Wavelets: Applications to Image Compression-I

Sachin P Nanavati and Prasanta K Panigrahi

Sachin P Nanavati has joined the Scientific and Engineering Computing Group of C-DAC, Pune. He enjoys trekking and bird watching.

Prasanta K Panigrahi is currently with the Quantum Information and Quantum Optics Division at PRL. His current research interests are in the areas of quantum information, solitons and wavelets.

Digital imaging has had an enormous impact on scientific and industrial applications. Uncompressed images require considerable storage capacity and transmission bandwidth. The solution to this problem is to compress an image for desired application. Wavelet transform has recently emerged as the tool of choice for image compression. In this article, we discuss the basic principles underlying compression of images and point out the advantages of wavelet transform over the previously used discrete cosine transform.

Introduction

Digital images have become an important source of information in the modern world of communication systems. In their raw form, these images require an enormous amount of memory. In fact, according to a recent estimate, 90% of the total volume of traffic in the internet is composed of images or image related data [1]. With the advent of multimedia computing, the demand for processing, storing and transmitting images has increased exponentially. Considerable amount of research has been devoted in the last two decades to tackle the problem of image compression. There are two different compression categories: *lossless* and *lossy*. Lossless compression preserves the information in the image. Thus an image could be compressed and decompressed without loosing any information. Applications requiring this type of compression include medical and legal records' imaging, military and satellite photography. In lossy compression information is lost but is tolerated as it gives a high compression ratio. Lossy compression is

Keywords

JPEG-2000, image compression, discrete cosine transform (DCT), discrete wavelet transform (DWT).

useful in areas such as video conferencing, fax or multimedia applications and is the focus of this article.

We begin with an introductory note on data compression. We then give a brief description of the various attributes of images, which make them amenable to compression and how to evaluate the compression performances. Discrete cosine and wavelet transforms, the workhorses of *JPEG-93* and *JPEG-2000*, respectively are then presented and the advantages of wavelets pointed out. In this article, we make use of the two dimensional wavelet transforms over images, which are represented as two dimensional arrays of numbers. The readers unfamiliar with wavelet transform are referred to our previous *Resonance* articles on this subject [2] and [3].

Let us clearly understand the difference between *information* and *data*. Data refers to the *means* with which the given information is conveyed. It may be in the form of symbols, ASCII characters or numbers. Various amount of data can be used to represent the same amount of information.

Data Compression

Data compression refers to the process in which a given information is represented with reduced number of data points. Let us clearly understand the difference between information and data. Data refers to the *means* with which the given information is conveyed. It may be in the form of symbols, ASCII characters or numbers. Various amount of data can be used to represent the same amount of information. For example, consider two individuals telling the same story in different number of words. Then there is a possibility that one individual among them will convey the same story with more number of words as compared to the other. This *non-essential* data introduces redundancy and hence the scope for compression. In this article, we will be dealing with lossy compression of images, which tries to achieve a high compression at the cost of some loss of information.

This *non-essential* data introduces redundancy and hence the scope for compression.

A digital image can be represented by a two dimensional (2-D) array i.e., a matrix, each of whose element $f(i,j)$ corresponds to the value of the (i,j)th pixel in the

While compressing an image, two important objectives are kept in mind. On one hand, the compressed image should not be distorted and on the other, it should require minimum number of bytes to store.

image. If each pixel represents a shade of gray in monochrome images, we need to allocate only one *byte* or 8 *bits* per pixel (bpp). With $2^8 = 256$ combinations, one can represent numbers ranging from 0 to 255. Thus, a gray scale image when displayed, will have shades of gray, ranging from 0 (black) to 255 (white). An uncompressed, say 800 $\times$ 800 pixel image, will need 64×10^4 bytes $\simeq$ 0.64 MB. If 1000 images are to be stored, we need 640 MB!. Obviously, we then need to compress the image, i.e. represent the same image with reduced number of bits, possibly with some loss, without changing the original size of the image.

Mathematically, a measure of compression is given by the *compression ratio* (C_R) defined as,

$$C_R = \frac{\text{number of bits in the original image}}{\text{number of bits in the compressed image}}. \quad (1)$$

While compressing an image, two important objectives are kept in mind. On one hand, the compressed image should not be distorted and on the other, it should require minimum number of bytes to store. Typically, these two objectives are conflicting, thus a suitable criterion is needed to reach a compromise. This criterion depends upon the particular application. As will be seen soon, it is possible to have a visually pleasing image quality, with a highly lossy compression!

It is possible to have a visually pleasing image quality, with a highly lossy compression!

Before proceeding further, let us define the mathematical quantity which measures the quality of the reconstructed image compared with the original image. It is called *peak signal to noise ratio* (*PSNR*) [4], measured in decibel (dB) and is defined as:

$$PSNR = 20\ log_{10}\left(\frac{255}{RMSE}\right) \quad (2)$$

where $RMSE$ is the root mean-squared error defined as

$$RMSE = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[f(i,j) - \hat{f}(i,j)\right]^2} \quad (3)$$

Here M and N are the width and height, respectively (in pixels), of the image array. f is the original image and $\hat{f}$ is the reconstructed image. Note that the original and the reconstructed images must be of the same size.

Images can be compressed, primarily by eliminating the following types of redundancies:

Spatial Redundancy: In most of the natural images, the values of the neighbouring pixels are strongly correlated i.e., the value of any given pixel can be reasonably predicted from the values of its neighbours. Thus the information carried by individual pixels is relatively small. This type of redundancy can be reduced, if the 2-D array representation of the image, is transformed into a format which only keeps differences in the pixel values. From our previous discussions on wavelet [2,3], we are familiar with the ability of the wavelet transforms to effectively capture variations at different scales. Hence, wavelets are ideally suited for this purpose.

Coding Redundancy: For a given image stored as a matrix with values ranging from 0 to 255, one can find out the number of times each digit occurs in the image. This frequency distribution of pixel values from 0 to 255 is called a histogram. In typical images, a few pixel values have greater frequency of occurrence, as compared to the others. Hence, if the same code word size is assigned to each pixel (8 bits in case of grayscale images), then coding redundancy is incurred. This can be removed, if fewer bits are assigned to the more probable gray scale values than the less probable ones. This redundancy can be reduced using *Huffman coding*.

Psychovisual Redundancy: This takes into account the properties of human vision. Human eyes do not respond with equal sensitivity to all visual informations. Certain information has less relative importance as compared to other informations in normal visual processing.

Human eyes do not respond with equal sensitivity to all visual informations. Certain information has less relative importance as compared to other informations in normal visual processing. This information is said to be *psychovisually* redundant.

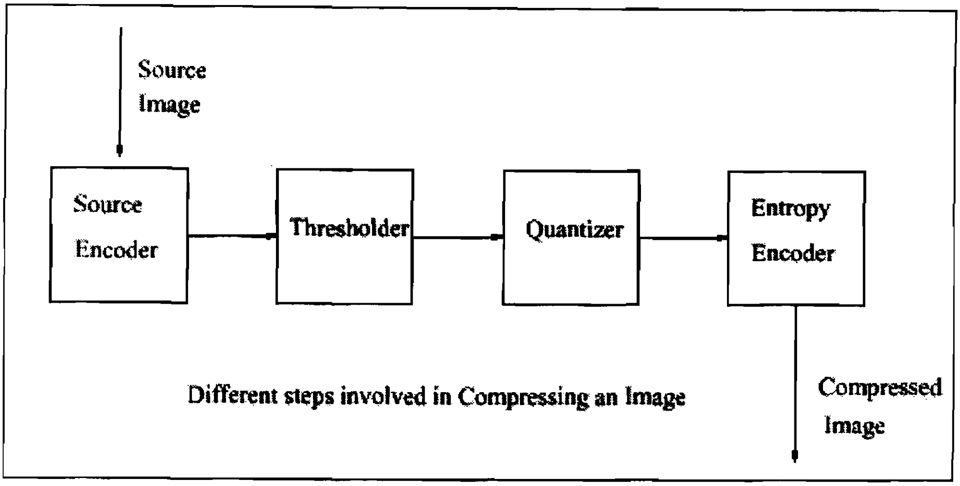


Figure 1. Lossy compression model.

This information is said to be *psychovisually* redundant. Broadly speaking, an observer searches for distinguishing features like edges or textural regions and mentally combines them into recognizable groupings. The brain then relates these groupings, with its prior knowledge, in order to complete the image interpretation process. This redundancy can be overcome by the process of *thresholding* and *quantization* of the wavelet coefficients, to be discussed later.

Lossy Image Encoder

Typical lossy image compression system contains four closely related components as shown in *Figure* 1. They are *(a) source encoder (b) thresholder (c) quantizer, and (d) entropy encoder.* Compression is accomplished by first applying a linear transform to decorrelate the image data, the transformed coefficients are then thresholded and quantized and finally, the quantized values are entropy coded. The decompression is achieved by applying the above four operations in the reverse order.

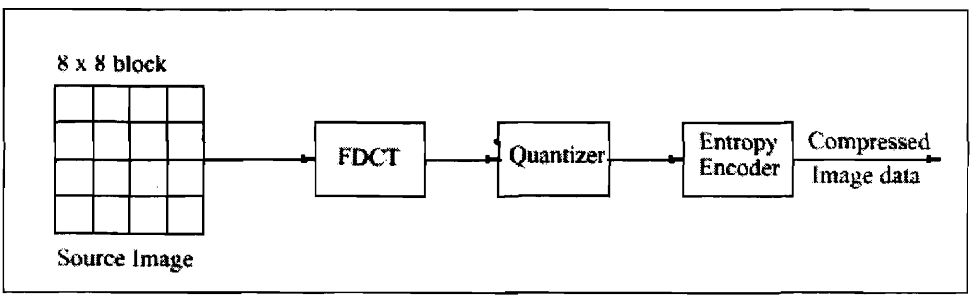
Source Encoder: This refers to the linear transforms, which are used to map the original image into some transformed domain. Popular transform techniques used are discrete Fourier transform (DFT), discrete cosine transform (DCT) or discrete wavelet transform (DWT).

Each transform has its own advantages and disadvantages. In this article we describe DCT and DWT. JPEG-93 was based on DCT, while the recent JPEG-2000 will be totally based on DWT. We will briefly review the pros and cons of each method and the advantages of DWT over DCT.

Discrete Cosine Transform (DCT) Based Coding: An important discovery of mid 1970's, DCT gives an approximate representation of DFT considering only the real part of the series. For a data of N values, DCT's time complexity (broadly speaking, amount of computational time) is of the order of $Nlog_2N$ similar to DFT. But DCT gives better convergence, as compared to DFT.

The block diagram of DCT based coding is shown in *Figure* 2. First, a given image is divided into 8×8 blocks and *forward discrete cosine transform* (FDCT) is carried out over each block. Since the adjacent pixels are highly correlated, the FDCT processing step lays the foundation for achieving data compression. This transformation concentrates most of the signal in the lower spatial frequencies, whose values are zero (or near zero). These coefficients are then quantized and encoded (which we will discuss later) to get a compressed image. The decompression is obtained by applying the above operations in reverse order and replacing FDCT by *inverse discrete cosine transform* (IDCT). Interested readers can refer [5] for more details.

Figure 2. DCT based compression model.



Wavelets are the probing functions, which give optimal *time-frequency* localization of a given signal.

Discrete Wavelet Transform (DWT) Based Coding: From a practical point of view, wavelets [6,7,8] provide a basis set which allows one to represent a data set in the form of differences and averages, called the *high-pass* or *detail* and *low-pass* or *average* coefficients, respectively. The number of data points to be averaged and the weights to be attached to each data point, depends on the wavelet one chooses to use. Usually, one takes $N = 2^n$ (where n is a positive integer), number of data points for analysis. In case of the simplest, Haar wavelet, the level-1 high-pass and low-pass coefficients are the nearest neighbour differences and nearest neighbour averages respectively, of the given set of data with the alternate points removed. Subsequently, the level-1 low pass coefficients can again be written in the form of level-2 high-pass and low-pass coefficients, having one-fourth number of points of the original set. In this way, with 2^n number of points, at the n^{th} level of decomposition, the low-pass will have only one point. For the case of Haar, modulo a normalization factor, the n^{th} level low-pass coefficient is the average of all the data points. In principle, an infinite choice of wavelets exist. The choice of a given wavelet depends upon the problem at hand. As one can easily imagine, wavelets are ideal to find variations at different scales present in a data set. This procedure can be easily extended to two dimensional case, for applications to image processing.

As discussed in earlier articles, wavelets are the probing functions, which give optimal *time-frequency* localization of a given signal. Due to its flexible mathematical modelling, it has certain distinct advantages over DCT. Below, we point out the relative merits of DCT and DWT.

1) ***Gibbs' Phenomenon:*** In the transformed domain, one usually performs a thresholding i.e., discarding of all the coefficients, whose values are below the threshold. In case of DCT, thresholding is done in the frequency

domain, where the time information about the signal is hidden in the relative phases of different frequency modes. Hence, the effect of chopping certain coefficients (local phenomenon), manifests itself throughout the signal (global phenomenon) after reconstruction. This effect is called Gibbs' phenomenon. Its ill-effects become obvious, if the selected threshold introduces large errors, which will eventually corrupt the entire signal.

In case of DWT, we get the so called *time-frequency* localization. Consider the (1,1) element of the low pass sub-matrix (after applying 2-D DWT over the image) [3]. In case of the Haar basis, this element is the result of an average, performed on the first 2 × 2 (parent) elements of the original image. Similarly, the origin of the remaining three high pass elements could be traced to differences. This procedure clarifies the notion of localization *i.e.*, each and every point in the sub-matrices can be attributed to particular sets of points in the original image. If thresholding is to be performed on these coefficients, they will affect only the corresponding parent elements of the original image and error (if any) will be local in nature.

Each and every point in the sub-matrices can be attributed to particular sets of points in the original image. If thresholding is to be performed on these coefficients, they will affect only the corresponding parent elements of the original image and error (if any) will be local in nature.

2) ***Time Complexity:*** Time complexity (broadly speaking, amount of computational time) of DCT is of O ($N\log_2 N$) [5] while many wavelet transforms can be calculated with $O(N)$ operations. More general wavelets require $O(N\log_2 N)$ calculations, same as that of DCT [8].

3) ***Blocking Artifacts:*** In DCT, the given image is sub-divided into 8×8 blocks. Due to this, the correlation between adjacent blocks is lost. This result is noticeable and annoying, particularly at low bit rates. This is shown in *Figure* 3. In DWT, no such blocking is done and the transformation is carried over the entire image.

Figure 3. Original image (left) and DCT reconstructed image (right) showing blocking artifacts [9].

4) ***Advantage of a Designed Wavelet Set:*** It is possible to construct our own basis function, for wavelets, depending upon the application in hand. Thus, if a suitable basis function is designed for a given task, then it will capture most of the energy of the function with very few coefficients. This freedom is curtailed in DCT, where one has only cosine functions as the basis set.

5) ***Compression Performance:*** The DCT based JPEG-93 compressor performs well for a compression ratio of about 25:1. But the quality of image rapidly deteriorates above 30:1, while wavelet based coders degrade gracefully, well beyond ratios of 100:1 [4].

Once DWT is performed, the next task is thresholding, which is neglecting certain wavelet coefficients. The basic idea is to truncate the insignificant coefficients, since the amount of information contained in them is negligible.

6) ***Disadvantages:*** The biggest disadvantage of the wave- let based coding technique is the problem of selecting basis function for a particular operation. This is because, a particular wavelet is suited for a particular purpose. So the property of each wavelet should be known earlier. Most of the time, the selection is done after experimenting with different sets of wavelets for a given application.

Once DWT is performed, the next task is thresholding, which is neglecting certain wavelet coefficients. For doing this one has to decide the *value of a threshold* and *how to apply the same.* This is an important step which affects the quality of the compressed image. The basic idea is to truncate the insignificant coefficients, since the amount of information contained in them is negligible. In the second part of this article, we will describe in detail the procedure for selecting the desired threshold and other aspects of image compression.

We are thankful to Arun K Pujari for bringing to our notice many helpful references and for many useful discussions and Rajneesh Atre for a careful reading of the manuscript.

Suggested Reading

[1] P N Topiwala (Editor), *Wavelet Image and Video Compression*, Kluwer Academic, Norwell, USA, 1998.

[2] S Nanavati and P Panigrahi, Wavelet transforms: A new mathematical microscope, *Resonance*, Vol.9, No.3, pp.50-64, March 2004.

[3] J Modi, S Nanavati, A Phadke and P Panigrahi, Wavelet transforms: Application to Data Analysis, *Resonance*, Vol.9, No.11, pp.10-22, November 2004 and Vol.9, No.12, pp.8-13, December 2004.

[4] M L Hilton, B D Jawerth and A Sengupta, Compressing Still and Moving Images with Wavelets, *Multimedia Systems*, Vol.2, No.3, April 1994.

[5] A K Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall of India, New Delhi, 1995.

[6] B B Hubbard, *The World According to Wavelets*, 2nd edition, Universities Press (India), Hyderabad, 2003.

[7] R M Rao and A S Bopardikar, *Wavelet Transforms: Introduction To Theory and Applications*, Pearson Education Inc., Delhi, India, 2000.

[8] C S Burrus, R A Gopinath and H Guo, *Introduction to Wavelets and Wavelet Transforms – A Primer*, Prentice-Hall, New Jersey, USA, 1998.

[9] S Saha, Image compression – from DCT to Wavelets: A Review, *ACM Crossroads Students Magazine* available at http://www.acm.org/crossroads/xrds6-3/sahaimgcoding.html

Address for Correspondence
Sachin P Nanavati
National PARAM Supercomputing Facility, Centre for Development of Advanced Computing (C-DAC), Pune University Campus, Ganesh Khind, Pune 411 007, India.
Email: sachinn@cdacindia.com

Prasanta K Panigrahi
Physical Research Laboratory
Navrangpura
Ahmedabad 380 009, India.
Email: prasanta@prl.ernet.in
URL:www.prl.ernet.in/~prasanta