Neural network and wavelet transform for scale-invariant data classification

Harold H. Szu
Naval Surface Warfare Center, Dahlgren Division Code R44, Silver Spring, Maryland 20903-5000

Xiang-Yang Yang
Quantex Corporation, 2 Research Court, Rockville, Maryland 20805

Brian A. Telfer
Naval Surface Warfare Center, Dahlgren Division Code R44, Silver Spring, Maryland 20903-5000

Yunlong Sheng
Department of Physics, University of Laval, Ste-Foy, Quebec, Canada G1K 7P4

(Received 19 February 1993; revised manuscript received 16 April 1993)

Given an astrophysical observation with an arbitrary carrier frequency and an unknown scale under an additive white noise, $s'(t) \equiv s(\alpha t) + n(t)$, its wavelet transform is $W'(a,b) \equiv (s'(t), h_{ab}(t))$, as computed by the inner product with a daughter wavelet $h_{ab}(t) \equiv h((t-b)/a)/a$. $W'(a,b)$ equals the original transform $W(a,b) \equiv (s(t), h_{ab}(t))$ displaced along the radial direction $W'(a,b) = W(\alpha a, \alpha b)$ plus noise in the time-scale joint-representation plane. A bank of wedge-shaped detectors collects those displaced transforms $W'(a,b)$ to create a set of invariant features. These features are fed into a two-layer feed-forward artificial neural network, to interpolate discrete sampling, as demonstrated successfully for real-time-signal automatic classification. Useful wavelet applications in turbulence onset, spectrum analyses, fractal aggregates, and bubble-chamber particle-track pattern-recognition problems are indicated but are modeled, in the interest of simplicity, in a one-dimensional example.

PACS number(s): 87.10.+e, 89.90.+n, 85.60.−q

I. INTRODUCTION

There are many interests in applying neural-network technology to experimental measurements as an automation tool, e.g., to analyze the astrophysical observation data of a black hole, a red giant, any evidence of extraterrestrial life, and the gigabit data rate anticipated from the future superconducting-supercollider experiments. Similarly, the usefulness of the wavelet transform (WT) to replace the Fourier transform (FT) has been demonstrated in data compression as well. This paper contributes toward solving a generic data processing problem based on the synergism of both technologies. One of the challenges in experimental diagnoses is to achieve distortion-invariant classification. Traditional approaches are based on time-frequency joint representations (TFJR) such as the Wigner distribution based on the quantum-mechanical uncertainty principle and the Woodward ambiguity function based on the Doppler-shift uncertainty. Unfortunately, both involve a second-order convolution and correlation integral of quadratic order, such that multiple pulses produce a double amount of spectrum pulses that complicate the identification task for both TFJR's [1]. Recently, a first-order time-scale joint representation (TSJR), called the wavelet transform [2–4], was developed to replace the traditional FT by computing the TSJR for noisy wideband transients. Fractal-aggregate [5] and turbulence data [6,7] have been analyzed by WT, giving efficient interpretations. Motivated by the high-definition-television (HDTV), multimedia satellite communication [B-ISDN (Broadband-Integrated Services Digital Network)], as well as the need for lossless fingerprint compression to be used in the National Crime Information Center, the wavelet transform is being widely adopted for signal and image processing. Moreover, such a WT-reduced set of data may be called a wavelet-feature set, which can be used to train and implement electronically a smaller size of the artificial neural network (ANN) for pattern classification [8]. Thus we begin with the comparison between the WT and the FT as follows.

II. SCALE-INVARIANT WAVELET TRANSFORMS

The WT generalizes the one-dimensional (1D) Fourier basis, $e_f(t) = \exp(2\pi i f t)$ and harmonics, to a wideband transient 2D basis, generated by an affine group of scale and shift operations,

$$h_{ab}(t) \equiv h((t-b)/a)/a \ . \tag{1}$$

Usually, digital implementation of the WT uses a discrete basis, $a/a_0 = 2^{\pm I}$ and $b/b_0 = \pm I$ (integer I) for a constant resolution for different scales. In Eq. (1), the normalization constant $1/a$ is preferred [9] for the subsequent scale-invariant applications. The admissible condition of a basis is that a square-integrable kernel $h(t)$ must have zero constant component and a sufficiently fast decay at high frequencies [10]. The WT is defined in the square-integrable Hilbert space:

$$\mathcal{T}_{ab}^{\mathrm{WT}}\{s(t)\}\equiv(h_{ab}(t),s(t))$$
$$=\int_{-\infty}^{+\infty}h_{ab}^{*}(t)s(t)dt\equiv W(a,b)\ . \tag{2}$$

The FT is known to be an angle-preserving, or conformal, mapping, which has suggested the design of wedge-ring detectors in the Fourier plane [11]. This device works for scale-rotation invariance based on shift invariance because, despite the motion of the object, the square-law detectors guarantee the object centroid alignment at the origin of the Fourier domain. This shift invariance is based on the modulus-invariant phase information in the following straightforward mathematics:

$$|\mathcal{T}_{f}^{\mathrm{FT}}\{g(t)\}|^2=|(e_f(t),g(t))|^2$$
$$=|G(f)|^2$$
$$=|G(f)\exp(-2\pi jfb)|^2$$
$$=\left|\int_{-\infty}^{+\infty}\exp(-2\pi jft)g(t-b)dt\right|^2\ . \tag{3}$$

On the other hand, the invariant property of the WT is based on the intrinsic scale property in the TSJR domain (a,b). The idea is simple. To investigate the invariant WT is to compute the WT of various scales of the identical signal. Hopefully, those scale-related WT coefficients organize themselves in such a fashion that can be easily collected to produce scale-invariant features. Let a generic signal under additive white noise be given by

$$s_i'(t)=s_i(\alpha_i t)+n(t),\quad i=1,2,\ldots\ , \tag{4}$$

where the unknown scale α's (suppressing class index i) are equivalent to unknown frequency compaction or hopping of similar waveforms $s(t)$. The associated WT coefficient denoted by the prime is computed

$$W'(a,b)\equiv\int_{-\infty}^{+\infty}dt\,s'(t)h^{*}((t-b)/a)/a$$
$$=\int_{-\infty}^{+\infty}dt's'(t')h^{*}((t'-b')/a')/a'\ . \tag{5}$$

Use is made of the change of variables: $t'=\alpha t$, $a'=\alpha a$, and $b'=\alpha b$, and Eq. (5) becomes exactly equal to the scaled original $W(a,b)$, Eq. (2), located radially by a factor of α in both the a and b planes [cf. Eq. (31) of Ref. [10] for the conventional wavelet normalization of the inverse square root of a], plus noise:

$$W'(a,b)=W(\alpha a,\alpha b)/\alpha+(\text{noise})\ . \tag{6}$$

Example: If $\alpha=2$, then the signal $s(2x)$ is shrunk by half (where, for example, the peak value of $s(x_0)$ at x_0 is shrunk to a location $x=x_0/2$). Therefore the WT coefficient $W'(a,b)=W(2a,2b)$ is shifted toward the origin by a factor of 2, so that for a compacted signal the WT locates toward the center by the factor 2 along both the a and b axes.

III. DESIGNS AND APPLICATIONS

This observation makes the geometric meaning of Eq. (6) clear and our design of a wedge detector on the TSJR plane follows. The wedge filter bank consists of N radial wedges [see Fig. 3(a)] that have N equally spaced angles in the upper half plane where $a\geq 0$,

$$b_n\equiv a\tan(n\theta_0)\ , \tag{7}$$

$$b_{n+1}\equiv a\tan[(n+1)\theta_0]\ , \tag{8}$$

where $\theta_0\equiv 2\pi/N$ for a total of N wedges. The total output value collected through the wedge-shaped filter denoted by w_n' for $s'(t)$ is integrated

$$w_n'\equiv\int_0^{\infty}da\sum_{b=b_n}^{b_{n+1}}W'(a,b)=w_n/\alpha \tag{9}$$

to give its value $1/\alpha$ multiplying with that w_n for $s(t)$. Since this formula (9) is true for every wedge-extracted value, a simple normalization of all values gives us a set of scale-invariant feature values. This fact is obvious, as often is the case in hindsight, in that the size information of the object is mapped to the value of some area integral and then a simple normalization can ignore the size information.

The present technique based on wavelets and neurons seems to be an ideal tool to solve real-time physics application problems that require scale invariance. For example, wavelet transforms have been applied to the onset of turbulent flow by Frisch and Orzag [7], and the Kolmogorov turbulence cascade by Meneveau [6], who have demonstrated the self-similar nature of the turbulent flow. Thus the present technique might be used to enhance the turbulence onset signal giving better diagnosis in real fluids, because it collects scale-invariant information at multiple resolution scales. A similar application by Freysz *et al.* [5] that can be extended by invariant wavelet technique was to investigate the scale-invariant fractal aggregation in diffusion-limited cases. Another application of invariant wavelet transform is the bubble-chamber particle-track-recognition problem with a large throughput rate in the superconducting-supercollider experiments. The challenge there is perhaps similar to the star track minutiae associated with the task of compressing Federal Bureau of Investigation (FBI) fingerprint files by a lossless wavelet transform which is approved by the FBI to be better than the traditional discrete cosine transform used in HDTV and data compression. However, the particle tracking is not stationary and is further compounded by the time-dependent event under electromagnetic field fluctuations, for which a scale-invariant wavelet transform automation may prove to be useful for minor changes in the measurement setup. In the interest of simplicity to bring forward the use of this synergism between wavelets and neurons we consider a one-dimensional example. An important measurement problem in astrophysics is detecting a weak signal under an unknown Doppler shift. This type of frequency modulation by environmental perturbations is modeled as follows.

IV. AUTOMATIC CLASSIFICATION

As a generic illustration, we consider two classes of pulse signals that have arbitrary carrier frequency and different frequency modulations and hopping. One class

has high-, low-, and medium-frequency modulated pulses, denoted as HLM, and the other class has three modulated pulses of equal frequency, denoted as MMM for a medium carrier frequency. A natural wavelet to analyze the change of pulses is the bipolar Haar wavelet $h(t)=\{-1,+1\}$ [12]. The HLM and MMM signal templates are shown in Fig. 1. The first and third HLM segments in Fig. 1(a) have an amplitude of 2 while the middle segment has 1.75 amplitude, and altogether there are 512 data points. The MMM signal template in Fig. 1(b) has one medium frequency in three pulses. Training and test vectors with differing scale and noise are normalized so each feature has a zero mean and 0.75 standard deviation. The training set consisted of $s_1(t)$ and $s_2(t)$ (subscripts denote the two classes), while the test set consisted of $s_1(2t)$, $s_2(2t)$, $s_1(2t)+n(t)$, and $s_2(2t)+n(t)$, where $n(t)$ is 20 dB noise. Figures 2(a) and 2(b) show the two noisy test signals. Figures 3(b) and 4 show simulated wedge outputs for the original HLM signal $s_1(t)$ and scaled and shifted versions. The wedges in Fig. 4(a) of the scaled signal are essentially identical to the original in Fig. 3(b). The wedges in Fig. 4(b) of the shifted signal show graceful degradation for shifts as well. To capture shift invariance, one can likewise design a horizontal bar filter bank (space does not permit us to discuss this further).

This invariant feature set can be furthermore fed into a two-layer feed-forward ANN for both the interpolation of discrete-sampled wedges and subsequent classification as demonstrated in the right-hand side of Fig. 5. Each signal is wavelet transformed by the Haar wavelet, and collected through 32 wedge detectors and fed into 32 input neurons of the input layer of the ANN. An $M\times N$ two-layer feed-forward ANN has $M=1$ single output-layer neuron for two classes (target or clutter) and $N=32$ neurons at the input layer, one for each wedge detector

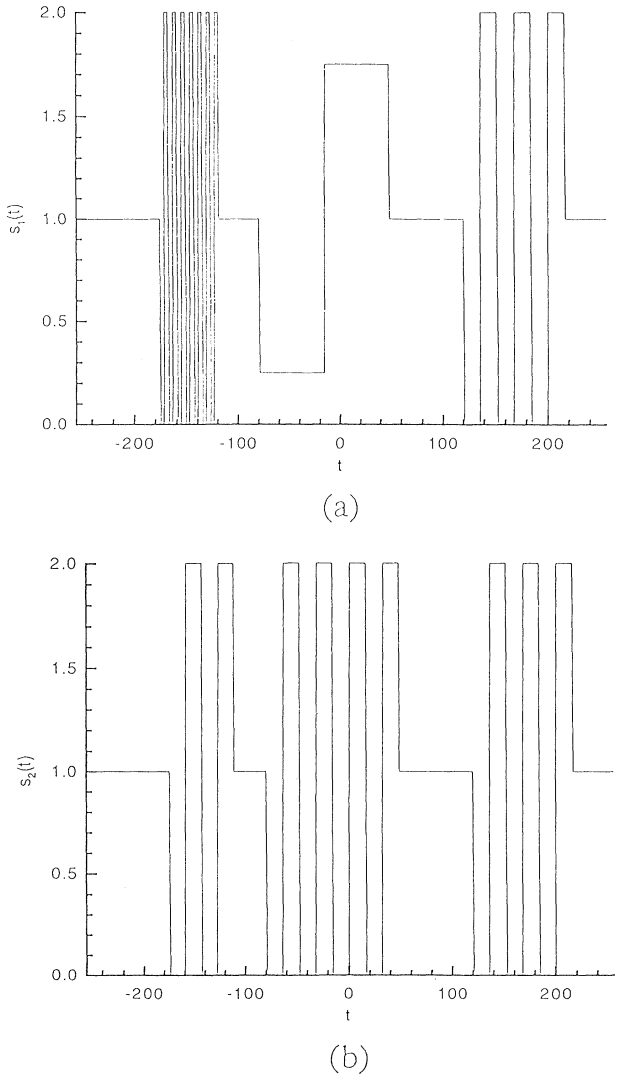

FIG. 1. Training set for two signal classes: (a) HLM $[s_1(t)]$, (b) MMM $[s_2(t)]$.

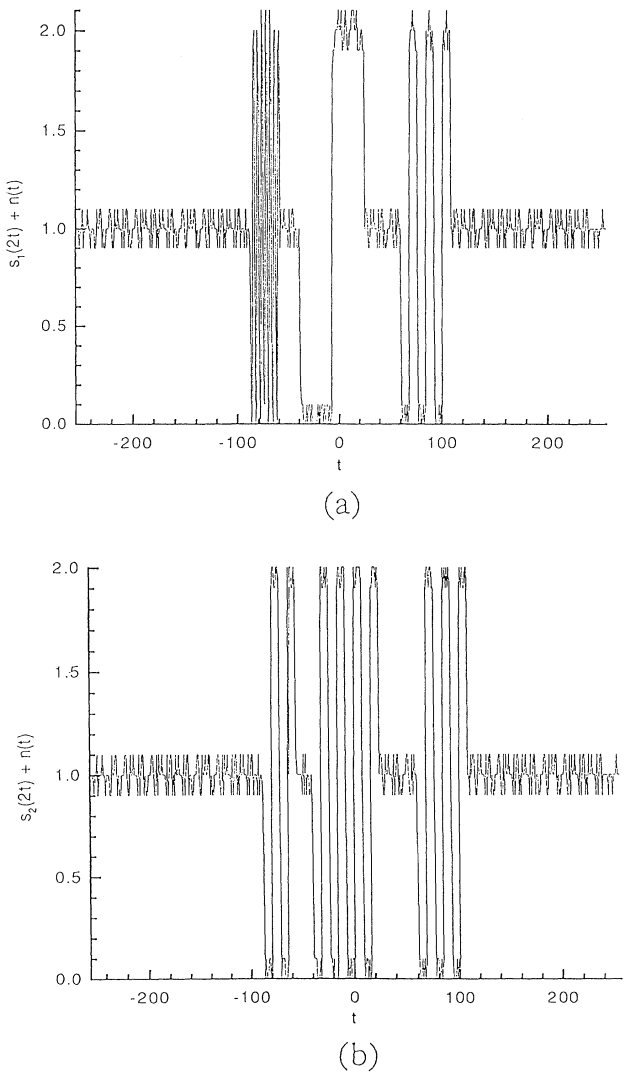

FIG. 2. Two noisy scaled signals from the test set: (a) HLM $[s_1(2t)+n(t)]$, (b) MMM $[s_2(2t)+n(t)]$.

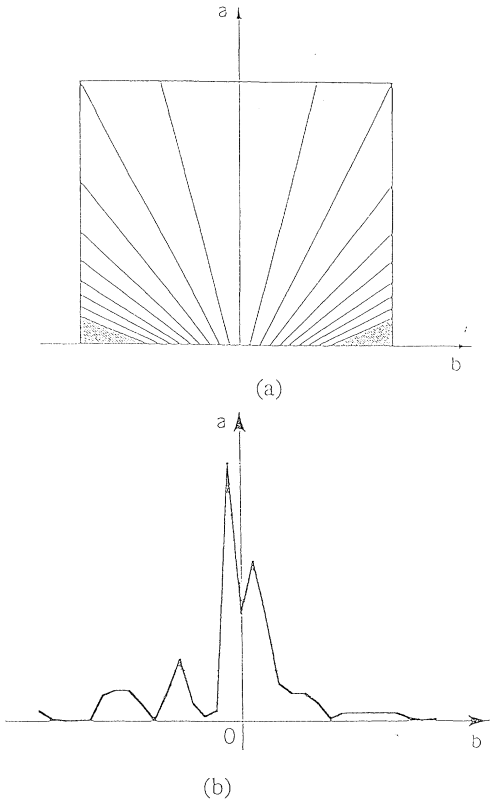

FIG. 3. (a) Wedge detector, (b) wedge outputs for $s_1(t)$.

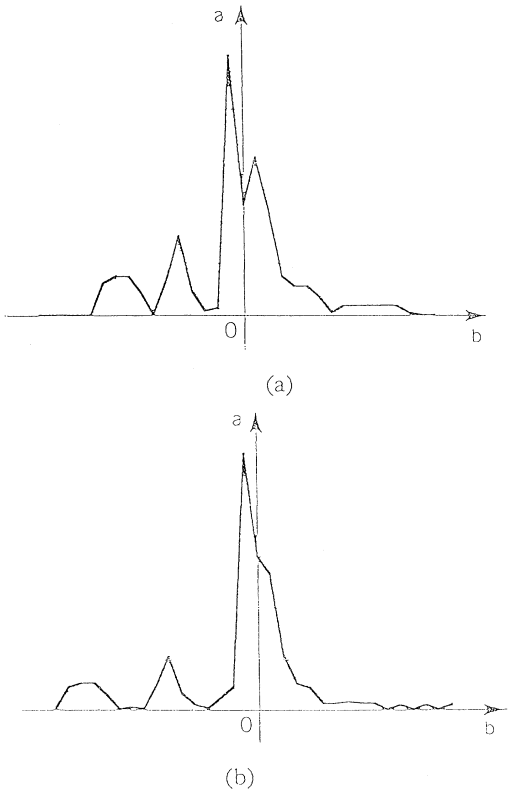

FIG. 4. Wedge outputs for (a) $s_1(2t)$ and (b) $s_1(t+64)$.

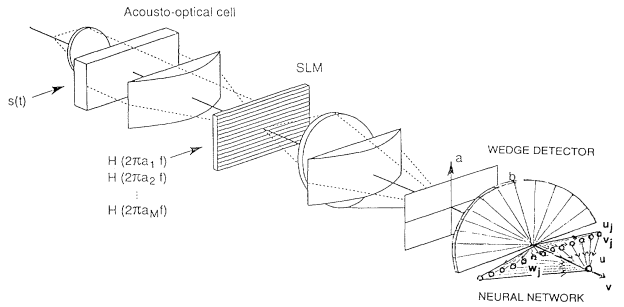

FIG. 5. Two-dimensional optical correlator with cylindrical FT lenses and a bank of the wavelet filters (SLM) for the WT of a one-dimensional signal (Ref. [10]) followed by the wedge detector, in the time-b and scale-a domain, and a two-layer feed-forward neural network for the noisy- and distorted-signal classification.

output. There are several approaches to treat the neural-network postprocessing (cf. any textbook for the Hopfield topdown energy design, the Kohonen self-organization feature map, the multiple-layer perceptions, the Grossberg adaptive resonance theory, etc.). We are not going to discuss the pros and cons of any specific architecture, except for the sake of completeness a simple gradient descent learning algorithm, well known to the neural-network community as the "backprop", is described as follows.

V. NEURAL-NETWORK LEARNING ALGORITHM

A performance or energy function $E(v)$ is a function of the output neuron activity function v. The usual choice is the squared difference of desired output d (during the training, say one for class 1 and zero for class 2) and the actual output v.

$$E(v)=(v-d)^2/2 \ . \tag{10}$$

The first layer has $N=32$ neurons whose input u_j and output v_j are denoted by the subscript j (note that the i index of the second-layer neuron is suppressed for a single neuron at the second layer considered here and thus the traditional double indices of the interconnect weight w_{ij} are also reduced to one index w_j for a single neuron at the second layer). The net input to the second-layer neuron is a w_j weighted sum from all outputs from the first layer:

$$u=\sum_{j=1}^{32} w_j v_j-\theta \ . \tag{11}$$

The soft threshold logic is defined by the logistic function

$$v=1/[1+\exp(-u)] \ , \tag{12a}$$

which has a non-negative logic with more input u implying more output v,

$$\left[\frac{dv}{du}\right]=v(1-v)\geq 0 \ . \tag{12b}$$

TABLE I. Network outputs for training and test set.

Training			Testing without noise		Testing with noise	
$s_1(t)$	HLM	0	$s_1(2t)$	5×10^{-6}	$s_1(2t)+n(t)$	0.023
$s_2(t)$	MMM	1	$s_2(2t)$	0.999	$s_2(2t)+n(t)$	0.990

The learning of the interconnect weight value is achieved by a local gradient descent,

$$\frac{\partial w_j}{\partial t}=-\frac{\partial E}{\partial w_j}\ . \tag{13}$$

Then, the standard gradient descent algorithm is used to determine the weights

$$\Delta w_j\equiv\left[\frac{\partial w_j}{\partial t}\right]\Delta t=-\left[\frac{\partial E}{\partial v}\right]\left[\frac{dv}{du}\right]\left[\frac{\partial u}{\partial w_j}\right]\Delta t$$
$$=(d-v)v(1-v)v_j\tau\ , \tag{14}$$

where use is made of the gradient descent to compute the chain rule of differentiations from Eqs. (10)–(12), and $\Delta t=\tau$. Given signal templates $s_i(t)$, $i=1,2$, for the training by Eq. (14), we have achieved perfect classification of two noisy contracted signals, Eq. (4), with unknown compaction scale α_i. The network outputs in Table I can be seen to give the desired results.

VI. OPTOELECTRONIC IMPLEMENTATIONS

An optoelectronic architecture for optical implementation [10] of the WT preprocessing and the neural-network classification is given in Fig. 5 [13]. Input data is represented by an acousto-optical transducer (with a speed exceeding 1 GHz). Daughter wavelets are encoded in the form of a film mask, or holographic matched filters called the spatial light modulator (SLM). The output is a two-dimensional plane. The horizontal direction corresponds to the shift b (discrete sampled), and the vertical direction corresponds to the dilation factor a (discrete sampled). Discrete sampling is taken care of by the fault tolerance of perceptron neural-network interpolation. Furthermore, we assume that a small angular separation $\theta_0\equiv2\pi/N$ between the width between $b_n\equiv a\tan(n\theta_0)$ and $b_{n+1}\equiv a\tan[(n+1)\theta_0]$ is approximated by a constant width, $a\theta_0$, for a large N detector bank, as depicted in Fig. 3(a). This is easier for the layout of the optoelectronics device so that each row of detectors has an equal spacing proportional to the location a of the row. A more expensive monolithic design of the wedge detectors (where each wedge is a single large detector) can also be used but the integration time will be longer to slow down the real-time operation. Then, a simple neural network is used to classify the yes and no decision about the detection of a specific event under arbitrary distortion and noise in the domain of time-scale joint representation.

In an attractive alternative approach, the wavelets at different shifts and dilations associated with a single wedge can instead be summed to simplify the detector. This can be done because the mathematical transform is linear. In this case, the transform consists simply of computing inner products of each wavelet sum (corresponding to each wedge-shaped detector). Only a single point detector is then needed for each inner product (one inner product per wedge), simplifying implementation.

VII. SUMMARY

In conclusion, the alluring synergism expressed in the Introduction is the linear superposition principle of the WT that allows a nonsinusoidal kernel to form a discrete but complete orthonormal set, that happens to match nicely with the discrete input nature of artificial neural networks. Thus one suspects that this enhanced capability to cope continuously with the real world variability is perhaps mathematically similar to the human sensor systems, which, of course, have major design detail differences due to the adaptation evolution processes. The bipolar Haar wavelet is chosen for the multiresolution change detection, via the $+$, $-$, and summation operations over different sizes of grids. Our purpose is to demonstrate the wedge-shaped detectors of different orientations, similar to the Hubel and Wiesel complex cortical cells, in order to extract invariant and multiscale wavelet features from noisy transients. We wish to emphasize that this invariant-feature extraction is mathematically valid for any wavelet prepossessing scheme and for any neural-network postprocessing architecture that are chosen appropriately to solve a specific problem at hand.

ACKNOWLEDGMENTS

The support of NSWCDD Independent Research Fund for H.H.S. and ONR for B.A.T. are acknowledged.

[1] H. Szu, Opt. Eng. **21**, 804 (1982).
[2] Opt. Eng. **31** (9) (1992), special issue on wavelet transform, edited by H. Szu and H. J. Caulfield.
[3] IEEE Trans. Inf. Theory **IF-38** (2) (1992), special issue on wavelet transform, edited by I. Daubechies, S. Mallat, and A. Wilsky.
[4] IEEE Trans. Signal Process. **SF-41** (12) (1993), special issue on wavelet transform, edited by P. Duhamel, P. Flandrin, T. Nishitani, A. Tewfik, and M. Vetteri.
[5] E. Freysz, B. Pouligny, F. Argoul, and A. Arneodo, Phys. Rev. Lett. **64**, 745 (1990).
[6] C. Menevaeu, Phys. Rev. Lett. **66**, 1450 (1991).
[7] U. Frisch and S. A. Orzag, Phys. Today **43** (1), 24 (1990).
[8] H. Szu, B. Telfer, and S. Kadambe, Opt. Eng. **31**, 1907 (1993).
[9] B. Telfer and H. Szu, Opt. Eng. **31**, 1830 (1993).
[10] H. Szu, Y. Sheng, and J. Chen, Appl. Opt. **31**, 3267 (1992).
[11] G. G. Lendaris and G. L. Stanley, Proc. IEEE **58** (2), 198 (1970).
[12] A. Haar, Math. Ann. **69**, 331 (1910).
[13] H. Szu, B. Telfer, X. Yiang, and Y. Sheng, NSWC Invention Disclosure, Case No. 75380, April 1993.