ISSN No: - 2456- 2165

A Review on Study and Analysis of various Compression Techniques

Rashmi Sharma
M.Tech Scholar
Department of Electronics and Communication Engineering
DCRUST, Murthal, Sonepat, Haryana,India
rashmisharma8081@gmail.com

Priyanka
Associate Professor
Department of Electronics and Communication Engineering
DCRUST, Murthal, Sonepat, Haryana,India
priyankaiit@yahoo.co.in

***Abstract*—This paper entails the study and analysis of various image compression techniques. Compression plays an important role in today's world for efficient transmission and storage of data. Different techniques for digital image compression have been reviewed and presented that include Huffman Coding, Lemphel-Ziv-Welch (LZW) Coding, Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT). Image compression may be lossy or lossless. Lossless Compression is preferred for archival purposes and often for medical imaging, technical drawings, etc. Lossy compression are especially suited for natural images such as photographs or where low bit rates are used. Lossless Compression techniques include huffman coding and Lemphel-Ziv-Welch (LZW). Lossy Compression techniques include Discrete Cosine Transform (DCT), Discrete Wavelet transform (DWT). Finally a performance comparison is made between these techniques based on different parameters like Peak Signal to Noise Ratio (PSNR), Compression ratio (CR), Root Mean Square Error (RMSE), etc.**

Keywords—Compression ratio, Discrete Cosine Transform, Discrete Wavelet Transform, Huffman Coding, Lemphel-Ziv-Welch Coding, Peak Signal to Noise Ratio, Root Mean Square Error.

I. INTRODUCTION

An image usually consists of enormous amount of data and requires large number of space in the memory. If more number of data is required for transmission then it takes much time to deliver the data to the receiver. Thus by using image compression techniques the time consumption can be greatly reduced [1]. The image can be compressed when the correlation between one pixel and its neighbor pixels is very high, or the values of one pixel and its adjacent pixels are very similar.The objective is to reduce redundancy of the image data in order to store or transmit data in an efficient form. It helps in reducing the no. of bytes without making any changes in its original image, so that it will take less time, and less hard disk space to send a data from one place to another [2]. There are two types of image compression: Lossless Image Compression and Lossy Image Compression.Lossless compression technique is one of the most suitable image compression technique where compressed image will be same as original image. With a loss-less compression and decompression, the original and decompressed files are identical bit per bit. It has much more importance in different areas like medical diagnosis, remote sensing, weather forecasting, aircraft, radar etc [2]. Lossy compression is used when the user can tolerate some difference between original and reconstructed representation of data. It transforms and simplifies the data in a much larger reductions in file size than lossless compression. But it can distort the file's content due to the higher reduction [3].

II. COMPRESSION- TECHNICAL CHALLENGES

Choosing an efficient and cost-effective compression algorithm to provide an optimal solution is very challenging and complex and thus involves performing tradeoff analysis on various factors which are important to the application. Some of the challenges in the compression are:

A. *Compression Efficiency*

High compression efficiency provides high data rates. It is required in the applications like video-on -demand, streaming multimedia, video conferencing etc.

B. Resolution

Higher the compression ratio, lesser the space required for the compressed data, but they fail to get high resolution. Application like HDTV, medical imaging mobile data requires high quality images but they require higher bandwidth.

C. Error Resiliency

Applications like satellite imaging as in [4], remote sensing, requires robust error resiliency as they uses air transmission methods.

D. Power Consumption

Low power consumption is ensured in the battery operated devices used in the application like mobile phones, wireless sensor network [5] . This requires low computation memory for encoding and decoding.

E. Data Rates

A high data rate requires high compression efficiency and high computational memory and low latency. It is highly needed in networks to download files.

III. EVALUATION CRITERIA

There are various parameters to evaluate compression, like complexity of the algorithm, memory required to implement the algorithm, how fast the algorithm performs on a given machine (compression speed), the amount of compression (compression ratio), how closely the reconstruction resembles

ISSN No: - 2456- 2165

the original (image quality), reduced energy consumption etc [3].

A. Compression Ratio (CR)

It is defined as the ratio of number of bits in the original image to the number of bits in the compressed image. High compression ratios are desired. It is represented as 5:1, if 1MB file is compressed to 200KB file. With lossless compression, limited amount of compression is achieved.

B. Image Quality (IQ)

I) PSNR:

The quantity which measures the quality of the Reconstructed image compared to the original image is called Peak Signal to Noise ratio (PSNR), measured in decibel(dB) and is defined as :

$$PSNR = 20\log_{10}\left(\frac{255}{\sqrt{MSE}}\right) \quad (1)$$

where MSE is Mean Squared Error and 255 is the maximum possible pixel value in 8 bits.

2) MSE:

MSE is defined as cumulative square of the difference between original and reconstructed image. It is given as

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-\hat{I}(i,j)]^2 \quad (2)$$

where I is original image, $\hat{I}$ is approximation of reconstructed image. m and n are dimensions (width and height) of the image. Thus lower the MSE, higher is PSNR for an image.

C. Compression Rate

It is defined as the average number of bits required to represent a single sample (or pixel). It is measured in bits per pixel(bpp) for coding an image.

D. Distortion

It is defined as the difference between the original image and the reconstructed image. In loss less compression distortion toleration is always zero for any application. There is a trade off between minimizing the rate and keeping the distortion small. Rate distortion theory is the entropy theory.

E. Coding Complexity and Compression Speed

The coding complexity of a compression algorithm is a measure of computational requirement in terms of computational requirement in terms of software (instructions per second) and hardware (low power). Complex algorithms can reduce the compression speed.

F. Latency

Latency is the time taken to compress, send, decompress and display a file. The more advanced the compression algorithm, higher the latency.

IV. LOSSLESS COMPRESSION TECHNIQUES

Lossless image compression has been a significant issue in recent years due to the increasing demand for storing huge amounts of high quality multimedia data in a small storage. Applications like medical imaging, satellite imagery, FAX transmission etc. uses lossless compression techniques to store and transmit data [3]. As presented in Fig.l, the basic principle of Lossless compression algorithm is that any non- random file will contain duplicated information that can be reduced using some modeling as presented in Fig. 1 .

This model determines the probability of a bit or character appearing in the data set. It assigns shortest code to the most common data and longer code for uncommon data. Such coding techniques are called entropy coding.

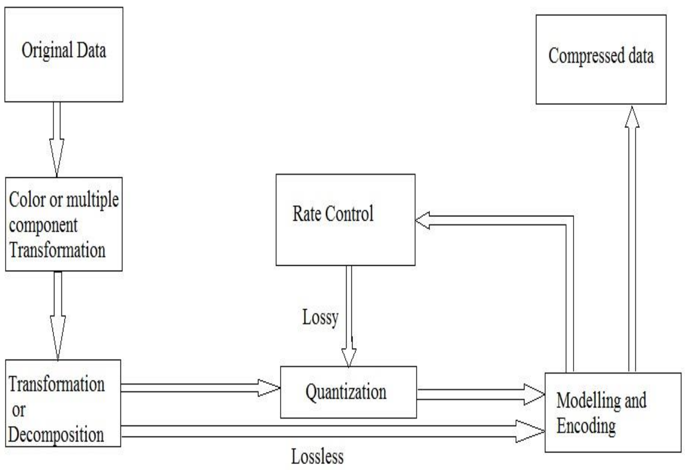

Fig. 1 Basic Structure of Lossless Compression

A) Huffman Coding

Huffman coding is an entropy encoding algorithm used for lossless image compression. Huffman coding is efficient technique for image compression to some extent [6]. The last two decades has seen considerable improvements in image and video compression techniques. Variable length coding, such as Huffman code, is widely used to increase coding efficiency. It uses the Huffman source-coding algorithm to generate the uniquely decipherable Huffman code with a minimum expected codeword length when the probability distribution of a data source is known to the encoder [7]. Entropy can be defined as a measure of information content; it will be able to represent the amount of bits used in the data in particular given image. Huffman coding uses a specific method for choosing the representation for particular images which results in a prefix code. The Compression of Images and data is both possible using Huffman Coding Algorithm. By using this Huffman algorithm we can be able to design the most efficient compression method. Huffman Coding comes under lossless technique here in lossless compression no information is lost during Image Compression.

In Huffman coding, more frequently occurred symbols will have shorter code words than symbol that occur less frequently & the two symbols that occur least frequently will have the same length. The first step in Huffman‘s approach is to create a series of source reductions by ordering the probabilities of the symbols in decreasing order and combining the two lowest probability symbols into a single symbol that replace them in the next source reduction. This

ISSN No: - 2456- 2165

compound symbol and its associated probability are placed in the first source reduction column so that the probabilities of the reduced source are also ordered from most to least probable. 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.

B) LZW

LZW compression technique is used to compress the images, characters, videos, audio, etc. Limitations in the image compression is that the number of colors is mainly considered and it is used for some cases. These limitations will be removing by using the bit plane slicing technique for both the RGB and gray scale images. Dictionary based approaches are used to scan a file and then perform searching on sequences of data or strings which occurs more than once in a scanned file. LZW compression works on single codes after replacing the strings of characters and it also adds a new character of strings in the dictionary in the encoding process. But in decoding we convert that single code into the characters of strings by using the static dictionary [16].

Compression in LZW

In this phase there are two synchronized events: building an indexed dictionary and compressing a string of symbols. The algorithm extracts the smallest substring that cannot be found in the dictionary from the remaining uncompressed string. It then stores a copy of this substring in the dictionary as a new entry and assigns it an index value. Compression occurs when the substring, except for the last character, is replaced with the index found in the dictionary. The process then inserts the index and the last character of the substring into the compressed string. As described in Fig 2, compression of string by LZW is done.

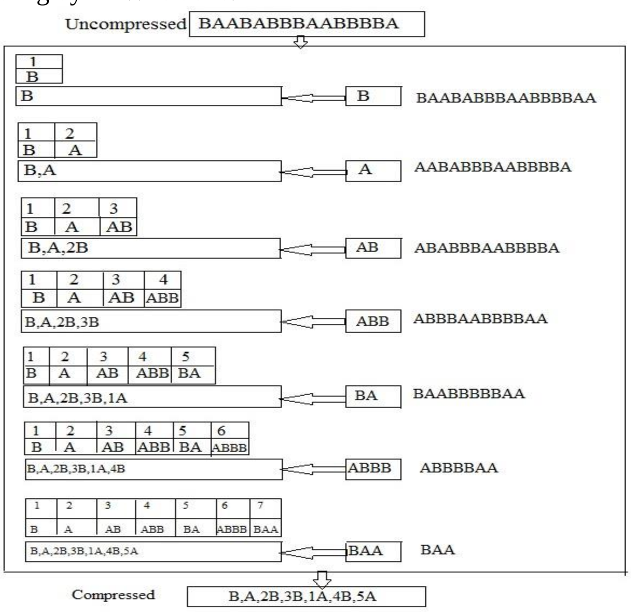

Fig. 2 Compression in LZW

Decompression in LZW

Decompression is the opposite of the compression process. The process extracts the substrings from the compressed string and tries to replace the indexes with the corresponding entry in the dictionary, which is empty at first and built up gradually . The idea is that when an index is received, there is already an entry in the dictionary corresponding to that index. As described in Fig. 3, the compression of string has been decompressed.

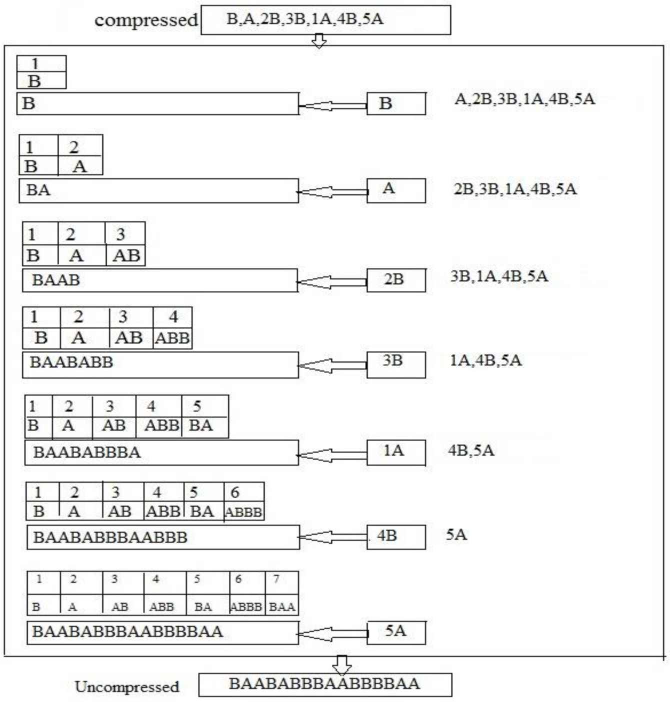

Fig. 3 Decompression in LZW

V. LOSSY COMPRESSION TECHNIQUES

Lossy Compression uses inexact approximations and partial data discarding for the representation of content. These techniques are used to reduce size of data for storage, handling and transmitting content. This is opposite to lossless compression, which doesnot degrade data.

Well designed lossy compression technique often reduces file sizes, such that the reproduced image includes degradation in respect to the original. Thus, all lossy techniques utilized for image compression have large compression rate as compared to lossless compression methods. Methodologies used for lossy compression consists of transform and predictive coding. Transform Coding techniques basically applies a Fourier- related transform; i.e. DCT and DWT.

A) DCT

The Discrete Cosine Transform (DCT) attempts to decorrelate the image data. After decorrelation each transform coefficient can be encoded independently without losing compression efficiency [9].

The DCT equation (Eq. 3) computes the i, j^{th} entry of the DCT of an image.

$$D(i,j)=\frac{1}{\sqrt{2N}}C(i,j)C(i,j)\sum_{x=o}^{M-1}\sum_{y=0}^{N-1}p(x,y)\cos\left[\frac{(2x+1)i\pi}{2M}\right]\cos\left[\frac{(2y+1)j\pi}{2N}\right] \quad (3)$$

$$C(u)=\begin{cases}1/\sqrt{2} & if \quad u=0\\ 1 & if \quad u>0\end{cases} \quad (4)$$

ISSN No: - 2456- 2165

Because the DCT use cosine functions, the resulting matrix depends on the horizontal, diagonal, and vertical frequencies. Therefore and image black with a lot of change in frequency has a very random looking resulting matrix, while and image matrix of just one color, has a resulting matrix of a large value for the first element and zeroes for the other elements [9].

DCT has possessions that, for an ordinary image, nearly all part of visually critical information of an image are concentrated in only couple of coefficient. All processing of coefficients are normalized by quantization process using quantization table with diverse scale provided by JPEG standard. The estimation of quantization is inversely proportional to quality of reproduced image. During quantization, the less significant frequencies components are discarded, and essential frequencies components that remain are make use of to recover image in decomposition process, after quantization, quantized coefficients are adjusted in a zigzag way for further compressed by an proficient coding algorithm as described in Fig. 4 and 5.

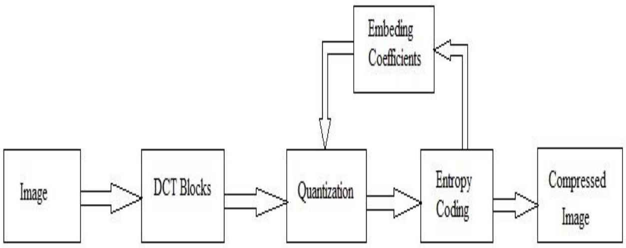

Fig. 4 Compression using DCT

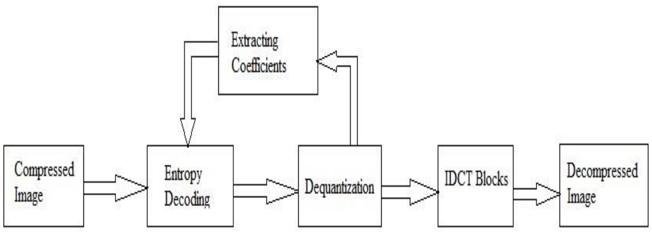

Fig. 5 Decompression using DCT

Discrete Cosine Transform (DCT) has many advantages:
(1) DCT can pack most data in least number of coefficients.
(2) DCT decreases the piece like form called blocking object that outcomes while limits among sub-images get to be noticeable . An image is represented to be as a two dimensional grid, 2-D DCT is utilized to process the DCT Coefficients of an image.

B) DWT

DWT plays an important role to compress the given image without the loss of any information in that particular image. DWT comes under lossless type of image compression. Here, DWT can be mainly used in the transformation of a discrete time signal to Discrete Wavelet Representation. DWT usually based on time-scale representation, which can be able to provide multi-resolution. Wavelets have more advantages over compressing signals. The wavelet transform is considered as the most advantageous and useful computational tools for a multiplicity of signal and image processing applications. Wavelet transforms are mainly used for images to reduce unwanted noise and blurring [1]. Wavelet transform has emerged as most powerful tool for both data and image compression. Wavelet transform performs multi resolution image analysis. The DWT has successfully been used in many image processing applications including noise reduction, edge detection, and compression. Indeed, the DWT is an efficient decomposition of signals into lower resolution and details. From the deterministic image processing point of view, DWT may be viewed as successive low-pass and high-pass filtering of the discrete time-domain signal [10]. In 2D image, the images are generally considered to be matrices with N rows and M columns. In wavelet transform, the decomposition of a particular image consists of two parts, one is lower frequency or approximation of an image (scaling function) and an other is higher frequency or detailed part of an image (wavelet function). Figure 6 explains Wavelet Filter decomposition of an image where four different sub-images are obtained; the approximation (LL), the vertical detail (LH), the horizontal detail (HL) and the diagonal detail (HH).

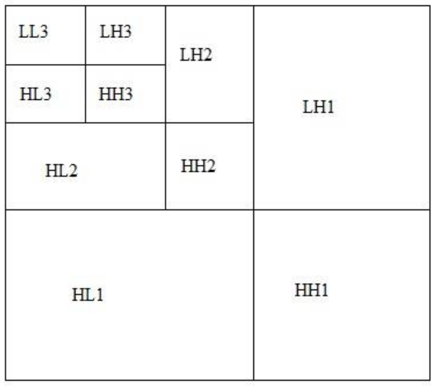

Fig. 6 Wavelet Filter Decomposition

VI. COMPARISON OF COMPRESSION TECHNIQUES

1) Lossless Compression Techniques

A. Huffman Coding

TABLE 1

Original File		Huffman Coding					
File	File Size	Comp File Size	Compression Ratio	Comp. Time	Decomp. Time	Entropy	Code Efficiency
1	22,094	13,826	62.5780755	16141	16574	4.788069	99.2827
2	44,355	27,357	61.6773757	54719	20606	4.810933	99.3618
3	11,252	847,5	67.4013509	3766	6750	5.026545	99.4386
4	15,370	8,961	58.3018868	5906	9703	4.354940	99.6695
5	78,144	45,367	58.0556409	156844	224125	4.540845	99.0655

The results in Table 1 show that entropy values lay between 4.3 and 5.1. To compress a single character of 1 byte, this algorithm needs only 4-5 bits. Code efficiencies are greater than 98% for all the cases. Thus this algorithm can be

ISSN No: - 2456- 2165

considered as an efficient algorithm [12]. According to the definition of the Code Efficiency, used code words can be further improved.

B. LZW Coding

TABLE 2

Original File			LZW			
File	File Size	No. of characters	Compressed File Size	Compression Ratio	Compression Time	Decomp. Time
1	22,094	21,090	13,646	61.76337	51906	7000
2	44,355	43,487	24,938	56.22365	167781	7297
3	11,252	10,848	7,798	69.30323	15688	3422
4	15,370	14,468	7,996	52,02342	21484	3234
5	78,144	74,220	24,204	30.97358	279641	11547

A dynamic dictionary is used by this algorithm and gives good compression ratios for the source text files. The disadvantage is that the size of the dictionary got increased with the size of the file since more and more entries are added by the algorithm. Table 3 shows low efficiency, because lot of resources is required to process the dictionary. This algorithm gives a good compression ratio which lies between 30% and 60% [12]. This is a reasonable value when it compared with the other algorithms. The compression ratio decreases as the file size increases, since the number of words can be represented by shorter dictionary entries.

TABLE 3 Saving Percentages

	HUFFMAN CODING	LZW CODING
1.	37.42	38.24
2.	38.32	43.78
3.	32.60	30.70
4.	42.70	47.98
5.	41.94	63.03

Saving percentages of all selected algorithms are shown in Table 3. LZW Coding shows the best results and average values are given by Huffman Coding.

The compression time is increased as file size increases. Compression times is average for Huffman approaches. The LZW algorithm works well for small files but not for the large files due to the size of the dictionary. Decompression times of all the algorithms are less than 500000 milliseconds except the LZW.

2) Lossy Compression Techniques

For Lossy Compression, Compression Ratio (CR) and Peak Signal to Noise ratio (PSNR) are measured. In Lossy Compression we analyse graphically various parameters like Compression ratio Analysis, PSNR Analysis and Saving Storage Analysis.

A) Compression Ratio Analysis

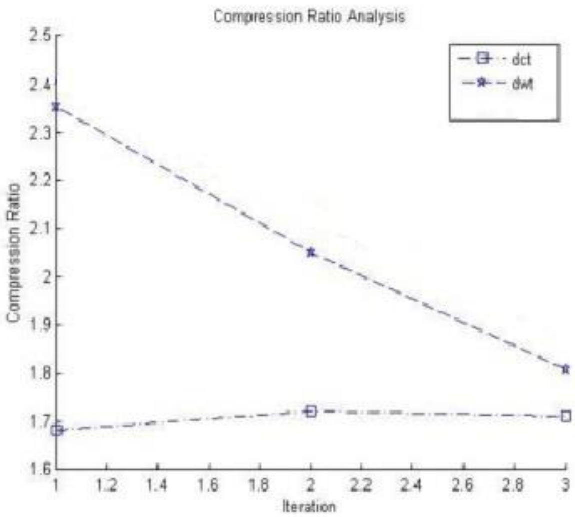

Fig. 7. Compression Ratio Analysis

B) PSNR Analysis

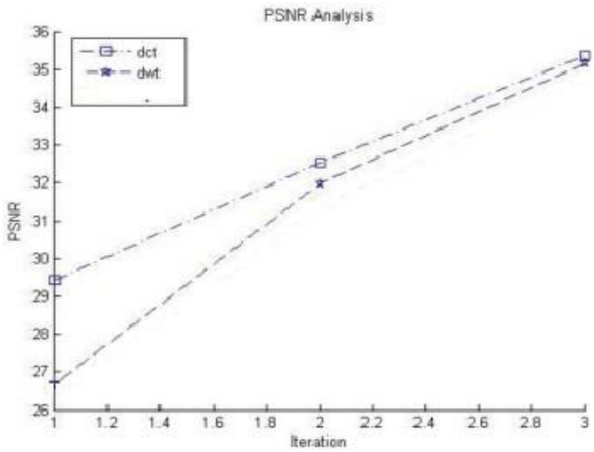

Fig.8. PSNR Analysis

C) Storage Saving Analysis

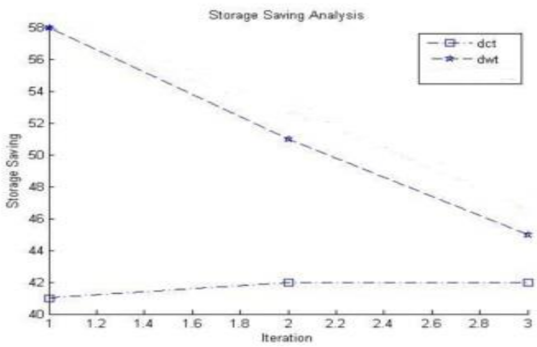

Fig.9. Storage Saving Analysis

ISSN No: - 2456- 2165

Our result demonstrates that DWT gives better compression ratio without losing more data of image. DCT defeats this inconvenience since it needs less handling force, but it gives less compression ratio [12]. DCT based standard JPEG uses blocks of images, but there are still relationship exits crosswise over blocks. Block limits are observable. Blocking artifacts can be seen at low bit rates. In wavelet, there is no compelling reason to block the image. It facilitates dynamic transmission of the image (adaptability).

VII. CONCLUSION

In this paper, a comparative analysis is done for various compression techniques. Compression techniques for lossless compression includes huffman coding and LZW coding. LZW approach works better as the size of the file grows up to a certain amount, because there are more chances to replace identified words by using a small index number. However, it can not be considered as the most efficient algorithm, because it can not be applied for all the cases. Compression techniques for lossy compression includes DCT, DWT. Out of these, DWT is best and gives good compression ratio without much data loss.

REFERENCES

[1] R.Praisline Jasmi, Mr.B.Perumal, Dr.M.Pallikonda Rajasekaran, "Comparison of Image Compression Techniques using Huffman Coding ,DWT, Fractal Algorithm" in *International Conference on Computer Communication and Informatics* (ICCCI -2015), Jan. 08 – 10, 2015, Coimbatore, INDIA, 978-1-4799-6805-3/15.

[2] Krishan Gupta, Mukesh Sharma, Kirti Saneja, "A New Lossless KMK Technique for Image Compression" in *International Conference on Reliability, Optimization and Information Technology* - ICROIT 2014, India, Feb 6-8 2014, 978-1-4799-2995-5/14.

[3] Smitha Rao, Pratima Bhat, "Evaluation of Lossless Compression Techniques" in *IEEE ICCSP 2015 Conference,* 978-1-4799-8081-9/15.

[4] L.N. Faria, L.M.G. Fonseca and M.H.M. Costa, " Performance Evaluation of Data Compression Systems Applied to Satellite Imagery", *Hindawi Publishing Corporation, Journal of Electrical and Computer Engineering*, Volume 2012, Article ID 471857, 15 pages, doi: 10,1155/2012/471857.

[5] P.Kumsawal, N.Pimpru, K.Attakimongeol and A.Srikaew, " Wavelet Based Data Compression Technique for Wireless Sensor Networks", published in WASET, vol:7, 2013-05-21.

[6] Jagdish 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, 2010.*

[7] Jian-Jiun Ding, Hsin-Hui Chin, Wei-Yi Wei, " Adaptive Golomb Code for Joint Geometrically" in *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 25, 4,April,2013.

[8] Sadhana Singh, Preeti Pandey, " Enhanced LZW Technique for Medical Image Compression" in *International Conference on Computing for Sustainable Global Development (INDIACom),* 978-9-3805-4421-2/16, February, 2016.

[9] Nusrat Ahmed Surobhi, "Amalgamation of SVD to JPEG for Enhanced Performance", *International conference on Information & Communication Technology-2007.*

[10] Karami A, Yazdi M, Mercier G, "Compression of Hyper Spectral Images Using Discrete Wavelet Transform and Tucker Decomposition" Selected Topics in *Applied Earth Observations and remote Sensing, IEEE Journal*, Vol:5, no.: 2, pp.444,450,April,2012.

[11] Singh H.K., Tomar S.K., Singh P, " Analysis of Multispectral Image using discrete wavelet Transform", in *Advanced Computing and Communication Technologies* (ACCT), 2013 Third International Conference, Vol., no., pp. 59,62,67, April,2013.

[12] S.R. Kodituwakku,U. S.Amarasinghe, "*Comparison of Lossless Data Compression Algorithms for text data*",Indian Journal of Computer Science and Engineering, Vol 1 No 4 416-425, ISSN : 0976-5166.

[13] Xingsong Hou, Jing Yang, Guifeng Jiang, Xueming Qian, "Complex SAR Image Compression Based on Directional Lifting Wavelet Transform With High Clustering Capability," *Geosciences and Remote Sensing, IEEE Transactions* on , vol.51, no.1, pp.527,538, January 2013.