Time–Frequency Analysis of Biosignals

A Wavelet Transform Overview

BY PAUL S. ADDISON, JAMES WALKER, AND RODRIGO C. GUIDO

The wavelet transform has emerged over recent years as a powerful time–frequency analysis and signal-coding tool suitable for use in the manipulation of complex nonstationary signals. It has been at the forefront of many recent developments in biomedical and biological signal processing, where it has been particularly useful in the study of numerous problematic signals. This overview article covers the emerging role of the wavelet transform in detail, where both continuous and discrete transforms are considered in turn, promoting new methods that may prove useful for further research in this area.

As a result of infinite extent of Fourier integral, analysis is time averaged. Thus, it contains only globally averaged information and so has the potential to obscure transient or location-specific features within the signal. This limitation can be partly overcome by introducing a sliding time window of fixed length to localize the analysis in time. This local or short-time Fourier transform (STFT) provides a degree of temporal resolution by highlighting changes in spectral response with respect to time. A number of alternative time–frequency and timescale methods are now available for signal analysis. Of these, the wavelet transform has emerged over recent years as the most favored tool by researchers for analyzing problematic signals across a wide variety of areas in science, engineering, and medicine [3], [25], [28], [47], [48]. It is especially valuable because of its ability to elucidate simultaneously local spectral and temporal information from a signal in a more flexible way than the STFT by employing a window of variable width. Thus, wavelet transforms produce a time–frequency decomposition of a signal over a range of characteristic frequencies that separates individual signal components more effectively than do STFT. This flexible temporal–spectral aspect of the transform allows a local scale-dependent spectral analysis of individual signal features. In this way, both short-duration high-frequency and long-duration low-frequency information can be captured simultaneously. Hence, the method is particularly useful for the analysis of transients, aperiodicity, and other nonstationary signal features, where, through the interrogation of the transform, subtle changes in signal morphology may be highlighted over the scales of interest. Another key advantage of wavelet techniques is the variety of wavelet functions available, thus allowing the most appropriate to be chosen for the signal under investigation. This is in contrast to Fourier analysis, which is restricted to one-feature morphology: the sinusoid. In its discrete form using orthogonal wavelet bases, the wavelet transform is particularly useful in signal coding, allowing information within the signal to be localized within a number of pertinent coefficients for compression purposes. Wavelet-transform analysis has now been applied to a wide variety of biomedical signals, including electroencephalogram (EEG), electrocardiogram (ECG), clinical sounds, respiratory patterns, blood pressure trends, and DNA sequences [5], [20], [30], [36], [42], [44].

This article provides an overview of the emerging role of wavelet-transform analysis in biomedical signal processing and analysis. It also provides a brief overview of the theory of the transform in its two distinct and very different forms: continuous and discrete.

The Wavelet Transform

Time–frequency signal analysis methods offer simultaneous interpretation of the signal in both time and frequency, which allows local, transient, or intermittent components to be elucidated. Such components are often obscured because of the averaging that is inherent within spectral-only methods, i.e., the fast Fourier transform (FFT). A number of time–frequency methods are currently available for the high-resolution decomposition in time–frequency plane useful for signal analysis, including the STFT, Wigner-Ville transform (WVT), Choi-Williams distribution (CWD), and the continuous wavelet transform (CWT). Of these, the CWT has emerged as the most favored tool by researchers as it does not contain the cross terms inherent in the WVT and CWD methods while possessing frequency-dependent windowing, which allows for arbitrarily high resolution of the high-frequency signal components, unlike the STFT.

Many of the ideas behind wavelet transforms have been in existence for a long time. However, wavelet-transform analysis really began in the mid-1980s, where they were developed to interrogate seismic signals [27]. Interest in wavelet analysis remained within a small, mainly, mathematical community during the rest of the 1980s, with only a handful of scientific

Digital Object Identifier 10.1109/MEMB.2009.934244

 0739-5175/09/$26.00©2009IEEE

The CWT is a time–frequency analysis method that differs from the STFT by allowing high localization in time of high-frequency signal features.

papers coming out each year. The application of wavelet-transform analysis in science and engineering really began to take off at the beginning of the 1990s, with a rapid growth in the number of researchers turning their attention to wavelet analysis during that decade. The last few years have each seen the publication of more than 1,000 refereed journal articles concerning applications of the wavelet transform, and these covering many disciplines.

Wavelet transforms, as they are in use today, come in essentially two distinct varieties or classes: the CWT and the discrete wavelet transform (DWT).

The Continuous Wavelet Transform

The CWT is a time–frequency analysis method that differs from the STFT by allowing high localization in time of high-frequency signal features. The CWT does this by having a variable window width, which is related to the scale of observation, a flexibility that allows for the isolation of the high-frequency features. Another important distinction from the STFT is that the CWT is not limited to using sinusoidal analyzing functions. Rather, a large selection of localized waveforms can be employed as long as they satisfy predefined mathematical criteria described later. The wavelet transform of a continuous time signal, $x(t)$, is defined as

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt, \tag{1}$$

where $\psi^*(t)$ is the complex conjugate of the analyzing wavelet function $\psi(t)$, a is the dilation parameter of the wavelet, and b is its location parameter. To be classified as a wavelet, a function must satisfy certain mathematical criteria. These are as follows:

1) It must have finite energy:

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty. \tag{2}$$

2) If $\hat{\psi}(f)$ is the Fourier transform of $\psi(t)$, i.e.,:

$$\hat{\psi}(f) = \int_{-\infty}^{\infty} \psi(t) e^{-i(2\pi f)t} dt, \tag{3}$$

then the following condition must hold:

$$C_g = \int_0^{\infty} \frac{\left| \hat{\psi}(f) \right|^2}{f} df < \infty. \tag{4}$$

This implies that the wavelet has no zero-frequency component, i.e., $\hat{\psi}(0) = 0$, or to put it another way, it must have a zero mean. Equation (4) is known as the *admissibility condition* and C_g is called the *admissibility constant*. The value of C_g depends on the chosen wavelet.

3) For complex (or analytic) wavelets, the Fourier transform must both be real and should vanish for negative frequencies.

The contribution to the signal energy at the specific a scale and b location is given by the two-dimensional (2-D) wavelet energy density function known as the *scalogram*, analogous to the *spectrogram*, the energy density surface of the STFT

$$E(a,b) = |T(a,b)|^2. \tag{5}$$

In practice, all functions that differ from $|T(a,b)|^2$ by only a constant multiplicative factor are also called *scalograms*, e.g., $|T(a,b)|^2/C_g$, $|T(a,b)|^2/C_g f_c$, etc., where f_c is a characteristic frequency of the wavelet function. The scalogram can be integrated across a and b to recover the total energy in the signal using the admissibility constant, C_g, as follows:

$$E = \frac{1}{C_g} \int_{-\infty}^{+\infty} \int_0^{\infty} |T(a,b)|^2 \frac{da}{a^2} db \quad \left[= \int_{-\infty}^{+\infty} x(t)^2 dt \right]. \tag{6}$$

The relative contribution to the total energy contained within the signal at a specific a scale is given by the scale-dependent energy distribution:

$$E(a) = \frac{1}{C_g} \int_{-\infty}^{\infty} |T(a,b)|^2 db. \tag{7}$$

Peaks in $E(a)$ highlight the dominant energetic scales within the signal. We may convert the scale-dependent wavelet energy spectrum of the signal, $E(a)$, to a frequency-dependent wavelet energy spectrum $E_W(f)$ to compare directly with the Fourier energy spectrum of the signal $E_F(f)$. To do this, we must convert from the wavelet a scale, which can be interpreted as a representative temporal, or spatial, period for physical data, to a characteristic frequency of the wavelet, such as the spectral peak frequency, the passband center, and the central frequency. The spectral components are inversely proportional to the dilation, i.e., $f \propto 1/a$. The frequency associated with a wavelet of arbitrary a scale is given by

$$f = \frac{f_c}{a}, \tag{8}$$

where the characteristic frequency of the mother wavelet, i.e., the archetypal wavelet at scale $a = 1$ and location $b = 0$, f_c, becomes a scaling constant, and f is the representative or frequency for the wavelet at arbitrary scale a. As we see from

The Morlet wavelet is one of the most popular complex wavelets used.

(8), each scale of the wavelet corresponds to a unique characteristic frequency. There is a very important distinction to be made here. The characteristic frequency f_c of the wavelet used in the wavelet transform is representative of the whole frequency makeup of the wavelet. The wavelet does not contain a single frequency, and the signal is not decomposed according to numerous single (sinusoidal) frequencies; this is not Fourier analysis!

Finally, as with the Fourier transform, the original signal may be reconstructed using an inverse transform

$$x(t) = \frac{1}{C_g}\int_{-\infty}^{\infty}\int_{0}^{\infty} T(a,b)\,\psi_{a,b}(t)\;\frac{da\,db}{a^2}. \tag{9}$$

In practice, a fine discretization of the CWT is computed, where, usually, the b location is discretized at the sampling interval and the a scale is discretized logarithmically.

As the wavelet transform given by (1) is a convolution of the signal with a wavelet function we can use the convolution theorem to express the integral as a product in Fourier space, i.e.,

$$T(a,b) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{x}(\omega)\,\hat{\psi}^*_{a,b}(\omega)\,d\omega,$$

where

$$\hat{\psi}^*_{a,b}(\omega) = \sqrt{a}\,\hat{\psi}^*(a\omega)\,e^{i\omega b} \tag{10}$$

is the Fourier spectrum of the analyzing wavelet at scale a and location b. In this way, the FFT algorithm can be employed to speed up the computation of the wavelet transform.

For its practical implementation, the CWT is computed over a finely discretized time–frequency grid. This discretization involves an approximation of the transform integral, i.e., a summation, computed on a discrete grid of a scales and b locations. In general, the wavelet transform is approximated in this way over each time step for a range of wavelet scales, there is therefore a heavy computational burden involved in the generation of the CWT. A vast amount of repeated information is contained within this redundant representation of the CWT $T(a,b)$. This can be considerably condensed by considering only local maxima and minima of the transform. Two definitions of these maxima are commonly used in wavelet analysis:

1) Wavelet ridges defined as

$$\frac{d\left(\left|T(a,b)^2\right|/a\right)}{da} = 0 \tag{11}$$

are used for determining instantaneous frequencies and amplitudes of signal components [11], [17]. Notice that this definition of a ridge uses the rescaled scalogram, $S(a, b) = |T(a, b)|^2/a$, as it leads to a simpler analytical solution relating the ridge locus to the instantaneous frequency. In addition, the amplitude of the ridge can be used to compute the amplitude of the instantaneous frequency component. Further, it has been shown recently how secondary wavelet transforms of ridges can be used to provide information on the frequency and amplitude modulation of the primary signal components [3]. The method, called *secondary wavelet feature decoupling* (SWFD), requires neither the primary signal nor the secondary modulations to be strictly stationary in nature.

2) Wavelet modulus maxima defined as

$$\frac{d|T(a,b)|^2}{db} = 0 \tag{12}$$

are used for locating and characterizing singularities in the signal [9], [35]. Note that (12) also includes inflection points with zero gradient. These can be easily removed when implementing the modulus maxima method.

There are many continuous wavelets; however, by far, the most popular are the Mexican hat wavelet and the Morlet wavelet.

The Mexican Hat Wavelet

The Mexican hat wavelet is the second derivative of a Gaussian function given by

$$\psi(t) = (1 - t^2)e^{-\frac{t^2}{2}}. \tag{13}$$

This wavelet, shown in Figure 1(b), has been used in practice for a number of data-analysis tasks in science and engineering, including the morphological characterization of engineering surfaces [37], the interrogation of laser-induced ultrasonic signals used to measure stiffness coefficients in a viscoelastic composite material [29], and the analysis of turbulent flows (e.g., [2], [14], [33]). In addition, the Mexican hat is used extensively in studies requiring the use of modulus maxima methods as its maxima lines, and those of all other derivatives of Gaussian functions are guaranteed continuous across scales for singularities in the signal [41]. An example of a Mexican hat-based wavelet transform of a chirp signal is shown in Figure 1(c). The increase in frequency of the components through time in the time–frequency plane is evident in the plot. Figure 2 shows the wavelet transform of an exponential discontinuity, a sudden spike in the signal halfway along its length followed by a smooth exponential decay. As the transform plot has been orientated with the smallest scales at the top, it points to the signal discontinuity in the signal.

A complex version of the Mexican hat function can easily be constructed by simply setting the negative part of its Fourier frequency spectrum to zero before performing an inverse Fourier transform to get the analytic version of the Mexican hat shown [4]. However, in practice, the Morlet wavelet is used when a complex wavelet function is required. Note that the Mexican hat normally used, i.e., that given by (13) and shown in Figure 1(b), is actually the negative of the second derivative of the Gaussian function. All derivatives of the Gaussian function may be employed as a wavelet. The most appropriate one depends on the application. The first and second derivatives of the Gaussian are shown in Figure 1(a) and (b), respectively. Higher order derivatives are less commonplace.

The Morlet Wavelet

The Morlet wavelet is one of the most popular complex wavelets used. The complete Morlet wavelet is defined as

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}} \left(e^{i\omega_o t} - e^{-\frac{\omega_o^2}{2}} \right) e^{-\frac{t^2}{2}}, \tag{14}$$

where ω_o is the central frequency of the mother wavelet. The central frequency of the wavelet is so called, because it is the peak frequency of a Gaussian distribution of components at a range of frequencies of the Morlet wavelet, i.e., it is at the center of the Gaussian. The second term in the brackets is known as the correction term, as it corrects the nonzero mean of the complex sinusoid of the first term. In practice, it becomes negligible for values of $\omega_o > 5$. Previous investigators have concentrated on wavelet transforms with ω_o in the range of 5–6, where it can be performed without the correction term, since it becomes very small. In this case, the Morlet wavelet becomes

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}} e^{i\omega_o t} \; e^{-\frac{t^2}{2}}. \tag{15}$$

This truncated Morlet wavelet is almost invariably used in literature and is often referred to as *Morlet wavelet*. It has been shown that lowering of ω_o below 5 allows an interrogation that is more temporal than spectral, which can be useful for some data-analysis tasks [4]. Two examples of Morlet wavelets are shown in Figure 3.

Figure 4(a) contains the same chirp signal as shown in Figure 1. The Morlet wavelet with $\omega_o = 5.33$ rad/s (i.e., $f_0 = 0.849$ Hz) was used to transform the signal. The real part of the transformed signal is plotted in Figure 4(b) and has similarities with the Mexican hat transform plot in Figure 1(c). The discontinuities at the beginning and end of the chirp segment are picked up in the phase plot of Figure 4(c). These are located using arrows at the top of the phase plot. The continuous increase in instantaneous frequency associated with the chirp is highlighted in the modulus plot of Figure 4(d). The instantaneous frequency associated with a signal component can be found from its wavelet transform ridges. These are the maxima found in the rescaled wavelet-transform scalogram. The ridge associated with the chirp signal is shown schematically in Figure 4(e), where the instantaneous scale a_R at time b_R can be used to find the instantaneous frequency $f_R = f_0/a_R$. The instantaneous amplitude and phase can also be found from the ridge. Further, if we plot the rescaled scalogram in terms

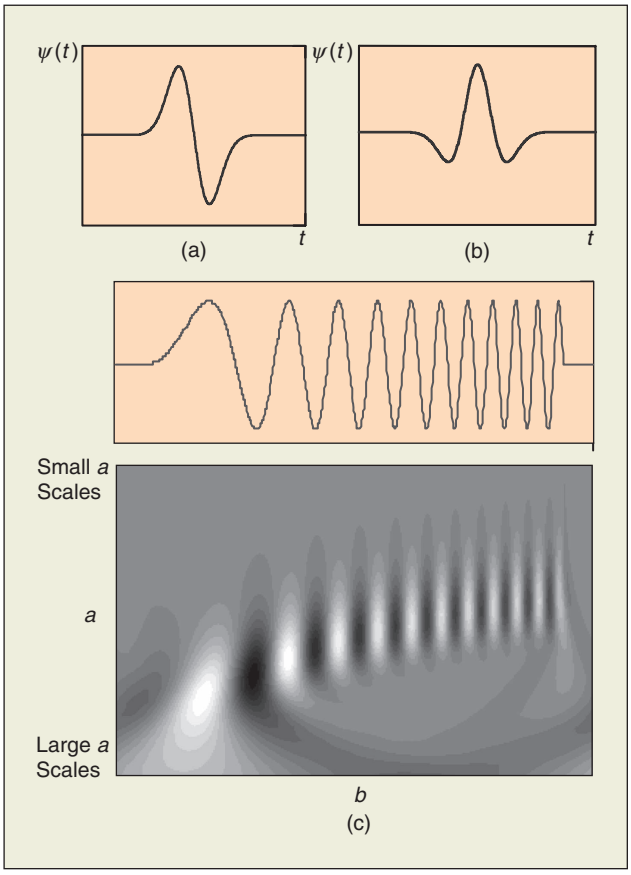

Fig. 1. Segment of a chirp signal with associated transform plot. (a) Gaussian wave (first derivative of a Gaussian). (b) Mexican hat (second derivative of a Gaussian). (c) Mexican hat-based transform of chirp signal (small/large *a* scales correspond to high-/low-frequency components, respectively). (Image reprinted from (59). © 2002 by Taylor & Francis Group LLC—Books. Reproduced with permission in the format Journal via Copyright Clearance Center.)

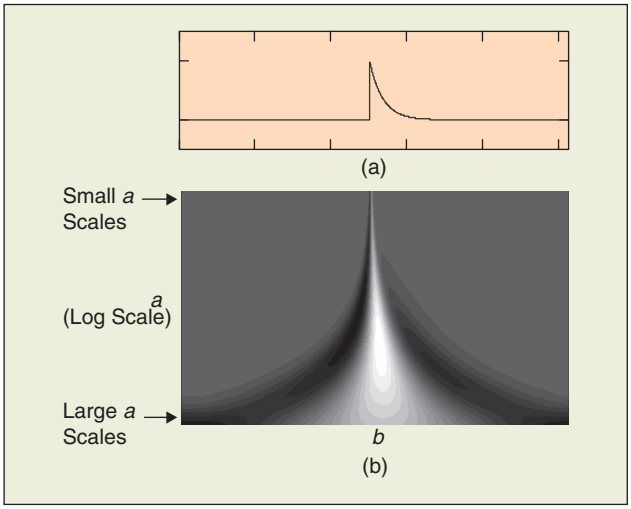

Fig. 2. Pointing to an exponential discontinuity. (a) A sudden spike with an exponential tail. (b) The transform plot for the discontinuity. (Image reprinted from (59). © 2002 by Taylor & Francis Group LLC—Books. Reproduced with permission in the format Journal via Copyright Clearance Center.)

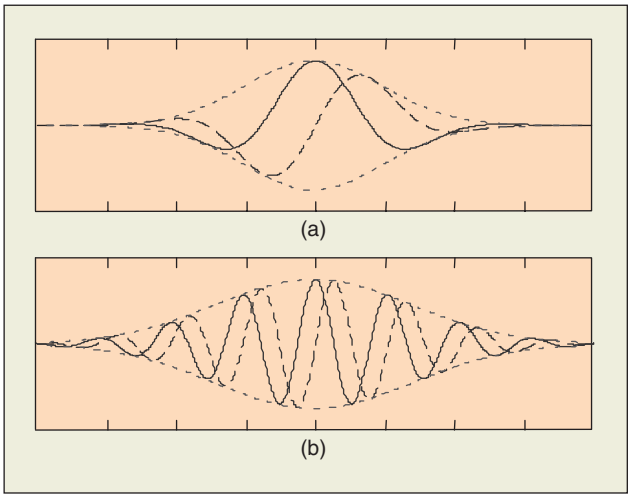

Fig. 3. Two Morlet wavelets: (a) $\omega_0 = 2.0$ ($f_0 = 0.318$) and (b) $\omega_0 = 12$ ($f_0 = 1.909$).

of a characteristic wavelet frequency, where $f = f_0/a$, then the instantaneous frequency can be read directly from this plot. The instantaneous frequency at time b_R can be found from a_R. We can see the relationship between maxima in the rescaled scalogram and instantaneous frequency by using a Morlet wavelet. Then, using a change of variable $t' = (t-b)/a$, it can be shown that maxima in the rescaled scalogram correspond to the instantaneous frequencies through their associated scales.

The standard Morlet wavelet has a form very similar to the analyzing function used for the STFT employing a Gaussian window, sometimes called a *Gabor transform*. The important difference is that, for the Morlet wavelet transform, the window and enclosed sinusoid are scaled together, whereas for the STFT, we keep the window length constant and scale only the enclosed sinusoid. The wavelet can therefore localize itself in time for a short duration, i.e., high-frequency fluctuations. There is, however, an associated spreading of the frequency distribution associated with wavelets of short duration. Conversely, there is a spreading in temporal resolution at low frequencies.

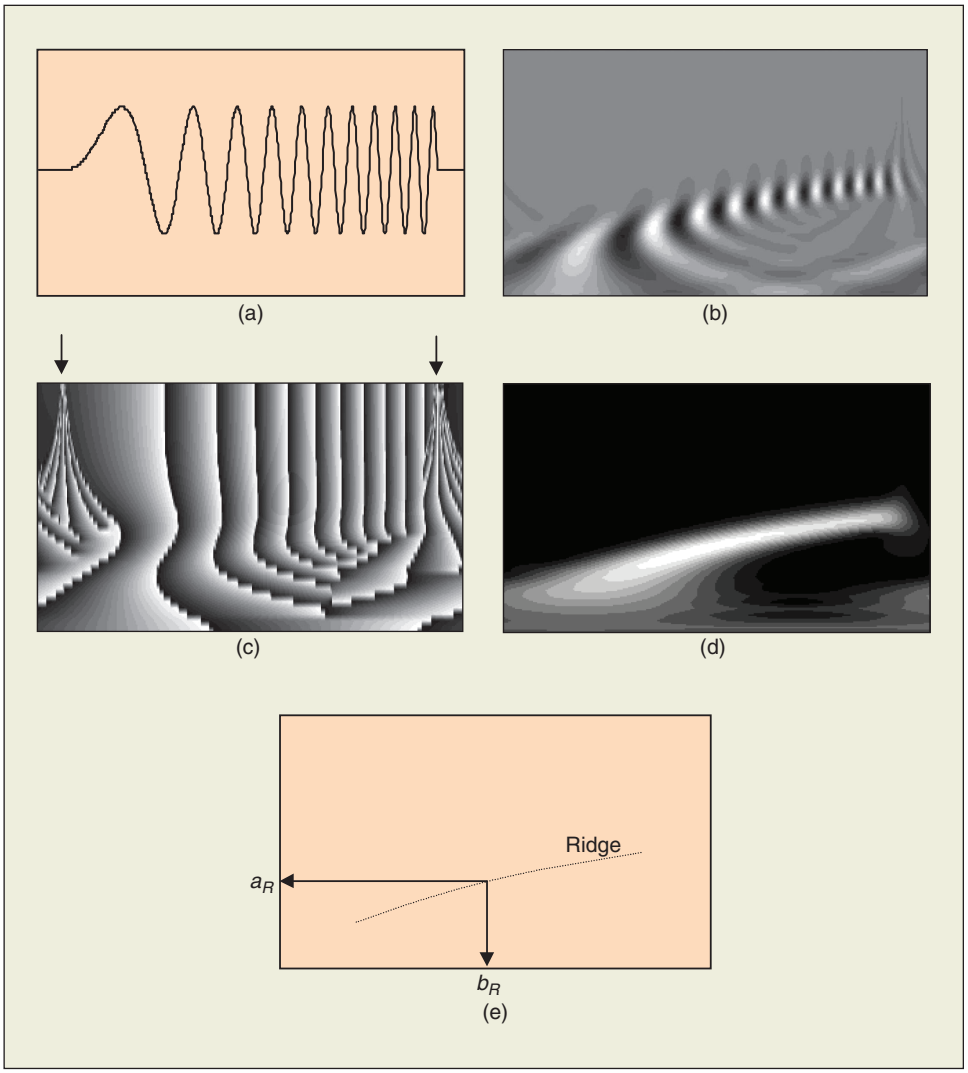

Fig. 4. Segment of a chirp signal with associated transform plots: Morlet wavelet. (a) Chirp signal segment. (b) Real part of Morlet wavelet transform. (c) Phase. (d) Modulus. (e) A schematic of the ridge found from the maxima of the rescaled scalogram $|T(a,b)|^2/a$. (Image reprinted from (59). © 2002 by Taylor & Francis Group LLC—Books. Reproduced with permission in the format Journal via Copyright Clearance Center.)

The Discrete Wavelet Transform

In its most common form, the DWT employs a dyadic grid (integer power of two scaling in a and b) with orthonormal wavelet basis functions and exhibits zero redundancy. Actually, the transform integral remains continuous for the DWT but is determined only on a discretized grid of a scales and b locations. In practice, the input signal is treated as an initial wavelet approximation to the underlying continuous signal, from which, using a multiresolution algorithm, the wavelet transform and its inverse can be computed discretely, quickly, and without loss of signal information. A natural way to sample the parameters a and b is to use a logarithmic discretization of the a scale and link this, in turn, to the size of steps taken between b locations. To link b to a, we move in discrete steps to each location b, which are proportional to the a scale. This kind of discretization of the wavelet has the form

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}}\psi\left(\frac{t - nb_0a_0^m}{a_0^m}\right), \tag{16}$$

where the integers m and n control the wavelet dilation

Wraparound is the simplest and one of the most common treatments of the boundary for a finite length signal.

and translation, respectively; a_0 is a specified fixed dilation step parameter set at a value greater than 1, and b_0 is the location parameter that must be greater than zero. A common choice for discrete wavelet parameters a_0 and b_0 are 2 and 1, respectively. This power-of-two logarithmic scaling of both the dilation and translation steps is known as the *dyadic grid* arrangement. The dyadic grid is perhaps the simplest and most efficient discretization for practical purposes and lends itself to the construction of an orthonormal wavelet basis. Substituting $a_0 = 2$ and $b_0 = 1$ into (16), we see that the dyadic grid wavelet can be written compactly as

$$\psi_{m,n}(t) = 2^{\frac{-m}{2}}\psi(2^{-m}t - n). \tag{17}$$

Note that this has the same notation as the general discrete wavelet given by (16). From here on, $\psi_{m,n}(t)$ will be used only to denote dyadic grid scaling with $a_0 = 2$ and $b_0 = 1$. Discrete dyadic grid wavelets are usually chosen to be orthonormal, i.e., they are both orthogonal to each other and are normalized to have unitary energy. This is expressed as

$$\int_{-\infty}^{\infty} \psi_{m,n}(t)\,\psi_{m',n'}(t)dt = \begin{cases} 1 & \text{if } m = m' \text{ and } n = n', \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

This means that the information stored in a wavelet coefficient $T_{m,n}$ obtained from the wavelet transform is not repeated elsewhere and allows for the complete regeneration of the original signal without redundancy. The corresponding family of orthonormal wavelets is an orthonormal basis. A basis is a set of vectors, a combination of which can completely define the signal, $x(t)$. An orthonormal basis has component vectors which, in addition to being able to completely define the signal, are perpendicular to each other. Figure 5 shows a number of examples of orthonormal wavelets.

Using the dyadic grid wavelet of (17), the DWT can be written as

$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt, \tag{19}$$

where $T_{m,n}$ is known as the wavelet (or detail) coefficient at scale and location indices (m,n).

Before continuing, it is important to make clear the distinct difference between the DWT and the discretized approximations of the CWT used in practice. The discretization of the CWT, required for its practical implementation, involves a discrete approximation of the transform integral (i.e., a summation) computed on a discrete grid of a scales and b locations. The inverse CWT is also computed as a discrete approximation. How close an approximation to the original signal is recovered mainly depends on the resolution of the discretization used and, with care, usually a very good approximation can be recovered. On the other hand, for the DWT, as defined in (19), the transform integral remains continuous but is determined only on a discretized grid of a scales and b locations. We can then sum the DWT coefficients to infinity over m and n to get the original signal back exactly. We will see later in this section how, given an initial discrete input signal, which we treat as an initial approximation to the underlying continuous signal, we can compute the wavelet transform and inverse transform discretely, quickly, and without loss of signal information.

Orthonormal dyadic discrete wavelets are associated with scaling functions and their dilation equations. The scaling

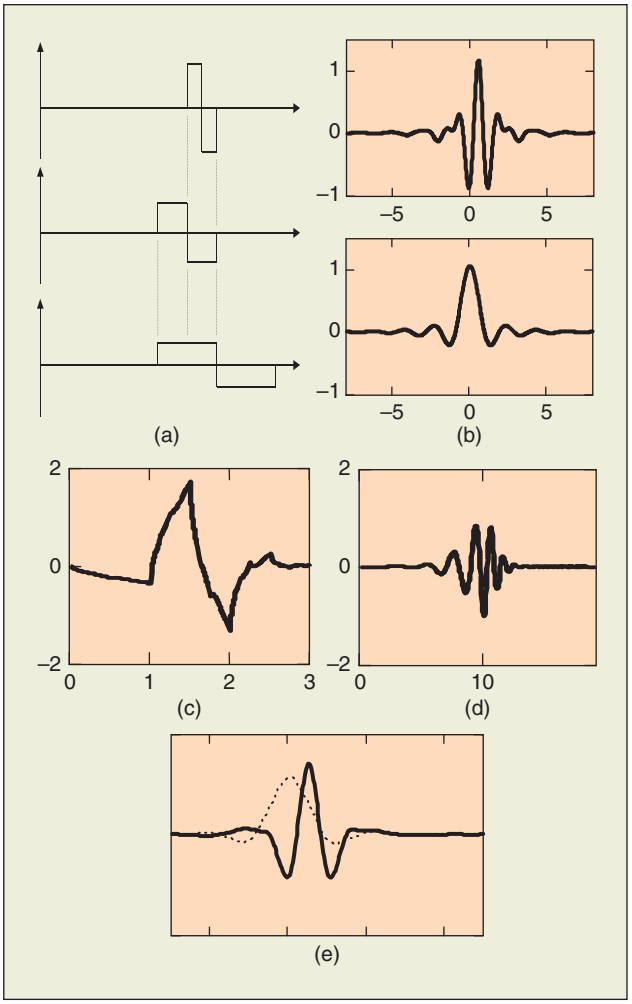

Fig. 5. Examples of discrete orthogonal wavelets. (a) Three Haar wavelets at three consecutive scales on a dyadic grid. (b) A Meyer wavelet and associated scaling function (right). (c) Daubechies D4 wavelet. (d) Daubechies D20 wavelet. (e) C18 Coiflet wavelet (scaling function shown dotted).

function is associated with the smoothing of the signal and has the same form as the wavelet, given by

$$\phi_{m,n}(t) = 2^{\frac{-m}{2}}\phi(2^{-m}t - n). \tag{20}$$

They have the property

$$\int_{-\infty}^{\infty} \phi_{0,0}(t)dt = 1, \tag{21}$$

where $\phi_{0,0}(t) = \phi(t)$ is sometimes referred to as the father-scaling function or father wavelet. The scaling function is orthogonal to translations of itself, but not to dilations of itself. The scaling function can be convolved with the signal to produce *approximation coefficients* as follows:

$$S_{m,n} = \int_{-\infty}^{\infty} x(t)\phi_{m,n}(t)dt. \tag{22}$$

From above, we can see that the approximation coefficients are simply weighted averages of the continuous signal factored by $2^{m/2}$. The approximation coefficients at a specific scale m are collectively known as *discrete approximation* of the signal at that scale. A *continuous approximation* of the signal at scale m can be generated by summing a sequence of scaling functions at this scale factored by the approximation coefficients as follows:

$$x_m(t) = \sum_{n=-\infty}^{\infty} S_{m,n}\ \phi_{m,n}(t), \tag{23}$$

where $x_m(t)$ is a smooth, scaling-function-dependent version of the signal $x(t)$ at scale index m. This continuous approximation approaches $x(t)$ at small scales, i.e., as $m \to -\infty$. A signal $x(t)$ can then be represented using a combined series expansion, with both the approximation coefficients and the detail coefficients as follows:

$$x(t) = \sum_{n=-\infty}^{\infty} S_{m_0,n}\ \phi_{m_0,n}(t) + \sum_{m=-\infty}^{m_0}\sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t). \tag{24}$$

We can see from this last equation that the original continuous signal is expressed as a combination of an approximation of itself, at arbitrary scale index m_0, added to a succession of signal details from scales m_0 down to negative infinity. The *signal detail* at scale m is defined as

$$d_m(t) = \sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t). \tag{25}$$

Hence, we can write (24) as

$$x(t) = x_{m_0}(t) + \sum_{m=-\infty}^{m_0} d_m(t). \tag{26}$$

From this equation, it is easy to show that

$$x_{m-1}(t) = x_m(t) + d_m(t), \tag{27}$$

which tells us that if we add the signal detail at an arbitrary scale (index m) to the approximation at that scale, we get the signal approximation at an increased resolution, i.e., at a smaller scale, index $m-1$. This is called a *multiresolution representation*.

Coefficients from Coefficients: Multiresolution and the Fast Wavelet Transform

The approximation coefficients at scale index $m+1$ can be generated using the scaling coefficients at the previous scale:

$$S_{m+1,n} = \frac{1}{\sqrt{2}}\sum_k c_k\ S_{m,2n+k} = \frac{1}{\sqrt{2}}\sum_k c_{k-2n}\ S_{m,k}. \tag{28}$$

Similarly, the wavelet coefficients can be found from the approximation coefficients at the previous scale using the reordered scaling coefficients b_k as follows:

$$T_{m+1,n} = \frac{1}{\sqrt{2}}\sum_k b_k\ S_{m,2n+k} = \frac{1}{\sqrt{2}}\sum_k b_{k-2n}\ S_{m,k}. \tag{29}$$

We can see now that, if we know the approximation coefficients $S_{m_0,n}$ at a specific scale m_0, then, through the repeated application of (28) and (29), we can generate the approximation and detail wavelet coefficients at all scales larger than m_0. Notice that, to do this, we do not even need to know exactly what the underlying continuous signal $x(t)$ is, only $S_{m_0,n}$. Equations (28) and (29) represent the multiresolution decomposition algorithm. The decomposition algorithm is the first half of the fast wavelet transform, which allows us to compute the wavelet coefficients in this way, rather than computing them laboriously from the convolution of (19). Iterating (28) and (29) performs, respectively, a high-pass and low-pass filtering of the input (i.e., the coefficients $S_{m,2n+k}$) to get the outputs ($S_{m+1,n}$ and $T_{m+1,n}$). The vectors containing the sequences $1/\sqrt{2}c_k$ and $1/\sqrt{2}b_k$ represent the filters: $1/\sqrt{2}c_k$ is the low-pass filter, letting through low-signal frequencies and hence a smoothed version of the signal, and $1/\sqrt{2}b_k$ is the high-pass filter, letting through the high frequencies corresponding to the signal details. The filter coefficients determine the wavelet used. The reader is referred to [15] for more information on coefficients (including listings) and the resulting properties of their associated wavelets.

We can go in the opposite direction and reconstruct $S_{m,n}$ from $S_{m+1,n}$ and $T_{m+1,n}$ using the reconstruction algorithm:

$$S_{m-1,n} = \frac{1}{\sqrt{2}}\sum_k\ c_{n-2k}S_{m,k} + \frac{1}{\sqrt{2}}\sum_k b_{n-2k}T_{m,k}, \tag{30}$$

where we have reused k as the location index of the transform coefficients at scale index m to differentiate it from n, the location index at scale $m-1$. Hence, at the smaller scale, $m-1$, the approximation coefficients can be found in terms of a combination of approximation and detail coefficients at the next scale, m. Note that if there are only a finite number of nonzero scaling coefficients ($=N_K$), then c_{n-2k} has nonzero values only when in the range 0 to N_k-1. The reconstruction algorithm is the second half of the multiresolution algorithm.

Discrete Input Signals of Finite Length

In practice, discrete input signals are analyzed. This is generally taken to be the signal approximation coefficients at scale index $m = 0$. In addition, this series of coefficients, $S_{0,n}$, is of finite length N, which is an integer power of 2: $N = 2^M$. Thus, the range of scales that can be investigated is $0 < m < M$.

The DWT has emerged as a particularly powerful tool for the encoding of data required for compression systems.

Substituting both $m = 0$ and $m = M$ into (24) and noting that we have a finite range of n which halves at each scale, we can see that the signal approximation scale $m = 0$, i.e., the input signal, can be written as the smooth signal at scale M plus a combination of detailed signals as follows:

$$\sum_{n=0}^{2^{M-m}-1} S_{0,n}\phi_{0,n}(t) = S_{M,n}\phi_{M,n}(t) + \sum_{m=1}^{M}\sum_{n=0}^{2^{M-m}-1} T_{m,n}\psi_{m,n}(t). \tag{31}$$

This is the form used to describe our finite length discrete signal in terms of its discrete wavelet expansion. The covering of a finite length time segment with wavelets is illustrated in Figure 6 for Daubechies D4 wavelets at two successive scales. The lower scale covers the time window using eight wavelets, and the larger scale uses four wavelets. One of the wavelets in each plot is shown in bold for clarity. The wavelets shown, which spill over the end of the window, have been wrapped around back to the beginning. Known as *wraparound*, it is the simplest and one of the most common treatments of the boundary for a finite length signal. However, note that by employing this method, we assume that the signal segment under investigation represents one period of a periodic signal and we are in effect sending the end of the signal back onto the beginning. Obviously, if the signal is not periodic, and in practice it is usually not, then we create artificial singularities at the boundary, which results in large detail coefficients generated near to the boundary: a point very often overlooked in practice. Details of other methods for treating the boundaries can be found in [3] and [52].

We can rewrite (31) as

$$x_0(t) = x_M(t) + \sum_{m=1}^{M} d_m(t), \tag{32}$$

where the mean signal approximation at scale M is

$$x_M(t) = S_{M,n}\phi_{M,n}(t). \tag{33}$$

When wraparound is employed to deal with the boundaries, the single approximation component $S_{M,n}$ is related to the mean of the input signal through the relationship $\overline{S_{0,n}} = S_{M,n}/\sqrt{2^M}$, where the overbar denotes the mean of the sequence $S_{0,n}$. In addition, when wraparound has been used to deal with the boundaries, the mean signal approximation at the largest scale, $x_M(t)$, is a constant-valued function equal to the input signal mean.

After a full decomposition, the energy contained within the coefficients at each scale is given by

$$E_m = \sum_{n=0}^{2^{M-m}-1} (T_{m,n})^2. \tag{34}$$

A wavelet-based power spectrum of the signal may be produced using these scale-dependent energies. To do so, we require a frequency measure that is a reciprocal of the wavelet dilation, e.g. the passband center of the power spectrum of the wavelet. A wavelet power spectrum can then be produced for the signal, which is directly comparable with both its Fourier and continuous wavelet counterparts. The sum of the energies contained in detail approximation coefficients is equal to the energy in the discrete input signal, and this is true for the energy at all stages of the multiresolution decomposition.

The term on the far right of (32) represents the series expansion of the fluctuating components of the finite length signal at various scales in terms of its detail coefficients. The detail signal approximation corresponding to scale index m is defined for a finite length signal as

$$d_m(t) = \sum_{n=0}^{2^{M-m}-1} T_{m,n}\psi_{m,n}(t). \tag{35}$$

As we saw in (32), adding the approximation of the signal at scale index M to the sum of all detail signal components across scales $0 < m < M$ gives the approximation of the original

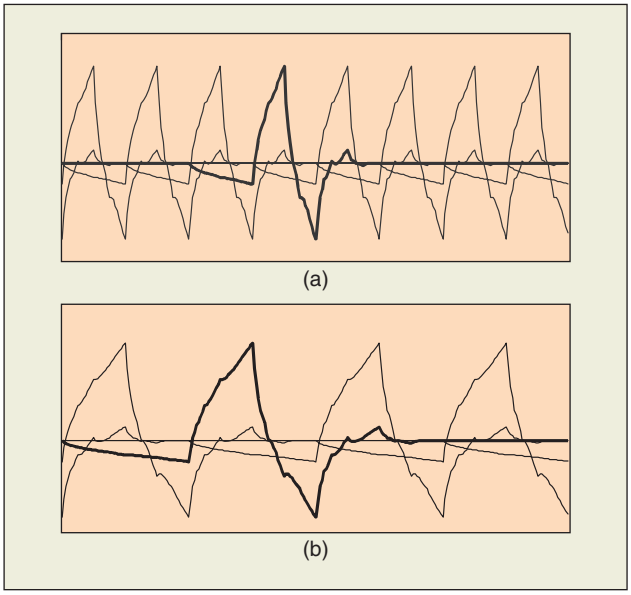

Fig. 6. Covering the time axis with dyadic grid wavelets. (a) Eight Daubechies D4 wavelets covering the time axis at scale m. (b) Four Daubechies D4 wavelets covering the time axis at scale $m+1$. These wavelets are twice the width of those in (a). (Image reprinted from (59). © 2002 by Taylor & Francis Group LLC—Books. Reproduced with permission in the format Journal via Copyright Clearance Center.)

signal at scale index 0. Figure 7 shows the details of a chirp signal with a short burst of noise added to the middle of it. A Daubechies D20 wavelet was used in the decomposition. The original signal is shown at the top of the plot. Below the signal, the details for ten wavelet scales, $d_1(t)$ to $d_{10}(t)$, are shown. The bottom trace is the remaining signal approximation $x_{10}(t)$. Adding together all these details plus the remaining approximation, which is the signal mean, returns the original signal. Two things are noticeable from the plot. First, there is a shift to the left of the large amplitude details with increasing scale, as we would expect as the chirp oscillation increases in frequency from left to right. The second thing to notice is that the high-frequency burst of noise is captured at the smallest scales, again as we would expect.

Figure 8 contains an example of a wavelet decomposition of a test signal using a discrete wavelet. The original composite signal is composed of three segments: a sinusoid, a uniformly distributed noise, and a flat line. The signal is decomposed using a Daubechies D6 wavelet. A member of this family is shown in Figure 8(b). The discrete-transform plot is shown in Figure 8(c), where the dyadic grid arrangement may be clearly seen. This plot is simply a discretized dyadic map of the detail coefficients, $T_{m,n}$, where the coefficients at larger scales have correspondingly longer boxes (as the wavelets cover larger segments of the input signal). In addition to the detail coefficients, $T_{m,n}$, the remaining approximation coefficient $S_{M,0}$ is added to the bottom of the plot. As we would expect, it covers the whole time axis. We can see from the transform plot that the dominant oscillation is picked up at scale index $m = 6$ and the high-frequency noise is picked up within the middle segment of the transform plot at the smaller scales. We can use the reconstruction algorithm (30) to get back the original input signal $S_{0,n}$ from the array of detail coefficients shown in Figure 8(c). Alternatively, we can reconstruct a modified version of the input signal by using only selected coefficients in the reconstruction. This is shown in Figure 8(d) and (e), where only the coefficients corresponding to scales $m = 5 - 8$ are kept (the others are set to zero), and the signal is reconstructed. Note that a gray scale is used to depict the coefficient values, where the maximum value is white and the minimum value is black. This has removed a significant amount of noise from the signal although the sinusoidal waveform is less smooth than the original in Figure 8(a). There are many more sophisticated methods to remove noise, which retain the most significant signal coefficients before reconstructing. A number of thresholding methods have been developed to allow the optimum selection of the most pertinent components. A full treatment of wavelet thresholding methods is out of the scope of this text. The reader is referred to [1] for a concise summary of the commonly used wavelet thresholding methods together with a comprehensive list of references on the subject. Data-compression methods require that only those wavelet coefficients that carry most of the signal information are identified and retained for use in the reconstruction of the signal. Thresholding methods have also been used to determine the significant coefficients pertaining to a pertinent signal feature, e.g., for ECG analysis, they can be used to determine the location of the QRS peak. A deeper mathematical treatment of the DWT can be found in [8], [41], [46], and [58].

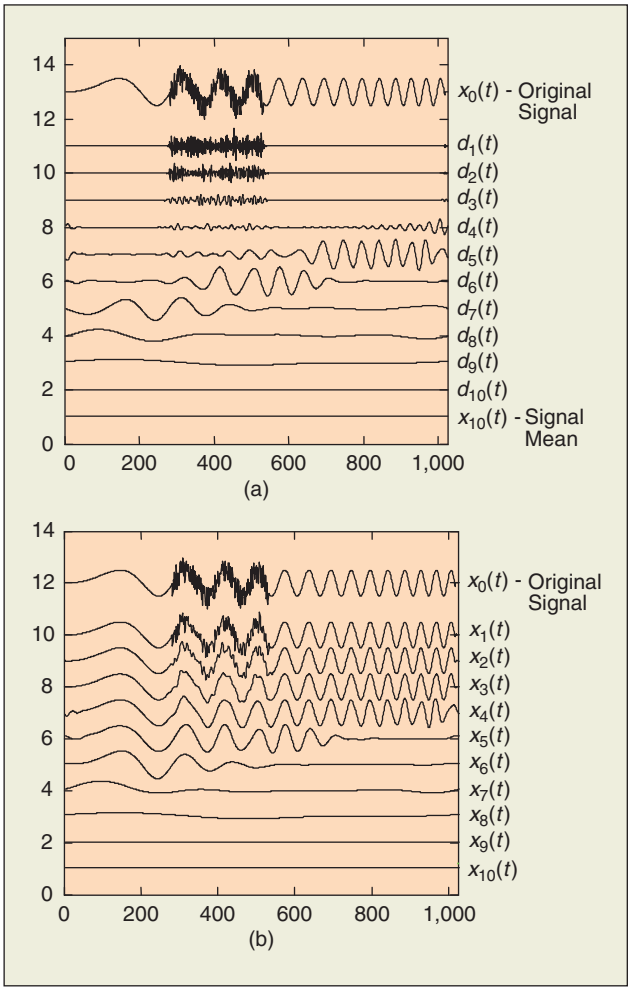

Fig. 7. Multiresolution decomposition of a chirp signal containing a short burst of noise. (a) Signal details $d_m(t)$ (the signals have been displaced from each other on the vertical axis to aid clarity). (b) Signal approximations $x_m(t)$. (Image reprinted from (59). © 2002 by Taylor & Francis Group LLC—Books. Reproduced with permission in the format Journal via Copyright Clearance Center.)

Methods Derivative of the DWT

The DWT has emerged as a particularly powerful tool for the encoding of data required for compression systems. Wavelet packet transforms (WPTs) are generalizations of the DWT that involve particular linear combinations of discrete wavelets, and the decomposition of a signal is performed in a manner similar to the multiresolution algorithm given earlier for the DWT, the difference being that, in the WPT signal decomposition, both the approximation and detailed coefficients are further decomposed at each level. At each stage, the wavelet packet algorithm partitions the time–frequency plane into rectangles of constant aspect ratio. The optimal WPT coefficient selection is chosen to represent the signal based on some predefined criterion. This criterion is normally based on an information cost function that aims to retain as much information in a few coefficients as possible. The most common measure of information used is the Shannon entropy measure [3]. Low entropies occur when the larger coefficient energies are concentrated at only a few discrete locations. In practice, the set of N wavelet packet coefficients that contain the least entropy are selected to represent the signal. That is, we want as much of the signal information to be concentrated within as

few coefficients as possible. To find these coefficients, the WPT coefficients are inspected, and at each scale, each pair of partitioned coefficients sets is compared to those from which they were derived. Once the whole wavelet packet array has been inspected and the minimum entropy criterion employed, an optimal representation of the time–frequency plane, with respect to the localization of coefficient energies, is obtained. This scheme provides the best basis for the signal decomposition.

Both the DWT and WPT lack translation invariance. Translation invariance simply means that if you shift the analysis along the signal by an arbitrary amount, all your transform coefficients simply move along with the same amount. For the dyadic grid structure of the DWT, this is clearly not the case: only if you shift along by the grid spacing at that scale do the coefficients become translation invariant at that scale and below. For the discretization of the CWT used in practice, the transform values are translation invariant only if shifted by any integer multiple of the discrete time steps, i.e., it is effectively translation invariant as the minimum shift would be the sampling interval of the signal. The stationary wavelet transform (SWT) is an offshoot of the DWT, whereby the scales are dyadic but the time steps are not subsampled at each level and, hence, are not dyadic. This destroys the orthogonality in the transform and also leads to the generation of many more coefficients but does provide translation invariance. Although orthogonality is destroyed, the SWT is very useful for some statistical applications [13]. The SWT is also known by a variety of names in the literature, including the dyadic wavelet transform (i.e., dyadic in scales), maximal overlap transform, and the redundant wavelet transform. It is out of scope of this article to consider the WPT and the SWT in more detail. For more information concerning the theory of these derivatives of the DWT, the reader is referred to [3, chapter 3].

Application of Wavelet Analysis to Biosignals

The wavelet transform has been applied to a wide variety of biosignals, including analysis of neurophysiological signals, EEG analysis, ECG analysis, electromyography (EMG) analysis, electrogastrogram analysis, cellular vibrations in the living cochlea, denoising for medical signals and images, estimation of pulmonary microvascular pressure, mammograms analysis, medical image enhancement, magnetic resonance and tomography medical images analysis, medical image and signal compression, brain signals analysis, phonocardiogram analysis, speech enhancement for hearing aids, Doppler ultrasound, classification of voice disorders, denoising of medical images, compression of medical data, and so on.

In the next articles of this special issue, some of these specific problems are treated, which illustrate best the diverse application of the wavelet transform to biosignal analysis.

2-D Wavelets and Image Processing

The field of image processing is a huge one. It encompasses, at the very least, the following areas: 1) image compression, 2) image denoising, 3) image enhancement, 4) image recognition, 5) feature detection, and 6) texture analysis. Wavelet-based techniques apply to all of these topics. One reason that wavelet analysis provides such an all-encompassing tool for image processing is that a similar type of analysis occurs in the human visual system. To be more precise, the human visual system performs hierarchical edge detection at multiple levels of resolution, and the wavelet transforms perform a similar analysis (more on this is detailed later).

Rather than attempting to describe in detail how wavelets apply to all of the areas listed earlier (that would take an entire book), we focus instead on the first two topics, image compression and image denoising. For those readers who desire to study other topics, they are explained at an elementary level in [52], where further references can be found.

2-D Wavelet Transforms for Images

We shall briefly outline the theory of 2-D wavelet transforms, focusing on their applications to discrete images. A more complete discussion can be found in [53].

Wavelet series for 2-D images are a simple generalization from one-dimensional (1-D) series. The wavelet basis functions consist of three kinds: $\psi_{m,j}(x)\phi_{m,k}(y), \phi_{m,j}(x)\psi_{m,k}(y)$, and

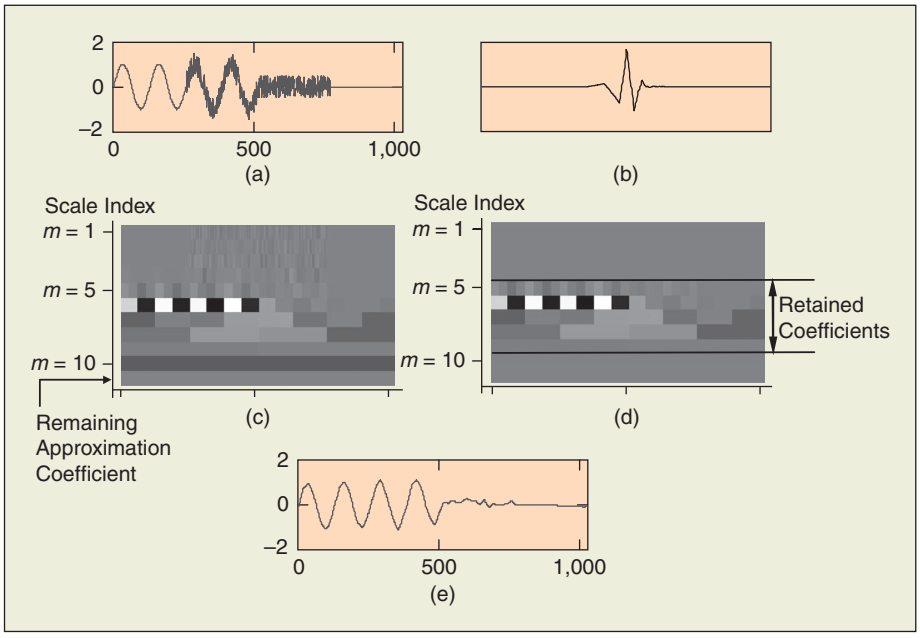

Fig. 8. Discrete wavelet transform of a composite signal. The (a) signal is decomposed using a member of the Daubechies D6 wavelet family (b) to give the dyadic array of transform coefficients plotted in discrete transform plot (c). (Note dyadic structure: large positive coefficient values are white and large negative values are black.) The coefficients corresponding to scales 5–9 are kept (the same is indicated in coefficient removal (d)) and used to reconstruct the signal using only the retained coefficients in (e). (Image reprinted from (59). © 2002 by Taylor & Francis Group LLC—Books. Reproduced with permission in the format Journal via Copyright Clearance Center.)

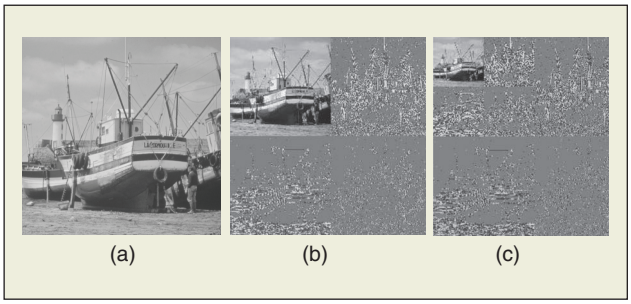

Fig. 9. (a) Image. (b) First-level transform. (c) Second-level transform.

$\psi_{m,j}(x)\psi_{m,k}(y)$. A 2-D wavelet series for $F(x,y)$ is expressible as

$$F(x,y) = \sum_{m=-\infty}^{\infty} \sum_{l \in J} T_{m,l} \Psi_{m,l}(x,y) \quad \text{with}$$

$$T_{m,l} = \int_{\mathbf{R}^2} F(x,y) \Psi_{m,l}(x,y)\, dx\, dy.$$

Here, $\Psi_{m,l}(x,y)$ stands for any one of the three types of basis functions, and is the index set needed for labeling all of these basis functions. There is also a scaling series, using $\{\phi_{m,j}(x)\phi_{m,k}(y)\}$ as basis, which is an obvious generalization of the 1-D scaling series. The theory of wavelet series for 2-D images is essentially the same as the 1-D theory mentioned earlier, with only slight notational modifications, so we now turn to a discussion of 2-D DWTs. Such transforms will be applied to matrices, just as in 1-D, they were applied to vectors.

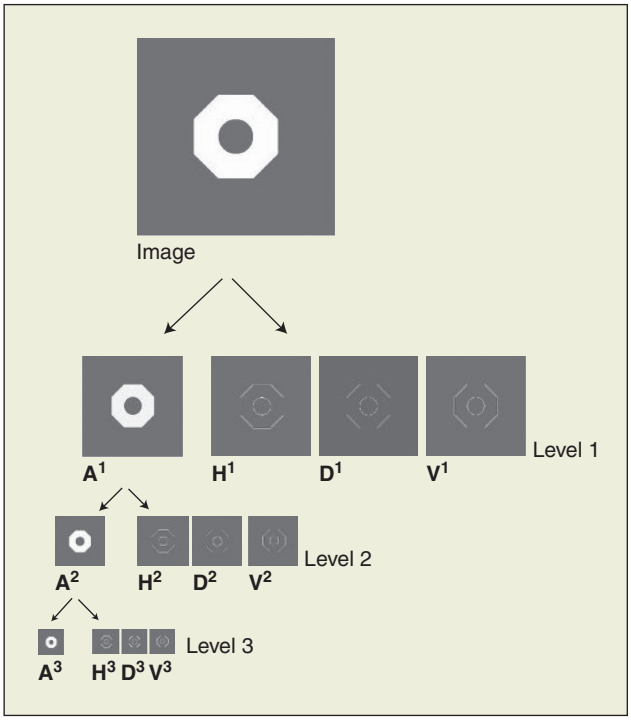

Fig. 10. Example of a wavelet transform. $\mathbf{A^1}$: smoothed subimage obtained from local averaging of the image, one-fourth resolution. $\mathbf{H^1}$: horizontal component subimage $\mathbf{D^1}$: diagonal component subimage $\mathbf{V^1}$: vertical component subimage $\mathbf{A^2}$, $\mathbf{H^2}$, $\mathbf{D^2}$, $\mathbf{V^2}$: iteration of localized averaging and localized differencing applied to $\mathbf{A^1}$ subimage. $\mathbf{A^3}$, $\mathbf{H^3}$, $\mathbf{D^3}$, $\mathbf{V^3}$: iteration of localized averaging and localized differencing applied to $\mathbf{A^2}$ subimage.

A DWT of a J by K matrix $\mathbf{F}$, where J and K are both even, is obtained in two steps [39]: 1) transform each row of $\mathbf{F}$ by a 1-D wavelet transform, obtaining a matrix $\mathbf{F}$; 2) transform each column of $\mathbf{F}$ by the same 1-D transform. Steps 1 and 2 are independent and may be performed in either order.

Step 1 of this transform process produces rows of 1-D transforms:

$$\mathbf{F} \mapsto \begin{pmatrix} \{\mathbf{S_{1,\ell}}\} | \{\mathbf{T_{1,\ell}}\} \\ \{\mathbf{S_{2,\ell}}\} | \{\mathbf{T_{2,\ell}}\} \\ \vdots \\ \{\mathbf{S_{J,\ell}}\} | \{\mathbf{T_{J,\ell}}\} \end{pmatrix}.$$

Step 2 then produces the following first-level transform of the matrix $\mathbf{F}$:

$$\mathbf{F} \mapsto \begin{pmatrix} \mathbf{A^1} & \mathbf{V^1} \\ \mathbf{H^1} & \mathbf{D^1} \end{pmatrix},$$

where $\mathbf{A^1}$, $\mathbf{V^1}$, $\mathbf{H^1}$, and $\mathbf{D^1}$ are each $J/2$ by $K/2$ submatrices. The trend $\mathbf{A^1}$ consists of scaling coefficients, while the fluctuations $\mathbf{V^1}$, $\mathbf{H^1}$, and $\mathbf{D^1}$ consist of wavelet coefficients for each of the three kinds of wavelet basis functions. For example, we show in Figure 9(b), a transform of the image shown in Figure 9(a). The trend $\mathbf{A^1}$ appears in the upper left quarter of the image in (b), and we see that it is a lower resolution version of the image in (a). The fluctuations $\mathbf{V^1}$, $\mathbf{H^1}$, and $\mathbf{D^1}$ appear in the upper right, lower left, and lower right quarters, respectively.

The trend $\mathbf{A^1}$ contains the scaling coefficients for the scaling basis $\{\phi_{M-1,j}(x)\phi_{M-1,k}(y)\}$. Hence, it is a lower resolution version of $\mathbf{F}$. For example, consider the image of an octagonal ring shown at the top of Figure 10. A first-level wavelet transform is shown in the second row of the figure. The trend $\mathbf{A^1}$ is clearly a lower resolution version of the original image. The vertical fluctuation $\mathbf{H^1}$ contains wavelet coefficients for the basis elements $\{\phi_{M-1,j}(x)\psi_{M-1,k}(y)\}$; i.e., trends along rows and fluctuations along columns. Wherever there are horizontal edges in an image, the fluctuations along columns are able to detect these edges. This tends to emphasize horizontal edges or edges containing a horizontal component. This can be seen clearly in the example of $\mathbf{H^1}$ in the figure. Notice also that the vertical edges, where the octagonal ring image is constant over long stretches, are removed from $\mathbf{H^1}$. The vertical fluctuation $\mathbf{V^1}$ (obtained from localized horizontal differencing and localized vertical averaging of the image) is similar to $\mathbf{H^1}$ (obtained from localized vertical differencing and localized horizontal averaging of the image), except that the roles of horizontal and vertical are reversed. We see this in the example of $\mathbf{V^1}$ shown in the second row of Figure 10. Finally, there is the diagonal fluctuation $\mathbf{D^1}$ (obtained from localized vertical differencing and localized horizontal differencing of the image), which contains wavelet coefficients for the basis elements $\{\psi_{M-1,j}(x)\psi_{M-1,k}(y)\}$. This fluctuation tends to emphasize diagonal features, as can be seen clearly in the example shown in the second row of Figure 10. Diagonal features are emphasized, because fluctuations along rows and columns tend to erase horizontal and vertical edges.

It is interesting to note that this decomposition of the image into a lower resolution subimage, along with several subimages reflective of responses to different edge orientations, is

The dyadic grid is the simplest and most efficient discretization for practical purposes and lends itself to the construction of an orthonormal wavelet basis.

analogous to operations performed by mammalian vision systems. Reference [57] is perhaps the first article to establish a close connection between wavelets and vision. A precursor to that article in the field of image processing was [10]. Several articles by [21]–[24] explore this connection completely. A good summary of wavelets and vision can be found in [56].

As with 1-D transforms, higher levels of 2-D wavelet transforms are computed by iterating the first-level transform on the trends. For example, in the third row of Figure 10, we show a second-level transform of the octagonal ring image. The trend $\mathbf{A^1}$ from the second row of the figure has been transformed into trend $\mathbf{A^2}$ and fluctuations $\mathbf{H^2}$, $\mathbf{D^2}$, and $\mathbf{V^2}$. The trend $\mathbf{A^2}$ is an even lower resolution version of the original octagon image, while the fluctuations reflect edge details in the lower resolution image $\mathbf{A^1}$. In Figure 9(c), we show another example of a two-level wavelet transform, which illustrates the recursive nature of the calculation. An example of a third-level transform is shown in the last row of Figure 10.

The transform used in Figure 10 is based on the Daub 9/7 wavelet system [12]. This system is a very popular one in wavelet-based image processing. For instance, it is used in the Joint Photographic Experts Group (JPEG) 2000 image compression standard [26].

The essential idea behind wavelet transforms is to iterate invertible smoothing and differencing operations and, at each iteration, maintain the same number of pixels as the original image. In Figure 10, we show how these smoothing and differencing operations apply to a test image. At Level 1, the image is smoothed (via local averages) and reduced to a one-fourth size lower resolution version of the original image, and three different local differencing operations are performed, providing edge detection of three kinds: horizontal component images (vertical edges suppressed), diagonal component images (horizontal and vertical edges suppressed), and vertical component images (horizontal edges suppressed). The total number of pixels at this level is equal to that of the original image. Moreover, this Level 1 transform can be inverted: from the four subimages at Level 1, the original image can be reconstructed.

Level 2 of a wavelet transform repeats the averaging and differencing operations on the Level 1 smoothed subimage. This iterative process continues. We have stopped in Figure 10 at Level 3. The Level 3 transform consists of a 1/64 size smoothed subimage, and all of the edge subimages created by local differencing at each level. The original image can be recovered from this three-level transform.

There are three reasons why wavelet transforms have had such a profound effect on image processing. First, the process of creating edge subimages at multiple resolutions is analogous to a process performed by mammalian vision systems (including human vision systems) (see [23], [24], [40], and [57]). Second, the process by which a wavelet transform is constructed (local averaging and differencing operations at multiple resolutions) is akin to some important methods for analyzing images. It is akin to the Laplacian pyramid method of [10] (see also [34]) and the Mumford-Shah theorem concerning edges and smooth background in images [43]. Third, there is an analogy between wavelet transforms and fractal theory. See [16] for an excellent discussion. The fractal-like nature of the wavelet transform is particularly evident in Figure 10.

Image Denoising

We shall briefly outline the theory of wavelet-based image denoising and provide references where a more thorough discussion can be found. Wavelet-based image denoising is based on a few basic ideas: 1) the wavelet transform of the image is an orthogonal (energy-preserving) transform, 2) the wavelet transform of the image performs energy compaction, most of the energy of the transform is concentrated in relatively few transform values, 3) there is a high degree of correlation between high-magnitude transform values at interscale (same relative positions but at different levels of the transform) and intrascale positions (same relative positions, including adjacent neighbor values, at the same level of the transform). We now discuss how these three ideas are used.

First, because the wavelet transform of the image is orthogonal, when the additive noise is independently identically distributed Gaussian normal with mean 0 and standard deviation σ, independent identically distributed (i.i.d.) $\mathcal{N}(0, \sigma^2)$, then the transform of the noise is also $\mathcal{N}(0, \sigma^2)$. Hence, with a probability of 1, the transformed noise values will lie below the threshold $T = \sigma\sqrt{2 \log M}$, where M is the larger of the row and column dimensions (because the probability of $\mathcal{N}(0, \sigma^2)$ noise lying above T in magnitude is much less than $1/M$, and because there are no more than M values along any row or column, it follows that the probability is 1 that no noise value's magnitude lies above T). Second, because of energy compaction, there will (assuming that is not extraordinarily high) be a large proportion of image-transform values that lie above the threshold T. Consequently, thresholding (removing all transform values of magnitude above T) will eliminate essentially all of the noise while retaining enough image-dominated transform values to reconstruct a good approximation of the original image prior to the addition of the noise. This thresholding method, with the added technique of shrinking the transform values with magnitude above T by subtracting T from their magnitudes, is due to [18] and [19]. These workers have also shown that the mean square error (MSE) is asymptotically less than that obtained using the classic Wiener method of denoising. (The MSE between images f and g is $(\text{MSE}) = 1/p \sum_{i,j} (f_{i,j} - g_{i,j})^2$, where the sum is over all P pixels.) Or, put in another way, the peak signal-to-noise ratio (PSNR) is asymptotically higher for this wavelet thresholding

Table 1. PSNRs for three denoising methods.

Image	Noisy	Wiener	TAWS	TAWS-Spin
Noise std. dev. $\sigma = 8$				
Barbara	30.06	30.43	31.61	**32.33**
Boats	30.07	32.90	33.66	**34.23**
Goldhill	30.06	32.45	32.55	**33.21**
Lena	30.06	34.54	34.61	**35.42**
Noise std. dev. $\sigma = 16$				
Barbara	24.08	27.96	27.98	**28.20**
Boats	24.08	29.84	29.90	**30.74**
Goldhill	24.08	29.66	29.43	**30.21**
Lena	24.05	30.65	31.23	**31.88**
Noise std. dev. $\sigma = 32$				
Barbara	18.27	24.05	24.40	**24.52**
Boats	18.22	25.04	26.39	**27.32**
Goldhill	18.22	24.99	26.46	**27.28**
Lena	18.17	25.31	27.87	**28.95**

method. (The PSNR between two 8-b gray-scale images is $(\text{PSNR}) = 10\log_{10}(255^2/(\text{MSE}))$.

Unfortunately, in practice, this basic shrinkage method removes too many image-dominated values, producing images lacking sharp details. See [53] for examples, and further discussion of this point. To remedy this defect of the basic shrinkage method, state-of-the-art denoising methods use the third idea mentioned earlier (the interscale and intrascale correlations of image-transform values) to extract more image-dominated transform values from below the threshold T. These methods can be quite elaborate. For simplicity, we shall only describe a rather elementary, but nevertheless quite effective, technique known as tree-adapted wavelet shrinkage

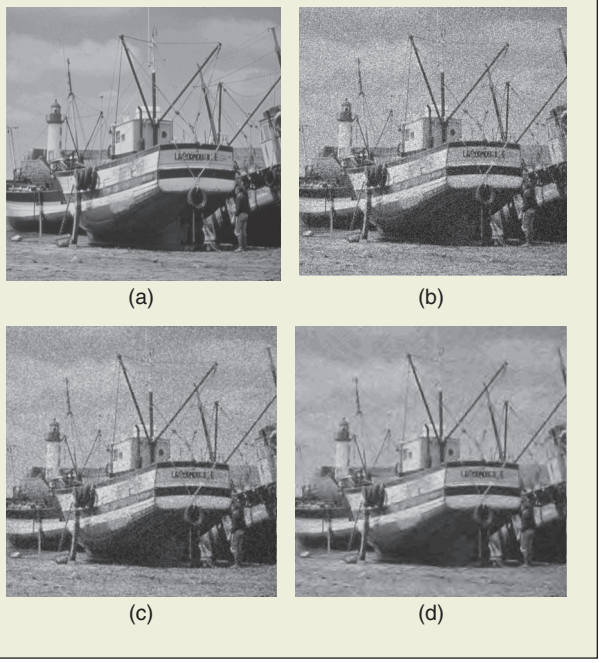

Fig. 11. Comparison of Wiener and TAWS-spin denoising. (a) Boats image. (b) Noisy image, $\sigma = 32$. (c) Wiener denoising. (d) TAWS-spin denoising.

(TAWS) [53], and then give references for some of the later, more advanced techniques.

The TAWS method is based on two facts. First, as noted in 3), there are interscale correlations between higher-magnitude wavelet coefficients in the image. For example, see the transforms of the test image shown in Figure 10. The edges shown in the fluctuation subimages $\mathbf{H^1}$, $\mathbf{H^2}$, and $\mathbf{H^3}$ are highly correlated, as are those in $\mathbf{V^3}$, $\mathbf{V^2}$, and $\mathbf{V^3}$, and those in $\mathbf{D^3}$, $\mathbf{D^2}$, and $\mathbf{D^1}$. These correlations allow one to identify image-dominated transform values at a lower threshold than T. Second, as noted in 3), there are intrascale correlations between higher-magnitude wavelet coefficients in the image. TAWS uses the fact that adjacent coefficients are correlated to further distinguish image-dominated transform values lying below the threshold T. TAWS uses a threshold of $T/8$, instead of T, and uses these two correlation methods to further extract image-dominated transform values that lie between $T/8$ and T, and applies shrinkage (by subtracting $T/8$ from the magnitudes of thresholded-transform values) as well.

A more complete discussion of TAWS can be found in [53]. We now show some data illustrating the superiority of TAWS to the classic Wiener method of denoising. The well-known Wiener method is the optimal (highest PSNR) method when a linear transform is used to filter out the noise. The TAWS method (and all other wavelet methods) are nonlinear, so the door is open for them to outperform the Wiener method. In Table 1, we show the PSNRs for the four test images, with three different noise standard deviations, using the Wiener and TAWS method. The Wiener method was performed via the MATLAB wiener2 function. As one can see from this table, the TAWS method generally produces higher PSNRs than the Wiener method.

There is one further denoising method summarized in this table, which is overwhelmingly superior to Wiener and TAWS. It is called *TAWS-spin*. The method of TAWS-spin is a simple variant of TAWS. It consists of taking the arithmetic mean of denoisings of shiftings (by both left and right, and up and down) of the noisy image. The reason for performing this averaging process is that a discrete 2-D wavelet transform is not shift invariant, and by averaging shiftings, we create shift invariance, and this improves denoising performance (more details are provided in [53]). The superiority of TAWS-spin is evident from the PSNRs in Table 1. We also provide a visual comparison of TAWS-spin and the Wiener method in Figure 11.

The methods of TAWS and TAWS-spin have been discussed here because they are relatively simple to implement and yet provide excellent denoising performance (the method of TAWS is also adaptable to combined denoising and compression, as discussed in [53]). There are a number of more advanced methods, which also take advantage of interscale and intrascale correlations of wavelet-transform values [31], [38], [45].

Image Compression

We shall outline the basic theory underlying wavelet-transform-based image compression. A brief description of some state-of-the-art image compression algorithms will be given.

Before we look at some particular wavelet-based image compression algorithms, it is helpful to look at a specific image and examine a wavelet transform of it. In Figure 12(a), we show an image commonly used in image processing, known as *airfield.* Figure 12(b) shows a histogram for the intensities of the pixels of airfield. Notice that the histogram is widely dispersed over the range from 0 to 255. A measure of this dispersion is *entropy,*

which for this image is 7.1. (See [6] for the definition of entropy.) This shows that there is almost no redundancy in the 8-b image values in airfield. A zip compression of airfield, for example, yields very little compression (about 15% savings in file size). After using a wavelet transform, however, we obtain Figure 12(c). Figure 12(c) has an entropy of 4.4, which by the principles of information theory (see [6]) can be compressed by a factor of about 8/4.4, a savings of slightly less than 50%. When the decompressed file, corresponding to Figure 12(c), is inverted, the resulting image is exactly the same as the original. Considerably more compression—with some loss of exactness, but still perceptually the same as the original—is obtained if the wavelet-transform values are not encoded to exact accuracy. For example, we show in Figure 12, reconstructions of 50:1 compressions of the airfield images using wavelet methods. For comparison, we show a reconstruction of a 50:1 compression using the JPEG algorithm.

The goal of wavelet-transform encoding is to take advantage of redundancy in the transformed image and obtain a good reconstruction upon decompression. The entropy of 4.4% of the histogram of the wavelet transform shows that there is redundancy in the wavelet transform, visible both in terms of gray background matching up at multiple resolutions (zerotrees), and in terms of nonzero values also overlapping at multiple resolutions (fractal-like aspect). [These zerotrees and the fractal-like aspect of wavelet transforms can be seen even more clearly in the transforms shown in Figures 10 and 11]. Each of the algorithms discussed later are similar in that they transmit code for locations of significant values in the wavelet transform and successively encode binary expansions, relative to a base threshold, of the nonzero values (by sending the most significant bits, then the next significant bits, etc.). See [54] for more details.

Adaptively Scanned Wavelet Difference Reduction

The simplest state-of-the-art algorithm for image compression is the adaptively scanned wavelet difference reduction (ASWDR) algorithm. The basic structure of ASWDR is the following:

- Wavelet-transform image.
- Initialize scan order and threshold.
- *Significance pass:* Encode new significant values using difference reduction. (More on this is discussed later.)
- *Refinement pass:* Generate refinement bits for old significant values.
- Update scan order to search through coefficients that are more likely to be significant at half threshold. (More on this is discussed later).
- Divide threshold by 2, repeat Steps 3 and 4.

The method of difference reduction is the following. Compute binary expansions of the number of steps between new significant values (skipping over old ones) as one scans through the transform. Replace the most significant bit by the sign of the transform value. Use these signs as delimiters between expansions. For example, suppose new significant values are $x[2]$, $x[3]$, and $x[14]$ at positions 2, 3, and 14 in the scan order, and

$$x[2] = +17,\ x[3] = -14,\ x[14] = +18.$$

These new values are at indices 2, 3, and 14. The steps between new values are $2 = (10)_2, 1 = 1(1_2)$, and $11 = (1011)_2$. Difference reduction encoding is then

$$0 + -011 +$$

For more details on difference reduction, see [54]. An elementary arithmetic coding, using Markov-1 statistics only, is also performed on the data generated by difference reduction. (See [7] for a description of Markov-1 arithmetic coding.) It is an open problem to find a more effective arithmetic coding procedure for ASWDR.

The unique feature of ASWDR is its creation of a new scan order. Because of space limitations, we shall not discuss this in detail here. The creation of a new scan order is discussed in detail in [53]–[55]. ASWDR creates a new scan order for each level by first scanning through insignificant children of significant parents at previous thresholds (taking advantage of interscale correlations), then scanning through insignificant children (of insignificant parents at previous thresholds) that have at least one significant adjoining transform value, and then completing the scan by scanning through the remaining insignificant transform values. For each transform value at level L, there are four transform values at level $L-1$ that occupy the same relative position in the image, these four values are the children of the parent value at level L. See [53] for more details. (See the previous three references for more complete discussions of how new scans are created in ASWDR.) By creating new scans in this way, ASWDR takes advantage of both zerotrees and the tendency of new significant children to have old significant parents, and increases the number of transform values which can be encoded with as few steps as possible with difference reduction.

If no new scan order is created, then the ASWDR algorithm reduces to the simpler WDR algorithm [50]–[52]. The WDR algorithm is extremely fast and is suitable for applications where maximum speed is needed [55].

JPEG 2000

The JPEG 2000 codec, which is the new ISO standard for photographic image compression, is described in great detail in [49]. Here, we shall only briefly summarize its main

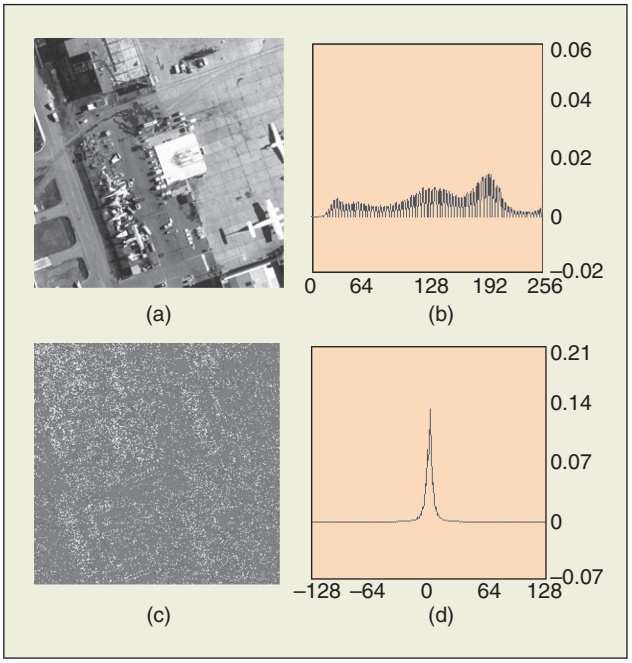

Fig. 12. Airfield image and reduction of entropy via wavelet transform. (a) Airfield image. (b) Image histogram. (c) Airfield transform. (d) Transform histogram.

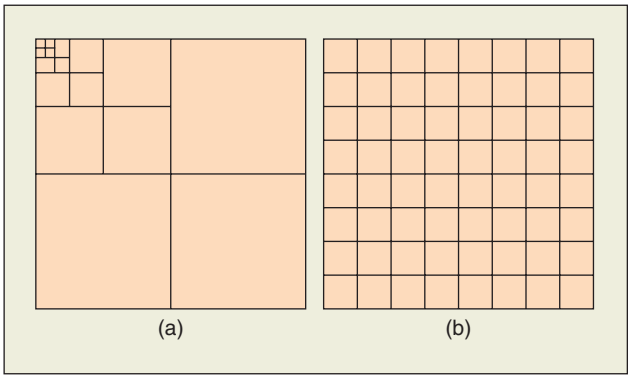

Fig. 13. (a) Subimages in a five-level wavelet transform. (b) Division of transform values into 64 blocks.

features. JPEG 2000 is a block-based method like JPEG , but instead of dividing the image into blocks (subimages), JPEG 2000 divides the wavelet transform into blocks. See Figure 13.

Because the top-left corner block inverts to a low-resolution version of the original image, JPEG 2000 avoids the blocking artifacts of the older JPEG algorithm. JPEG 2000 is described in great detail in [49], and comparisons of it with ASWDR and with JPEG can be found in [52]. We only have space here to provide a single comparison, shown in Figure 14. As we can see in this Figure, both of the wavelet methods provide much better results, both in subjective visual appearance and in objective PSNRs. Although ASWDR performs slightly better then JPEG 2000 for this example, there are other examples where the situation is reversed. Further examples and discussion are in [52].

Other Applications in Image Processing

In the short space we have here, we could only describe a couple of applications of wavelets to image processing, the applications of image denoising and image compression. In addition, there are also applications such as image enhancement, image recognition, feature detection, and texture analysis. The first three of these additional topics are explained at an elementary level in [52], where further references can be found. The last topic is discussed in [32]. All of these other applications involve, as with the two we described, making use of interscale and intrascale correlations of wavelet-transform values. The existence of these interscale and intrascale correlations are a consequence of the prevalence of multiple-edge features at multiple resolution scales, a property that wavelet transforms share with the processing in the human visual system.

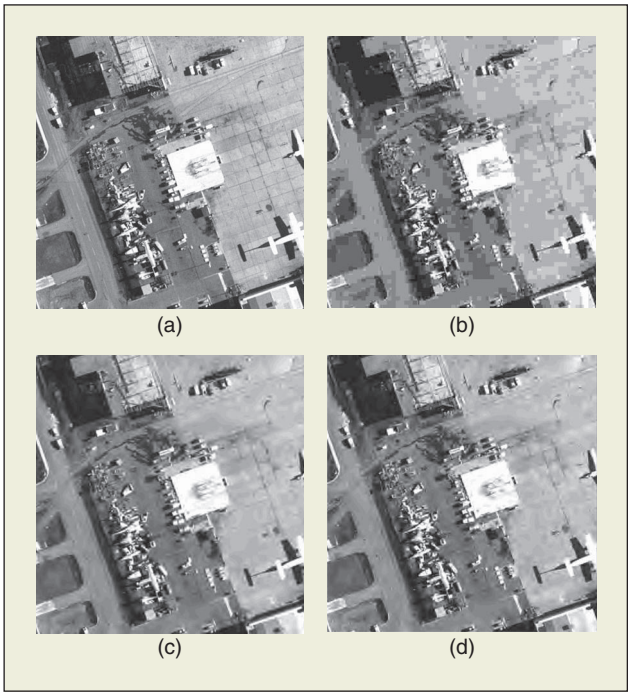

Fig. 14. Airfield image and various decompressions. The numbers in parentheses are the PSNRs. (a) Airfield image. (b) 50:1 JPEG (22.64). (c) 50:1 JPEG 2000 (24.10). (d) 50:1 ASWDR (24.34).

Discussion and Concluding Remarks

The wavelet transform has emerged over recent years as a key time–frequency analysis and coding tool for the investigation of biosignals. Its ability to separate pertinent signal components has led to a number of wavelet-based techniques which, with their characteristic frequency makeups and localized temporal aspects, supersede those based on traditional Fourier methods. In its continuous form, the CWT allows a powerful analysis of nonstationary signals, making it ideally suited for the high-resolution interrogation of the signals mentioned earlier, over a wide range of applications. In its discrete form, the DWT and its offshoots, the SWT and WPT, provide the basis of powerful methodologies for partitioning pertinent signal components, which serve as a basis for potent compression strategies. In conclusion, it has been shown that the wavelet transform is a flexible time–frequency decomposition tool that can form the basis of useful signal analysis, and coding schemes. It is envisaged that the future will see further application of the wavelet transform to biomedical signal analysis, as the emerging technologies based on them are honed for practical purposes.

Paul S. Addison is a R&D specialist within the Respiratory and Monitoring Division of Covidien. He has written numerous technical papers and two books in the field of signal analysis, including the *Illustrated Wavelet Transform Handbook*. He cofounded CardioDigital Ltd. in the United Kingdom and more recently set up its U.S. subsidiary CardioDigital Inc., Portland, Oregon. His skill set includes a range of commercial expertise, R&D team leading, out-of-the-box thinking for signal-processing algorithm design, and the development of novel mathematical methods in signal analysis. His research interests include the implementation of novel technology algorithms within next-generation medical devices.

James Walker received his doctorate degree from the University of Illinois, Chicago, in 1982. He has been a professor of mathematics at the University of Wisconsin-Eau Claire since 1982. He has published articles on Fourier analysis, wavelet analysis, complex variables, and logic, and he is the author of five books on Fourier analysis, fast Fourier transforms (FFTs), and wavelet analysis.

Rodrigo C. Guido received his B.Sc. degrees in computer science and computer engineering from São Paulo State University (UNESP), São José do Rio Preto, Brazil, and Educational

Foundation, Votuporanga (FEV), Brazil, respectively, in 1998, his M.Sc. degree in electrical engineering from Campinas State University (UNICAMP), Brazil, in 2000, and his Ph.D. degree in computational applied physics from the University of São Paulo, São Carlos (USP), Brazil, in 2003. He has already participated in two postdoctoral programs in signal processing at USP, from 2003 to 2007, and obtained the title of associate professor (Livre-Docência) in digital signal processing in 2008. He has served as a guest editor for many special issues of IEEE and Elsevier journals, as an organizer and chair for many IEEE conferences, and as a member of the editorial board of scientific journals, including *IEEE Signal Processing Magazine*, *Pattern Recognition Letters*, *Neurocomputing*, and *International Journal of Semantic Computing*. The main objective of his group's work is concentrated on digital speech processing, specially based on wavelets associated with algorithms for machine learning.

Address for Correspondence: Paul S. Addison, CardioDigital Ltd., Elvingston Science Centre, East Lothian, EH33 1EH Scotland, U.K. E-mail: p.addison@cardiodigital.com.

References

[1] Abramovich, et al., "Wavelet analysis and its statistical implications," *Statistician*, vol. 49, no. 1, pp. 1–29, 2000.
[2] P. S. Addison, "Wavelet analysis of breakdown of a pulsed vortex flow," *Proc. Inst. Mech. Eng. C, Mech. Eng. Sci*, vol. 213, pp. 217–229, 1999.
[3] P. S. Addison, J. N. Watson, and T. Feng, "Low-oscillation complex wavelets," *J. Sound Vib.*, vol. 21, pp. 58–61, 2002.
[4] J. N. Watson, P. S. Addison, G. R. Cleg, M. Holzer, F. Sterz, and C. E. Robertson, "Evaluating arrhythmias in ECG signals using wavelet transforms," *IEEE Eng. Med. Biol. Mag.*, vol. 19, no. 5, pp. 104–109, 2002.
[5] A. Arneodo, B. Audit, E. Bacry, S. Manneville, J. F. Muzy, and S. G. Roux, "Thermodynamics of fractal signals based on wavelet analysis: Application to fully development turbulence data and DNA sequences," *Phys. A*, vol. 254, pp. 24–45, 1998.
[6] R. B. Ash, *Information Theory*. New York: Dover, 1990.
[7] T. C. Bell, J. G. Cleary, and I. H. Witten, *Text Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
[8] C. Blatter, *Wavelets: A Primer*. Natick, MA: Peters, 1998.
[9] Bruce and Adhami, "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," *IEEE Trans. Med. Imag.*, vol. 18, no. 12, pp. 1170–1177, 1999.
[10] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 532–540, 1983.
[11] R. Carmona and W. L. Hwang, "Characterization of signals by the ridges of their wavelet transform," *IEEE Trans. Signal Processing*, vol. 45, no. 10, pp. 2586–2590, 1997.
[12] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 45, pp. 485–560, 1992.
[13] R. R. Coifman and D. L. Donoho, "Translation invariant de-noising," *Lect. Not. Stat.*, vol. 103, pp. 125–150, 1995.
[14] S. Collineau and Y. Brunet, "Detection of turbulent coherent motions in a forest canopy—Part 1: Wavelet analysis," *Boundary Layer Meteorol.*, vol. 65, pp. 357–379, 1993.
[15] I. Daubechies, *Ten Lectures on Wavelets* (CBMS-NSF Regional Conference Series in Applied Mathematics). Philadelphia, PA: SIAM, 1992.
[16] G. M. Davis, "A wavelet-based analysis of fractal image compression," *IEEE Trans. Image Processing*, vol. 7, pp. 141–154, 1998.
[17] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, Ph. Tchamitchian, and B. Torrésani, "Asymptotic wavelet and gabor analysis: Extraction of instantaneous frequencies," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 644–664, 1992.
[18] D. Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data," in *Different Perspectives on Wavelets*, S. Mallat, Ed. Providence, RI: AMS, 1993, pp. 173–205.
[19] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1995.
[20] P. Dupuis and C. Eugene, "Combined detection of respiratory and cardiac rhythm disorders by high-resolution differential measurement," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 3, pp. 498–502, 2000.
[21] D. J. Field, "Scale invariance and self-similar "wavelet" transforms: An analysis of natural scenes and mammalian visual systems," in *Wavelets, Fractals and Fourier Transforms*. Oxford: Clarendon Press, 1993, pp. 151–193.
[22] D. J. Field, "What is the goal of sensory coding?" *Neural Comput.*, vol. 6, pp. 559–601, 1994.
[23] D. J. Field, "Wavelets, vision and the statistics of natural scenes," *Philos. Trans. R. Soc.*, vol. 357, pp. 2527–2542, 1999.
[24] D. J. Field and N. Brady, "Wavelets, blur and the sources of variability in the amplitude spectra of natural scenes," *Vision Res.*, vol. 37, pp. 3367–3383, 1997.
[25] Fonseca, et al., "Wavelet time-frequency analysis and least-squares support vector machine for the identification of voice disorders," *Comput. Biol. Med.*, vol. 37, pp. 571–578, 2007.
[26] M. J. Gormish, D. Lee, and M. W. Marcellin, "JPEG 2000: Overview, architecture, and applications," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, Sept. 2000, vol. 2, pp. 29–32.
[27] P. Goupillaud, A. Grossman, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, pp. 85–102, 1984.
[28] Guido, et al., "A new technique to construct a wavelet transform matching a specified signal with applications to digital, real-time, spike and overlap pattern recognition," *Digital Signal Process*, vol. 16, pp. 24–44, 2006.
[29] S. Guilbaud and B. Audoin, "Measurements of the stiffness coefficients of a viscoelastic composite material with laser generated and detected ultrasound," *J. Acoust. Soc. Amer.*, vol. 105, no. 4, pp. 2226–2235, 1999.
[30] L. J. Hadjileontiadis and S. M. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 12, pp. 1269–1281, 1997.
[31] D. K. Hammond and E. P. Simoncelli, "Image denoising with an orientation-adaptive Gaussian scale mixture model," in *Proc. 13th IEEE Int. Conf. Image Processing*, Atlanta, GA, 2006.
[32] B. Hendricks, H. Choi, and R. Baraniuk, "Analysis of texture segmentation using wavelet-domain hidden Markov trees," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 1999.
[33] H. Higuchi, J. Lewalle, and P. Crane, "On the structure of a two-dimensional wake behind a pair of flat plates," *Phys. Fluids*, vol. 6, no. 1, pp. 297–305, 1994.
[34] B. Jähne, *Digital Image Processing*, 5th ed. New York: Springer-Verlag, 2002.
[35] S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels, "Wavelet transform-based QRS complex detector," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 7, pp. 838–848, 1999.
[36] M. Khalil and Duchene, "Uterine EMG analysis: A dynamic approach for change detection and classification," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 6, pp. 748–755, 2000.
[37] S.-H. Leea, H. Zahouania, R. Caterini, and T. G. Mathiaa, "Morphological characterisation of engineered surfaces by wavelet transform," *Int. J. Mach. Tools Manufact.*, vol. 38, no. 5/6, pp. 581–589, 1998.
[38] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Inter-scale orthonormal wavelet thresholding," *IEEE Trans. Image Processing*, vol. 16, pp. 593–606, 2007.
[39] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, 1989.
[40] D. Marr, *Vision.* San Francisco, W. H. Freeman, Ed., 1982.
[41] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.
[42] Marrone, et al., "Wavelet analysis of blood pressure waves in vasovagal syncope," *Phys. A*, vol. 271, pp. 458–469, 1999.
[43] D. Mumford and J. Shah, "Boundary detection by minimizing functionals. I," in *Image Understanding*. New York: Ablex Press, 1989.
[44] Petrosian, et al., "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," *Neurocomputing*, vol. 30, pp. 201–230, 2000.
[45] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, pp. 1338–1351, 2003.
[46] T. K. Sarkar, C. Su, R. Adve, M. Salazar-Palma, G. L. Arcia-Castillo, and R. R. Boix, "A tutorial on wavelets from an electrical engineering perspective—Part 1: Discrete wavelet techniques," *IEEE Antennas Propagat. Mag*, vol. 40, no. 5, pp. 49–70, 1998.
[47] Scalassara, et al., "Relative entropy measures applied to healthy and pathological voice characterization," *Appl. Math. Comput.*, vol. 207, pp. 95–108, 2009.
[48] Spadotto, et al., "Classification of normal swallowing and oropharyngeal dysphagia using wavelet," *Appl. Math. Comput.*, vol. 207, pp. 75–82, 2009.
[49] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Boston, MA: Kluwer, 2002.
[50] J. Tian and R. O. Wells, "A lossy image codec based on index coding," in *Proc. IEEE Data Compression Conf., DCC '96*, 1996, p. 456.
[51] J. Tian and R. O. Wells, Jr., "Embedded image coding using wavelet difference reduction," in *Wavelet Image and Video Compression*, P. Topiwala, Ed. Norwell, MA: Kluwer, 1998, pp. 289–301.
[52] J. S. Walker, *A Primer on Wavelets and their Scientific Applications*, 2nd ed. Boca Raton, FL: CRC, 2008.
[53] J. S. Walker, "Tree-adapted wavelet shrinkage," *Adv. Imaging Electron Phys.*, vol. 104, pp. 343–394, 2002.
[54] J. S. Walker and T. Q. Nguyen, "Wavelet-based image compression," in *Handbook of Transforms and Data Compression*, P. C. Yip and K. R. Rao, Eds. Boca Raton, FL: CRC, 2000, ch. 6, pp. 267–312.
[55] J. S. Walker, T. Q. Nguyen, and Y.-J. Chen, "A low-power, low-memory system for wavelet-based image compression," in *Recent Research Developments in Optical Engineering*, vol. 5, 2002, pp. 111–126.
[56] B. A. Wandell, *Foundations of Vision*. Sunderland, MA: Sinauer Associates, 1995.
[57] A. B. Watson, "Efficiency of a model human image code," *J. Opt. Soc. Amer.*, vol. 4, pp. 2401–2417, 1987.
[58] Williams and Armatunga, "An introduction to wavelets in engineering," *Int. J. Numer. Methods Eng.*, vol. 37, pp. 2365–2388, 1994.
[59] P. S. Addison, *Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine, and Finance*. London: Taylor and Francis, 2002, ch. 2–3.