

Mixed-Domain Signal Analysis

This final chapter explains the methods for using time-frequency or time-scale transforms to segment, classify, and interpret signals. The previous two chapters introduced these *mixed-domain* transforms and their application to elementary analysis tasks. The short-time Fourier (or Gabor) transform (Chapter 10) and the wavelet transform (Chapter 11) are the main tools for the applications we discuss. The applications explain their practical and efficient use, spotlight their strengths and weaknesses, and contrast them with pure time- and frequency-domain techniques.

This chapter covers three methods that, together with the local frequency or scale information given by the transforms, are capable of elucidating signal structure:

- A type of structured neural network, which we call the *pattern recognition network*;
- The *hidden Markov model* (HMM), which has become very popular for speech, text, and biological sequence analysis;
- The *matching pursuit*, a Hilbert space search technique for efficient signal description using a dictionary of signal models.

In place of a summary, there is an afterword to the entire book.

12.1 WAVELET METHODS FOR SIGNAL STRUCTURE

This section follows up on the theoretical work of the previous chapter. There we discovered a special tool for describing multiscale signal structure, the multiresolution analysis (MRA) of finite-energy signals. Now we want to explain how, working within the framework of a chosen MRA, we can:

- (i) Develop a discrete version of the wavelet transform.
- (ii) Show how an efficient algorithm for signal decomposition arises from the MRA of finite-energy signals.

- (iii) Link this result to the perfect reconstruction filter banks covered in Chapter 9.
- (iv) And, finally, show how to employ these methods for analyzing signal shape across many scales.

12.1.1 Discrete Wavelet Transform

Let us assume that we have selected a multiresolution analysis of square-integrable analog signals. Section 11.4 introduced this theory. To recapitulate, an MRA [1] is a chain of subspaces $\{V_i: i \in \mathbb{Z}\}$ in $L^2(\mathbb{R})$ such that:

- (i) The V_i are closed and nested: $\dots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots$
- (ii) Their union is dense in $L^2(\mathbb{R})$: $\overline{\bigcup_{n=-\infty}^{\infty} V_i} = L^2(\mathbb{R})$.
- (iii) The only signal common to all the V_i is the signal that is zero almost everywhere: $\bigcap_{i=-\infty}^{\infty} V_i = \{0\}$.
- (iv) Dilation by a factor of two links the closed subspaces: $x(t) \in V_i \Leftrightarrow x(2t) \in V_{i+1}$.
- (v) The root space V_0 is closed under integral translation: If $x(t) \in V_0$ and $k \in \mathbb{Z}$, then $x(t-k) \in V_0$.
- (vi) There is a bounded, one-to-one, linear map, with a bounded inverse $I: V_0 \rightarrow l^2$ that commutes with integral translation: If $k \in \mathbb{Z}$, and $I(x(t)) = s(n) \in l^2$, then $I(x(t-k)) = s(n-k)$.

Property (vi) is equivalent to the existence of a Riesz basis within V_0 (Section 11.4.1.4). The previous chapter provided examples of MRAs: step functions, piecewise linear functions, and cubic spline functions. Most importantly, associated with an MRA $\{V_i: i \in \mathbb{Z}\}$ are three special signals:

- (i) An analog *scaling function*, $\phi(t) \in V_0$, such that $\{\phi(t-k): k \in \mathbb{Z}\}$ is an orthonormal basis of V_0 .
- (ii) A discrete *associated lowpass filter* H_ϕ with impulse response, $h(n) = h_n$, given by $h_n = \langle \frac{1}{2}\phi(\frac{t}{2}), \phi(t-n) \rangle$, where $\phi(t)$ is the scaling function in (i).
- (iii) An analog *orthogonal wavelet* $\psi(t)$, defined by its Fourier transform as follows: $\Psi(2\omega) = e^{-j\omega} \overline{H(\omega + \pi)} \Phi(\omega)$, where $\phi(t)$ is the scaling function of (i), $\Phi = \mathcal{F}(\phi)$ is the Fourier transform of ϕ , and $H = H_\phi$ is the associated low-pass filter of (ii).

To make the MRA signal decomposition discrete, we assume that the analog source signals reside in root space $x(t) \in V_0$ and that we sample them at unit distance to get $x(n)$. The scaling function is a lowpass filter, and its expanding dilations—for example, $\phi(t/2)$, $\phi(t/4)$, $\phi(t/8)$, and so on—have successively narrower passbands. Filtering $x(t)$ by these dyadic dilations of $\phi(t)$ produces approximate versions of $x(t)$

which are increasingly smooth. Next, since some high frequency detail has been removed from $x(t)$ by the filtering, we may select samples that are further apart. For example, after the convolution $x * \phi(t/4)$ removes noise and sharp transitions from $x(t)$, sampling occurs on quadruple unit intervals. This idea comes from the Laplacian pyramid decomposition, covered at the end of Chapter 9 [2].

Let us formalize these ideas for the case of the MRA. Following the notation of Ref. 1, let $x_a(t) = ax(at)$ be the scaled dilation of $x(t)$ by factor a . Typically, $a = 2^i$ for $i \in \mathbb{Z}$. Then the *discrete approximate* representation of signal $x(t)$ at resolution 2^i is

$$(A_i^d x)(n) = (x(t) * \phi_{2^i}(-t))(2^{-i}n). \quad (12.1a)$$

On the other hand, the orthogonal wavelet $\psi(t)$ is a bandpass filter. Its dilation by various dyadic factors results in filters with narrower passbands and lower center frequencies. Thus, we define the *discrete detail* representation of $x(t)$ at resolution 2^i :

$$(D_i^d x)(n) = (x(t) * \psi_{2^i}(-t))(2^{-i}n). \quad (12.1b)$$

Although (12.1a) and (12.1b) discretize the decomposition of a square-integrable signal $x(t)$, it remains to see how to compute the various analog convolutions that are required.

12.1.2 Wavelet Pyramid Decomposition

The *discrete orthogonal wavelet representation* or *wavelet pyramid decomposition* consists of the following filtered and coarsely sampled discrete signals:

$$A_{-J}^d x, D_{-J}^d x, D_{-J+1}^d x, \dots, D_{-1}^d x. \quad (12.2)$$

Notice in (12.2) that only the pyramid maintains all of the detail signals, up to the decomposition level $-J$, but only the coarsest approximate representation of $x(t)$. Let us turn our attention to the convolution operations needed to derive this special structural description.

12.1.2.1 Coarse Signal Structure: The Approximate Signal. We concentrate on deriving the coarse signal approximation, $A_{-J}^d x$ in the pyramid decomposition (12.2). The next section explains how to derive the detailed structural descriptions.

Again, let $\{V_i\}_{i \in \mathbb{Z}}$ be a multiresolution analysis of $L^2(\mathbb{R})$, $\phi \in V_0$ be its scaling function, and $\psi \in V_1$ be its orthonormal wavelet (Section 11.4.4). We define $\tilde{h}(n) = h(-n)$ to be the reflection of $h(n)$, the impulse response of the associated

low-pass filter $H = H_\phi$. (Since a scaling function ϕ is known for the MRA, we drop the subscript.)

We rewrite the convolution (12.1a) as an inner product:

$$(A_i^d x)(n) = 2^i \langle x(t), \phi(2^i t - n) \rangle, \quad (12.3)$$

Now let $p > 0$ and i be integers, let $x \in L^2(\mathbb{R})$, and let H_p be the discrete filter with impulse response h_p :

$$h_p(n) = 2^{-p} \langle \phi(2^{-p} t), \phi(t - n) \rangle = \langle \phi_{2^{-p}}(t), \phi(t - n) \rangle. \quad (12.4)$$

Note that $h_1(n) = h(n)$, the impulse response of the quadrature mirror filter associated to the multiresolution analysis $\{V_i\}$.

Then we claim that the decomposition for the discrete approximate representation of signal $x(n)$ at level i is given in terms of the approximate representation of $x(n)$ at level $i + p$ by

$$(A_i^d x)(n) = \sum_{k=-\infty}^{\infty} \tilde{h}_p(2^p n - k) (A_{i+p}^d x)(k). \quad (12.5)$$

This means that we can get completely rid of the analog convolutions through which we originally defined the pyramid. Indeed, if we take $p = 1$ in (12.5), then each approximate level of representation comes from convolution with $\tilde{h}_p(n)$ followed by dyadic subsampling. This continues recursively for p levels to produce the approximate signal structure of $x(n)$ at level i from the approximate representation at level $i + p$.

To show how this works, we consider the expansion of V_i signals on the orthogonal basis elements of V_{i+p} . For any i , the signals $\{\phi_{2^i} t - 2^{-i} n\}_{n \in \mathbb{Z}}$ span $V_i \subset V_{i+p}$. Indeed, an orthonormal basis of V_{i+p} is $\{2^{-(i+p)/2} \phi_{2^{i+p}}(t - 2^{-i-p} n)\}_{n \in \mathbb{Z}}$. Consequently,

$$\phi(2^i t - n) = 2^{i+p} \sum_{k=-\infty}^{\infty} \langle \phi(2^i s - n), \phi(2^{i+p} s - k) \rangle \phi(2^{i+p} t - k). \quad (12.6)$$

With a change of variables $s = 2^{-i-p}(t + 2^p n)$, the inner product in (12.6) is

$$\begin{aligned} \int_{-\infty}^{\infty} \phi(2^i s - n) \phi(2^{i+p} s - k) ds &= 2^{-i-p} \int_{-\infty}^{\infty} \phi(2^{-p} t) \phi(t + 2^p n - k) dt \\ &= 2^{-i-p} \int_{-\infty}^{\infty} \phi(2^{-p} s) \phi(s - (k - 2^p n)) ds. \end{aligned} \quad (12.7)$$

Putting (12.6) back into (12.7), it follows that

$$\langle x(t), \phi(2^i t - n) \rangle = \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(2^{-p} s) \phi(s - (k - 2^p n)) ds \langle x(t), \phi(2^{i+p} t - k) \rangle. \quad (12.8)$$

From the definition of the impulse response h_p (12.4), we get

$$\langle x(t), \phi(2^i t - n) \rangle = 2^p \sum_{k=-\infty}^{\infty} \tilde{h}_p(2^p n - k) \langle x(t), \phi(2^{i+p} t - k) \rangle. \quad (12.9)$$

But the inner products in (12.9) are in fact the discrete representations of signal $x(t)$ at levels i and $i + p$. So (12.5) follows directly.

Let us find the impulse response of the discrete filter H_p . Since $\{\phi(t - k)\}_{k \in \mathbb{Z}}$ is a basis for V_0 ,

$$\phi(2^{-p} t) = \sum_{k=-\infty}^{\infty} \langle \phi(2^{-p} s), \phi(s - k) \rangle \phi(t - k). \quad (12.10)$$

Taking radial Fourier transforms on both sides of (12.10) and simplifying,

$$\Phi(2^p \omega) = \sum_{k=-\infty}^{\infty} h_p(k) \int_{-\infty}^{\infty} \phi(t - k) e^{-j\omega t} dt = H_p(\omega) \Phi(\omega). \quad (12.11)$$

The discrete-time Fourier transform of the filter H_p is

$$H_p(\omega) = \frac{\Phi(2^p \omega)}{\Phi(2\omega)}, \quad (12.12)$$

where $\Phi(\omega)$ is the radial Fourier transforms of scaling function $\phi(t)$. Applying the inverse discrete time Fourier transform to (12.12), gives $h_p(n)$ (Figure 12.1).

Figure 12.2 shows the $H_p(\omega)$. (12.18) lists filter values for the cubic spline MRA. Note that the $h_p(n)$ are even.

To extract a coarse approximation of a signal's structure using a given MRA, then, the steps are:

- (i) Select a resolution step factor $p > 0$.
- (ii) Compute the impulse response $h_p(n)$ of the filter with discrete time Fourier transform given by (12.12).

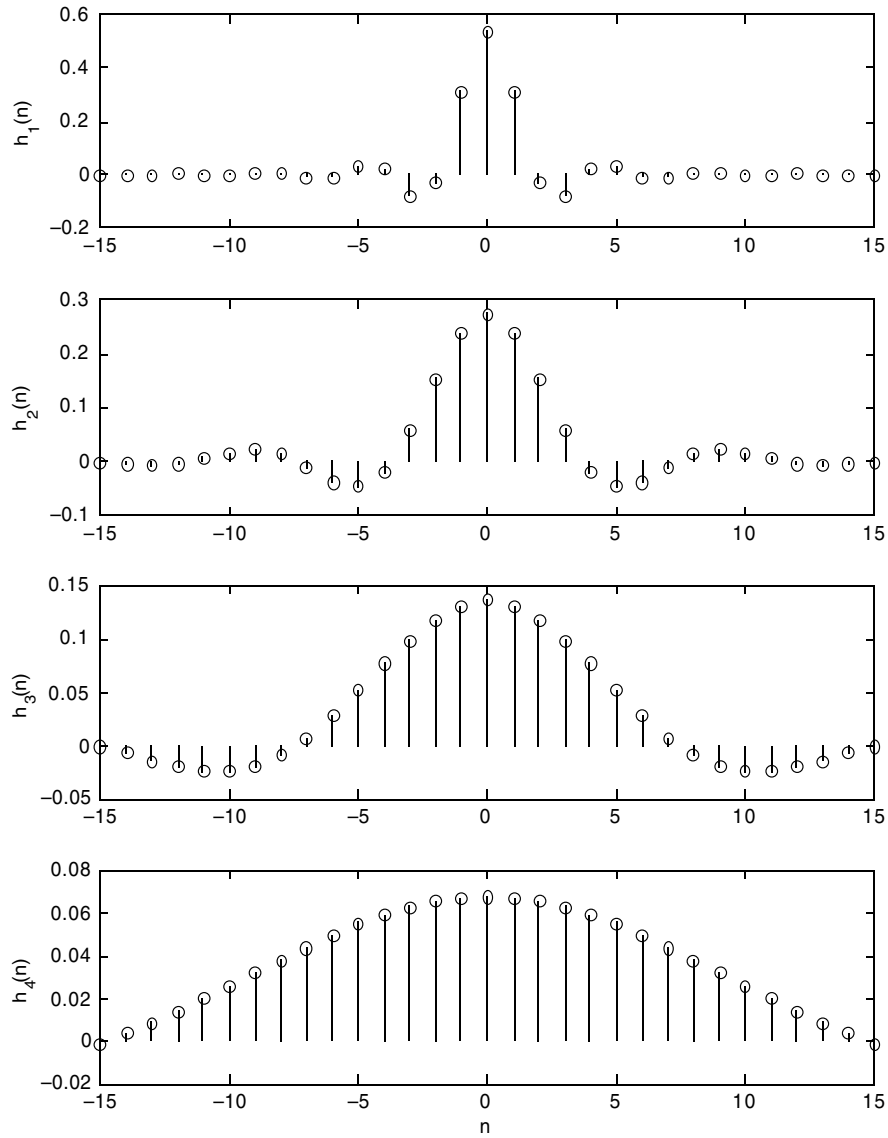


Fig. 12.1. Discrete filter impulse response signals h_1 , h_2 , h_3 , and h_4 for the cubic spline MRA.

- (iii) Compute the convolution (12.5) on 2^p -wide intervals to get a coarse approximation $(A_i^d x)(n)$ from $(A_{i+p}^d x)(n)$.
- (iv) Employ one of Chapter 4's thresholding methods to the magnitude of the decomposition coefficients, identifying large and small coefficients with significant signal features and background noise, respectively.

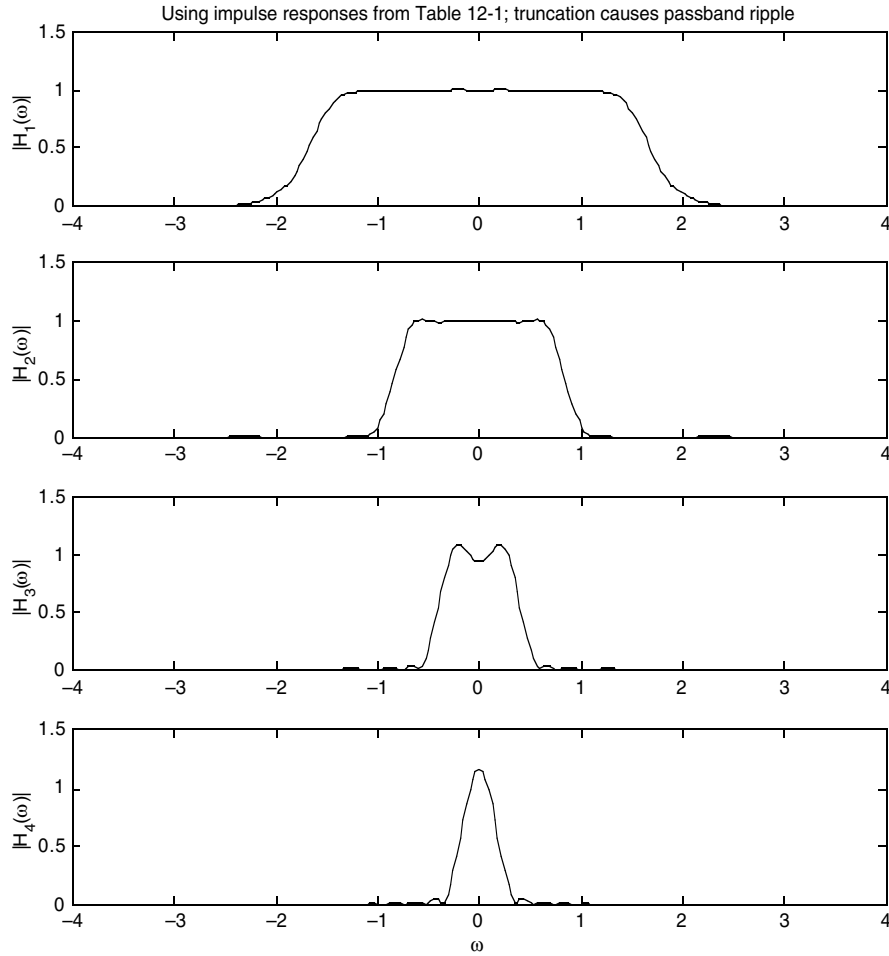


Fig. 12.2. Discrete-time Fourier transforms H_1 , H_2 , H_3 , and H_4 of h_1 , h_2 , h_3 , and h_4 for the cubic spline MRA.

12.1.2.2 Fine Signal Structure: The Detail Signal. A similar derivation gives the signal details at various resolutions. Let G_p be the discrete filter with impulse response g_p ,

$$g_p(n) = 2^{-p} \langle \psi(2^{-p}t), \phi(t-n) \rangle, \quad (12.13)$$

and note again that (12.1b) expresses an inner product:

$$(D_i^d x)(n) = 2^i \langle x(t), \psi(2^i t - n) \rangle. \quad (12.14)$$

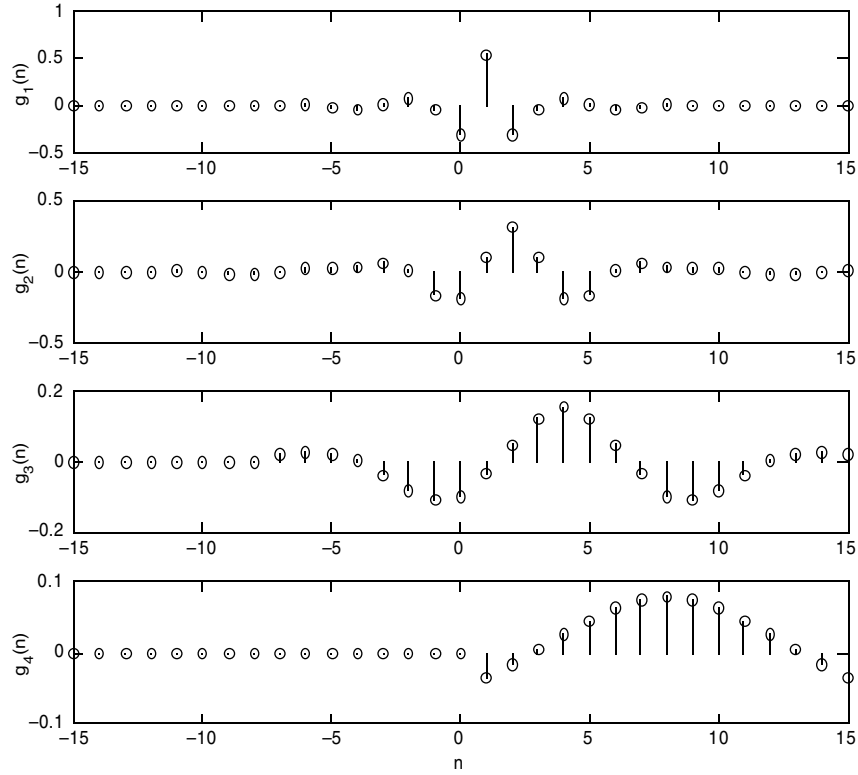


Fig. 12.3. Discrete filter impulse response signals g_1 , g_2 , g_3 , and g_4 for the cubic spline MRA.

We claim that $(D_i^d x)(n)$, the detail signal at level i , is given in terms of the approximate signal at level $i + p$ by

$$(D_i^d x)(n) = \sum_{k=-\infty}^{\infty} \tilde{g}_p(2^p n - k) (A_{i+p}^d x)(k). \quad (12.15)$$

To verify this, let O_i be the orthogonal complement of V_i inside V_{i+1} : $V_i \perp O_i$ and $V_{i+1} = V_i \oplus O_i$. The shifted, dilated orthogonal wavelets $\{\psi_{2^i t - 2^{-i} n}\}_{n \in \mathbb{Z}}$ span $O_i \subset V_{i+p}$. Since $\{2^{-(i+p)/2} \phi_{2^{i+p}(t - 2^{-i-p} n)}\}_{n \in \mathbb{Z}}$ is an orthonormal basis of V_{i+p} ,

$$\psi(2^i t - n) = 2^{i+p} \sum_{k=-\infty}^{\infty} \langle \psi(2^i s - n), \phi(2^{i+p} s - k) \rangle \phi(2^{i+p} t - k). \quad (12.16)$$

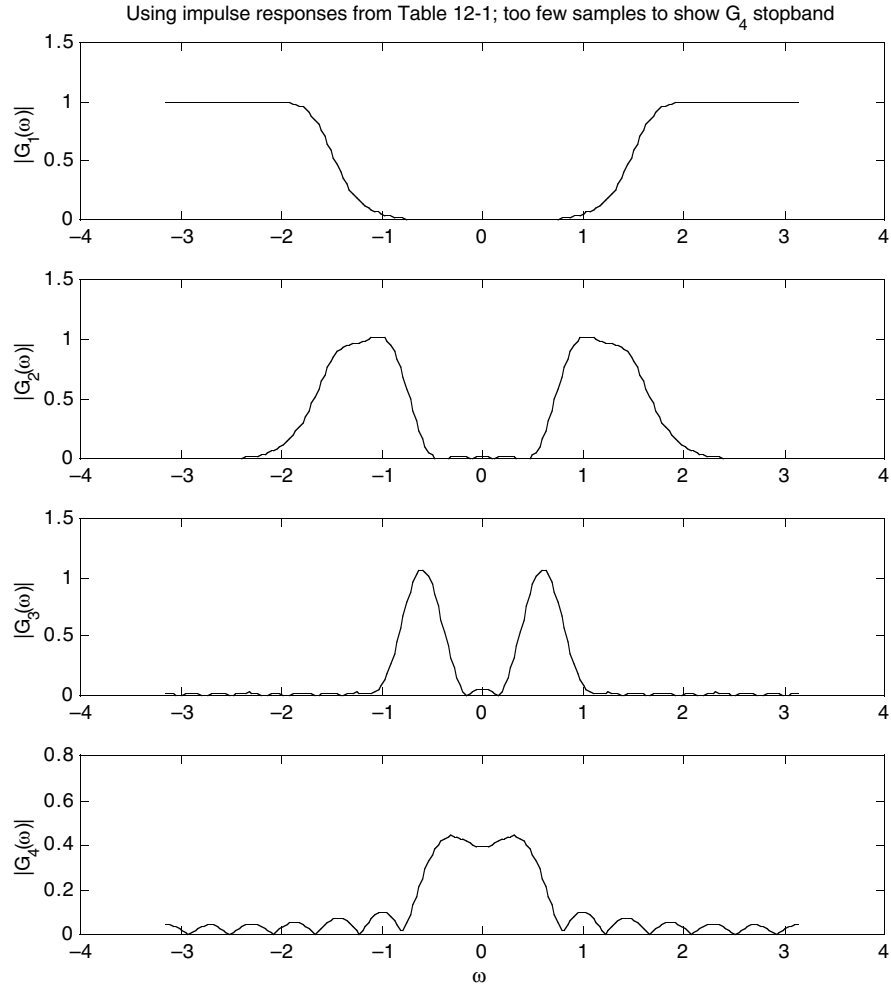


Fig. 12.4. Discrete-time Fourier transforms G_1 , G_2 , G_3 , and G_4 of g_1 , g_2 , g_3 , and g_4 for the cubic spline MRA.

Then, following the argument about approximate signals from the previous section (exercise) gives

$$\langle x(t), \psi(2^i t - n) \rangle = 2^p \sum_{k=-\infty}^{\infty} \tilde{g}_p(2^p n - k) \langle x(t), \phi(2^{i+p} t - k) \rangle, \quad (12.17)$$

and, consequently,

$$(D_i^d x)(n) = \sum_{k=-\infty}^{\infty} \tilde{g}_p(2^p n - k) (A_{i+p}^d x)(k). \quad (12.18)$$

TABLE 12.1. Cubic Spline MRA Orthogonal Wavelet Pyramid Filters

n	$h_1(n)$	$h_2(n)$	$h_3(n)$	$h_4(n)$	$g_2(n)$	$g_3(n)$	$g_4(n)$
0	0.542	0.272	0.136	0.068	-0.189	-0.095	-0.048
1	0.307	0.237	0.131	0.067	0.099	-0.035	-0.035
2	-0.035	0.153	0.118	0.066	0.312	0.049	-0.018
3	-0.078	0.057	0.099	0.063	0.099	0.125	0.003
4	0.023	-0.019	0.077	0.059	-0.189	0.157	0.025
5	0.030	-0.047	0.052	0.055	-0.161	0.125	0.045
6	-0.012	-0.039	0.028	0.050	0.005	0.049	0.062
7	-0.013	-0.013	0.007	0.044	0.054	-0.035	0.074
8	0.006	0.012	-0.009	0.038	0.027	-0.095	0.079
9	0.006	0.020	-0.019	0.032	0.018	-0.107	0.074
10	-0.003	0.015	-0.024	0.026	0.017	-0.080	0.062
11	-0.003	0.004	-0.023	0.020	0.000	-0.037	0.045
12	0.002	-0.006	-0.019	0.014	-0.018	0.003	0.025
13	0.001	-0.009	-0.013	0.009	-0.016	0.023	0.003
14	-0.001	-0.006	-0.006	0.004	-0.004	0.027	-0.018
15	-0.001	-0.001	0.000	-0.001	0.003	0.021	-0.035

Seeking the impulse response of the filter G_p , we expand on $\{\phi(t-k)\}_{k \in \mathbb{Z}}$, an orthonormal basis for V_0 :

$$\psi(2^{-p}t) = \sum_{k=-\infty}^{\infty} \langle \psi(2^{-p}s), \phi(s-k) \rangle \phi(t-k). \quad (12.19)$$

Fourier transformation of (12.19) produces

$$G_p(\omega) = \frac{\Psi(2^p \omega)}{\Phi(2\omega)}. \quad (12.20)$$

Filters for generating the detail structure of signals via the orthogonal wavelet decomposition are shown in Table 12.1(12.18). We set $g(n) = g_1(n) = (-1)^{1-n}h(1-n)$, so it is not shown. Observe that $g_p(n)$ is symmetric about 2^{p-1} .

12.1.2.3 Quadrature Mirror Filters. We have shown that discrete filters, H_p and G_p , with impulse responses h_p and g_p , respectively, are all that we need for the wavelet pyramid decomposition (12.2). Since we know the Fourier transforms of the wavelet and scaling function, we can compute these impulse responses from the inverse discrete-time Fourier transforms of (12.12) and (12.20). Figure 12.5 shows how the pyramid decomposition occurs by successive filtering and subsampling operations.

From Chapter 11's theoretical development, we know that the discrete low-pass filter $H(\omega)$ associated to an MRA satisfies $|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1$. Within an amplification factor, this is precisely the perfect reconstruction criterion of Chapter 9. In fact, $\sqrt{2}h(n)$ is a quadrature mirror filter. We can decompose the signal using

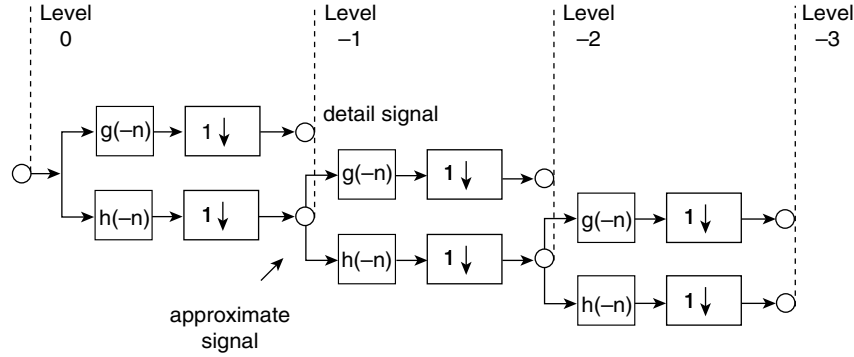


Fig. 12.5. Signal decomposition using the orthogonal wavelet pyramid.

this scaled filter, or we can slightly modify the perfect reconstruction scheme of Chapter 9 by supplying an additional amplification factor upon reconstruction [1, 3].

Consider the QMF pyramid decomposition in Figure 12.5. Let $\tilde{h}(n) = h(-n)$ be the reflection of $h(n)$ and

$$\tilde{H}(z) = \sum_{n=-\infty}^{\infty} \tilde{h}(n)z^{-n} \quad (12.21)$$

be the z -transform of $\tilde{h}(n)$. Subsampling by two followed by $\tilde{H}(z)$ filtering is the same discrete system as $\tilde{H}(z^2)$ filtering followed by subsampling [4] (exercise). Applying the same idea to $\tilde{g}(n)$, note that we can directly obtain level -2 coefficients by filtering with $\tilde{H}(z)H(z^2)$ and $\tilde{H}(z)\tilde{G}(z^2)$ and subsampling by four. We can compute the impulse response of the filter with transfer function $\tilde{H}(z)\tilde{H}(z^2)$ by convolving $h(n)$ with the filter obtained by putting a zero between every $h(n)$ value.

12.1.3 Application: Multiresolution Shape Recognition

This section shows how to use the orthogonal wavelet pyramid decomposition to recognize signal patterns at varying scales. This is one type of pattern recognition application [5]. We seek a known signal pattern, the *model*, in a *sample* signal that—perhaps along with background shapes and noise—contains a dilated version of the model. Moreover, if our pattern comparison algorithm gives localized information, we can attempt to *register* the time-domain position of the model as well. *Registration* is the process of finding the position of a *prototype*, or model, signal within a *candidate*, or sample, signal. For these tasks, signal analysts have relied upon multiple resolution methods for the following reasons.

- If the comparisons between the model and sample proceed pointwise, then the number of computations may become prohibitive—an especially acute problem in multiple dimensions (image analysis, computer vision, video analysis). Hierarchical structures that analyze signals at several resolutions can make the number of computations tractable [6]. Comparisons at coarse scales are iteratively improved in transition to fine scales.
- Coarse representations of signal structure can isolate significant features that are apparent only at certain resolutions [7–9].

When applying pyramid techniques to register a prototype object in a candidate signal, we first decompose both the model pattern and the sample. At the coarsest scale of decomposition, the algorithm compares the model and sample at all possible relative positions. The decomposition coefficients should not change as the model's offset into the sample varies. For otherwise, the decomposition would need to be recomputed for each location; computation time then increases drastically. The model's coefficients will not change if the decomposition procedure is translation-invariant. Of course, the coefficients could change in some simple way that is comparatively inexpensive to compute. Eventually, this produces a set of sufficiently good—or, *feasible*—comparison locations between prototype and candidate. The search continues at the next higher resolution with—hopefully—a greatly confined set of feasible registrations. The best acceptable match at the finest scale gives the final result.

We apply the multiresolution analysis (MRA) of $L^2(\mathbb{R})$ to the multiscale shape recognition problem [1]. Once a particular MRA is chosen, it leads to simple, compact, and efficient pyramid decompositions using quadrature mirror filter (QMF) banks (Figure 12.5). The algorithms do not increase the amount of memory space required for storing the representations, yet exactly reconstruct the original signal.

We note that the MRA concept extends to two (or more) dimensions for image analysis. In this case, the separable two-dimensional pyramid decomposition distinguishes between horizontal and vertical spatial frequencies, which is useful for texture and image analysis in artificial environments [1]. Supplementary orientation tunings are possible too [10].

Orthogonal wavelet pyramids suffer from the following difficulties in registration and matching applications:

- (i) A registration problem is that the lower resolution coefficients do not translate as the original signal is shifted; in fact, the decomposition coefficients change drastically (Figure 12.6). This greatly complicates the tasks of matching and registration and has inspired research into alternative representations that support pattern matching [11].
- (ii) A second registration difficulty arises from the orthogonality of the representation, a consequence of its derivation from an MRA of $L^2(\mathbb{R})$. A registration between prototype and candidate at one resolution may not indicate any correlation between them at a finer scale. Whether this second difficulty appears depends on the nature of the signals acquired by the processing

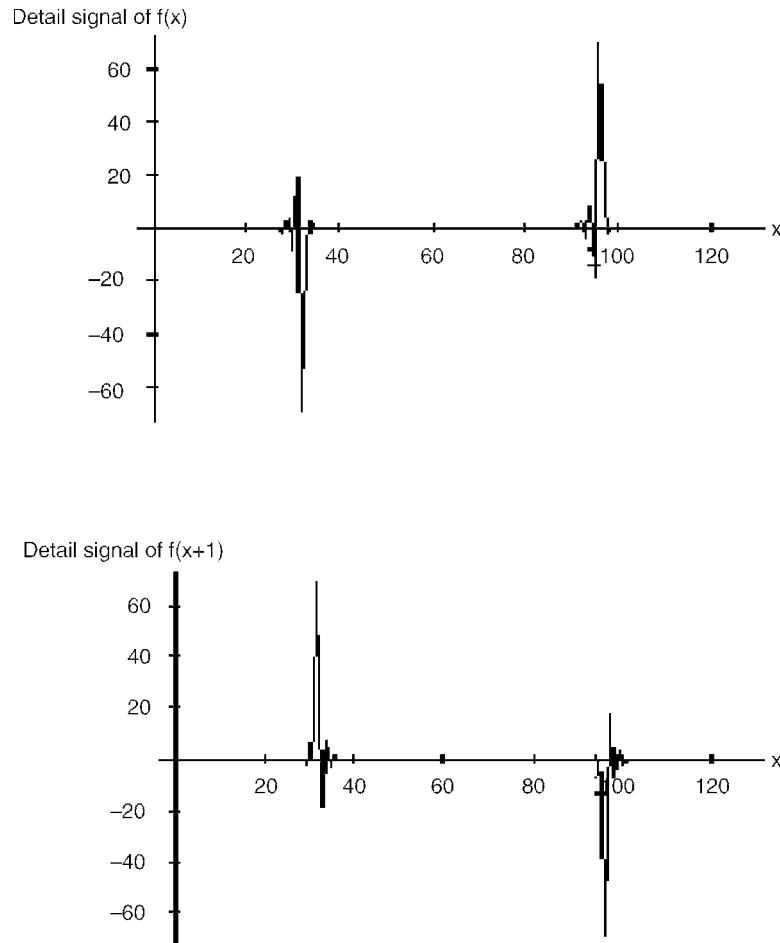


Fig. 12.6. Decomposing a prototype square pulse one resolution level using the orthogonal wavelet pyramid decomposition (a). The same pulse, shifted by a unit distance, represents the candidate. Its first coarse representation is shown in (b). This confounds the basic pyramid registration algorithm when it uses raw detail signals. The best registration position of candidate with respect to prototype is to align the left edge of one with the right edge of the other. On the other hand, the registration algorithm succeeds when using the magnitudes to measure of local energy.

system. Below, we describe an algorithm for coarse-to-fine tracking of registrations in orthogonal wavelet pyramids. The exercises suggest a comparison using the Laplacian pyramid [2] representation with the same registration algorithm. We report some earlier results here [5].

- (iii) A problem in multiscale matching is the dyadic dilation factor between pyramid levels. If the modeled object does not happen to be a dyadically scaled version of the candidate's pattern, then it is possible to overlook a match.

In matching applications, we seek instances of model patterns in an acquired signal. Suppose that prototypes are decomposed and stored in a phoneme recognition system, for example. The large number of prototypes needs a compact representation for the model pyramids to trim memory requirements. Applying orthogonal wavelets for such recognition systems is attractive and is a motivation for trying to circumvent the translation of coefficients problem. It is desirable to quickly reject candidate patterns that are not represented by any model. Decomposition of a full pyramid for a candidate pattern is costly when, for instance, only the coefficients at the fourth level of decomposition (1/16 of the total values) are used for comparison with prototypes. If the pattern is accepted, then the time constructing a full pyramid is not lost. It helps obtain a precise registration. But if the pattern is rejected, a full pyramid is built for a candidate even though a tiny fraction of the coefficients find use. One need not derive all pyramid levels, in sequence from finest to coarsest scale of representation, in order to reject or tentatively accept a candidate pattern. We derived formulas for filters that allow us to directly compute coarse pyramid levels in the orthogonal wavelet representation in Sections 12.1.2.1.2.

We use the well-known cubic spline MRA for registration experiments [12]. If $\{V_i : i \in \mathbb{Z}\}$ is the MRA, then the root space V_0 is all finite-energy continuously differentiable functions that are cubic polynomials on intervals $[k, k+1]$. We studied this example of an MRA in Section 11.4. It has a particularly suitable scaling function $\phi(t) \in V_0$ for signal analysis, with exponential decay in the time domain and polynomial decay of ω^{-4} in the frequency domain. We recall that the associated discrete low-pass filter is $h(n) = \langle \frac{1}{2}\phi(\frac{t}{2}), \phi(t-n) \rangle$. We set $g(n) = (-1)^{1-n}h(1-n)$, which is a discrete high-pass filter.

The registration algorithm begins with a set of *feasible points*, where sufficient correlation exists between the candidate and prototype signals, at the coarsest level of representation. Feasible points at finer scales are found, furnishing *feasible paths* up the pyramid. In order for a feasible point to continue a feasible path from a lower level, it must be close to a previous, coarser registration value. The best complete feasible path to the finest level of representation gives the registration between candidate and prototype patterns.

The algorithm uses the limited shift-invariance in the orthogonal wavelet pyramid representation. The coefficients of the wavelet representation at level $l < 0$ translate by amount k when the original signal is translated by amount $k2^{-l}$. The steps are:

- (i) The candidate signal is decomposed with the wavelet pyramid (Figure 12.5).
- (ii) The minimum of the registration cost function m over all registrations r ,

$$m(r, l) = \sum_i [X_{l,c}(i) - X_{l,p}(i-r)]^2, \quad (12.22)$$

is computed for level $l = -L$. $X_{l,c}$ is the candidate signal decomposition at level l , $X_{l,p}$ is prototype signal decomposition at level l , and i varies over candidate signal values. Let r_{-L} be a registration at which the minimum occurs, and call the minimum $M_{-L} = m(r_{-L}, -L)$. We pad arrays with zeros when endpoints overlap.

- (iii) All registrations s such that $m(s, -L) \leq Tm(r_{-L}, -L)$, where $T > 1$ is a threshold, are the feasible points at level $-L$. Call this set of registrations FR_{-L} , the starting points of feasible paths up the pyramid levels.
- (iv) Steps (ii) and (iii) repeat at higher levels $-L + k$, $1 \leq k < L$.
- (v) We prune the feasible point sets at the higher levels to retain only feasible points that continue, within an offset tolerance factor, a feasible path coming from a lower level. Thus, $s \in FR_{-k+1}$ only if $m(s, -k+1) \leq TM_{-k+1}$ and $t - \tau \leq s \leq t + \tau$ for some t in FR_{-k} and offset tolerance τ .
- (vi) Finally, if at level -1 no full feasible path has been found, then registration failed. If at least one feasible path is found, the best is selected as the final registration. In a local neighborhood of the best registration found at level -1 , the original prototype signal and original candidate signal are examined for the best correspondence value.
- (vii) To extend the algorithm to matching, where the scale of the candidate object is unknown, then we allow feasible paths to start and stop at intermediate levels.
- (viii) For matching applications based on coarse signal structure, it is useful to generate low-resolution pyramid levels directly rather than iteratively. The wide filters we developed above (Table 12.1(12.18)) allow us to jump many scales, quickly compare, and tentatively accept or reject candidate signals. This saves the time and memory cost of performing full pyramid decompositions. (12.18) Table 12.1 gives sample coefficients for $h_p(n)$ and $g_p(n)$. We use filters $g_1(n) = g(n)$, $g_2(n)$, ..., $g_p(n)$, and $h_p(n)$ for an orthogonal wavelet pyramid to level $-p$.

For experimentation, we register sections of the Australian coastline. Digitized, it contains 1090 points for the experiments herein. The scale-space representation has been studied in similar experiments [13]. Beginning from a reference zero position, we plot the outline of the Australian coast (Figure 12.7) as ordered pairs $(x(t), y(t))$, where t is the distance along the coastline to the reference position.

To find derivatives, we approximate $x(t)$ and $y(t)$ to a quadratic using Lagrange's interpolation formula. A signed curvature function [14],

$$\kappa(t) = \frac{(x'y'' - y'x'')}{[(x')^2 + (y')^2]^{3/2}}, \quad (12.23)$$

gives the coastline curvature at $(x(t), y(t))$ (cf. Figure 12.8a). Prototype signal $\kappa(t)$ is decomposed using the orthogonal wavelet decomposition (Figure 12.9a). The

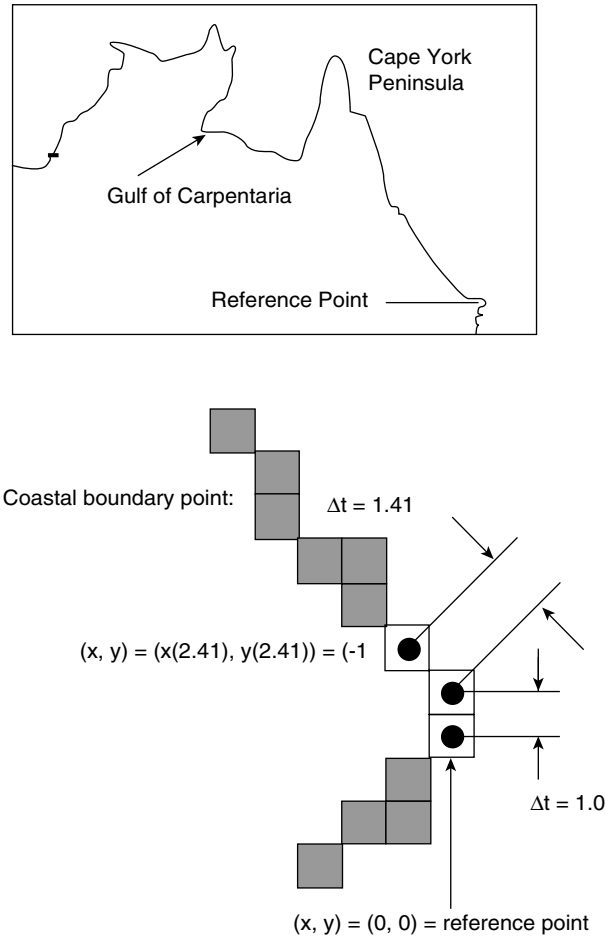


Fig. 12.7. Curvature maps of sections of coastline such as the Gulf of Carpentaria and the Cape York Peninsula are registered within an entire boundary curvature map of Australia. The independent variable of the curvature map is the arc-length distance from the reference point.

candidate signals for the experiments are partial curvature maps of coastal sections, for example the Gulf of Carpentaria (Figure 12.8b, Figure 12.9b).

The registration algorithm generates feasible paths through the pyramid (Figure 12.10). Thresholds T and τ depend on the application. For the boundary matching shown here, $T = 1.2$, and τ in Step (v) was chosen to be 2^{k+1} at level $-k$. If the value of T is too small, the registration can fail at the coarsest level; the r_L is incorrect, and the neighborhood of the correct registration holds no feasible point. When τ is too large, the number of feasible paths to check increases, slowing the coarse-to-fine

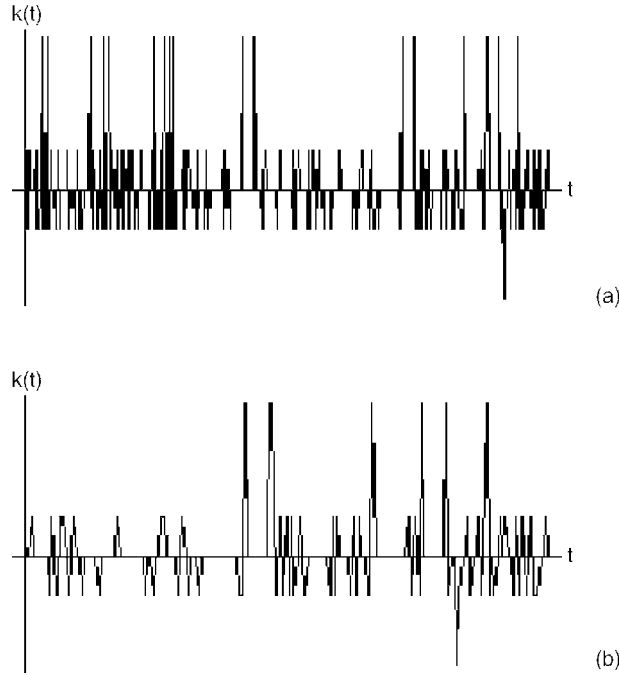


Fig. 12.8. Curvature maps. (a) Partial curvature map of Australian coastline. (b) Curvature map of Gulf of Carpentaria. The correct registration is evident.

algorithm. If τ is too small, initially correct feasible paths terminate at coarse resolutions. It turns out that this holds for both the Laplacian- and wavelet-based registration schemes [5]. The exercises suggest a comparison of these methods along with the piecewise continuous MRA [15].

Table 12.2 shows experimental results in registering continental boundaries. Over all but the lowest resolution level we rely on the detail signal magnitudes of the candidate and prototype signals. For the coarsest comparisons of structure, the approximate signals are used for matching. We add noise to the candidate curvature maps in some experiments. Note that the mean-square signal-to-noise ratio (SNR) employed here is $SNR = \sum s^2(t) / \sum N^2(t)$. Experiments reported in Ref. 5 were performed in which the signals $X_{l,c}$ and $X_{l,p}$ in (12.22) were taken to be either the approximate signal, the detail signal, or the absolute value of the detail signal coefficients. The outcome is problematic registration with the raw detail signals, but satisfactory convergence if $X_{l,c}$ and $X_{l,p}$ are approximate signals [5]. This is not unexpected, as the algorithm then compares successively low-pass filtered signals. The approximate signals are always used at the coarsest level (here, level -4) to generate the initial list of feasible points FR_{-4} . The feasible paths are robust when the candidate has large support, but smaller candidates of nearly straight coastline can fail to correctly register [5].

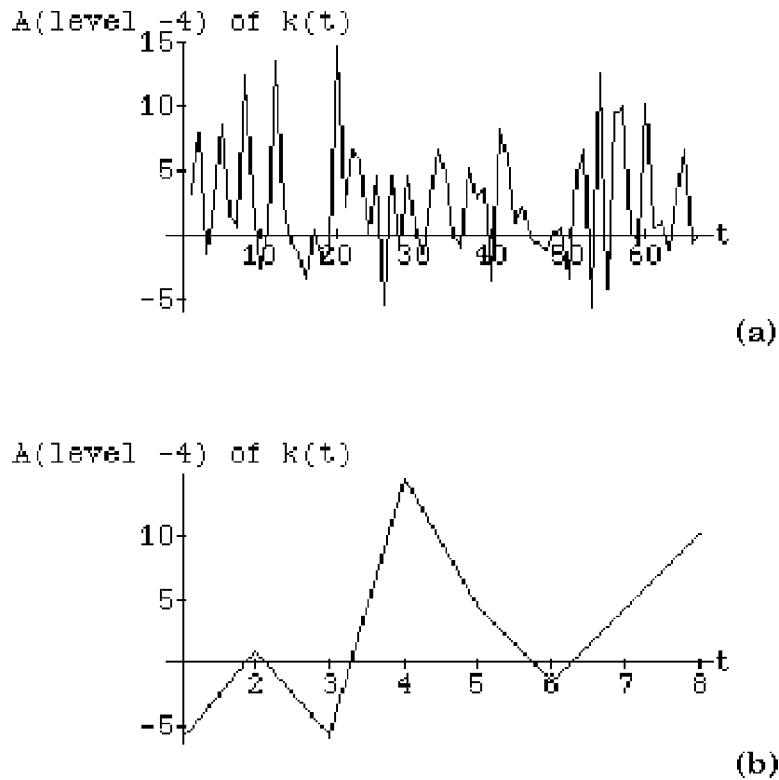
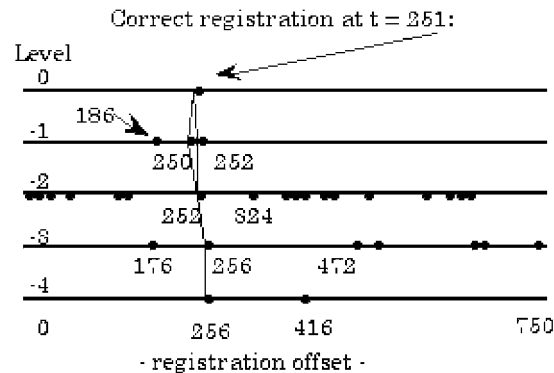


Fig. 12.9. Beginning registration of Gulf of Carpentaria at level -4. (a) Australian coastline curvature at resolution 1/16. (b) Gulf of Carpentaria curvature at resolution 1/16. Feasible points are not so evident.



Approximate signals compared at level -4.
Detail signal magnitudes compared at levels -3, -2, and -1.

Fig. 12.10. Registering Gulf segment against Australia boundary: two feasible paths, both ending at the correct registration offset of 251, are found.

TABLE 12.2. Orthogonal Wavelet Pyramid Algorithm Performance

Run	Candidate	Correct Offset	Match Measure	Pyramid Levels	Feasible Points	Success
1.0W	128 point segment of Gulf of Carpentaria	250	Detail signal magnitude	4	19	Yes
1.1W	Same as 1.0W, except shifted	251	Same	4	90	Yes
2.0W	Same as 1.0W, except uniform noise with mean-square SNR of 10.0 added to candidate signal	250	Same	4	87	Yes
2.1W	Same as 2.0W, except shifted	251	Same	4	111	Yes
2.2W	Same as 2.1W, except shifted	252	Same	4	79	Yes

Now, consider the Laplacian pyramid. It analyzes N -point signals into approximately $2N$ coefficients and images into $4N/3$ coefficients. Correlation between levels causes the larger pyramid sizes. Since the orthogonal wavelet and Laplacian pyramid representations are computationally quite alike [1], it is natural to study the registration algorithm using the Laplacian pyramid decomposition. For these experiments, we implemented the Laplacian pyramid using the approximately Gaussian [2] low-pass filter $\{0.05, 0.25, 0.4, 0.25, 0.05\}$. We find that raw difference signals of the pyramid (not their magnitudes) suffice for good registration. Figure 12.11 shows the result of the experiment of Figure 12.10 using the Laplacian pyramid decomposition.

Table 12.3 shows the results of the Laplacian-based registration algorithm on the same battery of experiments for which we used orthogonal wavelets.

To summarize the results, registering curvature maps with wavelet pyramids produces many more feasible points and paths. However, some Laplacian pyramid runs produce many feasible points too. Both methods are robust in the presence of noise, although the Laplacian pyramid suffers from very large feasible point counts at coarse resolutions. It turns out that candidate signals with small support often make the wavelet registration fail, whereas the Laplacian pyramid algorithm withstands these same small structures [5]. Overall, the Laplacian pyramid decomposition is somewhat better, since the correlation between levels stabilizes the coarse-to-fine tracking.

It appears that these results do not depend substantially on the type of MRA—and hence the discrete pyramid decomposition filters—chosen for deriving the coarse resolution signal structures. Both the Haar [16] and the Daubechies [17] compactly supported wavelets were used in similar registration experiments [18].

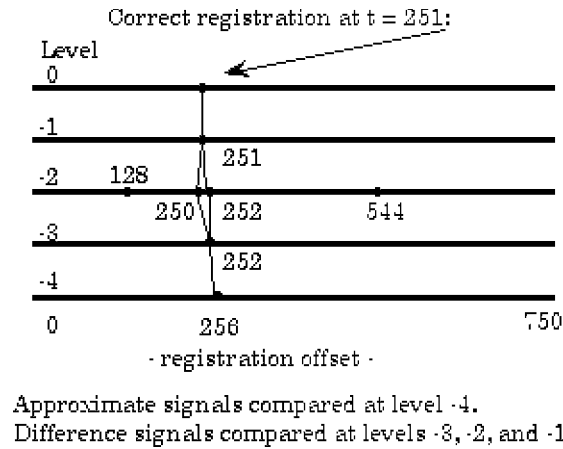


Fig. 12.11. Registering Gulf segment against Australia boundary using the Laplacian pyramid decomposition. Again, two feasible paths arise. The correct registration is found very quickly, at level -1 , where the unit distance of the difference signal is the same as the original curvature maps.

These researchers concluded that as long as two or more samples support the signal structure of interest, the approximate signal structures suffice for coarse-to-fine matching and registration. The same authors advise caution with the pyramid's detail signals. Neither method—Haar or Daubechies wavelet pyramids—significantly outperformed the other, although the Haar structures were slightly better when using detail signals [18].

TABLE 12.3. Laplacian Pyramid Algorithm Performance

Run	Candidate	Correct Offset	Match Measure	Pyramid Levels	Feasible Points	Success
1.0L	128-point segment of Gulf of Carpentaria	250	Detail signal	4	17	Yes
1.1L	Same as 1.0L, except shifted	251	Same	4	7	Yes
2.0L	Same as 1.0L, except uniform noise with SNR = 10.0 added to candidate	250	Same	4	954	Yes
2.1L	Same as 2.0L, except shifted	251	Same	4	1090	Yes
2.2L	Same as 2.1L, except shifted	252	Same	4	942	Yes

12.2 MIXED-DOMAIN SIGNAL PROCESSING

Mixed-domain signal transformations provide some new insights into signal processing tasks. Although this section concerns applications that refine rather than interpret a signal, we observe that these steps are often important ancillary feature of a signal analysis application. Here, we confine our remarks to three areas:

- Compression;
- Filter methods and filter banks;
- Enhancement.

Compression is necessary for constructing large signal databases, such as a model-based recognition system might employ. Filtering is important at the front end of an analysis application. Enhancement can be essential for building signal prototypes. In fact, although it is a purely signal-in, signal-out technology, compression is perhaps the most important commercial application of the orthogonal wavelet transform. We hasten to add that a very large research literature continues to flourish in all of these areas; our present assessment by no means constitutes the final word.

Good compression methods currently employ either the discrete cosine transform (DCT) or orthogonal wavelet transform coding. The idea is that the transform coefficients are statistically far simpler than the original signals or images, and therefore the transformed data can be described with fewer numerical values. Of course, the orthogonality of the wavelet transform and the efficient, perfect reconstruction filter banks that it provides promote its use in compression. After compression, a handful of transform coefficients nonlinearly encode complex signal and image patterns. Perfect reconstruction is, in principle, possible. Even with lossy compression, ratios of one-bit compressed versus one-byte (8-bit) original signal gives excellent reconstruction. When more decomposition coefficients are discarded in compressing the signal and there remains only a single bit versus 4 bytes of source signal, the reconstruction is still fairly good for human perception.

Digital sound, image, and video databases are huge. Whether they support a signal analysis system or not, compression is essential for economy of storage, retrieval, and transmission. The earliest wavelet compression methods were based on the orthogonal wavelet pyramid decomposition (12.2), shown in Figure 12.5 [1]. New compression methods—some of which provide compression ratios of two orders of magnitude—based on wavelet transforms have been reported in the research literature [19–22]. The basic idea is that many detail coefficients carry no useful signal information and can be set to zero without appreciably affecting the result of the pyramid reconstruction algorithm. There are many variations, but typically the steps are as follows:

- (i) Select a multiresolution analysis and a final level of pyramid decomposition $-L$, where $L > 0$.
- (ii) Decompose the signal $x(n)$ into its pyramid decomposition (12.2), producing detail coefficients for levels $-L \leq l \leq -1$ and approximate coefficients for level $l = -L$.

- (iii) Apply a threshold to the fine structure signals $D_{-L}^d x, D_{-L+1}^d x, \dots, D_{-1}^d x$, so that small magnitude coefficients are set to zero. This is typically a *hard threshold*: If $s(t)$ is a signal, then its hard threshold by $T > 0$ is given by

$$s_{T,h}(t) = \begin{cases} s(t) & \text{if } |s(t)| > T, \\ 0 & \text{if } |s(t)| \leq T. \end{cases} \quad (12.24)$$

- (iv) Apply a standard compression technique to the small coarse-resolution trend signal $A_{-L}^d x$. Examples include the Karhunen–Loeve compression [23] or—for images, especially—the Joint Photographic Experts Group (JPEG) standard algorithm [24], which derives from the discrete cosine transform (DCT) [25].
- (v) Apply an entropy coding technique, such as simple run-length encoding [26] to the detail signals.
- (vi) Decode the compressed pyramid levels from (iv) and (v) and reconstruct the original signal (with some loss, principally due to the thresholding operations) using the exact reconstruction afforded by the MRA's quadrature mirror filters;
- (vii) A conservative guideline for hard threshold selection is

$$T = \sigma \sqrt{2 \frac{\log(N)}{N}}, \quad (12.25)$$

where σ^2 is the variance of $x(n)$ at level $l = 0$ and N is the number of samples.

Compression ratios of about 25:1 on natural images are possible with the above method. The extension of the orthogonal wavelet pyramid decomposition to two dimensions (images) is necessary for this technique [1], but is unfortunately beyond our present scope. However, Figure 12.12 gives the idea. As with signals, the detail

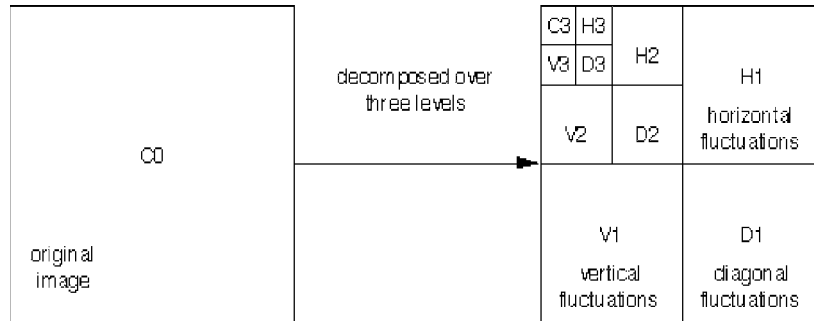


Fig. 12.12. Orthogonal wavelet image compression decomposes the original into four uncorrelated subimages. The trend image C_1 is analogous to the one-dimensional approximate signal. The three detail images contain the direction-sensitive fluctuations of the image C_0 . These include vertical details, horizontal details, and diagonal (corner) details. The image pyramid decomposition applies the same algorithm to each coarse structure trend image, C_1 , C_2 , and so on. This produces more detail images and a final trend image, C_3 above.

images tend to have simple statistics. They can be modeled, quantized, and individually compressed to impressive ratios. The decomposition allows the algorithm designer to tune the vertical, horizontal, and diagonal quantizations so that the direction-sensitive human visual system perceives minimum distortion in the reconstructed image [1, 27].

More recent approaches to signal compression are as follows:

- *Malvar “wavelets”* are in fact a time-frequency signal decomposition [28]. They are an alternative to the time-scale decomposition using wavelet pyramids. This time-frequency compression technique breaks down an image into a sequence of sinusoidally textured atoms, with adjustable leading and trailing borders. The overall size of the atoms is also tunable. Lastly, as we noted briefly in Chapter 10, as an atomic signal decomposition that uses sinusoids instead of complex exponentials may avoid the limitation of the Balian–Low theorem. This flexibility allows Malvar wavelet compression schemes to beat others when tested in constrained problem domains, such as fingerprint images [29].
- *Wavelet packets* are functions of the form $2^{m/2}W_n(2^m t - k)$, where m , n , and k are integers and $n > 0$ [30,31]. W_n extends only over a finite interval $[0, N]$, and it contains the root frequency of its family of atoms. The decomposition scheme is similar to the orthogonal wavelet pyramid, except that the detail signal structures are also composed at every level. Image decomposition uses tensor products of the translations and dilations of the W_n . A distinct advantage of wavelet packets over Malvar wavelets is that each set of atoms is generated by translation, dilation, and modulation of a single function. This simplifies the construction of algorithms and special-purpose compression hardware. Wavelet packets offer excellent compression ratios, in the realm of 100:1 [19].
- *Structural approaches* to wavelet-based image compression take a two-stage approach. This scheme first extracts edges across several scales. It then encodes the texture representing the difference between the original and the reconstruction from edge information. Combining the texture-coded error image with the edge-coded image gives a perceptually acceptable rendition of the original [20]. Closely related to transform signal compression are mixed-domain processing techniques for noise removal and enhancement.

12.2.1 Filtering Methods

Wavelet decompositions provide for methods that remove background noise from a signal but preserve its sharp edges. This can be especially valuable in signal analysis applications where it is necessary to identify local shapes, for example, that may be corrupted by noise, but still obtain a precise registration [29]. We recall from Chapter 9 that low-pass and bandpass filters removed high-frequency components from signals, but as a rule, these convolutional systems blur the edges as well. Once again, there are many alternatives for wavelet-based noise removal. The typical approach follows the compression algorithm, with a twist at the thresholding step [32–34]:

- (i) Select an MRA and final level of decomposition $-L$, where $L > 0$.
- (ii) Decompose $x(n)$ according to (12.2).
- (iii) Retain the coarse structure approximate coefficients at $l = -L$, but apply a *soft threshold* to $D_{-L}^d x, D_{-L+1}^d x, \dots, D_{-1}^d x$: If $s(t)$ is a signal, then its *soft threshold* by $T > 0$ is given by

$$s_{T,s}(t) = \begin{cases} \operatorname{sgn}[s(t)](|s(t)| - T) & \text{if } |s(t)| > T, \\ 0 & \text{if } |s(t)| \leq T. \end{cases} \quad (12.26)$$

Soft thresholding results in a continuous signal (Figure 12.13).

- (iv) Reconstruct the original signal.
- (v) Soft threshold selection is either heuristic, based on the hard threshold selection (12.25), or extracted via the Stein unbiased risk estimate (SURE) [33, 35].

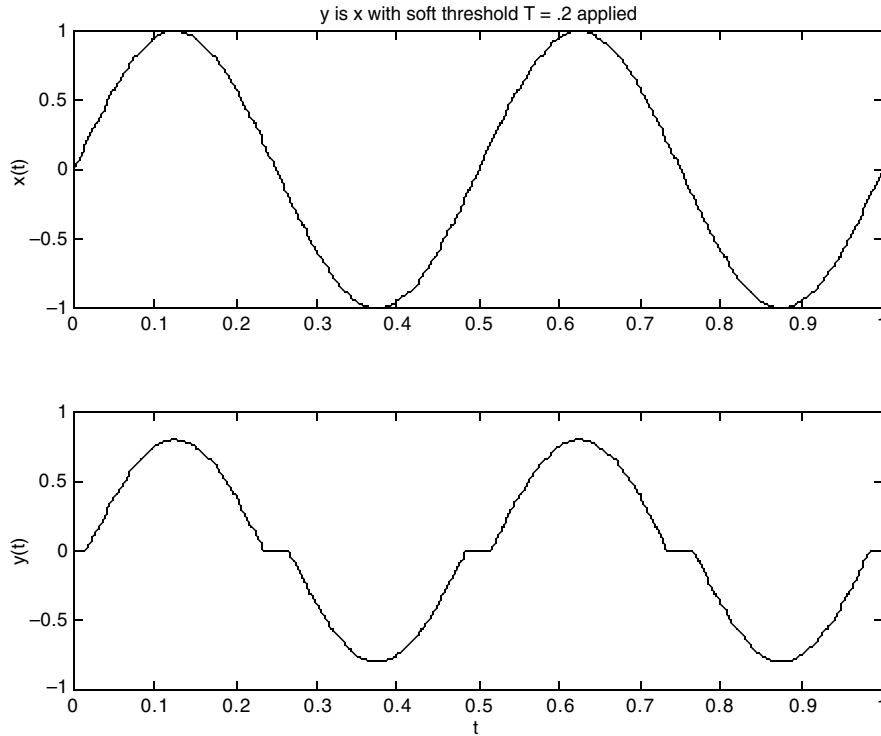


Fig. 12.13. Soft thresholding a sinusoidal signal.

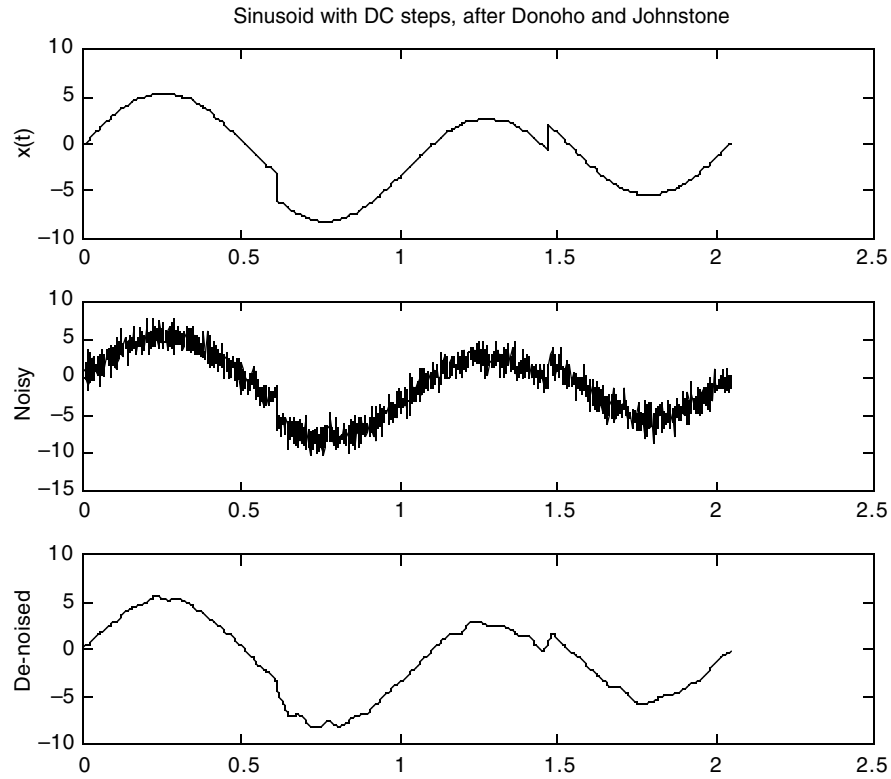


Fig. 12.14. Noise removal filtering using wavelets, soft thresholding, and the SURE threshold selection.

Figure 12.14 provides an example.

12.2.2 Enhancement Techniques

Let us consider a third processing technique using wavelet transforms. Many signal acquisition and imaging systems use photon detectors, such as the popular charge-coupled device (CCD) or the sensitive photomultiplier tube (PMT), as their data source. Examples are spectrometers used to control plasma etchers in the semiconductor manufacturing industry, an application we considered in Chapter 4; astronomical instruments; remote sensing devices; and photo-optical sensors in general.

CCDs now have become the digital image acquisition device of choice [36]. A *charge well* corresponds to a single pixel. They are small, holding perhaps 800 electrons per micron² (μ , 10^{-6}m). This is called the *well capacity*, W_c . Factors affecting CCD sensor performance include the following:

- Wells have finite capacity, so if exposure time is too long, electrons spill over to adjacent wells, causing *blooming* of the image (bright blurry spots).
- There are three noise sources. These are due to thermal effects, N_θ , (also called *dark current*); the quantum nature of light, N_ϕ ; and logic noise during readout, N_p .
- Total noise within the image at a pixel is therefore $N = N_\phi + N_\theta + N_p$.
- *Dynamic range* of the well is defined as (capacity)/(readout noise level) = W_c/N_p .
- Thermal effects may be mitigated by cooling the sensor; typically, 6 degrees C warmer means twice as much thermal noise; in other words,

$$N_\theta(t) = \int_{t_0}^t K_\theta 2^{\frac{T}{6}} dt, \quad (12.27)$$

where T is the temperature in degrees Celsius, K_θ is a constant, and t_0 is the dark current.

- Readout noise rate increases with readout frequency.
- Light flux striking sensor obeys a Poisson distribution, where α is the parameter of the distribution:

$$p_I(k) = e^{-\alpha} \frac{\alpha^k}{k!} = \text{Prob}(I = k). \quad (12.28)$$

- The mean of a Poisson distributed random variable is $\mu = \alpha$ and its standard deviation is $\sigma = \alpha^{1/2}$ in (12.28).

Of the diverse image noise sources, the most troublesome is *quantum noise*, N_ϕ , which arises from the discrete nature of light quanta detection. Its magnitude changes with the light intensity and is thus image-dependent. Toward eliminating this pernicious source of noise in photon imaging systems—so-called *photon noise*—Nowak and Baraniuk [37] have applied an adaptive filtering method, based on the wavelet pyramid decomposition.

We have already observed that in some applications, wavelet pyramids furnish an especially convenient and powerful tool for suppressing noise in signals and images. The decomposition of a signal into a pyramid allows us to design algorithms that eliminate apparent noise in certain frequency bands by simply attenuating (or even zeroing) the coefficients in the suspect pyramid level. When the signal is reconstructed from the modified pyramid, troublesome noise is then absent. Better compression ratios are also obtained, without significant loss in perceptual signal quality.

In Ref. 37, the idea is to adjust pyramid level intensities according to the energy in other parts of the representation. For example, we might decompose to two coarse levels of representation, then attenuate the detail coefficients selectively, where the approximate signals have large magnitudes. Detail signals corresponding to regions where the approximate signal is weak remain the same or are amplified. Then, we reconstruct the profile using the pyramid scheme. A number of the computational experiments in Ref. 37 markedly improve faint, low-contrast medical

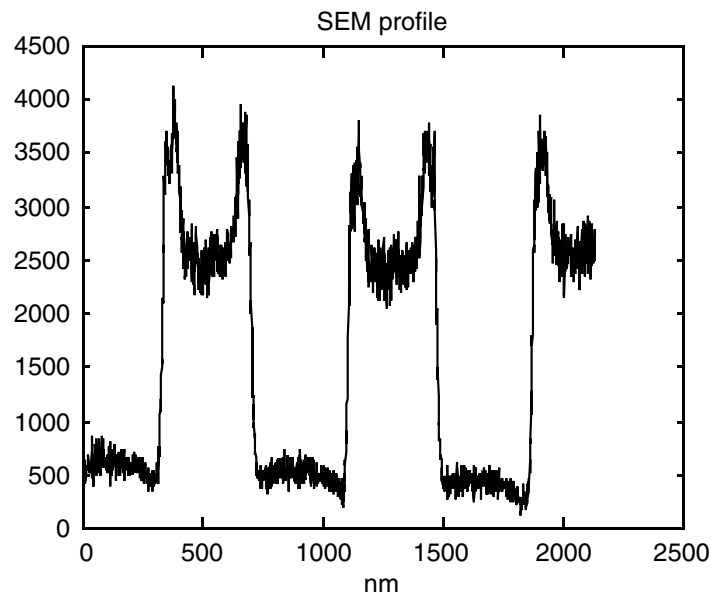


Fig. 12.15. Example of scanning electron microscope profile of lines on a test wafer. Noise magnitudes are roughly equal on tops of the high-magnitude lines and at the bottoms of the low-magnitude spaces between them.

and astronomical images. There is also a promise of better histogramming, edge detection, and higher-level image interpretation results. Figure 12.15 and Figure 12.16 show an application of this technique to raster lines from a scanning electron microscope (SEM) image.

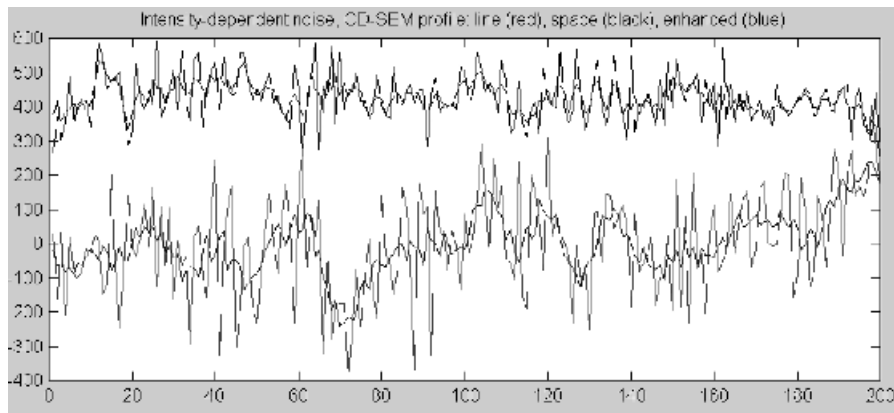


Fig. 12.16. After enhancement using the method of Nowak and Baraniuk. Note that noise on the wafer lines (*lower trace*) is reduced, whereas the details on the low magnitude spaces (*upper trace*) tends to be preserved.

12.3 BIOPHYSICAL APPLICATIONS

There was a surge of interest in these time-frequency and time-scale transforms in the mid-1980s. Psychophysicists noticed that Gabor elementary signals (Chapter 10) could model some aspects of the brain's visual processing. In particular, the *receptive fields* of adjacent neurons in the visual cortex seem to have profiles that resemble the real and imaginary parts of the Gabor elementary function. A controversy ensued, and researchers—electrical engineers, computer scientists, physiologists, and psychologists—armed with the techniques of mixed-domain signal decomposition continue to investigate and debate the mechanisms of animal visual perception [38, 39].

12.3.1 David Marr's Program

Signal concavity remains an important concept in analysis applications. Years ago, the psychologist Attneave [40] noted that a scattering of simple curves suffices to convey the idea of a complex shape, for instance, a cat. Later, computer vision researchers developed the idea of assemblages of simple, oriented edges into complete theories of low-level image understanding [41, 42]. Perhaps the most influential among them was Marr, who conjectured that understanding a scene depends upon the extraction of edge information over a range of visual resolutions [7]. Marr challenged computer vision researchers to find processing and analysis paradigms within biological vision and apply them to machine vision. Researchers investigated concavity and convexity descriptions as well as finding their boundaries at multiple scales. Thus, we might resolve an image into an intricately patterned structure at a fine scale, but coarse representation reveals just a butterfly wing. Marr speculated, but could not prove, that multiscale edges could uniquely describe signals and images. This would imply that the ultimate structural description of a signal would consist of its edge maps across all scales.

Two important early outcomes from Marr's program were scale space theory [8, 9, 43, 44] and optimal multiscale edge detectors [45–48] (Chapter 4). These theoretical results and the practical success of edge-based analysis and description of signal structure, bolstered Marr's conjecture. But wavelets weighed in as well. Mallat tried to use wavelet transform *zerocrossings* [49] as a multiscale structural signal characterization, and showed how the technique could be used for stereometry, but the method suffered from instability. Mallat and Zhong [50] changed strategies and showed that finding wavelet transform *maxima* across scales was equivalent to the Canny edge detector [45]. Then both Berman and Baras [51] and Meyer [29] found counterexamples to Marr's conjecture. In fact, Meyer's example gives a wavelet and a collection of sinusoidal sums that have the same zero crossings when convolved with the wavelet.

12.3.2 Psychophysics

Among the research efforts Marr's work inspired are comparisons between biological and computer vision. Such comparisons tend to support the notion that the

particular computer vision innovation being considered is more than an ad-hoc technological trick. When the biological analogies are clear, in fact, it is plausible that the technology is taking advantage of some fundamental physical properties of objects in the world and their possible understanding from irradiance patterns.

While most physiological studies of the visual brain have concentrated on cats and monkeys as experimental subjects [52], the evidence shows many similarities between the function of the cat's visual cortex and that of the monkey. It is therefore reasonable to assume—and the psychophysical studies done on human subjects support this—that the human visual system implements these same principles of cortical organization. Visual information arrives through the retina, and then passes down the optic nerve to the lateral geniculate nucleus (LGN), from which it is relayed to the visual part of the brain, variously known as V1, *area 17*, the *striate cortex*, or the *primary visual cortex*. Some two or three dozen separate visual areas of the brain have been identified according to their visual function [52].

Light impinging on the retina directly stimulates V1 neurons, as well as some other cortical areas. The area of the retina upon which a pattern of irradiance may stimulate a neuron is called the neuron's *receptive field* (RF). In their pioneering work, Hubel and Wiesel [53] differentiated between *simple* cortical neurons and *complex* cortical neurons. A simple cell tends to assume one of two states, “on” or “off,” according to whether special light patterns were directed within its RF. Complex cells, on the other hand, do not exhibit this binary behavior, are prone to have larger RFs than the simple cells, and can be stimulated over a much broader RF area [53]. Studying the cat's visual cortex the researchers further demonstrated that both the simple and complex cells within area V1 have a very high orientation selectivity. Monitoring the responses of cells while slits of light at various angles were flashed onto the RFs of the neurons demonstrated this. Interestingly, such orientation specificity is not shown in the retinal area of the eye or in the LGN, but only appears when the visual information finally reaches the cortex [53]. Campbell and Robson [54] confirmed this property for human subjects through a series of psychophysical experiments.

Campbell and Kulikowski [55] and Blakemore and Campbell [56] described another property showing independent vision channels to exist. These channels have an orientation selectivity in addition to the spatial frequency selectivity. The orientation selectivity exhibited by the independent channels is not well accounted for either in Marr's system [7] or in the Laplacian pyramid algorithms [2]. However, orientation selectivity as well as spatial frequency tuning is a feature of the channels in the two-dimensional wavelet multiresolution representation [1] as shown in Figure 12.12.

Originally, researchers in the animal vision physiology thought that cortical cells were *feature detectors*, activated by the presence of a dot, bar, edge, or corner. Orban [52] emphasizes that it is now clear that cortical cells are actually *filters* and not feature detectors. It is also now possible to identify some visual cortical areas with the animal's behavior. For cats, nearly all neurons of cortical areas 17 (V1), 18, and 19 are orientation-sensitive bandpass filters covering all orientations [52]. Cells in area 17 have the smallest orientation bandwidth and show strong preference for

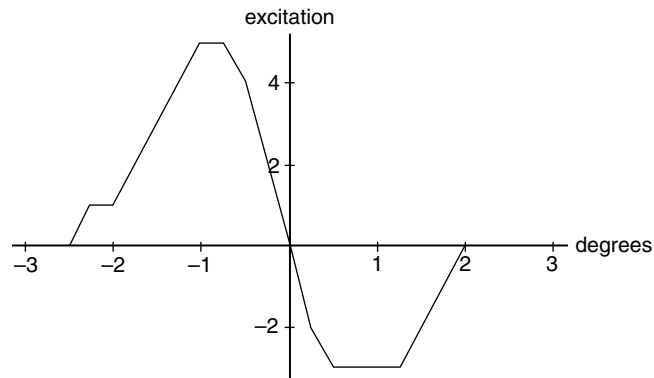


Fig. 12.17. RF profile of cortical simple cell #1316 of Ref. 57, odd symmetry.

horizontal and vertical directions. They are therefore most useful for observing stationary objects. The situation in area 19 is less clear. These cells have large receptive fields, broad orientation tuning, and little motion sensitivity. Area 18 cells have very large RFs, are sensitive only to low spatial frequencies, maintain a high orientation bandwidth, and have some velocity sensitivity. These cells work together to provide the animal with motion analysis [52].

To obtain an RF profile for a cortical cell, Jones and Palmer [57, 58] and Jones, Stepnoski, and Palmer [59] plotted neuron firing rate—the cell *activation level*—against the position of a spot of light within a small 16×16 grid. According to the widely held view that the simple cells are filters of varying orientation and spatial frequency sensitivity, it should be possible to model the impulse response of the simple cell filter by mapping firing rate versus stimulus position in the RF. In fact, these researchers were able to obtain plots of the spatial and spectral structure of simple RFs in cats. Figure 12.17 and Figure 12.18 are based on cross sections of contour plots of typical cat simple receptive fields provided in Ref. 57.

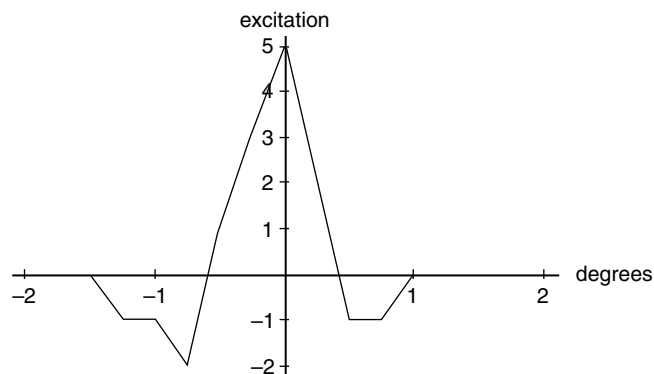


Fig. 12.18. RF profile of cortical simple cell #0219 of Ref. 57, even symmetry.

Many cortical simple cells have either an odd or even symmetry. Cells with odd symmetry, such as in Figure 12.17, have areas of excitation and inhibition on opposite sides of the RF center. On the other hand, those cells with even symmetry, such as in Figure 12.18, have a central excitatory (or inhibitory) region that is surrounded by an inhibitory (or excitatory) areas. The two basic types are illustrated in the above figures. In Figure 12.17 and Figure 12.18 the regions of positive excitation represent RF areas where responses to *bright stimuli* were obtained. The regions of negative excitation represent RF positions in which the response was found to correlate with *dark stimuli*. All of their excitation frequency measurements were made by microelectrodes inserted into cortical area 17.

Not long after the profiles of cortical neuron receptive fields became more widely understood, Marcelja [60] showed that the RF profiles were strikingly similar to the graphs of Gabor elementary functions (GEF). Marcelja's paper focused the attention of the computer and biological vision research communities onto the potential applications of these functions in vision research. We recall from Chapter 10 that a GEF is a sinusoid multiplied by a Gaussian. Gabor [61] showed that these functions are optimally localized in the time and frequency domains. The product of their spatial extent and bandwidth is minimal. Further, the functions form a complete mathematical set [62] for expansions of other signals.

When we choose the parameters appropriately and graph the GEF's real and imaginary parts separately, the functions resemble the RF profiles of cortical neurons. More remarkably, Pollen and Ronner [63] discovered that adjacent simple cells are often tuned to similar spatial frequencies, and have similar orientation selectivities, but appear to have phase difference of 90 degrees. That is, it appears that the cosine (even) and sine (odd) components of the GEFs are implemented in the visual cortex by pairs of adjacent simple cells [63]. Not surprisingly, these discoveries in physiological and psychophysical research aroused intense new interest in the study and application of the Gabor functions for computational vision.

Daugman [64] extended Gabor's results [61] to the case of two-dimensions, showing that the resulting two-dimensional elliptical Gaussians were optimally localized in the spatial and spatial frequency domains. Daugman suggested that filters based on the elliptical Gaussians modulated by sinusoids represent a necessary evolutionary compromise for a biological vision system with incompatible high-level goals. The organism must find both spatial information and spatial frequency information from its visual space. The way to accomplish this is to implement in the visual cortex the transform that has the best joint specificity in the spatial and frequency domains. This transform is the Gabor transform [65]. Further, the correspondence remains between the shapes of these two-dimensional Gabor functions and the cortical RFs considered as two-dimensional filters. Thus, by basing themselves on the GEFs as models, vision physiologists are evidently equipped with a formalism that explains the properties of orientation selectivity, spatial frequency selectivity, and the empirically observed quadrature relationship for pairs of cortical simple cells.

Does the visual cortex implement some kind of Gabor transform? The physiological experiments on animals and the psychophysical experiments on humans

seem to overwhelmingly support the view that the simple RFs are filters selective to orientation and frequency. Further, the spatial structure of the simple RFs closely resembles the GEFs. It is no wonder that the Gabor representation, with its optimal localization properties, was seized upon as a candidate model for the functioning of the visual cortex. Mallat questions the choice of the Gabor transform as a model for cortical functions, however [1]. He points out that the simple cells of the visual cortex do not have impulse responses which contain more cycles when the tuning is for a higher frequency. This would be the case if the cells were organized in the form of the logons of the Gabor representation. Instead, Mallat argues, the simple cells have larger RFs when the frequency tuning is lower and smaller RFs when the frequency tuning is correspondingly higher.

The experimental evidence is not completely clear on this issue however. The question at hand is whether simple cells with high-frequency tuning exhibit more cycles within their fields than cells selective of lower spatial frequencies. This is equivalent to saying that the bandwidth in octaves varies with the particular spatial frequency preferred by a cortical simple cell. Pollen and Ronner stress that a variety of bandwidths are typically associated with cells of a given frequency tuning [63].

Nevertheless, the correlations between preferred spatial frequency and bandwidth tend not to support the contention that a full Gabor transform, with its specific requirement of increased cycles in RFs with higher frequency tunings, is implemented in the visual cortex. A model counterposed to the Gabor model of cortical architecture, wherein the RF sizes vary inversely with the frequency tuning of the simple cells, is presented by Kulikowski, Marcelja, and Bishop [39].

Finally, some more recent studies of the visual cortex support the viewpoint that the receptive field sizes vary inversely with preferred spatial frequency. This would be the case if the visual cortex implements a kind of time-scale transform. A case in point is Anderson and Burr's investigations of human motion detection neurons in the visual cortex [66]. The authors discover a regular decrease in RF size as the observed preferred frequency tuning of these cells increases. The RF size was found to progressively diminish from as high as 7 degrees at low spatial frequencies to 2 minutes of arc for cells with the highest preferred frequencies.

As Gabor—and perhaps wavelet—transforms have proven useful in modeling aspects of the human visual system, so have they found applications in studying the auditory system. One can think of the cochlea as a bandpass filter bank (Chapter 9). From measurements of sound sensitivity above 800 Hz, it then turns out that the filter impulse responses are approximately *dilations* of one another [67]. Thus, cochlear sound processing roughly implements a wavelet transform.

12.4 DISCOVERING SIGNAL STRUCTURE

Time-frequency and time-scale transforms provide alternative tools for local signal description. The local descriptions can be merged and split, according to application design, resulting in a structural description of a signal. The motivation for this is that the physical processes that produced the signal changed over time and that the

structure of the signal, properly extracted, provides a means for identifying and understanding the mechanism that generated it. Structures may be simple time-ordered chains of descriptors. When there are long-term and short-term variations in the signal's generation, then it may be effective to build a hierarchical graph structure for describing the data.

12.4.1 Edge Detection

Let us examine how well the windowed Fourier and wavelet transforms can support basic edge detection signal analysis tasks. Signal edges represent abrupt changes in signal intensity and are a typical initial step to segmenting the signal.

12.4.1.1 Time-Frequency Strategies. A simple sawtooth edge experiment demonstrates that the windowed Fourier transform is a problematic edge detector.

The Gabor transform responses indicate local frequencies, and there are indeed high frequency components in the neighborhood of signal edges (Figure 12.19). Locality is poor, however. Shrinking the time width of the Gabor elementary functions provides better resolution. The problem is that this essentially destroys the nature of the transform. Using different window widths makes it more resemble the wavelet transform. Perhaps the most effective application for the short-time Fourier transforms is to indirectly detect edges by locating regions of distinct texture. Indeed, the windowed Fourier transforms are very effective for this purpose, and a number of research efforts have successfully applied them for texture segmentation [68–70]. The filter banks and time-frequency maps of Chapter 9 provide starting points for the spectral analysis of signal texture. The edges between differently textured regions are inferred as part of the higher-level interpretation steps.

12.4.1.2 Time-Scale Strategies. Better suited to edge detection are the time-scale transforms. Not only does the wavelet transform provide for narrowing the time-domain support of the analyzing wavelet, allowing it to zoom in on signal discontinuities, there are two theoretical results that support wavelet-based edge detection:

- (i) For certain continuous wavelet transforms, finding maximal response is identical to applying the optimal Canny edge detector [45].
- (ii) The decay of the wavelet transform maxima across scales determines the local regularity of the analyzed signal.

However, we also now know that edge-based descriptions of signal structure are not the final answer:

- (iii) Marr's conjecture is false [29, 51].

Yet, structural description by edges and extrema (ridge edges) remains a powerful tool for understanding signals. Both continuous and discrete wavelet transforms are closely related. For example, a wavelet transform can be built around spline wavelets

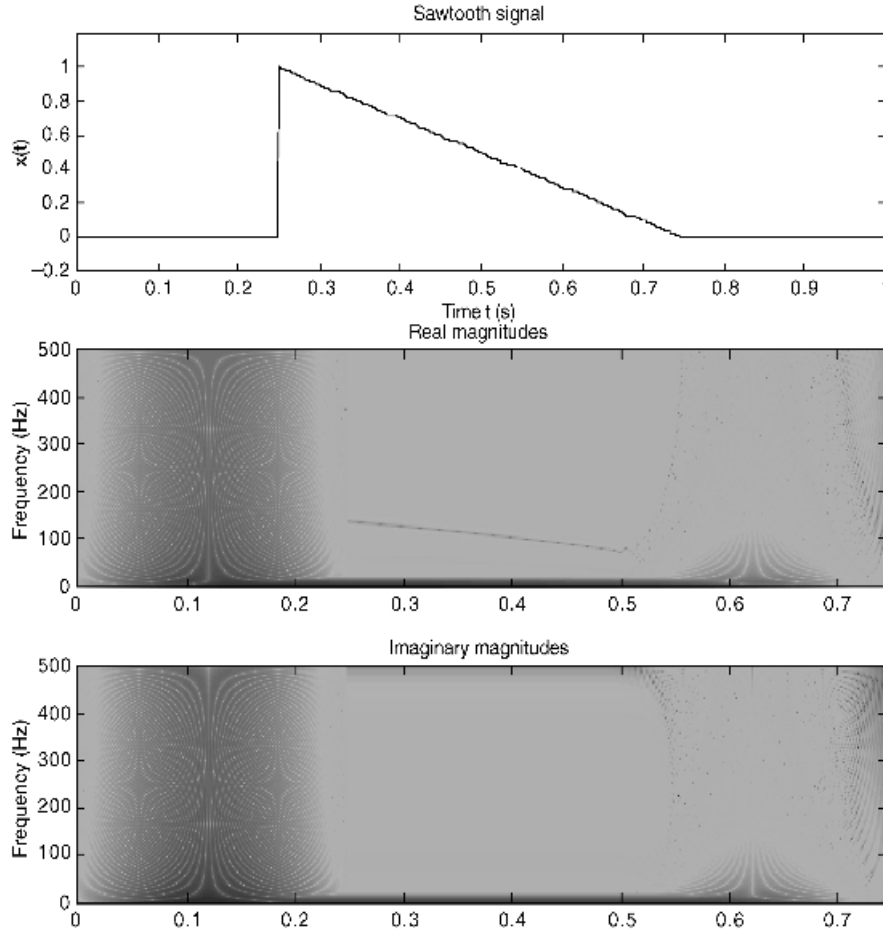


Fig. 12.19. A sawtooth edge (*top*); real and imaginary Gabor transform coefficient magnitudes (*middle*); and convolution of the discrete high-pass filter associated with the cubic spline orthogonal wavelet representation.

that mimic the optimal derivative of Gaussian [48]. The multiresolution decomposition of the signal supports a multiscale edge detector [50]. This is useful for discriminating background noise from substantive signal features according to perceptual criteria [71]. The discrete high-pass filters $g_p(n)$ given in Figure 12.3 and Figure 12.4 function as edge detectors for the orthogonal wavelet pyramid (Figure 12.5).

Let us turn to the continuous wavelet transform. Suppose $g(t) = Ae^{-Bt^2}$ is a Gaussian of zero mean. Set $\psi(t) = \frac{d}{dt}g(t)$. Then $\psi(t)$ is a wavelet because it has zero mean, is integrable, and has finite energy. Let $x_a(t) = ax(at)$ be the scaled dilation of $x(t)$ by factor a . Typically, $a = 2^i$ for $i \in \mathbb{Z}$, with $i < 0$ in our notation [1].

Then, changing notations from Chapter 11 slightly, the wavelet transform of an analog signal x is

$$(W_{-i}x)(t) = (\psi_{2^{-i}} * x)(t) \quad (12.29)$$

at time instant t . Consequently,

$$(W_{-i}x)(t) = (\psi_{2^{-i}} * x)(t) = \left(2^{-i} \frac{d}{ds} g(2^{-i}s) * x\right)(t) = 2^{-i} \frac{d}{dt} (g_{2^{-i}} * x)(t). \quad (12.30)$$

So where $|W_{-i}x|$ is large, the version of x , smoothed to resolution 2^{-i} , is changing rapidly [72].

Let us recount the result (ii) above [72]. Suppose $x(t)$ is a signal defined in a neighborhood of t_0 , $0 \leq \alpha \leq 1$, and there is a constant c such that for all t in an interval about t_0 , we have

$$|x(t) - x(t_0)| \leq c|t - t_0|^\alpha. \quad (12.31)$$

Then $x(t)$ is *Lipschitz*¹ α at $t = t_0$. The *Lipschitz regularity* of $x(t)$ at $t = t_0$ is the least upper bound of all α such that (12.31) holds. If there is an $0 \leq \alpha \leq 1$ and an open interval such that (12.31) holds for all $t \in (a, b)$, then the signal $x(t)$ is *uniformly Lipschitz* α on (a, b) . In other words, $x(t)$ is uniformly Lipschitz if it is as tame as an exponential function in some region.

Now suppose we have a continuous wavelet $\psi(t)$ that decays at infinity as $1/(1+t^2)$ and a square-integrable signal $x(t)$. Then it can be shown [73] that $x(t)$ is uniformly Lipschitz α on (a, b) if and only if there is a $c > 0$ such that for all $t \in (a, b)$

$$|(W_{-i}x)(t)| \leq c2^{i\alpha}. \quad (12.32)$$

The decay of wavelet transform maxima over many resolutions is essentially a study of the degree of singularity of the original signal. An extensive study of continuous wavelet transformation as a characterization of regularity is Ref. 74.

12.4.1.3 Application: The Electrocardiogram. Biomedical technology has investigated almost every avenue of signal analysis in order to improve electrocardiogram (ECG) interpretation. Researchers have experimented with time-domain, frequency domain, time-frequency domain, and now time-scale domain methods [75, 76]. Chapter 1 introduced ECG signal processing and analysis.

¹Analyst Rudolf Lipschitz (1832–1903) was professor at the University of Bonn.

Frequency-domain methods are effective for many important tasks in computerized electrocardiography, such as convolutional noise removal and band rejection of noise from (50 or 60 Hz, for example) alternating current power sources [77]. Edges and transients in the ECG are crucial to interpreting abnormalities. In order to preserve these features, yet remove noise, research has turned to mixed-domain filtering techniques, such as we covered in Section 12.2 [78]. Compression techniques using wavelet transforms are also known [79]. The most important task in automated ECG analysis is QRS complex detection [80], essentially a ridge edge detection problem. The foundation of these application is the characterization of signal regularity by wavelet transform maxima across scales [74]. Algorithms for QRS detection and time-scale decomposition of ECGs using the orthogonal wavelet decomposition are shown in Ref. 81. The continuous wavelet transform is studied for QRS characterization in Ref. 82. The wavelet transform is effective in revealing abnormalities, such as the ventricular late potential (VLP) [83]. For example, in Ref. 84 a synthetic VLP is introduced into the ECG. The late potential is difficult to discern in the time-domain trace. However, wavelet transformation reveals that the defect is as an enlargement in the time-domain support of the QRS complex at certain scales.

12.4.2 Local Frequency Detection

Both the short-time Fourier and wavelet transforms perform local frequency detection. The STFT or Gabor transform relies on time-frequency cells of fixed size (Chapter 10). The wavelet transform adapts the time domain extent according to the frequency tuning (Chapter 11).

12.4.2.1 Mixed-Domain Strategies. The fixed window width of the short-time Fourier transform is useful when the range of frequencies in the analyzed signal is known to remain within fixed bounds (Figure 12.20). An example of this is in texture analysis, where the local frequencies of the signal pattern are expected within given spectral ranges. Small defects in the texture are not readily detected, but the time-frequency map displays the overall local pattern quite well.

On the other hand, the wavelet pyramid decomposition tends to mimic the coarse structure of the signal in the approximate coefficients and provides a range of highpass filters sensitive to local textures in the detail coefficients (Figure 12.21).

Finally, the continuous wavelet transform clearly shows the scale of the underlying pattern features in its amplitude (Figure 12.22).

12.4.2.2 Application: Echo Cancellation. One application of wavelet transform-based filter banks has been to improve echo canceller performance. Chapter 2 (Section 2.4.4) explained the need for echo cancellation in digital telephony. The echo arises from an impedance mismatch in the four-wire to two-wire hybrid transformer. This causes an echo, audible to the far-end listener, to pass into the speech signal from the near-end speaker. The classical time-domain approach for reducing the echo is to remove the echo by an adaptive convolutional filter [85]. One problem is getting the canceller to converge quickly to an accurate echo model when the

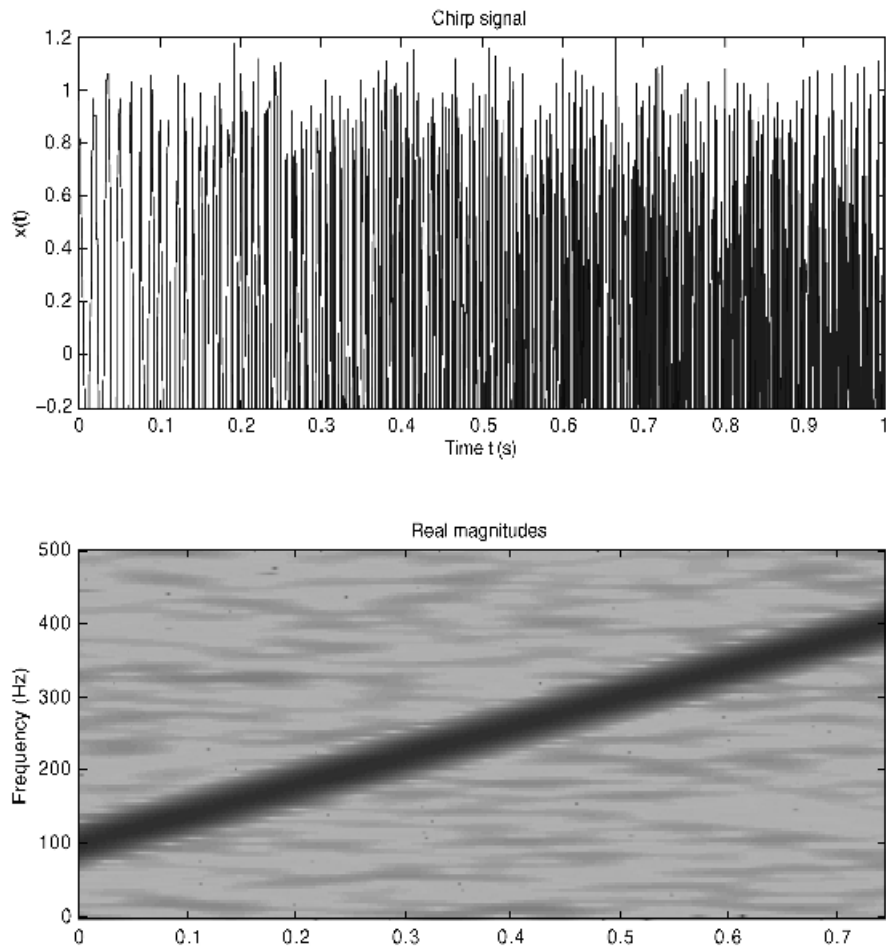


Fig. 12.20. A Gabor transform applied to a sinusoidal chirp signal embedded in noise.

echo path length varies. This can happen in digital telephony, and it is a problem in other applications, such as acoustic echo cancellers employed in teleconferencing systems. Both conventional quadrature mirror filter bank decompositions [86] and wavelet packet decompositions [87] have been used to replace the adaptive time-domain convolution in the classical echo canceller.

12.4.2.3 Application: Seismic Signal Interpretation. The continuous wavelet transform arose out of problematic attempts to use time-frequency methods in seismic data analysis [88]. Both the continuous wavelet transform and the discrete orthogonal pyramid decomposition are applicable to seismic signal interpretation. Early applications used the wavelet transform to improve visualization and interpretation of seismic sections [88–90].

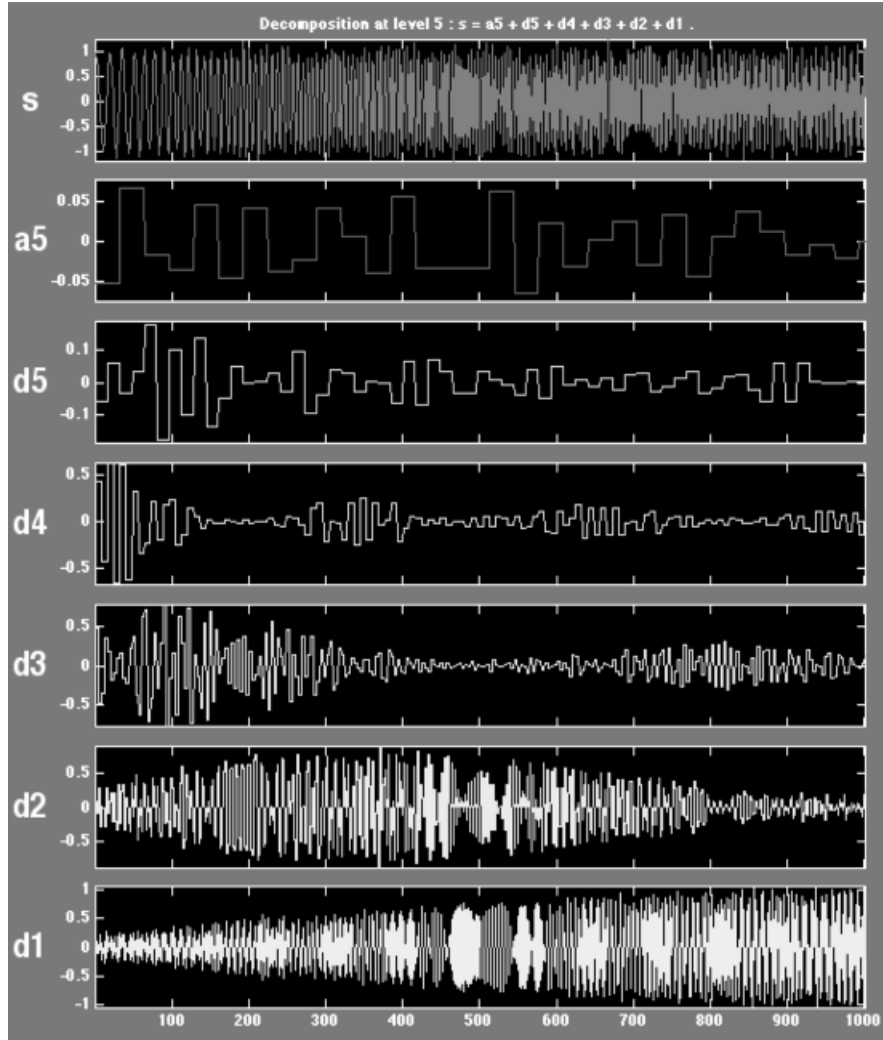


Fig. 12.21. An orthogonal wavelet transform applied to a sinusoidal chirp signal embedded in noise. Decomposition to five levels using the Haar wavelets.

Both the continuous and orthogonal wavelet transforms have been applied to seismic signal analysis. In Ref. 91, for example, researchers recommend the Morlet wavelet (12.33) for removing correlated ground roll noise from seismic exploration data sets.

$$\psi(t) = e^{-\frac{t^2}{2}} \cos(\omega_0 t). \quad (12.33)$$

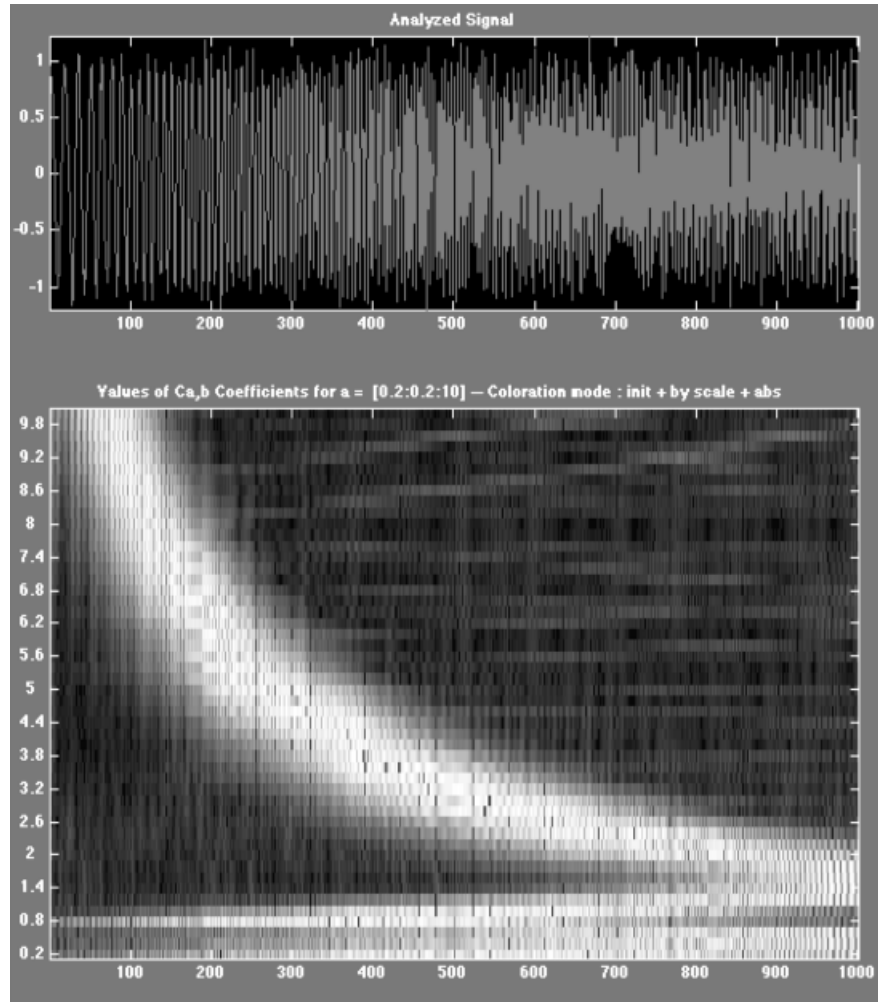


Fig. 12.22. A continuous wavelet transform applied to a sinusoidal chirp signal embedded in noise. Decomposition over 50 scale values using the Morlet wavelet.

The authors perform the continuous transformation by convolving the wavelet $\psi(t)$ with the raw data, exclude the bands containing the ground roll noise, and reconstruct the signal using the inverse transform.

In Ref. 92 the wavelet pyramid transform is considered for analyzing arrival times estimation in seismic traces. The researchers conclude that preliminary denoising is essential. The wavelet-based filtering preserves high-frequency components necessary for finding the boundaries between oscillatory components.

12.4.2.4 Application: Phoneme Recognition. The mixed-domain transforms offer different structural decomposition tools for the speech analyst. In Chapter 9, we considered two types of algorithm for localized frequency analysis: the filter bank and the time-frequency map. The wavelet pyramid decomposition (Figure 12.5) furnishes a filter bank scheme where the outputs are independent of one another and perfect reconstruction of the original signal is possible. In addition, both discrete time-frequency and time-scale transforms support a time-frequency map methodology. The difference between the two is that the time-frequency cells of the Gabor transform, for example, have a fixed time-domain extent (Chapter 10). The wavelet transform cells vary their time spans; cells tuned to higher frequencies have narrower time domain support. Local-frequency estimation, as a preliminary step for recognizing phonemes, remains an active area of research.

It is difficult to design a pitch detector that adapts to both high and low speech frequencies while maintaining adequate noise immunity [93–95]. Recently, the dyadic continuous wavelet transform, given by

$$F_{\psi}(a, b) = W[x(t)](a, b) = \int_{-\infty}^{\infty} x(t) \overline{\psi_{a,b}(t)} dt \equiv \langle x(t), \psi_{a,b}(t) \rangle, \quad (12.34a)$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad (12.34b)$$

$x(t)$ has finite energy, and $a = 2^i$ for some integer i , has been applied to this problem [96]. One advantage is that the analysis then corresponds to the apparent time-scale operation of the human auditory system [67]. Surprisingly, the researchers report that only a few scales $a = 2^i$ are necessary for accurate detection [96]. Compared to conventional time- and frequency-domain methods, the dyadic wavelet pitch detector:

- (i) Is robust to nonstationary signals within its analysis window;
- (ii) Works on a wide range of pitch signals, such as from male and female speakers;
- (iii) Can detect the beginning of the voiced segment of the speech sample, making it the possible basis for a pitch detection algorithm that operates synchronously with the pitch bearing event;
- (iv) Is superior within low frequencies to pitch determination by the time-domain autocorrelation method [95];
- (v) Is superior within high frequencies to the frequency-domain cepstrum method [97].

12.4.3 Texture Analysis

Until recently, texture has been a persistently problematic area for signal and image analysis. Although the human subject readily distinguishes visual textures, it has

hitherto not been possible to classify them with computerized algorithms, let alone provide theoretical models for synthesizing visually realistic textures.

12.4.3.1 Mixed-Domain Strategies. Some promising early applications of the wavelet transform were to texture analysis [98]. An important contribution to texture analysis and synthesis has come from applying overcomplete wavelet pyramids to the problem [99]. A variety of statistics on the transform coefficients are used to characterize textures. Deletion of certain groups of statistical parameters and subsequent flawed reconstruction of the original image demonstrates that the necessity of the chosen statistics.

12.4.3.2 Application: Defect Detection and Classification. It is possible to apply the statistics of overcomplete pyramids to the problem of texture flaw detection. This application is important in manufacturing defect detection systems, for example. The algorithm of Ref. 99 is capable of synthesizing textures that appear to lie in between two others and offers the promise of a statistical divergence measure for textures. Defects in local regions can be detected by developing the statistical parameters from a prototype sample and comparing them to statistics extracted from candidate textures.

12.5 PATTERN RECOGNITION NETWORKS

This section explains pattern recognition methods that are useful for analyzing signals that have been decomposed through mixed domain transforms.

12.5.1 Coarse-to-Fine Methods

Pattern recognition where the time-domain size of the recognized signal structures are unknown present a variety of problems for the algorithm designer. In particular, the shape recognition computations can require more time than is available in real-time. We have already reviewed a variety of multiresolution methods for this purpose:

- Multiscale signal edge operators [45–48];
- Time-scale representations such as the wavelet multiresolution analysis [1] and the Laplacian pyramid decomposition [2];
- Scale-space smoothing with a range of kernel sizes [5].

Such decompositions demand large numbers of floating-point multiplications and additions. However, online process control and speech recognition software must keep a real-time pace and make a recognition decision with a fraction of a second. Sometimes, cost constrains the type of processor. One way to stay within the paradigms provided by time-scale signal analysis, yet achieve a real-time recognition time is to employ the classic Haar wavelet approximation [16–18] (Figure 12.23). How to do this is described in process control applications [100, 101].

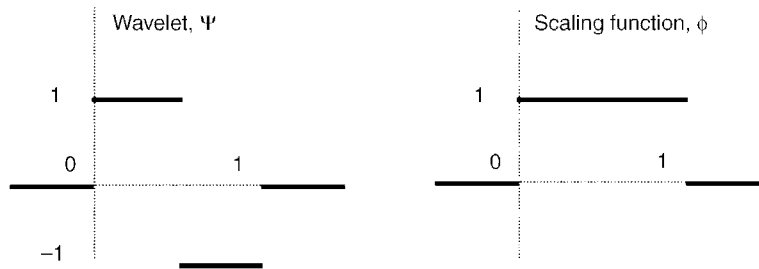


Fig. 12.23. The Haar MRA wavelet and scaling function. Translations and dilations of the wavelet form an orthonormal basis for all finite-energy signals, so the Haar representation can accommodate any signal shape. The scaling function can be used to develop step function approximations to signal shapes. The Fourier characteristics of these approximations are not as attractive as other decompositions, such as smooth spline functions. However, with this technique it is possible to implement a real-time signal decomposition with small industrial control computers.

It is possible to approximate signals by decomposing them into steps or to model them using the signal envelopes [101]. Figure 12.24 shows an example of an endpoint signal from the optical emission monitoring application in [101].

The idea behind using the Haar MRA for signal pattern recognition is that there is a simple relation between certain coarse and fine resolution patterns that allows the application to economize on matching.

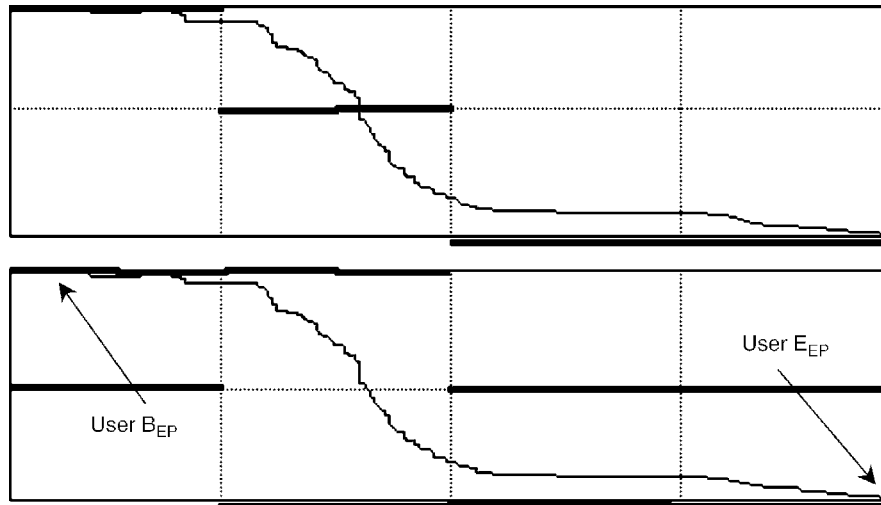


Fig. 12.24. Two methods to find Haar approximations for signal regions: projection to the mean (*top*) and the Haar envelope (*bottom*). In this figure, the shapes represent optical emission endpoint traces selected by users.

12.5.2 Pattern Recognition Networks

A pattern recognition network is a set of pattern detector elements or *neurons* tuned to different resolutions and different signal shapes [101].

Network *training* creates multiple resolution models of the signal shapes which are stored inside the neurons. The multiple resolution matching provides noise immunity during recognition, although linear and nonlinear smoothing operations help to clean acquired signals. To trim the number of pattern detectors in the network, the finest resolutions necessary for the model time and dynamic range divisions can be estimated as follows [101].

Given an input data file $x(n)$ with N values, one computes the discrete Fourier transform:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-2\pi jnk}. \quad (12.35)$$

Then, the non-DC coefficients representing the bulk of the signal's energy are selected: $k = 1, 2, \dots, k_c, N-1, N-2, \dots, N-k_c$. This selection is based on an energy percentage threshold. Then the signal detectors only need enough time-domain resolution so as to capture the shape of a single cycle of wavelength NT/k_c seconds, where T is the real-time application sampling rate. The length of a signal shape model, together with this minimum resolution value, determines the maximum length in samples of the discrete patterns stored in the detector elements.

The scheme of Ref. 101 also limits the dynamic range of signal models. Again using heuristic thresholds, typically modified by the application user, the amount of noise in the signal models is estimated. Then using a DFT approach again, the necessary dynamic range resolution in the step-shaped signal models is found. The result is a rectangular array of pattern detectors, from the lowest time resolution to the highest and from the lowest dynamic range division to the highest.

The projection of an acquired signal onto the step functions of the Haar representation are the coarse resolution representations used as models for pattern detection. Each "neuron" is a step function pattern detector. Before the network runs on real data, the patterns are checked against previously acquired data sets. One criterion is *stability*—how long a signal pattern persists in the data stream. Another criterion is the tendency to make false detections in the input data. Any neurons that fail to meet these criteria are disabled and not used by the network on real data.

The neurons of the network are interconnected with enabling and disabling links. When a coarse resolution node does not detect its established block shape pattern, it may disable certain finer resolution nodes. An example is shown in Figure 12.25.

At run time, the network presents the current and past signal data to each neuron. Each node computes the step pattern according to its particular resolutions. It compares the candidate pattern to its training pattern. When a node shows no match, higher-resolution nodes whose time or range resolutions are multiples of the non-matching unit cannot possibly activate and are disabled (Figure 12.25). When a node actually finds its pattern, all nodes of lesser resolution are disabled. The network continues to seek a more precise registration of the signal pattern. Any

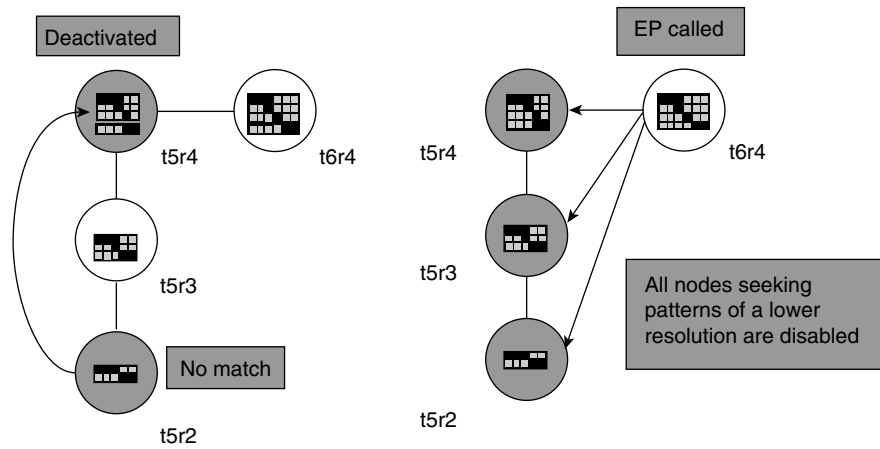


Fig. 12.25. Some interconnections of EP pattern detectors [101]. Node names indicate the time-domain and dynamic range resolutions of the Haar decomposition step functions that approximate the EP region. When a node shows no match, certain higher-resolution nodes cannot possibly activate and are disabled (*left*). When a node detects its pattern, the network disables all nodes of lesser resolution while it seeks a more precise registration of the signal shape. There is a single output node. Any enabled neuron that detects its own pattern can activate the output node to signal the detection of the prototype pattern.

enabled neuron that detects its own pattern can activate the output node to indicate a recognition success.

This is the key idea behind using the Haar step functions to model and compare signals. With a faster computer or freedom from the real-time processing requirement, other multiresolution approximations can be used. Of course, in this situation, the relation between detectors of different resolutions is not so easy to characterize and remains a potential problem for the algorithm design.

12.5.3 Neural Networks

Neural networks are an alternative to the structured design of the pattern recognition network above. Both supervised and unsupervised neural networks have been intensively studied in the last 20 years. An advantage of neural networks is that their training can be conditioned by training data, learning, as it were, the salient patterns present in the raw data [102]. The problem is that large amounts of data are sometimes necessary to train such networks. An example of applying neural networks to semiconductor process control is Ref. 103.

12.5.4 Application: Process Control

In semiconductor integrated circuit fabrication, plasma etch processes selectively remove materials from silicon wafers in a reactor [104]. The chemical species in the

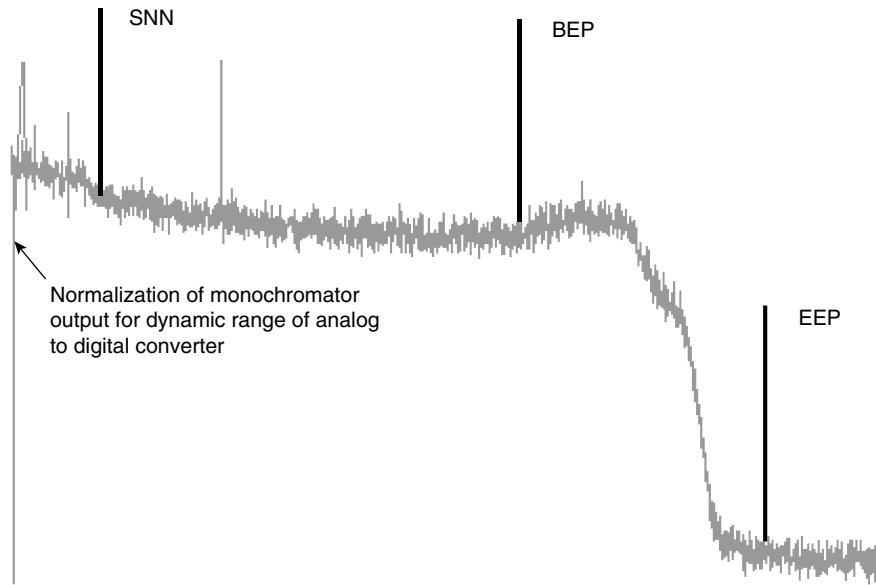


Fig. 12.26. Plasma etch optical emission trace from Ref. 101. The user acquires a sample trace and must indicate the points at which valid data begin (SNN), the endpoint is beginning (BEP), and the endpoint event is over (EEP). A Haar model can be made out of the data for training a pattern recognition network.

reaction emit characteristic wavelengths of light. It is typical to employ a monochromator and digitizer pass the light intensity signal to computer software algorithms to monitor and control the etch progress. When the target layer disappears, process endpoint occurs; the control computer extinguishes the plasma. For signal analysis, the problem is that process endpoints vary from wafer to wafer over a run. Both traditional neural networks [103] and the structured pattern recognition network [100, 101] have been used for this application. Extensive testing is reported in Ref. 101.

12.6 SIGNAL MODELING AND MATCHING

This final section mentions methods for extracting signal structure that have been particularly popular in conjunction with time-frequency and time-scale transforms.

12.6.1 Hidden Markov Models

The hidden Markov model (HMM) is a stochastic state machine that is especially useful for sequential analysis of data. Thus, it has been applied widely in speech recognition [105], handwriting recognition [106], and biological sequence analysis [107].

12.6.2 Matching Pursuit

The matching pursuit is a greedy, iterative algorithm [108]. In Ref. 109 the method is applied with overcomplete dictionaries of damped sinusoids. The method typically uses an overcomplete dictionary for numerical stability. It improves upon traditional techniques such as least squares, singular value decomposition, and orthonormal basis decompositions—for example, the orthogonal wavelet pyramid [1].

12.6.3 Applications

Two of the most important applications of hidden Markov models have been in speech recognition and biological sequence analysis.

12.6.3.1 Speech Analysis. Speech recognition applications are described in the classic tutorial [105].

12.6.3.2 Protein Analysis. A tutorial on protein analysis is Ref. 110.

12.7 AFTERWORD

There is no doubt that mixed-domain signal transforms, combining both time and either frequency or scale information, have altered the signal processing landscape. It is almost impossible today to browse an academic journal in the discipline and not find a contribution that concentrates on the theoretical or practical implications of these techniques. This text introduces the new methods into the mainstream of signal processing education.

Some of the trends we identified when contemplating the task of writing this book have become clearer and stronger. It is still true that learning windowed Fourier and wavelet transforms has its mathematical challenges. The entire signal processing research community has embraced the underlying mathematical tools—especially Hilbert space theory—even though they may entail a steep learning curve. Here we have developed the mathematics incrementally and colored it with terminology, notations, and concepts directly relevant to signal theory. This might relieve some anxiety and make the climb less daunting. Also, the selection of algorithms and applications in *signal analysis* not only reflects the modern mathematical slant but also emphasizes signal understanding as opposed to pure processing. We think that this too is timely, as more and more automated signal recognition technologies have intruded into our lives.

Of the new mixed-domain transforms, probably the most surprises came from orthogonal wavelets. These functions captivated researchers in so many areas. Today, however, the associated signal decomposition seem to be retreating into compression applications, with its analysis powers having been tested and found lacking for pattern recognition. But it does offer insights into texture characterization. On the other hand, the continuous wavelet transform rises up to be the better tool for

transient signals. The exponential short-time Fourier methods cannot be sparse and complete, as we now know, but this has not prevented them from becoming the tool of choice in a number of early processing applications. The time-frequency and time-scale tools are just alternative ways to break a signal into pieces and sort them out into a structural description. Beyond that, their importance is that they highlight the special global nature of the wellspring of them all—Fourier's transform.

REFERENCES

1. S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 674–693, July 1989.
2. P. J. Burt and E. H. Adelson, The Laplacian pyramid as a compact image code, *IEEE Transactions on Communication*, vol. 31, no. 4, pp. 532–540, April 1983.
3. S. Mallat, *A Wavelet Tour of Signal Processing*, San Diego, CA: Academic Press, 1998.
4. O. Rioul and M. Vetterli, Wavelets and signal processing, *IEEE SP Magazine*, pp. 14–38, October 1991.
5. R. L. Allen, F. A. Kamangar, and E. M. Stokely, Laplacian and orthogonal wavelet pyramid decomposition in coarse-to-fine registration, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3536–3543, December 1993.
6. R. Y. Wong and E. L. Hall, Sequential hierarchical scene matching, *IEEE Transactions on Computers*, vol. C-27, no. 4, pp. 359–366, April 1978.
7. D. Marr, *Vision*, New York: W. H. Freeman, 1982.
8. T. Lindeberg, *Scale-Space Theory in Computer Vision*, Hingham, MA: Kluwer, 1994.
9. B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever, eds., *Scale-Space Theory in Computer Vision* (Proceedings, First International Conference, Scale-Space '97, Utrecht, The Netherlands), Berlin: Springer-Verlag, 1997.
10. R. H. Bamberger and M. J. T. Smith, A filter bank for the directional decomposition of images: Theory and design, *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 882–893, April 1992.
11. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, Shiftable multiscale transforms, *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992.
12. P.-G. Lemarié, Ondelettes à localisation exponentielle, *Journal de Mathématiques Pures et Appliquées*, vol. 67, pp. 227–236, 1988.
13. F. Mokhtarian and A. Mackworth, Scale-based description and recognition of planar curves and two-dimensional shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 34–43, January 1986.
14. M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Englewood Cliffs, NJ: Prentice-Hall, 1976.
15. J.-O. Strömberg, A modified Franklin system and higher order spline systems on R^n as unconditional bases for Hardy spaces, in W. Becker, A. P. Calderon, R. Fefferman, and P.W. Jones, eds., *Proceedings of the Conference in Honor of Antoni Zygmund*, vol. II, New York: Wadsworth, pp. 475–493, 1981.

16. A. Haar, Zur theorie der orthogonalen Functionensysteme, *Mathematische Annalen*, vol. 69, pp. 331–371, 1910.
17. I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.
18. H. S. Stone, J. Le Moigne, and M. McGuire, The translation sensitivity of wavelet-based registration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1074–1081, October 1999.
19. M. Antonini, M. Barlaud, I. Daubechies, and P. Mathieu, Image coding using vector quantization in the wavelet transform domain, *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 2297–2300, April 1990.
20. S. G. Mallat and J. Froment, Second generation compact image coding with wavelets, in C. K. Chui, ed., *Wavelets: A Tutorial in Theory and Applications*, San Diego, CA: Academic Press, 1992.
21. D. Sinha, and A. H. Tewfik, Low bit rate transparent audio compression using adapted wavelets, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3463–3479, December 1993.
22. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Upper Saddle River, NJ: Prentice-Hall, 1995.
23. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, San Diego, CA: Academic Press, 1982.
24. G. K. Wallace, The JPEG still picture compression standard, *Communications of the ACM*, vol. 34, no. 4, pp. 31–44, April 1991.
25. N. Ahmed, T. Natarajan, and K. R. Rao, Discrete cosine transform, *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, January 1974.
26. A. Gersho and A. M. Gray, *Vector Quantization and Signal Compression*, Boston: Kluwer, 1992.
27. F. W. C. Campbell and J. J. Kulikowski, Orientation selectivity of the human visual system, *Journal of Physiology*, vol. 197, pp. 437–441, 1966.
28. H. Malvar, Lapped transforms for efficient transform/subband coding, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 969–978, June 1990.
29. Y. Meyer: *Wavelets: Algorithms and Applications*, Philadelphia: Society for Industrial and Applied Mathematics, 1993.
30. R. R. Coifman, Y. Meyer, and V. Wickerhauser, Wavelet analysis and signal processing, in M. B. Ruskai, et al., eds., *Wavelets and Their Applications*, Boston: Jones and Bartlett, pp. 153–178, 1992.
31. R. R. Coifman, Y. Meyer, and V. Wickerhauser, Size properties of wavelet-packets, in M. B. Ruskai et al., eds., *Wavelets and Their Applications*, Boston: Jones and Bartlett, pp. 453–470, 1992.
32. D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, vol. 81, pp. 425–455, 1994.
33. D. L. Donoho, De-noising by soft-thresholding, *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, March 1995.
34. D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, Density estimation by wavelet thresholding, *Annals of Statistics*, vol. 24, pp. 508–539, 1996.

35. C. M. Stein, Estimation of the mean of a multivariate normal distribution, *Annals of Statistics*, vol. 9, pp. 1135–1151, 1981.
36. K. R. Castleman, *Digital Image Processing*, Upper Saddle River, NJ: Prentice-Hall, 1996.
37. Nowak and Baraniuk, Wavelet-domain filtering for photon imaging systems, *IEEE Trans. IP*, pp. 666–678, May 1999.
38. D. A. Pollen and S. F. Ronner, Visual cortical neurons as localized spatial frequency filters, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 5, pp. 907–916, September–October 1983.
39. J. J. Kulikowski, S. Marcelja, and P. O. Bishop, Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex, *Biological Cybernetics*, vol. 43, pp. 187–198, 1982.
40. F. Attneave, Some informational aspects of visual perception, *Psychological Review*, vol. 61, pp. 183–193, 1954.
41. H. Asada and M. Brady, The curvature primal sketch, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 2–14, January 1986.
42. I. Biedermann, Human image understanding: Recent research and a theory, *Computer Vision, Graphics, and Image Processing*, vol. 32, pp. 29–73, 1985.
43. A. P. Witkin, Scale-space filtering, *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, W. Germany, 1983. See also A. P. Witkin, Scale-space filtering, in *From Pixels to Predicates*, A. P. Pentland, ed., Norwood, NJ: Ablex, 1986.
44. T. Lindeberg, Scale space for discrete signals, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 234–254, March 1990.
45. J. Canny, A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, November 1986.
46. H. D. Tagare and R. J. P. deFigueiredo, On the localization performance measure and optimal edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 12, pp. 1186–1190, 1990.
47. R. J. Qian and T. S. Huang, Optimal edge detection in two-dimensional images, *IEEE Transactions on Image Processing*, vol. 5, no. 7, pp. 1215–1220, 1996.
48. M. Gökmen and A. K. Jain, $\lambda\tau$ -space representation of images and generalized edge detector, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 545–563, June 1997.
49. S. Mallat, Zero-crossings of a wavelet transform, *IEEE Transactions on Information Theory*, vol. 37, no. 4, pp. 1019–1033, July 1991.
50. S. Mallat and S. Zhong, Characterization of signals from multiscale edges, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, July 1992.
51. Z. Berman and J. S. Baras, Properties of the multiscale maxima and zero-crossings representations, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3216–3231, 1993.
52. G. A. Orban, *Neuronal Operations of the Visual Cortex*, Berlin: Springer-Verlag, 1984.
53. D. H. Hubel and T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *Journal of Physiology*, vol. 160, pp. 106–154, 1962.

54. F. W. C. Campbell and J. Robson, Application of Fourier analysis to the visibility of gratings, *Journal of Physiology*, vol. 197, pp. 551–566, 1968.
55. F. W. C. Campbell and J. J. Kulikowski, Orientation selectivity of the human visual system, *Journal of Physiology*, vol. 195, pp. 437–441, 1966.
56. C. Blakemore and F. W. C. Campbell, On the existence in the human visual system of neurons selectively sensitive to the orientation and size of retinal images, *Journal of Physiology*, vol. 203, pp. 237–260, 1969.
57. J. P. Jones and L. A. Palmer, The two-dimensional spatial structure of simple receptive fields in cat striate cortex, *Journal of Neurophysiology*, vol. 58, pp. 1187–1211, 1987.
58. J. P. Jones and L. A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, *Journal of Neurophysiology*, vol. 58, pp. 1233–1258, 1987.
59. J. P. Jones, A. Stepnoski, and L. A. Palmer, The two-dimensional spectral structure of simple receptive fields in cat striate cortex, *Journal of Neurophysiology*, vol. 58, pp. 1212–1232, 1987.
60. S. Marcelja, Mathematical description of the responses of simple cortical cells, *Journal of the Optical Society of America*, vol. 70, pp. 1297–1300, 1980.
61. D. Gabor, Theory of communication, *Journal of the Institute of Electrical Engineers*, vol. 93, pp. 429–459, 1946.
62. M. Porat and Y. Y. Zeevi, The generalized Gabor scheme of image representation in biological and machine vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 452–468, July 1988.
63. D. A. Pollen and S. F. Ronner, Visual cortical neurons as localized spatial frequency filters, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 5, pp. 907–916, September/October 1983.
64. J. G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160–1169, July 1985.
65. J. G. Daugman, Spatial visual channels in the Fourier plane, *Vision Research*, vol. 24, no. 9, pp. 891–910, 1984.
66. S. J. Anderson and D. C. Burr, Receptive field size of human motion detection units, *Vision Research*, vol. 27, no. 4, pp. 621–635, 1987.
67. X. Yang, K. Wang, and S.A. Shamma, Auditory representation of acoustic signals, *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 824–839, March 1992.
68. M. R. Turner, Texture discrimination by Gabor functions, *Biological Cybernetics*, vol. 55, pp. 71–82, 1986.
69. A. C. Bovik, M. Clark, and W. S. Geisler, Multichannel texture analysis using localized spatial filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 55–73, January 1990.
70. T. Weldon and W. E. Higgins, Designing multiple Gabor filters for multitexture image segmentation, *Optical Engineering*, vol. 38, no. 9, pp. 1478–1489, September 1999.
71. J. Lu, J. B. Weaver, D. M. Healy, Jr., and Y. Xu, Noise reduction with a multiscale edge representation and perceptual criteria, *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Victoria, BC, Canada, pp. 555–558, October 4–6, 1992.

72. S. Mallat and S. Zhong, Wavelet transform maxima and multiscale edges, in M. B. Ruskai, et al., eds., *Wavelets and Their Applications*, Boston: Jones and Bartlett, pp. 67–104, 1992.
73. M. Holschneider, *Wavelets: An Analysis Tool*, New York: Oxford University Press, 1995.
74. S. Mallat and W. L. Hwang, Singularity detection and processing with wavelets, *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, March 1992.
75. M. Akay, ed., *Time-Frequency and Wavelets in Biomedical Signal Processing*, New York: Wiley-IEEE Press, 1997.
76. M. Unser and A. Aldroubi, A review of wavelets in biomedical applications, *Proceedings of the IEEE*, vol. 84, no. 4, pp. 626–638, April 1996.
77. J. R. Cox, Jr., F. M. Nolle, and R. M. Arthur, Digital analysis of electroencephalogram, the blood pressure wave, and the electrocardiogram, *Proceedings of the IEEE*, vol. 60, pp. 1137–1164, 1972.
78. P. E. Tikkanen, Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal, *Biological Cybernetics*, vol. 80, no. 4, pp. 259–267, April 1999.
79. B. A. Rajoub, An efficient coding algorithm for the compression of ECG signals using the wavelet transform, *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 4, pp. 255–362, April 2002.
80. B.-U. Kohler, C. Hennig, and R. Orglmeister, The principles of software QRS detection, *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, pp. 42–57, January/February 2002.
81. C. Li, C. Zheng, and C. Tai, Detection of ECG characteristic points using wavelet transforms, *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 21–28, January 1995.
82. S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels, Wavelet transformbased QRS complex detector, *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 7, pp. 838–848, July 1999.
83. L. Khadra, M. Matalgah, B. El_Asir, and S. Mawagdeh, Representation of ECG-late potentials in the time frequency plane, *Journal of Medical Engineering and Technology*, vol. 17, no. 6, pp. 228–231, 1993.
84. F. B. Tuteur, Wavelet transformations in signal detection, in *Wavelets: Time-Frequency Methods and Phase Space*, J. M. Combes, A. Grossmann, and P. Tchamitchian, eds., 2nd ed., Berlin: Springer-Verlag, pp. 132–138, 1990.
85. K. Murano, S. Unagami, and F. Amano, Echo cancellation and applications, *IEEE Communications Magazine*, vol. 28, no. 1, pp. 49–55, January 1990.
86. A. Gilloire and M. Vetterli, Adaptive filtering in sub-bands with critical sampling: Analysis, experiments and applications to acoustic echo cancellation, *IEEE Transactions on Signal Processing*, vol. 40, no. 8, pp. 1862–1875, August 1992.
87. O. Tanrikulu, B. Baykal, A. G. Constantinides, and J. A. Chambers, Residual echo signal in critically sampled subband acoustic echo cancellers based on IIR and FIR filter banks, *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 901–912, April 1997.
88. A. Grossmann and J. Morlet, Decomposition of Hardy functions into square Integrable wavelets of constant shape, *SIAM Journal of Mathematical Analysis*, vol. 15, pp. 723–736, July 1984.

89. P. Goupillaud, A. Grossmann, and J. Morlet, Cycle-octave and related transforms in seismic signal analysis, *Geoexploration*, vol. 23, pp. 85–102, 1984–1985.
90. J. L. Larsonneur and J. Morlet, Wavelets and seismic interpretation, in J. M. Combes, A. Grossmann, and P. Tchamitchian, eds., *Wavelets: Time-Frequency Methods and Phase Space*, 2nd ed., Berlin: Springer-Verlag, pp. 126–131, 1990.
91. X.-G. Miao and W. M. Moon, Application of wavelet transform in reflection seismic data analysis, *Geosciences Journal*, vol. 3, no. 3, pp. 171–179, September 1999.
92. G. Olmo and L. Lo Presti, Applications of the wavelet transform for seismic activity monitoring, in *Wavelets: Theory, Applications, and Applications*, C. K. Chui, L. Montefusco, and L. Puccio, eds., San Diego, CA: Academic Press, pp. 561–572, 1994.
93. M. Cooke, S. Beet, and M. Crawford, eds., *Visual Representations of Speech Signals*, Chichester: Wiley, 1993.
94. T. Parsons, *Voice and Speech Processing*, New York: McGraw-Hill, 1987.
95. W. Hess, *Pitch Determination of Speech Signals: Algorithms and Devices*, New York: Springer-Verlag, 1983.
96. S. Kadambe and G. F. Boudreaux-Bartels, Application of the wavelet transform for pitch detection of speech signals, *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 917–924, March 1992.
97. A. M. Noll, Cepstrum pitch determination, *Journal of the Acoustical Society of America*, vol. 41, pp. 293–309, February 1967.
98. T. Chang and C.-C. J. Kuo, Texture analysis and classification with tree-structured wavelet transform, *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 429–441, October 1993.
99. J. Portilla and E. P. Simoncelli, A parametric texture model based on joint statistics of complex wavelet coefficients, *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, October 2000.
100. R. L. Allen, R. Moore, and M. Whelan, Multiresolution pattern detector networks for controlling plasma etch reactors, *Process, Equipment, and Materials Control in Integrated Circuit Manufacturing*, Proceedings SPIE 2637, pp. 19–30, 1995.
101. R. L. Allen, R. Moore, and M. Whelan, Application of neural networks to plasma etch endpoint detection, *Journal of Vacuum Science and Technology (B)*, pp. 498–503, January–February 1996.
102. J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*, Redwood City, CA: Addison-Wesley, 1991.
103. E. A. Rietman, R. C. Frye, E. R. Lory, and T. R. Harry, Active neural network control of wafer attributes in a plasma etch process, *Journal of Vacuum Science and Technology*, vol. 11, p. 1314, 1993.
104. D. M. Manos and G. K. Herb, Plasma etching technology—An overview, in D. M. Manos and D. L. Flamm, eds., *Plasma Etching: An Introduction*, Boston: Academic Press, 1989.
105. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, February 1989.
106. N. Arica and F. T. Yarman-Vural, Optical character recognition for cursive handwriting, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 801–813, June 2002.

107. R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis*, Cambridge: Cambridge University Press, 1998.
108. S. Mallat and S. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing*, pp. 3397–3415, December 1993.
109. M. M. Goodwin and M. Vetterli, Matching pursuit and atomic signal models based on recursive filter banks, *IEEE Transactions on Signal Processing*, pp. 1890–1902, July 1999.
110. R. Karchin, *Hidden Markov Models and Protein Sequence Analysis*, Honors Thesis, Computer Engineering Department, University of California, Santa Cruz, June 1998.

PROBLEMS

1. Using material from Chapters 9 and 11, suppose we are given a multiresolution analysis of finite-energy signals.
 - (a) Show that the discrete lowpass filter $H(\omega)$ associated to the MRA satisfies $|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1$;
 - (b) Let $g(n) = g_1(n) = (-1)^{1-n}h(1-n)$ and $G(\omega) = e^{-j\omega}\overline{H(\omega + \pi)}$. Show that, indeed, $g(n)$ is the inverse discrete-time Fourier transform of $G(\omega)$.
 - (c) Show that $|H(\omega)|^2 + |G(\omega)|^2 = 1$.
 - (d) Using the perfect reconstruction criterion of Chapter 9, show that $\sqrt{2}h(n)$ is a quadrature mirror filter (QMF).
 - (e) Sketch a reconstruction diagram using $h(n)$ and $g(n)$ for the reconstruction of the original signal decomposed on the pyramid [1].
2. In the QMF pyramid decomposition (Figure 12.5), let $\tilde{h}(n) = h(-n)$ be the reflection of $h(n)$ and $\tilde{H}(z)$ be its z -transform. Similarly, let $\tilde{g}(n) = g(-n)$ and $\tilde{G}(z)$ be the transfer function of the filter with impulse response $\tilde{g}(n)$.
 - (a) Show that subsampling a signal $x(n)$ by two followed by $\tilde{H}(z)$ filtering is the same discrete system as $\tilde{H}(z^2)$ filtering followed by subsampling [4].
 - (b) Applying the same idea to $\tilde{g}(n)$, prove filtering with $\tilde{H}(z)\tilde{H}(z^2)$ and $\tilde{H}(z)\tilde{G}(z^2)$ and subsampling by four produces the level -2 approximate and detail coefficients, respectively.
 - (c) Show that we can compute the impulse response of the filter with transfer function $\tilde{H}(z)\tilde{H}(z^2)$ by convolving $\tilde{h}(n)$ with the filter obtained by putting a zero between every $\tilde{h}(n)$ value.
 - (d) State and prove a property similar to (c) for $\tilde{H}(z)\tilde{G}(z^2)$.
 - (e) State and prove properties for level $l = -L$, where $L > 0$, that generalize these results.
3. Suppose $p > 0$ and define the filter G_p as in (12.13). Let O_i be the orthogonal complement of V_i inside V_{i+1} : $V_i \perp O_i$ and $V_{i+1} = V_i \oplus O_i$.
 - (a) Show (12.16).

- (b) Show (12.17).
 - (c) Show (12.18).
 - (d) Since $\{\phi(t-k)\}_{k \in \mathbb{Z}}$ is an orthonormal basis for V_0 , explain the expansion (12.19).
 - (e) By Fourier transformation of (12.19), show that $G_p(\omega) = \frac{\Psi(2^p \omega)}{\Phi(2\omega)}$.
4. Suppose that $y(n) = x(n-2)$ and both signal $x(n)$ and $y(n)$ are decomposed using the orthogonal wavelet pyramid.
- (a) How do the first-level $L = -1$ coefficients for $y(n)$ differ from the first-level coefficients for $x(n)$?
 - (b) Generalize this result to delays that are higher powers of 2.
5. Show by simple convolutions on discrete steps and ridge edges that discrete highpass filters $g_p(n)$ given in Figure 12.3 and Figure 12.4 function as edge detectors for the orthogonal wavelet pyramid.
6. Suppose $g(t) = Ae^{-Bt^2}$ is a Gaussian of zero mean and $\psi(t) = \frac{d}{dt}g(t)$.
- (a) Show that $\psi(t)$ is a wavelet.
 - (b) Let $x_a(t) = ax(at)$ be the scaled dilation of $x(t)$ by factor $a = 2^{-i}$ for $i \in \mathbb{Z}$, with $i > 0$. Define the wavelet transform $(W_{-i}x)(t) = (\psi_a * x)(t)$. Show that

$$(W_{-i}x)(t) = a \frac{d}{dt}(g_a * x)(t). \quad (12.36)$$

- (c) Explain the significance of $|W_{-i}x|$ being large.
 - (d) Explain the significance of large $|W_{-i}x|$ when a is large. What if a is small?
7. Suppose $x(t)$ is discontinuous at $t = t_0$. Show that its Libschitz regularity at t_0 is zero.

Advanced problems and projects.

8. Implement the multiscale matching and registration algorithm of Section 12.1.3.
- (a) Use the cubic spline MRA as described in the text.
 - (b) Use the Laplacian pyramid.
 - (c) Use the MRA based on piecewise continuous functions.
 - (d) Develop matching and registration experiments using object boundaries or signal envelopes.
 - (e) Compare the performance of the above algorithms based on your chosen applications.
 - (f) Explore the effect of target shape support in the candidate signal data.

9. Derive the impulse responses for the $h_p(n)$ and $g_p(n)$ for the case where the MRA is
 - (a) Based on the Haar functions;
 - (b) The Stromberg MRA.
10. Compare linear and nonlinear filtering of the electrocardiogram to the wavelet de-noising algorithms.
 - (a) Obtain and plot an ECG trace (for example, from the signal processing information base; see Section 1.9.2.2).
 - (b) Develop algorithms based on wavelet noise removal as in Section 12.2.1. Compare hard and soft thresholding methods.
 - (c) Compare your results in (b) to algorithms based on edge-preserving nonlinear filters, such as the median filter.
 - (d) Compare your results in (b) and (c) to algorithms based on linear filters, such as the Butterworth, Chebyshev, and elliptic filters of Chapter 9.
 - (e) Consider the requirements of real-time processing and analysis. Reevaluate your comparisons with this in mind.
11. Compare discrete and continuous wavelet transforms for QRS complex detection [81, 82].
 - (a) Using your data set from the previous problem, apply a nonlinear filter to remove impulse noise and a convolutional bandpass filter to further smooth the signal.
 - (b) Decompose the filtered ECG signal using one of the discrete wavelet pyramid decompositions discussed in the text (the cubic spline multiresolution analysis, for instance). Describe the evolution of the QRS complexes across multiple scales [81]. Develop a threshold-based QRS detector and assess its usefulness with regard to changing scale and QRS pulse offset within the filtered data.
 - (c) Select a scale for decomposition based on a continuous wavelet transform [82]. Compare this method of analysis to the discrete decomposition in (b).
 - (d) Consider differentiating the smoothed ECG signals to accentuate the QRS peak within the ECG. Does this improve either the discrete or continuous algorithms?
 - (e) Consider squaring the signal after smoothing to accentuate the QRS complex. Does this offer any improvement? Explain.
 - (f) Do soft or hard thresholding with wavelet de-noising help in detecting the QRS complexes?
 - (g) Synthesize some defects in the QRS pulse, such as ventricular late potentials, and explore how well the two kinds of wavelet transform perform in detecting this anomaly.