



CONTRIBUTED ARTICLE

Wavelet Transforms and Neural Networks for Compression and Recognition

HAROLD SZU, BRIAN TELFER* AND JOSEPH GARCIA

Naval Surface Warfare Center

(Received and accepted 22 December 1994)

Abstract—Robust recognition for image and speech processing needs data compression that preserves features. To accomplish this, we have utilized the discrete wavelet transforms and the continuous wavelet transforms (CWT) together with artificial neural networks (ANN) to achieve automatic pattern recognition. Our approach is motivated by the mathematical analog of the CWT to the human hearing and visual systems, e.g., the so-called Mexican hat and Gabor functions, Gaussian window, respectively. We develop an ANN method to construct an optimum mother wavelet that can organize sensor input data in the multiresolution format that seems to become essential for brain-style computing. In one realization, the architecture of our ANN is similar to that of a radial basis function approach, except that each node is a wavelet having three learnable parameters: weight W_{ij} , scale a , and shift b . The node is not a McCulloch–Pitts neuron but a “wave-on”. We still use a supervised learning conjugate gradient descent algorithm in these parameters to construct a “super-mother” wavelet from a superposition of a set of waveons—mother wavelets. Using these techniques, we can accomplish the signal-enhanced and feature-preserving compression, e.g., on the infrared images, that avoids the overtraining and overfitting that have plagued ANN’s ability to generalize and abstract information. Published by Elsevier Science Ltd

Keywords—Adaptive wavelet transform, Discrete wavelet transform, Data compression, Pattern recognition, Neural networks.

1. INTRODUCTION AND REVIEW

Open-field automatic pattern recognition (APR) differs in several respects from commercial pattern recognition applications such as character or speech recognition. APR encompasses a wide variety of sensors, e.g., laser radar, radar, sonar, electro-optic, in both passive and active models, and producing a wide variety of signal types, e.g., one-dimensional (1-D) signals, two-dimensional (2-D) images, range images. The APR environment is difficult to quantify and noncooperative, with the hostile objects often actively attempting to avoid detection/recognition, as opposed to commercial applications which allow control over the environment to enhance recognizability. The uncontrolled environment leads to a

tremendous variety of types of background clutter, and to partially occluded patterns that can appear in a wide range of scales and aspects. Patterns can also vary greatly due to environmental effects, e.g., the diurnal and climatic variations in infrared imagery. Immense amounts of data must be processed in real-time. Because of the difficulties and expense involved in collecting real data, simulated data must often be used during system development. All of these factors cause APR to be a very difficult problem. For a further introduction to the difficulties of APR, see Roth (1990), Sadjadi (1991) and MIT Lincoln Laboratory Journal (1993).

In both military and commercial APR, there is an increasing need for data compression in our wireless and wire communication systems for transmitting digital audio and video (e.g., B-ISDN picture phone and two-way television for commerce, education, and polling services). Compressibility is one of the major issues in the development of certain real-time APR systems [e.g., FBI Program in Automatic Fingerprint Identification System (AFIS) at the National Crime Information Center, e.g., to screen traffic violators

Acknowledgement: Support from the NSWCDD Independent Research Program is gratefully acknowledged.

Requests for reprints should be sent to Harold Szu, Centre for Adv. Comp. Studies, P.O. Box 44330, University of Louisiana, Lafayette, LA 70504-4330, USA.

* Present address: MIT Lincoln Labs, 244 Wood Street, Lexington, MA 02173, USA.

nation-wide in real-time for those wanted by the FBI]. The other issue is the representation versus the classification (Telfer & Szu, 1993).

The traditional compression achieves a low bit per pixel (bpp) rate through the IEEE Joint Photographic Expert Group (JPEG) standard, i.e., implemented by an 8×8 pixel moving window using the discrete cosine transform and the Huffman statistical and other run-length compression coding. Recently, the FBI has demonstrated, for fingerprints consisting of ridge discontinuities, that the discrete wavelet transforms (DWT) have achieved a 20:1 compression ratio without any loss of fingerprint minutiae (Hopper, 1994). Furthermore, they have shown the DWT to be fast since the computational complexity is of the order $O(N)$ for N data. In Section 4, this DWT technology is introduced from the historical viewpoint of Haar as the finite impulse response filter bank. Since it consists of a pair of lowpass and highpass filters of finite lengths, it is fast and can filter noise and smooth the signal recursively, thereby permitting the traditional compression scheme. The inverse DWT, which is lossless if the bpp is identical, can synthesize the data and recover an enhanced version. Using multispectral infrared images, we apply the edge-preserving DWT algorithm adopted by FBI and demonstrate 59:1 compression ratio for the infrared imagery and then we apply the continuous wavelet transforms (CWT) to extract the features for artificial neural network (ANN) APR with a much reduced false alarm rate. Since the size of an ANN is usually small, we wish to have a small number of features, making fast data compression desirable.

The CWT is another compression method having a continuous and selective feature extraction. Because it is directly related to the so-called Mexican hat or the difference of Gaussian (DoG) or the on-center-off-surround complex cell response function in the human visual system, it is therefore more flexible in extracting features from clutter. We have proved a super-mother theorem that is based on a linear superposition of admissible mother wavelets being itself admissible, and that shows that a smaller number of transform coefficients are required to represent features.

Since we apply ANN learning to construct a super-mother wavelet and use it directly, we develop a recognition algorithm for example to differentiate the vowel of "a" from the other vowels (Szu et al., 1992). Given complex speech data, this concept may be extended if we apply the Morlet wavelet to extract important formant features and then for example we can use an ANN to spot specific phonemes spoken by either a male or female speaker. In this sense, the CWT preprocessing is robust and produces highly compressed features, as the input, without the actual

voiced and unvoiced sounds used in conventional ANNs for APR. Because of the simplicity, there is the possibility to have a real-time integrated system consisting of a wavelet preprocessing chip followed by an ANN post-processing chip.

Traditionally, APR approaches are both geometric model-based and statistical feature-based (or signature-based) techniques. Model-based methods that employ stored image models are useful for occlusion because parts imply the whole. Feature-based methods are useful if the underlying physics-chemistry, e.g., the fuel spectral line ratio, is fully understood and explored. In both cases, some form of data compression and feature extraction is used as an important component of any APR system. Numerous feature extraction methods have been used for automatic target recognition (Szu & Telfer, 1994): Edge detection and Neocognitron-based features (e.g., Daniell et al., 1992), local spatial features (e.g., Thoet et al., 1992), Fourier transform magnitudes (Gorman & Sejnowski, 1988), projection pursuit-based features (Bachmann et al., 1994), and polar-log transforms (Szu & Messner, 1986). For feature-based methods, we have developed new neural network tools employing minimum-misclassification-error (MME) performance (Telfer & Szu, 1994), the adaptive wavelet transform (Szu et al., 1992), and fast annealing techniques (Szu & Hartley, 1987 a,b). The feature-based approach is particularly useful if the feature extraction can be patterned after the human visual system with the capability of self-generation of critical features that are locally invariant under the limited variability of motion, intensity, scale and rotation. The wavelet transform, provides both an excellent basis for compression and for selecting time or space/scale features for APR.

The paper is organized as follows. We introduce the CWT preprocessing in Section 2 from a generalized Fourier transform viewpoint. We then introduce the combination of CWT with an ANN in Section 3. Here we describe how the neuron transfer logic function is replaced by a wavelet kernel function and how a supervised ANN learning may be used to determine an optimum super-mother wavelet when given training data and classification performance measure. We define the DWT from a digital filter theory in Section 4, and describe the data compression of multispectral infrared imagery and give an application in the multispectral infrared pattern recognition in Section 5. We conclude in Section 6.

2. CONTINUOUS WAVELET TRANSFORMS

Wavelet transform (WT) is a linear transform similar to the Fourier transform (FT), but generalized for any wideband transient. This generalization becomes efficient from the matched filter viewpoint when the

kernel is tailor-made to the signal itself. Also, from pedagogical viewpoint, both the WT signal and its wavelet transform happen to share an identical acronym WT. This coincidence says when WT should be used and when not, i.e., the WT is advantageous when applied to wideband transients. Almost all real-world signals are wideband transients, and thus utilization of WT is advantageous in many cases. We emphasize the ability to construct a linear transform kernel (other than sinusoidal functions) as a basis to learn how the brain processes information from the eyes and ears. The relationship to the other orthogonal feature extraction methodology, e.g., Karhunen-Loeve (KL) transform will be given immediately after we define WT.

Three properties of WT are given as its definition as follows: (i) The transform kernel is a small wave that has a compact support and therefore requires no fixed window. This window employed by the short-time FT can generate artifacts in case of the high resolution seismic imaging for which the French effort of WT was initiated in 1980. The kernel has no zero frequency, d.c., component, so that the kernel when multiplied by a constant is always integrated to produce a zero area. This attribute is typical of any bandpass filter in producing a change detection. In other words, this change detection is typical to that of human visual experience where any uniform background will not be detected since it will not contribute to the Wavelet expansion. (ii) Wavelet bases are all generated from single template called mother wavelet from a real affine scale a and shift b operations: $t \rightarrow t' = (t - b)/a$, of which, e.g., a logarithmic base 2 or dyadic representation means $a/a_0 = 2^n$, $b/b_0 = m$ for all integer n, m . Consequently, the bandwidth Δf_0 of a mother wavelet and its central frequency f_0 is maintained at the same value as for all daughters. This definite proportionality constant is called the constant fidelity requirement: $Q = f_0/\Delta f_0$. Since any group wave requires a large enough bandwidth Δf to produce through interference a small time-localized spread Δt , the time-bandwidth product of $\Delta t \Delta f$ must be greater than one. For example, for a high pitch organ pipe $f_1 = 2f_0$ the corresponding bandwidth must be also doubled $\Delta f_1 = 2\Delta f_0$. Consequently, the organ pipe length becomes shorter $\Delta t_1 = \Delta t_0/2$. That is why a shorter pipe is used for high pitch, in which all the group waves have identical shape with different sizes. (iii) The necessary and sufficient condition for an admissible mother wavelet $h(t)$, to be proved in this Section 2, is that the inversely linear weighted power spectral density $|H(f)|^2/|f|$ must be integrated over all frequencies to be a finite constant.

A method commonly used for feature extraction is to compute the spatial pattern covariance and perform a KL transform to find its eigenvectors. This allows the pattern of a certain window size to be

optimally represented (in the least squared sense) under a wide sense stationary assumption as the sum of a small number of eigenvectors with the largest eigenvalues. Potential disadvantages of KL feature extraction are: it relies on fixed windows, with the specified size affecting the results; it assumes stationary samples; it finds the best features for representing the clutter, but not necessarily the best for discriminating between the pattern and target; computing eigenvectors is slow [$O(N^3)$ for singular value decomposition of an $N \times N$ matrix]. In addition, the eigenvector associated with the largest eigenvalue normally corresponds to the low frequency components including d.c. term and is not best suited for target detection; it may not be a meaningful space-scale feature. Wavelets do provide such features. Wavelet-based feature extraction avoids these disadvantages: there is no fixed window width, since the wavelets are used at different scales; it does not assume stationarity; optimal feature for target detection rather than representation can be chosen; fast transforms exist [$O(N)$ for orthogonal and bi-orthogonal discrete wavelets (Mallat, 1989) and $O(N \log N)$ for continuous wavelets (Rioul & Duhamel, 1992)]. We have developed new methods (Szu et al., 1992; Telfer et al., 1994; Szu & Telfer, 1994) for adaptive wavelet feature extraction that have yielded better results than other common features in an acoustic detection application (Telfer et al., 1994).

The CWT is easier for adapting its kernel for feature extraction, and the DWT is faster in computation for multiresolution data compression. Since it is concise to introduce the CWT from the continuous FT viewpoint, we define the CWT below and give the DWT later in Section 4.

Signal analyses have used a windowed or short-time FT

$$G(\omega|t_0) \equiv \int_{-\infty}^{\infty} dt g(t) \exp(-j\omega t) w(t|t_0), \quad (1)$$

where the window function $w(t|t_0)$ has a fixed width for all frequencies and if it is Gaussian, eqn (1) is a Gabor transform. On the contrary, the WT utilizes a wavelet kernel having compact support of which the time width is a Δt reduced at smaller scale $a < 1$ which implies a higher frequency f/a and consequently higher bandwidth $\Delta f/a$ according to the constant fidelity $Q_0 = f/\Delta f$, in order to localize the sharp image edge better in Figure 1. Figure 1 shows a comparison between Gabor transform and Morlet wavelet transform for an ideal rooftop/ridge function.

While a real-world signal is often localized and transient, noise is present everywhere at all frequencies. Consequently, pushing the frequency limit to

Wavelet Transform vs. Gabor Transform

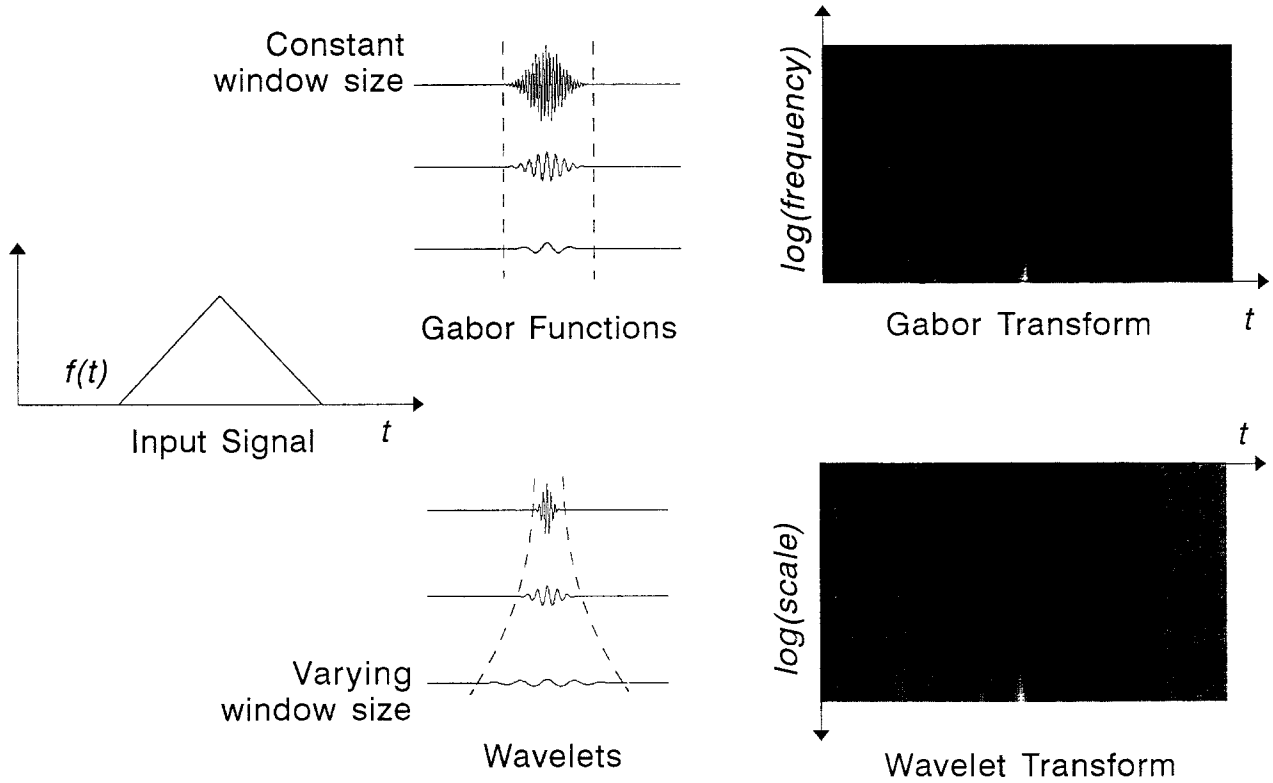


FIGURE 1. Continuous WT: Comparison between Gabor transform and Morlet wavelet transform for an ideal rooftop function on a flat floor without any noise. It shows the wavelet requirement of a constant fidelity $Q_0 = f/\Delta f$ which together with the Fourier reciprocity theorem for a group wave $\Delta f \Delta t \geq 1$ implies a high frequency by a factor $2f$ which must be associated with twice the bandwidth $2\Delta f$, and consequently the half the window time width $\Delta t \geq 1/2\Delta f$. The visible separation of three singularities of the rooftop function on a floor. A short-time Fourier transform that has a fixed Gaussian window is called the Gabor transform. Morlet wavelet transform is $h(t) = \exp(i\omega_0 t), \exp(-t^2)$ better in locating a roof-top seismic image intensity discontinuity.

achieve higher frequency or shorter wavelength resolution results in a poor signal-to-noise (SNR) ratio because the usual decrease in signal's energy at higher frequencies is compounded with the fixed window that admits an identical amount of (white) noise energy. This drawback was not discovered until the French seismic imaging exploration in the last decade (Combes et al., 1989), for which Morlet (Combes et al., 1989) proposed a wavelet kernel with $w_0 \geq 5$

$$h(t) = \exp(-j\omega_0 t) \exp(-t^2/2) = \exp(j5t) \exp(-t^2/64) \quad (2)$$

to achieve an approximately zero integrated area

$$\int dt \exp(-j(\omega = 0)t) h(t) = H(\omega = 0) = 0$$

i.e., no d.c. component, in order to nullify any uniform part of signal echo. The numerical value given in eqn(2) is used in Section 3 to extract the formant over a speech fields of 4096 data points.

Instead of increasing the frequency, the time was relatively shrunken by the affine (scale a , shift b) transform.

$$t \rightarrow t' = (t - b)/a \quad (3)$$

which can generate the complete basis functions

$$h(t) \rightarrow h(t')/\sqrt{a} \equiv h((t - b)/a)/\sqrt{a} \equiv h_{ab}(t). \quad (4)$$

The WT is defined similarly to the FT

$$\begin{aligned} FT\{g(t)\} &= \int_{-\infty}^{\infty} dt \exp(j\omega t) g(t) \equiv \langle \exp(j\omega t), g(t) \rangle \\ G(a, b) &= WT\{g(t)\} = \int_{-\infty}^{\infty} dt h_{ab}(t) g(t) \equiv \langle h_{ab}(t), g(t) \rangle, \end{aligned} \quad (5)$$

where the inner product angular brackets \langle, \rangle are defined with the complex conjugate*. Similarly, the inverse WT is defined as

$$\begin{aligned}
g(t) &= FT^{-1}\{FT\{g(t')\}\} \\
&= (1/2\pi) \int_{-\infty}^{\infty} d\omega \exp(jt) \langle \exp(j\omega t'), g(t') \rangle \\
g(t) &= WT^{-1}\{G(a, b)\} = (1/c) \int da/a^2 \int db G(a, b) h_{ab}(t) \\
&\equiv \langle G(a, b), h_{ab}(t) \rangle
\end{aligned} \tag{6}$$

in terms of 2-D Haar measure $[db da/a^2]$ which is equivalent to the dimensionless time-frequency product $[dt d\omega]$ where $dt \approx -db$, and $d\omega \approx d(1/a) = -da/a^2$.

Substituting eqn (5) in (6) gives completeness eqn (7) mapping $g^*(t')$ identically to $g(t)$. The kernel function satisfies the completeness condition in the time domain

$$\int da/a^2 \int db h_{ab}(t)^* h_{ab}(t') = c \delta(t - t') \tag{7}$$

or rewritten for either even or analytical Fourier amplitude in the Fourier domain:

$$\int_{-\infty}^{\infty} d\omega |H(\omega)|^2 / |\omega| = c < \infty, \tag{8}$$

where use is made of FT of a daughter wavelet, eqn (4)

$$H_{ab}(\omega) = \sqrt{(a)} H(\omega) \exp(j\omega b)$$

which is related to the FT of the mother wavelet. Note that the instantaneous frequency can be defined by taking the logarithmic derivative of $H_{ab}(2)$ with respect to the shift parameter b

$$j\omega = d(\log H_{ab}(\omega))/db$$

which is the basic formula for Maes and Daubechies (1995) to construct noisy speaker ID problem by the CWT modified Cepstrum approach. Thus, the admissible kernel $h(t)$ should avoid the division-by-zero divergence in eqn (6) implying d.c. $H(0)=0$ [with an oscillation decay faster than t^{-1} as proved by Daubechies (1993)]. The kernel $h(t)$ has been called a mother wavelet by Morlet, Grossmann, Meyer, Mallat, Coifman and Chui.

3. COMBINE ANN WITH WT—WAVENETS

Such wavelet features have been shown to be attractive for target detection. Wavelet features chosen to match an object or parts of an object, input to a multilayer perceptron, have been shown to improve detection accuracy in multispectral electro-optical images and in sidescan sonar images (Telfer et

al., 1994). A more powerful approach is to adaptively compute wavelet features in conjunction with the neural network to maximize detection accuracy. This idea has been successfully demonstrated for acoustic backscatter (Telfer et al., 1994) and infrared imagery (Casasent & Smokelin, 1994) detection applications.

In recognition, wavelet feature orthogonality is not as important as class separation. Since classifiers normally require relatively lengthy off-line training, adapting wavelet features together with the classifier during training is an attractive approach to minimize the misclassification rate. Given an unknown signal with an additive white noise

$$s(t) \rightarrow s'(t) = s(\alpha t)$$

which has an arbitrary α -scale change, the corresponding wavelet coefficient

$$S(a, b) \rightarrow S'(a, b) = S(\alpha a, \alpha b)$$

changes likewise along the radial direction of the (a, b) plane. A wedge-shaped filter in the a, b domain was used to collect CWT coefficients with each wedge, which were then passed to the ANN. This is a model of the neural network “optical cochlea” (Szu et al., 1992; Zhang & Benvensite, 1992; Pati & Krishnaprasad, 1993) for scale-invariant classification, as shown in Figure 2 for a real-time implementation.

Moreover, we wish to match signal class better by adaptively constructing a superposition of admissible mothers called a super-mother wavelet. In so doing the number of terms needed to represent the signals is greatly reduced and each becomes more robust in the sense of the redundancy and stable under perturbation. We state the Adaptive WT theorem of Szu and Telfer (1994) that the *linear superposition of admissible mothers is also admissible*. This theorem allows us to use both the noisy exemplars and the top-down performance measure, similar to a supervised training strategy used in artificial neural networks to construct a supermother for each phoneme Szu et al. (1992) and Zhang and Benveniste (1992) and Pati and Krishnaprasad (1993). One of our architectures is similar to a radial base function approach, except each base function has different scale size and shift to be determined from the supervised training. Conjugate gradient descent in three variables, the scale, the shift, and the synaptic weight are implemented to produce such “super-mothers” representing vowels: “a”, “i” and “e”, etc.

Lippman et al. (1993) have demonstrated using a multilayer perceptron, that the variations of the first and second formants are needed to classify ten different phonemes, such as “head, hod, heed, etc.” This fact suggests using the Morlet wavelet to extract

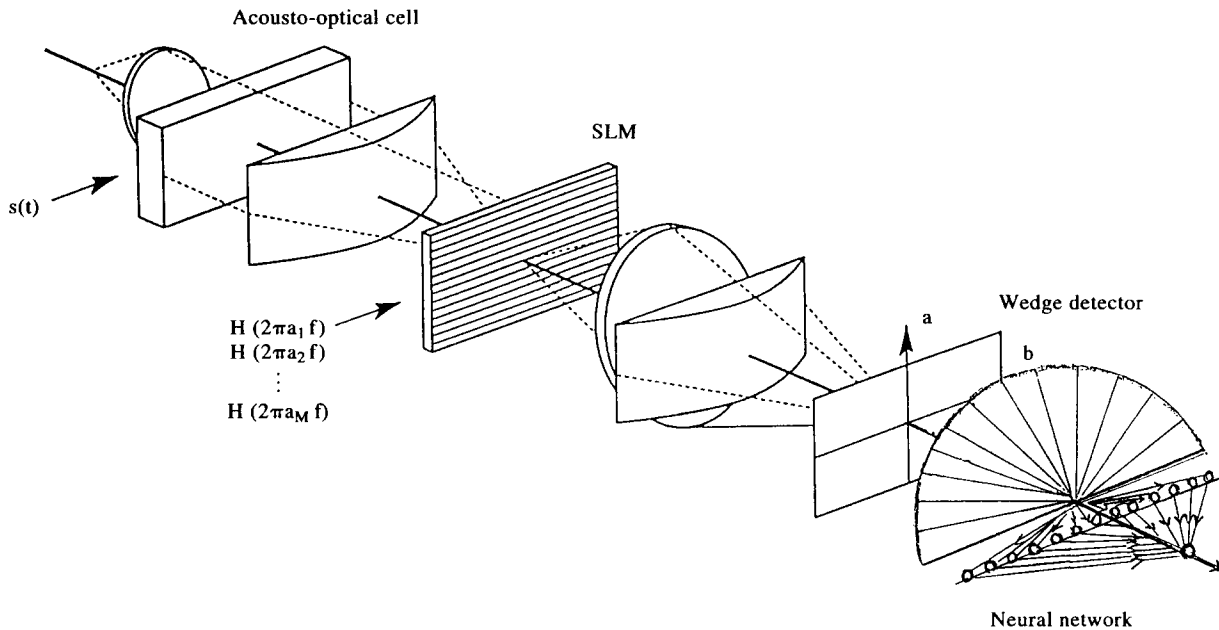


FIGURE 2. Real-time scale-invariant WT-ANN signal “optical cochlea” classifies. The sound signal $s(t)$ enters as an electrical signal into a acousto-optical cell, it generates an ultrasound inside the crystal, and it modulates the dielectric constant that affects the coherent plane wave passing through it. The cylindrical lens produces a 1-D FT to convert the signal to the Fourier domain $S(f)$, which is multiplied by several daughter wavelets scaled from a_1 to a_M coded in the spatial light modulator (SLM). Then, by the Fourier deconvolution theorem, the inverse FT converts a bank of correlations of all wavelets to the output domain scale parameter a versus the time-delay shift b parameter. Then, the theorem states that any scale change of the signal will be translated radially within its wedge.

from the speech local frequency at a constant fidelity Q . We hope to extrapolate the speaker variation of an ordinary sound, e.g., “the seat”, from those spoken by a male versus a female speaker. For this example, we found several formants are prominent for the phoneme “e” and thus constructed these formant wavelets correspondingly. This set of formant wavelets can form a bank of matched filters (Figure 3) that can be correlated in parallel with a speech signal, and different weights of each formant can have different importance involving its identification. This weight set could be obtained by a supervised training ANN over multiple speakers. Then, this bank can then be used to segment vowels from the continuous speech with different emphases on each formant for different speakers. All together, these correlation peaks would identify and single out the phonemes involved as shown in Figure 3 on the right-hand side. Different dilated versions of the supermothers could be used to identify speech at different speeds. This would serve as the front end of a wavelet neural network classifier, e.g., given a male and female speaker saying “the seat” as shown in the first column of Figure 3: Phoneme spotting by supermothers. This could be done by a single correlation of the input speech with a bank of supermothers (and column of Figure 3) which only kept those matched phonemes irrespective of a male or female speaker (in third column), and consequently achieving a feature-preserving data compression.

An interesting example is the cocktail party effect in which one can hear one’s own name no matter at what pitch it is whispered during a noisy drinking party. In that case, one’s name is a wideband transient mother wavelet $h(t)$. Since the daughters called the “wave-ons” are from a set like radial basis functions but related to one another by the affine transform, they can match in parallel the input sound, the name plus noise, with specific daughter neurons. When this happens a winner-take-all can be used to suppress all the other weak SNR contributions from other daughter neurons. This winner-take-all ANN approach has been implemented optically using electron trapping material (Athale et al., 1993). It eliminates other less-matched daughter neurons having more noise contribution than a signal, and thus the wavenet enhances the relative SNR, explaining the cocktail party effect as conjectured (Szu et al., 1992). We wish to comment that Maes and Daubechies (1995) have constructed a noisy speaker ID by the CWT modified Cepstrum approach, based on Section 2 the logarithmic derivative of daughter wavelet delay parameter b , of which an ANN post-processing would be a natural extension.

4. DISCRETE WT FROM DIGITAL FILTER BANK VIEWPOINT

Wavelets form a complete, orthogonal and normalized (CON) coordinate system satisfying a scaling

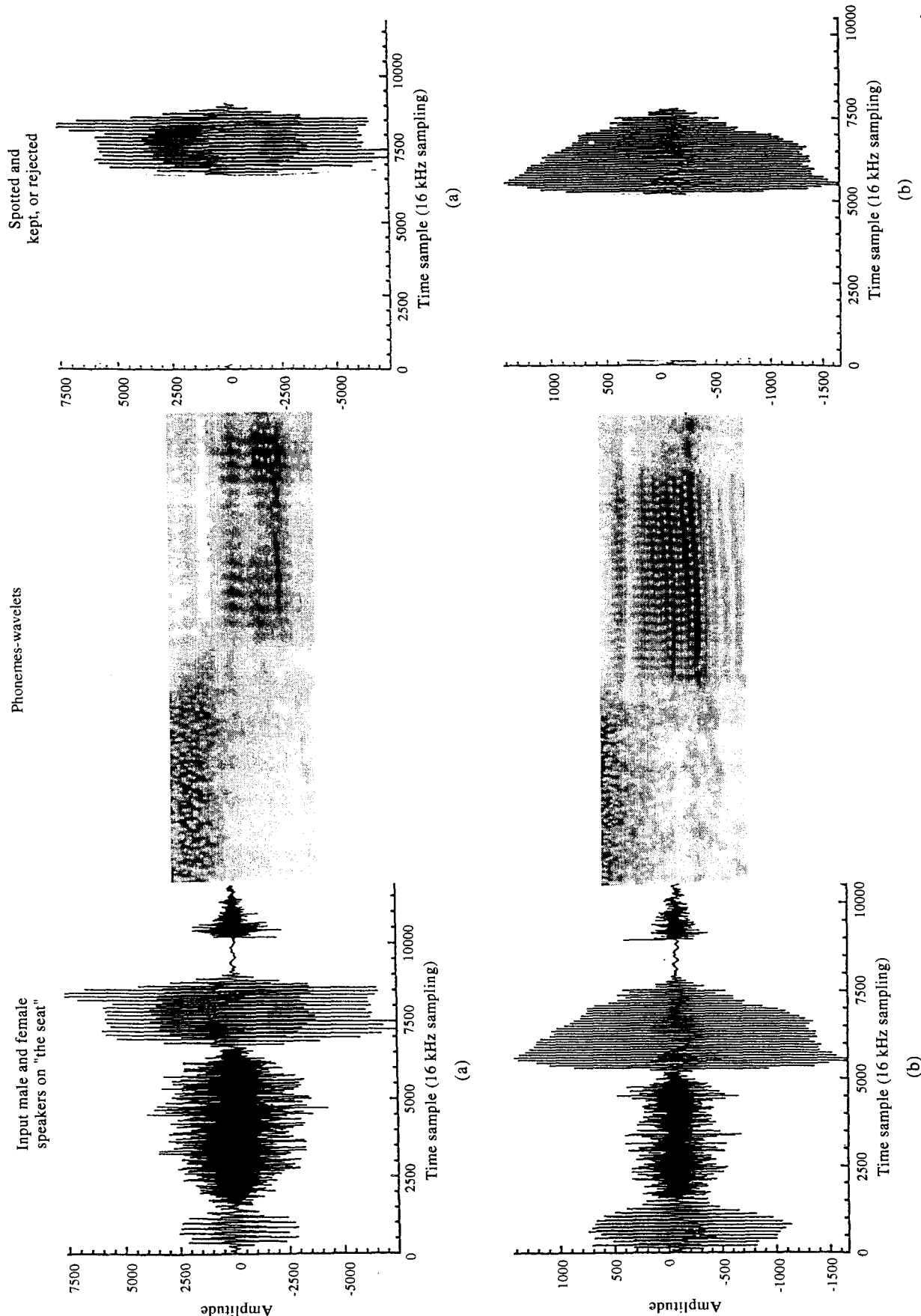


FIGURE 3. Continuous WT compression and recognition application. Phoneme spotting among various speakers. Continuous speech spoken by a male and female speaker represents the speaker variation. Example "the seat" is shown on the left-hand side. The CWT of the both male and female speeches is performed by a scaled bank of Morlet wavelet transform $h(t) = \exp(-t^2) \exp(-t^2)$ where $t \rightarrow t' = (t, 2t, 4t)$. The result is plotted, in the middle against the time axis showing the distinctive "the" and "s" high frequency sounds on the left-hand side and follows the important "e" vowel which is characterized with several formants in both male and female speakers. Thus, we construct three wavelets corresponding to three formants. The decision by means of an ANN can determine the relative importance of this training set when projected onto the formant wavelets. Likewise, the supervised training result is used to test the phoneme spotting among other less meaningful sounds so that we can keep the original sound with its original pitch and time ordering by simply deleting the other sound as shown on the right-hand side.

relationship, as suggested by Haar similar to the 45° rotation of the Cartesian coordinates as follows. The simplest orthogonal coordinates are the well-known Cartesian coordinates, e.g.

$$x \equiv (1, 0, 0, 0), y \equiv (0, 1, 0, 0), z \equiv (0, 0, 1, 0), \\ \text{and } z' \equiv (0, 0, 0, 1)$$

because the inner product of any pair of coordinate axes will be zero (by component $0 \times 1 = 0$), implying the orthogonality. Likewise, Haar chose the even-odd orthogonality, namely an even function

$$\phi(t) = (+, +, +, +)$$

pointing in the first orthant of four positive axes, and an odd function

$$\Psi(t) = (+, +, -1, -1)$$

pointing in the lower orthant of positive x , y , and negative z , and z' . Their inner product gives the result of zero implying the orthogonality in four dimensions (4-D). Furthermore, this property repeats itself in a half interval, if $\psi(2t) = (+, -, 0, 0)$ is chosen pointing in the plane bounded by $x \geq 0$ and $y \leq 0$, and $\psi(2t-1) = (0, 0, +, -)$ pointing in the plane bounded by $z \geq 0$ and $z' \leq 0$. For example, any 4-D vector, e.g., a ramp \mathbf{g} , can be identically expanded in Cartesian or Haar coordinates, because both are complete

$$\mathbf{g} \equiv \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} ++ +0 \\ ++ -0 \\ +- +0 \\ +- -0 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \\ -1 \\ -1 \end{bmatrix} \equiv [H4]\mathbf{G}, \quad (9)$$

where the expansion coefficient \mathbf{G} is easily computed because of the orthogonal property, i.e., the inverse matrix is the transpose matrix [superscript T indicates the matrix transpose operation if use is made of the basis normalization (4, 4, 2, 2)].

$$\mathbf{G} \equiv \{5, -2, -1, -1\}^T = [H4]^{-1} \mathbf{g} = [H4]^T \mathbf{g} \quad (10)$$

which is formally identical to the expansion of CWT: $G(a, b) = \langle h_{ab}(t), g(t) \rangle$.

The $O(N)$ computational complexity is due to the matrix reduction (Strang, 1994).

$$[H4] = \begin{bmatrix} [H2] & 0 & 0 \\ & 0 & 0 \\ 0 & 0 & [H2] \\ 0 & 0 & \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} [H2] & 0 & 0 \\ & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

in terms of three reduced matrices

$$[H2] = \begin{bmatrix} + & + \\ + & - \end{bmatrix}. \quad (12)$$

Similarly, $[H8]$ is reduced to $[H4]$. Zero entries reduce the dense matrix of $O(N^2)$ to $O(N)$ for large N data. Equation (12) reveals also a base-2 sub-band nature as below.

The lowpass L and highpass H filters define the scaling function ϕ and wavelet ψ

$$L\{c_k\} = \{+, +\} \leftrightarrow \phi(t) = \phi(2t) + \phi(2t-1) \\ = \sum_k c_k \phi(2t-k) \quad (13)$$

$$H\{d_k\} = \{+, -\} \leftrightarrow \psi(t) = \phi(2t) - \phi(2t-1) \\ = \sum_k d_k \phi(2t-k). \quad (14)$$

The scaling function eqn (13) of Haar is satisfied by $\text{rect}(t) = \text{rect}(2t) + \text{rect}(2t-1)$ which has $\{1, 1\}$ coefficients that are equivalent to averaging two neighborhood pixels via the lowpass filter L , eqn (13). Likewise, a Haar wavelet $H = \{+, -\}$, eqn (14) is a highpass filter in differencing two neighborhood pixels. Given an arbitrary input signal, say $\{1, 1\}$ fed in parallel to lowpass $L = \{+, +\}$, and to highpass $H = \{+, -\}$, then the convolution product gives $\{1, 1\} \times \{+, +\} = \{1, 2, 1\}$ and $\{1, 1\} \times \{+, -\} = \{1, 0, -1\}$. Because of the smoothing, a smaller number of pixels is needed, and a down sampling decimates every other pixel in both channels giving $\{1, 2, 1\} \rightarrow \{1, 1\}$; $\{1, 0, -1\} \rightarrow \{1, -1\}$. The receiver for reconstruction takes the up sampling by interpolation, fill-in zeros giving $\{1, 1\} \rightarrow \{1, 0, 1\}$; $\{1, -1\} \rightarrow \{1, 0, -1\}$. Since Haar synthesis filters $L^+ = \{-, -\} = -1 \times \{1, 1\}$ and $H^+ = \{+, -\}$ are within an overall constant identical to those of analysis filters, the reconstruction gives $\{1, 0, 1\} \times \{-, -\} = \{-1, -1, -1, -1\}$ and $\{1, 0, -1\} \times \{+, -\} = \{1, -1, -1, 1\}$, which are added: $\{-1, -1, -1, -1\} + \{1, -1, -1, 1\} = \{-2, -2\} = \{1, 1\} \times (-2)$ to reproduce the original $\{1, 1\}$ within a factor. Without the down and up sampling, the system remains lossless, as easily verified with the above example. Such a DWT via the sub-band coding was known in digital filter theory as lossless quadrature mirror filters coined for the mirror symmetry with respect to $2\pi/4$. Only the highpass outputs called detail signals are sent through the channel, the lowpass outputs called reference signals are usually fed back to the input to produce the next level highpass detail signal to be sent through recursively shown in Figure 2. Since the down sampling by two is at every level of recursion, the net number of pixels remains the same $N/2 + N/4 + \dots = N$. Otherwise, the amount of

TABLE 1
Analysis/Decomposition and Synthesis/Reconstruction Filters

System	L	H	L^+	H^+
Haar, H_2	1, 1	1, -1	-1, -1	1, -1
Hat, H_3	1, 2, 1	1, 2, -6, 2, 1	-1, 2, 6, 2, -1	1, -2, 1
Spline, S_4	1, α , α , 1	1, α , $-\alpha$, -1	-1, α , α , -1	1, $-\alpha$, α , -1
Spline, $S_{9,7}$	0.04, -0.02, -0.1, 0.4, 0.9, 0.4, -0.1, -0.02, 0.04; $-1 \times \text{even } L$; -0.06, -0.04, 0.8, 0.4, -0.04, -0.06; $-1 \times \text{odd } H$			
Daubechies $4 \times D_4$	$1 + \sqrt{3}$, $3 + \sqrt{3}$, $3 - \sqrt{3}$, $1 - \sqrt{3}$; $1 - \sqrt{3}$, $-3 + \sqrt{3}$, $3 + \sqrt{3}$, $-1 - \sqrt{3}$; identical			
Laplacian pyramid	$1 + 2a$, 1, $4a$, 1, $1 - 2a$; difference image;			identical

data to be stored in the original (Laplacian pyramid) algorithm would prohibitively increase. Such a down sampling, however, sensitizes the quantization or alignment errors. Daubechies (1992) was the first to go beyond Haar's system, by assuming analysis filters

L and H equal to synthesis filters L^+ and H^+ , and moreover made the simplification *Ansatz*

$$L = (c_0, c_1, c_2, c_3); H = (c_3, -c_2, c_1, -c_0) \quad (15)$$

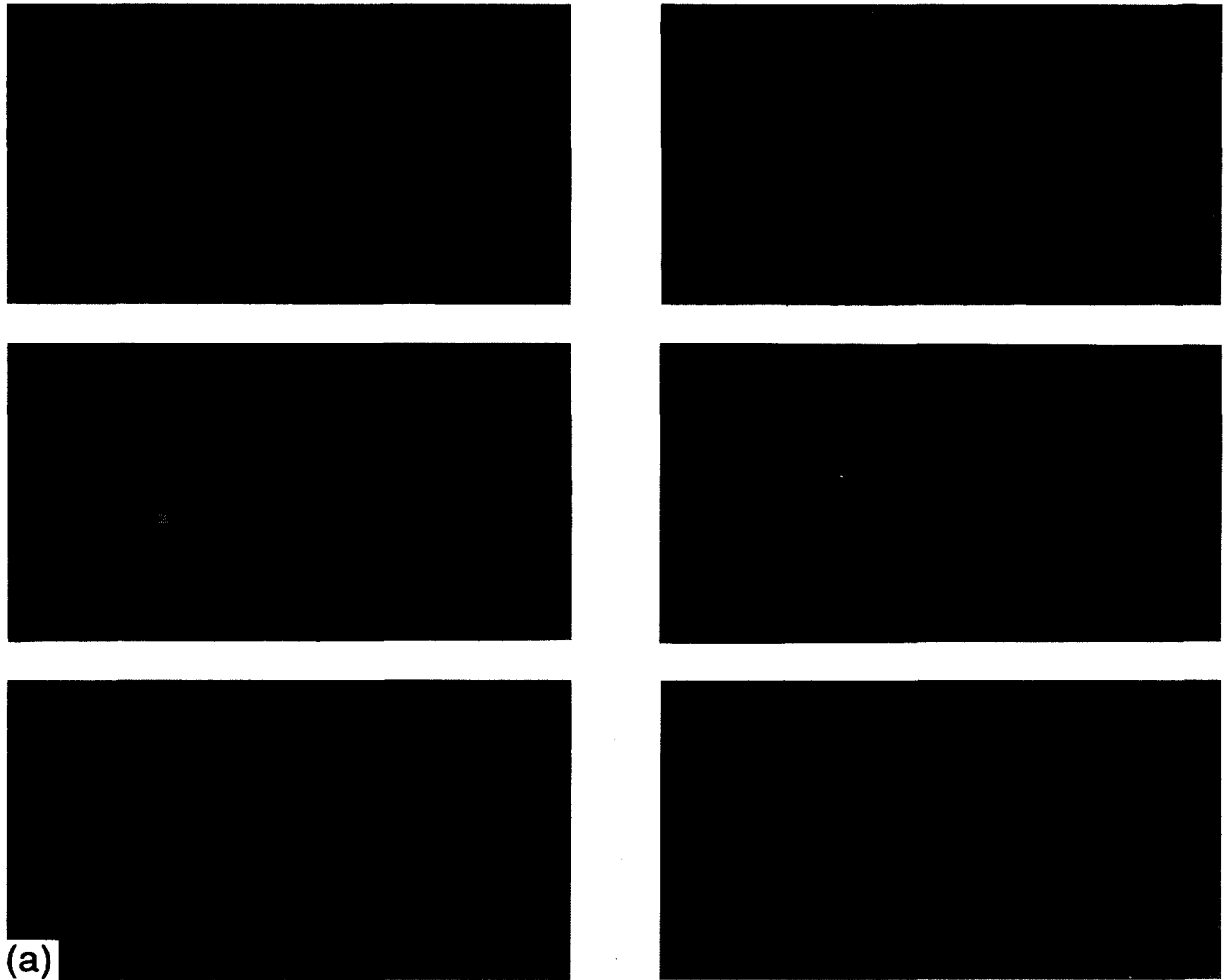


FIGURE 4. Discrete WT data compression and recognition application. Given the immense throughput rate of the 6 band multispectral infrared imagery shown in (a), the goal is to detect in real time the blob-like man-made objects on a beach. A single frame is given in (b) and the application of biorthogonal subband coding using $S_{9,7}$ (given in Table 1) in four levels iterative and decimation is shown in (c). The DWT is based on a biorthogonal subband coding system that has the standard FBI filter bank design: seven and nine coefficients for the respective low-pass and high-pass filters. The compression ratio of 59:1 is accomplished due to the nonuniform quantization to a few bits depending on the repeated filtering and decimation in half over four levels shown in (d). The reconstructed image is shown in (e). We have simultaneously achieved infrared sensor non-uniformity compensation, sensor denoising and edge-preserving data compression, as demonstrated by the frame differencing between (b) and (e) shown in (f). When the result of decompression via the inverse DWT synthesis is passed through a CWT feature extractor and then followed by an ANN APR, we obtain a good result shown in the right-hand side of (g) in contrast to the identical ANN APR without the CWT wavelet preprocessing shown on the left-hand side.



FIGURE 4b.

which guarantees the orthogonality between the low and high-pass filters, and also reduced eight unknowns c 's and d 's to four unknowns c 's, so that the self-adjoint system forms a unitary matrix (called paraunitary by Vetterli and Herley, 1992)

$$\begin{vmatrix} c_0 & c_1 & c_2 & c_3 \\ c_3 & -c_2 & c_1 & -c_0 \end{vmatrix} = \begin{vmatrix} c_0 & c_3 \\ c_1 & -c_2 \\ c_2 & c_1 \\ c_3 & -c_0 \end{vmatrix} = \begin{vmatrix} \sum_i c_i^2 & 0 \\ 0 & \sum_i c_i^2 \end{vmatrix}. \quad (16)$$

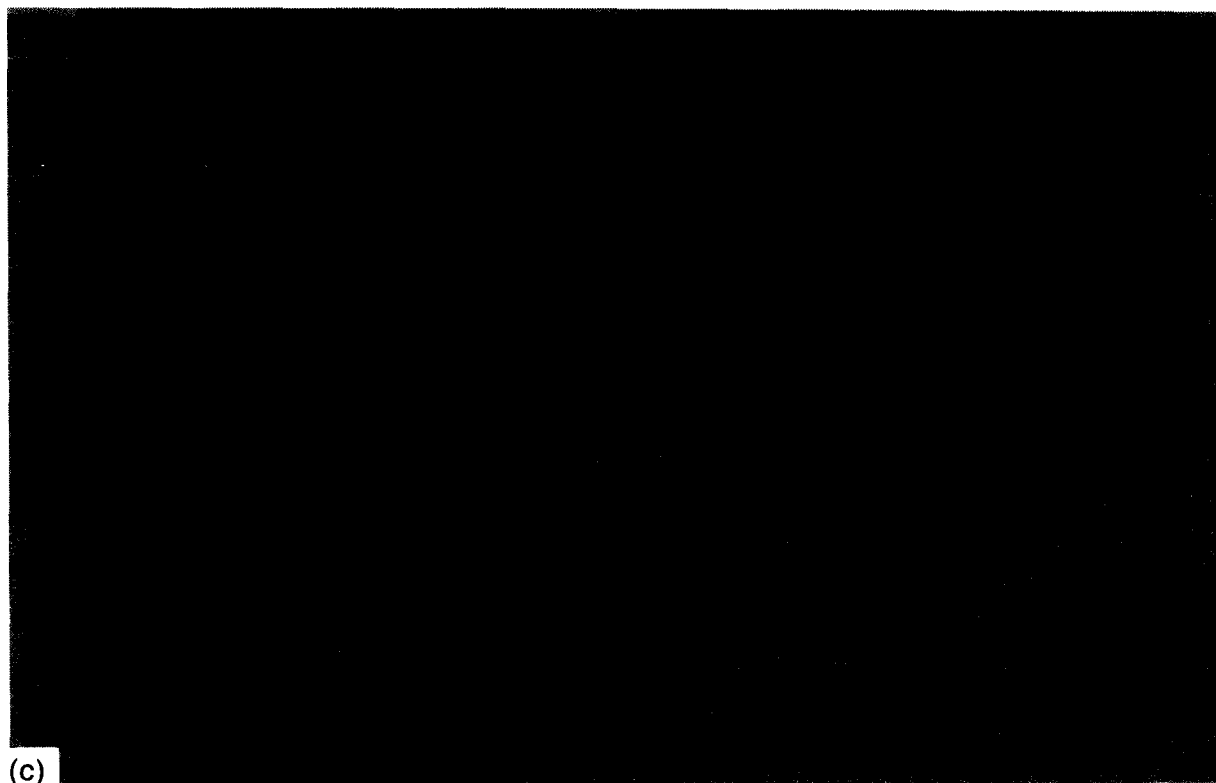


FIGURE 4c.

63	31	15	3	0
31	31			
15	7			
3		0		
0			0	

(d)

(d)

FIGURE 4d.

Only four equations are needed to determine four c 's:

(i) The low-pass scaling eqn (13) is normalized by $\int dt\phi(t) = 1$, which yields $c_1 + c_2 + c_3 + c_4 = 2$.
(ii) To extract the change, the high-pass wavelet filter H , eqn (14) should nullify a constant (1,1,1,1): $c_3 - c_2 + c_1 - c_0 = 0$, and also (iii) nullify a linear ramp (1,2,3,4): $c_3 - 2c_2 + 3c_1 - 4c_0 = 0$. (iv) Two daughter wavelets should be orthogonal in a shift by two: $(c_3, -c_2, c_1 - c_0, 0, 0)$ $(0, 0, c_3, -c_2, c_1 - c_0)^T = c_1c_3 + c_0c_2 = 0$. This set determines uniquely four filter coefficients $L = (1, 3, 3, 1)/4 + (1, 1, -1, -1)\sqrt{(3)/4}$ shown in Table 1. A contribution of WT to the filter bank theory was the discovery of the

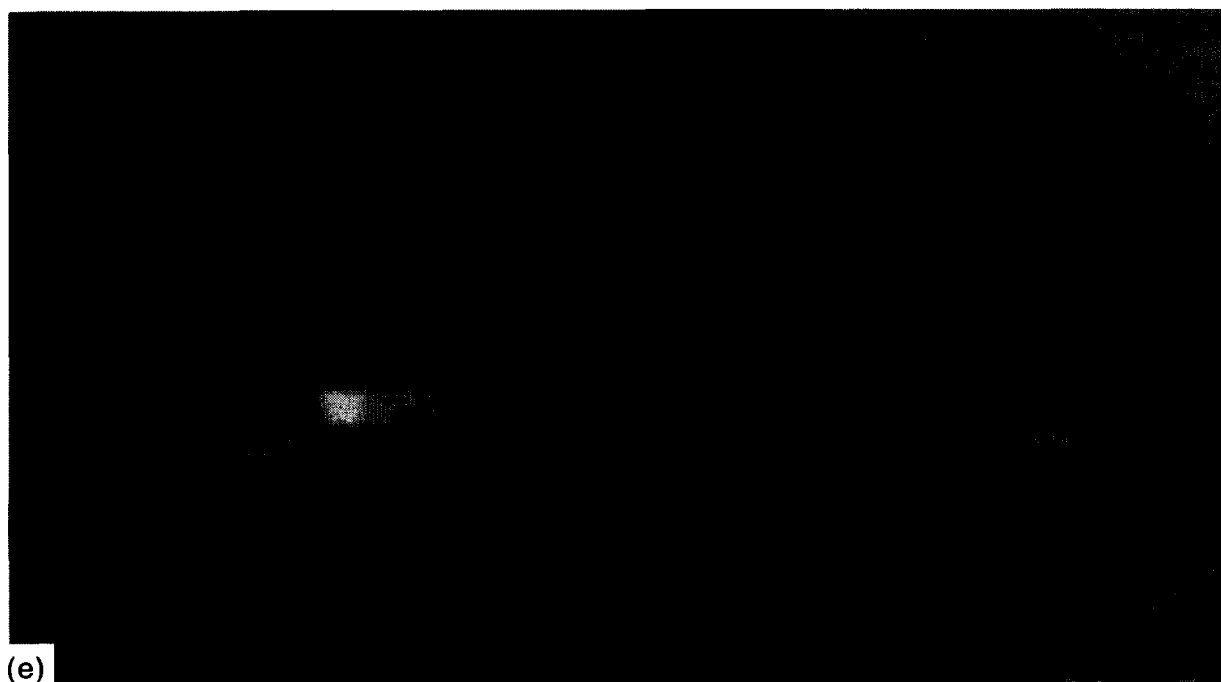
importance of the regularity condition which means that solving the scaling function $\phi(t)$ of eqn (13) by recursion should converge stably, and then a shifted version $\phi(2t - k)$ gives the wavelet eqn (14). Then, continuous WT: $\langle g(t), \psi(t) \rangle$ can be approximated by discrete WT: filter bank convolution product: $g * H$ recursively in an iterative fashion by feedback of the low-pass output to the pair of filters.

Note that for data compression application of the regularity condition is necessary for repeated usage of the filter, but not sufficient for it can still yield a unsymmetric scaling function (and sometimes even fractal-like) which is less suited for image compression. This is because for DWT the image value at a boundary is extrapolated by assuming usually a mirror-symmetric value with respect to the image boundary in order to minimize the discontinuity effect. A symmetric scaling function in four taps is still unitary in a biorthogonal system $L \neq L^+$; $H \neq H^+$ of eight coefficients (Table 1)

$$\begin{vmatrix} 1 & \alpha & \alpha & 1 \\ 1 & \alpha & -\alpha & -1 \end{vmatrix} \begin{vmatrix} -1 & 1 \\ \alpha & -\alpha \\ \alpha & \alpha \\ -1 & -1 \end{vmatrix} = \begin{vmatrix} 2\alpha^2 - 2 & 0 \\ 0 & -2\alpha^2 + 2 \end{vmatrix}.$$

5. IMAGE COMPRESSION AND ANN CLASSIFICATION

In Table 1, only the Spline $S_{9,7}$ system, whose L of nine coefficients and the adjoint L^+ of seven



(e)

FIGURE 4e.

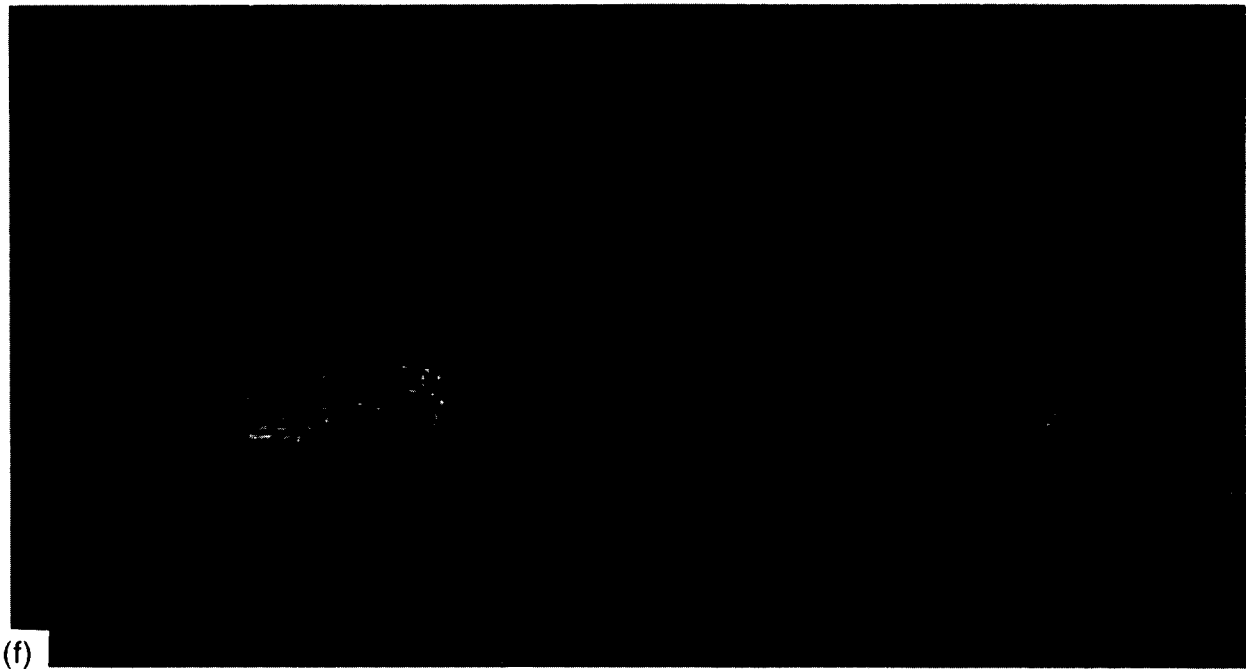


FIGURE 4f.

coefficients satisfies the regularity conditions is symmetric for easy extrapolation of mirror-symmetric boundary condition. Therefore it is suited for image compression, as demonstrated by Antonini et al. (1992). In fact, the FBI (Hopper, 1994) has developed a "lossless" 20:1 (approximate) fingerprint compression that will become an important part of a real-time nation-wide fingerprint ID system.

The synergism between (9, 7) biorthogonal FBI wavelet chips with neural chips may push the frontier of information and automation technologies. Since DWT is linear, it merely rearranges the information content. Data compression must result from non-uniform requirement of pixel dynamic range generated by successive filter bank operations. The data compression exploits the fact that a filtered image which has less detail requires fewer bbbp. In so doing the convergence of recursive filtering depends crucially on the stability of the filter under the iteration, the so-called regularity property. Since for a biorthogonal system analysis L synthesis L^+ , both must be regular. Unfortunately, the simple 4 tap L^+ is not regular. As a byproduct, one can achieve noise suppression or clutter rejection by means of selective reconstruction. In Figure 4 we have applied $S_{9,7}$ recursively to four levels to a multispectral infrared imagery and shown that when the detail output at the first level was eliminated, the non-uniform infrared sensor noise (appearing as diagonal waves in two sub-bands) has been eliminated while preserving the target features at 59:1 compression ratio. We have demonstrated that with and without wavelet feature extraction the performance of ANN classi-

fier of the mine fields is quite different (Telfer & Szu, 1993).

6. CONCLUSION

We have compared mathematically, the functionality of ears and eyes with the CWT parallel inputs to ANN to produce a redundant and yet complete multiresolution analysis input. The DWT is introduced to be CON basis used in digital filter theory as the subband coding with the added regularity condition derived from mathematical wavelet theory. Such a smart and fast preprocessing can accomplish a feature-preserving data compression that avoids the overtraining or overfitting plagued the ANN's ability of generalization and abstraction.

ANN is a democratic voting scheme among hidden layer neurons without hidden agenda. WT is a bank of constant Q matched filters, that is an efficient way to produce more correlation peaks than single matched filter peak. Then, the committee member consisting of daughter wavelets, one for each input neurons can be trained by ANN supervised learning to weight differently each neuron output voting.

Since WT is a systematic MRA scheme, and it produces local frequency content written conveniently in terms of the linear, first order time-scaling/frequency joint representation, as opposed to the second-order Wigner distribution and Woodward ambiguity function which tend to produce cross talks. More voter participation in ANN may help resolve the mini-max conflicting requirement that we wish to

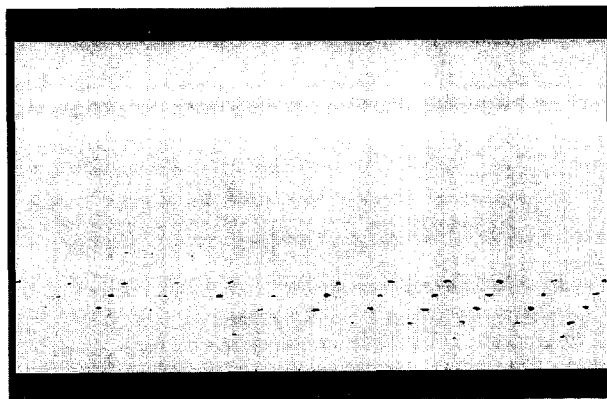
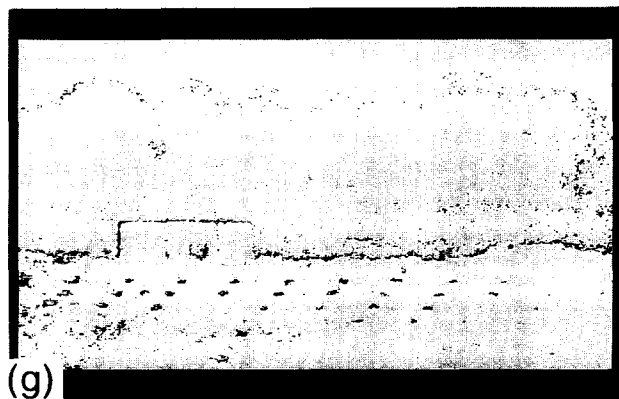


FIGURE 4g.

group alike as closely as possible and segregate the dislikes as far apart as possible (Szu, 1992).

We believe that the combination of Continuous WT and ANN is a powerful paradigm for pattern recognition, and the DWT is a fast MRA algorithm to provide different dynamic range requirement for bpp data compression accordingly. Together, we have demonstrated compression followed by recognition giving us a better SNR and an improved performance of the receiver operation characteristics of the detection probability against the false alarm rate.

REFERENCES

- Antonini, M., Barlaud, M., Mathieu, P. & Daubechies, I. (1992). Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2), 205–220.
- Bachmann, C. M., Musman, S. A., Luong, D. & Schultz, A. (1994). Unsupervised BCM projection pursuit algorithms for classification of simulated radar presentations. *Neural Networks*, 7, 709–728.
- Baum, E. B. & Haussler, D. (1989). What size net gives valid generalization? *Neural Computation*, 1, 151–160.
- Casasent, D. & Smokelin, J.-S. (1994). Neural net design of macro Gabor wavelet filters for distortion-invariant object detection in clutter. *Optical Engineering*, 33, 2264–2271.
- Combes, J. M., Grossmann, A. & Tchamitchian, P. (Eds.) (1989). *Wavelets: Time-frequency methods and phase space*. New York: Springer.
- Daubechies, I. (1992) *Ten lectures on wavelets*. Philadelphia, PA: SIAM.
- Daniell, C. E., Kemsley, D. H., Lincoln, W. P., Tackett, W. A. & Baraghimian, G. A. (1992). Artificial neural networks for automatic target recognition. *Optical Engineering*, 31, 2521–2531.
- Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematical Annals*, 69, 331–371.
- Hopper, T. (1994). Compression of grey-scale fingerprint images. In *Wavelet applications, Proceedings of SPIE*, Vol. 2242, pp. 180–187, 5–8 April, Orlando.
- Fogler, R. J., Koch, M. W., Moya, M. M., Hostetler, L. D. & Hush, D. R. (1992). Feature discovery via neural networks for object recognition in SAR imagery. *Proceedings of the International Joint Conference on Neural Networks*, Vol. IV, pp. 408–413.
- Gorman, R. P. & Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*, 1, 75–89.
- Hampshire, J. (1993). A differential theory of learning for efficient statistical pattern recognition. Ph.D. dissertation, Carnegie Mellon University, Electrical Engineering Department.
- Hampshire, J. & Waibel, A. (1990). A novel objective function for improved phoneme recognition using time delay neural networks. *IEEE Transactions on Neural Networks*, 1, 216–228.
- MIT Lincoln Laboratory Journal (1993). Special issue on automatic target recognition, Spring, Vol. 6.
- Rioul, O. & Duhamel, P. (1992). Fast algorithms for discrete and continuous wavelet transforms. *IEEE Trans. Info. Theory*, 38, 569–586.
- Roth, M. W. (1990). Survey of neural network technology for automatic target recognition. *IEEE Transactions on Neural Networks*, 1, 28–43.
- Ruck, D. W., Rogers, S. K., Kabrisky, M. & Miller, J. P. (1988). Multisensor target detection and classification. *Proceedings of SPIE*, 931, 14–21.
- Ruck, D. W., Rogers, S. K., Kabrisky, M., Oxley, M. E. & Suter, B. W. (1990). The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Transactions on Neural Networks*, 1, 296–298.
- Lippmann, R. P., Kukulich, L. & Singer, E. (1993). LNKnet: Neural network, machine-learning, and statistical software for pattern classification. *The Lincoln Lab Journal*, 6, 249–268.
- Mallat, S. (1989). A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11, 674–693.
- Mallat, S. & Zhang, Z. (1993). Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*.
- Pati, Y. C. & Krishnaprasad, P. S. (1993). Analysis and synthesis of feedforward neural networks using discrete affine wavelet transforms. *IEEE Transactions on Neural Networks*, 4, 73–85.
- Strang, G. (1994) Wavelets. *American Scientist*, 82, 150–255.
- Sheng, Y., Roberge, D., Szu, H. & Lu, T. (1993). Optical wavelet matched filters for shift-invariant pattern recognition. *Optics Letters*, 18(2), 209–301.
- Szu, H., Telfer, B. & Kadambe, S. (1992). Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31(9), 1907–1916.
- Szu, H. & Telfer, B. (1994). Mathematics of adaptive wavelet transforms: Relating continuous with discrete transforms. *Optical Engineering*, 33(7), 2111–2124.
- Szu, H., Hsu, C., Thaker, P. & Zaghloul, M. (1994). Image wavelet

- transforms implemented by discrete wavelet chips. *Optical Engineering*, **33**(7), 2310–2325.
- Szu, H., Sheng, Y. & Chen, J. (1992). Wavelet transform as a bank of matched filters. *Applied Optics*, **31**(6), 3267–3277.
- Szu, H., Yang, X.-Y., Telfer, B. & Sheng, Y. (1993). Neural network and wavelet transform for scale-invariant data classification. *Physical Reviews E*, **48**(2), 1497–1501.
- Szu, H. & Hartley, R. (1987a). Fast simulated annealing. *Physical Letters A*, **122**, 157–162.
- Szu, H. & Hartley, R. (1987b). Nonconvex optimization by fast simulated annealing. *Proceedings of the IEEE*, **75**, 1538–1540.
- Szu, H. & Messner, R. (1986). Adaptive invariant novelty filters. *Proceedings of the IEEE*, **74**, 519–520.
- Sadjadi, F. (1991). Automatic object recognition: Critical issues and current approaches. *Proceedings of SPIE*, **1471**, 303–313.
- Telfer, B. & Szu, H. (1994). Energy functions for minimizing misclassification error with minimum-complexity networks. *Neural Networks*, **7**, 809–818.
- Telfer, B., Szu, H., Dobeck, G., Garcia, J., Ko, H., Dubey, A. & Witherspoon, N. (1994). Adaptive wavelet classification of acoustic backscatter and imagery. *Optical Engineering*, **33**, 2192–2203.
- Thoet, W. A., Rainey, T. G., Brettle, D. W., Slutz, L. A. & Weingard, F. S. (1992). ANVIL neural network program for three-dimensional automatic target recognition. *Optical Engineering*, **31**, 2532–2539.
- Zhang, Q. & Benveniste, A. (1992). Wavelet networks. *IEEE Transactions on Neural Networks*, **3**, 889–898.
- Vetterli, M. & Herley, C. (1992). Wavelets and filter bank: theory and design. *IEEE Transactions on Signal Processing*, **40**(9), 2207–2232.
- Maes, S. & Daubechies, I. (1995). Private communication for SPIE Wavelet Applications Conference (Chair Szu) April 17–20, Orlando.