

Haar Wavelets, Image Compression, and Multi-Resolution Analysis

1/C Chris Eubanks
Mentor: Professor Popovici
Initial Report for Capstone Paper
SM422
April 4, 2007

Thesis: The purpose of this paper is to understand the implications of wavelet theory primarily through techniques of linear algebra with minimal emphasis on Fourier analysis. Much attention will be placed on the Haar wavelet (the simplest form of a wavelet) in one and two dimensions in order to understand its importance as a mathematical tool for hierarchically decomposing functions. Image compression and multi-resolution analysis will also be studied in order to fully understand wavelet applications.

Furthermore, on a more specific scale, given a function f , in W , we will identify elements in the nested subspaces that best approximate it in the L^2 norm. We are going to substitute the natural basis on W and its subspaces by one given by Haar wavelets. We will prove that in this new basis, the representation of f at various resolutions can be done by a very simple iteration process. Moreover, the reconstruction of the original function from the coefficients of this representation is equally simple and fast.

Introduction: Wavelets are at the forefront of both mathematics and engineering. By name, wavelets date back only to the 1980's and essentially provide an alternative to classical Fourier methods for both one and two dimensional data analysis and synthesis. Applications of wavelets are quite diverse and include astronomy, acoustics, nuclear engineering, sub-band coding, signal and image processing, neurophysiology, music, MRI, speech discrimination, optics, fractals, turbulence, earthquake prediction, radar, human vision, and solving partial differential equations. Ultimately, the emergence of wavelets compliments our fast paced, high speed, and information based lives.

A wave is usually referred to as an oscillating function of time and/or space. Fourier analysis is a type of wave analysis which expands signals in terms of sinusoids or complex exponentials. A wavelet, in a broad sense, is simply a small wave "with finite energy, which has its energy concentrated in time or space to give a tool for the analysis of transient, nonstationary, or time-varying phenomenon" (Reza 2). The wavelet still has the oscillating wavelike characteristics but it additionally has the ability to allow simultaneous time, or space, and frequency analysis.

The wavelet analysis procedure employs a wavelet prototype function, called mother wavelet. Temporal analysis is performed with a contracted high frequency version of mother-wavelet, while frequency analysis is performed with a dilated, low frequency

version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients. Smart choice of wavelet mother function results in wavelet basis yielding sparse data representation. In many cases wavelet coefficients truncated below a threshold. This sparse coding makes wavelets an excellent tool for data compression. (<http://www.smolensk.ru/user/sgma>)

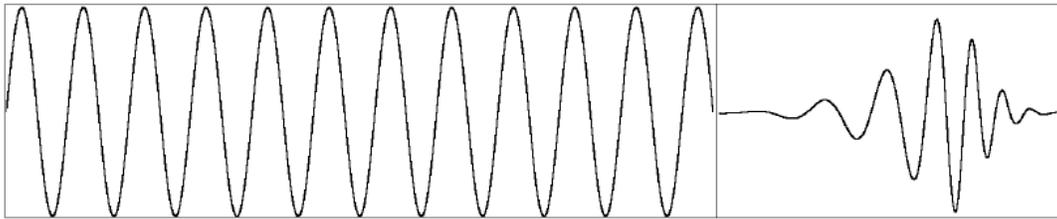


Figure 1- Comparison of a Wave and a Wavelet: Left graph is a Sine Wave with infinite energy and the right graph is a Wavelet with finite energy.

(Reza 2)

Before we hit the ground running, so to speak, with wavelets, it is imperative that a fundamental mathematical framework for wavelets be established first. The following will be a brief review of vector spaces, inner products and norms, and orthonormal bases, for each contribute greatly to wavelet theory.

A “vector space” is a set in which one can add two elements, and can multiply an element by a real number, in such a way that certain familiar properties hold. The following is more precise. A vector space is a set V (whose elements are called vectors) equipped with two binary operations, called vector addition and scalar multiplication.

If $x, y \in V$ there is an element $x+y$ in V called the vector sum of x and y . Also, if $\alpha \in \mathfrak{R}$ (the real numbers) and $x \in V$ there is an element αx in V , called the multiple of α and x . It is also important to note that if V is a vector space and we let $v_1, v_2, v_3, \dots, v_n$ be elements of V we say that a linear combination of $v_1, v_2, v_3, \dots, v_n$ is any vector of the form:

$$\sum_{j=1}^n \alpha_j v_j = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n, \text{ where } \alpha_1, \alpha_2, \dots, \alpha_n \in \mathfrak{R} \text{ (Bartle 52).}$$

Now we move onto inner products and norms. An inner product is a generalization of the dot product for vectors in \mathfrak{R}^n and it gives a generalized notion of perpendicularity, called

orthogonality. As we will see later on, orthogonal wavelets as well as localized basis are important in their contribution to data compression. If we let V be a vector space over C (the complex number field) then a complex inner product map $\langle \cdot, \cdot \rangle : V \times V \rightarrow C$ with the properties of additivity, scalar homogeneity, conjugate symmetry, and positive definiteness. A vector space V with a complex inner product is called a complex inner product space. (Frazier 80). Nevertheless, this study will use only the real domain.

We will take a quick look at the set of all square-summable real sequences such that

$$L^2_N = \left\{ \{x_j\}_{j=1}^{2^N} : x_j \in \mathfrak{R} \text{ for all } j, \text{ and } \sum_{j=1}^{2^N} |x_j|^2 < \infty \right\} .$$

With the obvious component-wise addition and scalar multiplication, L^2_N is a vector space over \mathfrak{R} where N consists of the natural numbers. N is also the number of resolutions that one could possibly obtain using wavelet theory.

The next two review topics will be norms and orthogonality (and what it means to be orthonormal). Letting V be a vector space over the real numbers with an inner product

$$\langle \cdot, \cdot \rangle \text{ and } v \in V \text{ we can define } \|v\| = \sqrt{\langle v, v \rangle} \text{ and } \langle u, v \rangle = \int_a^b u(x)v(x)dx .$$

For \mathfrak{R}^n this norm agrees with the usual notion of the length of a vector also known as the magnitude (Frazier 82). Some properties of the norm are as follows:

$$\|f + g\| \leq \|f\| + \|g\|$$

$$\|cf\| = |c| \|f\|$$

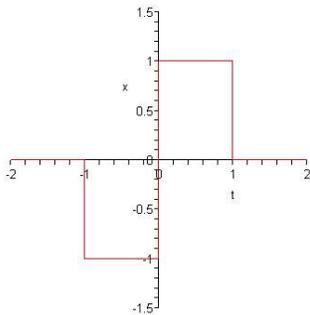
$$\|f\| = 0 \Leftrightarrow f \equiv 0$$

An extension of the norm is the distance between functions f and g which will come into play later when we try to find the error between two functions. By definition, the distance is $\|f - g\|$ and is denoted $d(f, g)$.

The definition of orthogonality is as follows. Suppose V is an inner product space. For $u, v \in V$, we say that u and v are orthogonal (written $u \perp v$) if $\langle u, v \rangle = 0$. Also, if we suppose that V is a real inner product space we can let B be a collection of vectors in V . B is an orthogonal set if any two different elements of B are orthogonal. B is an

orthonormal set if B is an orthogonal set and $\|v\|=1$ for all $v \in B$. Orthogonal sets of nonzero vectors are linearly independent (Frazier 84).

It is also important to note that up to this point, continuous functions have been used. Discontinuous functions are a bit harder to work with but many wavelets, such as the Haar wavelet, are discontinuous so it is important to note the concept of “almost equal.” Let f and g be piecewise continuous on $[a,b]$. We say that f and g are almost equal if there exists a finite set of points, S , such that for $x \in [a,b] \setminus S$. In the following case with the Haar wavelet, we are “ignoring” the points at $x=-1,0,1$. We will consider the set S the exception set. In summary, it is a fact that if f and g are almost equal on $[a,b]$ and if x_0 is a point of continuity for both f and g then $f(x_0) = g(x_0)$.



(constructed from Heavyside function from maple)

The following is a lemma contingent on what has just been discussed. (almost equal=a.e.)

If we let f,g,h be piecewise continuous on $[a,b]$ such that $f=g$ a.e. on $[a,b]$ and $g=h$ a.e. on $[a,b]$, then:

- (i) $f=h$ a.e.
- (ii) $f+g=g+h$ a.e.
- (iii) $f*g=f*h$ a.e.

Proof of (i): By definition of $f=g$ a.e. on $[a,b]$ there exists a finite set S_1 , such that for $x \in [a,b] \setminus S_1, f(x)=g(x)$. By definition of $g=h$, there exists a finite set S_2 such that for $x \in [a,b] \setminus S_2, g(x)=h(x)$. Now if we let $S_3 = S_1 \cup S_2$, it is considered to be a finite union since S_1, S_2 are finite. Let $x \in [a,b] \setminus S_3 \subset [a,b] \setminus S_2 \subset [a,b] \setminus S_1$. From this, $f(x)=g(x)$ since $x \in [a,b] \setminus S_1$ and $g(x)=h(x)$ since $x \in [a,b] \setminus S_2$. Therefore $f(x)=h(x)$. QED

Now that some of the basics are out of the way we turn our attention to wavelets using signal analysis as a mode of transition. When working in signal analysis, there are a number of different functions one can perform on that signal in order to translate it into different forms that are more suitable for different applications. The most popular function is the Fourier transform that converts a signal from time versus amplitude to frequency versus amplitude. This transform is useful for many applications, but it is not based in time. To combat this problem, mathematicians came up with the short term Fourier transform which can convert a signal to frequency versus time. Unfortunately, this transform also has its shortcomings mostly that it cannot get decent resolutions for both high and low frequencies at the same time.

The wavelet transform is a mechanism used to dissect or breakdown a signal into its constituent parts, thus enabling analysis of data in different frequency domains with each components resolution matched to its scale. Alternatively this may be seen as a decomposition of the signal into its set of basis functions (wavelets), analogous to the use of sines and cosines in Fourier analysis to represent other functions. These basis functions are obtained from dilations or contractions (scaling), and translations of the mother wavelet. The important difference that distinguishes the wavelet transform from Fourier analysis is its time and frequency localization properties. When analyzing signals of a non-stationary nature, it is often beneficial to be able to acquire a correlation between the time and frequency domains of a signal. In contrast to the Fourier transform, the wavelet transform allows exceptional localization in both the time domain via translations of the mother wavelet, and in the scale (frequency) domain via dilations.

Although the wavelet transform has come into prominence during the last decade, the founding principles behind wavelets can be traced back as far as 1909 when Alfred Haar discovered another orthonormal system of functions, such that for any continuous function $f(x)$, the series

$$f(x) = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} a_{2^j+k} w(2^j x - k), \quad \text{for } 0 \leq x < 1$$

converges to $f(x)$ uniformly over the interval $0 \leq x < 1$.

(www.wavelet.org/tutorial/gifs/haar.gif) Here k represents the translation of the wavelet

which is an indication of time and space. The integer k is also the center of the dyadic interval while j is referred to as the scale (dyadic in nature). W is the mother wavelet.

So how can a signal be converted and manipulated while keeping resolution across the entire signal yet still be based in time? This is where wavelets come into play. Wavelets are finite windows through which the signal can be viewed. In order to move the window about the length of the signal, the wavelets can be translated about time in addition to being compressed, widened, or shifted. The main problem that is being worked with in this paper is when we are given a function f which is piecewise continuous on $[a,b]$. We then want to approximate it with simpler functions (piecewise constants in our case) such that the approximation is the best possible, within the designated class of simple functions. Yet as we extend the class of simple functions, the approximations can be improved. Ultimately, it comes down to the concept of “reusing” the part of the previous approximation to rebuild the new, better approximation. So in order to increase the resolution of an image while still using the previous base, a better basis needs to be established. The basis we are looking for is of the wavelet type. The first component to wavelet and multiresolution analysis are vector spaces. The concept of high and low resolution comes into play when discussing vector spaces. If we start with the original space, the image is of low resolution and as we nest more vector spaces, the resolution continues to increase. This continues until one is satisfied with the compressed or approximated image or until there is no clear distinction between the raw image and what the wavelet basis has produced. The basis of each of these vector spaces is the scale function for the wavelet.

Aforementioned, the mathematical formulation of the signal-processing concept of high and low resolutions uses a sequence of nested subspaces of an inner product space of allowable functions (i.e. allowable signals). The functions we will be working with are either discrete signals or piecewise continuous functions, supported on the interval $J=[-1,1]$. Consider the dyadic subintervals of $[-1,1]$, i.e. the intervals whose end points are of the form $k/2^n$, for integer k and positive integer n , listed below by their “generation:”

$J_0=J$, generation zero

$J_1^0=[-1,0]$, $J_1^1=[0,1]$, first generation,

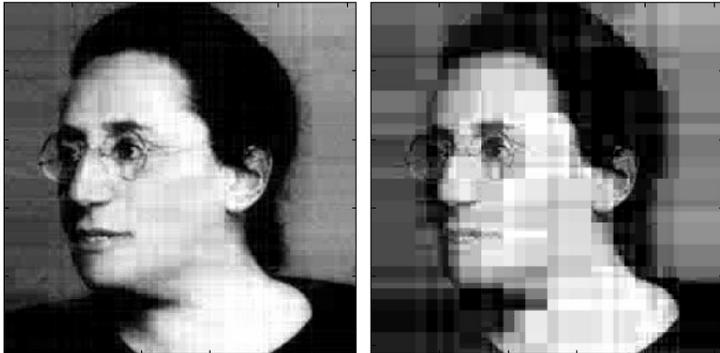
$J_2^0=[-1,-1/2]$, $J_2^1=[-1/2,0]$, $J_2^2=[0,1/2]$, $J_2^3=[1/2,1]$, second generation

...

In general; $J_n^k = [\frac{k}{2^{n-1}} - 1, \frac{k+1}{2^{n-1}} - 1]$, where $k=0$ to $k=2^n - 1$

For practical purposes one can think of an image as a vector space such as V_j which would be the perfectly normal image, and V_{j-1} would be that image at a lower resolution until you get to V_0 where you just have one pixel in the entire image. Considering the spaces just mentioned, V_0, V_1, \dots, V_n , with functions defined on $[-1,1]$ and piecewise on intervals of generations zero, generation one, generation n, respectively. In engineering language, these spaces are listed from low resolution to high resolution. These linear spaces have dimensions: $\dim V_0=1, \dim V_1=2, \dim V_2=4, \dots, \dim V_n=2^n$

For each vector space V_j , there is an orthogonal complement called W_j and the basis function for this vector space is the wavelet. (davis.wpi.edu)



(davis.wpi.edu)

Another theorem to be discussed is the following: let f be a piecewise continuous on $[a,b]$ and let $V_0([a,b])$ be the space of functions constant on $[a,b]$ (meaning $g \in V_0$ iff $\exists c \in \mathfrak{R}, g(x)=c$ for all x in $[a,b]$):

(i) Then there exists a unique function $f_0 \in V_0$ such that $\|f - f_0\| \leq \|f - g\|$ for all $g \in V_0$

(ii) The function f_0 is equal to $f(x)=$ average of f on $[a,b]$ for all $x \in [a,b]=\frac{\int_a^b f(t)dt}{b-a}$

The proof is as follows: Since function in V_0 are constant functions, all we need to show

is that $\forall c \in \mathfrak{R} \quad \|f - f_{ave}\|^2 \leq \|f - c\|^2$ where $f_{ave} = \frac{\int_a^b f(t)dt}{b-a}$. We thus want to show that

$\int_a^b |f(t) - f_{ave}|^2 dt \leq \int_a^b |f(t) - c|^2 dt$ (*). Consider the function $Q: \mathfrak{R} \rightarrow \mathfrak{R}$ defined by

$Q(c) = \int_a^b |f(t) - c|^2 dt$. Note that the left side of (*) represents $Q(f_{ave})$. We will show it is

a quadratic function, whose graph is a parabola. Its minimum is attained at $c = f_{ave}$ thus proving (*). Upon studying

$$Q(c) = \int_a^b (f^2(t) - 2f(t)c + c^2)dt = \left(\int_a^b f^2(t)dt\right) - 2c\left(\int_a^b f(t)dt\right) + c^2$$

So with this, $\frac{dQ}{dc} = -2\int_a^b f(t)dt + 2c(b-a)$. $\frac{dQ}{dc} = 0$ when $-2\int_a^b f(t)dt + 2c(b-a) = 0$ thus

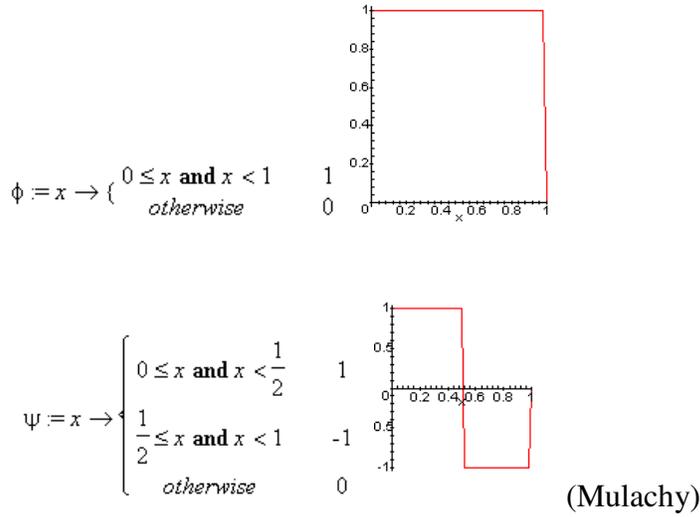
occurring when $c = \frac{\int_a^b f(t)dt}{b-a}$

Now we make an introduction to the Haar wavelet. Alfred Haar (1885-1933) was a Hungarian mathematician who worked in analysis studying orthogonal systems of functions, partial differential equations, Chebyshev approximations and linear inequalities. Although the wavelets discussed in this paper had their origins in the early work of Haar, the subject has only really gathered momentum in the last decade. The Haar measure, Haar wavelet, and Haar transform are named in his honor.



(O'Connor, Robertson 1)

The piecewise functions that make up the Haar wavelet and their graphs are as follows:



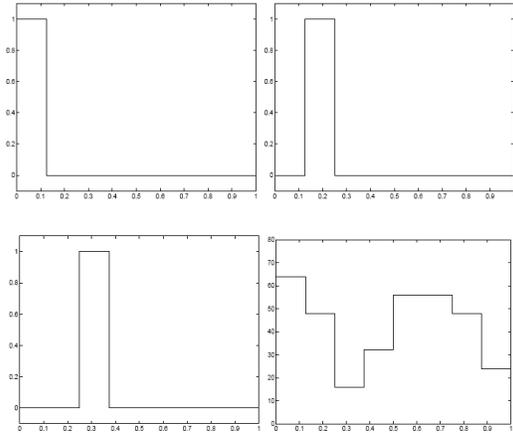
Given a function f , in W , we will identify elements in the nested subspaces that best approximate it in the L_2 norm. We are going to substitute the natural basis on W and its subspaces by one given by Haar wavelets. In this new basis, the representation of f at various resolutions can be done by a very simple iteration process. The Haar wavelet can be considered to be a special type of step function that can in turn be thought of as a linear combination of dyadically dilated and translated unit step functions on $[-1,1]$. It is important to note that the function ϕ satisfies a scaling equation of the form

$$\phi(x) = \sum_{i \in \mathbb{Z}} c_i \phi(2x - i) \text{ where in this case the only nonzero } c_i \text{'s are } c_0 = c_1 = 1, \text{ i.e.}$$

$$\phi(2x) + \phi(2x - 1).$$

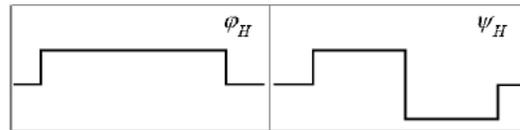
For each $0 \leq i \leq 2^3 - 1$, we get an induced (dyadically) dilated and translated scaling function: $\phi_i^3(x) = \phi(2^3 x - i)$. These eight functions form a basis for the vector space V^3 of piecewise constant functions on $[0,1)$ with possible breaks at $1/8, 2/8, 3/8, \dots, 7/8$. It is important to note that ϕ_0^3 is 1 on $[0, 1/8)$ only, ϕ_1^3 is 1 on $[1/8, 2/8)$ only, and so on. The following figure shows three of these basis functions together with a typical element of V^3 . The last plot shows the unique element of,

$$64\phi_0^3 + 48\phi_1^3 + 16\phi_2^3 + 32\phi_3^3 + 56\phi_4^3 + 56\phi_5^3 + 48\phi_6^3 + 24\phi_7^3 \in \mathcal{V}^3,$$



(Mulachy)

One of the reasons Haar wavelets are of preference when discussing wavelet theory, is not only because it is the simplest, but it is the only “orthogonal wavelet that has a symmetric analysis and synthesis filters” (Reza 4).



Left graph is the Haar scaling function and the right one is the Haar mother wavelet.

(Reza 4)

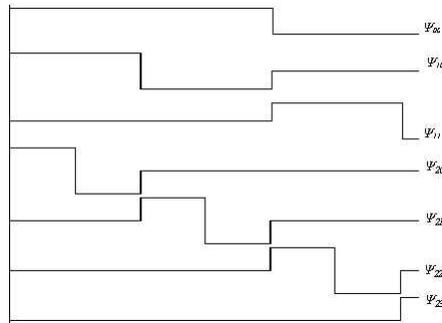


Fig.3. First 7 wavelets from Haar wavelet basis. (www.smolensk.ru/user/sgma)

Wavelet theory is based on analyzing signals to their components by using a set of basis functions. As seen above, one important characteristic of the wavelet basis functions is that they relate to each other by simple scaling and translation and that they recursively build upon each other. The original wavelet function, known as the mother wavelet (i.e. the Haar mother wavelet), which is generally designed based on some desired

characteristics associated to that function, is used to generate all basis functions. In most wavelet transform applications, it is required that the original signal be synthesized from the wavelet coefficients (the process of determining these coefficients will be discussed later). In general the goal of most modern wavelet research is to create a mother wavelet function that will give an “informative, efficient, and useful description of the signal of interest” (Reza 1). Some wavelets, simply based on their nature, are more suitable for particular applications than others. For example the Daubachy wavelet is suited for signals (such as noise signals) sharp spikes and peaks and in some cases discontinuities.

As an extension, the wavelet transform is a two-parameter expansion of a signal in terms of a particular wavelet basis functions or wavelets. If we let $\psi(t)$ represent the mother wavelet then all other wavelets are obtained by simple scaling and translation of $\psi(t)$ as follows: $\psi_{a,\tau}(t) = (1/\sqrt{a})\psi[(\frac{t-\tau}{a})]$. The scaling is discrete and dyadic meaning $a=2^{-j}$ and the translation is discretized with respect to each scale by using $\tau = k2^{-j}T$. In this case, the wavelet basis functions are obtained by $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - kT)$ where $k,j \in Z$ (integers). The integer k represents translation of the wavelet function and is an indication of time or space in wavelet transform. Integer j , on the other hand, is an indication of the wavelet frequency or spectrum shift. It is referred to as the scale. The following are two different scaled versions of a wavelet along with the mother wavelet.

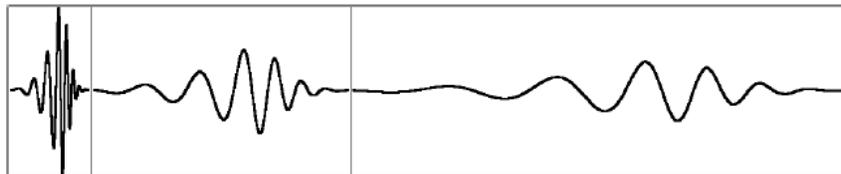


Figure 2- The left graph is the mother wavelet ψ_{D12} , the middle one is the wavelet at scale $j = -1$ and the right one is the wavelet at scale $j = -2$. The other way to look at these graphs is: to assume that the right graph is the mother wavelet, the middle one is the wavelet at scale $j = 1$ and the left one is the wavelet at scale $j = 2$.

(Reza 3)

Another aspect of the wavelet transform is that the localization or compactness of the wavelet increases as frequency or scale increases. In other words, higher scale corresponds to finer localization and vice versa (Reza 3). The paper titled *Wavelet Characteristics: What Wavelet should I Use*, written by Ali M. Reza provides elegant

illustrations of the fundamentals of wavelets as well as multi-resolution analysis. She states that the multiresolution formulation needs two closely related basic functions. “In addition to the wavelet $\psi(t)$, there is a need for another basic function called the scaling function which is denoted $\phi(t)$.” The two-parameter wavelet expansion for a signal designated as $x(t)$ is given by the following decomposition series:

$$x(t) = c_k \phi_{j_0,k}(t) + \sum_k \sum_{n=0}^N d_{n,k} \Psi_{n,k}(t)$$

The coefficients you see, c_k , are referred to as approximation coefficients at scale 0, and the set of $d_{n,k}$ considered to be detail coefficients. The relationship of these wavelet coefficients to the original signal (for real and orthogonal wavelets) are given as:

$$d_{n,k} = \int x(t) \Psi_{j,k}(t) dt$$

$$c_k = \int x(t) \phi_{j_0,k}(t) dt$$

Other examples of newly developed wavelets are shown in the following figure:

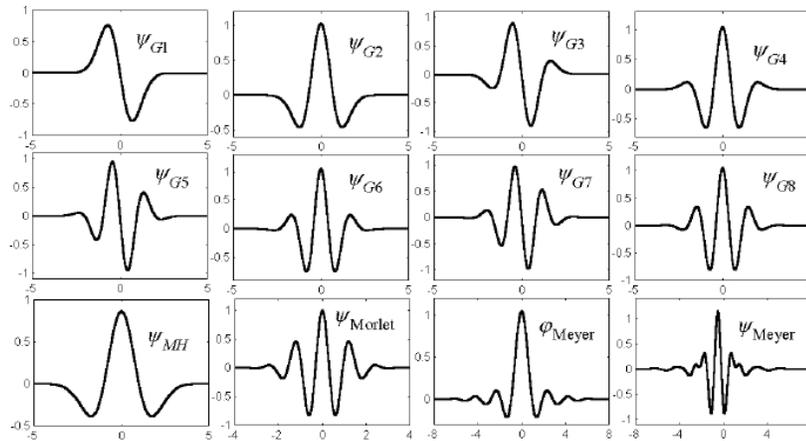
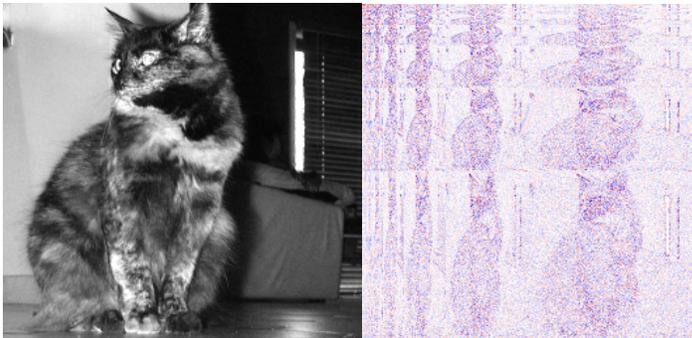


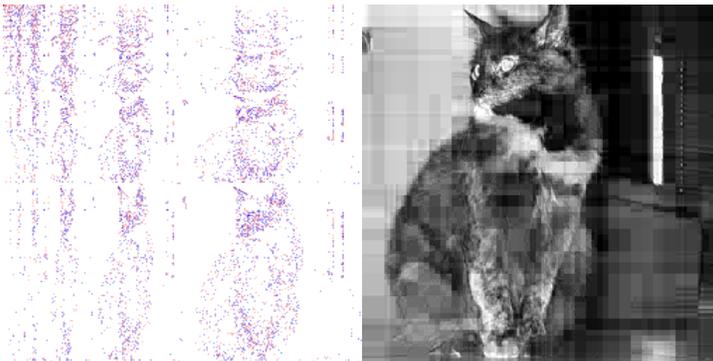
Figure 4- Different Gaussian wavelets obtained from derivatives of the Gaussian function along with Mexican Hat wavelet, Morlet wavelet and Meyer scaling function and wavelet. The order of the derivatives for Gaussian wavelets are shown as subscript for these wavelets.

The following “images” of a cat show the fundamental transition between an original image and the final compressed image. It begins with substituting the normal basis for the image with the Haar wavelet basis. From this, we then find the coefficients that best approximate the original image through the means of the basis. Then the coefficients are quantized, or averaged. This is where the true concept of image compression comes into play. The goal is to drop or filter the largest amount of coefficients possible yet still be able to create a reconstructed image that is represented by fewer coefficients still similar to the raw image. The images show how the the Haar transform can be used in lossy image compression:



1. Original image

2. Haar coefficients



3. Quantized coefficients

4. Reconstructed image

It is important to understand that wavelets are used in a series expansion of signals or functions much the same way a Fourier series used the wave or sinusoid to represent a signal or function. The big distinction between Fourier analysis and wavelet analysis is that for the Fourier series, sinusoids are chosen as basis functions, then the properties of the resulting expansion are examined. For wavelet analysis, one poses the desired properties and then derives the resulting basis functions (Burrus xi).

Additionally, in wavelet analysis, the scale that is used to look at data is of particular

importance. Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large “window,” we would notice gross features. Similarly, if we look at a signal with a small “window,” we would notice detailed features. The result in wavelet analysis is to see both the forest and the trees and be able to transfer this data quickly and efficiently. This is why wavelet analysis along with data or image compression is so unique and beneficial in our fast paced lives.

References

- Bartle, Robert G., *The Elements of Real Analysis*, 2nd edition, John Wiley & Sons, New York NY, 1976.
- Burrus, C. Sidney, *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice Hall, New Jersey. 1998.
- Fomitchev, Max. *An Intro to Wavelets and Wavelet Transforms*, <http://www.smolensk.ru/user/sgma>. 1998.
- Frazier, Michael W. *An Introduction to Wavelets Through Linear Algebra*. New York, 1991.
- Germano, Tom, *Wavelets in Multiresolution Analysis*, <http://davis.wpi.edu/~matt/courses/wavelets/>, February 15, 1999.
- J J O'Connor and E F Robertson. *Haar Biography*. <http://www-history.mcs.st-andrews.ac.uk/Biographies/Haar.html>, 2000.
- Mikulic, Emil. *How to do a Haar Transform*. <http://dmr.ath.cx/gfx/haar/>, 2001.
- Mulachy, Colm, *Plotting and Scheming with Wavelets*, <http://www.spelman.edu/~colm/wav.pdf>
- Resnikoff, L. Howard, Raymond O. Wells, *Wavelet Analysis: The Scalable Structure of Information*. Springer-Verlag New York, Inc. 1998.
- Reza, Ali M., *Wavelet Characteristics: What Wavelet Should I Use?* http://www.xilinx.com/products/logicore/dsp/wavlet_char.pdf, October 19, 1999.
- Stollnitz, Eric J., Tony D. DeRose, David H. Salesin, *Wavelets for Computer Graphics: A Primer*. <http://grail.cs.washington.edu/projects/wavelets/article/>.
- Walker, James. *Fast Fourier Transforms*, 2nd edition, CRC Press, 1996.
- www.wavelet.org/tutorial/gifs/haar.gif

