Tutorial review on existing image compression techniques

Bhavna Pancholi, Reena Shah, Mitul Modi

Department of Electrical Engineering
Faculty of Technology and Engineering
The M. S. University, Baroda

Abstract: *Image compression is a technique that can reduce the storage space of images and videos. Moreover it is helpful to increase storage and transmission process performance. Image compression reduces the image fidelity, when an image is compressed at low bitrates. Hence, the compressed images suffer from block artifacts. The prime focus of image compression technique is to reduce the number of the image pixel elements without affecting the original image. It can be achieved by removing the redundancy present in the image. This paper shows review of different existing techniques which is available lossless and lossy compression such as Huffman coding, LZW coding, DCT, DWT, VQ compression etc. and based on the review guidelines are provided to choose best algorithm for an image.*

Keywords: image compression, lossless and lossy compression, LZW, DC, DWT, VQ compression.

1. Introduction

An image is an artifact that records visual perception. Images are vital documents today's digital world. And to work with them in some applications there is need to be compressed. Image compression plays a very important role in the transmission and storage of image data. The main aim of image compression is to represent an image in the fewest number of bits without losing the required information content within an original image.

Now a days many compression techniques [1] are rapidly developed for compress large data such as images duo to advancement in technology. Large amount of data must be stored and handle in such a way using efficient algorithm usually succeed in compression.

The prime focus of image compression is to reduce redundancy of image before it transmit and store to appropriate place. The compression ratio [2] defined as follow:

$$C = N_1 / N_2$$

Where N_1 = Actual Image
   N_2 = compressed Image

1.1 Image Format
Images are classified by following.
   A) *JPEG:* Joint Photographic Expert Group (JPEG) is an excellent way to store 24-bit photographic images, such as those used for imaging and multimedia *applications*. JPEG 24-bit (16 million color) images are superior in appearance to 8-bit (256 color) images on a Video Graphics Array (VGA) display and are at their most spectacular, when using 24-bit display hardware (which is now quite inexpensive). JPEG was designed to compress, color or gray-scale continuous-tone images of real-world subjects, photographs, video stills, or any complex graphics, that resemble natural subjects. Animations, ray tracing, line art, black-and-white documents, and typical vector graphics don't compress very well under JPEG and shouldn't be expected to. And, although JPEG is now used to provide motion video compression, the standard makes no special provision for such an application.

   B) *PNG:* Portable Network Graphics (PNG) is a file format for lossless image compression. Typically, an image in a PNG file can be 10% to 30% more compressed than in a GIF format. It allows to make a trade-off between file size and image quality when the image is compressed. It produces smaller files and allows more colors. PNG also supports partial transparency. Partial transparency can be used for many useful purposes, such as fades and antialiasing for text.

   C) *GIF:* Graphics Interchange Format (GIF) is useful for images that have less than 256-(2^8) colors, grayscale images and black and white text. The primary limitation of a GIF is that it only works on images with 8-bits per pixel or less, which means 256 or fewer colors. Most color images are 24 bits per pixel. To store these in GIF format that must first convert the image from 24 bits to 8 bits. GIF is a lossless image file format. Thus, GIF is "lossless" only for images with 256 colors or less. For a rich, true color image,

GIF may "lose" 99.998% of the colors. It is not suitable for photographic images, since it can contain only 256 colors per image.

D) *TIFF:* The TIFF (Tagged Image File Format) is a flexible format that can be lossless or lossy compression [14]. It normally saves 8 bits or 16 bits per color (red, green, blue) for 24-bit and 48-bit totals, respectively. The details of the image storage algorithm are included as part of the file. In practice, TIFF is used almost exclusively as a lossless image storage format that uses no compression at all. TIFF files are not used in web images. They produce big files, and more importantly, most web browsers will not display TIFFs.

E) *BMP:* The Bitmap (BMP) file format handles graphics files within the Microsoft Windows OS. Typically, BMP files are uncompressed, hence they are large; advantage is that their simplicity, wide acceptance, and use in Windows program

F) *RAW:* RAW refers to a family of raw image formats (output) that are options available on some digital cameras. These formats usually use a lossless or nearly-lossless compression, and produce file sizes much smaller than the TIFF formats of full-size processed images from the same cameras. The raw formats are not standardized or documented. Though lossless, it is a factor of three or four smaller than TIFF files of the same image. The disadvantage is that there is a different RAW format for each manufactures and so has to use the manufacturer's software to view the images

1.2 Types of Redundancy:
Redundancy means to reduce some data which is not relevant or provide no information. There are three types of redundancy

A) *Coding Redundancy:* A code is a system of symbols used to represent a body of information. Each piece of information or event assigned a sequence of code symbols called code word. This type of coding is always reversible and usually implemented using lookup tables (LUTs). Examples of image coding schemes that explore coding redundancy are the Huffman codes and the arithmetic coding technique.

B) *Inter – pixel Redundancy:* In image neighboring pixels are not statistically independent. It is due to the correlation between the neighboring pixels of an image. This type of redundancy is called Inter-pixel redundancy. This type of redundancy is sometime also called spatial redundancy. This redundancy can be explored in several ways, one of which is by predicting a pixel value based on the values of its neighboring pixels.

C) *Psycho- Visual Redundancy:* Many experiments on the psycho physical aspects of human vision have proven that the human eye does not respond with equal sensitivity to all incoming visual information; some pieces of information are more important than others. Most of the image coding algorithms in use today exploit this type of redundancy, such as the Discrete Cosine Transform (DCT) based algorithm at the heart of the JPEG encoding standard.

2. Basic Image Compression Model

Image compression systems are composed of two distinct structural blocks: an encoder and a decoder. As shown in the fig 1, the encoder is responsible for reducing the coding, inter pixel and psycho visual redundancies of input image. In first stage, the mapper transforms the input image into a format designed to reduce inter pixel redundancies. The second stage, qunatizer block reduces the accuracy of mapper's output in accordance with a predefined criterion. In third and final stage, a symbol decoder creates a code for quantizer output and maps the output in accordance with the code. These blocks perform, in reverse order, the inverse operations of the encoder's symbol coder and mapper block. As quantization is irreversible, an inverse quantization is not included.
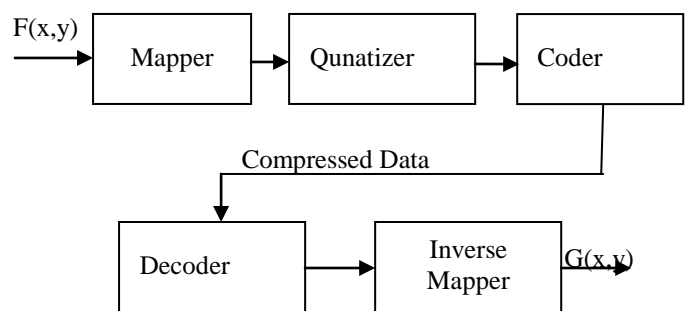

Fig.1 Compression System

Benefits of Compression:
- It provides a potential cost savings associated with sending less data over switched telephone network where cost of call is really usually based upon its duration.
- It not only reduces storage requirements but also overall execution time.
- It also reduces the probability of transmission errors since fewer bits are transferred.
- It also provides a level of security against illicit monitoring.
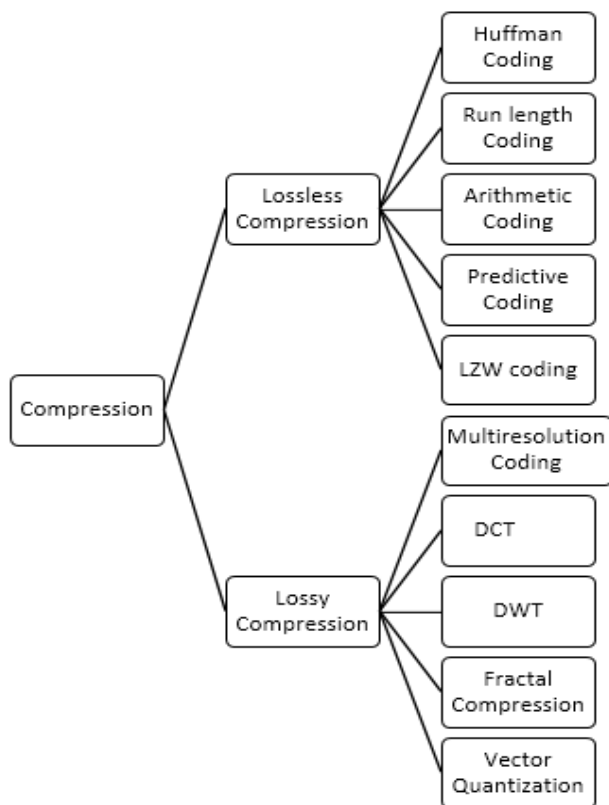
3. Image Compression Techniques

Fig.2 Existing Techniques

Digital image compression can be divided mainly in two categories: lossless and lossy compression. When lossless data is decompressed, the resulting image is identical to the original. Lossy compression algorithms result in loss of data and the decompressed image is not exactly the same as the original.

3.1 Lossless Compression

This is an image compression technique in which the image produced after applying compression method [3] is numerically identical to the original image i.e. there is no loss of image information. It is used in the field of medical imaging, for clip arts etc.
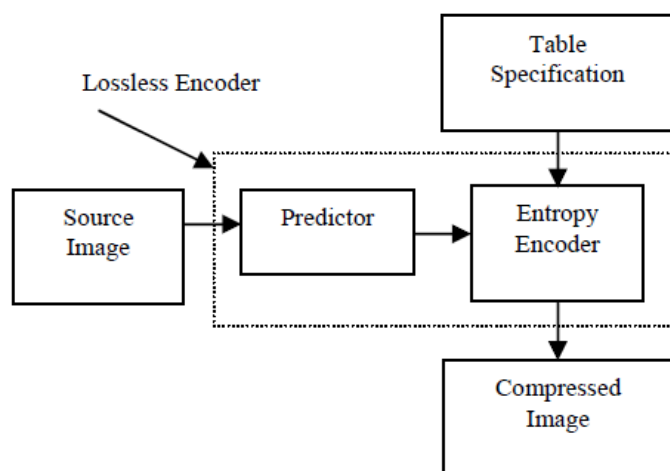

Fig. 3 Block diagram of Lossless compression Model

Various techniques under this category are:

A) *Huffman Coding:* In computer science and information theory, Huffman coding is an entropy encoding algorithm used for lossless data compression. It was invented by Huffman. Huffman coding [4] today is often used as a "back-end" to some other compression methods. The term refers to the use of a variable-length code table for encoding a source symbol where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol. The pixels in the image are treated as symbols. The symbols which occur more frequently are assigned a smaller number of bits, while the symbols that occur less frequently are assigned a relatively 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 other symbol.

Original   Compressed
Fig.4 Huffman Coding

Huffman has some limitations. So, a better version called Double Huffman coding was developed which provides better results.

B) *Run Length Coding:* RLE is a lossless data compression technique which replaces data by a (length, value) pair, where value is the repeated value and length is the number of repetitions. RLE is a simple form of data compression in which data is in the form of runs. Runs is sequences in which the same data value occurs in many consecutive data elements are stored as a single data value and count, rather than as the original run.

For example, consider a screen containing plain black text on a solid white background. There will be many long runs of white pixels in the blank space, and many short runs of black pixels within the text. Consider a single scan line, with B representing a black pixel and W white pixel:

WWWWWWWWWWWWBBBWWWWWWW
WWWBBBWWWWWWWWWWWWWWWW
WWWWWWWBWWWWWWWWWWWWWW

If we apply the run-length encoding (RLE) data compression algorithm to the above hypothetical scan line, we get the following:

12W3B10W3B23W1B13W

65 characters are represented using only 18 characters. But, the actual format used for the storage of images is generally binary rather than ASCII characters like this, but the principle remains the same.

C) *Arithmetic Coding:* AC is the compression technique for lossless encoding that represents a message as some finite intervals between 0 and 1 on the real number line. Basically, it divides the intervals between 0 and 1 into a number of smaller intervals corresponding to probabilities of the message's

symbols. Then, the first symbol selects an interval, which is further divided into smaller intervals. The next input symbol selects one of these intervals, the procedure is repeated. As a result, the selected interval narrows with every symbol, and in the end, any number in the final interval can be used to represent a message. Each bit in the output code refines the precision of the value of input code in the interval. AC is the most efficient method to code symbols according to the probability of their occurrence. A variation of arithmetic coding is the Q-Coder, developed by IBM in the late 1980's.

D) *Predictive Coding:* Lossless predictive coding predicts the value of each pixel by using the value of its neighboring pixels. Therefore, every pixel encoded with a prediction error rather than its original value. These errors are much smaller compared with original value so that fewer bits are required to store them. Such as DPCM (Differential Pulse Code Modulation) is a lossless coding method, which means that the decoded image and the original image have the same value for every corresponding element. A variation of the lossless predictive coding is adaptive prediction that splits the image into blocks and computes the prediction coefficients independently for each block to high prediction performance.

E) *LZW coding:* Ziv and Lempel can be developed in 1977 and 1978. Terry Welch improved the scheme in 1984 (called LZW compression).Lempel Ziv (LZ) encoding is an example of dictionary-based encoding. LZW (Lempel-Ziv–Welch) is a dictionary based coding. Dictionary based coding can be static or dynamic. In static dictionary coding, during the encoding dictionary is fixed and decoding processes. In dynamic dictionary coding, the dictionary is updated on fly. LZW [5] [6] is broadly used in computer industry and is implemented as compress command on UNIX.The main idea is to create a table (a dictionary) of strings used for transmission session. If both the sender and the receiver have a copy of the table, then previously occurred strings can be replaced by their index in the table to minimize the amount of information transmitted. It works on file formats such as GIF, TIFF, and PDF.
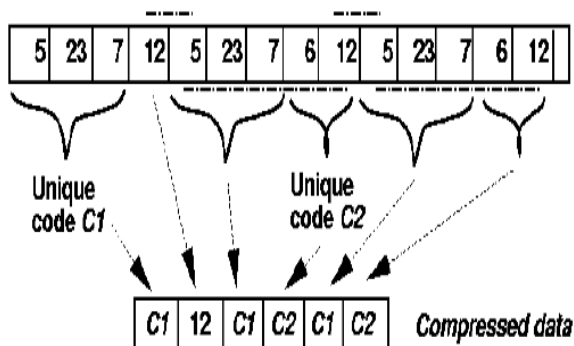

Fig. 5 LZW Coding

3.2 Comparison of Lossless Compression Techniques

Method	Pros.	Cons.	Application
RLE	It can be used to compress data made of	It does not work well at all on	Used in making Palette and in

	any combination of symbols. It does not need to know the frequency of occurrence of symbols.	continuous-tone images such as photographs,	fax machine
Huffman Coding	High transmission speed Easy to implement	Compression of image files that contain long runs of identical pixels by Huffman is not as efficient when compared to RLE.	It used as a "back-end" to some other compression method.
LZW Coding	No need to Analyse the incoming text	Only good at text files but not on other files	used in the GIF and TIFF image format

Table No. 1 Comparison

3.3 Lossy Compression

Unlike lossless technique, the image produced after compression is not identical to the original image. So, it is called lossy technique. It provides high compression ratio than lossless technique [7].It is further categorized into following types:
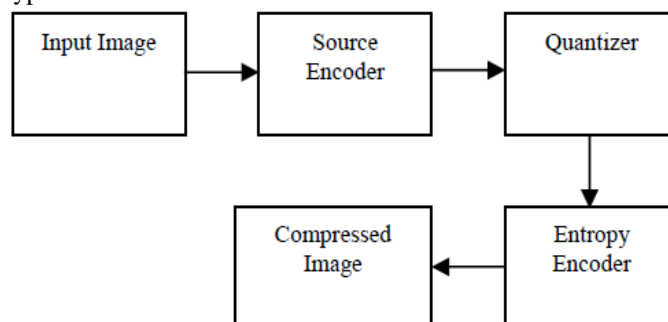

Fig. 6 Block diagram of Lossy Compression Model

The most common example of lossy compression is JPEG. An algorithm that restores the presentation to be the same as the original image. Reconstruction of the image is an approximation of the original image, therefore the need of measuring of the quality of the image for lossy compression technique. Lossy compression technique provides a higher compression ratio than lossless compression.

Major performance considerations of a lossy compression scheme include:
- Compression ratio
- Signal to noise ratio
- Speed of encoding & decoding

A) *Multiresolution Coding:* HINT (hierarchical interpolation) is a multiresolution coding scheme based on sub-samplings. It starts with a low resolution

version of the original image, and Interpolates the pixel values to successively generate higher resolutions. The errors between the interpolation values and the real values are stored, along with the initial low-resolution image. Compression is achieved since both the low-resolution image and the error values can be stored with fewer bits than the original image. Laplacian

Pyramid is another multiresolution image compression method developed by Burt and Adelson. It successively constructs lower resolution versions of the original image by down sampling so that the number of pixels decreases by a factor of two at each scale. The differences between successive resolution versions together with the lowest resolution image are stored and utilized to perfectly reconstruct the original image. But it cannot achieve a high compression ratio because the number of data values is increased by 4/3 of the original image size.

B) *Discrete Cosine Transform (DCT):* The JPEG/DCT still image compression has become a standard recently. JPEG is designed for compressing full-color or grayscale images of natural, real-world scenes.

To exploit this method, an image is first partitioned into non overlapped 8×8 blocks. A discrete Cosine transform (DCT) [8, 9] is applied to each block to convert the gray levels of pixels in the spatial domain into coefficients in the frequency domain. The coefficients are normalized by different scales according to the quantization table provided by the JPEG standard conducted by some psycho visual evidence. The quantized coefficients are rearranged in a zigzag scan order to be further compressed by an efficient lossless coding strategy such as run length coding, arithmetic coding, or Huffman coding. The decoding is simply the inverse process of encoding. So, the JPEG compression takes about the same time for both encoding and decoding. The encoding/ decoding algorithms provided by an independent JPEG group [9] are available for testing real-world images. The information loss occurs only in the process of coefficient quantization. The JPEG standard defines a standard 8×8 quantization table [9] for all images which may not be appropriate. To achieve a better decoding quality of various images with the same compression by using the DCT approach, an adaptive quantization table may be used instead of using the standard quantization table.
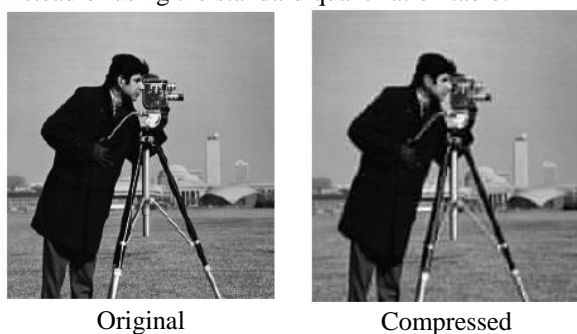
Original   Compressed

Fig. 7 Reconstructed image after applying DCT

C) *Discrete Wavelet Transform:* Wavelets are functions outlined over a finite interval and having an average value of zero. The fundamental plan of the wavelet transform is to represent any arbitrary function (t) as a superposition of a set of such wavelets or basis functions.[15] Image compression supported based on transforms has recently received an increasing interest [10] [12] [13][14]. The current progressive wavelet approach applies a wavelet rework on pictures in a pyramid fashion up to the required scale victimization the idea of multiresolution signal decomposition with the wavelet illustration [11] and also the conception of embedded zero tree wavelet (EZW) based on the decaying spectrum hypothesis. [14] during a pyramidical structure once a particular scale of wavelet transforms on an image, an algorithmic rule [14] in turn determines if a constant is critical in the spatial-frequency domain to make a significance map consisting of the sign (+ or −) of a major coefficient, an insignificant image, and a zero tree image.[15]
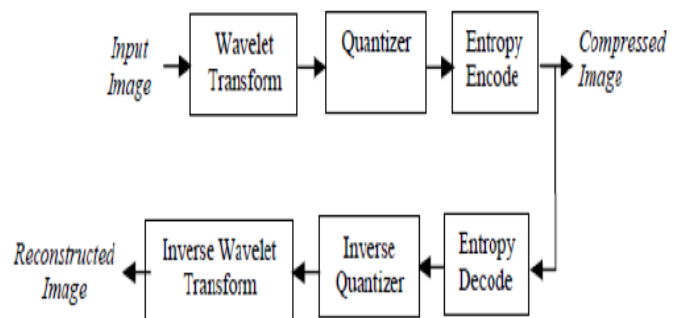
Fig. 8 DWT Model

Why DWT ?

Despite all the advantages of JPEG compression schemes based on DCT namely simplicity, satisfactory performance, and availability of special purpose hardware for implementation; these are not without their shortcomings. Since the input image needs to be ``blocked,'' correlation across the block boundaries is not eliminated. This results in noticeable and annoying ``blocking artifacts'' particularly at low bit rates as shown in Fig.1. Lapped Orthogonal Transforms (LOT) [16] attempt to solve this problem by using smoothly overlapping blocks. Although blocking effects are reduced in LOT compressed images, increased computational complexity of such algorithms do not justify wide replacement of DCT by LOT.

| Original | Compressed |

Fig. 9 Reconstructed image after applying DWT

D) Vector Quantization (VQ) Compression: A vector quantizer is composed of two operations. The first is the encoder, and the second is the decoder. The encoder takes an input vector and outputs the index of the codeword that offers the lowest distortion. In this case the lowest distortion is found by evaluating the Euclidean distance between the input vector and each codeword in the codebook. Once the closest code word is found, the index of that codeword is sent through a channel (the channel could be computer storage, communications channel, and so on). When the encoder receives the index of the code word, it replaces the index with the associated code word.

The fundamental idea of VQ[16] for image compression is to establish a codebook consisting of code vectors such that each code vector can represent a group of image blocks of size m × m, (m=4 is always used). An image or a set of images is first partitioned into m × m non overlapping blocks which are represented as m2-tuple vectors, called training vectors. The size of training vectors can be very large. For example, a 512 × 512 image contributes 16,384 training vectors.

The goal of codebook design is to establish a few representative vectors, called code vectors of size 256 or 512, from a set of training vectors. The encoding procedure is to look for a closest code vector in the codebook for each non overlapped 4 × 4 block of an image to be encoded. The most important work is to design a versatile codebook. Nasrabadi and King [17] give a good review of VQ. Chen's comparison [19] indicates that a codebook developed based on LBG [18] algorithm generally has higher PSNR values over some other schemes despite its slow off-line training. In this paper, we adopt LBG algorithm for training a codebook of size 256×256 to meet a desired 0.5 bpp compression ratio.

E) Fractal Compression: Fractal theories are totally different from the others. M. Barnsley introduced the fundamental principle of fractal image compression in 1988 [21]. Fractal image compression is also called as fractal image encoding because compressed image is represented by contractive transforms and mathematical functions required for reconstruction of original image. Contractive transform ensures that, the distance between any two points on transformed image will be less then the distance of same points on the original image [21]. These transforms are composed of the union of a number of affline mappings on the entire image, known as iterated function system (IFS) [20], [21]. Barsnley has derived a special form of the Contractive Mapping Transform (CMT) applied to IFS's called the College Theorem [20, 21]. The usual approach of fractal image compression is based on the college theorem, which provides distance between the image to be encoded and the fixed point of a transform, in terms of the distance between the transformed image and the image itself. This distance is known as college error and it should be as small as possible. A.E. Jacquin gave first publication on Fractal image compression with partitioned IFS (PIFS) in 1990 [20], [22], [23]. In Jacquin's method the image is partitioned in sub images called as 'Range blocks' and PIFS are applied on sub-images, rather than the entire image. Locating the range blocks on their respective position in image itself forms the entire image. Temporary images used to form range blocks are known as domain blocks.
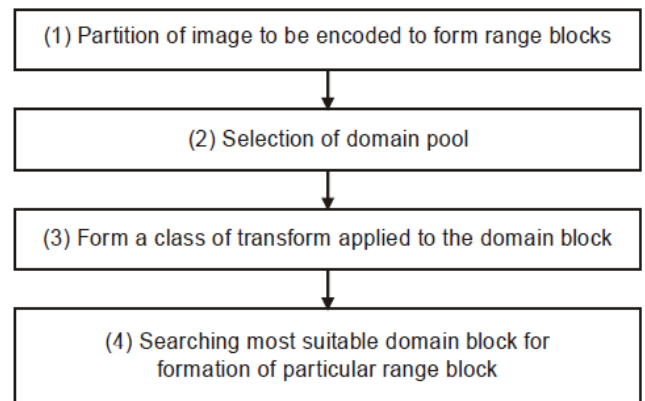

Fig. 10 Domain Blocks

Experimental Results:

Original Image

Fig. 11 Original Image (a) Wavelet (b) DCT (c) VQ (d) Fractal Algorithms

Method	PSNR Values (in dB)	CPU Time	
		Encoding	Decoding
Wavelet	34.66	0.35 sec	0.27 sec
DCT	31.73	0.12 sec	0.12 sec
VQ	29.28	2.45 sec	0.18 sec
Fractal	29.04	5.65 sec	1.35 sec

Table No. 2 Performance Evaluation [15]

3.4 Comparison of Lossy Compression Techniques

Method	Pros.	Cons.	Application
DWT	High Compression Ratio	Coefficient quantization Bit allocation	photographic images
DCT	Current Standard	Coefficient (dct) quantization Bit allocation	photographic images
VQ	Simple decoder No-coefficient	Slow codebook generation Small bpp	Used for lossy data compression, lossy data correction, pattern recognition and density estimation.
Fractal	Good mathematical Encoding-frame	Slow Encoding	Reproduction of damaged images

Table No. 3 Comparison of Lossy Compression Techniques.

3.5 Novel hybrid compression using DWT and Neural Network:

All the methods discussed above does not provide much compression ratio.T his, a hybrid approach was used which combines both DWT and Neural networks to provide substantial improvements [24].Wavelet based image compression provides higher compression ratios .Neural networks also produce most efficient results as compared to traditional approaches. With wavelet transform based compression, the quality of compressed images is usually high. Moreover wavelets do a better job of handling discontinuities in data. Neural networks have proved to be useful because of their parallel architecture and flexibility.
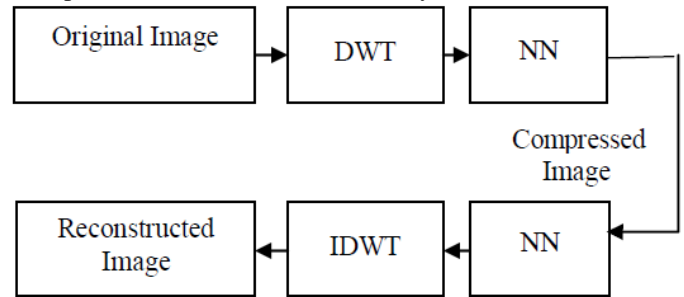

Fig. 12 Fusion Coompression Model

4. Conclusion

Images are vital documents these days to figure with them in some applications they have to be compressed, additional or less looking on the aim of the application such as transmission and storage in knowledge bases. In this paper we have reviewed different types of image format and their compression algorithms. Based on that, we find that lossy compression provided high compression ratio compared to lossless. Lossless compression is best suited for text compressions. All methods of lossy compression have good compression ratio, when images having more than 0.5 bpp .Moreover compression of image is mainly depended on quality of image. If the quality of image is very poor (below 0.25 bpp), DWT methods is superior to other methods.

References

[1] Mohammad Kabir Hossain, Shams MImam , Khondker Shajadul Hasan, and William Perrizo, "A Lossless Image Compression Technique Using Generic Peano Pattern Mask Tree," IEEE, pp. 317-322,2008.

[2] Uli Grasemann and Risto Miikkulainen," Effective Image Compression using Evolved Wavelets,"ACM, pp. 1961-1968, 2005.

[3] Ming Yang and Nikolaos Bourbakis, "An Overview of Lossless Digital Image Compression Techniques," IEEE, pp. 1099-1102,2005.

[4] Mridul Kumar Mathur, Seema Loonker and Dr. Dheeraj Saxena "Lossless Huffman Coding Technique For Image Compression And Reconstruction Using Binary Trees," IJCTA, pp. 76-79, 2012.

[5] Manjinder kaur, G. k., "A Survey of Lossless and Lossy Image Compression Techniques," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3,no. 2,pp. 323-326,February2013.

[6] Gonzalez, R. C, "Digital Image Processing ," Pearson Education, India,2009.

[7] Sonal, Dinesh kumar,"A study of various image compression techniques",Department of computer science and engineering,Unpublished.

[8] Rao, K. R., Yip, P., "Discrete Cosine Transforms - Algorithms, Advantages, Applications", Academic Press, 1990.

[9] Wallace, G. K. "The JPEG Still Picture Compression Standard", Comm. ACM, vol. 34, no. 4, April 1991, pp. 30- 44.

[10] Buccigrossi, R., Simoncelli, E. P., "EPWIC: Embedded Predictive Wavelet Image Coder".

[11] Mallat, S.G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Trans. PAMI, vol. 11, no. 7, July 1989, pp. 674-693.

[12] N.M. Nasrabadi, R.A. King, "Image coding using vector quantization: a review", IEEE Trans. On Communications,vol. 36, 957-571, 1988.

[13] Y.W. Chen, "Vector Quantization by principal component analysis", M.S. Thesis, National Tsing Hua University, June, 1998.

[14] H.S.Chu, "A very fast fractal compression algorithm", M.S. Thesis, National Tsing Hua University, June, 1997.

[15] Sachin Dhawan, "A Review of Image Compression and Comparison of its Algorithms", I S S N : 2 2 3 0 - 7 1 0 9 (On l i n e) | I S S N : 2 2 3 0 - 9 5 4 3, IJECT Vol. 2, Issue 1, March 2011.

[16] Malavar, H. S., "Signal Processing with Lapped Transforms", Norwood, MA, Artech House, 1992.

[17] Chan, Y. T., "Wavelet Basics", Kluwer Academic Publishers,Norwell, MA, 1995.

[18] N.M. Nasrabadi, R.A. King, "Image coding using vector quantization: a review", IEEE Trans. On Communications,vol. 36, 957-571, 1988.

[19] Y.W. Chen, "Vector Quantization by principal component analysis", M.S. Thesis, National Tsing Hua University, June, 1998.

[20] Y. Fisher, Fractal Image Compression: Theory and Application. New York: Springer-Verlag, (1994).

[21] M. Barnsley, Fractals Everywhere. New York: Academic, (1988).

[22] A.E. Jaquin, "Image coding based on a fractal theory of iterated contractive image transformation", IEEE Trans. On Image Processing, 1(1): (1992).

[23] A.E Jaquin, "Fractal image coding: A review", Proceeding of tile IEEE, 81(10): (1993).

[24] Er. Shruti Puniani, Er. Nishi Madaan. "Various Image Compression Techniques: A Review" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 4, April 2014