, so that the overall algorithm is better than the individual components.

"Hybrid algorithm" does not refer to simply combining multiple algorithms to solve a different problem-many algorithms can be considered as combinations of simpler pieces – but only to combining algorithms that solve the same problem, but differ in other characteristics, notably performance.

V. HUFFMAN ALGORITHM

Huffman coding was Proposed by DR. David A. Huffman in 1952. This is a general technique for coding symbols based on their statistical occurrence frequencies (probabilities). The pixels in the image are treated as symbols. The symbols that occur more frequently are assigned a smaller number of bits, while the symbols that occur less frequently are assigned a larger number of bits. Huffman code is a prefix code. This means that the (binary) code of any symbol is not the prefix of the code of any symbol. Most image coding standards use lossy techniques in the earlier stages of compression and use Huffman coding as the final step which is a method for the construction for minimum redundancy code. Huffman code is technique for compressing data.

VI. LEMPEL ZIV WELCH (LZW) ALGORITHM

There are generally two classes of lossless compressors: dictionary compressors and statistical compressors. Lempel-Ziv algorithm is a dictionary compressor. Dictionary compressors build dictionaries of strings and replaces entire groups of symbols. Dictionary based coding can further be divided into static dictionary coding and dynamic dictionary coding. In static dictionary coding, dictionary is fixed during the encoding and decoding processes. In dynamic dictionary coding, the dictionary is updated on fly

The compression algorithm is a dictionary based compressor that consists of a rule for parsing strings of symbols from a finite alphabet into substrings whose lengths do not exceed a predetermined integer, and a coding scheme which maps these substrings into decipherable code-words of fixed length over the same alphabet. As the algorithm runs, a dictionary of the strings which have appeared is updated and maintained. The dictionary is pre-loaded with the 256 possible bit sequences that can appear in a byte.

VII. METHODOLOGY

The methodology adopted for the lossless compression begins by reading the medical images on to the workspace of MATLAB R2013a. LZW encoding is initiated by calling a function which finds the symbols (i.e. pixel value which is non-repeated) as well as the function which calculates the probability of each symbols. The probability of the symbols are arranged in decreasing order and lower probabilities are merged and this step is continued until only two probabilities are left and codes are assigned according to rule that "the highest probable symbol will have a shorter length code". Further Huffman encoding is then performed i.e. mapping of the code words to the corresponding symbols which results in Huffman code-words.

- *Algorithm Implementation*

This section deals with the design of a lossless image compression using a LZW-Huffman hybrid algorithm. The method is based on LZW algorithm and the Huffman algorithm (an error correcting technique in order to improve the compression ratio of the image comparing to other compression algorithms). The method is a lossless image compression scheme which is applied to all types of medical image based on LZW algorithm that reduce the repeated value in image and Huffman codes that detect/correct the errors. The Huffman algorithm works by adding extra bits called parity bits, whose role is to verify the correctness of the original message sent to the receiver so, the system in this project work benefit from this feature. This method of Huffman converts blocks of size k pixels to n pixels by adding parity bits, depending on the size of the image, which is encoded into a code word of the length n.

This developed compression system improves the compression of the image through the implementation of LZW algorithm. First, the entered image is converted to the gray scale form and then converted from decimal to binary to be a suitable form to be compressed. The algorithm builds a data dictionary (also called a translation table or string table) using the in-built MATLAB functions of data occurring in an uncompressed data stream. Patterns of data are identified in the data stream and are matched to entries in the dictionary. If the patterns are not present in the dictionary, a code phrase is created based on the data content of that pattern, and it is stored in the dictionary. The phrase is written to the compressed output stream. When a reoccurrence of a pattern is identified in the data, the phrase of the pattern already stored in the dictionary is written to the output.

The binary input image is firstly divided into blocks of size 7 bits each; only 7 bits needed to represent in each byte, 128 value in total, while eighth bits represent sign of the number (most significant bit) that do not affect the total value of blocks, and converts it to a pixel field to be accepted as an input to the HUFFMAN algorithm. Each block is decoded using HUFFMAN, then checked if it is a valid code-word or not. The HUFFMAN converts the valid block to 4 bits. The method adds 1 as an indicator for the valid code-word to an extra file called (map), otherwise if it is not a valid code-word, it remains 7 and adds 0 to the same file. The benefit of the extra file (map) is that it is used as the key for image decompression in order to distinguish between compressed blocks and the not compressed ones (code-word or not). After the image is compressed, the file (map) is compressed to decrease its size, and then it is attached to the header of the image. This step is repeated twice, the HUFFMAN decoding repeated to improve the compression ratio.

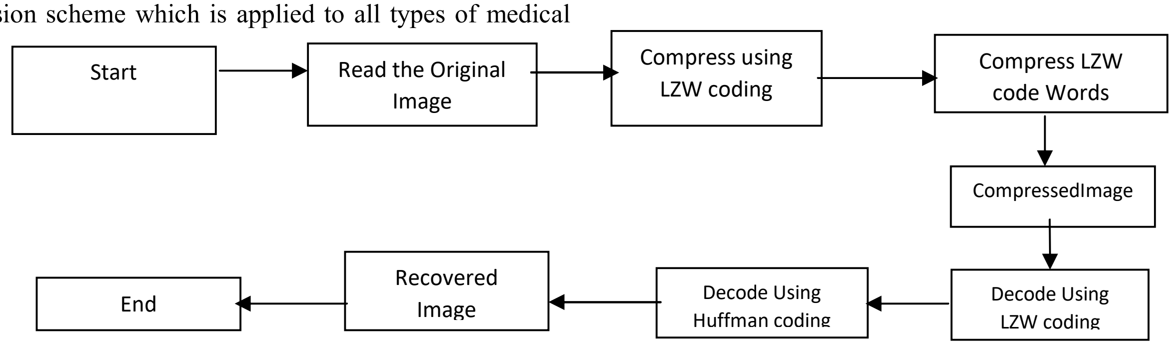

Fig 1.0 Medical Image Compression Dataflow Diagram

Concatenation of all the LZW code words is done and Huffman error detection encoding is applied. Final encoded values are the compressed data. Decoding process is applied on final encoded values and output the Huffman code words and Huffman encode value of the encoding process. In the last step the recovered image is generated, which is the compressed image.

The procedure can be summarized in the following brief steps and the data flow diagram in figure 3.1:

Step1: Read image on to the workspace of MATLAB

Step2: Call a function which finds the symbols (i.e. pixel value which is non-repeated).

Step3: Call a function which calculates the probability of each symbols.

Step4: Probability of symbols are arranged in decreasing order and lower probabilities are merged and this step is continued until only two probabilities are left and codes are assigned according to rule that the highest probable symbol will have a shorter length code.

Step5: Further LZW encoding is performed i.e. mapping of the code words to the corresponding symbols will result in a LZW code-words.

Step6: Concatenate all the LZW code words and apply Huffman Dictionary on final encoded values (compressed data) for error detection and correction.

Step7: Apply Huffman decoding process on final encoded values and output the Huffman code words

Step8: Apply Huffman encode value to the code words.

Step9: In the last step the Recovered image is generated as the compressed image.

VIII. COMPRESSION PSEUDO-CODE

This hybrid algorithm compresses the original image by implementing a number of steps described in the pseudo-code below. The algorithm steps are:

Input: image (f)

Output: compressed file

Begin

Initialize parameters

SET round to zero

READ image (f)

Convert (f) to gray scale

SET A = ( ) // set empty value to matrix A

A = image (f)

Convert matrix A into binary (Bn)

Convert matrix compress by LZW into binary

d2 = decoding by HUFFMAN decoder

INCREMENT the parameter test (the number of code-word found) by 1;

add the compressed block d2 to the matrix CmprsImg

add 1 to the map[round] matrix ELSE

add the original block (origin) to the matrix CmprsImg

add 0 to the map[round] matrix

Pad and Add remd bits to the matrix CmprsImg and encode it (Huffman encoding)

Final map file = map [round] to reuse map file in the iteration

FOR stp = 1 to 3

END FOR stp

INCREMENT round by 1

ENDWHILE

END

The algorithm was implemented on MATLAB IDE (using MATLAB functions as specified in the steps above).

IX. RESULTS AND DISCUSSION

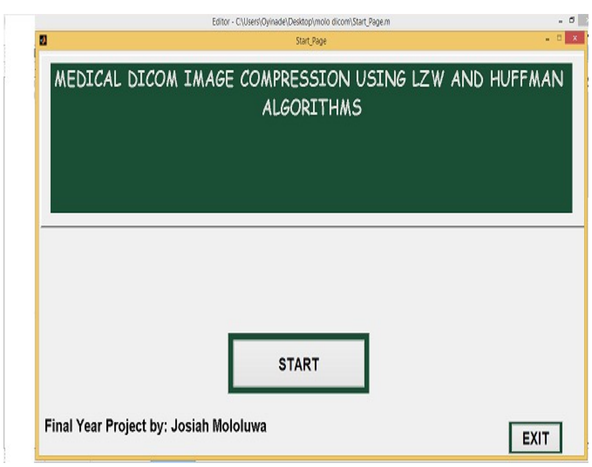

Fig 2.0: **The index window**

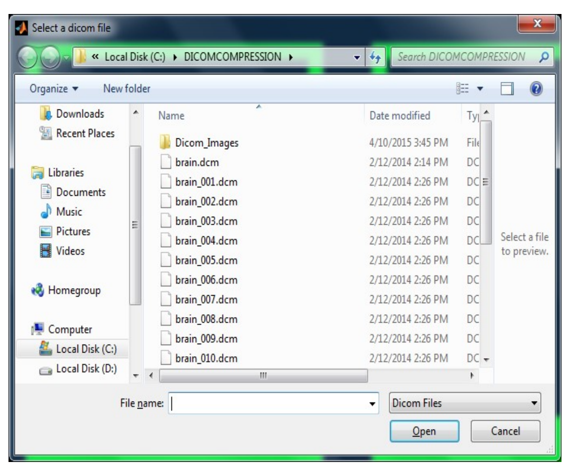

Fig 3.0: **The uploading of the medical image**

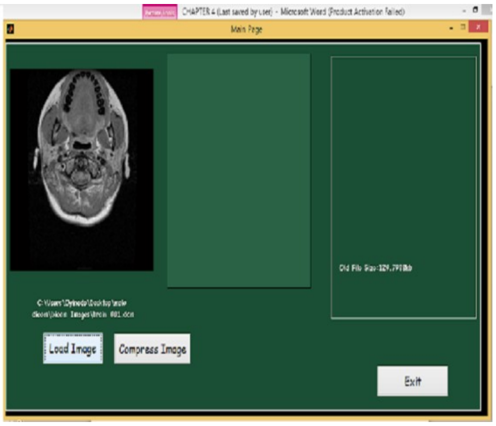

Fig 4.0: The original image to be compressed

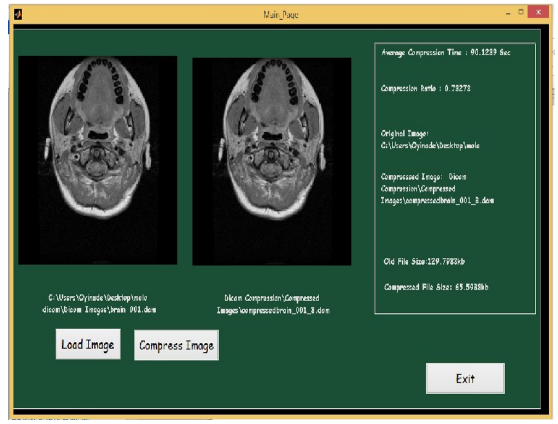

Fig 5.0: The compressed image and details

The software designed using MATLAB can be loaded by copying it into the HDD of the PC. The folder is accessed via the MATLAB IDE. The index page pops up once the user clicks on the 'start_page.m' in the folder. The starting point for the use of the software designed for the lossless compression of medical images is the index page as shown in figure 2.0, on the index page is the 'start' button. Once the user presses the start button the compression page will be opened. The compression page has three panes and three buttons; the 'load image', 'compress image' and 'exit page' buttons. Once the user clicks on the 'load image' button, the app opens the folder system where the user can pick the medical image to be compressed from the HDD as shown in figure 3.0, the file location of the medical image to be compressed is also displayed. The chosen image is displayed on the first pane of the compression page as shown in figure 4.0; 'Compression initialized' is displayed on the third pane. Compression takes between few seconds to few minutes, depending on the size of the medical image to be compressed.

Once the compression is done, the compressed image is displayed on the second pane of the compression page, the file location of the compressed image is also displayed. The compressed image is automatically saved in the HDD of PC. The compressed image still maintains the same quality as of the original image, the DICOM property maintained. The compression details are shown on the third pane of the compression page; the details are the compression ratio, the file locations of the original and the compressed medical images as well as the report (whether it is successful, aborted or not successful) as shown in figure 5.0. 'Exit page' button takes the user back to the 'index page'.

X. CONCLUSION

It has been established that image compression reduces the irrelevance and redundancy of an image data in order to be able to store or transmit data in an efficient form. During this research, the analysis and comparison of existing Medical Image Compression Algorithms was done and this resulted from the review of related works done on image compression. The critical analysis led to the adoption of the scheme of requirement specifications that highlighted the features of the Medical Image Compression software implemented in this project work.

This Medical Image compression project work is all about gathering of medical image dataset and compressing them without loss of DICOM properties (Digital Imaging and Communications in Medicine) and quality, in order to make it easier for transmission, transfer and storage.

The development of the medical image compression software involved different phases which include the experimental comparison of the features of Lempel Ziv Welch (LZW) and Huffman algorithms. LZW is first initiated by calling a function which finds the symbols as well as the function which calculates the probability of each symbol. The Huffman encoding is then performed to map the code words to the corresponding symbols which results in Huffman code-words.

REFERENCES

[1]. Alasdair McAndrew "An Introduction to Digital Image Processing with Matlab", September 1, 2004
[2]. Amit S. Tajnep, ProfPravin S., Kulkarnip (2015) Medical images compression using hybrid technique
[3]. Chaur-Chin Chen, Department of Computer Science, National TsingHua University "On the Selection of Image Compression Algorithms", June 2005
[4]. Guy E.Blelloch, "Introduction to Data Compression", Computer Science Department, Carnegie Mellon University, January 31, 2013
[5]. Jagannath.D.J.I and ShanthiniPandiaraj, Lossless Compression of a Desktop Image For Transmission", International Journal of Recent Trends in Engineerong, Volume 2, No.#, November 2009
[6]. Pabbisetti Sathyanarayana1, and Hammid R. Saeedipour, "Digital Image Compression and Decompression Using Three Different Transforms and Comparison of Their Performance" September 2006.
[7]. Priya C., Kesavamurthy T., Uma Priya M.(July 2014) An efficient lossless medical image compression using hybrid algorithm, Advanced materials research,.
[8]. Sonal, dinesh Kumar "A study of various image compression Techniques" Department of computer science & engineering and Guru jhambheswar university of science and technology, hisar.

[9]. Suryender Kumar and SonitSukhraj Singh, “Image compression technique for medical images: a review”, IJREAS Volume 2, Issue 2 (February 2012).
[10]. ThaneshwarKumal, Dr Ramesh Kumar (Feb 2015) Medical Image Compression Using Hybrid Techniques of DWT, DCT and Huffman Coding
[11]. Wei-Yi Wei “An Introduction to Image Compression” National Taiwan University University, Taipei, Taiwan, ROC, 2004.
[12]. Ajala F. A., Adigun A.A, Oke A.O, “Development of Hybrid Compression Algorithm for Medical Images using Lempel-Ziv-Welch and Huffman Encoding” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume 7, Issue 4, September 2011