Development of Hybrid Compression Algorithm for Medical Images using Lempel-Ziv-Welch and Huffman Encoding

Ajala F. A., Adigun A.A, Oke A.O

***Abstract:** Image compression is of utmost importance in data processing, because of the cost savings it offers and because of the large volume of data transferred from one end to the other. The smaller the size of the data the better transmission speed and it also saves time. In communication, transmission of data efficiently, fast and noise free is essential.. Both the LZW and Huffman image compression algorithm are lossless in manner and these methods and some versions of them are very common in use of compressing images. On the average Huffman gives better compression results, while LZW give a better signal-noise-ratio and when the compression efficiency gap between the LZW algorithm and its Huffman counterpart is the largest. In this work, Hybrid of LZW and Huffman image compression was developed and used. It gives better compression ratio and SNR than Huffman Encoding and LZW Algorithm. It also provides cheap, reliable and efficient system for image compression in digital communication system. The average result shows that Huffman encoding has 59.46% and LZW has1,99% of compression ratio, whereas the hybrid of Huffman and LZW has compression ratio of 47.61% but has 92.76% of Signal to Noise Ratio that produce better result of the original image.*

***Keywords:** Compression, Transmission, Huffman, LZW, Encoding.*

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. Loss of any Information is not acceptable and pleasant in some situation especially when dealing with medical information. Image compression techniques can be divided into either lossy or lossless. To avoid loss of critical medical information, lossless compressions of the medical images are indispensable. Many years ago, the generation of medical images in hospitals has been increased considerably and large amount of medical data are generated every year, therefore, storage and transmission of these medical information is necessary.

Revised Manuscript Received on 30 September 2018.
* Correspondence Author
Ajala F. A., Department of Computer Science and Engineering, LAUTECH, Ogbomosho, Oyo State, Nigeria. E-mail: faajala@lautech.edu.ng
Adigun A.A, Department of Computer Science and Engineering, LAUTECH, Ogbomosho, Oyo State, Nigeria. E-mail: aaadigun@lautech.edu.ng
Oke A.O, Department of Computer Science and Engineering, LAUTECH, Ogbomosho, Oyo State, Nigeria. E-mail: aooke@lautech.edu.ng

Complete data fidelity is ensured after reconstruction by lossless compression, but in general compression, ratios is limited to 2:1 to 3:1. However, only ordinary reduction is provided in file size used by lossless techniques. Lossy compression techniques are required to significantly affect storage costs. Medical images require large amounts of memory. (Meyer and Tischer, 2010)For medical imaging users, major issues are the consequent requirements of storage space. 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 is peak signal, noise 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.

II. IMAGE COMPRESSION

Compression refers to reducing the quantity of data used to represent a file, image or video content without excessively reducing the quality of the original data. Image compression is the application of data compression on digital images. The main purpose of image compression is to reduce the redundancy and irrelevancy present in the image, so that it can be stored and transferred efficiently. The compressed image is represented by less number of bits compared to original. Hence, the required storage size will be reduced, consequently maximum images can be stored and it can transferred in faster way to save the time, transmission bandwidth.

Image compression is the application of data compression on digital images. Image compression is used to minimize the amount of memory needed to represent an image (Subramanya, 2001). Images often require a large number of bits to represent them, and if the image needs to be transmitted or stored, it is impractical to do so without somehow reducing the number of bits. Compression of image plays an important role in medical field for efficient storage and transmission. There are many types of medical image compression techniques available. Different techniques used in different image like X-ray angiograms (XA), magnetic resonance image (MRI), etc.

Compression is achieved by the removal of one or more of three basic redundancies:

1. Coding redundancy, which is present when less than optimal (i.e. the smallest length) code words are used.
2. Inter pixel redundancy, which results from correlation between the pixels of an image.
3. Psycho visual redundancy which is due to data that is ignored by the human visual system (i.e. visually non-essential information.

Most of the existing image coding algorithm is based on the correlation between adjacent pixels and therefore the compression ratio is not high. (Mohammed, 2008)

Image compression is an inevitable solution for image transmission since the channel bandwidth is limited and the demand is for faster transmission. Storage limitation is also forcing to go for image compression as the color resolution and spatial resolutions are increasing according to quality requirements. A huge amount of online information is used either graphical or pictorial in nature. The requirements for storage and communications are high. Compressing the data is one of the ways out for this problem. Thus the methods of compressing the data prior to storage and/or transmission are of essential real-world and viable concern (Mohamed, 1997 IEEE).

Lossless and lossy Compression techniques are the two broad classes of image compression. And the lossy compression, affords higher levels of compression ratio but result in a less than perfect replica of the original image (Rajeswari and Rajesh, 2011).

The principal aim of image compression is to reduce the number of bits required to represent an image. Lossless compression is preferred for archival purposes and often for medical imaging, technical drawings, clip art. Though the reconstructed image is not identical to the original image, lossy compression algorithms attain high compression ratios by exploiting human visual properties (Ramya and Mala, 2007), (Suruil and Jain, 2007). Lossless algorithms, however, are limited by the low compression ratios they can attain.

Aim of Image Compression

The reasons for image compression are as follows;

- Digital images require a large amount of space for storage and large bandwidths for transmission. eg A 640 x 480 color image requires close to 1MB of space.
- It reduces the amount of data required to represent a digital image. e.g Reduce storage requirements and increase transmission rates.
- It reduces the number of bits needed to represent a given image or its information.
- Image compression exploits the fact that all images are not equally likely.

III. METHODOLOGY

In this lossless medical image compression implementation process, the procedure (process) was coded from scratch using MATLAB R2013a programming language. The implementation was done in two phases. The first phase was the application of Lempel–Ziv–Welch (LZW) encoding on the medical images to extract relevant features from the medical image dataset, the second phase is the application of Huffman algorithm on the same medical images in the same dataset. At the second stage, Huffman receives the reduced features obtained from the feature extraction in phase one. The algorithm used for the medical image compression is discussed below:

3.1. Huffman Encoding

Huffman coding was proposed by DR. David A. Huffman in 1952. Which is Huffman (W, n) Here, W means weight and n is the number of inputs

Step 1: Input: A list W of n (Positive) Weights.

Step 2: Output: An Extended Binary Tree T with Weights Taken from W that gives the minimum weighted path length.

Step 3: Procedure: Create list F from singleton trees formed from elements of W. **While** (F has more than 1 element) **do**

Step 4: Find T1, T2 in F that have minimum values associated with their roots // T1 and T2 are sub tree

Step 5: Construct new tree T by creating a new node and setting T1 and T2 as its children

Step 6: Let, the sum of the values associated with the roots of T1 and T2 be associated with the root of T Add T to F

Step 7: Do Huffman-Tree stored in F

3.2. Lempel-Ziv-Welch (LZW) Algorithm

In 1978, Jacob Ziv and Lempel created the first popular universal compression algorithm for data when no prior knowledge of the source was available. The algorithm goes thus

Step 1: At the start, the dictionary contains all possible roots, and P is empty;

Step 2: C: = next character in the char stream;

Step 3: Is the string P+C present in the dictionary?

(a) if it is, P := P+C (extend P with C); if not, – output the code word which denotes P to the code stream; – add the string P+C to the dictionary; –P := C (P now contains only the character C); (c) Are there more characters in the char stream? –if yes, go back to step 2; –if not:

Step 4: Output the code word which denotes P to the code stream;

Step 5: END.

3.3. The developed Hybrid Compression Algorithm

To develop the hybrid of Huffman Encoding and Lempel-Ziv-Welch (LZW) compression algorithm, the 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 steps

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.

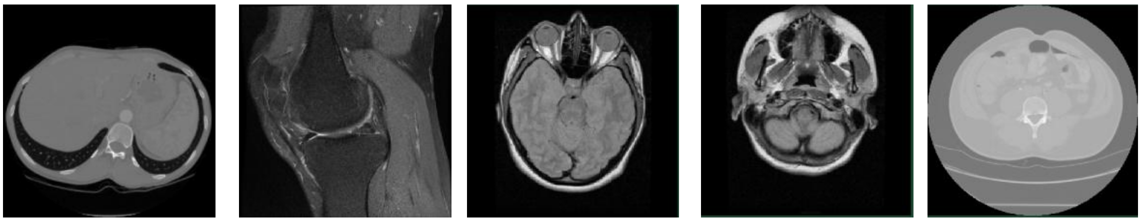
Step7: Apply Huffman decoding process on final encoded values and output the Huffman code words

Step8: Apply Huffman encode value to the code words.

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

3.4. Image Acquisition

Digital Imaging and Communications in Medicine (**DICOM**) is the standard for the communication and management of medical imaging information and related data. Five DICOM images are used for this project. The images are shown below:



(a) **(b)** **(c)** **(d)** **(e)**

Figure 3.1: Sample Images used for testing

IV. RESULT AND DISCUSSION

4.1. Result of DICOM Image Compression using Huffman Encoding

Table 4.1 Result of DICOM Image Compression using Huffman Encoding

Images/Parameter	SNR (%)	PSNR (%)	MSE (%)	Compression Ratio (%)	Old Size (kb)	New Size (kb)
Image 1	40.12	27.44	56.93	64.82	89.11	31.35
Image 2	31.98	26.13	31.93	83.57	285.86	46.98
Image 3	32.42	24.34	60.05	49.47	129.80	65.59
Image 4	30.62	23.76	60.85	49.46	129.80	65.60
Image 5	48.08	26.26	67.33	49.98	515.55	257.78

The result above shows that the DICOM images has reduced from its original size. The results show a high compression ratio, Mean Squared Error (MSE) but low Signal-Noise-Ratio (SNR). There are higher and lower compression ratio depends on the images.

4.2. Result of DICOM Image Compression Using Lempel-Ziv-Welch Algorithm

The result below shows that the DICOM images is closely identical to the original image. The result shows a low compression ratio, Mean Squared Error (MSE) but a high Signal-Noise-Ratio (SNR).

Table 4.2 Result of DICOM Image Compression using Lempel-Ziv-Welch Algorithm

Images/Parameter	SNR (%)	PSNR (%)	MSE (%)	Compression Ratio (%)	Old Size (kb)	New Size (kb)
Image 1	75.73	42.90	29.49	2.17	89.11	69.79
Image 2	59.11	37.91	31.92	1.86	285.86	232.78
Image 3	56.87	34.96	34.17	1.98	129.80	104.11
Image 4	55.65	34.63	38.08	1.97	129.80	104.12
Image 5	76.37	38.54	29.73	1.99	515.55	412.44

4.3. Result of DICOM Image Compression Using a Hybrid of Lempel-Ziv-Welch Algorithm and Huffman Encoding

The result below shows that the DICOM images is closely identical to the original image and has reduced from its original size. The result shows a high compression ratio, mean squared error (MSE) and a high signal-noise-ratio

(SNR). Table 4.4 and figure 4.1 shows the difference between the Huffman encoding, LZW algorithm and the hybrid.

Huffman generally has a high compression ratio but a low SNR. This means that the noise in the resulting image is greater than the signal making the image distorted. This image can't be used in a medical facility due to lack of accuracy.

LZW algorithm generally has a high SNR but a low compression ratio. This means that images compressed using LZW are lossless but the difference in size is little as this makes file transfer slow in medical facility.

The hybridization of Huffman Encoding and LZW algorithm give a better result. The resulting image gives a high compression ratio and high SNR. This means the image is lossless and the reduction in size is significant. Figure 4.6 shows the resulting images after the compression using the hybridization of LZW algorithm and Huffman encoding.

Table 4.3 Result of DICOM Image Compression Using a Hybrid of LZW algorithm and Huffman Encoding

Images/Parameter	SNR (%)	PSNR (%)	MSE (%)	Compression Ratio (%)	Old Size (kb)	New Size (kb)
Image 1	92.91	60.18	18.83	44.22	89.11	49.71
Image 2	89.97	58.76	21.16	44.89	285.86	157.51
Image 3	90.18	59.77	16.96	49.47	129.80	65.59
Image 4	96.54	59.66	28.42	49.46	129.80	65.59
Image 5	94.19	59.74	22.52	49.99	515.55	257.78

Table 4.4 Comparison of the Average Values of the Algorithm

Images/Parameter	SNR (%)	PSNR (%)	MSE (%)	Compression Ratio (%)
Huffman Encoding	36.64	25.59	55.42	59.46
LZW	64.75	37.79	32.68	1.99
Hybrid of LZW and Huffman	92.76	59.62	21.58	47.61

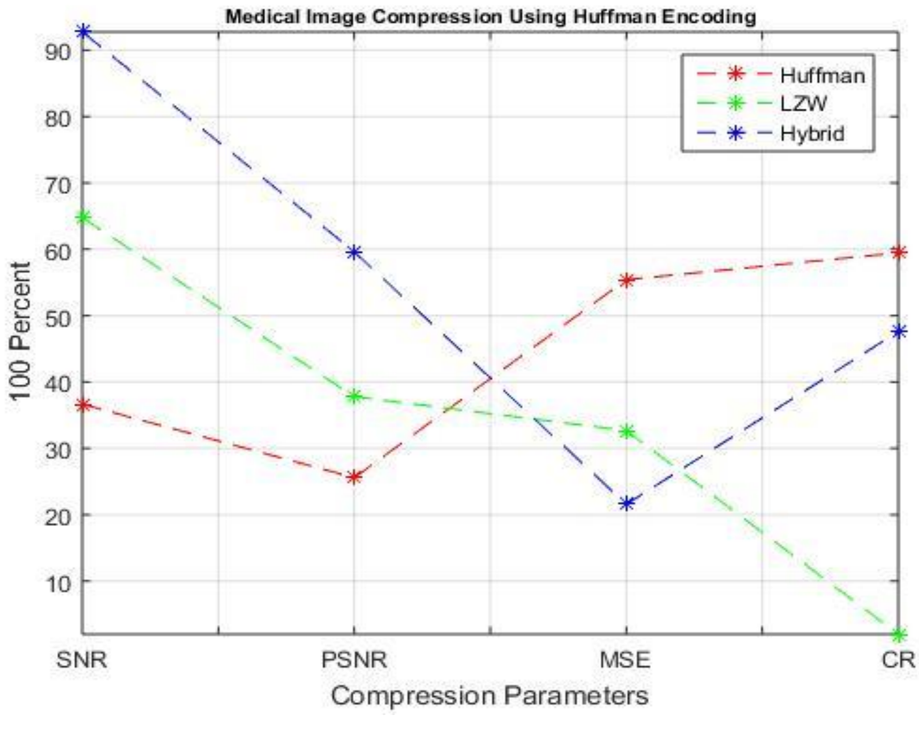


Figure 4.1: Graphical view of result of Average Values of The Algorithms

V. CONCLUSION

The main idea of using a hybridization of Huffman encoding and LZW algorithm to compress medical image is to increase the compression ratio while the image retains almost all the feature of the original image. Thus the hybridization consist of features from both Huffman encoding and LZW algorithm. Image compressed with Huffman encoding are always fast and have a massive reduction in size but some of the original data is always discarded making the construction of the original image not possible. The image is not clear as the original image.

Image compressed with LZW algorithm always have little reduction in size but the difference in the original image and compressed file are negligible to the human eyes. The process is a bit slower than the Huffman encoding.

The hybridization allows for features in both algorithms to be implemented. Images compressed from the hybridization of the algorithm show a high compression ratio and high signal-to-noise ratio.

REFERENCES

1. Subramanya A, (2002) "Image Compression Technique," Potentials IEEE, Vol. 20, Issue 1, pp 19-23,
2. Meyer B and P. Tischer, (2010) "TMW—A New Method for Lossless Image Compression," Australia,
3. Huffman D (1952), A Method for the Construction of Minimum Redundancy Codes,‖ Proc. IRE, pp. 1098-1101,
4. Rajeswari R and R. Rajesh, (2011) "WBMP Compression," International Journal of Wisdom Based Computing, Vol. 1, No. 2,
5. Ramya R, and K. Mala, (2007) A hybrid Compression Algorithm for Compound Images, IEEE International Conference on Computational Intelligence and Multimedia Applications, Vol.3, pp 68-72.
6. Sunil Kumar and R.C.Jain, (2007) "Low complexity fractal-based image compression techniques," IEEE Transactions on Consumer Electronics, Vol.43, no.4, pp 987-993,
7. Mohammed, (2008)"Highly scalable hybrid image coding scheme," Digital Signal Processing, Vol.18, no.3, pp 364-374,
8. Zhe-Ming L.U.and P.E.I.Hui, (2005) "Hybrid image compression scheme based on PVQ and DCTVQ," IEICE transactions on information and systems, Vol.88, no.10, pp 2422-2426,
9. ZIV, J., AND LEMPEL(1978), A. Compression of individual sequences via variable-rate coding. IEEE Transactions on Information Theory 24, 530–536