Image Compression

Using Discrete Wavelet Transform And

Predictive Techniques

A Thesis

Submitted By

Ishraq M. Hassan

To the Council of College of Science - University of Babylon in Partial Fulfillment of the Requirements for the Degree of M.Sc. in Computer Science

October - ٢٠٠٦   **Ramadhan - ١٤٢٧**

نظام ضغط الصور
باستخدام تقنيتي تحويل المويجة والمتنبأ

رســـالة

مقدمة من قبل

إشـراق مـرزه حسـن

الى مجلس كلية العلوم - جامعة بابل

وهي جزء من متطلبات نيل درجة الماجستير في علوم الحاسبات

تشرين الاول - ٢٠٠٦   رمضان -١٤٢٧

بِسم ِاللهِ الرَحّمن ِالرحّيم

وَقـُل ِاْعمَلوُا فسَيَرى َ اللهُ عَمَلـَكُمْ وَرَسـُوُلهُ

وَالمُؤْمِنـُونَ وَسَتـُرَدُّونَ اِلـى عَاِلـِم ٱلغيْـبِ

وَالشـَّـهَادَةِ فَيُنَبـِئـُكُمْ بمِا كُنتـُمْ تـَعْمَلـون ۞

صدَقَ اللهُ العَـلِيُ العَظِيمْ

سُوَرْةُ التوّبَة ْ (١٠٥)

الخلاصة:

إعتمـدت فكـرة نظـام ضـغط الصـور المطـور المقتـرح علـى تطبيـق التحويـل المـويجي المتقطـع علـى صـورة خطـأ التنبـؤ **(Error Predictive or Residual Image)** بـدلا مـن تطبيقـه علـى الصـورة الأصـلية (تحتـوي صـورة خطـأ التنبـؤ علـى قيـم بسـيطة مقارنـة بقيـم الصورة الأصلية) لتقليل قيم معاملات كل حزم المويجة, وكنتيجة يمكن الاستغناء عن حـزم التفاصيل **Details-Bands** للمستوى التحليلي الأول وكل/ أو بعض حزم التفاصيل للمستوى

التحليلي الثاني كما انه عند تقليل قيم معاملات المويجة فأن تأثير عملية التكماة **Quantization Process** على الحزم التي تُعتمد لاعادة بناء الصورة يكون أقل.

تم تطبيق النظام المقترح على نوعين من الصور هما, الصور الطبيعية **Natural Image (R,G,B)** والصور ذات التدرجات الرمادية **Gray-Scale Image**. وكانت النتائج متقاربة لهذين النوعين من الصور. أظهرت النتائج انه بالإمكان الاعتماد على حزم التقريبات **Approximation-Bands** للمستوى التحليلي الأول فقط لاسترجاع الصور المعقدة, أما الصور الأقل تعقيد فيمكن الاعتماد على بعض حزم التفاصيل وحزمة التقريبات للمستوى الثاني لاسترجاعها, ولاسترجاع الصور متوسطة التعقيد و الصور البسيطة فيمكن الاعتماد على تقريبات المستوى التحليلي الثاني فقط.

تمت مقارنة سبعة أنواع من المويجات المختلفة لأربعة عوائل هي:

db١, db٣, db٥, sym١, coif١, bior ٢.٢, bio٢.٤.

كانت النتائج متقاربة غالبا إلا إن مويجة **(db٣)** حققت أفضل **Peak Signal -to- Noise Ratio (PSNR)** في حين أعطت عائلة **(Bior)** أقل نسبة ضغط.

طريقة التكماة المستخدمة هي من نوع المكمأ المفرد **Scalar-Quantizer** حيث تم خزن أعلى قيمة لكل حزمة ومن ثم تم تخصيص (٥ بت) لتمثيل كل قيمة من قيم معاملات التقريبات و (٣ بت) لتمثيل كل قيمة من قيم معاملات التفاصيل.

أما طريقة الضغط المستخدمة لضغط المعاملات المكمأة فهي طريقة **Run-Length Encoding (RLE)** التي أعطت نتائج جيدة مع النظام المقترح عند استخدامها لضغط حزم التفاصيل المكمأة , في حين كانت نتائجها محدودة عند استخدامها لضغط حزم التقريبات المكمأة .

أعلى نسبة ضغط تم الحصول عليها للصور التي تم اختبارها هي (٣٨:١).

تم تطبيق النظام باستخدام لغة **MATLAB**.

ABSTRACT

The idea behind the suggested developed image compression system depends on the application of DWT on error prediction values (Residual Image) instead of the original image, the error prediction values have lower values compared to the original image, to reduce the coefficient values of all the wavelet bands. As a result the detail bands can be overcomed for the first decomposition level and all / some of detail bands for the second decomposition level. In addition the effect of quantization process is less.

The system has been applied on the natural and gray-scale images. The results obtained were similar for both image types. It is possible to depend on the approximation band of one decomposition level only to recover the complex images. For the less complicated images some of detail-bands in addition to the approximation band of two decomposition levels can be used to recover them. To recover less complicated and simple image the approximation band of two decomposition level only can be used.

Seven types of different wavelet for four families were compared:

db١, db٣, db٥, sym١, coif١, bior ٢.٢, bio٢.٤.

The results obtained were almost similar. The **db٣** wavelet produced the best (PSNR), which nearest from **Coif١** wavelet. Where as the **bior** family produced less compression ratio.

The suggested quantization method was of scalar-quantizer type, the highest value for each band was saved, then **(٥-bits)** have been used to represent each of approximation coefficients, and **(٣-bits)** for each of details coefficients.

The use **RLE** to compress the quantizer details coefficients produced good results with the suggested system, but it is limited in its performance when applied on the quantizer approximation coefficients.

The highest compression ratio obtained of the test images is about (١:٣٨).

The suggested system have been applied using **MATLAB**.

Supervisor Certification

I certify that this thesis was prepared under my supervision at the department of Computer Science/College of Science/Babylon University, by **Ishraq M. Hassan** as partial fulfillment of requirements for the degree of Master of Science in Computer Science.

Signature:

Name : **Dr. Eng. Sattar B. Sadkhan**

Title **: Assistant Professor**

Date : / /٢٠٠٦

In view of the available recommendations, I forward this thesis for debate by the examination committee.

Signature:

Name : **Dr. Abbas M. Albakry**

Title : **Head of Computer Science Department**

Date : / /٢٠٠٦

Acknowledgment

I am indebted to **MY GOD** for always helping me to finish what I start and for helping me to present this work.

I would like to express my deep gratitude and appreciation to my supervisor **Dr. Eng. Sattar B. Sadkhan** for his submitting to the title of this work , support and encouragement given during the course of this work. Their suggestions, guidance and moral support in difficult times have been essential for completing this work.

My deep appreciation goes to the department of computer science, The **head** of department and **members** of the staff.

Also I wish to thank my **friends** for their cooperation . And for **all the persons** that helped me to finsh this work.

LIST OF TABLES

List of Abbreviations

Abbreviation	Meaning
٢D-DCT	Two dimensional – Discrete Cosine Transform
٢D-DWT	Two dimensional- Discrete Wavelet Transform
٢D-IDWT	Two-dimensional Discrete Wavelet Transform
BMP	Bit Maps Image Format
C.R.	Compression Ratio
dB	Unit measurement of SNR(decibels)

Db٤	Daubechies ٤ wavelet type
DWT	Discrete Wavelet Transform
FIR	Finite Impulse Response
IDWT	Inverse Discrete Wavelet Transform
JPEG	Joint Photographic Experts
JPEG-DCT	**DCT used in JPEG**
MRA	Multi-Resolution Analysis
MSE	Mean Square Error
PSNR	Peak-Signal-to- Noise-Ratio
QMF	Quadrature Mirror Filters
RGB	Red, Green, Blue
RLE	Run Length Encoding
RMSE	Root Main Square Error
SNR	Signal-to- Noise-Ratio
STFT	Short Time Fourier Transform
VQ	Vector Quantization.
WT	Wavelet Transform

LIST OF FIGURES

Contents

١.١ Introduction :

Visual information, transmitted in the form of digital images, is becoming a major method of communication in the modern age (information age).

Computer imaging can be separated into two primary categories: ١) Computer vision and ٢) image processing (include image restoration, image enhancement, and image compression).

Image compression involves reducing the typically massive amount of data needed to represent an image. This is done by eliminating data that are visually unnecessary and by taking advantage of the redundancy that is inherent in most images [١][٢].

Researches in data compression has grown rapidly producing a large number of compression methods. Some of these methods are based on transforms [١].

In general, a transform maps image data into a different mathematical space via a transformation equation. Here, the image data are mapped from the spatial domain to the frequency domain (spectral domain).

These transformations are used as tool in many areas of engineering and science, including computer imaging.

The most widely used are, Fourier transforms, cosine transforms, Walsh-hadmard transforms, and wavelet transforms.

A wavelet transform has been developed in the ١٩٨٠, as an alternative to the windowed Fourier transform for digital signal processing. It has many applications in signal processing and computer graphics, in addition to their use in data compression [٣].

The principle of wavelet transform coding is to decompose an image into multi-resolution subimages. Each subimage corresponds to a different frequency band by using multiple scale wavelet bases, which result in flexible time and frequency resolution. Adaptive quantization can then be used which preserve the statistical property of each upper band subimage to a chivied good visual reconstruction quality. This technique is free from blocking artifacts, mosquito noise, which exist in block transform coding schemes like JPEG when high compression ratio is required [٤].

Considerable interest has been arisen in recent years regarding wavelet as a new transform technique for both speech and image processing applications. This technique has shown effective results in several applications such as edge deduction, feature extraction, nonlinear noise filtering, and image compression [٥][٦].

١.٢ Literature Review:

Image compression is a well-studies field, in which it is difficult to chive significant improvements. This opens the door to new and interesting research avenues [٦].

The wavelet transformation collected with image compression by using **Subband Coding**.The work in image processing by using subband coding techniques was first used by (J. Wood) in ١٩٨٦, who applied subband coding on original image then apply **Differential Predictive Coding (DPC)** on the coefficients of subband to reduce the coefficients values, which now can compressed well with **Entropy Coding** [٧].

In ١٩٨٨, Hammed used the Subbands Coding to break down the gray-scale image. The original image was break down in to (٧) subbands (٢ levels of analysis). Then the **Differential Predictive Coding (DPC)** was applied on the bands produced to reduced the subband values, and then the **Entropy Coding** was used to compresse these simple values [٨].

Wen-chuang, ١٩٩٦, used compression algorithm that depends on estimation of edge location of the subbands. The wavelet transform was applied on gray-scale images using (db٢ wavelet) and the coefficient was determined whether it represents edge, will conserved otherwise it will be neglected. Then **Layd-Max-Quantizer** was used to quantization the wavelet coefficients, which now can be compressed with **Entropy Coding** [٩].

Christopher J.C. Burges and others, ٢٠٠١, explored the use of neural networks to predicate wavelet coefficients for image compression. They showed that by reducing the variance of the residual coefficients, the nonlinear prediction could be used to reduce the length of the compressed bitstream. They obtained results on several network artitctures (fully connected net (box, and pyramidal), and convolutional net) and training methodologies(on-the-fly training, optimal brain damage and fixed-weights); some pitfalls of the approach are examined and explained. A two layer fully connected network trained off-line, applied to a test set consisting of seven ٥١٢ by ٧٦٨ images from the Kodak database, and gives ٥ consequent overall improvement in the bit rate between ٤٪ and ٧٪ [١٠].

Michael B. Martin and Amy E. Bell,٢٠٠١, they claimed that the advances in wavelet transforms and quantization methods have produced algorithms capable of surpassing the existing image compression standards like the Joint Photographic Experts Group (JPEG) algorithm. And they mentioned that for best performance in image compression, wavelet transforms require filters that combine a number of desirable properties, such as orthogonality and symmetry. However, the design possibilities for wavelets are limited because they cannot simultaneously possess all of the desirable properties. They claimed (also) that the relatively new field of M**ultiwavelets** shows promise in obviating some of the limitations of wavelets, since the multiwavelets offer more design options and are able to combine several desirable transform features. The few previously published results of multiwavelet-based image compression have mostly fallen short of the performance enjoyed by the current wavelet algorithms. In the paper [١١], the auther presented new decomposition scheme in multiwavelet transform (iterating on only the $L_{١}L_{١}$- subband instead of iterating over the four L_iL_i- subbands). Also they claimed that the SPIHT quantizers hold well with scalar wavelet but they do not hold for multiwavelet. Therefore they introduced a new quantization methods work well with multiwaveletes that allows multiwavelets decomposition to receive most of the benefits of SPIHT-like quantizer, and referred to this method as **" Shuffling"**. Also they introduced an approach that combines wavelet packet decomposition with multiwavelet filters; they call it "multiwavelet packet decomposition". According to their results, the multiwavelet packets typically give the best results for the synthetic images (which tend to have more high-frequency content than natural images and they often don't compressed as well with traditional image compression transform like the Discrete Cosine Transform (DCT) and Scalar Wavelets). The wavelet packets give the best results for the natural images (which typically have a large amount of low-frequency content) with few exceptions. Where they suggested to use Uniform Quantization with a basis selection method based on a rate-distortion approach to reach a better performance in multi wavelet packets.

Lakshmi R. Iyer and Amy.E.Bell, ٢٠٠١ [١٢], claimed that the multiwavelet transform unlike the scalar wavelet transform allows orthogonality and symmetry to exist. Also they mentioned that for lossy images compression, the balancing order of the multiwavelet filter bank dictates energy compaction. But balancing alone dose not guarantee good compression performance. Filter bank characteristics such as shift-variance and magnitude response also influence peak signal-to-noise ratio (PSNR) and perceived image quality. They analyzed the effete of multiwavelet properties on image compression, and presented the comparison between the balance performance and unbalance wavelet by using three balanced, orthogonal multi-wavelets SA٤^{b}, BMW٨ and BMW١٢ and unbalanced multi-wavelet SA٤^{u}. The suggested image compression scheme employ symmetric signal extension and ٥ level multiwavelet decomposition with shuffling. They compared the proposed system performance with the performance of the biorthogonal scalar wavelet B٩/٧. The results showed that the unbalanced multiwavelet usually perform about ٠.٣ – ١.٢ db worst than balanced multiwavelets, and SA٤^{b} typically outperforms the other balanced multiwavelets but the difference in their PSNRs is typically less than ٠.٥ db. A comparison with scalar wavelets shows that B٩/٧ a chives ٠.٢ - ٠.٥ db higher PSNRs than SA٤^{b}. They used Lena, Goldhill, and Barbara images as a test images. The results indicated the best multiwavelet image compression performance reported to that date.

Panrong Xiao, ٢٠٠١ [١٣], presented a study to compare wavelet methods, a software tool called “Min Image” was used .The “Min Image “ apply one type of wavelet and uniform quantizer (UQ). He modified Min Image software by adding additional functionally to support other wavelet types. The results of different wavelet were compared. Another modification is using EZW coding algorithm instead of UQ and their results were compared. The compression results were analyses and compared. These include the subjective and objective qualities of reconstructed images, timing of composition and decomposition for different wavelet types, and timing of EZW coding algorithm for different compression ratio.

Francois G. Meyer A.Z. Averbuch and R.R.Coifman,٢٠٠١[١٤], provided paradigm for image representation and image compression as a solution to the image coding problem (any image is parsed into a superposition of coherent layers: smooth-regions layer, textures layer, etc.) referred to it “ Multi-Layered” representation or decomposition. The multi-layered decomposition algorithm consists in a cascade of compressions applied successively to the image itself and to the residuals that resulted from the previous compressions. During each iteration of the algorithm, they code the residual part in a lossy way. They only retaine the most significant structures of the residual part, which results in a sparse representation. Each layer is encoded independently with a different transform, or basis, at a different bit rate; and the combination of compressed layers can always be reconstructed in a meaningful way. The strength of the multi-layer approach comes from the fact that the base functions will give reasonable account of the large trend of the data, while others will catch the local transients, or the oscillatory patterns. They claimed that this multi-layered representation has a lot of beautiful applications in image understanding, and image and video coding. They presented studies for its capabilities, also they showed the results of implementing the algorithm on Barbara, and Houses images and it's comparison with well-known SPHIT algorithm.

Suhad Ahmed, ٢٠٠٢ [١٥], provided an analysis for using wavelet transformation families (db١,٢,٣ ; Sym٥,٦ ; bior٢.٢,٣.٣,٢.٤ and cofi١) in compressing various types of images, also comparisons among different results of these wavelets. She used the scalar quantizer for the subbands groups result from applying the wavelet transformation on images. Then, using the Run Length Coding (RLE) for the resulted subbands. She had been a chivied compression ratio (٧٢٪ up to ٨٥٪ with peak signal –to- noise ratio PSNR (٢٢.٠٣ - ٢١.٠٥)) respectively, for gray level images, (٧٧٪ up to ٨٩٪ with PSNR (٢١.٠٧ - ١٩.٥٥)) respectively, for RGB color images and (٧٥٪ up to ٨٨٪ with PSNR (١٥.٩٣ - ١٦.٣٨)) respectively, for color images with ٢٥٦ color.

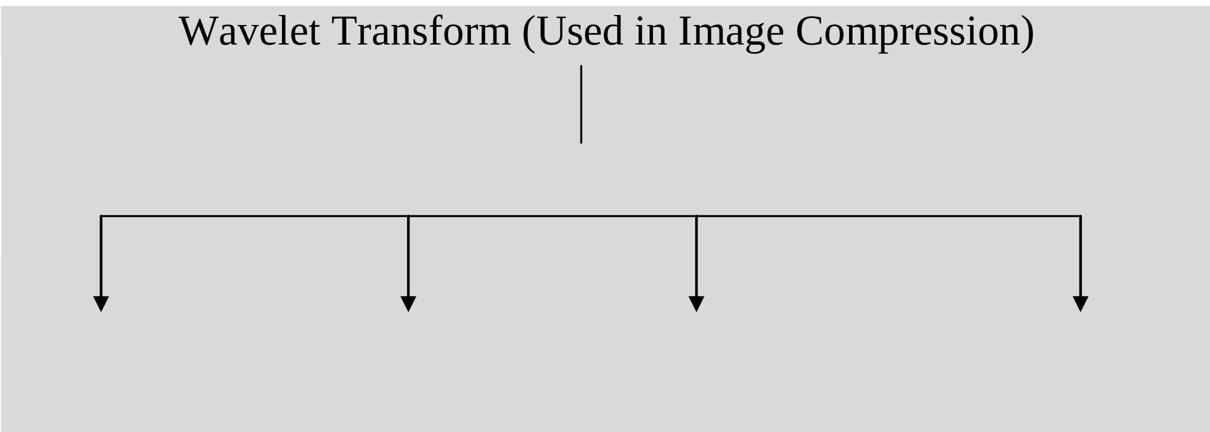
S.Rout and Amy.E.Bell, ٢٠٠٢ [١٦], compare the compression performance of ٣ scalar wavelets and ٥ balanced multi-wavelet on ٧ color images, where the previous compaction of some scalar wavelets and multi-wavelet properties illustrated their importance for grayscale images. Also they presented another property; the perfect reconstruction (PR). The observed that SA٤^{b} depicted the best performance amongst the multi-wavelets in terms of both subjective quality and PSNR. These result matches their exception based on the result, which indicated that only BMW satisfies the (PR) conditions. Also they claimed that although SA٤ rivaled the biorthagonale scalar wavelets in term of many desirable properties, it suffers from lower balancing vanishing order, this difference explained the (٠.٠١- ٠.٤٧) db performance gab between the best scalar wavelet and best balanced multi-wavelet.

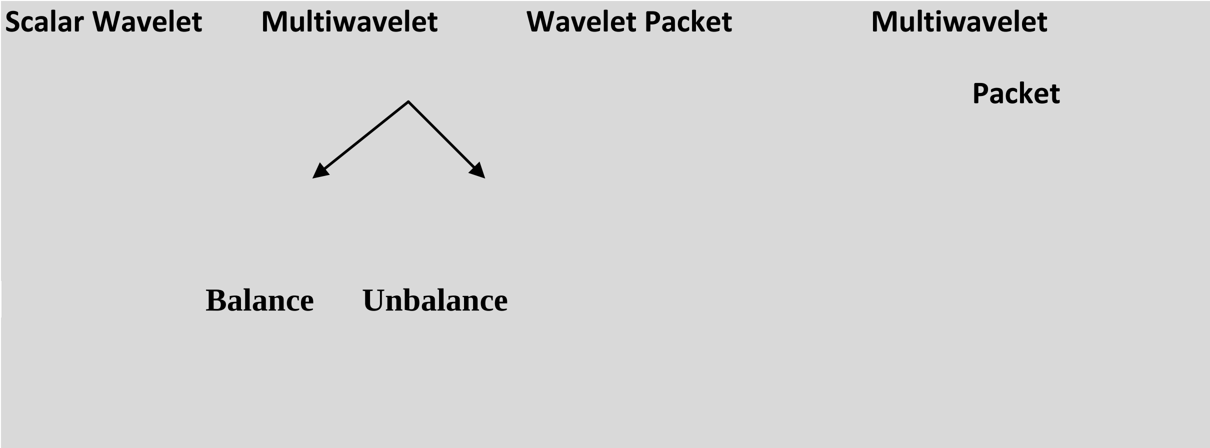
Patrice Y.Simard, Christopher and David in ٢٠٠٢ [٦], proposed two algorithms based on the assumption that typical images are made of smooth regions of low activity, which are easily compressed by linear transform, and high activity regions (edges, textures), which are harder to compress. The first algorithm, the Tarp Filter, produce a simple but efficient recursive density estimator of wavelet coefficients, has very low capacity, fast on-line adaptation, and an infinite-size context. It is best suited for regions of low activity, which are essentially composed of zeros. In contrast the second algorithm used to predict wavelet coefficients based on neural networks, which has high capacity, a finite-size context, and is trained Off-line. The size of the context is determined by the capacity needed for learning and by the desired computational speed. Also they claimed that the two encoders could be used separately or in combination.

K. A. Kotteri Amy. E. Bell and J. E. Carletta, ٢٠٠٤, showed that the biorthogonal ٩/٧ wavelet is used for lossy compression in the JPEG٢٠٠٠ image coding standard. A hardware implementation of the standard must consider the accuracy and the efficiency with which the quantized filter coefficients are represented; filter structure is also an important design consideration. A high precision representation ensures compression performance close to the unquantized, infinite precision filter bank, but this comes at the cost of increased hardware resources and processing time. They proposed a gain compensation method that significantly improves the performance of the quantized filters without sacrificing efficiency. Their method achieved performance close to the unquantized filter case while also realizing a fast, efficient hardware implementation [١٩]. In another study they claimed that the filter bank approach for computing the discrete wavelet transform (DWT), which called the convolution method, can employ either a “ nonpolyphase” or “Polyphase” structure. They introduced a comparison for filter banks with an alternative polyphase structure for calculating the DWT- the lifting method. They look at the traditional lifting structure and a recently proposed “ flipping” structure for implementing lifting. The quantization of the coefficients (for implementation in fixed-point hardware) plays a crucial role in the performance of all structures, affecting both image compression quality and hardware metrics. They designed several quantization methods and compared the best design for each approach: the non-polyphase filter bank, the polyphase filter bank, the Lifting and recently proposed “Flipping” structure. The results indicated that; for the same image compression performance, the flipping structure gives the smallest and fastest, low-power hardware [١٨].

In ٢٠٠٥, D.A. Karras, S.A.Karkanis and B.G.Mertizos suggested a image compression scheme, using the discrete wavelet transformation (DWT), which is based on attempting to preserve the texturally important image characteristics. They claimed that the main point of the proposed methodology lies on that, the image is divided into regions of textural significance employing textural descriptors as criteria and fuzzy clustering methodologies. These textural descriptors include concurrence matrices based measures and coherence analysis derived features. While rival image compression methodologies utilizing the DWT apply it to the whole original image. They presented an approach that involves a more sophisticated schema in the application of the DWT. More specifically, the DWT is applied separately to each region in which the original image is partitioned, and depending on how it has been texturally clustered, its relative number of the wavelet coefficients is determined. Therefor, different compression ratios are applied to the above specified image regions. An experimental study is conducted to qualitatively assessing the proposed compression approach. Moreover, this experimental study aimed at comparing different textural measures in terms of their results concerning the quality of the reconstructed images [١٩].

As a conclusion from literature review, we can summarize the different wavelet transformation and quantization methods used within image compression activities in the following three figures.





Figure(١.١): Wavelet Transform used in Image compression.

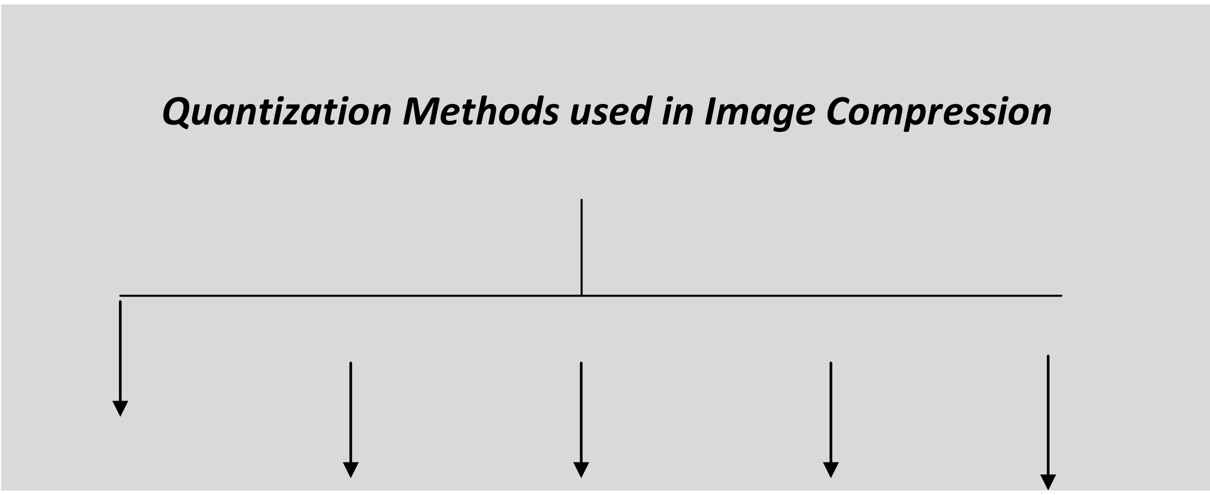
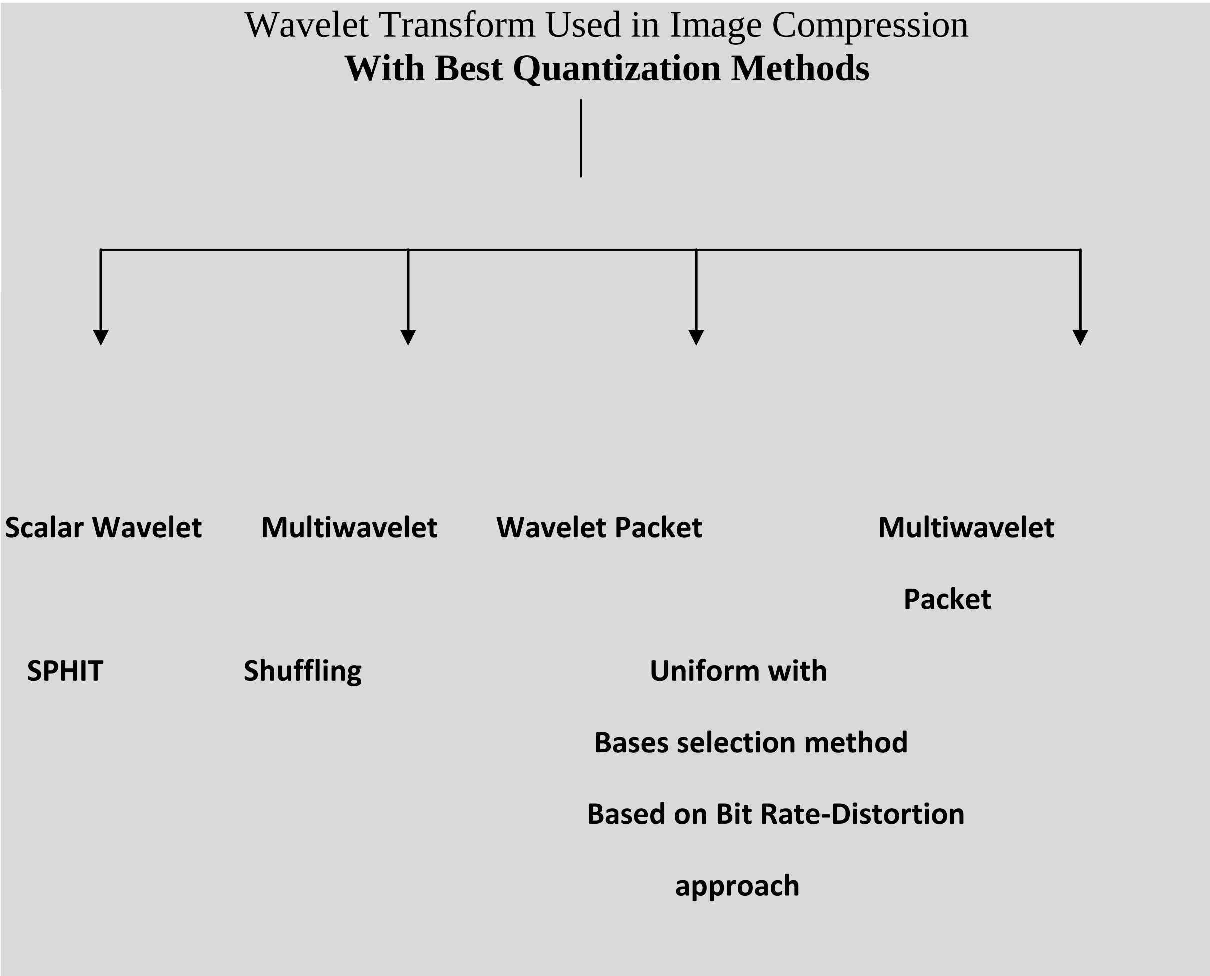


Figure (١.٢): Quantization Methods Used in Image Compression.



Natural Images	Synthetic Images	Natural Images	Synthetic Images

Figure (١.٣): Wavelet Transforms Used in Image Compression with Best Quantization Method Suggested to Each Type(Natural & Synthetic Images).

١.٣: Aim of the thesis:

The thesis aims towards the following aspects:

١. Designing a developed image compression system using wavelet transform technique.

٢. Studies different possible cases to analysis it and then proposed an general system.

٣. Software implementation of the designed system using a appropriate programming language.

٤. Evaluating the performance of the developed algorithm.

١.٤: Outline of the Thesis

This thesis presents a developed image compression technique based on wavelets. It is organized as follows:

Chapter two presented images compression techniques, also the criteria using to evaluate the performance of any compression technique.

Chapter three includes the theoretical background of the mathematical material used in the thesis. It begins with an introduction to the wavelet followed by the presentation of the wavelet theory and wavelet analysis.

Chapter four is devoted to the developed image compression system, case studies with their discussion, and general suggested image compression system derived from the results. The obtained simulation results and discussion, conclusions and suggestions for future works are introduced in ***Chapter Five***.

٢.١:Definitions:

Image compression has been pushed to a head of image processing field, which allows an efficient transmission or storage of digital images with a minimum [١].

Image compression involves reducing size of image data files, while retaining necessary information. The reduced file is called the compressed file, and it used to reconstruct the original image, resulting in the decompressed image. The original image, before any compression is performed it is called uncompressed image file [٢].

Taking advantage of the redundancy that is inherent in image data develops compression algorithms. Three primary types of redundancy can be found in images:

Coding Redundancy: This type occurs when the data used to represent the image are not utilized in an optimal manner.

Interpixel Redundancy: Occurs because adjacent pixels tend to be highly correlated (This in most images the brightness levels do not change rapidly, but gradually). So that adjacent pixel values tend to be relatively close to each other in value.

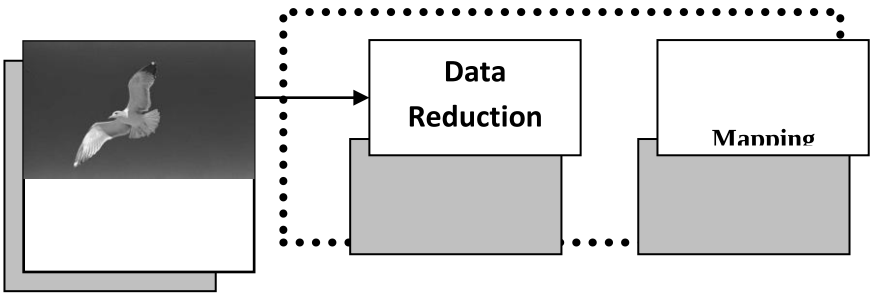
Psychovisual redundancy: Some information is more important to the human visual system to another type of information [١].

The key in image compression algorithm development is to find out exactly minimal data required to retain the necessary information. This is accomplished by taking advantage of the redundancy that exists in images.

٢.٢: Compression System Model:

The compression system model is composed of two parts: The compressor and decompressor.

- **The Compressor Part :** Is composed of a preprocessing stage and encoding stage as shown in figure (٢.١).



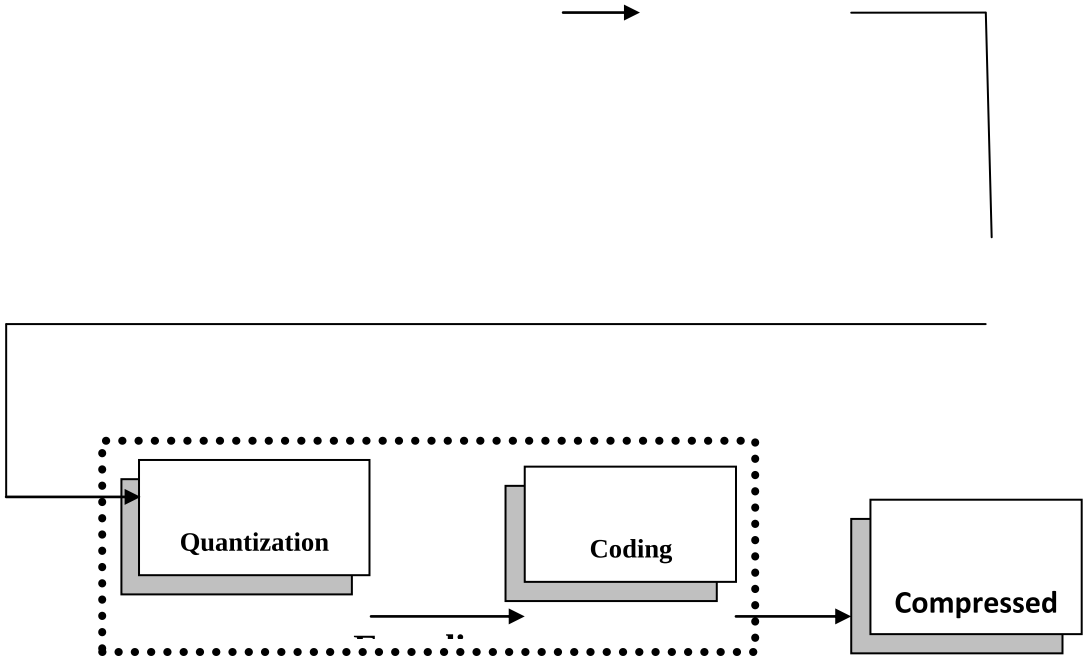


Figure (٢.١): Compressor

The first step in preprocessing is data reduction. This may be done by image enhancement (or by Gray-level and/or Spatial Quantized). The second step is mapping process, which maps the original image data in to another mathematical space where it is easier to compress the data.[١] The encoding stage can be also broken down in to two processes: Quntization and Encoding. The quantization is a process that's its output can have only a limited number of possible values. Each input is forced to one of the allowable output values, while the coding process maps the output data from the quantization process onto a code in optimal manner. A compression algorithm may consist of all the stage, or it may consist of only one or two of the stages.

- ***The Decompressor Part:*** Can be further broken down into the stages shown in figure (٢.٢). The first, the decoding stage, takes the compressed file and reverses to original coding by mapping the codes processed by a stage that performs an inverse mapping to reverse the original mapping process. Finally, the image may be postprocess to enhance the look of the final image [١].

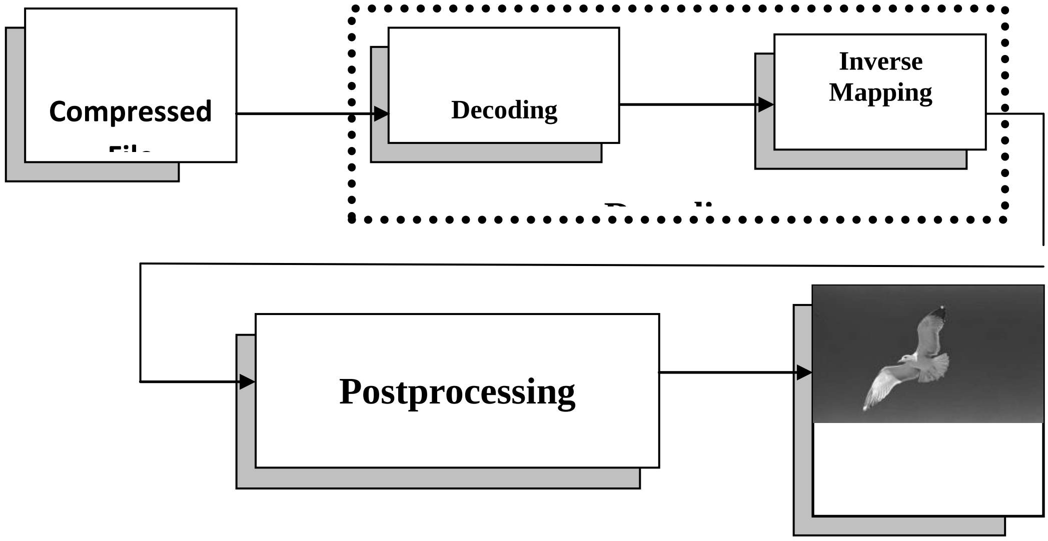


Fig (٢.٢) Decompressor

<u>٢.٣: Compression Techniques:</u>

Based on the requirements of reconstruction, data compression schemes can be divided into two broad classes. One is lossless compression, in which reconstruction is identical to input. Examples of lossless methods are Run length coding, Huffman coding, and Arithmetic coding. The other is lossy compression, which generally provides much higher compression than lossless compression but allows the reconstruction to be different from input [١٣].

٢.٣.١:Lossless Compression:

If data have been lossless compressed, the original data can be recovered exactly from the compressed data. It is generally used for applications that cannot allow any difference between the original and reconstructed data [١٣].

An important concept here is the idea of measuring the average information in an image, referred to as the ***entropy***. The entropy for a NxN image can be calculated by this equation:

$$\textbf{Entropy} = -\sum P_i \log_{٢} (Pi) \qquad (٢.١)$$

This measure provides us with a theoretical minimum for the average number of bits per pixel that could be used to encode the image. This number is theoretically optimal and can be used as a metric for judging the success of coding scheme.[١]

A. Length Encoding (RLE):

Run length encoding, sometimes called recurrence coding, is one of the simplest data compression algorithms. It is effective for data sets that are comprised of long sequences of a single repeated character [١٣].

In images coding, these method works by counting the number of adjacent pixels with the same gray-level value. There are several methods of RLE, basic methods that are used primarily for binary (two-valued) images and extended versions for gray-scale images.

Another way to extend basic RLE to gray-level images include the gray level of a particular ran as part of the code. This technique is only effective with images containing a small number of gray levels [١].

B. Predictive Coding

Predictive coding has been used extensively in image compression. It works by predicting the next pixel value based on the previous values and encoding the difference between the predicted value and the actual value. This technique takes advantage of the fact that adjacent pixels are highly correlated, which means that the difference between adjacent pixels is typically small. Because this difference is small, it will take only a small number of bits to represent it. The use of a predictor can reduce the amount of information to be encoded [١].

The values of some near-neighbors of a pixel are subtracted from the pixel to get a small number, which is then compressed further using Huffman or Arithmetic coding. Figure (٢.٣ a) shows a pixel X and three neighbor pixels A, B and C. Figure (٢.٣ b) shows eight possible ways (Predictions) to combine the values of the tree neighbor. In lossless mode, the user can select one of these predictions, and the encoder then uses it to combine the three neighbor pixels and subtract the combination form the value of X. The result is normally a small number which is then entropy-coded. The lossless mode of JEPG uses differencing to reduce the values of pixels before they are compressed.

Predictor ٠ is used only in hierarchical mode of JPEG. Predictors ١,٢, and ٣ are called “One-dimensional “. Predictors ٤,٥,٦ and ٧ are “Two-dimensional” [٢].

Selection Value	Prediction
١	**A**
٢	**B**
٣	**C**
٤	**A+B-C**
٥	**A+((B-C)/٢)**

	C	B	
	A	X	

(a) *(b)*

Figure (٢.٣): shows eight possible ways (Predictions) to combine the values of the tree neighbor.

<u>C. Huffman Coding:</u>

Huffman coding, developed by D.A. Huffman, is a classical data compression technique. It has been used in various compression applications, including image compression. It uses the statistical property of characters in the source stream and then produces respective codes for these characters. These codes are of variable code length using an integral number of bits. The codes for characters having a higher frequency of occurrence are shorter than those codes for characters having lower frequency. This simple idea causes a reduction in the average code length, and thus the overall size of compressed data is smaller than the original. Huffman coding is based on building a ***binary tree*** that holds all characters in the source at its ***leaf nodes***, and with their corresponding characters' probabilities at the side [١٣]. The Huffman algorithm can be described in five steps [١]:

١. Find the gray-level probabilities for the image by finding the histogram.

٢. Order the input probabilities from smallest to largest.

٣. Combine the smallest two by addition.

٤. GOTO step ٢, until only two probabilities are left.

٥. By working backward along the tree, generate code by alternating assignment of ٠ and ١.

By using this procedure, the characters are naturally assigned codes that reflect the frequency distribution.

Highly frequent characters will be given short codes, and infrequent characters will have long codes. Therefore, the average code length will be reduced. If the count of characters is very biased to some particular characters, the reduction will be very significant[١]. An example of how Hoffman coding work shown in example (٢.١).

Example ٢.١:

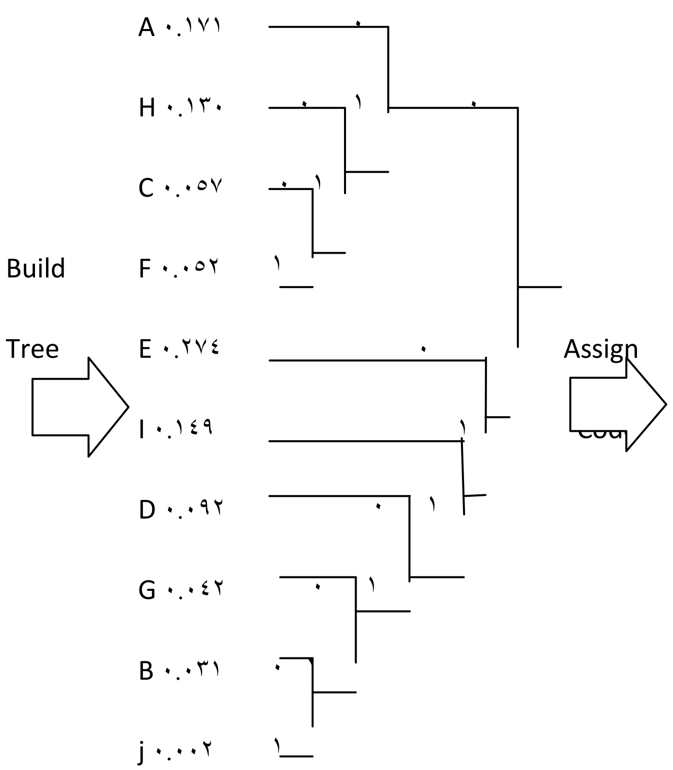
If we have the probability distribution of characters as shown in table (٢.١). [٢٠]

Table (٢.١): Example of Huffman input with probabilities.

X	A	B	C	D	E	F	G	H	I	J
P(x)	٠.١٧١	٠.٠٣١	٠.٠٥٧	٠.٠٩٢	٠.٢٧٤	٠.٠٥٢	٠.٠٤٢	٠.١٣٠	٠.١٤٩	٠.٠٠٢

The final static code tree is given below in figure (٢.٤):

X	P(X)
A	٠.١٧١
B	٠.٠٣١
C	٠.٠٥٧
D	٠.٠٩٢
E	٠.٢٧٤
F	٠.٠٥٢
G	٠.٠٤٢
H	٠.١٣٠
I	٠.١٤٩
J	٠.٠٠٢



X	Code
A	٠٠
B	١١١١١٠
C	٠١١٠
D	١١١٠
E	١٠
F	٠١١١
G	١١١١٠
H	٠١٠
I	١١٠
J	١١١١١١

Figure (٢.٤) Example of Huffman with its code.

D. Arithmetic Coding:

The Huffman method is more efficient method, but rarely produces the best variable-size code. Arithmetic coding is also a kind of statistical coding algorithm similar to Huffman coding. However, it uses a different approach to utilize symbol probabilities, and performs better than Huffman coding. This method read the input stream symbol by symbol and appends more bits to the code each time [٢]. It can assign a code as a fraction of a bit [١٣]. Therefore, when the symbol probabilities are more arbitrary, arithmetic coding has a better compression ratio than Huffman coding. In brief, this is can be considered as grouping input symbols and coding them into one long code. Therefore, different symbols can share a bit from the long code.

Although arithmetic coding is more powerful than Huffman coding in compression ratio, arithmetic coding requires more computational power and memory. Huffman coding is more attractive than arithmetic coding when simplicity is the major concern [٢][١٣].

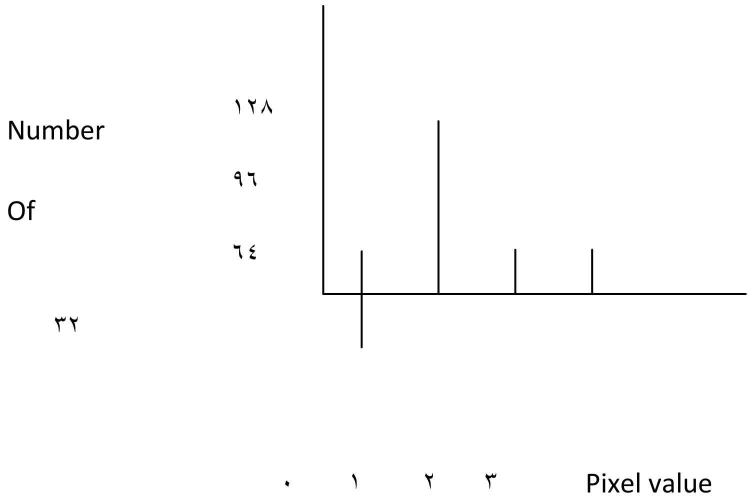
Example ٢.٢[١]:

Given a ١٦×١٦, ٢-bit image with the histogram shown in Figure (٢.٥)a, we can define an arithmetic coding probability table shown in figure (٢.٥)b. The initial subinterval specifies how the ٠ to ١ interval is divided based on the distribution, where the width of the sub interval is equal to the probability, and the subinterval starts where the previous one stops. In figure (٢.٥)c, the actual arithmetic coding process is illustrated, with an example pixel value sequence of ٠, ٠, ٣, ١.

١. Starting on the left, the initial ٠ to ١ interval is subdivided, based on the probability distribution.

٢. The first pixel value ٠ is coded by extracting the subinterval corresponding to the ٠ and subdividing it again, based on relative distribution.

This process is repeated for each pixel value in the sequence until a final interval is determined, in this case from ٥٨/١٠٢٤ to ٦٢/١٠٢٤ or ٠.٠٥٦٦٤٠٦٢٥ to ٠.٠٦٠٥٤٦٨٧٥. Any value within this subinterval, such as ٠.٥٧ or ٠.٠٦٠, can be used to represent this sequence of gray-level values.

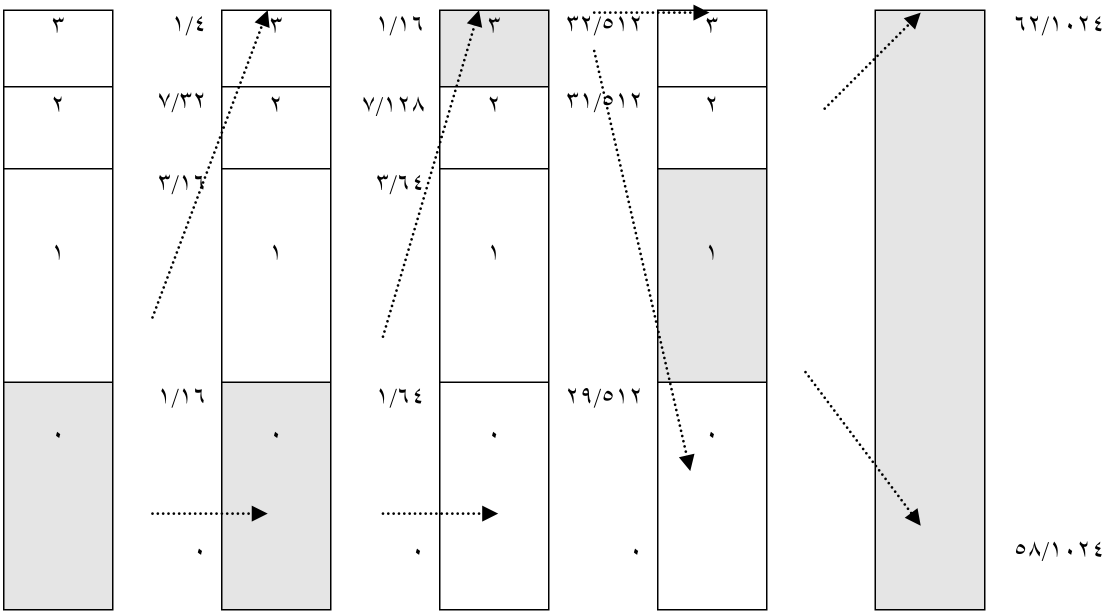


a. Histogram

Pixel value	Probability	Initial Subinterval
٠	٦٤/٢٥٦=١/٤	٠ - ٠.٢٥
١	١٢٨/٢٥٦=١/٢	٠.٢٥ — ٠.٧٥
٢	٣٢/٢٥٦=١/٨	٠.٧٥ — ٠.٨٧٥

٣	٣٢/٢٥٦=١/٨	٠.٨٧٥ - ١

b. Probability Table



c. Coding process for ٠, ٠, ٣, ١.

Figure (٢.٥):Example of Arithmetic Coding.

<u>٢.٣.٢:Lossy Compression Methods:</u>

Lossy compression techniques involve some loss of information, and data can not be recovered or reconstructed exactly. In some applications, exact reconstruction is not necessary. For example, it is acceptable that a reconstructed video signal is different from the original as long as the differences do not result in annoying artifact [١٣]. However, we can generally obtain higher compression ratios results in a poorer image, but the results are highly image dependent. A technique that works well for one application may not be suitable for another.

Lossy compression is performed in both the spatial and transform domains. There are some of these methods [١].

A. Vector Quantization

Vector Quantization (VQ) is a lossy compression method. It uses a codebook containing pixel patterns with corresponding indexes on each of them. The main idea of VQ is to represent arrays of pixels by an index in the codebook. In this way, compression is achieved because the size of the index is usually a small fraction of that of the block of pixels.

The main advantages of VQ are the simplicity of its idea and the possible efficient implementation of the decoder. Moreover, VQ is theoretically an efficient method for image compression, and superior performance will be gained for large vectors. However, in order to use large vectors, VQ becomes complex and requires many computational resources (e.g. memory, computations per pixel) in order to efficiently construct and search a codebook. More research on reducing this complexity has to be done in order to make VQ a practical image compression method with superior quality [١٢].

B. Fractal Compression:

The application of fractals in image compression started with M.F. Barnsley and A. Jacquin [٢١]. Fractal image compression is a process to find a small set of mathematical equations that can describe the image. By sending the parameters of these equations to the decoder, we can reconstruct the original image.

C. Transform Based Image Compression

The basic encoding method for transform based compression works as follows:

١. **Image Transform:** Divide the source image into blocks and apply the transformations to the blocks.

٢. **Parameter Quantization:** The data generated by the transformation are quantized to reduce the amount of information. This step represents the information within the new domain by reducing the amount of data. Quantization is in most cases not a reversible operation because of its lossy property.

٣. **Encoding :** Encode the results of the quantization. This last step can be error free by using Run Length encoding or Huffman coding. It can also be lossy if it optimizes the representation of the information to further reduce the bit rate.

Transform based compression is one of the most useful applications. Combined with other compression techniques, this technique allows the efficient transmission, storage, and display of images that otherwise would be impractical [٢٢].

DCT-Based Transform Coding

Ahmed, Natarajan, and Rao first applied the Discrete Cosine Transform (DCT) [٢٣] to image compression in the work. It is a popular transform used by the JPEG (Joint Photographic Experts Group) image compression standard for lossy compression of images. Since it is used so frequently, DCT is often referred to in the literature as JPEG-DCT, DCT used in JPEG.

JPEG-DCT is a transform coding method comprising four steps. The source image is first partitioned into sub-blocks of size ٨x٨ pixels in dimension. Then each block is transformed from spatial domain to frequency domain using a ٢D-DCT-basis function. The resulting frequency coefficients are quantized and finally output to a lossless entropy coder. DCT is an efficient image compression method since it can decorate pixels in the image (since the cosine basis is orthogonal) and compact most image energy to a few transformed coefficients. Moreover, DCT coefficients can be lossy quantized according to some human visual characteristics. Therefore, the JPEG image file format is very efficient. The JPEG ٢٠٠٠ use wavelet transform instead of DCT transform [١٣].

Wavelet Transform

Wavelets are functions defined over a finite interval. The basic idea of the wavelet transform is to represent an arbitrary function *f(x)* as a linear combination of a set of such wavelets or basis functions. These basis functions are obtained from a single prototype wavelet called the mother wavelet by dilations (scaling) and translations (shifts).

The purpose of wavelet transform is to change the data from time-space domain to time-frequency domain that makes better compression results [١٣].

We will discuss more detail on wavelet theory in Chapter ٣.

Reason to Use Wavelet Based Compression

As discussed earlier, for image compression, loss of some information is acceptable. Among all of the above lossy compression methods, vector quantization requires many computational resources for large vectors; predictive coding has inferior compression ratio and worse reconstructed image quality than those of transform based coding. So, transform based compression methods are generally best for image compression.

For transform based compression, JPEG compression schemes based on DCT (Discrete Cosine Transform) have some advantages such as simplicity, satisfactory performance, and availability of special purpose hardware for implementation. However, because the input image is blocked, correlation across the block boundaries cannot be eliminated.

This results in noticeable and annoying "blocking artifacts" particularly at low bit rates as shown in figure (٢.٦).

(a) (b)

Figure (٢.٦) : (a) Original Lena Image (b) Reconstructed image to show blocking artifacts

Over the past ten years, the wavelet transform has been widely used in signal processing research, particularly, in image compression. In many applications, wavelet-based schemes achieve better performance than other coding schemes like the one based on DCT. Since there is no need to block the input image and its basis functions have variable length, wavelet based coding schemes can avoid blocking artifacts. Wavelet based coding also facilitates progressive transmission of images [٢٤].

٢.٤: Fidelity Criteria

To determine exactly what information is important and to be able to measure image fidelity, we need to define an image fidelity criterion. Fidelity criteria can be divided into two classes:[١]

٢.٤.١:Objective Fidelity Criteria:

A natural way to determine the fidelity of a recovered image is to find the difference between the original and reconstructed values. The most often used average measure is the average of squared error. This measure is called the *Root mean squared error (RMSE*) and it often given by :

$$e_{RMS} = \sqrt{\frac{1}{N^2}\sum_{r=0}^{N-1}\sum_{c=0}^{N-1}\left[\hat{I}(r,c) - I(r,c)\right]^2} \qquad (٢.٢)$$

where:

$I^{\wedge}(r,c)$: Reconstructed Image.

$I\ (r,c)$: Original Image.

N : Samples number.

The smaller the value of the error metrics, the better the compressed image represents the original image. Alternately, with the signal-to-noise (SNR) metrics, a larger number implies a better image. The SNR metrics consider the decompressed image I(r,c) to the “signal” and the error to be “noise”. We can define the root-mean-square signal-to-noise ratio as:

$$SNR_{RMS} = \sqrt{\frac{\sum_{r=0}^{N-1}\sum_{c=0}^{N-1}\left[\hat{I}(r,c)\right]^2}{\sum_{r=0}^{N-1}\sum_{c=0}^{N-1}\left[\hat{I}(r,c) - I(r,c)\right]^2}} \qquad (٢.٣)$$

Another related metric, the peak signal-to nose ratio, is defined as

$$PSNR(dB) = 10\log_{10} \frac{(L-1)^2}{\frac{1}{N}\sum_{r=0}^{N-1}\sum_{c=0}^{N-1}\left[\hat{I}(r,c) - I(r,c)\right]^2} \qquad (٢.٤)$$

Where L= the number of gray levels

(e.g., for ٨ bits L=٢٥٦)

PSNR is the most commonly used value to evaluate the objective image compression quality.

٢.٤.٢:Subjective Fidelity Criteria:

The subjective measures are better method for comparison of compression algorithms, if the goal is to achieve highly quality images as defined by our visual perception.

Subjective testing is performed by creating a database of images to be tested, gathering a group of people that are representative of the desired population, and then having all the test subjects evaluate the images according to predefined scoring criterion. The results are then analyzed statistically, typically using the averages and standard deviations as metrics [١].

٢-٥:Compression Ratio:

The ratio of the original uncompressed image file and the compressed file is referred to as the compression Ratio. The compression Ratio is denoted by [١٥]:

$$\textbf{Compression Ratio} = \frac{\textbf{Uncompressed File Size}}{\textbf{Compressed File Size}} \qquad (٢.٥)$$

Another way to state the compression is to use the terminology of bits pits per pixel. For an N x N image :

$$\textbf{Bits Per Pixel} = \frac{\textbf{Number of Bits}}{\textbf{Number of Pixels}} = \frac{\textbf{Number of Bytes x ٨}}{\textbf{N x N}} \qquad (٢.٦)$$

Also, it can be written by [١٥]:

$$\text{Compression Ratio} = \left(١ - \left(\frac{\textbf{Compressed\_Size}}{\textbf{Original \_Size}}\right)\right) \textbf{X ١٠٠ \%} \qquad (٢.٧)$$

٣.١: Development of wavelet based images coding in brief

State of the art image coder today is based on wavelet theory. The wavelet theory is purely mathematical creation and its applications have been found very useful for image compression. Wavelet theory consists of a transform quite similar to the fourier transform or discrete cosine transform, DCT, and has nothing to do with image compression in itself. When the discrete wavelet transform, DWT, and its extensions like wavelet packets (flat decomposition instead of hierarchical), are used to derive the FIR (Finite Impulse Response) filters actually used in practical application the compression algorithm is referred to as being wavelet based. The filters can then be analyzed in a more traditional way.

One of the fundamental differences between wavelet based coding and traditional sub-band coding, lies in the interpretation of the wavelet transform into classic sub-band terminology. Instead of discussing signal analysis in terms of filters and frequency bands, the wavelet transform uses a mathematical terminology such as wavelet functions bases and subspace. Approximation space and detail space splitting is viewed as low pass and high pass filtering. The cascaded use of the very same filter on incrementally subsampled versions of the original image is one of the multi resolution analyses. From a principle point of view, this idea of several scales is a property that promotes the wavelet transform to something slightly more than just another transform [٢٥].

٣.١.١:Fourier Transform:

In ١٩th century, the French mathematician, J. Fourier, showed that any periodic function could be expressed as an infinite sum of periodic complex exponential functions. Many years after this remarkable property of periodic m functions was discovered, the ideas were generalized to non-periodic functions, and then to periodic or non-periodic discrete time signals. After this, Fourier transform became a very famous tool for computer calculations.

$$F(w) = \int_{-\infty}^{\infty} f(t) e^{-j^{wt}} dt \qquad (٣.١)$$

is generally called the Fourier Transform. The equation

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(w) e^{j^{wt}} dw \qquad (٣.٢)$$

is called the inverse fourier Transform.

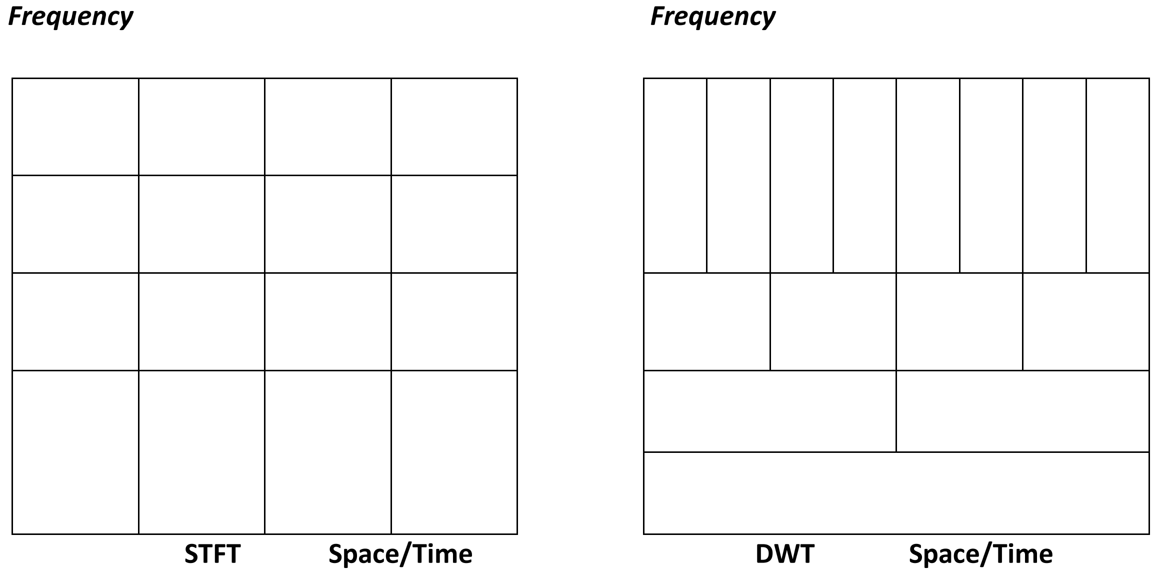
Note that in the Fourier Transform equation, the integration is from minus infinity to plus infinity over time. So, no matter when the component with frequency w appears in time, it will affect the result of the integration equally as well. The lack of time information is one serious weakness of Fourier Transform. That is why Fourier transform is not suitable if the signal has time varying frequency, i.e., the signal is non-stationary [١٣].

٣.١.٢:Windowed Fourier Transform

To solve the above problem, Gabor in ١٩٤٥ introduced the Windowed Fourier Transform. The basic idea is to divide the signal into small enough segments, where these segments can be assumed to be stationary. The width of this window must be equal to the segment of the signal where this assumption is valid.

The Windowed Fourier Transform has several problems. If we use a window of infinite length, we get the Fourier Transform, which gives perfect frequency resolution, but no time information. On the other hand, in order to obtain a stationary sample, we must have a small enough window in which the signal is stationary. The narrower we make the window, the better the time resolution, and better the assumption of stationary, but poorer the frequency resolution [١٣]. However, the Wavelet has this window built in and its size varies depending on the scale. This is very important since it is not possible to measure frequency exactly and at the same time its location. The figure (٣.١) shows the tiling of the space-frequency plan and the area is held constant in the wavelet case, thereby trying to be as precise as possible in both domains. In other words, to measure high frequencies only a short filter is needed while low frequencies, slowly varying, require larger filter kernels and thereby affecting the precision in space or time.

The use of a fixed windowing function is equivalent to cutting the image into fixed blocks before analyzing the frequency. By decomposing the image using wavelets at different scales, block related artifacts are avoided. After decomposition into a approximation part (low-pass) and a detail part (high-pass) there is nothing stopping us from repeating the process once more. If we choose to only split the approximation part, again with the same filter the scheme is called ***wavelet decomposition*** and if we also split the detail part, the scheme is referred to as ***wavelet packet decomposition*** [٢٥].



Figure(٣.١): **The tiling of the space - frequency plane for a one-dimensional signal differs between STFT and DWT (discrete wavelet transform). The window fixates the width in the short-time-fourier-transform, but in the wavelet transform case, the window varies with the frequency. The uncertainty relation sets the lower boundary of the tile size.**

٣.٢: Wavelet Transform

J.Morlet was investigating a way to analyses seismological data and introduced the term Wavelet in early eighties. Wavelet means small wave and the figure below is one example of Haar wavelet, note that its mean is zero [٢٦]. The original term was “Ondeletts”, which is French. A few years later Ingrid Daubechies constructed families of orthonormal wavelet function with compact support and Stephane Mallet put the wavelet transform in the framework of multi-resolution signal decomposition [٢٥].

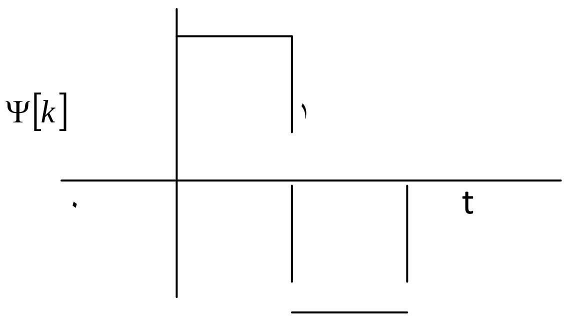


Figure (٣.٢) Haar Wavelet

The fundamental idea behind wavelets is to analyze the signal at different scales or resolutions, which is called multi-resolution. Wavelets are a class of functions used to localize a given signal in both space and scaling domains. A family of wavelets can be constructed from a mother wavelet. Compared to Windowed Fourier analysis, a mother wavelet is stretched or compressed to change the size of the window. In this way, big wavelets give an approximate image of the signal, while smaller and smaller wavelets zoom in on details. Therefore, wavelets automatically adapt to both the high frequency and the low-frequency components of a signal by different sizes of windows. Any small change in the wavelet representation produces a correspondingly small change in the original signal, which means local mistakes will not influence the entire transform. The wavelet function may be either continuos or discrete four our purposes, the discrete form is suited for non-stationary signals [١٣].

The wavelets analyze the input signal in sections by translation of analysis function. With **WT**, the analysis function is a wavelet function, Ψ. The wavelet function is scaled (or expanded or dilated) in addition to being translated in time. The Ψ is often called the mother wavelet [٢٧] because it **“ gives birth “** to a family of wavelets through the dilation and translations. A generalized wavelet family, $\Psi_{a,b}$, described in the normalized form is [٢٨]:

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}}\Psi\left(\frac{x-b}{a}\right) \qquad (٣.٣)$$

where **a** represents the scale and **b** represents the translation (shifting) parameters, and the constants the scale (**١/√a**)is used for energy normalization across different scales.

The scale parameters, **a**, indicates the level of analysis. Small values of **a** provide a local, fine grain or high frequency, analysis while large values corresponding to large scale, coarse grain or low-frequency, analysis. Changing the **b** parameter moves the time localization center of each wavelet.

Typically, the scale factor between levels increase by two. Thus, scaling is also known as **dilation**. Widely used **a** and **b** parameter setting that create an orthonormal basis are **a=** ٢ **ʲ** and **b =** ٢ **ʲ k** $(j,k \in Z)$. The wavelet family then becomes [٢٩]:

$$\Psi_{j,k}(x) = 2^{-\frac{j}{2}}.\Psi\left[2^{-j}x - k\right] \qquad (٣.٤)$$

A requirement on $\Psi[k]$ is that it is square inegrable and orthonormal (The same function is used in both, the transform and inverse).

Two functions are orthonormal. When they are inner products is:

$$\left\langle \Psi_{j,b}, \Psi_{j1,b1} \right\rangle = \int \Psi_{j,b} \Psi_{j1,b1} = \begin{cases} 1\, for j = j1, b = b1 \\ 0 otherwise \end{cases} \qquad (٣.٥)$$

If the mother wavelet satisfy these equations, it forms an orthonormal basis into which we may transform our function and an inverse transform may uniquely be found. This is not different from any other transform. However, by adding two extra requirements to the mother wavelet $\Psi[k]$ the wavelet functions get properties that make the transform useful for practical applications like signal processing.

The mother wavelet should have compact support, which means that only a finite number of terms are non-zero.

The second requirement is called the admissibility condition and implies:

$$\sum_{k} \Psi[k] = 0 \qquad (٣.٦)$$

i.e. the wavelet functions mean value is zero.

This simply means that the wavelet function is a zero mean function with compact support and hence exhibits some oscillatory behavior. It is this fact that has given the function its name: one small wave, wavelet.

Where functions that matched all these requirements were developed and the wavelet transform was put in the framework of multi-resolution analysis the importance of these wavelet bases became clear [٢٤].

٣.٣: The Discrete Wavelet Transform (DWT)

The wavelet transform calculates wavelet coefficients by taking the inner product of an input signal, $f(x)$, with a function, that is in this case the wavelet family, $\Psi_{j,k}(x)$. The discrete wavelet transform (DWT) is:

$$D_{j,k} = \left\langle f, \Psi_{j,k} \right\rangle = 2^{-\frac{j}{2}} . \int_{-\infty}^{\infty} f(x) \Psi(2^{-j} x - k) dx \qquad (٣.٧)$$

where $D_{j,k}$ are the **wavelet coefficients**. In wavelet vernacular, the wavelet coefficients are called details. From an intuitive perspective, the wavelet coefficients are measures of the goodness of fit between the signal and the wavelet [٢٩].

Wavelet function are constructed from a father wavelet, or scaling function, Φ. From the scaling function, Φ, it is possible to construct an orthonormal wavelet, Ψ, such that a signal can be decomposed (analyzed) and reconstructed exactly and efficiently [٣٠].The development of this relationship is briefly summarized here [٢٩] :

There is existing a **twin scale relation** (also know as the **dilation** or **refinement** equation) that relates **Multi Resolution Analysis, MRA,** functions at successive levels:

$$\Phi(x) = \sqrt{2} \sum_{k \in Z} h_k \Phi(2x - k) \qquad (٣.٨)$$

where $\Phi(x)$ is the scaling function and ***h***h_k is a square-summable sequence whose elements are obtained from the inner product of two levels of scaling functions:

$$h_k = \left\langle \Phi_{j+1,0}, \Phi_{j,k} \right\rangle \qquad (٣.٩)$$

The sequence **{** h_k **}**represents the coefficients of the scaling function filter. If scaling function is selected from one of the known family of wavelets, the scaling filter coefficients are known. The scaling filter is a low-pass, **FIR** filter. The filter has the properties of $\sum h_k = 1$ and normalization of $\sqrt{\sum h_k^2} = \frac{1}{\sqrt{2}}$.

Using the twin–scale relation and the MRA properties, a general equating for calculating the scaling function at any level j+١ given level j by the equation:

$$\Phi_{j+1,0}(x) = \sum_{k \in Z} h_k \Phi_{j,k}(x) \qquad (٣.١٠)$$

where $\Phi_{j,k}(x)$ is the scaling function at level **j** with translation index **k**, and $\Phi_{j+1,0}$ is the next lower level scaling function.

From the scaling function, the wavelet function, Ψ, is calculated as follows [٣١]:

$$\Psi(x) = \sqrt{2} \sum_{k \in Z} g_k \Phi_{j,k}(2x - k) \qquad (٣.١١)$$

where $\Psi(x)$ represents the mother wavelet (top most wavelet) and g_k represents the wavelet filter coefficients defined by :

$$g_k = (-1)^k h_{1-k} \qquad (٣.١٢)$$

Thus, the wavelet function is obtained by convolving the scaling function with the reversed and alternating signed form of the scaling filter. The wavelet calculated by Equation (٣.١١) that is orthogonal to the scaling function. The general equation for calculating the wavelet function at any level j is given by :

$$\Psi_{j+1,0}(x) = \sum_{k \in Z} g_k \Phi_{j,k}(x) \qquad (٣.١٣)$$

Figure (٣.٣) shows an example of wavelet and corresponding scaling function, specifically the example shows the Daubechies٤ **(db٤)** functions. As the figure shows, the wavelet function has high-frequency oscillations and the scaling function is lower in frequency. Thus, the wavelet function creates a high-pass wavelet filter **(** g_k **)** that provides the **detail coefficients**. The scaling function creates a low-pass wavelet filter **(** h_k **)** that provides the **approximation coefficients** [٢٩] .

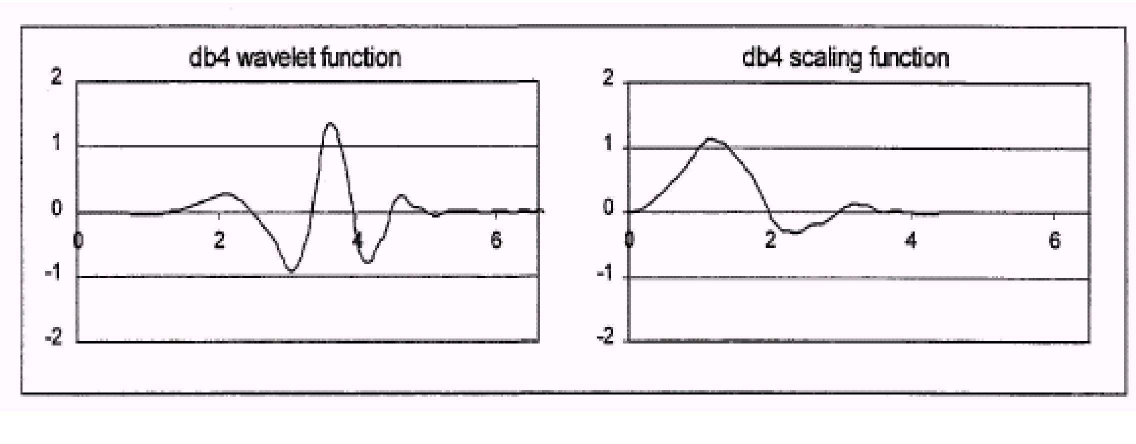


Figure (٣.٣): Daubechies٤ (db٤) Wavelet and Scaling Function.

Deriving the wavelet filter coefficients by Equation (٣.١٢) forces the wavelet and scaling filters to be Quadrature Mirror Filters (QMF) of each other and makes perfect signal reconstruction possible. The wavelet filter is the mirror reflection of the scaling filter with alternating signs. For example, if the scaling filter coefficients are $\mathbf{h_k} = \{\mathbf{a, b, c, d}\}$, then the wavelet filter coefficients are $\mathbf{g_k} = \{\mathbf{d, -c, b, -a}\}$ [٢٨][٣٠].

Recall from Equation (٣.٧) that convolving the input signal with the wavelet function creates the detail coefficients. The approximations coefficients are calculated in the same way, by taking the inner product of the signal, f , and the family of dilated, j, and translated, k, scaling functions:

$$A_{j,k} = \left\langle f, \Phi_{j,k} \right\rangle = 2^{-\frac{j}{2}} . \int_{-\infty}^{\infty} f(x) \Phi(2^{-j} x - k) dx \quad (٣.١٤)$$

Equation (٣.٧) and (٣.١٤) define the procedure for complete signal decomposition using wavelet. Although these equations can be implemented algorithmically and would proved accurate results, they don’t provide efficient signal decomposition [٢٩].

The next section describes an efficient decomposition procedure that uses convolution rather than integration for calculating the detail and approximation coefficients.

٣.٤: Efficient Wavelet Decomposition Algorithm

As described in the pervious sections, a signal is decomposed using the wavelet transform into two sets of coefficients called approximation $A_{j,k}$ and details $D_{j,k}$. The approximation coefficients represent the low frequency and the detail coefficients represent the high-frequency signal components. Calculating the detail and approximation coefficients through integration as shown in Equations ٣.٧ and ٣.١٤ is time consuming, especially as the decomposition algorithm is applied repeatedly to intermediate sets of coefficients (known as multi-level decomposition)[٢٩]. A more efficient algorithm results from calculating the ($A_{j,k}$) and ($D_{j,k}$) coefficients through convolution respectively. This recursive decomposition algorithm is sometimes referred to as the cascade algorithm [٢٨] or the pyramid algorithm [٢٧]. It is key to the fast wavelet algorithm.

Using the relation in Equation (٣.١٤) and the dilation Equation in (٣.١٠), an efficient decomposition algorithm for computing the $A_{j,k}$ coefficients is obtained [٣٢]:

$$A_{j+1,k} = \sum_{n} h_{n-2k} A_{j,n} \qquad (٣.١٥)$$

where j is the level (or scale), k is the translation index, and h_k is the scaling function filter coefficients as in Equation (٣.٩). This equation says that lower-level approximation coefficients $A_{j+1,k}$ are computed recursively given the approximation coefficients at a higher level A_j .

Similarly, a twin-scale relationship for computing the $D_{j,k}$ coefficients is obtained using the relation in Equation (٣.١٣) resulting in the decomposition formula:

$$D_{j+1,k} = \sum_{n} g_{n-2k} A_{j,n} \qquad (٣.١٦)$$

where g_k is the wavelet filter coefficients as defined in Equation (٣.١٢). this equation says that lower-level detail coefficients ($D_{j+1,k}$) are computed from higher-level approximation coefficients, A_j. Recursive application of these decomposition formulas provides a means for obtaining lower level detail and approximation coefficients once the highest level approximation coefficients are calculated. The input signal provides the top level (finest grain) approximation coefficients, A_0. Figure (٣.٤) Shows a pictorial of the decomposition process [٢٩].

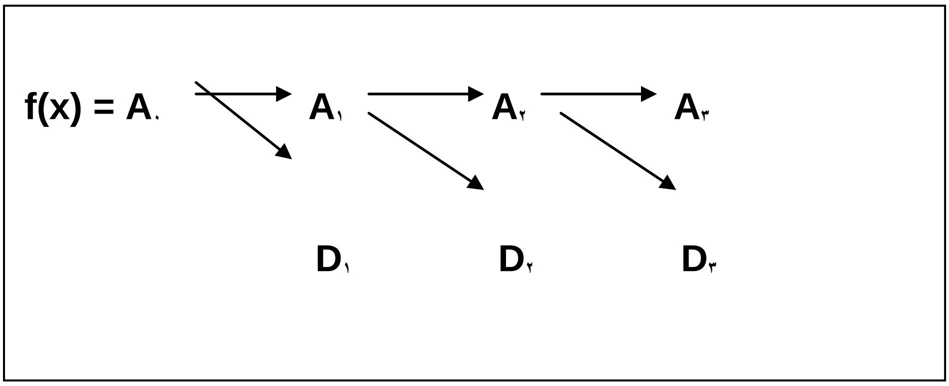
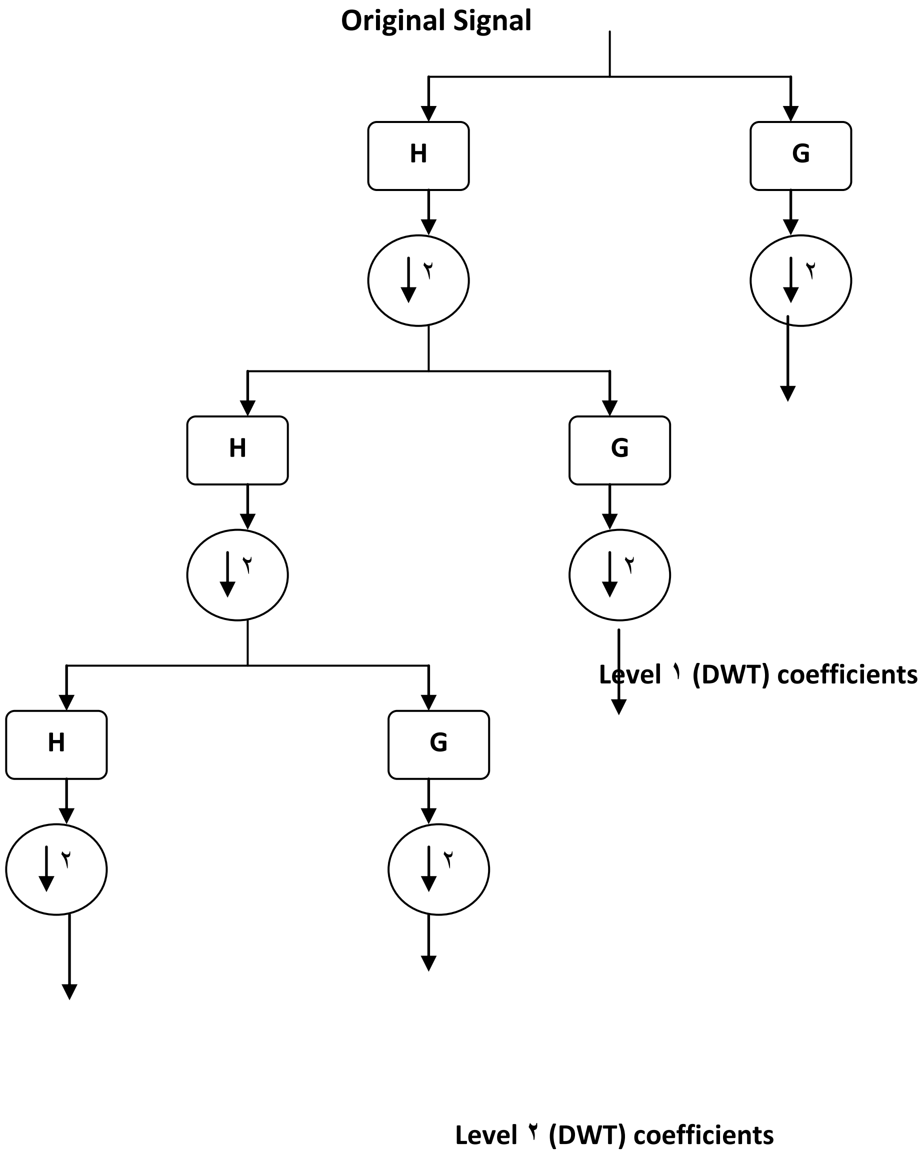


Figure (٣.٤): Schematic of wavelet Decomposition Algorithm. Lower Level Approximation (A_j) and Detail (D_j) Coefficients are Obtained From the Highest Level A٠ Coefficients. A٠ is the Input Signal f(x).

The result is a down-sampling of the coefficient vectors by a factor of two in the decomposition algorithm. The down-sampling and recursive natures of the algorithm are important components of the fast wavelet transform algorithm [٣٣].

Figure (٣.٥) shows DWT-Sub band coding scheme (or Filter Bank) with ٣ levels, where H and G denote the low pass and high pass filters respectively [٣٤]:



..... Level ٣ (DWT) coefficients

Figure (٣.٥): Filter Bank Tree of The DWT.

The decomposition of the signal into wavelet space is done by successive low-pass and high-pass filtering of the time domain signal and down-sampling the signal after each filtering. The filtering operation corresponds to the convolution of discrete signal with filter. Note that, only the low-frequency sub bands, while the high-frequency sub bands were untouched [٣٣][٣٥].

The high-frequency sub band at stage (level) ١ is termed level-١ detail coefficients. The process is repeated a given number of times on the low-frequency sub band at each level. The low-frequency sub band of the last stage called approximation coefficients. The combination of approximation coefficients and the detail coefficients of all levels is termed DWT coefficients [٣٦][٣٧].

٣.٥: Inverse Discrete Wavelet Transform

The wavelet decomposition algorithm is reversible and provides signal reconstruction. The inverse discrete wavelet transform (IDWT) provides exact signal reconstruction or synthesis. Lower level approximation and detail coefficients combine to create higher level coefficients. Figure (٣.٦) shows the reconstruction process [٢٩].

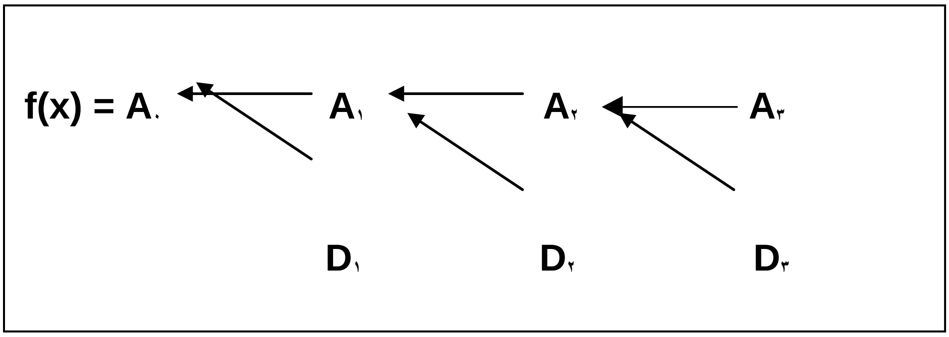


Figure (٣.٦): Schematic of wavelet Reconstruction Algorithm.

Lower Level Approximation (A_j) and Detail (D_j) Coefficients

Combine to Reconstruct Signal

The discrete wavelet reconstruction formula using the wavelet filter { g_k } and scaling filter { h_k } is as follows [٣٨]:

$$A_{j+1,k} = \sum_n h_{n-2k} A_{j,n} + g_{k-2m} D_{j+1,m} \qquad (٣.١٧)$$

Thus, the approximation coefficients ($A_{j,k}$) at any level can be computed from one set of low-level scaling function coefficients ($A_{j+1,m}$) and all the intermediate wavelet coefficients ($D_{j+1,k}$).

In order to proved prefect reconstruction, the **h** and **g** reconstruction filters are mirror images of the decomposition filters. For example, if the decomposition filter **H = {a, b, c, d}**, then the reconstruction filter **H'=h_K= {d, c, b, a}**.

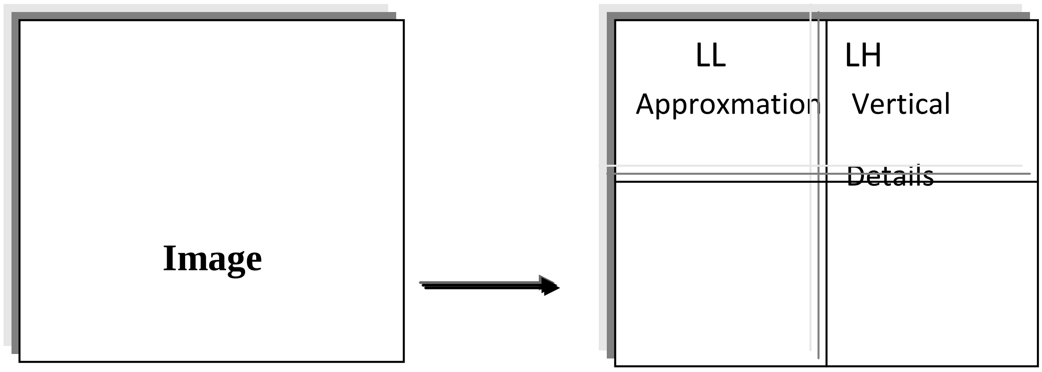
Consequently, the pair of reconstruction filters also **QMFs**. The coefficient vectors ***A*** and ***D*** are up-sampled (zeros instead at every other location) prior to convolution with the filters. This is analogous to the down-sampling operation in the decomposition process. The up-sampled and filtered coefficient vectors are then added together to create the next higher level $A_{j,k}$ vector. This process is repeated recursively to recreate the original input signal [٢٩][٣٧].

٣.٦: ٢D-Discrete Wavelet Transform

A ٢D separable discrete wavelet transform is equivalent to two consecutive ١D transforms. For an image, a ٢D is implemented as a ١D-row transform followed by a ١D-column transform. Transform coefficients are obtained by projecting the ٢D input image **x (u,v)** onto a set of ٢D basis functions that are expressed as the product of two ١D basis shown in the following equation:

$$\begin{aligned}
\Phi(u,v) &= \Phi(u)\Phi(v) \equiv LL_Band \\
\Psi_1(u,v) &= \Psi(u)\Phi(v) \equiv HL_Band \\
\Psi_2(u,v) &= \Phi(u)\Psi(v) \equiv LH_Band \\
\Psi_3(u,v) &= \Psi(u)\Psi(v) \equiv HH_Band
\end{aligned} \qquad (٣.١٨)$$

The ٢D-DWT (analysis) can be expressed as the set of equation (٣.١٨). The scaling function and wavelet functions corresponding to different sub-bands in the decomposition.



Figure(٣.٧):Single-level ٢D wavelet decomposition (Analysis).

The synthesis bank performs the **٢D-IDWT** to reconstruct **x(u,v)** $\in$ **V_m.**

A single stage of a ٢D-filter bank is shown in figure (٣.٨). First, the rows of the input image are filtered by the high pass and low pass filters. The outputs from these filters are down sampling by two, and then the columns of the outputs are filtered and down sampled to decompose the image into four sub-bands. The synthesis stage performs up-sampling and filtering to reconstruct the original image.

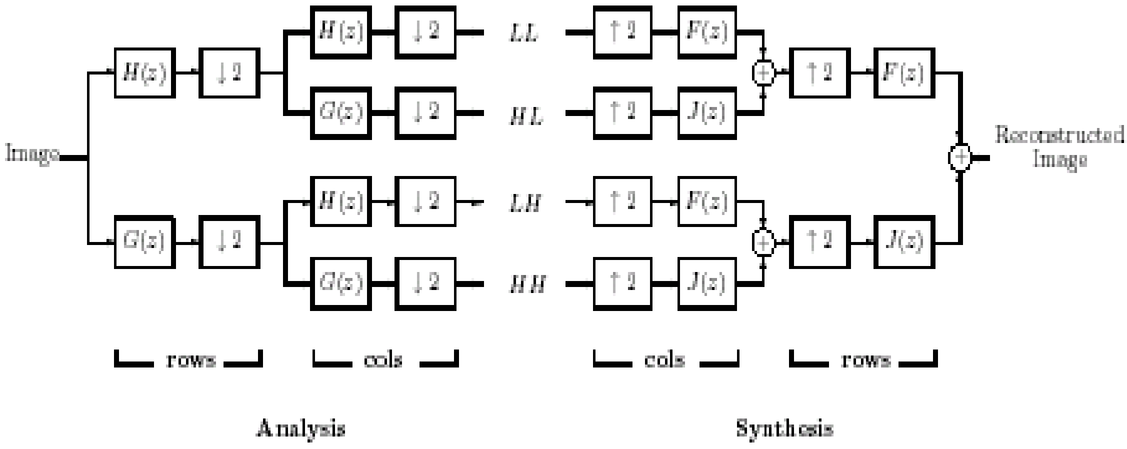


Figure ٣.٨: One level filter bank for computation of ٢D-DWT and IDWT.

Multiple levels of decomposition are achieved by iterating the analysis stage on only the LL band. For (**i**) levels of decomposition, the image is decomposed into **٣i+١** sub-bands [٣٧].

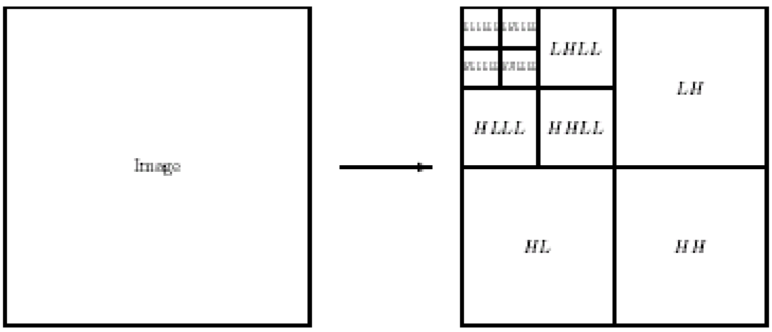


Figure ٣.٩: Three level wavelet decomposition of an image.

The importance of wavelet transform comes from the principle that wavelet transform coding decompose an image into multi-resolution sub-images. Each sub-image corresponds to a different frequency and energy distribution bands by using multiple scale wavelet bases, which results in flexible time and frequency resolution. These bands are high frequency band with little energy information (corresponding Details bands) and low frequency bands with high-energy information (corresponding Approximation bands). The energy for each band is computed according to the following equation[١٥]:

$$Energy = \sqrt{\frac{1}{NxN}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} f(n_1,n_2)^2} \qquad (٣.١٩)$$

Where :

f(n١,n٢): the sub-band values (coefficients).

٣.٧: Wavelet Families

The implementation of filter bank algorithm using a two channels perfect reconstruction filter bank. These filters have the following conditions[١٥]:

$$\sum_{n=0}^{N1-1} h(n) = \sqrt{2}$$
$$and \qquad (٣.٢٠)$$
$$\sum_{n=0}^{N1-1} h^2(n) = 1$$

Where :

h(n): Wavelet filters.

N١: Length of filter.

There are four standard families of filters:

٣.٧.١:Dabuchies Family:

Ingrid Daubechies, one of the brightest stars in the world of wavelet research, invented what are called compactly supported orthonormal wavelet, thus making discrete wavelet analysis practicable [٣٧].

The name of the Daubechies family wavelet are written dbN١, where N١ is the order, and **db** the “surname” of the wavelet, some authors use ***DM***, where ***M*** is the length of the filter which relates with the order via the following relation[٢٨][٣٥]:

N١ =M / ٢ (٣.٢١)

Therefore, db١ and D٢ is the same, the range of N١ is ١<N١<٤٥. The db١ wavelet is the same as Haar wavelet. Some Daubechies wavelets are shown in figure (٣.١٠)[٣٩].

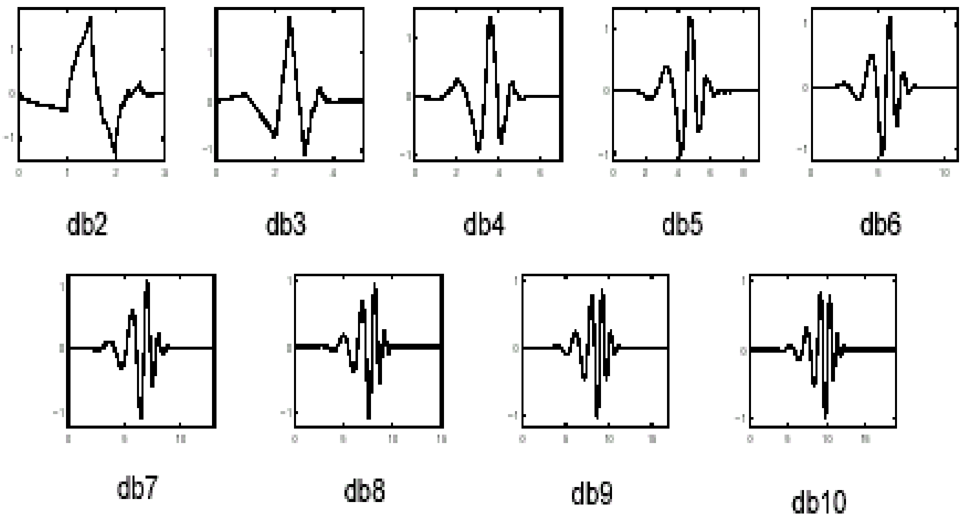


Figure (٣.١٠):Forms of Some Daubechies Wavelets.

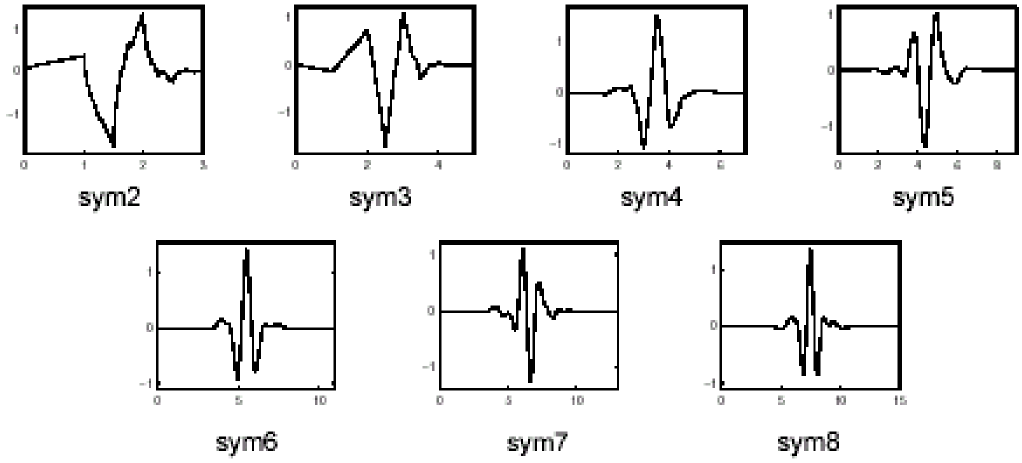
٣.٧.٢: Symlets wavelet

The Symlets are nearly symmetrical wavelets proposed by Daubechies as modifications to the db family. The properties of the two wavelet families are similar.

Referred to this family by (sym N١), where N١ is the order of the filter which has the a direct relation with the length of the filter as followes:

Filter length L = ٢N١ ***(٣.٢٢)***

Some Symlets wavelets are shown in figure (٣.١١)[٣٧].



Figure(٣.١١):Forms of Some Symlets Wavelets.

٣.٧.٣: Coiflets Filters

R. Coifman produced a new family of filters called coifelts filters, and referred to it by coifN١, where; N١ is the order of filter and the filter length gives by the following relation[١٥]:

Filter length = ٦N١. ***(٣.٢٣)***

Some of this wavelet family showed in figure (٣.١٢).

Figure(٣.١٢):Forms of Some Coiflets Wavelets.

٣.١٠.٤: Biorthogonal Filters

Biorthogonals use a pair of filters where one is used for decomposition and the other for reconstruction and the decomposition filters orthogonal on composition filters. Referred to it by **(bior d/r)**, where **(d)** represent the length of lowpass filter of decomposition stage, and **(r)** represent the length of filters of composition stage. Figure (٣.١٣) shows some of these wavelets[١٥].

Figure(٣.١٣):Forms of Some Biorthogonals Wavelets.