A Review: Various Approaches for JPEG Image Compression

Chandra Dev[1], Chandra Shekhar Rai[2], Krishna Gopal Bajpai[3] , Avinash Kaushal[4]

[1,2,3] Students B.Tech, [4] Assistant Professor

Department of EIE, Galgotias College of Engineering & Technology

Greater Noida, Uttar Pradesh- 201308, India

Abstract **-Image compression is used specially for the compression of images where tolerable degradation is required. With the wide use of computers and consequently need for large scale storage and transmission of data, efficient ways of storing of data have become necessary. With the growth of technology and entrance into the Digital age, the world has found itself amid a vast amount of information. Dealing with such enormous information can often present difficulties. Image compression is minimizing the size in bytes of a graphics file without degrading the quality of the image to an unacceptable level. It also reduces the time required for images to be sent over the Internet or downloaded from Web pages. JPEG and JPEG 2000 are two important techniques used for image compression. JPEG image compression standard use DCT. The discrete cosine transform is a fast transform. It is a widely used and robust method for image compression. It has excellent compaction for highly correlated data.**

Keywords—JPEG Compression Note: There should no nonstandard abbreviations, acknowledgments of support, references or footnotes in the abstract.

1. INTRODUCTION

Processing of images which are digital in nature by digital computer is known as digital image processing. Image processing is used for improvement of pictorial information for human perception. An image can be represented as a function $f(x,y)$ expressed in a two dimensional spatial co-ordinate system, where $f(x,y)$ must be non zero and finite. Image compression is a process to compress the image for reducing the storage required to save an image or bandwidth required to transmit. Image compression system is composed of two distinct structural blocks. They are encoder and decoder. When original image is fed into the encoder a set of symbols from the input data are created and are used to represent the image. The best image quality at a given bit-rate (or compression rate) is the main goal of image compression, however, there are other important properties of image compression schemes.

Scalability generally refers to a quality reduction achieved by manipulation of the bitstream or file (without decompression and re-compression). Other names for scalability are progressive coding or embedded bits stream. Despite its contrary nature, scalability also may be found in lossless codec, usually in form of coarse-to-fine pixel scans. Scalability is especially useful for previewing images while downloading them (e.g., in a web browser) or for providing variable quality access to e.g., databases. There are several types of scalability:

1.1 PROCESSING TECHNIQUE:

This processing technique may be, Image enhancement, Image reconstruction, and Image compression

1.1.1 Image Enhancement:

This process does not increase the inherent information content in data. It includes gray level & contrast manipulation, noise reduction, edge crispening and sharpening, filtering, interpolation and magnification, pseudo coloring, and so on.

1.1.2 Image Restoration:

It is concerned with filtering the observed image to minimize the effect of degradations. Effectiveness of image restoration depends on the extent and accuracy of the knowledge of degradation process as well as on filter design. Image restoration differs from image enhancement in that the latter is concerned with more extraction or accentuation of image features.

1.1.3 Image Compression:

It is concerned with minimizing the no of bits required to represent an image. Application of compression are in broadcast TV, remote sensing via satellite, military communication via aircraft, radar, teleconferencing, facsimile transmission, for educational & business documents , medical images that arise in computer tomography, magnetic resonance imaging and digital radiology, motion , pictures ,satellite images, weather maps, geological surveys and so on.

Work on international standards for image compression started in the late 1970s with the CCITT (currently ITU-T) need to standardize binary image compression algorithms for Group 3 facsimile communications. Since then, many other committees and standards have been formed to produce de jure standards (such as JPEG), while several commercially successful initiatives have effectively become de facto standards (such as GIF). Image compression standards bring about many benefits, such as:

- Easier exchange of image files between different devices and applications;

- reuse of existing hardware and software for a wider array of products;
- Existence of benchmarks and reference data sets for new and alternative developments.

2. Types of digital Images

The toolbox supports four types of images:
- Gray-scale images
- Binary images
- Indexed images
- RGB images

2.1 Gray-scale Images

A gray-scale image is a data matrix whose values represent shades of gray. When the elements of a gray-scale image are of class uint8 or uint16, they have integer values in the range [0, 255] or [0, 65535], respectively. If the image is of class double or single, the values are floating-point numbers Values of double and single gray-scale images normally are scaled in the range [0, 1], although other ranges can be used.

2.2 Binary Images

Binary images use only a single bit to represent each pixel. Since a bit can only exist in two states—on or off, every pixel in a binary image must be one of two colors, usually black or white. This inability to represent intermediate shades of gray is what limits their usefulness in dealing with photographic images.

2.3 Indexed Images:

Some color images are created using a limited palette of colors, typically 256 different colors. These images are referred to as indexed color images because the data for each pixel consists of a palette index indicating which of the colors in the palette applies to that pixel. There are several problems with using indexed color to represent photographic images. First, if the image contains more different colors than are in the palette, techniques such as dithering must be applied to represent the missing colors and this degrades the image. Second, combining two indexed color images that use different palettes or even retouching part of a single indexed color image creates problems because of the limited number of available colors.

2.4 RGB Images:

A color image is made up of pixels each of which holds three numbers corresponding to the red, green, and blue levels of the image at a particular location. Red, green, and blue (sometimes referred to as RGB) are the primary colors for mixing light—these so-called additive primary colors are different from the subtractive primary colors used for mixing paints (cyan, magenta, and yellow). Any color can be created by mixing the correct amounts of red, green, and blue light. Assuming 256 levels for each primary, each color pixel can be stored in three bytes (24 bits) of memory. This corresponds to roughly 16.7 million different possible colors.

3. Need for image compression:

The needs for image compression becomes apparent when number of bits per image are computed resulting from typical sampling rates and quantization methods. For example, the amount of storage required for given images is

a low resolution, TV quality, color video image which has 512 x 512 pixels/color,8 bits/pixel, and 3 colors approximately consists of 6 x 10^6 bits;

a 24 x 36 mm negative photograph scanned at 12 x 10^{-6}mm:3000 x 2000 pixels/color, 8 bits/pixel, and 3 colors nearly contains 144 x 10^6 bits;

a 14 x 17 inch radiograph scanned at 70 x 10^{-6}mm: 5000 x 6000 pixels, 12 bits/pixel nearly contains 360 x 10^6 bits.

4. Types of compression:

Compression can be categorized in two broad ways:

4.1 Lossless Compression:

Data is compressed and can be reconstituted (uncompressed) without loss of detail or information. These are referred to as bit-preserving or reversible compression systems also lossless compression frequently involves some form of entropy encoding and are based in information theoretic techniques. Methods for lossless image compression are:

a) RUN-LENGTH ENCODING :

Run-length encoding is a very simple form of data compression in which runs of data (that 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.
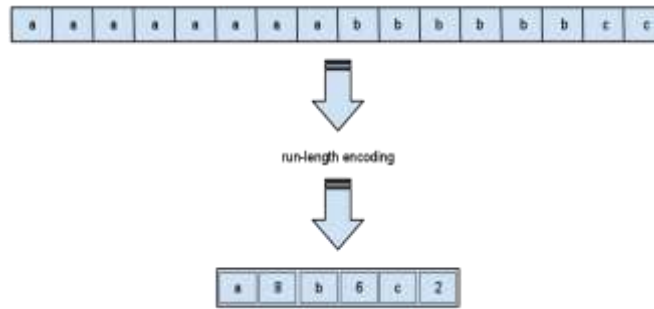
Fig 4.1 Run length encoding

b)   *EntropyEncoding:*

In information theory an entropy encoding is a lossless data compression scheme that is independent of the specific characteristics of the medium. One of the main types of entropy coding creates and assigns a unique prefix-free code to each unique symbol that occurs in the input. of each codeword is approximately proportional .

c)   *Huffman Coding*:

When coding the symbols of an information source individually, Huffman coding yields the smallest possible number of code symbols per source symbol. Huffman coding is an entropy encoding algorithm used for lossless data compression.The term refers to the use of a variable-length code table for encoding a source symbol (such as a character in a file) 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.

d)   *Arithmetic Coding:*

Arithmetic coding generates non block codes. In this coding a one-to-one correspondence between source symbols and code words does not exist. Instead, an entire sequence of source symbols is assigned a single arithmetic code word. The code word itself defines an interval of real numbers between 0 and 1.

4.2 Lossy Compression:

The aim is to obtain the best possible fidelity for a given bit-rate or minimizing the bit-rate to achieve a given fidelity measure. Video and audio compression techniques are most suited to this form of compression. If an image is compressed it clearly needs to uncompressed (decoded) before it can viewed/listened to. processing of data may be possible in encoded form however. Lossy compression use source encoding techniques that may involve transform encoding, differential encoding or vector quantization. Figure 4.2 shows the lossless compression [8].
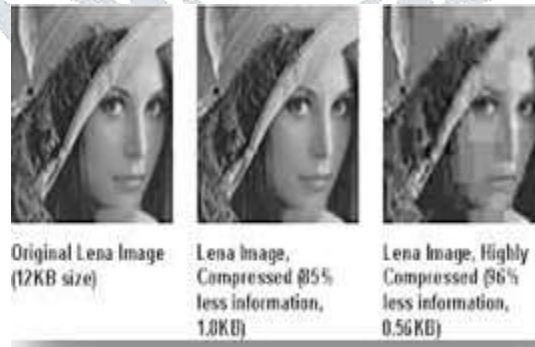
Fig 4.2: Examples of Lossy Compression

4.2.1 Methods for compression:

Reducing the color space to the most common colors in the image. The selected colors are specified in the color palette in the header of the compressed image. Each pixel just references the index of a color in the color palette, this method can be combined with dithering to avoid posterization Chroma subsampling. This takes advantage of the fact that the human eye perceives spatial changes of brightness more sharply than those of color, by averaging or dropping some of the chrominance information in the image.

a)   *Transform coding.*

This is the most commonly used method. In particular, a Fourier-related transform such as the Discrete Cosine Transform (DCT) is widely used.The DCT is sometimes referred to as "DCT-II" in the context of a family of discrete cosine transforms; e.g., see discrete cosine transform

.The more recently developed waveletextensively, followed by quantization and entropy coding. transform is also used

b) *Lossless versus Lossy compression:*

In lossless compression schemes, the reconstructed image, after compression, is numerically identical to the original image. However lossless compression can only achieve a modest amount of compression. Lossless compression is preferred for archival purposes and often medical imaging, technical drawings, clip art or comics. This is because lossy compression methods, especially when used at low bit rates, introduce compression artifacts. An image reconstructed following lossy compression contains degradation relative to the original. Often this is because the compression scheme completely discards redundant information. However, lossy schemes are capable of achieving much higher compression. Lossy methods are especially suitable for natural images such as photos in applications where minor (sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate. The lossy compression that produces imperceptible differences can be called visually lossless.

4.3 Advantage of lossy method.

The advantage of lossy methods over losslessmethods is that in some cases a lossy method can produce a much smaller compressed file than any known lossless method, while still meeting the requirements of the application. Lossy methods are most often used for compressing sound, images or videos. The compression ratio (that is, the size of the compressed file compared to that of the uncompressed file) of lossy video codecs are nearly always far superior to those of the audio and still-image equivalents. Audio can often be compressed at 10:1 with imperceptible loss of quality, video can be compressed immensely (e.g. 300:1) with little visible quality loss. Losssly compressed still images are often compressed to 1/10th their original size, as with audio, but the quality loss is more noticeable, especially on closer inspection.

5. Model For Lossless Compression Using Discrete Cosine Transform (DCT):
Figure below shows the model for Lossless image compression and Decompressin. The building block for the Compression process are
- a) Disctrete Cosine Transform
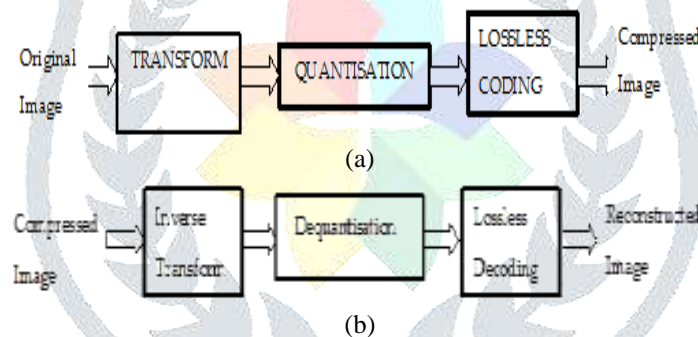- b) Quantizer
- c) Lossless Coding Method (Encoder)

(a)

(b)

Fig 5.1:: Image compression(a) /decompression (b) model.

5.1 Disctrete Cosine Transform
The general equation for a 2D (N by M image) DCT is defined by the following equation:

$$f(u,v) = \alpha(u)\alpha(v)\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} f(i,j)\left(\cos[\frac{(2x+1)u\pi}{2N}]\cos[\frac{(2y+1)v\pi}{2M}]\right)$$

Where, $\alpha(u)=\alpha(v)$
u,v= 0,1,2,3,4,………,N-1
The inverse 2D-DCT transformation is given by the following equation:

$$f(u,v) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} D(i)D(j)D(i,j)\cos\left[\frac{(2x+1)i\pi}{2N}\right]\cos[\frac{(2y+1)j\pi}{2N}]$$

Where D (i) = (1/N) ^1/2 for i=0
D (j) = (2/N) ^1/2 for j=1, 2, 3……., (N-1)

The DCT is similar to the discrete Fourier transform: it transforms a signal or image from the spatial domain to the requency domain. DCTs are important to numerous applications in science and engineering, from lossy compression of audio (e.g. MP3) and images (e.g. JPEG) (where small high-frequency components can be discarded), to spectral methods for the numerical solution of partial differential equations.

The use of cosine rather than sine functions is critical in these applications: for compression, it turns out that cosine functions are much more efficient (as described below, fewer functions are needed to approximate a typical signal), whereas for differential equations the cosines express a particular choice of boundary conditions. In particular, a DCT is a Fourier-related transform similar to the discrete Fourier transform (DFT), but using only real numbers. DCTs are equivalent to DFTs of roughly twice the length, operating on real data with even symmetry (since the Fourier transform of a real and even function is real and even), where in some variants the input and/or output data are shifted by half a sample.

There are eight standard DCT variants, of which four are common. DCT-based image compression relies on two techniques to reduce the data required to represent the image. The first is quantization of the image's DCT coefficients; the second is entropy coding of the quantized coefficients. Quantization is the process of reducing the number of possible values of a quantity, thereby reducing the number of bits needed to represent it. Entropy coding is a technique for representing the quantized data as compactly as possible.

5.2 Quantizer:

The first is quantization of the image's DCT coefficients; the second is entropy coding of the quantized coefficients. Quantization is the process of reducing the number of possible values of a quantity, thereby reducing the number of bits needed to represent it. Quantization reduces the accuracy of the transformer's output in accordance with some pre-established fidelity criterion. Reduces the psycho visual redundancies of the input image.

Quantization at the encoder side, means partitioning of the input data range into a smaller set of values. There are two main types of quantizers: scalar quantizers and vector quantizers. A scalar quantizer partitions the domain of input values into a smaller number of intervals. If the output intervals are equally spaced, which is the simplest way to do it, the process is called uniform scalar quantization; otherwise, for reasons usually related to minimization of total distortion, it is called non uniform scalar quantization.

Fig.5.2 Quantized/dequantized cameraman image using the quantization matrix.

Table. 5.1 Ordering coefficients using zigzag scanning

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

Table.5.2 Quantized 8 × 8 DCT coefficients.

55	-43	7	6	-3	4	0	-4
0	1	1	0	0	1	0	0
0	0	0	1	1	1	1	1
0	0	0	1	1	1	1	1
0	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
0	1	1	1	1	0	0	1
0	0	0	0	1	1	1	0

5.3Zigzag Scanning: Due to efficient energy compaction property of the DCT, many coefficients, especially the higher frequency coefficients become zero after quantization. Therefore, zigzag scanning the N × N
DCT array is used to maximize the zero run-lengths. The zigzag scanning pattern used in
JPEG is shown in figure.

Fig 5.3 A zigzag scanning pattern

Encoding (Lossless)
Symbol (entropy) encoder:
JPEG Coding of DC Coefficients The DC coefficient in each 8 × 8 block of DCT is a measure of the average value of that block of pixels—actually it is 8 times the average. Since there is considerable correlation in the DC values between adjacent blocks, we can gain additional compression if we code the DC differential values than if we code the DC values individually. This amounts to simple integer prediction of the current block DC (j) from the previous block DC (j − 1) as describe by

$$e_{dc}(j) = DC(j) - DC(j - 1)$$

In JPEG baseline, the DC differential is divided into 16 size categories with each size category containing exactly 2size number entries. From the definition of 2D DCT, we see that the maximum DC value of an 8 × 8 block of pixels is 8 × (255–128) = 1016 for 8-bit images. Since lossless coding is one mode of JPEG, the maximum value of the DC coefficient difference could be 2032, which is size category 11. For 12-bit images, the DC differential will have a range between −32,767 and +32,767, which is size category 15—the maximum size allowed. Table 7.14 shows the 16 sizes and the corresponding amplitude range for the DC differentials. To code a DC differential, we first determine its size category from

$$size = [log_2(|val| + 1) + 0.5]$$

Where val is the DC differential value. The code for the DC differential Val is then the code for its size followed by the binary code for the amplitude. If the differential value is negative, then the code for the amplitude is the 2s complement of Val-1.

JPEG Coding of AC Coefficients The quantized AC coefficients are coded slightly differently in the JPEG standard than the DC coefficient because of the possibilities of run lengths and nonzero amplitudes. The run-length/amplitude or level pair is coded using variable-length codes (VLC).
However, only codes for a subset of the run-length/amplitude pairs are used and any pair not in the allowed range is coded using an escape code followed by a 6-bit code for the run-length and an 8-bit code for the amplitude. In many quantized DCT blocks, there may be a large run of zeros up to the end of the block. In such cases, an end-of-block (EOB) code of 4 bits for the luma and 2 bits for the chroma is used to signal the end of the block. Having 2 or 4 bits for the EOB is an indication that EOB occurs most often.

6.Conclusion

 We have reviewed and summarized the characteristics of image compression, need of compression, principles behind compression, different classes of compression techniques and various image compression algorithms based on Wavelet, JPEG/ DCT, VQ, and Fractal approaches. Experimental comparisons on 256×256 commonly used image of Lenna and one 400×400 fingerprint image suggest a recipe described as follows. Any of the four approaches is satisfactory when the 0.5 bits per pixel (bpp) is requested. However, for a very low bit rate, for example 0.25 bpp or lower, the embedded zero tree wavelet (EZW) approach is superior to other approaches. For practical applications, we conclude that

 (1) Wavelet based compression algorithms are strongly recommended,

 (2) DCT based approach might use an adaptive quantization table,

 (3) VQ approach is not appropriate for a low bit rate compression although it is simple,

(4) Fractal approach should utilize its resolution-free decoding property for a low bit rate compression.

7. REFERENCES

[1] Rafel C.Gonzalez And Richard E. Woods " Digital Image Processing" , 3rd ed. New Delhi, Prentice-hall Of India,2008,ch. 8,pp.547-644.

[2] Rafel C.Gonzalez ,Steven L. Eddins And Richard E. Woods " Digital Image Processing Using Matlab", 2nd ed. New Delhi, Prentice-hall Of India, 2009.

[3]Alaukika Nayak, Rashmi Rekha Sahoo and Prabin Kumar Bera, "Performance Evaluation of DCT-JPEG and DWT-JPEG" .

[4] P. S. Tejyan and P. Singh, "A novel approach to the 2D-DCT and 2D-DWT based JPEG image compression," (IJAEST) J. vol . 1, Issue No. 2,pp. 079 – 084.

[5] Subramanya A,"IMAGE COMPRESSION TECHNIQUE," Potential IEEE,Vol.20 Issue1,pp.19-23, Feb-March 2001.

[6] A study of various image compression techniques.

[7] Ming Yang & Nikolaos Bourbakis, "An Overview of Lossless Digital Image Compression Techniques," Circuits& Systems, 2005 48th Midwest Symposium , vol. 2 IEEE, pp 1099-1102, 7 –10 Aug, 2005

[8] Sachin Dhawan,"A Review of Image Compression and Comparison of its Algorithms", IJECT Vol. 2, Issue 1, March 2011, 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 (P r i n t), pp 22-27.