K. DİMİLİLER

ANALYSIS AND DESIGN OF AN IMAGE COMPRESSION SYSTEM BASED ON OBJECTIVE AND SUBJECTIVE QUALITY MEASUREMENTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY

by

KAMİL DİMİLİLER

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL
AND ELECTRONIC ENGINEERING

NICOSIA 2014

NEU, 2014

ANALYSIS AND DESIGN OF AN IMAGE COMPRESSION SYSTEM BASED ON OBJECTIVE AND SUBJECTIVE QUALITY MEASUREMENTS

**A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY**

by

KAMİL DİMİLİLER

**In Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in
Electrical and Electronic Engineering**

NICOSIA 2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Kamil Dimililer

Signature :

Date: 10.03.2014

ACKNOWLEDGEMENTS

I would like to thank everyone who provided help and advice during the preparation of this dissertation.

First of all, I would like to thank to my supervisor Prof. Dr. Adil Amirjanov for his invaluable advice and belief in my work and myself over the course of the PhD. Research.

Secondly, I would like to express my deepest gratitude to the dean of Engineering Faculty Prof. Dr. Adnan Khashman and chairman of Electrical Engineering Department Assist. Prof. Dr. Ali Serener who provided this opportunity for me. Their assistance, guidance, contribution and unconditional support was very precious for me.

Thirdly, I would also like to extend my appreciation to Near East University and Thesis Supervision Committee Members for their advice.

Finally, I would like to thank my wife, my daughter and my family for their constant encouragement, support and patience during the preparation of this dissertation.

Hamile olan eşime ve tatlı kızıma ...

To my pregnant wife and my sweet daughter ...

ABSTRACT

With the development of communication technology the applications and services of health telemetics are growing. In view of the increasingly important role played by digital medical imaging in modern health care, it is necessary for large amount of image data to be economically stored and/or transmitted. A need for the development of image compression systems arises to combine high Relative Data Redundancy with preserving the critical information.

Image characteristics affect the amount of compression to be applied to an image. One of the most important characteristic that affects the amount of compression on medical images is the contrast between the pixels of an image.

There are various kinds of image compression methods that are applied in order to increase the Relative Data Redundancy by using image transforms such as discrete cosine transform and wavelet transform.

In this thesis, a new algorithm is proposed with the two criteria. Objective Criteria and Subjective Criteria in order to determine the Relative Data Redundancy and compression method that should be applied to a medical image is automated.

Linear Regression Analysis has been used as a statistical approach that includes the characteristics of images, quality of compression, compression method and Relative Data Redundancy as objective criteria.

A new image characteristic has been established that combines the entropy of an image with the contrast that is more efficient in determining the Relative Data Redundancy.

Back Propagation Neural Networks is used as a decision tool and it is based on the subjective criteria as empirical analysis of human in order to determine the compression method to be applied to a medical image.

Analysis of the two methods of assessment showed that Linear Regression Analysis is accurate for the assessment of Relative Data Redundancy and Back Propagation Neural Networks is accurate for the assessment of Compression Method. Combination of two algorithms gives comparable results with the state of the art methods in the literature.

Keywords: Image Compression, Neural Networks, Linear Regression Analysis, Relative Data Redundancy, Compression Ratio, Compression Method

ÖZET

Haberleşme Teknolojisinin gelişmesiyle telematik alanındaki uygulamalar ve servisler büyümeye başlamıştır. Yüksek miktarda sağlık alanında bulunan bilginin en ekonomik şekilde saklanıp veri yoluyla en verimli şekilde gönderilmesi büyük önem kazanmaya başladı. Bir resimdeki kritik bilginin önemini de düşünerek yeterli oranda sıkıştırma sistemlerine ve bu yöntemlerin geliştirilmesine ihtiyaç duyulmaktadır.

Bir resmin karakteristik özellikleri, o resmin hangi oranda sıkıştırılması gerektiğini etkiler. Bir resmin ne kadar sıkıştırılacağı konusunda, resmin karakteristik özellikler arasında en önemli karakteristiklerden biri de pikseller arasındaki kontrast değerleridir.

Sıkıştırma oranını bulmada ayrık kosinüs dönüşümü ve dalgacık dönüşümü gibi birkaç çeşit resim sıkıştırma yöntemi bulunmaktadır. Bu dönüşümler bir resme blok blok veya tüm resme direk olarak uygulanıp sıkıştırma oranı bulunabilmektedir.

Bu tezde, iki kriter göz önünde bulundurulup yeni bir algoritma önerilmiştir. Objektif kriter ve subjektif kriter kullanılıp medikal resimlere uygulanmış olup sıkıştırma yöntemi ve sıkıştırma oranı bulunmuştur.

Lineer regresyon analizi kullanarak istatistiksel yöntemle resimlerin karakteristik özellikleri, sıkıştırma kalitesi, sıkıştırma yöntemi ve sıkıştırma oranı, objektif kriter olarak kullanılıp değerlendirilmiştir.

Yeni bir karakteristik özellik olarak resmin entropisi ile kontrast değeri birleştirilip resim sıkıştırma oranını tesbit etmede daha etkili bir karakteristik özellik olduğu görülmüştür.

Geri yayılım sinir ağları, medikal resim sıkıştırma yöntemine karar vermede kullanılmış olup, subjektif kriter olarak insan tecrübesine dayanan analizde kullanılmıştır.

İki yöntemin birleşimi analiz edilerek lineer regresyon analizinden elde edilen sonuçlarla bir resmin hangi oranda sıkıştırılması gerektiğine ve geri yayılım sinir ağlarından elde edilen sonuçlarla ise bir resmin hangi yöntemle sıkıştırılması gerektiğine karar verilmiştir. İki algoritmanın birleşiminden iyi sonuçlar elde edilmiştir.

Anahtar Kelimeler: İmge Sıkıştırma, Yapay Sinir Ağları, Lineer Regresyon Analizi, Relatif Artık Bilgi, Sıkıştırma Oranı, Sıkıştırma Yöntemi

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ABBREVIATIONS

ANN : Artificial Neural Network

BP : Back Propagation

BPNN : Back Propagation Neural Network

BWT : Biorthogonal Wavelet Transform

CM : Compression Method

CR : Compression Ratio

DCT : Discrete Cosine Transform

DFT : Discrete Fourier Transform

DWT : Daubechies Wavelet Transform

FFT : Fast Fourier Transform

FT : Fourier Transform

GUI : Graphical User Interface

HWT: Haar Wavelet Transform

LRA : Linear Regression Analysis

MOS : Mean Opinion Score

MRI : Magnetic Resonance Image

MRNN: Multi Resolution Neural Network

MSE : Mean Square Error

NN : Neural Network

PR : Pattern Recognition

PSNR : Peak Signal-to-Noise Ratio

CHAPTER 1

INTRODUCTION

Two dimensional signal processing includes image processing where the input is an image, such as a photograph or video frame; the output of image processing may be either an image or, a set of parameters or characteristics related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Digital image processing is the application of different computer algorithms to perform processing of digital images.

The graphics are employed in modern computers extensively. Computer's file directory is displayed graphically in Window-based operating systems. The progress of many system operations, such as uploading or downloading a file, may also be represented graphically. Many applications provide a graphical user interface (GUI), which makes it easier to use the program and to interpret displayed results. Computer graphics is used in many areas in everyday life to convert many types of complex information to images. Thus, images are important, but they tend to be large. Modern hardware can display many colors, which is why it is common to have a pixel represented as a 24-bit number, where the red, green, and blue components occupy 8 bits each. Such a 24-bit pixel can specify one of 16.78 million colors. As a result, an image at a resolution of 512×512 that consists of such pixels occupies 786,432 bytes. At a resolution of 1024× 1024 it becomes four times as big, requiring 3,145,728 bytes. Videos are also generally used in computers, causing even larger sizes of images. So image compression is an important factor.

An important feature of image compression is that the compression can be lossy. An image, exists for people to look at, so, when it is compressed, it is acceptable to lose image information to which the human eye is not very sensitive. This is one of the main ideas behind the many lossy image compression methods that have been developed in recent decades.

The information can be compressed if it contains redundancy. Data compression amounts to removing or reducing redundancies that exist within the image data. Using lossy compression techniques, however we have a new concept, namely compressing by removing irrelevancy data. An image can be lossy and compressed by removing irrelevant data, even if the original image does not have any redundant information.

There are many methods applied in image compression which automatically establish Relative Data Redundancy such as .gif, .jpg and jpeg2000. Maximum Relative Data Redundancy can be achieved by JPEG.

The JPEG compression algorithm was originally created in order to serve as a standard in the image compression. The baseline JPEG image compression algorithm is the most basic type of sequential DCT based image compression. By using transform coding, quantization, and entropy coding with an 8 bit pixel resolution, a high-level data compression is possible to be achieved. However, the Relative Data Redundancy achieved is due to sacrifices made in quality. The baseline specification assumes that 8-bit pixels are the source image, but extensions can use higher pixel resolutions. JPEG assumes that each block of data input is 8x8 pixels, which are serially input in raster order. Similarly, each block is sequentially input in raster order. Baseline JPEG image compression has some configurable portions, such as quantization tables and Huffman tables, which can be specified individually in the JPEG file header. By studying the source images to be compressed, Huffman codes and quantization codes can be optimized to reach a higher level of compression without losing more quality than is acceptable. Although this mode of JPEG is not highly configurable, it still allows a considerable amount of compression. Furthermore, compression can be achieved by sub-sampling chrominance portions of the input image, which is a useful technique playing on the human visual system.

JPEG 2000 is a wavelet-based image compression standard and coding system. It was created by the Joint Photographic Experts Group committee in 2000 with the intention of superseding their original discrete cosine transform-based JPEG standard (created in 1992) with a newly designed, wavelet-based method. While there is a modest increase in compression performance of JPEG 2000 compared to JPEG, the main advantage offered by JPEG 2000 is the significant flexibility of the code stream. The code stream obtained after compression of an image with JPEG 2000 is scalable in nature, meaning that it can be decoded in a number of ways; for instance, by truncating the code stream at any point, one may obtain a representation of the image at a lower resolution, or signal-to-noise ratio.

An image can be compressed lossless or lossy depending on the field of application. Lossless methods are the methods that if applied to an image, the original image can be retrieved after the reconstruction process. Lossy image compression methods are the methods that when the compression process has been applied, some of the data in the image is lost and the reconstructed image will have less details than the original image. The details are important in some applications, so lossy compression with low Relative Data Redundancy can be applied in order to keep the details.

The suggested system will be applied on the medical images. For medical images, the contrast of the pixels within the image is very important in order to represent a crack on a bone.

Teleradiology, which is the term used for using technology to send radiographic images or x-rays across distances from one location to another, has become lately the most preferred and used clinical aspects in the field of telemedicine. Telemedicine refers to the use of communication and information technologies for the delivery of clinical care, such as the transfer of radiological images from a site of image acquisition to a remote location for interpretation in hospitals such as (CT) scans which is the computerized tomography, (MRI) that stands for magnetic resonance imaging, (US) scans which is the ultrasonography and x-ray scans. These radiological images must be compressed before transmission or due to the bandwidth or due to storage limitations (Singh et al., 2007).

A rapid development has come out in data compression methods to compress huge data files such as images where the compression of data in various applications has become vital (Nadenau et al., 2003). Efficient methods of compression, to compress and store or transfer image data files while retaining high image quality and marginal reduction in size are needed due to the improvements of technology (Ratakonda & Ahuja, 2002) .

To evaluate the quality of compression and recovery, there are two different approaches for assessment which are objective and subjective criteria (Dimililer, 2013).

Objective criterion uses PSNR and is determined by using the mean square error of the images. PSNR depends on the Relative Data Redundancy, compression method and the characteristics of images such as contrast, brightness, variance of intensity in the decision of the Relative Data Redundancy with optimal compression method which is based on the statistical analysis of the characteristics of the images.

Subjective criteria uses the human expert in deciding the optimum Relative Data Redundancy with optimal compression method. The experts are the doctors in their field to check the original images with the compressed set of images using three compression methods and find the optimum Relative Data Redundancy with optimal compression method that is based on the empirical analysis.

For medical images, the contrast between the pixels are very important. When the Medical images are considered, the color of the bones are white. If there is a crack or if a bone is broken, the white bone is affected by black color in grayscale. So the contrast between the bone and the crack can be easily chosen by an expert.

To express the difference of medical images for the Relative Data Redundancy, some new characteristic of image related to Relative Data Redundancy need to be established. This characteristic will be discussed in this thesis which combines the contrast of the image with the entropy of the image. The usefulness of this new established characteristic of the image is shown in chapter 2.

Subjective criteria is completely expert based system which decides the optimum Relative Data Redundancy and optimal compression method based on the idea of the expert. The experts are asked to choose the optimum Relative Data Redundancy and optimal compression method upon presenting the original images and compressed set of images. As the Relative Data Redundancy increases, the blocking artifacts comes on the reconstructed image or blurring effect comes on the reconstructed image. The expert is asked to choose a Relative Data Redundancy that the cracks within the bones are not lost due to lossy compression.

The methods of compression used within the thesis are Discrete Cosine Transform, Daubechies Wavelet Transform and Biorthogonal Wavelet Transform. Discrete Cosine Transform is block by block compression with low Relative Data Redundancy because as the Relative Data Redundancy increases, blocking artifacts appears in the reconstructed images. This type of compression is block by block compression algorithm. The advantage of this algorithm is that the compression is applied block by block so that the details within the blocks are not lost due to low Relative Data Redundancy. However, Wavelet Transform which JPEG2000 is based on image compression is applied to the whole image and higher Relative Data Redundancy can be achieved. The disadvantage of Wavelet Transform based image compression is that due to high Relative Data Redundancy, the details such as the cracks on the bones can be lost after reconstruction.

Linear Regression Analysis has been applied using the characteristics of images to make a relationship between the PSNR value and Optimum Relative Data Redundancy. With the suggested equation and the quality of the image needed, the Optimum Relative Data Redundancy of a given image is found using the characteristics of images.

Back Propagation Neural Network has been applied using the pixel values of the images in the decision of the optimum Relative Data Redundancy and optimal compression method. The system has been trained using a training set and the system of multilayer perceptron with back propagation learning algorithm is capable of choosing the optimum Relative Data Redundancy and optimal compression method based on the training patterns that the empirical analysis based approach has been used to decide the optimum Relative Data Redundancy and optimal compression method.

The suggested system can be applied on medical images. The objective criteria based approach will give a result of optimum Relative Data Redundancy and optimal compression method based on the characteristics of the images and the subjective criteria based approach will give the result of optimum Relative Data Redundancy and optimal compression method using the supervised training algorithm with express assessment of the patterns for training BPNN. When the system will give the outputs of objective criteria and subjective criteria, particular algorithm of selecting Relative Data Redundancy and compression method will be discussed.

1.1 Contributions

- Suggesting a new feature of an image which is Contrast Weighted Entropy that can be applied efficiently in the decision of the optimum Relative Data Redundancy and optimum compression method.

- Establishing a relationship between the quality of the compression and characteristics of images by linear regression analysis (LRA).

- Establishing Relative Data Redundancy and compression method with BPNN which is trained by a set of images selected from the database by subjective criteria as human expert.

- Creating a linearity between the Relative Data Redundancy and compression method of an image using LRA that the characteristics of the images has been considered as objective criteria.

- Designing an algorithm to provide a range of applicability of the Back Propagation Neural Network and Linear Regression Analysis for compression of images.

1.2 Thesis Overview

The remaining chapters of the dissertation are given as follows.

Chapter 2 will give a general information about the criteria to the image compression. A new feature of image which is Contrast Weighted Entropy is introduced. The importance of

Relative Data Redundancy with the quality of compression will also be explained using objective and subjective assessment. The charasteristics of an image will also be given.

Chapter 3 will cover the most common lossy image compression techniques that are applied for medical images in the thesis. Mathematical background of the Discrete Cosine Transform (DCT), Daubechies Wavelet Transform and Biorthogonal Wavelet Transform will be explained.

Chapter 4 will cover the express assessment of Relative Data Redundancy using linear regression analysis to find the optimum Relative Data Redundancy and optimal compression method by using objective criteria.

Chapter 5 will cover the application of neural network approaches in image compression. Back propagation neural network will be considered to find the optimum Relative Data Redundancy by using subjective criteria.

Comparison and range of applicability of objective and subjective criteria for establishing Relative Data Redundancy and compression method are discussed in chapter 6. The chapter 7 represents concluding remarks and discussions of future work.

CHAPTER 2

IMAGE COMPRESSION

2.1 Overview

Image Compression is one of the applications of data compression on digital images. The objective of image compression is to reduce the redundancy of the image data in order to be able to store or transmit the data in an efficient form. The smaller file size that compression provides can take up much less space in your hard drive, web site or digital camera. It will also allow for more images to be recorded on other media, such as a photo and CD. Compressed images also take less time to load than their originals, making it possible to view or transmit more images in a shorter period of time.

In chapter 2, Image characteristics with the criteria of efficiency of compression will be studied in details. Relative Data Redundancy is the ratio of the total number of bits needed to code the original image to the total number of bits needed to code the compressed image. Quality of compression using objective and subjective assessment and time consumption in image compression will be studied. The characteristics affecting the Relative Data Redundancy will be selected.

2.2 Image Characteristics

The characteristic of an image represents the numerical and statistical values within an image. Contrast, change in contrast, brightness, change in brightness, resolution of image, entropy, contrast weighted entropy and variance intensity are included in the characteristics, (Pardo, 2003).

2.2.1 Contrast

Contrast is the difference in visual properties that makes an object (or its representation in an image) distinguishable from other objects and the background.

Root mean square (RMS) contrast is defined as the standard deviation of the pixel intensities (Peli, 1990). Root mean square (RMS) contrast depends on the spatial distribution of contrasts in an image. Equation 2.1 gives the RMS Contrast (Gonzalez & Woods, 2008).

$$Ct = \sqrt{\frac{1}{mn}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\left(I_{ij} - \bar{I}\right)^2} \tag{2.1}$$

where intensities I_{ij} are the *i*-th *j*-th element of the two dimensional image of size *m* by *n*. $\bar{I}$ is the average intensity of all pixel values in the image. The image *I* is assumed to have its pixel intensities normalized in the range [0,1].

2.2.1.1 Change in Contrast

Change in contrast is the difference of the contrast value between the original image and the reconstructed image. When the compression and reconstruction process takes place, the contrast value of the reconstructed image is also affected. When the contrast values of the original and reconstructed images are considered, the more Relative Data Redundancy, the more change of the contrast level of the reconstructed image due to lossy compression. The change in contrast of the original image and reconstructed image can be given in equation 2.2 where Ct_{org} represents the contrast value of the original image, Ct_{rec} represents the contrast value of the reconstructed image and ΔCt represents the change in contrast levels of an image.

$$\Delta Ct = \left|Ct_{org} - Ct_{rec}\right| \tag{2.2}$$

2.2.2 Brightness

Brightness or luminance, the mean of intensity is the average of total grey pixels (Ramírez et al., 2002). The brightness of an image can be estimated using the equation 2.3, where Br represents the brightness of an image, I(i,j) represents the intensity of pixels, m and n are used for the dimensions of the two dimensional image;

$$Br = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} I(i,j)}{m*n} \tag{2.3}$$

2.2.2.1 Change in Brightness

Change in brightness is the brightness difference between the original image and the reconstructed image. Equation 2.4 is used to calculate the change in contrast where, Br_{org} , is the brightness of the original image, Br_{rec} , is the brightness of the reconstructed image and ΔBr is the change in brightness between the original and reconstructed image.

$$\Delta Br = \left| Br_{org} - Br_{rec} \right| \tag{2.4}$$

2.2.3 Resolution

Image resolution describes the detail an image holds. The term applies to digital images, film images, and other types of images. The term *resolution* is often used for a pixel count in digital imaging (Standardization Committee, 2005). Higher resolution means more image details. An image resolution *Rsn* of *N* pixel by *M* pixels can be expressed in equation 2.5.

$$Rsn = m * n \tag{2.5}$$

2.2.4 Entropy

Entropy is a statistical measure of randomness within an image. It is defined in equation 2.6, (Gonzalez & Woods, 2008). Entropy or uncertainty is the average information per source output. p_i represents the probability of intensity within each pixel and L represents the number of intensity values. Larger magnitudes of H indicate that the output source has more entropy or uncertainty. Thus each output provides the end user with more information. Entropy coding is applied in lossy compression after the quantization step which is a lossless form of compression performed on a particular image for more efficient storage, (Song, 2008)

Digital images contain large amount of information that need evolving effective techniques for storing and transmitting the ever increasing volumes of data, (Rouse & Hemami, 2006).

$$H = -\left(\sum_{i=0}^{L} p_i \log_2 p_i \right) \tag{2.6}$$

2.2.5 Contrast Weighted Entropy

Contrast weighted entropy is a new term that combines the change of contrast values between the neighbor pixels with the entropy of the image which includes the information present within an image (Dimililer, 2013). An image with high contrast weighted entropy should be compressed with a low amount of Relative Data Redundancy not to make the information within the pixels lost. However an image with low contrast weighted entropy can be compressed with higher amount of Relative Data Redundancy because there is not enough details that can be lost during the compression process. Equation 2.7 represents the Contrast Weighted Entropy of an image.

$$CWH = -\sum_{L} (p_i - \mu) \times p_i \times \log_2 (p_i) \tag{2.7}$$

where Tot represents the total number of pixels within the image, L represents the number of intensity values of the pixels within the image, p_i represents the probability of the intensity values and μ represents mean value of the intensity values. (Dimililer, 2013)

2.2.6 Variance of Intensity

The Variance is a measure of the spread values of any pixel intensity z about the mean μ and it is a useful measure of image contrast, where z_i, { i=0, 1, 2… L – 1}, denotes the values of all possible intensities in an $m * n$ digital image. The smaller the variance, the more similar the pixels, (Khalil, 2010).

The probability, $p(z_k)$, of intensity level z_k occurring in a given image is estimated using the equation 2.8 as follows;

$$p(z_k) = \frac{n_k}{mn} \tag{2.8}$$

Where n_k is the number of times that intensity z_k occurs in the image and mn is the total number of pixels. The mean (average) intensity is given in equation 2.9;

Figure 2.1: An Original X-ray Image

Figure 2.1 represents an original image named *K15_Org* from the image database.

The characteristics of original image, contrast, brightness, resolution, entropy, visual entropy and variance of intensity of the x-ray image named *K15_Org* from the database is given in Table 2.1

Table 2.1 Characteristics of the Original Image

	RMS Contrast	Brightness	Resolution	Entropy	CWE	Variance of intensity
K15_Org	0.2708	76.040253	65536	5,8572	-43995,949	4805,1

$$\text{«} = \sum_{k=0}^{L-1} z_k \, p(z_k) \tag{2.9}$$

Similarly, the variance of the intensities $\pm^2$ is estimated using the equation 2.10

$$\pm^2 = \sum_{k=0}^{L-1} (z_k - \text{«})^2 \, p(z_k) \tag{2.10}$$

Table 2.2 represents the characteristics of the original image and reconstructed images after DCT compression has been applied.

Figure 2.2 represents the original image and the reconstructed set of a compressed hand images using discrete cosine transform based image compression with the relative data redundancy of 10% up to 90%.

Table 2.2 Characteristics of a DCT Compression based Image Set with Relative Data Redundancy of 10% up to 90%

	RD	RMS Contrast	Change In Contrast	Brightness	Change In Brightness	Resolution	Entropy	CWE	Variance of intensity
K15_Org		0.2708	-	**76.040253**	-	**65536**	5,891	-43995,949	**4805,1**
K15_10_dct	**10%**	0,2709	0,0001	75,99025	0,050003	65536	5,86933	-45465,05	4809,7
K15_20_dct	**20%**	0,2709	0,0001	75,945435	0,094818	65536	5,770828	-52417,25	4810,1
K15_30_dct	**30%**	0,2707	0,0001	75,943695	0,096558	65536	5,774827	-51519,81	4803
K15_40_dct	**40%**	0,2704	0,0004	75,945892	0,094361	65536	5,768996	-51072,21	4792,2
K15_50_dct	**50%**	0,2698	0,0010	75,951508	0,088745	65536	5,767518	-50434,83	4769,3
K15_60_dct	**60%**	0,2683	0,0025	75,959442	0,080811	65536	5,747524	-50038,93	4718,7
K15_70_dct	**70%**	0,2784	0,0076	72,762726	3,277527	65536	5,7997	-47152,8	5081
K15_80_dct	**80%**	0,2878	0,0170	69,717804	6,322449	65536	5,70033	-57631,42	5427
K15_90_dct	**90%**	0,2954	0,0246	66,992096	9,048157	65536	5,421081	-99241,04	5719,5

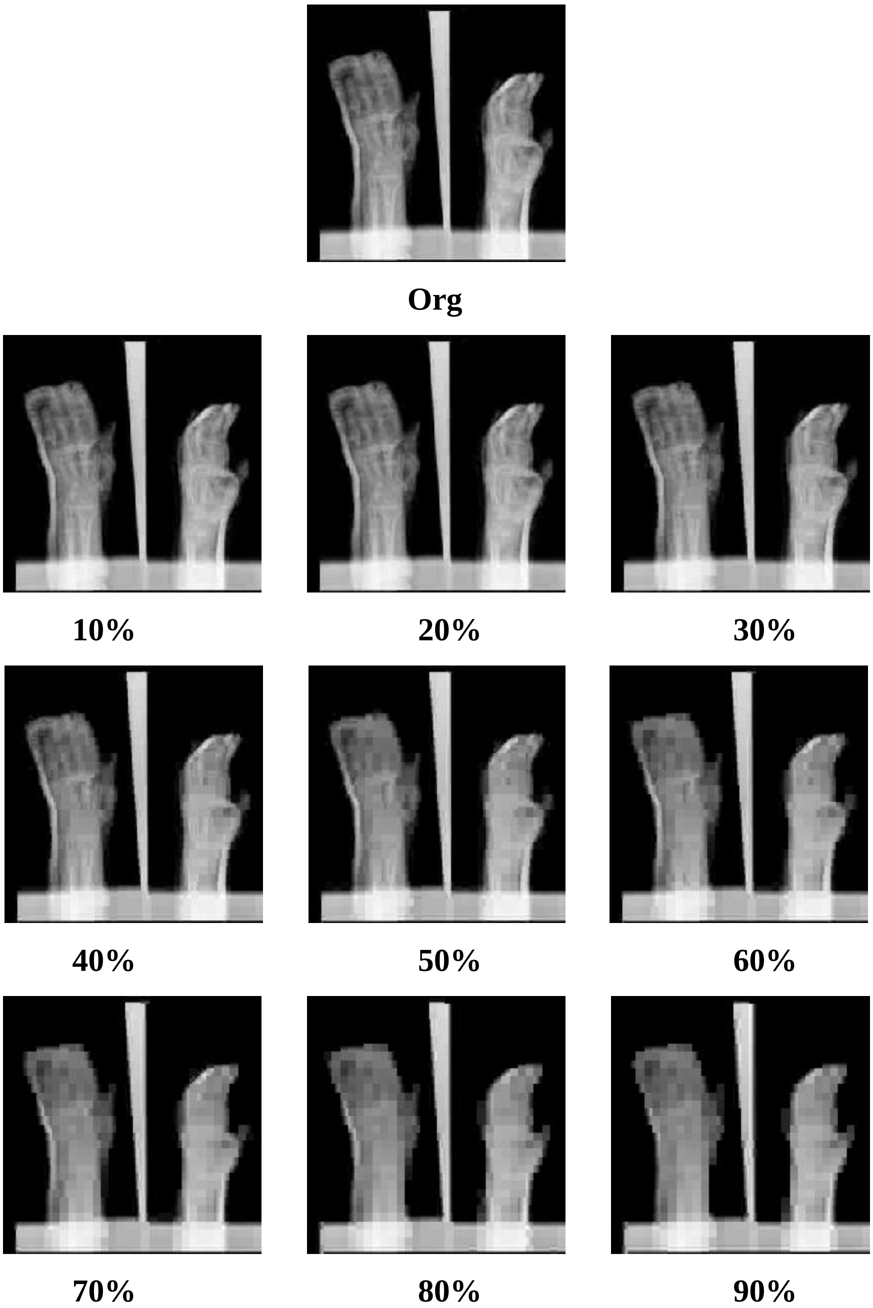


Figure 2.2: Original Image and DCT-based Compression Images with Relative Data Redundancy of 10% up to 90%

2.3 Criteria of Efficiency of Compression

Data Compression represents the process of reducing the amount of data required to represent a given quantity of information. Various amounts of data may be used to represent the same amount of information.

When the future is considered, the need to store image data and transmit images will increase. Even with rapid growth in computer power and the increase in internet bandwidth, the ability to process and transmit the desired amount of image data continues to be problematic.

Image compression involves reducing the size of image data, while retaining necessary information. So the excess amount of data included at an image is removed using the image compression methods. For digital images, data refers to the pixel grey level values that correspond to the brightness of a pixel at a point in space. Information which is an interpretation of data in a meaningful way is an elusive concept. A binary image represents only black and white pixels that can be represented just like a text image which the necessary information may only involve the text being readable are included whereas for a medical image, the necessary information data may involve each detail in the original image.

Image compression can be lossless or lossy. Lossless image compression is the method that is used to compress the image data and when uncompressed, the original image without any data loss can be achieved whereas lossy image compression is the method used to remove the excess data which are considered as details of the image in order to save more space.

Quality, entropy, intensity, variance of intensity, change in brightness and contrast of the original and reconstructed images, variance of intensities with the frequency that represents the details are the important factors affecting the Relative Data Redundancy of an image.

Relative Data Redundancy is one of the important criteria that affect the quality of compression. Relative Data Redundancy is analogous to the physical Relative Data Redundancy used to measure physical compression of substances, and is defined in the same way, as the ratio between the compressed size and the uncompressed size, (Salomon, 2006).

The quality of compression can be measured using the original image and the compressed image. Peak Signal to Noise Ratio and Root Mean Square are the most important measures used to find the quality of the compression, (Sheikh & Bovik, 2005).

The time used to compress and decompress an image is also an important criterion. Processing Time is the total time interval between image acquisition and getting the reconstructed image. Processing time may vary depending on the hardware and software that are used for the implementation, (Khashman & Dimililer, 2005).

2.3.1 Relative Data Redundancy

Digital images have excess data within the image that the human eye does not feel. The importance of Relative Data Redundancy comes out when preserving good perceptual quality. The higher Relative Data Redundancy, the higher loss of details of an image. There are important measures that can be helpful in deciding how much to compress an image. A representation that compresses a 10 Mb file to a 2Mb file has a Relative Data Redundancy of 10/2, often notated as an explicit ratio, 1:5 which can be written as “one to five”. Equation 2.11 represents the relative data redundancy and equation 2.12 represents the Relative Data Redundancy.

$$RD = 1 - \frac{1}{CR} \tag{2.11}$$

$$CR = \frac{UCS}{CS} \tag{2.12}$$

where, RD represents relative data redundancy, CR represents the Compression Ratio, CS represents the compressed size of an image and UCS represents the uncompressed size of an image.

2.3.2 Quality of Compression

There are two kinds of assessments to compare original and reconstructed images after compression that can be used to find the quality of compression of an image. Objective assessment includes the characteristics of an image such as RMS and PSNR values that can be used to find the signal to noise ratio between an original image and a compressed image. Subjective assessment includes human as an expert. In real life, the experts are the doctors in this field who decides whether if there is any problem within the x-ray image.

2.3.2.1 Objective Assessment

The quality of an image after compression can be calculated using the most common methods such as Mean squared error (MSE) and Peak signal to noise ratio (PSNR), (Pardo, 2003).

2.3.2.1.1 Root Mean Square Error

Root mean square error (RMSE) is a frequently-used measure of the differences between pixel values predicted by a model or an estimator and the values actually observed from the ground truth being modelled or estimated, (Pardo, 2003). Equation 2.13 represents the calculation of the Mean Squared Error between an original and a reconstructed image where $I_{i,j}$ represents the intensity of the original image, $K_{i,j}$ represents the intensity of the reconstructed image after the compression process has been applied, *m* and *n* shows the pixel sizes of the image, (Saifuddin, 2010).

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_{i,j} - K_{i,j}\right]^2 \qquad (2.13)$$

2.3.2.1.2 Peak Signal to Noise Ratio

PSNR is the peak signal to noise ratio which gives a measure of the level of information bearing signal power to the level of noise power, (Gonzalez & Woods, 2008). It is the standard method for quantitively comparing a reconstructed image with the original image.

Table 2.3 Relative Data Redundancy and PSNR Values

Relative Data Redundancy (%)	**Peak Signal to Noise Ratio**	**Relative Data Redundancy (%)**	**Peak Signal to Noise Ratio(dB)**
Org.	**-**	**50%**	30,88368
10%	65,716832	**60%**	29,159137
20%	57,399182	**70%**	27,555793
30%	38,327157	**80%**	26,077194
40%	33,609323	**90%**	24,690228

Hence the PSNR of a greyscale image x and its reconstruction y can be calculated as in equation 2.14, (Shi & Sun, 2000). Table 2.3 shows the Relative Data Redundancy and the peak signal-to-noise ratios of an x-ray image randomly selected from the database, when discrete cosine transform based image compression has been applied. Peak Signal-to-Noise Ratio will be considered in order to create the relationship between n original image and a reconstructed image.

$$PSNR = 10\log_{10}\left(\frac{(255)^2}{MSE}\right) \qquad (2.14)$$

Table 2.3 shows the relationship between Relative Data Redundancy and the PSNR values for an image from the database when Discrete Cosine Transform compression has been applied.

2.3.2.2 Subjective Assessment

Human being is considered to decide upon optimum Relative Data Redundancy and optimum compression method in subjective assessment.

When comparing compression codecs it is used as an approximation to human perception of reconstruction quality, therefore in some cases one reconstruction of an image after the compression process has been applied, may appear to be closer to the original than another, even though it has a lower PSNR (a higher PSNR would normally indicate that the reconstruction is of higher quality). One has to be extremely careful with the range of validity of this metric; it is only conclusively valid when it is used to compare results from the same codec (or codec type) and same content, (Huynh-Thu & Ghanbari, 2008).

2.3.2.2.1 Human (Expert) Assessment

The human eye is sensitive to subtle variations in luminance of an image. So the human eye is a very important subjective assessment in decision of optimum Relative Data Redundancy using the visual details of an image, (Eskicioglu, 2010;Nadenau & Reichel, 1999;Westen et al., 1995).

2.3.3 Time Analysis of Relative Data Redundancy

It is important to calculate the time of image compression, because when the medical images are used, the result of compression should be handled in minimum computational cost of time not to affect the patient.

In the process of compression, the total time to compress an original image and get the reconstructed image for any ratio using discrete cosine transform is 10 seconds and using wavelet transform is 4 seconds. These results were obtained by using a 2.8 GHz PC with 8 GB of RAM, Windows 7 Ultimate 64-bit OS and Matlab R2010b software..

When Supervised Back Propagation Neural Network has been considered, the time to train a set of images to the network takes approximately 10 minutes and the time to test an image of a patient takes approximately 0.3 seconds using Matlab Software. However, when the linear regression has been considered, 0.1 seconds is enough to decide the Relative Data Redundancy upon presenting the image to the system using Matlab Software.

2.4 Summary

Image compression is minimizing the size in bytes of a graphics file without degrading the quality of the image to an unacceptable level. The objective of image compression is to reduce the redundancy of the image data in order to be able to store or transmit the data in an efficient form.

In this chapter, criteria of efficiency of compression are studied. Relative Data Redundancy, quality of compression using objective and subjective assessment and time consumption in image compression is studied.

In this thesis, Discrete Cosine Transform based image compression, Daubechies Wavelet Transform based image compression and Biorthogonal Wavelet Transform based image compression will be used in order to compress the images and the results of compression will be applied using Back Propagation Neural Networks and Linear Regression Analysis using the Image brightness, variance of intensity of the image and visual entropy of the image in order to decide the Relative Data Redundancy.

CHAPTER 3

IMAGE COMPRESSION METHODS

3.1 Overview

This chapter gives general information about the theory of Discrete Cosine Transform, Daubechies Wavelet Transform and Biorthogonal Wavelet Transform applied in the field of lossy image compression. The reason to compress the images using DCT is that the compression algorithm has been applied block by block, the reason to compress the images using Daubechies Wavelet Transform is the orthogonal property of the Daubechies and the reason to compress the images using Biorthogonal Wavelet Transform is the property of biorthogonal. The applied algorithms to compress an image using these methods will be given briefly.

3.2 Discrete Cosine Transform (DCT)

Discrete Cosine Transform (DCT) has been widely used in image processing, such as in the field of image compression (Gonzalez & Woods, 2008). Some of the applications of two-dimensional Discrete Cosine Transform involve still image compression and compression of a set of video frames, however multidimensional DCT is generally used for compression of streams of video. DCT is also useful for transferring multidimensional data to frequency domain, where different operations, like spread spectrum, data compression, data watermarking, can be performed in easier and more efficient manner (Martucea, 1994).

The JPEG standard has come towards the late 1980's and has been an effective solution to the standardization of compression of images. Although JPEG has some very useful strategies for DCT quantization and compression, it was only developed for low Relative Data Redundancy and 8×8 DCT block size was used for speed (Gonzales & Woods, 2004).

The following is the general overview of the DCT process.

1. The image is segmented into 8 × 8 blocks of pixels.
2. DCT has been applied starting from left to right and from top to bottom to each block.
3. Each block is compressed through quantization (Hard Threshold coding).
4. The array of compressed blocks that constitute the image is stored in a drastically reduced amount of space.

5. When desired the image is constructed through decompression, which uses the Inverse Discrete Cosine Transform (IDCT).

The DCT is a widely used transformation in images in the field of compression of data. It is an orthogonal transform, which has a fixed set of basis functions that are image independent, an efficient algorithm for computation, good compaction of energy and reduction of correlation properties. Karhunen Loeve Transform (KLT) is a basis function of a first order Markov image closely resembles those of the DCT (Ahmed et al., 1989). They become identical as the correlation between the adjacent pixel approaches to one. The DCT belongs to the family of discrete trigonometric transform, which has 16 members (Gonzales & Woods, 2004).

The 1D DCT of a 1× N vector x (n) is defined in equation 3.1

$$Y[k] = C[k]\sum_{n=0}^{N-1} x[n]\cos\left[\frac{(2n+1)k®}{2N}\right] \tag{3.1}$$

where k = 0,1,2,..., N −1 and N represents the size of the size of the image in one dimensional.

$$C[k] = \begin{bmatrix} \sqrt{\frac{1}{N}},\ k = 0 \\ \sqrt{\frac{1}{N}},\ k = 1,2,3 \ldots, N-1 \end{bmatrix} \tag{3.2}$$

The original signal vector x (n) can be reconstructed back from the DCT coefficients Y[k] using the Inverse DCT (IDCT) operation and can be defined as

$$x[n] = \sum_{k=0}^{N-1} C[k]Y[k]\cos\left[\frac{(2n+1)k®}{2N}\right] \qquad \text{where n = 0,1,2,..., N −1} \tag{3.3}$$

The DCT can be extended to the transformation of 2D signals or images. This can be achieved in two steps: by computing the 1D DCT of each of the individual rows of the two dimensional image and then computing the 1D DCT of each column of the image. If represents a 2D image of size x (n1, n2) N × N, then the 2D DCT of an image is given by:

$$Y[j,k] = C[j]C[k]\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x[m,n]\cos\left(\frac{(2m+1)\,j®}{2N}\right)\cos\left(\frac{(2n+1)k®}{2N}\right) \tag{3.4}$$

that j, k, m, n= 0,1,2,...,N −1 and [m, n] represents the columns or rows of the image matrix.

$$C[j] \quad \text{and} \quad C[k] \;\; C[k] = \begin{bmatrix} \sqrt{\frac{1}{N}},\ j,k = 0 \\ \sqrt{\frac{1}{N}},\ j,k = 1,2,3\ldots,N-1 \end{bmatrix} \tag{3.5}$$

Similarly the 2D IDCT can be mentioned as

$$x[m,n] = \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} C[j]C[k]Y[j,k]\cos\left(\frac{(2m+1)\,j®}{2N}\right)\cos\left(\frac{(2n+1)k®}{2N}\right) \tag{3.6}$$

The DCT is a real valued transform and is closely related to the Discrete Fourier Transform (Dimililer, 2012). In particular, an N × N DCT of x (n1, n2) can be given in terms of DFT of its even-symmetric extension, that leads to a very fast computational algorithm. Because of the even-symmetric extension process, no artificial discontinuities are introduced at the block boundaries. The computation of the DCT requires only real arithmetic. Because of the given properties above the DCT is very popular and its widely used for data compression operation.

The DCT that is presented in equations (3.1) and (3.3) is orthonormal and perfectly reconstructing provided the coefficients are represented to an infinite precision. This means that when the coefficients are compressed it is possible to obtain a full range of compressions and image qualities. The coefficients of the DCT are always quantized for high compression, but DCT is very resistant to quantization errors due to the statistics of the coefficients it produces. The coefficients of a DCT are usually linearly quantized by dividing by a predetermined quantization step.

The DCT is applied to image blocks N x N pixels in size (where N is usually multiple of 2) over the entire image. The size of the blocks used is an important factor since they determine the effectiveness of the transform over the whole image. If the blocks are too small then the images is not effectively de-correlated but if the blocks are too big then local features are no longer exploited. The tilling of any transform across the image leads to artifacts at the block boundaries. The DCT is associated with blocking artifacts since the JPEG standard suffers heavily from this at higher compressions. However the DCT is protected against blocking artifact (Fig 3.1) as effectively as possible, without interconnecting blocks, since the DCT basis functions all have a zero gradient at the edges of their blocks. When edges occur in an image DCT relies on the high freq. parts to make the image more shaper. However these high freq. parts persist across the whole block and although they are good at improving the quality of edges that they tend to 'ring' in the flat areas of the block.

This ringing effect increases, when larger blocks are used, but larger blocks are better in terms of compression, so a trade-off is generally established (Gonzales & Woods, 1992).

(a) **(b)**

Figure 3.1:

DCT Transform of X-ray Images **(a)** X-ray Image, (**b)** Blocking Artifacts

3.3 Wavelet Transform

Transform coding of images is performed by the projection of an image on some basis. The basis is chosen so that the projection will effectively decorrelate the pixel values, and thus, represent the image in a more compact form. The transformed (decomposed) image is then quantized and coded using different methods such as scalar and vector quantization, arithmetic coding, run length coding, Huffman coding, and others.

The best way to introduce the wavelet transform mathematical derivation is to review the familiar Fourier transform. The Continuous Time Fourier Transform ($CTFT = F(w)$) is defined as the inner product or projection of the function f(x) onto the basis function e^{-jwx}. In mathematical terms, the CTFT is (Gonzales & Woods, 1992; Jain, 1989).

$$CTFT : F(w) = \int_{-\infty}^{\infty} f(x) e^{-iwx} dx = \left\langle e^{-iwt}, f(x) \right\rangle \qquad (3.7)$$

where < , > denotes the inner product, x is the location variable, and w is the frequency variable. The inner product in equation 3.7 denotes the Fourier coefficients. Therefore, a function f(x) can be written as

$$f(x) = \int_{-\infty}^{\infty} F(w)e^{iwx}dw \tag{3.8}$$

Equation 3.8 represents the function f(x) is the integration of e^{iwx} multiplied or weighed by the appropriate Fourier coefficient. Although the Fourier transform is a very strong mathematical and analytical tool, it has some properties which are not desirable for image compression. Equation 3.8 shows all past and future time information is needed to accurately reconstruct a signal f(x). If all this time information is not known, the reconstructed signal will only be an approximation of the original. In addition, in order for the Fourier transform not to have an infinite number of coefficients, the signals being transformed must be stationary. Since the Fourier transform can be written as the projection of the signal into cosines in the real plane and sines in the imaginary plane, discontinuities in a signal would only be represented with many sine and cosine terms.

Therefore, the decomposition of a discontinuous signal would contain many coefficients which are spread out over the entire frequency axis. (Barnard, 1994).

In order to mitigate this coefficient frequency spreading, the CTFT was modified to form the Short Time Fourier transform (STFT). The STFT, which is sometimes referred to as the windowed Fourier transform, is similar to the CTFT except for the addition of a window function (w) that limits the existence of the transform. The window function is usually a Gaussian, and it's shifting over the time axis results in a time frequency description of the signal. In general, the STFT can be mathematically given in equation 3.9.

$$STFT : F(w,b) = \int_{-\infty}^{\infty} f(x)e^{-iwx}w(x-b)dx \tag{3.9}$$

where x is the location variable, w is the frequency variable, w() is the window function, and b is the location shifting parameter for the window function (Gonzales & Woods, 1992).

The STFT eliminates the need for complete past and future signal information. However, the STFT still contains a drawback. The window function used in the transform has a fixed window size. Therefore, it cannot adapt to the characteristics of signals at certain points. That is, it does not give a good description of signals with widely changing frequency spectra.

The continuous time wavelet transform CTWT mitigates the limitations of the CTFT and STFT by having a basis function which can be both shifted and dilated or contracted. The CTWT basis function is

$$\Psi_{a,b}(x) = a^{-\frac{1}{2}}\Psi\left(\frac{x-b}{a}\right) \tag{3.10}$$

where Ψ(x) is a zero mean band pass function, and the transform is defined as

$$CTWT(a,b) = a^{-\frac{1}{2}}\int_{-\infty}^{\infty}\Psi\left(\frac{x-b}{a}\right)f(x)dx = \left\langle \Psi_{a,b}(x), f(x)\right\rangle \tag{3.11}$$

where $\Psi_{a,b}$ are the wavelets basis function, a is the scaling factor, and b is the shifting factor (Barnard, 1994).

$\Psi_{a,b}$ are real and oscillatory and fade away when they approach plus or minus infinity (Daubechies, 1994). Ψ itself is called the mother wavelet, and the shifted and scaled conglomeration of Ψ form an orthonormal basis which as mentioned above are called wavelets. A close examination of equation 3.11 shows that increasing 'a' causes T to be stretched, and thus, the formed wavelets can act as low frequency windows. On the other hand, decreasing 'a' cause's Ψ to be shrunk, and thus, these wavelets act as high frequency windows. Consequently, one can narrow and widen the time frequency window as well as shift it over the time domain via 'b' in order to adjust to the frequency characteristics of any signals. In addition, due to the decay property of the wavelets, one does not need all past and future signal information in order to accurately represent any signal (Mallat, 1989; Antonini et al., 1992; Akansu, 1994; Rao et al., 1994).

In general, any signal f(x) can be represented as

$$f(x) = \frac{1}{C_{\Psi}}\int_{-\infty}^{\infty}\int_{0}^{\infty} CTWT(a,b)a^{-\frac{1}{2}}\Psi\left(\frac{x-b}{a}\right)\frac{dadb}{a^2} \tag{3.12}$$

where a, b are continuous and C_Ψ is a constant. Since there are no restrictions on a and b, there exist infinite combinations of possible wavelets. In order to reduce this redundancy, a and b are discretized by letting

$$a = a_0{}^m \tag{3.13}$$

and

$$b = na_0{}^m b_0 \tag{3.14}$$

where n, m, a_0, and b_0 are constants. By substituting these equalities into equation 3.10, the basis functions now become

$$\Psi_{m,n}(x) = a_0{}^{-\frac{m}{2}} \Psi\left(a_0{}^{-m}x - nb_0\right) \tag{3.15}$$

and the new "sampled" CTWT, now called the continuous time wavelet series (CTWS), is defined as (Mallat, 1989), (Antonini et al., 1992). (Akansu, 1994), (Rao et al., 1994), (Gonzales & Woods, 1992).

$$CTWS_{m,n}(x) = d_{m,n} = \int_{-\infty}^{\infty} \Psi_{m,n}(x) f(x)dx = \left\langle \Psi_{m,n}(x), f(x) \right\rangle \tag{3.16}$$

and

$$f(x) = \sum_m \sum_n d_{m,n} \Psi_{m,n}(x) = \sum_m \sum_n \left\langle \Psi_{m,n}(x), f(x) \right\rangle \Psi_{m,n}(x) \tag{3.17}$$

As mentioned above, wavelets are zero mean band pass functions and have zero DC coefficients. Thus, in order to completely represent a function f(x) via a pure wavelet expansion, one needs an infinite number of wavelet resolutions in equation 3.17 (Gonzales & Woods, 1992). This is impracticality in image compression where the goal is to reduce the amount of data needed to represent the total image.

To avoid using an infinite number of coefficients for the wavelet representation, f(x) is represented by a low pass or low resolution version, $f_j(x)$ and a detail version, D. The detail version D is simply a finite wavelet representation. The low pass version, on the other hand, is a result of the inner product between f(x) and a scaling function Φ (t). Thus, if the scaling function weight coefficients are

$$c_{j,n} = \int_{-\infty}^{\infty} f(x) a_0^{-\frac{j}{2}} \Phi\left(a_0^{-j} x - n b_0\right) dx = \left\langle \Phi_{j,n}(x), f(x) \right\rangle \tag{3.18}$$

and the function at resolution j (i.e., $f_j(x)$) is

$$f_j(x) = \sum_{n=-\infty}^{\infty} c_{jn} a_0^{-\frac{j}{2}} \Phi\left(a_0^{-j} x - n b_0\right) = \sum_{-\infty}^{\infty} \left\langle \Phi_{j,n}(x), f(x) \right\rangle \Phi_{j,n}(x) \tag{3.19}$$

then any arbitrary function f(x) can be represented as (Mallat, 1989), (Antonini et al., 1992). (Akansu, 1994), (Rao et al., 1994), (Gonzales & Woods, 1992).

$$f(x) = \sum_{n=-\infty}^{\infty} c_{j,n} a_0^{-\frac{j}{2}} \Phi\left(a_0^{-j} x - n b_0\right) + \sum_{m=-\infty}^{j} \sum_{n=-\infty}^{\infty} d_{m,n} a_0^{-\frac{m}{2}} \Psi\left(a_0^{-m} x - n b_0\right) \tag{3.20}$$

$$= \sum_{n=-\infty}^{\infty} c_{j,n} \Phi_{j,n}(x) + \sum_{m=-\infty}^{j} \sum_{n=-\infty}^{\infty} d_{m,n} \Psi_{m,n}(x) \tag{3.21}$$

$$= f_j(x) + D \tag{3.22}$$

Furthermore, by letting $a_0 = 2$ and $b_0 = 1$, one can pick special Ψ(x) such that for fixed m, $\Psi_{m,n}$ (x) and $\Phi_{m,n}$ (x) individually form an orthonormal basis. In addition, both $\Psi_{m,n}$ (x) and $\Phi_{m,n}$ (x) are orthogonal complements of one another, and < $\Psi_{m,n}$ (x) , f(x) > represents the information lost by going to a lower resolution via < $\Phi_{m,n}$ (x) , f(x) > (Gonzales & Woods, 1992).

Now that the wavelet representation of a function has been established by the above, one should be aware of the containment restrictions placed on both *Φ(x)* and Ψ(x). Since the scaling function projects the function f(x) to a lower resolution plane (i.e., resolution j in the above discussion), it is easy to visualize that all of the information in any lower resolution is contained by the top resolution of the scaling function. Thus, the containment restriction on *Φ(x)* is

$$\Phi(x) = 2\sum_n h_0(n)\Phi(2x-n) \tag{3.23}$$

where $h_0(n)$ are called interscale basis coefficients. Similarly, the containment restriction on Ψ(x) is

$$\Psi\left(x\right) = 2\sum_n h_1(n)\Phi(2x-n) \tag{3.24}$$

where $h_1(n)$ are the expansion coefficients. In general, it can be proved that

$$h_n = 2^{\frac{1}{2}}\int \Phi(x-n)\Phi(2x)dx \tag{3.25}$$

In addition if one defines

$$g_l = (-1)^l h_{-l+1} \tag{3.26}$$

then the scaling coefficients $C_{j,n}$ and wavelet coefficients $d_{m,n}$ can be represented by

$$c_{j,n} = \sum_k h_{2n-k} c_{j-1,k} \tag{3.27}$$

and

$$d_{m,n} = \sum_k g_{2n-k} c_{m-1,k} \tag{3.28}$$

Equations 3.27 and 3.28 are used for the computation of the multiresolution wavelet transform, and h is a low pass filter and g is a band pass filter (Mallat, 1989), (Antonini et al., 1992). (Akansu, 1994), (Rao et al., 1994), (Gonzales & Woods, 1992).

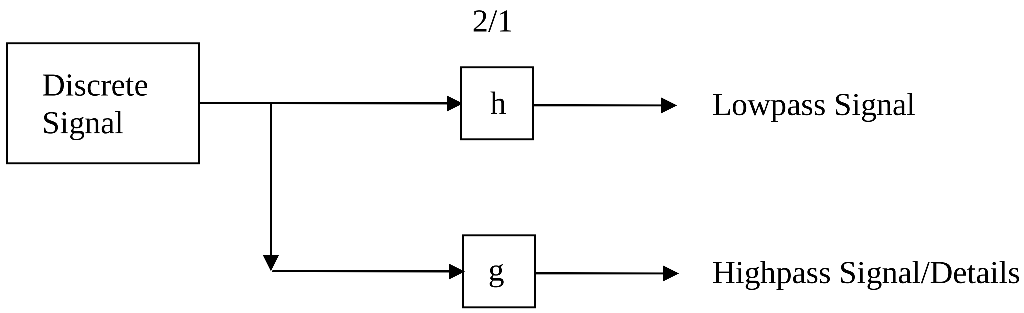


Figure 3.2: Implementation of the 1-D Wavelet Transform

3.3.1 Daubechies Wavelet Transform

Daubechies found a general algorithm that finds different wavelet coefficients based on two properties. The first property is the orthogonality condition placed on the wavelet filters. Simply stated, the orthogonality condition requires that, given the wavelet filter W, then WW = 1. The second property defined by Daubechies is that the filter g should have a zero response to a smooth data vector since g is a band pass filter. To get the zero response, g has to have a certain number of vanishing moments. The advantage of Daubechies Wavelet is the shifting property and orthogonality property and the disadvantage of Daubechies Wavelet is that the function is not symmetric.

Using the above properties, Daubechies defined the h and g filters by keeping only the first n coefficients and arranging these coefficients in a matrix as follows:

$$W = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & & & & & & \\ g_0 & g_1 & g_2 & g_3 & & & & & & \\ & & h_0 & h_1 & h_2 & h_3 & & & & \\ & & g_0 & g_1 & g_2 & g_3 & & & & \\ & & & & & \cdots & & & & \\ & & & & & \cdots & & & & \\ & & & & & & h_0 & h_1 & h_2 & h_3 \\ & & & & & & g_0 & g_1 & g_2 & g_3 \\ h_2 & h_3 & & & & & & & h_0 & h_1 \\ g_2 & g_3 & & & & & & & g_0 & g_1 \end{bmatrix} \tag{3.29}$$

where the blank matrix entries represent zeros. Due to equations 3.27 and 3.28, W becomes

$$W = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & & & & & & \\ h_3 & -h_2 & h_1 & -h_0 & & & & & & \\ & & h_0 & h_1 & h_2 & h_3 & & & & \\ & & h_3 & -h_2 & h_1 & -h_0 & & & & \\ & & & & & \cdots & & & & \\ & & & & & \cdots & & & & \\ & & & & & & h_0 & h_1 & h_2 & h_3 \\ & & & & & & h_3 & -h_2 & h_1 & -h_0 \\ h_2 & h_3 & & & & & & & h_0 & h_1 \\ h_1 & -h_0 & & & & & & & h_3 & -h_2 \end{bmatrix} \tag{3.30}$$

By applying the orthogonality condition, one finds that

$$h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1 \tag{3.31}$$

and

$$h_2 h_0 + h_3 h_1 = 0 \tag{3.32}$$

Since four coefficients were used, one of them must set the moments of the matrix to zero. This condition yields

$$h_3 - h_2 + h_1 - h_0 = 0 \tag{3.33}$$

$$0h_3 - 1h_2 + 2h_1 - 3h_0 = 0 \tag{3.34}$$

Since there are four equations with four unknowns, Daubechies solved them to find her four wavelet coefficients that are referred to as DAUB4. These coefficients are

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}} \tag{3.35}$$

$$h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \tag{3.36}$$

$$h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}} \tag{3.37}$$

and

$$h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}} \tag{3.38}$$

By increasing the number of coefficients and letting n moments go to zero, Daubechies found the values for 6, 8, 10... 2n wavelet coefficients. (Mallat, 1989).

The Analysis and The Synthesis filters that are used in the thesis are shown in Figure 3.3.

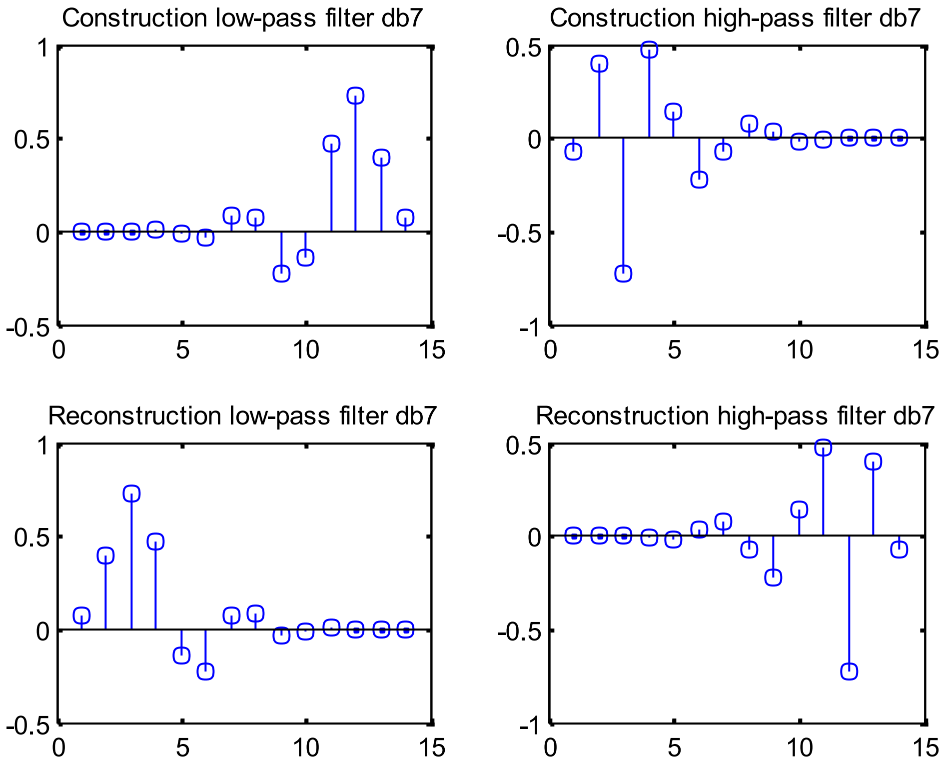


Figure 3.3: Analysis and Synthesis Filters for Daubechies 14 Wavelet

3.3.2 Biorthogonal Wavelet Transform

The biorthogonal wavelets introduced by Cohen, Daubechies, and Feauveau contain in particular compactly supported biorthogonal spline wavelets with compactly supported duals. In biorthogonal wavelets, separate decomposition and reconstruction filters are defined and hence the responsibilities of *analysis* and *synthesis* are assigned to two different functions (in the biorthogonal case) as opposed to a single function in the orthonormal case (Loganathan R et. al, 2010). The advantage of Biorthogonal Wavelet is that the symmetric wavelet will not weight the signal to any direction and thus no shifting will occur. The biorthogonal scaling function is given by

$$\phi(t) = 2\sum_{n=-\infty}^{\infty} h(n)\phi(2t-n) \qquad \underline{\text{dual}} \qquad \tilde{\phi}(t) = 2\sum_{n=-\infty}^{\infty} \tilde{h}(n)\tilde{\phi}(2t-n)$$

$$\left\langle \phi(t), \tilde{\phi}(t-k) \right\rangle = \delta(k) \qquad \left\langle \phi(2^{-k}t), \tilde{\phi}(2^{-k}t-n) \right\rangle = 2^{k}\delta(n) \tag{3.39}$$

Where $h(n)$ and $\tilde{h}$(n) serve as impulse response of FIR filters and two sets of scaling functions $\phi(t)$ and $\tilde{\phi}$(t) generate subspaces V_k and $\tilde{V}_k$ respectively. Unlike the orthogonal case, it is possible to synthesize biorthogonal wavelets and scaling functions which are symmetric or antisymmetric and compactly supported.

The Analysis and The Synthesis filters that are used in the thesis are shown in Figure 3.4.

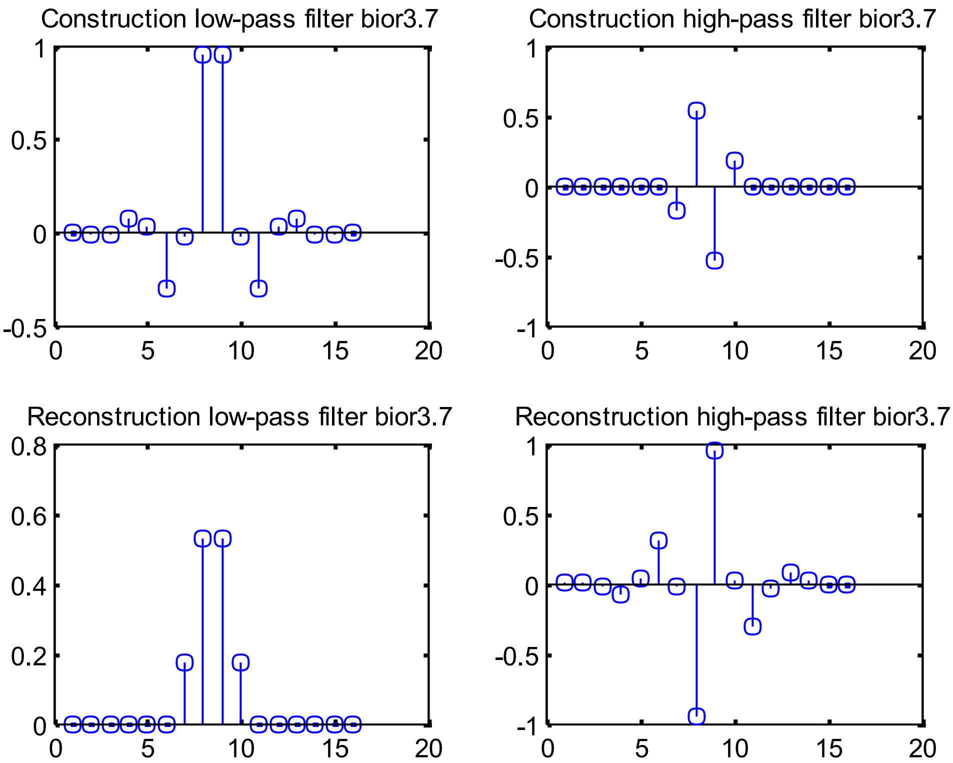


Figure 3.4: Analysis and Synthesis Filters for Biorthogonal 3.7 Wavelet

3.4 Multiresolution or Pyramidal Decomposition

The Wavelet Transform represents any arbitrary function, f(x), in terms of two sets of basis functions $\Psi_{m,n}$ (x) and $\Phi_{j,n}$ (x) where

$$f(x) = \sum_{n=-\infty}^{\infty} c_{j,n} \Phi_{j,n}(x) + \sum_{m=-\infty}^{j} \sum_{n=-\infty}^{\infty} d_{m,n} \Psi_{m,n}(x) \tag{3.40}$$

and $\Psi_{j,n}(x)$ and $\Phi_{j,n}$ span orthogonally complimentary spaces (Antonini et al., 1992). The detail or information lost going to the lower approximation of the original function, f(x), by taking the inner product of f(x) with $\Phi_{j,n}$, is given by the projection (or inner product) of the original function in the $\Psi_{j,n}(x)$ subspace. Thus, the discrete representation of a discrete function $a_{j-1,1}(f)$ at resolution j-1 is

$$c_{j-1,1}(f) = \sum_{n} (g_{2n-1} d_{m,n}(f) + h_{2n-1} c_{j,n}(f)) \tag{3.41}$$

where g_{2n-1} and h_{2n-1} are the discrete high and low pass filters, respectively, and $c_{j,n}$ and $d_{m,n}$ are the lower j resolution discrete signal and the details lost by going to the lower resolution signal, respectively (Akansu, 1994).

It follows that one can represent an arbitrary signal by its lower resolution and the detail lost going to the lower resolution with proper choices of the basis functions. Furthermore, one can extend this concept to a multiresolution approach where the original signal is represented by consecutively decomposing it to lower resolutions, while at the same time storing all the detail information between each consecutive resolution. This multiresolution approach can be extended to a 2-D image and is quite suitable for progressive transmission of images as in figure 3.5. In the 2-D multiresolution WT, one must introduce a scaling function Φ(x,y) and a wavelet function Ψ(x,y) that are separable. That is

$$\Phi(x,y) = \Phi(x)\Phi(y) \qquad (3.42)$$

and

$$\Psi(x,y) = \Psi(x)\Psi(y) \qquad (3.43)$$

In addition, the combinations of these functions are defined as

$$\Phi(x,y)_l = \Phi(x)\Phi(y) \qquad (3.44)$$

$$\Psi(x,y)_h = \Phi(x)\Psi(y) \qquad (3.45)$$

$$\Psi(x,y)_v = \Psi(x)\Phi(y) \qquad (3.46)$$

$$\Psi(x,y)_d = \Psi(x)\Psi(y) \qquad (3.47)$$

The filters h and s can be defined for the basis functions. The filters are h_r,h_c,g_r and g_c for Φ(x), Φ(y), Ψ(x), and Ψ(y), respectively. The subscripts r and c stand for row and column and indicate the direction in which the filters are applied (Mallat, 1989), (Akansu, 1994). So, the low pass and detail images are given by

$$c_l = h_r h_c c_0 \tag{3.48}$$

$$d_h = d_{1,1} = h_r g_c c_0 \tag{3.49}$$

$$d_v = d_{1,2} = g_r h_c c_0 \tag{3.50}$$

and

$$d_d = d_{1,3} = g_r g_c c_0 \tag{3.51}$$

Computationally, the 2-D WT of an image is taken in two parts. First, the 1-D WT is taken along the image pixel rows by multiplying each row by the appropriate low and high pass filters h and g. The low pass and detail groups are then down sampled by two. The second step in the 2-D WT is accomplished by taking the 1-D WT along the columns of each of the reordered low and high pass filtered groups.

This is again accomplished by multiplying both the low pass and high passes filtered groups by the same low pass and high pass filters discussed above. The columns are again down sampled by two. The result of this operation is a decomposition which has a low pass L image in quadrant one, a vertical error image D_v in quadrant two, a horizontal error image D_h in quadrant three, and a diagonal error image D_d in quadrant four. Thus, the first wavelet decomposition of the image is complete. Further wavelet transformations of the resulting low pass image will result in a multiresolution WT decomposition. Figure 3.5 and figure 3.6 represent the graphical representation of the multiresolution WT results and implementation.

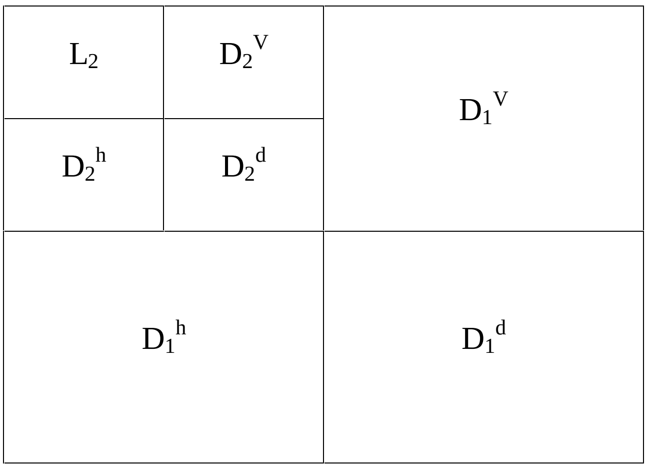


Figure 3.5: Multiresolution Decomposition

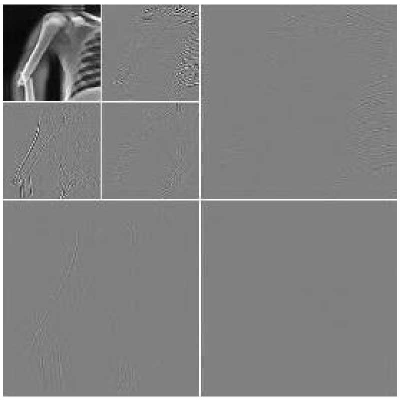


Figure 3.6: Two Level Wavelet Decomposition 256x256 X-ray Image using Daubechies' Wavelet

The wavelet type to be used in the image compression system depends on the images. The compression performance for images with different spectral activity decides the wavelet function from wavelet family. (Pradhan et al., 2010). The properties of the Daubechies Wavelets are asymmetric and orthogonal. The properties of Biorthogonal wavelets are symmetric but not orthogonal. Biorthogonal Wavelet with 3 coefficients in the construction part and 7 coefficients in the reconstruction part has been applied in the thesis, known as Biorthogonal 3.7. The Daubechies with 14 filter coefficients has been preferred to be applied in the image compression system because of the minimal retained energy, the better performance of PSNR has been reached and as the number of filter coefficients increase, the regularity increases (Grgic et al., 2001), (Lees, 2002). Biorthogonal Wavelet type 3.7 has been chosen among the wavelets because the best performance has been reached using the medical images. (Sadashivappa & Babu, 2008).

3.5 Summary

In this chapter, lossy image compression using Discrete Cosine Transform, Daubechies Wavelet Transform and Biorthogonal Wavelet Transform are discussed giving brief information about their algorithms. These algorithms have been preferred to be used because of different properties between each other. The Discrete Cosine Transform has been preferred to be applied on the images using the segment by segment compression to be compared with the Daubechies Wavelet Transform and Biorthogonal Wavelet Transform which the compression process has been applied to the whole image. The blocking artifacts comes out as the Relative Data Redundancy increases in DCT so, low Relative Data Redundancy can be achieved using DCT. Higher Relative Data Redundancy can be achieved using The Wavelet Transform applied in image compression.

CHAPTER 4

ASSESSMENT OF RELATIVE DATA REDUNDANCY

4.1 Overview

This chapter will cover the Linear Regression Analysis method applied in the decision of Relative Data Redundancy. An introduction to Linear Regression Analysis will be covered and the application in the field of image compression will be studied.

4.2 Multiple Linear Regression Analysis

Statically, linear regression is a method of relationship modelling between many scalar variables. In the linear regression, the unknown parameters can be modelled based on the estimation of relationships between the information data using linear relations or functions.

These models are known as “linear models”. Mostly, linear regression refers to a model where the mean of one variable is given as an affine function of another variable. Less common definitions of the linear regression appears in the cases where the median or another quantity of the conditional distribution of a variable is given as a linear function of the other variable. Like most types of regression analysis, the linear regression concentrates mainly on the probability distribution of two variables more than the joint probability distribution, which is the domain of multivariate analysis (Montgomery, Peck, Vining, 2013).

The multiple linear regressions can be simply considered as direct extensions of the simple regression. It allows for the use of multiple independent variables at once. The aim of multiple regression is the same as that of the simple regression. It is the prediction or explanation of the relationship between behaviour or response of a dependent variable to the variations of its independent variables under some conditions. However, the computations are considerably more complicated and must be performed by computers, and the inferential procedures may be more difficult to interpret, primarily because there may be relationships among the independent variables.

Linear regression was one of the first types of regression to be studied strictly, and to be employed widely in practical applications (Montgomery, Peck, Vining, 2013). This is because models which are linearly dependent on their unknown variables are easier to be predicted than models which are non-linearly correspondent to their unknown variables and because the statistical properties of the outcoming predictors are simpler to be determined. Linear regression has widely many different practical applications. Most of the linear regression applications situate in one of the following two broad categories:

- For prediction and forecasting aims, linear regression can be employed to suit a predictive model to a set of observed information data. After the development of such a model, any additional value of the independent variable can then be given without its related value of dependent variable; the model can be used to predict the dependent variable value. A dependent variable value can then be predicted easily without having previous idea about its value.
- Given a variable y and a set of variables $x_1, ..., x_p$ that may be correspondent to it, then linear regression analysis can be used to determine the relationship potency between y and the x_j, to measure which x_j is not having relationship with y at all, and to identify which subsets of the x_j have information about y, thus once one of them is known, the others are no longer informative.

There are a number of variables that can affect the Relative Data Redundancy. These variables include the Brightness, Variance of intensity, Contrast Weighted Entropy. Relative Data Redundancy and PSNR values will be applied with the variables to predict the Relative Data Redundancy for an x-ray image.

The models of linear regression are mostly predicted using the least squares method approach. Other approaches such as minimizing the “lack of fit” in other norms, or minimizing a penalized version of the least squares loss function can be employed for regression. On the contrary, non-linear models can be predicted and fitted using the least squares method. Inspite of the closed relationship between linear modelling and “least squares”, they are not totally synonymous

Linear Regression has been applied in image processing and image compression in the literature. Wei Gaofeng et al applied the characteristics of images such as the brightness of an image, the contrast, the ability of filter to focus energy, and entropy distribution of wavelet transform constants in order to provide accurate and intention way to evaluate image compression using wavelet filter without encoding or decoding, and has certain importance on the study of assessment and choice of wavelet filters (Gaofeng et al., 2010).

Adaptive linear regression has been applied in the enhancement of JPEG still image compression for archiving higher Relative Data Redundancy. Golomb-Rice coding is proposed to ameliorate the JPEG quality standards. As a result, the coding scheme that is not based on frequency analysis are encoded with this method and guaranteed in average quality. (Chang & Lai, 2009).

If the value of a variable y depends on the two independent variables x_1 and x_2, the equation for a multiple regression analysis can be given as follows;

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \tag{4.1}$$

where, the equation 4.1 shows a three dimensional plane on the space.
If the expected value of y on a three dimensional plane is accepted as $\hat{y}$, the equation 4.2 can be drawn as;

$$\hat{y} = \hat{\ }_0 + \hat{\ }_1 x_1 + \hat{\ }_2 x_2 \tag{4.2}$$

where, y is the regression equation on x_1 and x_2 and equation 4.3 represents the estimate prediction of mass regression equation as follows;

$$\text{«}_{y/x_1,x_2} = {}_0 + x_1 + {}_2 x_2 \tag{4.3}$$

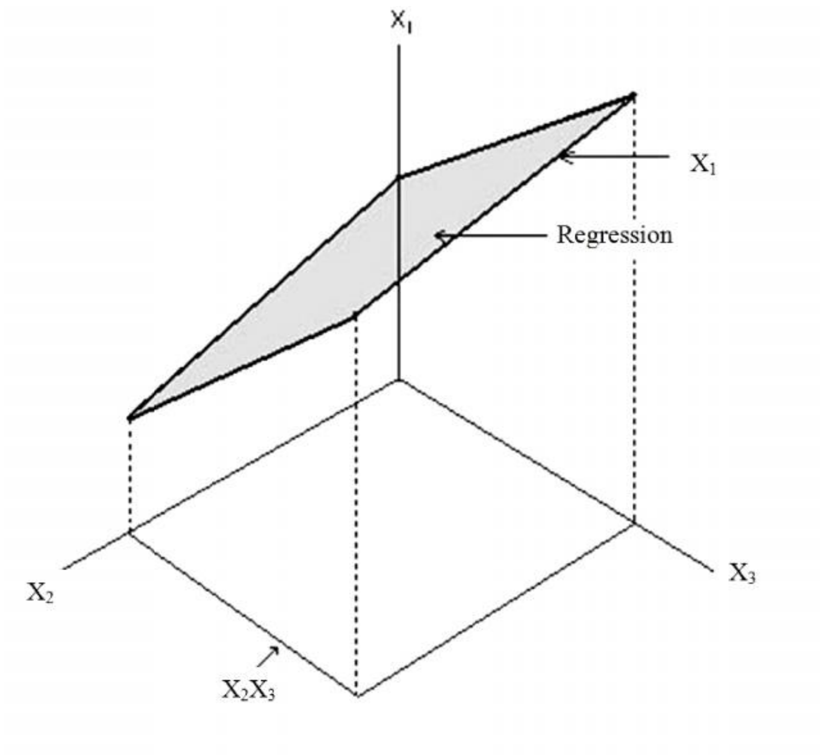


Figure 4.1: An Example of Regression Analysis of Two Variables

$$\sum_{i=1}^{n} e^2 = \sum_{i=1}^{n} (y - y')^2 = \sum_{i=1}^{n} (y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2)^2 \tag{4.4}$$

Mean square error estimation in linear regression can be estimated by taking the derivatives of the equation 4.4 with respect to β_0 , β_1 and β_2 using the equations below and getting the equations 4.5, 4.6 and 4.7 respectively;

$$\frac{\partial e}{\partial \hat{\beta}_0} = 2\sum(-1)\left(y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta} x_2\right) = 0$$

$$= -\sum y + n\hat{\beta}_0 + \hat{\beta}_1 \sum x_1 + \hat{\beta}_2 \sum x_2 = 0 \tag{4.5}$$

$$\frac{\partial e}{\partial \hat{\beta}_1} = 2\sum(-x_1)\left(y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2\right) = 0$$

$$= -\sum yx_1 + \hat{\beta}_0 \sum x_1 + \hat{\beta}_1 \sum x_1^2 + \hat{\beta}_2 \sum x_1 x_2 = 0 \tag{4.6}$$

$$\frac{\partial e}{\partial \hat{\beta}_2} = 2\sum(-2)\left(y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2\right) = 0$$

$$= -\sum yx_2 + \hat{\beta}_0 \sum x_2 + \hat{\beta}_1 \sum x_1 x_2 + \hat{\beta}_2 \sum x_2^2 = 0 \tag{4.7}$$

In order to remove (-)'ve signs in the equations, Σy, Σyx_1 and Σyx_2 are taken to the left hand side of the equations and the following equations found that are equation 4.8, 4.9 nd 4.10 respectively:

$$\sum y = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_1 + \hat{\beta}_2 \sum x_2^2 \tag{4.8}$$

$$\sum yx_1 = \hat{\beta}_0 \sum x_1 + \hat{\beta}_1 \sum x_1^2 + \hat{\beta}_2 \sum x_1 x_2 \tag{4.9}$$

$$\sum yx_2 = \hat{\beta}_0 \sum x_2 + \hat{\beta}_1 \sum x_1 x_2 + \hat{\beta}_2 \sum x_2^2 \tag{4.10}$$

Instead of y, x_1 and x_2 in the equations above, the standart deviation of the arithmetic means of these values are given in order to get the following equations 4.11, 4.12 and 4.13.

$$\sum y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_{1i} + \hat{\beta}_2 \sum x_{2i} \tag{4.11}$$

$$\sum y_i x_{1i} = \hat{\beta}_0 \sum x_{1i} + \hat{\beta}_1 \sum x_{1i}^2 + \hat{\beta}_2 \sum x_{1i} x_{2i} \tag{4.12}$$

$$\sum y_i x_{2i} = \hat{\beta}_0 \sum x_{2i} + \hat{\beta}_1 \sum x_{1i} x_{2i} + \hat{\beta}_2 \sum x_{2i}^2 \tag{4.13}$$

In order to find β_0, β_1 and β_2, x_i and y_i should be replaced by the difference between these values and their rithmetic mean as in equation 4.14 and eqution 4.15.

$$x_i' = x_i - \overline{x_i} \tag{4.14}$$

$$y_i' = y_i - \overline{y_i} \tag{4.15}$$

So, the equation 4.16, equation 4.17 and equation 4.18 are derived as.

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}_1 - \hat{\beta}_2 \overline{x}_2 \tag{4.16}$$

$$\hat{\beta}_1 = \frac{\left(\sum x_{1i}' y_i'\right)\left(\sum x_{2i}'^2\right) - \left(\sum x_{2i}' y_i'\right)\left(\sum x_{1i}' x_{2i}'\right)}{\left(\sum x_{1i}'^2\right)\left(\sum x_{2i}'^2\right) - \left(\sum x_{1i}' x_{2i}'\right)^2} \tag{4.17}$$

$$\hat{\beta}_2 = \frac{\left(\sum x_{2i}' y_i'\right)\left(\sum x_{1i}'^2\right) - \left(\sum x_{1i}' y_i'\right)\left(\sum x_{1i}' x_{2i}'\right)}{\left(\sum x_{1i}'^2\right)\left(\sum x_{2i}'^2\right) - \left(\sum x_{1i}' x_{2i}'\right)^2} \tag{4.18}$$

4.2.1 Hypothesis Testing

In statistical significance testing, the p-value is the probability of gaining a test statistic at any rate as excessive as the one that was really noticed, assuming that the null hypothesis is correct. The lower the p-value, the less likely the result is if the null hypothesis is correct, and by consequence the more "significant" the outcome is, in the sense of statistical meaning or significance. The null hypothesis can be often rejected when the p-value is less than 0.05 or

0.01, corresponding respectively to a 5% or 1% chance of rejecting the null hypothesis while it is correct. In this thesis, the significance values of less than 5% are accepted and applied. The coefficients of the equations are found using the SPSS and the results for each compression method are shown in Table 4.1, Table 4.6 and Table 4.11 considering the significance of 0.5 or less than 0.5. The results using SPSS using different variations shows that the hypothesis testing is correct for the suggested variations.

In multiple linear regression issues, certain tests of hypotheses about the model parameters are helpful in measuring the usefulness of the model. In this part, we describe several important hypothesis-testing methods. These methods imply that the errors $£_i$ in the model be normally and independently distributed with null mean and variance σ^2 , NID (0, σ^2). As a result of this assumption, the observations $¥_i$ are normally and independently distributed with mean $\;_0 +$ $\Sigma_{j=1}^{k} \;_j ¡_{ij}$ and variance σ^2.

4.2.1.1 Test for Significance of Regression

The test for significance of regression is an analysis carried out to determine whether a linear relationship exists between the response variable ¥ and a subset of variables $¡_{1,}\, ¡_{2.........,}\, ¡_{k.}$ The appropriate hypotheses are

$$\begin{aligned} &H_0 : \;_1 = \;_2 = \ldots = \;_{k=} 0 \\ &H_1 = \;_j \neq 0 \text{ for at least one } j \end{aligned} \tag{4.19}$$

Rejection of H_0 in Equation 4.19 requires that at least one of the regressor variables $x_{1,} x_{2,} ...., x_k$ contributes significantly to the model. The test process involves an analysis of variance partitioning of the total sum of squares SS_T into a sum of squares due to the model (or to regression) and a sum of squares due to remaining (or error);

$$SS_T = SS_R + SS_{\mathrm{E}} \tag{4.20}$$

Now if the null hypothesis $H_0 : \;_1 = \;_2 = ... = \;_k = 0$ is true, then $SS_R / \pm^2$ is distributed as X_k^2, where the number of freedom degrees for X^2 is equal to the number of regressor variables in the model. $SS_{\mathrm{E}} / \pm^2$ can be distributed as X_{n-k-1}^2 and that SS_E and SS_R are independent. The test procedure for $H_0 : \;_1 = \;_2 = ... = \;_k = 0$ is to compute

$$F_0 = \frac{SS_R / k}{SS_E l(n-k-1)} = \frac{MS_R}{MS_E} \tag{4.21}$$

and to reject H_0 if F_0 exceeds $F_{\sim,k,n-k-1.}$ Alternatively, $P-$value approach is applied in hypothesis testing and, thus, reject H_0 if the $P-$value for the statistic F_0 is less than $\sim$. The test can be generally summarized in an investigation of variance table like in Table 4.1.

A computational formula for SS_R may be found easily. The computational formula for SS_E in Equation 4.22;

$$SS_E = y'y - \hat{\ }'X'y \tag{4.22}$$

Now, because

$$\begin{aligned} SS_T &= \Sigma_{i=1}^{n} y_i^2 - (\Sigma_{i=1}^{n} y_i)^2 / n \\ &= y'y - (\Sigma_{i=1}^{n} y_i)^2 / n, \end{aligned} \tag{4.23}$$

The foregoing equation can be rewritten as;

$$SS_E = y'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - \left[\hat{\ }'X'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \right] \tag{4.24}$$

Or

$$SS_E = SS_T - SS_R \tag{4.25}$$

Therefore, the regression sum of squares is;

$$SS_R = \hat{\ }'X'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{4.26}$$

and the error sum of squares is;

$$SS_E = y'y - \hat{\ }' X'y \tag{4.27}$$

and the total sum of squares is;

$$SS_T = y'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{4.28}$$

Table 4.1 Variance Analysis for Significance of Regression

Analysis of Variance for Significance of Regression in Multiple Regression				
Source of Variance	**Squares sum**	**Freedom Degrees**	**Mean Square**	F_0
Regression	SS_R	k	MS_R	MS_R / MS_E
Error or remaining	SS_E	$n-k-1$	MS_E	
Total	SS_T	$n-1$		

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{4.29}$$

R^2 is a measure of the amount of reduction in the variability of y obtained by using the regressor variables $x_{1,} x_{2,} ..., x_k$ in the model. However, as we have noted previously, a large value of R^2, apart from whether the additional variable is statistically significant or not. Thus, it is possible for models that have large values of R^2 to give poor predictions of new observations or estimates of the mean response.

as R^2 always increases as with the addition of new conditions to a model, some experts of regression modeling prefer to use an adjusted R^2 statistic that can be given by;

$$R^2_{adj} = 1 - \frac{SS_E l(n-p)}{SS_T l(n-1)} = 1 - \left(\frac{n-1}{n-p}\right)(1-R^2) \tag{4.30}$$

In general, the adjusted R^2 statistic will not always increase as variables are added to the model. On the contrary, the value of R^2_{adj} will often decrease when any unnecessary terms are added.

4.2.1.2 Tests on Individual Coefficients of Regression and Group of Coefficients

Researchers are mostly interested in trying hypotheses and applying tests on the individual regression coefficients. Such tests are useful for finding the value of each regressor variable in the regression model. As an example, the model can be more valuable and efficient if additional variables were included or even if one or more variables of the model were deleted. The addition of variables to the regression model most probably increases the sum of squares for regression and decreases the sum of the squared errors. It is very important to take the decision whether the increase in the regression squares sum is sufficient to deserve including the additional variables in the model. Over more, including an unimportant variable to the model can in fact increase the mean square error MSE, which means that the model is becoming less useful.

The hypotheses for testing the significance of any individual regression coefficient, say $_{j,}$ are

$$\begin{aligned} H_0 &: \ _j = 0 \\ H_1 &: \ _j \neq 0 \end{aligned} \tag{4.31}$$

If $H_0 : \ _j = 0$ is not rejected, then this indicates that x_j can be deleted from the model. The test statistic for this hypothesis is

$$t_0 = \frac{\hat{\ }_j}{\sqrt{\hat{\pm}^2 C_{jj}}} \tag{4.32}$$

Where C_{jj} is the diagonal element of $(X'X)^{-1}$ corresponding to $\hat{\ }_j$. The null hypothesis $H_0 : \ _j = 0$ is rejected if $|t_0| > t_{\sim/2,n-k-1.}$ Note that this is really a partial or marginal test because the regression coefficient $_j$ depends on all the other regressor variables $x_i (i \neq j)$ that are in the model.

The denominator of Equation 4.32. $\sqrt{\hat{\pm}^2 C_{jj}}$, is often called the standard error of the regression coefficient $\hat{\ }_j$. That is,

$$se(\hat{\ }_j) = \sqrt{\hat{\pm}^2 C_{jj}} \tag{4.33}$$

Therefore, an equivalent way to write the rest statistic in Equation 4.32 is

$$t_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{4.34}$$

4.3 Estimation of Relative Data Redundancy

The multiple linear regression model is defined as a direct extension of the simple linear model defined in equation 4.35. The model is given by,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \tag{4.35}$$

Where, y is the dependent variable, x_j , j = 1, 2, . . . , m, represent m different independent variables, β_0 is the intercept (value when all the independent variables are 0), β_j , j = 1, 2, . . . , m, represent the corresponding mregression coefficients and ε is the random error, mostly considered to be a normal distribution with zero mean and a variance of σ^2.

Multiple Linear Regression Analysis has been applied in this thesis in Lossy image compression techniques; Discrete Cosine Transform, Daubechies Wavelet Transform and Biorthogonal Wavelet Transform to estimate the Relative Data Redundancy using SPSS 17.

4.3.1 Estimation of Relative Data Redundancy of DCT based Image Compression

The estimation of Relative Data Redundancy using Discrete Cosine Transform can be calculated using the equation 4.36;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{4.36}$$

where y represents the PSNR value that is denoted as P, β_0 represents the constant variable after linear regression has been applied, β_1 represents the coefficient for Relative Data Redundancy which is defined as *R*, β_2 represents the coefficient for the relationship of the characteristics of the image given in the equation below in equation 4.37.

$$P = \beta_0 + \beta_1 R + \beta_2 X \tag{4.37}$$

and *X* represents the relationship between the characteristics of an image given below in equation 4.38 and equation 4.39 denoted as X_1 and X_2;

$$X_1 = \frac{V * B}{E} \tag{4.38}$$

where V represents the logarithm of the variance of the image and B represents the logarithm of intensity of the image, E represents the logarithm of entropy of the image and X_1 represents the relationship between these characteristics of an original image in equation 4.39.

$$X_2 = \frac{V * B}{CWE} \tag{4.39}$$

where V represents the logarithm of the variance of the image and B represents the logarithm of intensity of the image, CWE represents the logarithm of contrast weighted entropy of the image and X_2 represents the relationship between these characteristics of an original image in equation 4.39. Table 4.2 represents the results for LRA based DCT Compression using two equations.

Table 4.2 Results for LRA based DCT Compression using two equations

DCT	R Squre	Sig1	Sig2	Sig3
X_1	0,437	0,000	0,000	0,327
X_2	0,452	0,000	0,000	0,001

Table 4.3 Variables used in Finding the DCT Relative Data Redundancy

Variables Entered			
Model	**Variables Entered**	**Dependent Variable**	**Method**
1	X_1 or X_2 LOG CR	LOG PSNR Y	Enter

Table 4.3 represents the variables entered to the SPSS in applying the multiple linear regression analysis. There are two variables applied in finding the linear regression which are X in equation 4.22 and LOG CR, the logarithm of Relative Data Redundancy. The logarithm of PSNR has been used as dependent variable as LOG Y.

Table 4.4 Model Summary of the DCT Compression using X_1

Model Summary for X_1				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
1	.661	.437	.434	.0853610607

Table 4.5 Model Summary of the DCT Compression using X_2

Model Summary for X_1				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
1	.672	.452	.449	.0842225608

Table 4.4 and Table 4.5 represents the model summary of the DCT compression using two variances of X_1 and X_2. In the column labelled R are the values of multiple correlation coefficient between the predictors and the outcome. The predictors are equation 4.22 and the logarithm of the Relative Data Redundancy used in the model. The next column labelled R^2 is a measure of how much of the variability in the outcome is accounted for by the predictors. The adjusted R^2 gives the idea of how well the model generalizes and ideally the value of adjusted R^2 should be close to R^2. The last column used above represents the standard error of the estimate.

Table 4.6 Analysis of Variance of DCT based Image Compression using X_1

ANOVA					
Model 1	**Sum of Squares**	**df**	**Mean Square**	**F-Ratio**	**Sig.**
Regression	2.107	2	1.054	144.587	.000[a]
Residual	2.718	373	.007		
Total	4.825	375			

Table 4.7 Analysis of Variance of DCT based Image Compression using X_2

ANOVA[b]					
Model 1	**Sum of Squares**	**df**	**Mean Square**	**F-Ratio**	**Sig.**
Regression	2.179	2	1.090	153.599	.000[a]
Residual	2.646	373	.007		
Total	4.825	375			

Table 4.6 and Table 4.7 represents the analysis of variance of the DCT based image compression using X_1 and X_2. The analysis of variance table ckecks whether the model is better at predicting the results than using the mean. particularly, F-Ratio represents the ratio of the prediction enhancement resulting from fitting the regression model in the table, in comparison to the lack of accuracy that exists in the model labeled as residual.

Table 4.8 Coefficients of the Model of DCT based Image Compression using X_1

Coefficients[a]					
Model 1	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
	B	**Std. Error**	**Beta**		
(Constant)	2.121	.050		42.615	.000
LogR	-.331	.019	-.660	-16.981	.000
X_1	-.005	.005	-.038	-0.981	.327

Table 4.9 Coefficients of the Model of DCT based Image Compression using X_2

Coefficients[a]					
Model 1	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
	B	**Std. Error**	**Beta**		
(Constant)	2.073	.026		78.937	.000
LogR	-.325	.019	-.648	-16.844	.000
X_2	-.001	.000	.128	3.338	.001

Table 4.8 and Table 4.9 gives the coefficients used in the linear regression model using X_1 and X_2. The unstandardized coefficients are B value and Standard error for constant value, logarithm of the Relative Data Redundancy and the equation 4.22. Each one of the beta values has a standard connected error representing to what degree these values would vary through different samples, these standard errors can be used to determine whether or not the b value differs significantly from zero using the t-statistic. Therefore, weather the t-test connected with a b value is considerable and less than 0.05, then the predictor is making important involvement to the model. The equation 4.40 gives us the result of the DCT Model of the Image compression system using the coefficients considering X_1 and X_2.

$$LogP = 2.073 - 0.325(LogR) + 0.01(X_2) \tag{4.40}$$

4.3.2 Estimation of Relative Data Redundancy of Daubechies Wavelet based Image Compression

The estimation of Relative Data Redundancy using Discrete Wavelet Transform can be calculated using the equation 4.41;

$$y = {}_0 + {}_1x_1 + {}_2x_2 + \ldots + {}_nx_n \tag{4.41}$$

where y represents the Logarithm of the PSNR value, β_0 represents the constant variable after linear regression has been applied, β_1 represents the coefficient for Relative Data Redundancy which is defined as *LogCR*, β_2 represents the coefficient for the relationship of the some characteristics of the image given in the equation below in quation 4.42.

$$P = \ _0 + \ _1(R) + \ _2 X \tag{4.42}$$

and X represents the relationship between the charasteristics of an image given below in equation 4.43 and equation 4.44 denoted as X_1 and X_2;

$$X_1 = \frac{V * B}{E} \tag{4.43}$$

where V represents the logarithm of the variance of the image and B represents the logarithm of intensity of the image, E represents the logarithm of entropy of the image and X_1 represents the relationship between these characteristics of an original image in equation 4.44.

$$X_2 = \frac{V * B}{CWE} \tag{4.44}$$

where V represents the logarithm of the variance of the image and B represents the logarithm of intensity of the image, CWE represents the logarithm of contrast weighted entropy of the image and X_2 represents the relationship between these characteristics of an original image in equation 4.44. Table 4.10 represents the results for LRA based DBW Compression using two equations.

Table 4.10 Results for LRA based DBW Compression

DAUB	R Squre	Sig1	Sig2	Sig3
X_1	0,705	0,000	0,000	0,277
X_2	0,705	0,000	0,000	0,212

Table 4.11 Variables used in Finding the Daubechies Wavelet Relative Data Redundancy

Variables Entered/ Removed[b]			
Model	**Variables Entered**	**Dependent Variable**	**Method**
2	*X* LOG CR	LOG PSNR (LOG Y)	Enter

Table 4.11 represents the variables entered to the SPSS in applying the multiple linear regression analysis. There are two variables applied in finding the linear regression which are X in equation 4.25 and LOG CR, the logarithm of Relative Data Redundancy. The logarithm of PSNR has been used as dependent variable as LOG Y. Table 4.12 and Table 4.13 represents model summary of the daubechies wavelet relative data redundancy using X_1 and X_2.

Table 4.12 Model Summary of the Daubechies Wavelet Relative Data Redundancy using X_1

Model Summary				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
2	.839	.705	.703	0.4332949589

Table 4.13 Model Summary of the Daubechies Wavelet Relative Data Redundancy using X_2

Model Summary				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
2	.840	.705	.704	0.4331203932

In the column labelled R are the values of multiple correlation coefficients between the predictors and the results. The predictors are equation 4.24 and the logarithm of the Relative Data Redundancy used in the model. The next column labelled R^2 is a measure of how much of the variability in the results is considered by the predictors. The adjusted R^2 gives the idea of how well the model generalizes and how the value of adjusted R^2 must be close to R^2. The last column used above represents the error of the estimate.

Table 4.14 and Table 4.15 represents the ANOVA of Daubechies Wavelet based image compression using X_1 and X_2.

Table 4.14 Analysis of Variance of Daubechies Wavelet based Image Compression using X_1

ANOVA[b]					
Model 2	**Sum of Squares**	**Df**	**Mean Square**	**F**	**Sig.**
Regression **Residual** **Total**	2.091 .877 2.967	2 467 469	1.045 .002	556.794	.000[a]

Table 4.15 Analysis of Variance of Daubechies Wavelet based Image Compression using X_2

ANOVA[b]					
Model 2	**Sum of Squares**	**Df**	**Mean Square**	**F**	**Sig.**
Regression **Residual** **Total**	2.091 .876 2.967	2 467 469	1.045 .002	556.431	.000[a]

The analysis of variance table shows whether the model is better at predicting the outcome than using the mean as a reference method. Generally, the ratio between the evolutions in prediction as a result of fitting the regression model in the table in comparison with the error labeled as residual can be given by F-Ratio.

Table 4.16 Coefficients of the Model of Daubechies Wavelet based Image Compression using X_1

Coefficients[a]					
Model 2	**Unstandardized Coefficients**		**Standardized Coefficients**		
	B	**Std. Error**	**Beta**	**t**	**Sig.**
(Constant) **LOG CR (LOG CR)** **Equation 4.6**	3.060 -.739 .002	.045 .022 .002	 -.839 .027	68.243 -33.361 1.089	.000 .000 .277

Table 4.17 Coefficients of the Model of Daubechies Wavelet based Image Compression using X_2

Coefficients[a]					
Model 2	**Unstandardized Coefficients**		**Standardized Coefficients**		
	B	**Std. Error**	**Beta**	**t**	**Sig.**
(Constant) **LOG CR (LOG CR)** **Equation 4.6**	3.077 -.737 .0001	.041 .022 .000	 -.836 .032	75.381 -33.167 1.250	.000 .000 .212

Table 4.16 and Table 4.17 gives the coefficients used in the linear regression model. The unstandardized coefficients are B value and Standard error for constant value, logarithm of the Relative Data Redundancy and the equation 4.26. Each one of the beta values has a corresponding error representing the degree to which they would vary through different samples. These errors are used to find out if the B values are considerably different from zero using the t-statistic. By consequent, to find out whether the t-test related with a b value is significant and less than 5%. That is, the predictor is considered to be contributing effectively in the model. The equation 4.45 gives us the result of the Daubechies Wavelet Model of the Image compression system using the coefficients above.

$$LogPSNR = 3.077 - 0.737(LogCR) + 0.0001(X_2) \tag{4.45}$$

4.3.3 Estimation of Relative Data Redundancy of Biorthogonal Wavelet based Image Compression

The estimation of Relative Data Redundancy using Discrete Wavelet Transform can be calculated using the the equation 4.46;

$$y = {}_0 + {}_1 x_1 + {}_2 x_2 + \ldots + {}_n x_n \tag{4.46}$$

where y represents the Logarithm of the PSNR value, β_0 represents the constant variable after linear regression has been applied, β_1 represents the coefficient for Relative Data Redundancy which is defined as *LogCR*, β_2 represents the coefficient for the relationship of the some characteristics of the image given in the equation below in quation 4.47.

$$P = {}_0 + {}_1(R) + {}_2 X \tag{4.47}$$

and X represents the relationship between the charasteristics of an image given below in equation 4.48 and equation 4.49 denoted as X_1 and X_2;

$$X_1 = \frac{V * B}{E} \tag{4.48}$$

where V represents the logarithm of the variance of the image and B represents the logarithm of intensity of the image, E represents the logarithm of entropy of the image and X_1 represents the relationship between these characteristics of an original image in equation 4.48.

$$X_2 = \frac{V * B}{CWE} \tag{4.49}$$

where V represents the logarithm of the variance of the image and B represents the logarithm of intensity of the image, CWE represents the logarithm of contrast weighted entropy of the image and X_2 represents the relationship between these characteristics of an original image in equation 4.49. Table 4.18 represents the results for LRA based BWT Compression using two equations.

Table 4.18 Results for LRA based BW Compression

BIOR	R Squre	Sig1	Sig2	Sig3
X_1	0,620	0,000	0,000	0,987
X_2	0,628	0,000	0,000	0,001

Table 4.19 Variables used in Finding the Biorthogonal Wavelet Relative Data Redundancy

Variables Entered/ Removed[b]			
Model	**Variables Entered**	**Dependent Variable**	**Method**
3	X LOG CR	LOG PSNR (LOG Y)	Enter

Table 4.19 represents the variables entered to the SPSS in appying the multiple linear regression analysis. There are two variables applied in finding the linear regression which are X in equation 4.29 and LOG CR, the logarithm of Relative Data Redundancy. The logarithm of PSNR has been used as dependent variable as LOG Y. Table 4.20 and Table 4.21 represents model summary of the biorthogonal wavelet relative data redundancy using X_1 and X_2.

Table 4.20 Model Summary of the Biorthogonal Wavelet Relative Data Redundancy using X_1

Model Summary				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
3	.787	.620	.618	.05794578736

Table 4.21 Model Summary of the Biorthogonal Wavelet Relative Data Redundancy using X_2

Model Summary				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
3	.792	.628	.626	.05731070893

In the columns R of the table 4.20 and 4.21 are the values of multiple correlation coefficients between the predictors and the results. The predictors are equation 4.28 and the logarithm of the Relative Data Redundancy used in the model. The next column labelled R^2 is a measure of how much of the variability in the result is considered for by the predictors. The adjusted R^2 gives the idea of how the model can generalize and the value of adjusted R^2 is to be approaching to R^2. The last column shown in the table above is the standard estimate error.

Table 4.22 and Table 4.23 represents the ANOVA of Biorthogonal Wavelet based image compression using X_1 and X_2.

Table 4.22 Analysis of Variance of Biorthogonal Wavelet based Image Compression using X_1

ANOVA[b]					
Model 3	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Regression	2.555	2	1.277	380.462	.000[a]
Residual	1.568	467	.003		
Total	4.123	469			

Table 4.23 Analysis of Variance of Biorthogonal Wavelet based Image Compression using X_2

ANOVA[b]					
Model 3	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Regression	2.589	2	1.295	394.144	.000[a]
Residual	1.534	467	.003		
Total	4.123	469			

The analysis of variance table tests if the model is really and noticeably better in the result prediction than considering the mean as a 'best guess'. The F-Ratio which is the ratio of the amelioration in prediction that results from fitting the regression model, as a relation with the inaccuracy that exists in the model which is called residual.

Table 4.24 Coefficients of the Model of Biorthogonal Wavelet based Image Compression using X_1

Coefficients[a]					
Model 3	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
	B	**Std. Error**	**Beta**		
(Constant)	3.226	.060		53.594	.000
LOG CR (LOG CR)	-.817	.030	-.787	-27.585	.000
Equation 4.5	4.992E-5	.003	,000	0.017	.987

Table 4.25 Coefficients of the Model of Biorthogonal Wavelet based Image Compression using X_2

Coefficients[a]					
Model 3	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
	B	**Std. Error**	**Beta**		
(Constant)	3.213	.054		59.489	.000
LOG CR (LOG CR)	-.810	.029	-.780	-27.536	.000
Equation 4.5	.001	.000	.091	3.226	.001

Table 4.24 and Table 4.25 gives the coefficients used in the linear regression model using X_1 and X_2. The unstandardized coefficients are B value and Standart error for constant value, logarithm of the Relative Data Redundancy and the equation 4.35. Each of the beta values has a related standard error showing to what point these values can vary with different samples. These errors are used to find out whether the b value is considerably different from zero using the t-statistic. Therefore is the t-test connected with a b value is significant and less than 5%, and then the predictor is making good donation to the model. The equation 4.50 gives us the result of the DCT Model of the Image compression system using the coefficients above.

$$LogPSNR = 3.213 - 0.810(LogCR) + 0.001(X_2) \tag{4.50}$$

4.4 Graph and Results

Using the equations 4.23, 4.27 and 4.36, the following three dimensional graph has been drawn to mention that all the results of the equations are the points on the graph. Table 4.26 has been used to get the graph of the results of the LRA.

Table 4.26 Coefficients of the Equations

	B_0	B_1	B_2
DCT	2.073	-0,325	0,001
DWT	3.077	-0,737	0.0001
BWT	3.213	-0.810	0.001

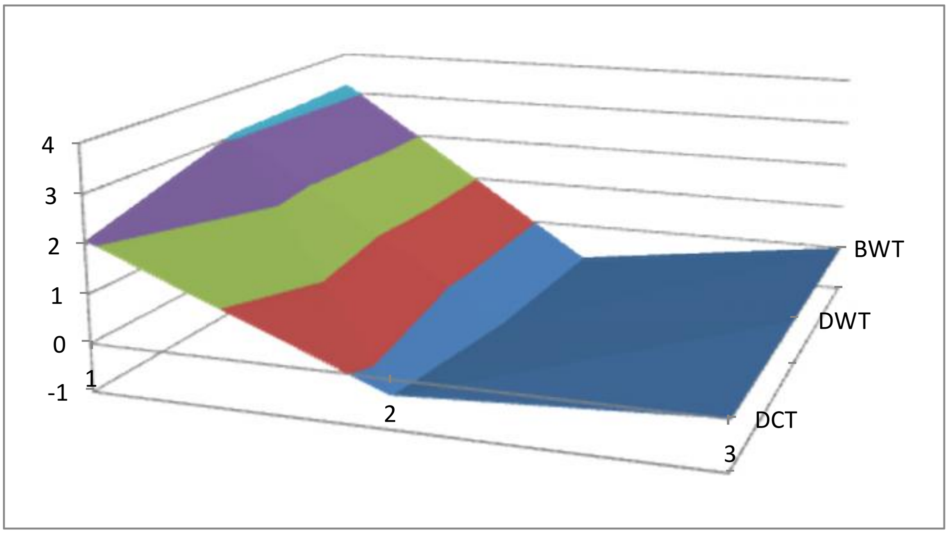


Figure 4.2: The 3-D Graph of the Results of the LRA

4.5 Summary

In this chapter, Linear Regression Analysis based image compression has been applied using the statistical values and characteristics of the images. The suggested system is capable of calculating the optimum Relative Data Redundancy for the x-ray images considering the characteristics without applying any compression method.

CHAPTER 5

NEURAL NETWORK APPROACHES TO IMAGE COMPRESSION

5.1 Overview

This chapter presents the use of back propagation neural networks in the field of image compression. The results suggest that back propagation neural networks can be efficiently used in deciding the suitable Relative Data Redundancy after presenting an image to suggested system.

5.2 Mean Opinion Score

The Mean Opinion Score (MOS) is the mathematical average of all the individual results, it ranges from 1 which is considered (worst) to 5 which is the (best). The MOS is the score applied by averaging the results of a set of standard, subjective tests. In multimedia especially when codecs are used to compress the bandwidth requirement, the MOS offers a numerical indication of the perceived quality from the users' perspective of received media after compression and/or transmission. The score is given as a single number in the range 1 to 5, where 1 is the worst in terms of quality, and 5 is the best measurement as shown in Table 5.1.

There are 10 experts who scored 90 images in total with 27 Relative Data Redundancy that involve Discrete Cosine Transform, Daubechies Wavelet Transform and Biorthogonal Wavelet Transform having Relative Data Redundancy of 10 and 90 percent for each method.

Table 5.1 Mean Opinion Score

MOS	Quality	Impairment
1	Bad	Very annoying
2	Poor	Annoying
3	Fair	Slightly annoying
4	Good	Perceptible but not annoying
5	Excellent	Imperceptible

MOS has been calculated using the average scores for each Relative Data Redundancy using 10 experts. The maximum MOS value for each image and each Relative Data Redundancy has been used as the MOS value of the image.

Equation 5.1 represents The Mean Opinion Score (MOS_{MR}), computed by taking the opinion score given by the n^{th} viewer to the MR^{th} image and n represents the number of viewers that participated in each experiment.

$$MOS_{MR} = \frac{1}{n}\sum_{i=1}^{n} S_{i,MR} \tag{5.1}$$

For each image with 27 Relative Data Redundancy, a MOS value of 3.75 which is the 75% of the highest score of 5 has been used as the minimum value for each image are considered as the targets for the Neural Network System. When MOS values has been considered for the image database, a maximum value of CR for DCT has been reached as 40% and a minimum value of CR for Wavelet has been reached as 50%.

Equation 5.2 represents that The Maximum target is the Mean Opinion Score for the images (k).

$$MaxS_j = MOS_k \tag{5.2}$$

5.3 Back Propagation Neural Networks

Back-propagation was created by generalizing the Widrow-Hopff learning rule to multiple-layer networks and nonlinear differentiable transfer functions (Haykin, 1999). Inputs and their correspondent outputs are used to train or learn a neural network until it can predict a function or an input output relationship, connecting input values with specific output values, or classify input sets in an appropriate way which is defined by the user. Networks are constructed using biases, sigmoid functions, and linear output layers are capable of approximating any relation with a finite number of discontinuities. Standard back propagation algorithm is a gradient descent algorithm, as in the Widrow Hopff learning algorithm, where the network weights are moved along the negative of the gradient of the error function. The word back-propagation refers to the way in which the gradient is computed for nonlinear multilayer networks and based on that gradient is employed for the computing of the weights in each layer.

Well trained back-propagation networks tend to give logical answers when fed with inputs that they have never been seen. Classically, a new input leads to an output similar to the correct output for inputs used in the learning algorithm that are similar to the new input being fed. This property makes it easy to train a network on a delegate set of input-output pairs and get fine results without training the network on all possible input-output pairs. There are two features of the Back-Propagation Neural Networks. These are the training or learning of the ANN and the generalization of the ANN.

5.3.1 Back Propagation Learning Algorithm in Neural Networks

The Back-propagation algorithm was initially proposed by Paul Werbos in the 1970's. However, it was re-discovered in 1986 by Rumelhart and McClelland and then the Back-Propagation algorithm became extensively used.

To describe the neural networks model based on the back propagation learning algorithm, various types of elements must be checked in order to see how they can join to form the general topology of back-propagation.

These elements that combine to form the back propagation topology of neural network, will be explained below. The neural network training operations is separated into two processes. First, the feed-forward processing or propagation. This is used both in the test operation of the trained neural networks and training process of the network. Second; the error processing of back-propagation and the back-propagation of this error. These two operations are applied only during the training or learning process of the neural networks. Before the two categories of processing are presented, another important element which is the activation function will be discussed and presented. Detailed discussions about the equations will be presented in this chapter, (Haykin, 1999).

5.3.1.1 The Activation Function

An artificial neuron which is the base structural block in a back-propagation network is shown in Figure 5.1. The input of the neuron named as "*net*" is obtained as a weighted sum and it is given by equation 5.3 where O_i represents the summation of the outputs for the input layer with weights w_i and n represents the number of the inputs.

$$net = \sum_{i=1}^{n} O_i w_i \tag{5.3}$$

In Figure 5.1, F(net) is the activation function in the form of a sigmoid function. The ease of the derivative of the sigmoid function justifies its attractiveness and use as an activation function in the learning algorithms.

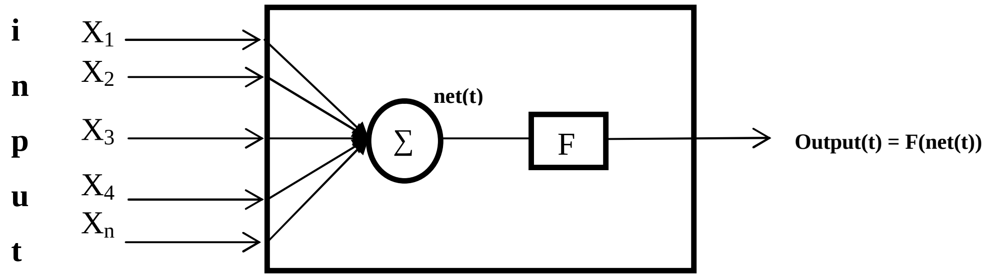


Figure 5.1: Artificial Neuron

Using a sigmoid activation function, the output of the artificial neuron can be given in equation 5.4 (Simon Haykin, 1999).

$$F(net) = \frac{1}{(1 + \exp(-net))} \tag{5.4}$$

The derivative of sigmoid function is found using the equation 5.5:

$$\frac{\partial F(net)}{\partial net} = \frac{\exp(-net)}{(1 + \exp(-net))^2} = F(net)([1 - F(net)]) \tag{5.5}$$

All differentiable functions everywhere can be used in the Back-Propagation Neural Network learning algorithm. Linear adjustable gain functions, linear threshold characteristic functions, relay functions with threshold characteristics and sigmoid functions with different gain values are suitable and common used activation functions.

5.3.1.2 Calculations of Feed Forward

Figure 5.2 is the most common configuration for a simple three layer back propagation model back-propagation neural network. Each neuron within the network is represented by a circle and each interconnection with the associated weight, using an arrow. The neurons that are labelled as h are the bias neurons. Normalization is necessary for the input data prior to training. The numerical values of the input data into the input layer preferred to be in the range of (0 - 1). The feed forward calculations for the stages can be described with respect to the layers.

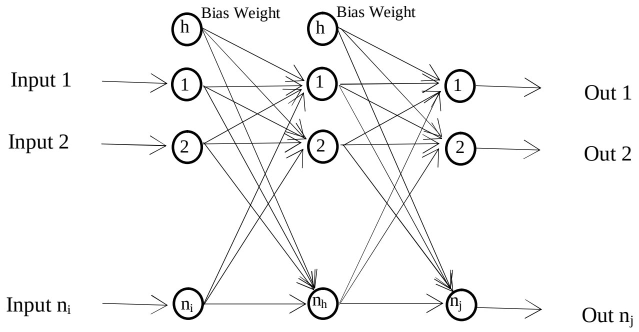


Figure 5.2: Back Propagation Network Structure Showing the Input-Output Relationship

5.3.1.3 Input Layer

In Figure 5.3, a neuron in the input layer is shown. L_i represents the input of the data to the layer of input. Each input layer's output neuron is equal to the normalized input. Each output of the input layer neuron O_i is connected to the hidden layer inputs which is described in Figure 5.4. Equation 5.6 represents the relationship of the input data and the output of input layer where I_i represents the input of the input layer and Oi represents the output of the input layer.

$$O_i = I_i$$

(5.6)

5.3.1.4 Hidden Layer

A neuron in the hidden layer is shown in Figure 5.4. The signal given to a neuron in the layer of hidden I_h is equal to the sum of the total outputs of the neurons in input layer O_i multiplied by the associated connection weights W_{hi}, as in equation 5.7.

$$\left(I_h = \sum_i W_{hi} O_i \right) \tag{5.7}$$

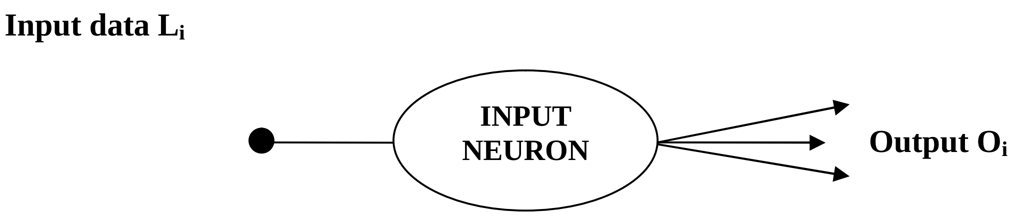


Figure 5.3: A Neuron of Input Layer

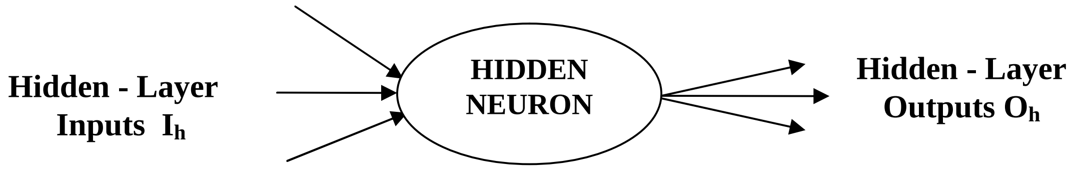


Figure 5.4: A Neuron in the Hidden Layer

The output of the hidden neurons O_h is found using the sigmoid function and is described in equation 5.8.

$$O_h = \frac{1}{1+\exp(-I_h)} \tag{5.8}$$

5.3.1.5 Output Layer

A neuron of the output layer is shown in Figure 5.5. The signal that is given to a neuron in the output layer I_j is equal to the total sum of all the outputs of the hidden layer neurons O_h who are multiplied by the their connected weights W_{jh} as in equation 5.9.

$$I_j = \sum_h W_{jh} O_h \tag{5.9}$$

The output of each output neuron O_j is found using the sigmoid function in a similar way as in the hidden layer and is shown in equation 5.10.

$$O_j = \frac{1}{1+\exp(-I_j)} \tag{5.10}$$

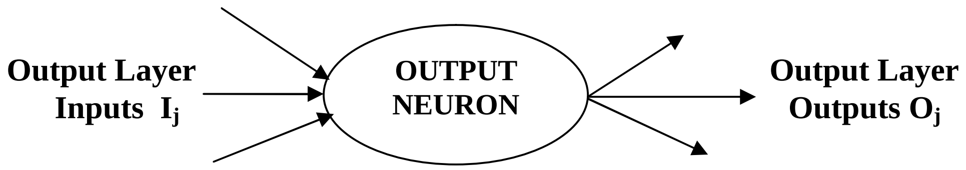


Figure 5.5: A Neuron of the Output Layer

The given set of equations and calculations that has been described earlier in the feed forward processing can be applied during the learning process and the testing process as well as running process.

5.3.2 Error Back Propagation Calculations

The back propagation error calculations are carried out during the training of the back propagation neural network which includes the error and some other important parameters with weights adjustment.

5.3.2.1 Signal Error

During the training of the neural network, the feed forward calculations are joint with back-propagation of the error and weight adjustments that represent the neural network's training process. Main process of training a neural network is the calcualtion of the network error. Rumelhart and McClelland defined an error that depends on the difference between the output value and the target value T_j. The error term represents a measure of how well a network is simulating the input output relationship of the given sets of data. The definitions for the error are presented in equation 5.11. The subscript p denotes that the value is for a given pattern.

$$\left(E_p = \sum_{j=1}^{n_j} (T_{pj} - O_{pj})^2 \right) \tag{5.11}$$

Using the training process, the goal is to reduce the maximum possible the mean squared error over all training patterns. In equation 5.10, the output of a neuron in the output layer that is a function of its input, or $O_j = f(I_j)$ can be seen. The derivative of this function, $f'(I_j)$ is very important in the error back-propagation algorithm. For output layers, a quantity which is known as the error signal is known by $_j$ and can be given in equation 5.12.

$$\left(\Delta_j = f'(I_j)(T_j - O_j) \right) \tag{5.12}$$

The value of error is propagated back and the appropriate weight adjustments to the back propagation neural network are performed. This process is carried out by finding the accumulative sum of the 's for each neuron in the whole training set, and back-propagating the error based on the grand total . This process is called batch (epoch) training.

5.3.2.2 Adjustment of Weights

Initially, each weight has to be given an initial random value. Whenever random initialization is done, re-adjustment of the Weights is applied in each stage, starting at the end of the process of feed forward, and going backward to the inputs in the hidden layer.

5.3.2.2.1 Output-Layer Weights Update

The weights W_{jh} are updated using equation 5.13 that feeds the output layer. This includes the bias weights at the output layer of neurons. To find the new weights of the network W_{jh}(new), the learning rate η should be multiplied with the change in output and output of the hidden layer and should be added with the old weights of the hidden layer W_{jh}(old).

$$W_{jh}(new) = W_{jh}(old) + \eta \Delta_j O_h \tag{5.13}$$

In order to avoid the risk of the back propagation neural network getting caught in a local minima, the momentum α can be added to the equation in 5.14.

$$W_{jh}(new) = W_{jh}(old) + \eta \Delta_j O_h + \alpha \left[\Delta W_{jk}(old)\right] \tag{5.14}$$

where W_{jh} (old) stands for the change in previous weight and α represents the momentum rate.

5.2.2.2.1 Hidden-Layer Weights Update

For the hidden layer, a definition for the error term is not simple to figure out. However, Rumelhart and McClelland (Rumelhart & McClelland, 1986) gave a definition that describes the error term for a hidden neuron in equation 5.15.

$$\Delta_h = f'(I_h) \sum_{j=0}^{n_j} W_{jh} \Delta_j \tag{5.15}$$

The adjustments of the weights for the connections of the hidden layers from the input layers are now processed in a similar way to those feeding the output layer and these adjustments are found using equation

$$W_{hi}(new) = W_{hi}(old) + \eta \Delta_h O_i + \alpha \left[\Delta W_{hi}(old)\right] \tag{5.16}$$

The bias weights at the hidden layer neurons are updated again using equation 5.16.

5.4 Neural Network for estimation of Relative Data Redundancy

Neural Networks applied in compression of images using discrete cosine transform as well as discrete wavelet transform, within the recent years. The use of DCT and artificial neural networks has also been investigated in search for optimal compression methods. In (Dokur, 2008) DCT and neural networks based medical image compression has been proposed & applied to MR and CT medical data images. In (Meyer-Base et al., 2005) a topology-preserving neural networks was developed in the field of medical image compression using a neural gas network. In (Ashraf & Akbar, 2007) adaptive architecture neural networks were implemented for medical image compression. In (Liying & Khashayar, 2005) a neural network classifier was used with a combination of different image compression techniques for various applications. In (Soliman & Omari, 2006) another model of neural network named as direct classification was applied to compress image data. In (Vilovic, 2006) a method with direct solution was applied to image compression using neural networks was previously suggested. More works using neural networks for image compression applications emerged lately, such as those in (Ashraf & Akbar , 2005) a neural network based image quantizer was used to get a high Relative Data Redundancy while maintaining good quality images. In (Northan & Dony, 2006) a Multi-Resolution Neural Network (MRNN) with a filter bank has been used as a transform for coding. In (Veisi & Jamzad, 2007) an image compression algorithm which is based on image blocks complexity measure methods by using a neural network was proposed. In (Osowski et al., 2006) image compression using discrete wavelet transform and a back propagation neural network was suggested. In (Khashman & Dimililer, 2007) a neural network based Discrete Cosine Transform compression system that finds the optimum compression ratio for a variety of images was also suggested.

Unlike the DCT, the wavelet transforms are not Fourier based transforms and therefore discontinuities in image data can be handled with better results using wavelet transform (Khashman & Dimililer, 2008). The Wavelets are the mathematical tools for hierarchically decomposing the functions. There is a preference to use wavelet transforms in image compression because higher Relative Data Redundancy and higher PSNR values can be obtained (Talukder & Harada, 2007). Haar wavelet transform (HWT) is one of the wavelet methods and it is applied in compressing the digital images. Previously, Haar image compression include an application that is applied to adaptive data hiding for the images segmenting the original image into 8x8 sub-blocks and reconstructing the images after compression with better quality (Lai & Chang, 2006).

The use of compression methods in general with wavelet compression in particular, using medical images has been previously investigated (Li et al., 2007; Shenbaga & Vidhya, 2007; Chen & Tseng, 2007; Karlik, 2006). More recently, a neural network model was used to obtain the compression ratio using Haar wavelet image compression that was applied to a variety of digital images (Khashman & Dimililer, 2008). The development and implementation of the proposed system using medical x-ray images with image compression applied uses 90 x-ray images from our medical image database (FGM, 2007) , contains x-ray images of broken, fractured, dislocated and healthy bones in different areas of human body. Initially, Discrete cosine transform, Biorthogonal wavelet transform and Daubechies wavelet transform were applied to x-ray images using nine Relative Data Redundancy (10%, 20%, ... , 90%) for each method as shown in examples in Figure 5.6 for Discrete cosine transform, Figure 5.7 for Daubechies Wavelet transform and Figure 5.8 for Biorthogonal Wavelet transform.

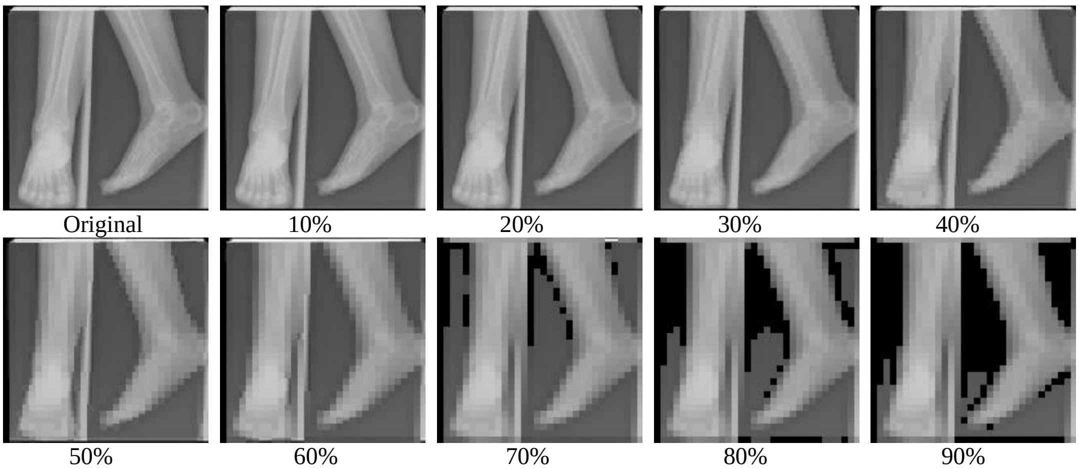


Figure 5.6: An Original Image and Discrete Cosine Transform based Compression with Nine Ratios

The Compression Ratio for DCT, DWT, BWT in the x-ray images were found using the compression criteria based on visual inspection and empirical analysis. Visual inspection includes 10 experts who were asked to check and observe the edge continuity and smoothness of certain areas within the set of reconstructed images (Khashman & Dimililer, 2005).
The database is organized as three sets that are the set of training used to train the database, a testing set with pre-determined compression raitos and testing set 2 with compression values not determined.

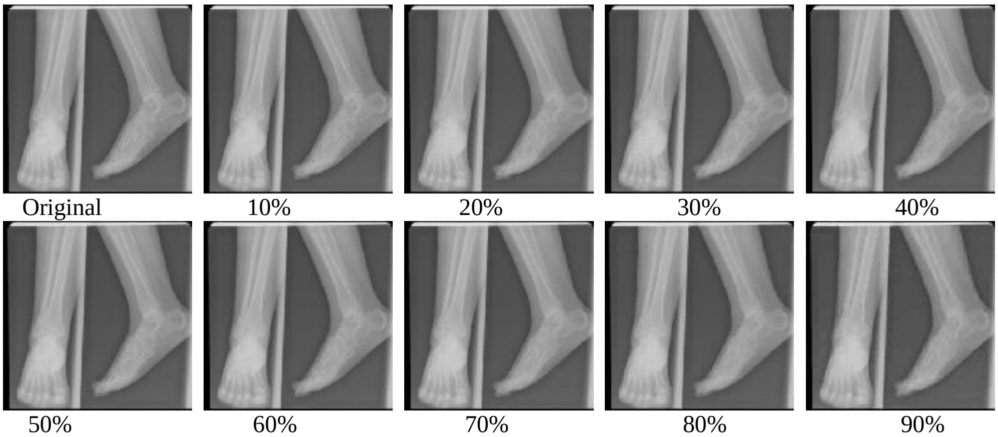


Figure 5.7: An Original Image and Daubechies Wavelet Transform based Compression with Nine Ratios

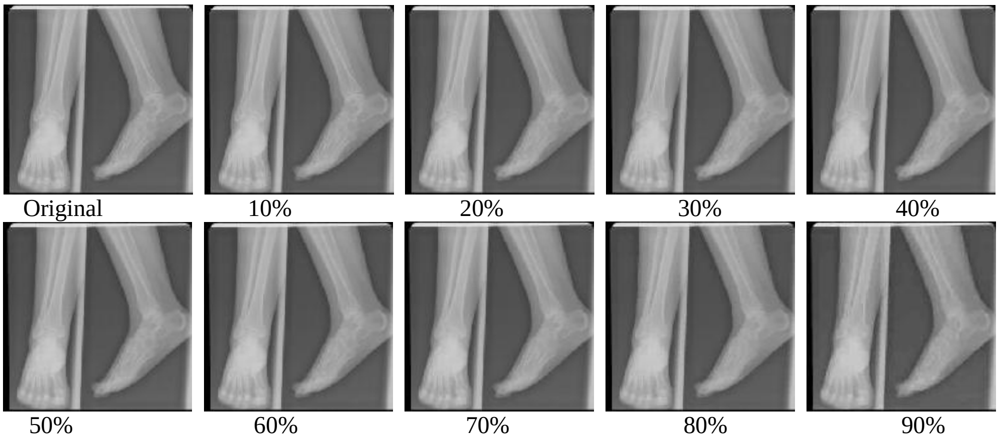


Figure 5.8: An Original Image and Biorthogonal Wavelet Transform based Compression with Nine Ratios

Table 5.2 Average PSNR values of Targets in BPNN

Target	**Train Images**			**Method**	**RD**	**Avg. PSNR**
1	**Image72**	**Image77**	**Image85**	**DCT**	**10**	**72,56**
2	Image32	Image4	Image57	DCT	20	46,72
3	Image10	Image22	Image58	DCT	30	40,20
4	Image65	Image71	Image83	DCT	40	38,53
5	**Image24**	**Image63**	**Image68**	**Daub**	**50**	**66,57**
6	Image45	Image59	Image81	Daub	60	64,43
7	Image14	Image36	Image73	Daub	70	55,16
8	Image11	Image28	Image3	Daub	80	50,39
9	Image38	Image42	Image74	Daub	90	44,72
10	**Image26**	**Image89**	**Image9**	**BIOR**	**50**	**58,40**
11	Image19	Image76	Image87	BIOR	60	64,80
12	Image21	Image35	Image51	BIOR	70	57,13
13	Image1	Image40	Image54	BIOR	80	54,14
14	Image18	Image7	Image86	BIOR	90	43,49

- Training Set: 42 images with their *known* ratios of compression used for training the neural network within the suggested compression system. There are 14 input layers used in output layer of the back propagation neural network consists of the Relative Data Redundancy that are chosen by the experts. A minimum Mean Opinion Score value of 3.75 has been used in the decision of number of input layers to the neural networks. Figure 5.9 represents the examples of training set images.

Table 5.2 represents the training image set from the database with their average PSNR values of each target.

Figure 5.9: Training Set Examples

- Testing Set 1 : 30 images with their *known* ratios of compression used to test and validate the efficiency of the trained back propagation neural network. Figure 5.10 represents the examples of testing set 1 images.

Figure 5.10: Testing Set 1 Examples

- Testing Set 2 : 18 images with un*known* compression are used to test and validate the efficiency of trained back propagation neural network. Figure 5.11 represents the examples of testing set 2 images.

Figure 5.11: Testing Set 2 Examples

Samples of original x-ray images with their compressed versions using the pre-determined image compression method and their Relative Data Redundancy while training the back propagation neural network is shown in Figure 5.12.

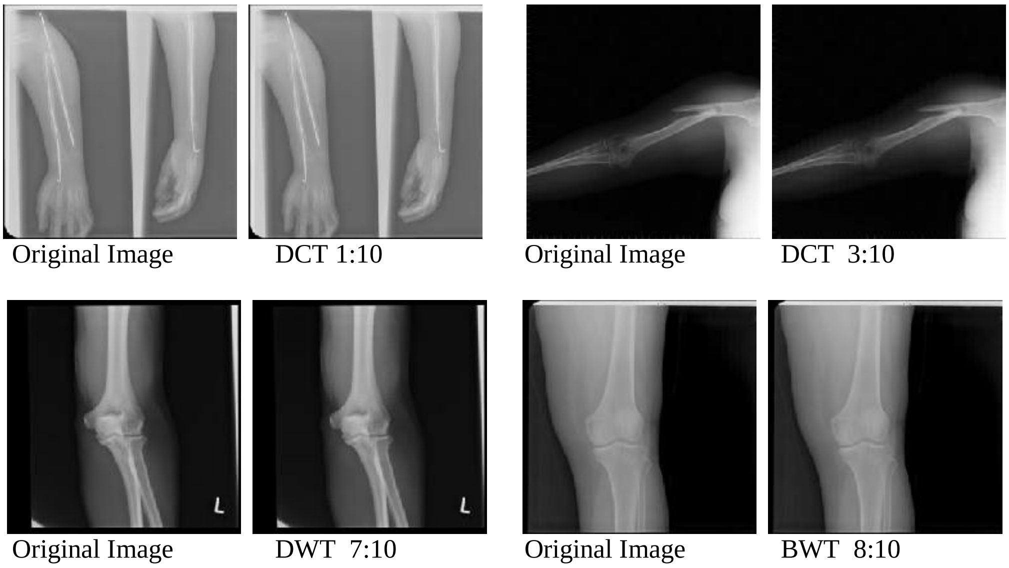


Figure 5.12: Training Set Image Samples with their Ideal Compression Method and Relative Data Redundancy

The suggested x-ray image compression system uses a supervised back propagation neural network learning algorithm, due to its simplicity in implementation, and the availability of sufficient “input/target” database for training the supervised learner. The neural network relates the x-ray image intensity that are pixel values, to the image compression method and its Relative Data Redundancy that are trained using images with predetermined Relative Data Redundancy and methods. After the training process, the neural network would choose the method and the Relative Data Redundancy for an x-ray image after presenting it to the BPNN by using the intensities.

Simple image resizing was used to resize the original images of size (256x256) pixels into (64x64) pixels, deleting the next three pixels while keeping the first pixel value. Further reduction to the size of the images was attempted in order to reduce the training time by decreasing the number of input layer neurons, however, a meaningful back propagation neural network training could is not achieved. So, the initial reduced size of 64x64 pixels has been preferred to be used. The size of the input images affects the choice of the number of neurons in the back propagation neural network’s input layer, which has three layers; input, hidden and output layers. Using one-pixel-per-node approach, 4096 neurons in the neural network’s input layer, with hidden layer of 46 neurons, assures meaningful training while keeping the time cost to minimum, and its output layer has 14 neurons representing the output classification of the image compression method and the Relative Data Redundancy as follows: output neurons {1-4} represent Discrete Cosine Transform based compression at ratios (1:10 – 4:10), output neurons {5-9} represent Daubechies Wavelet Transform based compression at ratios (5:10 - 9:10) whereas, output neurons {10-14} represent Biorthogonal Wavelet Transform based compression at ratios (5:10 – 9:10).

During the learning process, with various experiments, the learning coefficient and the momentum rate were adjusted in order to achieve the required minimum rms error value of 0.003. Figure 5.13 shows the topology of the back propagation neural network, within the suggested image compression system.

5.5 Results

The evaluation of the training and testing results were performed by using two kinds of measurements that are the recognition rate and the accuracy rate. The decision of the ideal compression method is made by the neural network, the recognition rate is defined in equation 5.17 as follows:

$$RR_{OC} = \left(\frac{I_{OC}}{I_T}\right) * 100 \tag{5.17}$$

where RR_{OC} represents the recognition rate for the back propagation neural network based intelligent x-ray compression system, I_{OC} is the amount of compressed x-ray images, and I_T is the total amount of x-ray images within the database set.

The accuracy rate RA_{OC} of the back propagation neural network output results are defined in equation 5.18 as follows:

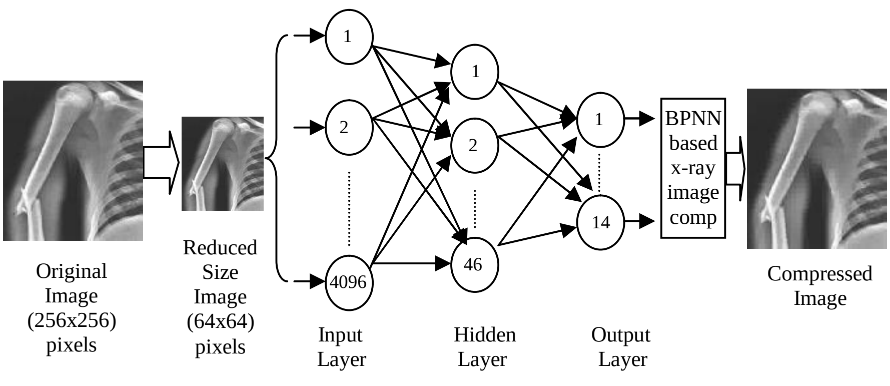


Figure 5.13: X-ray Image Compression System using BPNN

$$RA_{OC} = 100 + \left(1 - \frac{\left(\left(\left| S_p - S_i \right|\right) * 10\right) + S_T}{S_T} \right) \tag{5.18}$$

where S_P presents the pre-determined or expected Relative Data Redundancy in percentage, S_i presents the Relative Data Redundancy as determined by the trained back propagation neural network in percentage and S_T represents the total number of image Relative Data Redundancy used for all of the compression methods.

The Optimum Compression Deviation (OCD) is another term for the evaluation in our work. *OCD* is the error between the pre-set or expected optimum Relative Data Redundancy S_P and the optimum Relative Data Redundancy S_i as determined by the trained neural network, and is defined in equation 5.19 as follows:

$$OCD = \frac{\left(\left| S_p - S_i \right|\right)}{10} \tag{5.19}$$

The OCD is used to specify the accuracy level of the compression system, and regarding its value the recognition rates vary. Table 5.3 presents the three main values of OCD and their related accuracy and recognition rates. The evaluation of the system implementation outcomes implies the use of (*OCD = 2*) because it provides a minimum accuracy rate of 77.8% which is considered highly sufficient for this application.

Table 5.3. Accuracy and Recognition Rates According to OCD.

OCD	Accuracy Rate (RA_{OC})	Recognition Rate (RR_{OC})
0	100 %	17/30 (56%)
1	88.9 %	25/30 (83%)
2	77.8 %	26/30 (87%)
3	Below 77.8 %	30/30 (100%)

Table 5.4. Neural Network Final Training Parameters

Input nodes	4096
Hidden nodes	46
Output nodes	14
Learning rate	0.005
Momentum rate	0.5
Error	0.003
Iterations	2740
Training time (seconds)	498.89
Run time (seconds)	0.003

The neural network has learnt and converged after 2740 epochs, and the time of learning was 498.89 seconds, which can be considered fast. The running time for the generalized neural networks test after the training and using one forward propagation was 0.003 seconds. These results were obtained by using a 2.8 GHz PC with 8 GB of RAM, Windows 7 Ultimate 64-bit OS and Matlab R2010b software. Table 5.4 lists the final parameters of the trained neural network, whereas Figure 5.14 shows the mean squared error curve of the neural network during the process of learning. The trained neural network has recognized correctly the method and Relative Data Redundancy for all 42 training set images as would be expected, thus yielding 100% recognition of the training set. Testing the trained neural network using the 30 images from Test Set 1 that were not fed to the network before gave a recognition rate of 56%, where 17 out of the 30 images with known ideal method with 100% accuracy and optimum Relative Data Redundancy were correctly classified, 83% recognition rate with 25 out of 30 images with an accuracy level of 88.9%, 87% recognition rate with 27 out of 30 images and 100% with an accuracy level of 77.8% and 30 images out of 30 using an accuracy level of below 77.8%.

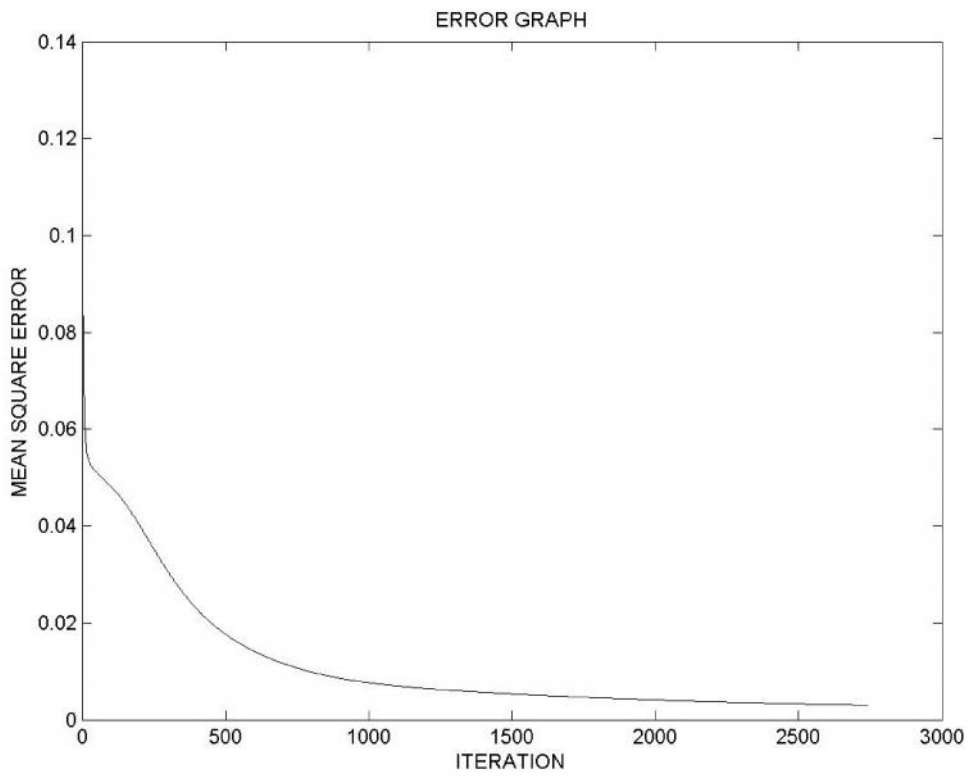


Figure 5.14: Neural Network's Learning Curve

The results of x-ray compression using Back Propagation Neural Networks are demonstrated in Figure 5.15 which shows examples of the optimally compressed x-ray images as determined by the trained neural network.

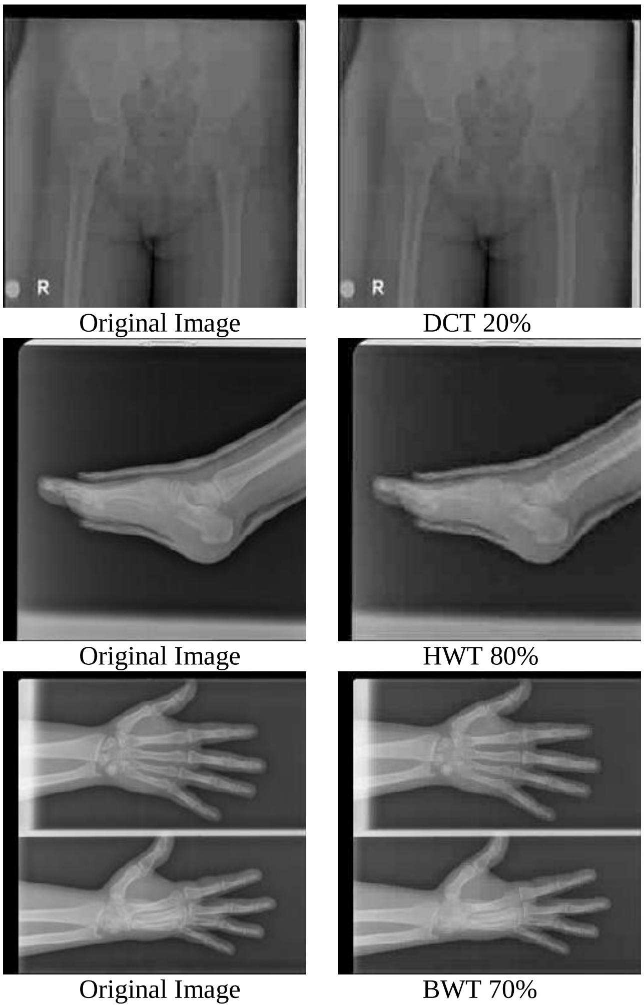


Figure 5.15: Testing Set 2 Image Compression using the developed Back Propagation Neural Network System

5.6 Summary

A back propagation method to medical x-ray image compression using a neural network based discrete cosine transform and daubechies wavelet transform and biorthogonal wavelet transform image compression is proposed. The method uses DCT, DWT and BWT based compression with three Relative Data Redundancy each and a back-propagation neural network that learns to establish the relationship and connect the grey x-ray image intensity (pixel values) with the compression method and a single optimum Relative Data Redundancy to be used with the selected method.

The implementation of the proposed system chooses DCT, DWT and BWT based image compression methods. The aim of an optimum Relative Data Redundancy is to join high compression with good quality compressed x-ray images, thus making the storage and transmission of these medical images more efficient, easy, and lossless.

The proposed system was studied, developed and implemented using 90 x-ray images of fractured, dislocated, broken, and healthy bones in different parts of the body. The neural network within the intelligent compression system learnt to associate the 30 training images with their predetermined ideal compression method and optimum Relative Data Redundancy within 498.89 seconds. Once trained, the neural network could recognize the ideal method and optimum ratio for an x-ray image within 0.003 seconds.

In this work, a minimum accuracy level of 77.8% was considered as acceptable. Using this accuracy level, the neural network yielded 87% correct recognition rate of ideal method and optimum ratios (Dimililer, 2013). The successful implementation of our proposed intelligent system was seen throughout the high recognition rates and the minimal training and test time when running the trained neural network.

CHAPTER 6

SETTING PARAMETERS OF COMPRESSION FOR MEDICAL IMAGE PROCESSING

6.1 Overview

Linear Regression Analysis has been applied in Chapter 4 in order to find the Relative Data Redundancy that can be applied to an image, using the characteristics of the images and Back Propagation Neural Networks has been applied in Chapter 5 in order to find Relative Data Redundancy with compression method of an image. In this chapter, comparative analysis of the approaches for medical image compression will be discussed and the algorithms for setting optimal compression method with Relative Data Redundancy will be given using the known set with the pre-determined Relative Data Redundancy set and unknown set as the undetermined Relative Data Redundancy set.

6.2 Algorithm for setting Optimal Compression Method and Relative Data Redundancy

Linear Regression Analysis uses the characteristics of the images that has been considered to find the Relative Data Redundancy of an image. The characteristics of the images such as the intensity of the images, variance of intensity of the images, entropy value of the images and contrast weighted entropy of the images has been considered.

Back Propagation Neural Network as a supervised algorithm based on the empirical analysis has been applied to find a relationship between the image pixels and their optimal compression methods with the Relative Data Redundancy. 10 experts has been considered in the empirical analysis part to find the optimal compression methods with Relative Data Redundancy.

Figure 6.1 represents The Compression System using Back Propagation Neural Networks and Linear Regression Analysis in order to find the Relative Data Redundancy and compression method.

The suggested image compression consists of two parts. One part is the back propagation neural networks part and the second one is linear regression part. Using the back propagation neural networks, the compression method has been chosen upon giving the image set to the system to be trained and tested by two another set of images.

The BPNN has been used to find a relationship between the compression method with the Relative Data Redundancy and the experts decisions. Regression Analysis has been used considering the characteristics of the images as X. So, The Back Propagation Neural Networks has been chosen as optimum in the decision of the Compression Method and Linear Regression Analysis has been found optimum in the decision of the Relative Data Redundancy.

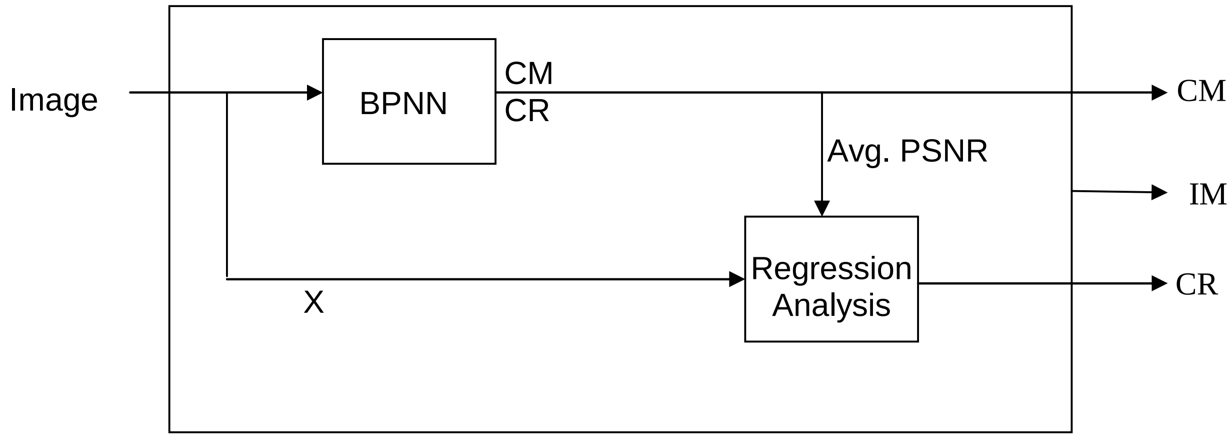


Figure. 6.1: Compression System

A set of images has been randomly chosen from the database considering the idea of the experts and they have been used as training set for the Back Propagation Neural Networks. Another set of images has been prepared to test the efficiency of the back propagation neural network system as well as linear regression analysis. 42 images has been used in the training set, 30 images has been used in the known set and 18 images has been used in the unknown set.

Linear regression analysis has been used in the decision of Relative Data Redundancy of an image considering the characteristics of the images. The average psnr values of the trained set with the characteristics of the images such as intensity with the variance of intensity and contrast weighted entropy of the images are used in the calculation of the Relative Data Redundancy. The average PSNR values of the training set in Table 5.2 using each method are applied in the equations that were created in Chapter 4. Equation 6.1 represents the calculation of Relative Data Redundancy CR using DCT Compression the results of PSNR values with the characteristics of the images, Equation 6.2 represents the calculation of Relative Data Redundancy CR using Daubechies Compression the results of PSNR values with the characteristics of the images, Equation 6.3 represents the calculation of Relative Data Redundancy CR using Biorthogonal Compression the results of PSNR values with the characteristics of the images.

$$Log(Avg.PSNR) = 2.073 - 0.325(LogCR) + 0.01(X_2) \quad (6.1)$$

$$Log(Avg.PSNR) = 3.077 - 0.737(LogCR) + 0.0001(X_2) \quad (6.2)$$

$$Log(Avg.PSNR) = 3.213 - 0.810(LogCR) + 0.001(X_2) \quad (6.3)$$

As a result, when an image is considered in the compression system, the back propagation neural networks gives a compression method for the suggested image and linear regression gives a Relative Data Redundancy suggested for the image.

6.3 Assessment of the established algorithm using known set

In pre-determined Relative Data Redundancy Set, the Relative Data Redundancy and the compression method for a set of images has been pre-determined using the experts. CR_{Pre} represents the pre-determined Relative Data Redundancy, CM_{Pre} represents pre-determined compression method.

BPNN represents the Back Propagation Neural Networks used to access the recognition features of the images applied in finding the Relative Data Redundancy with compression method. ΔCR_{BPNN} represents the absolute value of the change in Relative Data Redundancy using the pre-determined and LRA applied Relative Data Redundancy. Equation 6.4 represents the accuracy to find the Relative Data Redundancy using the change in Relative Data Redundancy using pre-determined Relative Data Redundancy and Back Propagation Neural Networks applied Relative Data Redundancy.

$$\Delta CR_{BPNN} = |CR_{Pre} - CR_{BPNN}| \quad (6.4)$$

Back Propagation Neural Networks has been applied in the decision of the Relative Data Redundancy with the compression method. Relative Data Redundancy (CR_{BPNN}) and Compression Methods (CM_{BPNN}) has been decided by the Back Propagation Neural Networks and the absolute value of the change in the Relative Data Redundancy using pre-determined and BPNN applied results has been found.

LRA represents Linear Regression Analysis applied Relative Data Redundancy and ΔCR_{LRA} represents the absolute value of the change in Relative Data Redundancy using the pre determined and LRA applied Relative Data Redundancy. Equation 6.5 represents the accuracy to find Relative Data Redundancy using the change in pre-determined Relative Data Redundancy and Linear Regression applied Relative Data Redundancy.

$$\Delta CR_{LRA} = |CR_{Pre} - CR_{LRA}| \quad (6.5)$$

The Relative Data Redundancy using Linear Regression Analysis has been found and the absolute value of the change in the Relative Data Redundancy using pre-determined and LRA applied results has been found.

The final decision of the assessment of the Relative Data Redundancy of suggested system using Linear Regression Analysis is denoted as CR_{Final} and the final decision of the assessment of the compression method of the suggested system using Back Propagation Neural Networks is denoted as $CM_{Final.}$

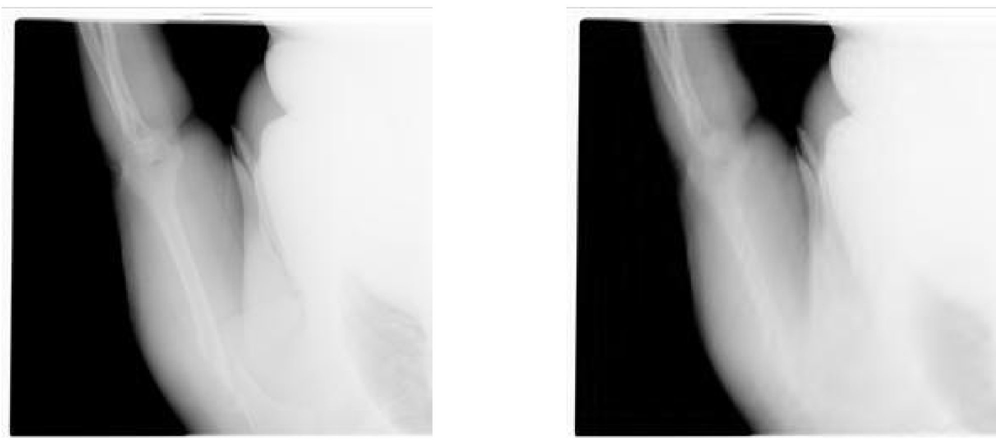
Table 6.1 Image Compression Results using Known Set

FILENAME	CR_{Pre}	CM_{Pre}	CR_{BPNN}	CM_{BPNN}	CR_{LRA} (%)	CR_{FINAL}	CM_{FINAL}
	Experts		**BPNN**		**LRA**	**FINAL DECISION**	
Image2	90	DAUB	90	DAUB	86,3347494	**86,3347494**	**DAUB**
Image5	30	DCT	30	DCT	28,55443736	**28,55443736**	**DCT**
Image6	90	BIOR	90	BIOR	88,76816302	**88,76816302**	**BIOR**
Image8	10	DCT	10	DCT	4,708002461	**4,708002461**	**DCT**
Image20	60	BIOR	60	BIOR	62,33843757	**62,33843757**	**BIOR**
Image12	50	DAUB	50	DAUB	50,32074755	**50,32074755**	**DAUB**
Image13	10	DCT	10	DCT	4,640304625	**4,640304625**	**DCT**
Image15	30	DCT	30	DCT	28,28469796	**28,28469796**	**DCT**
Image16	60	DAUB	60	DAUB	64,98144994	**64,98144994**	**DAUB**
Image17	50	DAUB	50	BIOR	50,31609567	**50,31609567**	**BIOR**
Image23	20	DCT	20	DCT	18,09204842	**18,09204842**	**DCT**
Image25	70	DAUB	70	DAUB	73,38675433	**73,38675433**	**DAUB**
Image27	50	DAUB	50	DAUB	52,57190267	**52,57190267**	**DAUB**
Image29	70	BIOR	70	BIOR	67,56980402	**67,56980402**	**BIOR**
Image30	70	DAUB	70	DAUB	73,3612944	**73,3612944**	**DAUB**
Image31	90	DAUB	90	DAUB	86,26098767	**86,26098767**	**DAUB**
Image33	30	DCT	40	DCT	28,22372874	**28,22372874**	**DCT**
Image34	90	BIOR	80	DAUB	88,84926776	**88,84926776**	**DAUB**
Image37	60	BIOR	70	BIOR	63,38963152	**63,38963152**	**BIOR**
Image39	90	DAUB	80	DAUB	86,30677922	**86,30677922**	**DAUB**
Image41	70	DAUB	80	DAUB	66,59489808	**66,59489808**	**DAUB**
Image43	20	DCT	30	DCT	17,87161699	**17,87161699**	**DCT**
Image44	50	DAUB	60	DAUB	52,57504995	**52,57504995**	**DAUB**
Image46	20	DCT	20	DCT	18,26127434	**18,26127434**	**DCT**
Image47	60	DAUB	90	DAUB	64,88617814	**64,88617814**	**DAUB**
Image48	30	DCT	20	DCT	28,32751444	**28,32751444**	**DCT**
Image50	60	BIOR	90	BIOR	63,63552558	**63,63552558**	**BIOR**
Image52	70	BIOR	90	BIOR	67,49332408	**67,49332408**	**BIOR**
Image55	50	DAUB	80	DAUB	52,55514333	**52,55514333**	**DAUB**
Image56	90	DAUB	50	DAUB	86,27444433	**86,27444433**	**DAUB**

Table 6.1 represents the results of image compression system using the 30 images in the known set, which includes the algorithm applied in finding the optimal compression method with the Relative Data Redundancy. Pre-determined Relative Data Redundancy and Pre-determined Compression Methods has been gathered using the mean of 10 experts in empirical analysis. Considering the compression method, It can be seen that Back Propagation Neural Networks is applicable in finding the suitable compression method for the system where 28 images out of 30 images with a ratio of 93.3% has been correctly found, and considering the Relative Data Redundancy, it can be seen that the linear regression based statistical presentation of the Relative Data Redundancy gives better performance than back propagation based image Relative Data Redundancy.

In order to arrange the parameters of compression using the medical images, 3 different approaches are combined and analysed. One of these approaches is the experts approach, second approach is back propagation neural networks and the third one is linear regression analysis. In order to improve the efficiency of the medical image compression system, the three approaches are considered and checked if there is a difference between each other. So the differences between the experts and back propagation neural networks as well as the differences between the experts and linear regression analysis are considered in order to apply hypothesis testing for the 30 images using SPSS.

Figure 6.2 represents the original image and the 86.33% Daubechies Wavelet compressed image using the established algorithm with the known set of image 2.



ORIGINAL IMAGE 86.33% Daubechies

Figure 6.2: An Original Image with 86.33% Relative Data Redundancy and Daubechies Wavelet Compression Method with a MOS value of 4.6 considering 90% of DWT Relative Data Redundancy

Figure 6.3 represents the original image and the 28.33% Discrete Cosine compressed image using the established algorithm with the known set of image 48.

ORIGINAL IMAGE

28.33% Discrete Cosine

Figure 6.3: An Original Image with 28.33% Relative Data Redundancy and Discrete Cosine Compression Method with a MOS value of 4.6 considering 30% of DCT Relative Data Redundancy

Figure 6.4 represents the original image and the 63.39% Biorthogonal Wavelet compressed image using the established algorithm with the known set of image 37.

ORIGINAL IMAGE

63.39 % Biorthogonal

Figure 6.4: An Original Image with 63.39% Relative Data Redundancy and Biorthogonal Wavelet Compression Method with a MOS value of 4.8 considering 60% of BWT Relative Data Redundancy

Levene's test is an inferential statistic used to assess the equality of two variances that are independent from each other (Navidi, 2012). The results of the differences are given to SPSS as one row and two factors are given in the table. The aim of the SPSS applied is to compare the two independent variables using compare means to get the results to be compared with each other by analysing the results to be represented. So Independent sample T test model has been applied using compare means in SPSS to get the following results of tables. Table 6.2 represents group statistics and Table 6.3 represents independent samples test results using Known set.

When we got the compare means results using SPSS, the number of images, the mean value of the results, standart deviation as well as the results of two independent samples T-test has been realized. It has been realized that the two kinds of results with the same amount of images are different from each other considering the 95% of significance.

When the two independent samples test are considered, it has been regarded that the Significance value for the given samples is 0.00 which is less than 0.05 means that the two population results which are ΔCR_{BPNN} and ΔCR_{LRA} are different from each other. And when the Mean values are considered, it can be seen that ΔCR_{LRA} is closer to zero than ΔCR_{BPNN}. Therefore ΔCR_{LRA} gives more accurate results.

Table 6.2 Group Statistics Results using Known Set

FACTOR		N	Mean	Std. Deviation	Std. Error Mean
CR	**Δ CR_{BPNN}**	30	7,3333	10,80655	1,97300
	Δ CR_{LRA}	30	2,7634	1,35171	0,24679

Table 6.3 Independent Samples Test Results using Known Set

	T-test for Equality of Means								
	F	**Sig.**	**T**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
Equal variances assumed	33,93	,000	2,298	58	,025	4,56993	1,98837	,58977	8,55009
Equal variances not assumed			2,298	29,907	,029	4,56993	1,98837	,50860	8,63125

Therefore we got a conclusion that if the significance is greater than 0.05, it means that the two populations are same and if the significance is less than 0.05, it means that we dont have confidence to say that they are equal. So they are different

In conclusion, after getting Relative Data Redundancy considering the experts, BPNN and LRA, it can be mentioned that these results are different from each other but they are not significant when equal variances for Relative Data Redundancy are not assumed. Therefore, the decisions of experts considering the Relative Data Redundancy as well as compression method are compared and analyzed with the results of BPNN and LRA.

6.4 Assessment of the established algorithm using unknown set

The unknown set represents the 18 images that are decided by the system. The established algorithm has been assessed in Unknown set which is an unknown set of images that the Relative Data Redundancy and compression methods were not pre-determined in order to see the efficiency of the established algorithm. Table 6.4 represents The Results of Image Compression applied in Unknown Set to find the compression method with the Relative Data Redundancy.

Table 6.4 The Results of Image Compression using Unknown Set

FILENAME	BPNN		LRA	FINAL DECISION	
	CR_{BPNN}	CM_{BPNN}	CR_{LRA} (%)	CR_{FINAL}	CM_{FINAL}
Image49	90	DAUB	52,48	**52,48**	**DAUB**
Image53	70	BIOR	56,44	**56,44**	**BIOR**
Image60	30	DCT	4,65	**4,65**	**DCT**
Image61	70	DAUB	64,94	**64,94**	**DAUB**
Image62	90	BIOR	88,9	**88,9**	**BIOR**
Image64	70	DAUB	54,14	**54,14**	**BIOR**
Image66	80	DAUB	73,39	**73,39**	**DAUB**
Image67	30	DCT	32,44	**32,44**	**DCT**
Image69	80	BIOR	64,91	**64,91**	**DAUB**
Image70	70	BIOR	54,18	**54,18**	**BIOR**
Image75	90	DAUB	73,39	**73,39**	**DAUB**
Image78	90	DCT	4,67	**4,67**	**DCT**
Image79	80	BIOR	67,68	**67,68**	**BIOR**
Image80	60	DAUB	50,31	**50,31**	**DAUB**
Image82	80	BIOR	68,15	**68,15**	**BIOR**
Image84	60	BIOR	54,15	**54,15**	**BIOR**
Image88	30	DCT	32,5	**32,5**	**DCT**
Image90	90	DAUB	86,32	**86,32**	**DAUB**

Table 6.5 Group Statistics Results using Unknown Set

FACTOR		N	Mean	Std. Deviation	Std. Error Mean
CR	**BPNN**	18	70.0000	20.86370	4.91762
	LRA	18	54.6467	23.57840	5.55748

Table 6.6 Independent Samples Test Results using Unknown Set

	T-test for Equality of Means								
	F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
Equal variances assumed	0.07	,782	2.069	34	,046	15.35333	7.42082	0.27241	30.43426
Equal variances not assumed			2.069	33.504	,046	15.35333	7.42082	0.26417	30.43426

When the two independent samples test are considered, it has been regarded that the Significance value for the given samples is 0.782 which is more than 0.05 meaning that the results of two populations which are BPNN and LRA have no difference between each other. Therefore we got a conclusion that the results of two groups are not different from each other. In order to decide whether this difference is valuable or not, Sig. 2-tailed values have been checked whether if the value of equal variances not assumed is less than 0.05. The value of using the equal variances assumed, 0.046 has been proved that the differences of the two groups are not significant with each other and they come from different populations.

The results of BPNN and LRA using the unknown set of images proved that the suggested algorithm is applicable in the decision of the Relative Data Redundancy and the compression method of an image.

In conclusion, after getting Relative Data Redundancy using BPNN and LRA, it can be mentioned that these statistics are different from each other. Therefore, the decisions of experts considering the Relative Data Redundancy as well as compression method are compared and analyzed with statistics of BPNN and LRA.

Figure 6.5 represents the original image and the 32.44% Discrete Cosine compressed image using the established algorithm with the Unknown set of image 67.

ORIGINAL IMAGE

32.44% DCT

Figure 6.5: An Original Image with 32.44% Relative Data Redundancy and Discrete Cosine Compression Method with a MOS value of 4.7 considering 30% of DCT Relative Data Redundancy

Figure 6.6 represents the original image and the 64.94% DWT compressed image using the established algorithm with the Unknown set of image 61.

ORIGINAL IMAGE

64.94% DWT

Figure 6.6: An Original Image with 64.94% Relative Data Redundancy and DWT Compression Method with a MOS value of 3.8 considering 60% of DWT Relative Data Redundancy

Figure 6.7 represents the original image and the 54.14% BWT compressed image using the established algorithm with the Unknown set of image 64.

ORIGINAL IMAGE 54.14% BWT

Figure 6.7: An Original Image with 54.14% Relative Data Redundancy and Biorthogonal Wavelet Compression Method with a MOS value of 4,3 considering 50% of BWT Relative Data Redundancy

6.5 Conclusions

The averages of the ΔLRA based image Relative Data Redundancy and ΔBPNN based image Relative Data Redundancy has been calculated and applied in SPSS to prove which method is suitable to find the optimum Relative Data Redundancy.

When the compression method has been considered, 28 images out of 30 images 93.3% were correctly recognized using Back Propagation Neural Networks.

Analysis of the quality of the two approach for the compression of images indicate that BPNN is very accurate (93.3%) to find the Compression Method of the known set where 28 images out of 30 are correct. But for finding the Relative Data Redundancy, LRA is needed which is significantly better and more accurate than BPNN. As a result, the following algorithm for the compressions of medical images can be established.

Step 1 : Assessment and setting of Compression Method by using Back Propagation Neural Networks

Step 2 : Evaluating Relative Data Redundancy by using Linear Regression Analysis

The suggested algorithm was used to assess the Unknown set which is an unknown set of medical images without pre-determined Relative Data Redundancy and compression methods.

The Figures 6.2 up to Figure 6.7 represents the Uncompressed Images with their Optimum Relative Data Redundancy and Optimal Compression Methods which has been found using the suggested algorithm.

The suggested image compression consists of two parts. One part is the back propagation neural networks part and the second one is linear regression part. Using the back propagation neural networks, the compression method has been chosen upon giving the image set to the system to be trained and tested by another set of images. Initially 30 images has been trained using the back propagation neural networks system and 28 of the images out of 30 has been correctly recognized by the back propagation neural networks giving a percentage of 93.3 % correctly recognized compression method.

Linear regression analysis has been used in the decision of Relative Data Redundancy of an image considering the characteristics of the images. The average psnr values of the trained set with the characteristics of the images such as intensity with the variance of intensity, entropy and contrast weighted entropy of the images are used in the calculation of the Relative Data Redundancy.

The suggested system in the decision of Relative Data Redundancy uses the characteristics of the images in the decision of the Relative Data Redundancy that can be applied to an image regarding the characteristics such as contrast, entropy, contrast weighted entropy, brightness of the images.

As a result, when an image is considered in the compression system, the back propagation neural networks gives a compression method for the suggested image and linear regression gives a Relative Data Redundancy suggested for the image.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

Image Compression is a branch of image processing that is applied on images such that excess amount of data within the image can be reduced or removed. With lossy compression, the concept is that, namely compressing the images by removing irrelevancy.

The aim of Lossy image compression is to reduce redundancy and irrelevance of the data present in an image in order to be able to transmit or store the data in an efficient form.

Three types of transform based image compression techniques has been applied to the image sets in order to compare the effect of Relative Data Redundancy and compression method with each other. Discrete Cosine Transform based image compression is a block-by-block compression technique that is applied on the images, Daubechies Wavelet Transform and Biorthogonal Wavelet Transform based image compression is a whole image applied compression technique and has been applied and discussed on Medical Images.

10 experts in this field has been considered using Mean Opinion Score are used in the decision of the Relative Data Redundancy for the suggested methods. A Relative Data Redundancy range of 10% up to 40% has been preferred by the experts decisions and a Relative Data Redundancy range of 50% up to 90% has been preferred by the experts decisions when considering a Mean Opinion Score value of 3.75 which is the 75% of the maximum score.

There are two criteria for assessment for the compressions of the images. Objective and Subjective. This thesis combine both criteria and establish the system to make automations for setting optimal Relative Data Redundancy and optimal compression method for the compressions of the medical images.

Objective assessment includes the characteristics of an image such as RMS and PSNR values that can be used to find the signal to noise ratio between an original image and a compressed image. PSNR depends on the Relative Data Redundancy, compression method and the characteristics of images such as contrast, brightness, variance of intensity in the decision of the Relative Data Redundancy with compression method which is based on the statistical analysis of the characteristics of the images.

The contrast of image, variance of intensity of the image and the entropy value of the image affects the compression method and Relative Data Redundancy. Contrast is the most important characteristic for medical images. But there is a need of some investigation of characteristic of an image for amplifying the dependence to Relative Data Redundancy.

To express the difference of medical images for the Relative Data Redundancy, a new characteristic of image has been established. This characteristic has been applied in this thesis and has been used to combine the contrast of the image with the entropy of the image and it is named as Contrast Weighted Entropy (CWE). CWE has more correlation related to PSNR value than classical entropy of the image. It has been shown the usefulness of this new established characteristic of the image. It has been proved that the suggested new characteristic of the image can be efficiently used in the decision of the Relative Data Redundancy.

Linear Regression Analysis has been employed for expressions which combine peak signal to noise ratio, Relative Data Redundancy, intensity, variance of intensity, contrast weighted entropy, compression Method. The Relative Data Redundancy of the images has been found using the characteristics of the images.

Subjective criteria uses the human as an expert in deciding the Relative Data Redundancy with optimal compression method. The experts are the doctors in their field to check the original images with the compressed set of images using three compression methods and find the Relative Data Redundancy with compression method that is based on the empirical analysis. The train set for neural networks has been established by the experts.

Back Propagation Neural Network was used as the recognition device and has been applied using the pixel values of the images in the decision of the optimum Relative Data Redundancy and optimal Relative Data Redundancy. The system has been trained using the train set and the system of back propagation neural network is capable of choosing the Relative Data Redundancy and compression method based on the training patterns that the empirical analysis based approach has been used to decide the Relative Data Redundancy and compression method.

Analysis of results combined the objective and subjective criteria using linear regression analysis in the assessment of Relative Data Redundancy and back propagation neural networks in the assessment of the compression method. LRA is accurate in Relative Data Redundancy as objective criteria and BPNN is accurate in compression method as subjective criteria. So the following algorithm has been established; First Assessment and setting of compression method by using back propagation neural networks and next, evaluating Relative Data Redundancy by linear regression analysis using the characteristics of the images. It was proved that the suggested system is applicable for medical images.

The contribution of the thesis includes;

- Design and development of the contrast-weighted entropy, where the energy of particular gray level in an image is weighted by the different configuration of neighbouring pixels.
- Creating an image compression system for medical x-ray images using a neural network classifier which uses subjective criteria for learning the non-linear relationship between characteristics of an x-ray image, most relevant compression method and optimal compression ratio.
- Establishing an analytical expression between characteristics of an image, a compression method and a compression ratio by using objective criteria for measuring the quality of a compressed image.
- Implementing of the image compression system with 3 image compression methods and 10 levels of the compression ratio.
- Investigating the degree of a correlation between objective and subjective quality measures of x-ray images by using the proposed image compression system.

7.2 Recommendations

The future work will include the investigation of some other characteristics of images which completely influence the Relative Data Redundancy and Compression Method of an image and the algorithm will be applied on the images using segmentation to compress the images segment by segment.

In order to increase the quality of the system, new image dataset will be used for re-processing of the LRA and BPNN in order to improve the effectiveness of the system.

REFERENCES

Journal

Singh, S., Kumar, V., Verma, H.K. (2007). Adaptive Threshold-based block classification in medical image compression for teleradiology. Computers in Biology and Medicine, *37*, 811-819.

Nadenau, M.J., Reichel, J., Kunt, M. (2003). Wavelet Based Color Image Compression: Exploiting the Contrast Sensitivity Function. *IEEE Transactions on Image Processing, 12(1)*, 58-70

Ratakonda, K., Ahuja, N. (2002). Lossless Image Compression with Multiscale Segmentation. *IEEE Transactions on Image Processing, 11(11)*, 1228-1237.

E. Peli (1990). Contrast in Complex Images. *Journal of the Optical Society of America A 7 (10)*, 2032–2040.

Boris Escalante-Ramírez, Mauricio Moreno-Gutiérrez, José L. Silván-Cárdenas (2002). Vector quantizer based on brightness maps for image compression with the polynomial transform. *SPIE Proceedings Vol. 4790,* 454-464.

M. I. Khalil (2010). Image Compression Using New Entropy Coder. *International Journal of Computer Theory and Engineering, 2(1),* 1793-8201.

H. Lee, S. Jeon, and Sanghoon Lee (2008). A Cross-Layer Approach for Maximizing Visual Entropy Using Closed-Loop Downlink MIMO. *EURASIP Journal on Advances in Signal Processing, 173,* 1-14.

Q. Huynh-Thu and M. Ghanbari (2008). Scope of validity of PSNR in image/video quality assessment. *Electronics Letters, 44,* 800–801.

S. Martucea (1994). Symmetric convolution and the discrete sine and cosine transform. *IEEE Transaction on Signal Processing, 42*, 1038-1051.

N. Ahmed, T. Natrajan, and K. R. Rao (1989). Discrete Cosine Transform. *IEEE Transactions on Computers, 23(1),* 90-93.

I. Daubechies (1988). Orthonormal bases of compactly supported wavelets. *Communication Pure Applied Mathematics., 41(7),* 909-996.

Stephane G. Mallat (1989). A theory for muhiresolution signal decomposition: the wavelet representation. *IEEE Trans, on Pattern Analysis and Machine Intelligence, 11(7),* 674-693.

Antonini M., Bariaud M., Mathieu P., and Daubechies I. (1992). Image coding using wavelet transform, *IEEE Trans. on Image Processing,1(2),* 205-220.

Ali N. Akansu (1993). Wavelets and Filter Banks, *IEEE Circuit and Devices, 10(6),* 14-18.

Loganathan R et. al. (2010). Medical Image Compression Using Biorthogonal Spline Wavelet With Different Decomposition. *International Journal on Computer Science and Engineering, 2(9),* 3003-3006.

Dokur, Z. (2008). A unified framework for image compression and segmentation by using an incremental neural network. *International Journal on Expert Systems with Applications 34(1)*, 611-619.

Meyer-Base, A., Jancke, K., Wismuller, A., Foo, S., Martientz, T. (2005). Medical image compression using topology-preserving neural networks. *Engineering Applications of Artificial Intelligence 18(4),* 383--392.

Liying, M., Khashayar, K. (2005). Adaptive Constructive Neural Networks using Hermite Polynomials for Image Compression. *Lecture Notes in Computer Science, 3497*, 713-722.

Soliman, H.S., Omari, M. (2006). A neural networks approach to image compression. *Journal of Applied Soft Computing 6(3),* 258-271.

Northan, B., Dony, R.D. (2006). Image Compression with a multi resolution neural network. *Canadian Journal of Electrical and Computer Engineering, 31(1),* 49-58.

Osowski, S., Waszczuk, R., Bojarczak, P. (2006). *Image compression using feed forward neural networks — Hierarchical approach Lecture Notes in Computer Science, 3497,* 1009-1015.

Khashman, A., Dimililer, K. (2008). Image Compression using Neural Networks and Haar Wavelet. *WSEAS Transactions on Signal Processing, 4(5),* 330-339.

Talukder, K.H., Harada, K. (2007). Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image. *International Journal of Applied Mathematics, 36(1),* 1-8.

Lai, B.L., Chang, L.W. (2006). Adaptive Data Hiding for Images Based on Haar Discrete Wavelet Transform. *Lecture Notes in Computer Science, 4319*, 1085-1093.

Chen, Y.T., Tseng, D.C. (2007). Wavelet-based medical image compression with adaptive prediction. *Computerized Medical Imaging and Graphics, 31(1),* 1-8.

Karlik, B. (2006). Medical Image Compression by Using Vector Quantization Neural Network Neural Network. *World International Journal on Neural and Mass - Parallel Computing and Information Systems, 16(4),* 341-348.

Dimililer K. (2012). Neural Network Implementation for X-Ray Image Compression. *Electronics World, 118(1911),* 26-29.

Pradhan C., Patra P. K., Maharatha S. (2010). Power Saving of DHT in Digital Image Photography. *International Journal of Computer Applications, 8(1),* 35-40.

Sonja Grgic, Mislav Grgic, Branka Zovko-Cihlar (2001). Performance Analysis of Image Compression Using Wavelets. *IEEE Transactions on Industrial Electronics, 48(3)*, 682-695.

Sadashivappa, G., Babu, A. (2008). Performance Analysis of Image Coding using Wavelets. *International Journal of Computer Science and Network Security, 8(10),* 144-151.

Singh, P., Singh, P., Sharma, R. K. (2011). JPEG Image Compression based on Biorthogonal, Coiflets and Daubechies Wavelet Families. *International Journal of Computer Applications , 13(1)*, 1-7.

Dimililer, K.. (2013). Backpropagation Neural Network Implementation for Medical Image Compression. *Journal of Applied Mathematics*, *2013*, 1-8.

Book

Camera & Imaging Products Association (2005). Guideline for Noting Digital Camera Specifications in Catalogs. Japan

David Salomon. (2006). Data Compression: The Complete Reference. USA, Springer-Verlag.

Sheikh, H.R. & Bovik A.C.yung-Sin Song. (2005). Handbook of Image and Video Processing. Information Theoretic Approaches to Image Quality Assessment. USA, Academic Press

Sheikh, H.R. & Bovik A.C.yung-Sin Song. (2008). Digital Image Processing. USA, Addison-Wesley Press.

Yun Q. Shi & Huifang Sun. (2000). Handbook of Image and Video Processing. Boca Raton FL, CRC Press

Anil K. Jain. (1989). Fundamentals of Digital Image Processing. Englewood Cliff, NJ, Prentice Hall Press

Simon Haykin. (1999). Neural Networks, a comprehensive foundation. USA, NJ: Prentice Hall Press

Sheikh, H.R. & Bovik A.C.yung-Sin Song. (2005). Handbook of Image and Video Processing. Information Theoretic Approaches to Image Quality Assessment, Academic Press

Alvaro Pardo. (2003). Tools for image analysis, representation and compression PhD. Dissertation.

Saifuddin, M. (2010). Image Deblurring (BARCODE), BSc. Thesis.

H. J. Barnard. (1994). Image and Video Coding Using Wavelet Transforms, PhD. Dissertation.

Karen Lees. (2002). Image Compression using wavelets. MSc. Thesis.

D. C. Montgomery, E. A. Peck, G. G. Vining. (2013). Linear Regression Analysis. Wiley

Navidi W & Monk B. (2012). Elementary Statistics, UK, McGraw-Hill Press

Book Chapter

Myung-Sin Song. (2008). Applied and Numerical Harmonic Analysis. In, John J. Benedetto (Eds.*). Entropy Encoding in Wavelet Image Compression* (pp. 293-311). Springer-Verlag Press

Rumelhart, D. E. and McClelland, J. L. . (1986). Psychological and Biological Models, *In* chapter On learning the past tenses of English verbs. (pp. 216-271). Cambridge, MA, MIT Press

International Conference

Kljun, M., Vicic, J., Kavsek, B., & Kavcic, A. (2007). Evaluation comparisons and evaluations of learning management systems. *In Proceedings of the 29th International Conference on Information Technology Interfaces (ITI 2007)* (pp. 363–368). Zagreb: University of Zagreb.

David M. Rouse and Sheila S. Hemami (2006). Quantifying the Use of Structure in Cognitive Tasks. In Proceedings of the IEEE Western New York Image Processing Workshop (WNYIP). San Jose, CA

Ahmet Eskicioglu (2000). Quality measurement for monochrome compressed images in the past 25 years. In Proc. of the International Conference on Acoustics Speech and Signal Processing (ICASSP) (pp. 1907-1910). San Diego

M. Nadenau and J. Reichel (1999). Compression of color images with wavelets under consideration of the hvs. In Proceedings SPIE Human Vision and Electronic Imaging Conference (pp. 129-140). San Jose, CA

S.J.P. Westen, R.L. Lagendijk, and J. Biemond (1995). Perceptual image quality based on a multiple channel hvs model. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 2351-2354). Detroit, Michigan

A.V. Rao, M. Thomas, S. Mitra, and M.E. Parten (1994). Compression of high resolution images with visually transparent techniques. In SPIE, Proc. On Applications of Digital Image Processing. (pp. 71-82). San Diego

Wei Gaofeng; Jiang Hongxu; Yang Rui (2010).Linear-Regression Model Based Wavelet Filter Evaluation for Image Compression. In Asia-Pacific Conference on Wearable Computing Systems. (pp. 315-318). Shenzen

Chin-Chen Chang, Yeu-Pong Lai (2009). An Enhancement of JPEG Still Image Compression with Adaptive Linear Regression and Golomb-Rice Coding. In Proceedings of the 2009 Ninth International Conference on Hybrid Intelligent Systems. (pp. 35-40). Shenyang

Ashraf, R., Akbar, M. (2007). Adaptive Architecture Neural Nets for Medical Image Compression. In IEEE International Conference on Engineering of Intelligent Systems. (pp. 1-4). Islamabad

Vilovic, I. (2006). An Experience in Image Compression Using Neural Networks. In Proceedings of IEEE 48th International Symposium on Multimedia Signal Processing and Communications. (pp. 95-98). Croatia

Ashraf, R., Akbar, M. (2005). Absolutely lossless compression of medical images. In Proceedings of IEEE 27th Conference on Engineering in Medicine and Biology. (pp. 4006-4009). Shanghai

Veisi, S., Jamzad, M. (2007). Image Compression with Neural Networks Using Complexity Level of Images. In Proceedings of the 5th IEEE International Symposium on image and Signal Processing and Analysis. (pp. 282-287). Istanbul

Khashman, A. Dimililer, K. (2007). Neural Networks Arbitration for Optimum DCT Image Compression. In Proceeding of IEEE International Conference on 'Computer as a Tool'. (pp. 151-156). Poland

Li, W., Wu, X., Zhang, Y., Yu, M. (2007).Wavelet-based Far Infrared Medical Image Compression with Histogram Shifting. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedical Engineering. (pp. 1025-1027). China

Shenbaga, S., Vidhya, D.K. (2007). Development of Medical Image Compression Techniques. In Proceedings of International Conference on Computational Intelligence and Multimedia Applications. (pp. 97-101). Sivakasi

Khashman, A. Dimililer, K. (2005). Comparison Criteria for Optimum Image Compression In Proceeding of IEEE International Conference on 'Computer as a Tool'. (pp. 935-938). Serbia