Automated Detection of Subpixel Hyperspectral Targets With Continuous and Discrete Wavelet Transforms

Lori Mann Bruce, *Member, IEEE*, Cliff Morgan, *Student Member, IEEE*, and Sara Larsen, *Student Member, IEEE*

Abstract—A major step toward the use of hyperspectral sensors to detect subpixel targets is the ability to detect constituent absorption bands within a pixel's hyperspectral curve. This paper introduces the use of multiresolution analysis, specifically wavelet transforms, for the automated detection of low amplitude and overlapping constituent bands in hyperspectral curves. The wavelet approach is evaluated by incorporating it into an automated statistical classification system, where wavelet coefficients' scalar energies are used as features, linear discriminant analysis is used for feature reduction, and maximum likelihood (ML) decisions are used for classification. The system is tested using the leave-one-out procedure on a database of 1000 HYDICE signals where half contain a subpixel target or additive Gaussian absorption band. Test results show that the continuous and discrete wavelet transforms are extremely powerful tools in the detection of constituent bands, even when the amplitude of the band is only 1% of the amplitude of the background signal.

I. INTRODUCTION

WITH THE increasing availability of remotely sensed hyperspectral imagery, the potential for accurate, automated scene analysis and target recognition systems is ever-expanding. Hyperspectral reflectance curves can provide insight into the on-ground (or near ground) constituent materials in a single remotely sensed pixel. Assuming a nonhomogeneous mixture of subpixel materials, a fundamental goal would be to decompose the single hyperspectral curve and extract information about the presence and concentration of each of the materials. This capability has many applications. One example is the detection of weakly concentrated, near-ground aerosols in hyperspectral imagery. This application is particularly challenging because aerosols can cause increased or decreased absorption and, when weakly concentrated, can have extremely low amplitudes relative to the background clutter's hyperspectral curve.

A. Linear Mixing and Gaussian Absorption Bands

Many approaches to resolving subpixel targets utilize linear unmixing models, where the methods are based on the assumption that the hyperspectral curve for a given pixel is a linear combination of the endmember spectral curves. The endmembers correspond to hyperspectral curves that are unmixed spectra of pure materials. Methods such as constrained least squares, are then utilized to determine the fractional contribution of each endmember to each pixel. Most automated methods for finding endmember curves are essentially means of finding the purest pixels in the image. However, it may be impossible to find completely pure pixels.

When linear unmixing approaches are not appropriate, one may detect subpixel targets through feature extraction and classification. High spectral resolution, as with hyperspectral imagery, provides the opportunity to detect some materials by the presence of their specific absorption bands. Many materials have diffuse reflectance spectra where the absorption band is actually a distribution function that contains contributions from a large set of transmitted light orientations. Consequently from the central limit theorem, the absorption band shape will tend toward a Gaussian distribution, and the absorption band's center wavelength will be the mean of the band centers over all transmitted light orientations [1].

B. Derivative Analysis

For target detection purposes, the amplitude, mean, and variance of the Gaussian absorption band could be used as quantitative features in an automated classification system. However, this approach is complicated by factors such as unresolved composite absorption bands and low SNRs in the hyperspectral curve. Multiple high-order derivative analysis has been utilized to help overcome this problem [1], [2]. Derivative-based methods typically use a preprocessing, smoothing filter followed by a derivative operator, e.g., mean filter smoothing followed by the finite approximation derivative algorithm. Another commonly used method is that of Sovitzky-Golay, where the smoothing and derivative operations are combined into one calculation [3]. The finite approximation algorithm estimates derivatives as follows:

$$\frac{df(\lambda)}{d\lambda}\Big|_i = \frac{f(\lambda_{i+k}) - f(\lambda_i)}{\Delta\lambda} \tag{1}$$

where $f(\lambda)$ is a digital hyperspectral curve. The separation between wavelengths $\Delta\lambda = \lambda_{i+k} - \lambda_i$ is a variable parameter that determines the resolution of the results. Generally, $k = 1$ but can be varied. The derivative will extract different information for different band separations [1]–[7]. By increasing the band

separation or the width of the smoothing filter, the effects of high-frequency components in the signal are decreased. However, the challenge is to appropriately select the width of the derivative window and/or the width of the smoothing prefilter.

In previous work, the appropriate width of the smoothing prefilter has been determined experimentally, on a case-by-case basis. Huguenin and Jones optimized the width of the smoothing prefilter via a tradeoff between noise and signal attenuation, requiring an estimate or prediction of the full-width-half-maximum of the constituent reflectance bands contained in the hyperspectral curve [1]. In Demetriades–Shah *et al.*'s use of the Savitsky–Golay method [4] to detect the "red edge" and to decorrelate soil background from vegetation spectra, they experimentally determined the width of the smoothing operation by maximizing the SNR of the smoothed curve [4]. Tsai [2] examined the effects of changing the width of the smoothing operation and the tradeoff between noise removal and ability to resolve fine spectral details [2]. Their conclusion was that the smoothing width should be as large as possible but remain less than the width of the spectral feature of interest.

Piech and Piech [5], [6] have conducted investigations into the use of scale-space images for systematically varying the smoothing width in derivative analysis of hyperspectral curves [5], [6]. Scale-space images are produced by convolving the hyperspectral curve with Gaussian derivative filters which have been scaled, or stretched, to have increasing widths. This method is closely related to the wavelet transform.

C. Wavelet Approach

The wavelet transform is a relatively new signal analysis tool that provides a systematic means for analyzing signals at various scales or resolutions. With continuous wavelet transforms (CWT), one can analyze signals, including multidimensional signals such as image cubes, across a continuum of scales. With discrete wavelet transforms (DWT), signals are analyzed over a discrete set of scales. Typically, the discrete scales are dyadic (2, 4, 8, 16, …) and the transform can be implemented using a variety of fast algorithms and customized hardware, e.g.,the multiresolutional dyadic filter tree implementation.

The CWT and the DWT have been applied to various areas of remotely sensed signal processing. To date in remote sensing, wavelets have mainly been used for data compression [8], [9]. Other remote sensing applications of wavelets include image texture analysis [10], image edge detection [11], image denoising [12], and data fusion [13]. Wavelets have proven to be quite powerful in these applications. However, the use of wavelet transforms for analyzing hyperspectral data has been extremely limited. In this paper, the authors propose the use of the CWT and the DWT for the automated detection of constituent bands in hyperspectral curves. This approach offers the ability to vary the widths of signal processing operators, such as smoothing and derivative operators, in an automated fashion. As a result, the hyperspectral curve is analyzed over a range of resolutions, and the user need not select or assist in the selection of a specific operator width.

The primary goal of this paper is to demonstrate the use of wavelet-based methods for detection of weak subpixel targets in hyperspectral imagery. An automated hyperspectral curve constituent band-detection system is designed, and the performance of the system is evaluated. The automated system consists of three modules: feature extraction, feature reduction, and classification. The feature extraction module is based on the wavelet transform, and the feature reduction and classification modules are based on traditional statistical methods. The wavelet-based methods are dependent on the type of wavelet transform and the selection of the mother wavelet function. Thus, a secondary goal of this study is to determine which type of wavelet transform (CWT or DWT) and which standard mother wavelets perform best in the automated detection system.

Section II of this paper provides an overview of the wavelet-based system designed and tested for automated detection of constituent bands in a hyperspectral curve. Section III briefly overviews necessary background material on the CWT and the DWT. Section IV provides details about the methodologies used for the wavelet-based feature extraction and reduction, hyperspectral curve classification, and system performance evaluation. The results from analyzing the performance of the system on a database of HYDICE curves are discussed in Section V, and final conclusions are provided in Section VI.

II. SYSTEM OVERVIEW

Fig. 1 provides an overview of the automated detection system. For the first module of the automated detection system, two feature extraction methods are implemented and tested, and they are based on the CWT and the DWT of hyperspectral curves. The features can be computed using information from the entire or a particular range of the curve. If the entire curve were used, it would represent a scenario where the automated detection system had no *a priori* information concerning the spectral location of the suspected target/anomaly. However, the situation could occur where a specific target were suspected, and hence the system would know a spectral range over which the anomaly might exist. In this case, the automated detection system would extract features using information from a particular range of the curve, or region of interest (ROI), as shown in Fig. 2.

The second module of the automated detection system is feature reduction/selection. Linear discriminant analysis (LDA) is used for the feature reduction process. With LDA, the features are weighted according to their discriminating ability, and the dimensionality of the feature space is reduced to one, thus enabling a simple graphical representation of class separation. The third module of the automated detection system is classification, and maximum likelihood (ML) classification is utilized. The classification is either "target" or " no target" depending on whether or not the constituent absorption band is detected. The classification method is supervised and thus requires training prior to its use in the automated detection system. The classifier is trained and tested on a database of HYDICE curves in a non-biased manner, i.e.,the curves used for training and the curves used for testing are randomly selected and are mutually exclusive.

The performance of the system is evaluated in several ways. First, the reduced features, resulting from LDA, are evaluated
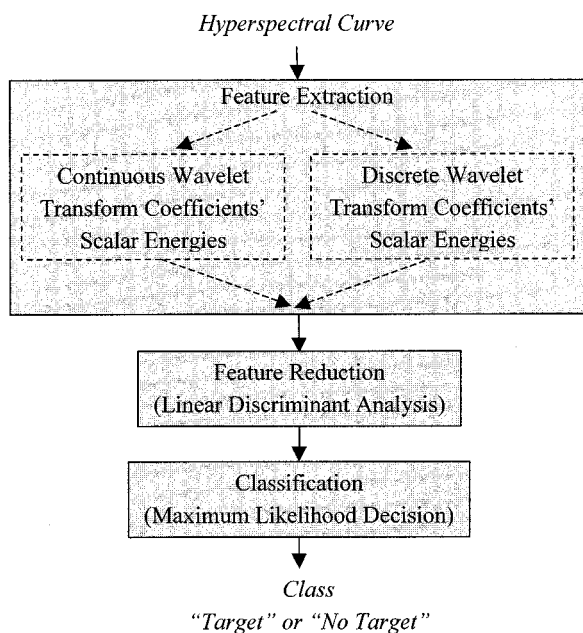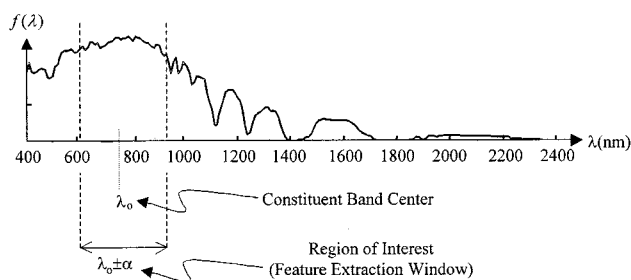
Fig. 1. Automated detection system overview.

Fig. 2. Region of interest (ROI) determination for feature extraction.

using receiver operating characteristics (ROC) curves. The area under the ROC curve is used to evaluate how well the automated detection system can be trained. This is useful since the evaluation is not dependent on the type of classifier being used; it simply provides a means for evaluating the discriminating ability of the wavelet-based features. The performance of the entire automated system is also evaluated using ML classification accuracies.

III. WAVELETS BACKGROUND

Wavelet analysis is based on the idea of projecting a signal onto a set of basis functions. A set of wavelet basis functions, $\{\psi_{a,b}(\lambda)\}$, can be generated by shifting and scaling the basic or mother wavelet, $\psi(\lambda)$, according to the following:

$$\psi_{a,b}(\lambda) = \frac{1}{\sqrt{a}}\psi\left(\frac{\lambda - b}{a}\right) \tag{2}$$

where $a > 0$ and b are real numbers. The variable a is the scaling factor of a particular basis function and b is the translation variable along the function's range. When $a > 1$, the functions are dilated and when $a < 1$, the functions are contracted. The coefficient $1/\sqrt{a}$ is included to normalize the energy of the wavelets.

All of the wavelets $\{\psi_{a,b}(\lambda)\}$ generated by shifting and scaling the mother wavelet $\psi(\lambda)$ have the same basic shape.

All wavelet functions must oscillate, have an average value of zero, and have finite support. This "admissibility condition" can be represented by

$$\int_{-\infty}^{+\infty} \frac{|\Im(\psi(\lambda))|^2}{|s|}ds < \infty \tag{3}$$

where $\Im(\cdot)$ denotes the Fourier transform (FT) and s is the Fourier domain variable. An important property of many wavelet systems is the multiresolution analysis (MRA) property, where the decomposition of a signal is in terms of the resolution of detail signals [14]. If a wavelet basis satisfies the MRA criteria, its associated DWT can be implemented with a multiresolutional dyadic filter tree which is shown in Fig. 3. This type of implementation is very useful because it allows for fast algorithms (similar to the well-known fast Fourier transform (FFT) to be utilized for the DWT. For this study, Gaussian derivative wavelet functions are constructed for use in feature extraction. However, these wavelet functions do not possess the MRA property. Consequently, a dyadic filter tree cannot be implemented, and the CWT is utilized instead.

There exist many different types of mother wavelets and wavelet bases. The Haar wavelet is one of the simplest examples. The Haar wavelet is discontinuous, and it resembles a step function

$$\psi(\lambda) = \begin{cases} 1, & 0 \leq \lambda \leq \frac{1}{2} \\ -1, & \frac{1}{2} \leq \lambda \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Also, a well-known family of wavelets was developed by I. Daubechies [15], and they are generally referred to as Daubechies–n, where n is the "order" of the mother wavelet. The order corresponds to the regularity of the mother wavelet, and the Daubechies-1 wavelet is equivalent to the Haar wavelet. For this study, the authors investigate the use of many different mother wavelets, including the Gaussian derivative, Coiflet, symlet, biorthogonal, and Daubechies family of wavelets for automated detection of constituent absorption bands.

The CWT, denoted by $W_f(a, b)$, of a function, $f(\lambda)$, with respect to the wavelet basis function, $\psi_{a,b}(\lambda)$, can be defined as

$$W_f(a, b) = \int_{-\infty}^{\infty} f(\lambda)\,\psi_{a,b}(\lambda)d\lambda \tag{5}$$

where the wavelet function $\psi_{a,b}(\lambda)$ is given by the previous equation. For the CWT, the scale parameter a and the shift parameter b are specified as real numbers. Hence, the transform coefficients $W_f(a, b)$ are continuous with respect to the variables a and b [14]. For the DWT, the wavelet basis functions are represented as

$$\psi_{j,k}(\lambda) = 2^{\frac{-j}{2}}\psi\left(2^{-j}\lambda - k\right) \tag{6}$$
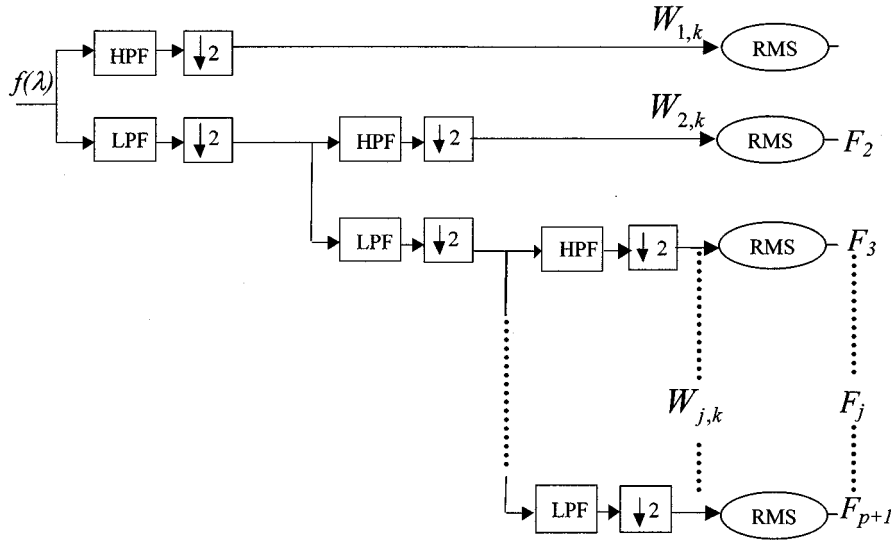
Fig. 3.   Dyadic filter tree implementation used for DWT coefficient scalar energy feature extraction.

and the wavelet coefficients are obtained by

$$W_{j,k} = \langle f(\lambda), \psi_{j,k}(\lambda) \rangle. \tag{7}$$

Thus, the scales are $a = 2, 4, 8, \ldots, 2^j, \ldots, 2^p$.

The DWT has been extensively used in the development of fast wavelet algorithms. The most common implementation of the DWT, the well-known dyadic filter tree is shown in Fig. 3 [15]. The highpass and lowpass filters correspond to the selection of mother wavelet function. At each stage of the filter tree, a set of approximation and detail coefficients are produced, corresponding to the input signal's large and small scale behavior, respectively. In theory, the decomposition can be extended to $p = \log_2(N)$ levels, where N is the length of the input signal. In practice, however, the maximum number of decomposition levels also depends on the choice of mother wavelet. For this study, the number of levels is chosen such that p is maximized for each mother wavelet investigated.

IV. METHODOLOGIES

The CWT and DWT are used to obtain features related to the fine detail and large scale behavior of the hyperspectral curves. The features are computed by taking the root mean square (RMS) energy of the wavelet coefficients at each scale a. A region of interest determines which portion of the hyperspectral curve is used for the wavelet decomposition and consequently, which coefficients are used in the calculation of the RMS energy. The region of interest (ROI) is defined as the constituent band's center location λ_0, plus or minus an uncertainty value α as illustrated in Fig. 2. For this paper, several uncertainty values are investigated, including $\alpha = 4, 5, 7, 10, 15,$ and 25 spectral bands. For example, HYDICE spectral bands are approximately 10 nm wide, so an uncertainty value of $\alpha = 4$ corresponds to a 90 nm-wide ROI. The RMS energy is also computed using all coefficients. This case represents the scenario where the system has no indication as to potential spectral location of the target's absorption band.

A. Continuous Wavelet Transform Feature Extraction

The 1×10 CWT feature vector $\bar{F} = [F_a]$ is computed as

$$F_a = \sqrt{\frac{1}{N} \sum_{b=1}^{N} [W_f(a,b)]^2} \tag{8}$$

where $W_f(a,b)$ is the CWT of the hyperspectral curve $a = 1, \ldots, 10$ are the scales. The CWT was implemented such that the endpoints of the hyperspectral curve were mirrored and only the center N samples of the output were retained when the length of the original hyperspectral curve was N samples.

The wavelet functions used for the CWT are derivative descendants of a Gaussian function. This is done for several reasons. First, convolution with a derivative of a Gaussian function is equivalent to convolution with a Gaussian function followed by the derivative operation. As a result, the Gaussian derivative CWT is equivalent to Gaussian smoothing followed by a derivative algorithm as was used by Huguenin and Jones [1] and Tsai and Philpot [2]. Second, the Gaussian functions are used since they match the functions traditionally used to model absorption and reflection anomalies of spectra materials. Wavelet coefficients are analogous to correlation coefficients in that large amplitude coefficients result when the shifted and dilated mother wavelet is similar in shape to the signal being analyzed. By selecting a Gaussian based mother wavelet, wavelet coefficients may be used to detect the presence of a Gaussian shaped target.

The wavelet functions used in the CWT were n th derivatives of a Gaussian function

$$\Psi_n(\lambda) = \frac{d^n}{d\lambda^n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\lambda-\lambda_0)^2}{2\sigma^2}} \right]. \tag{9}$$

For this study, seven different Gaussian-derivative mother wavelets were investigated, i.e., $n = 1, 2, \ldots, 7$. All of these functions fulfill the wavelet function admissibility criteria [14]. Note that the wavelet function named $\Psi_2(\lambda)$ is equivalent to

the well-known *Mexican-hat wavelet*. All seven are analyzed for their ability to detect whether or not a target is present in hyperspectral curves.

B. Discrete Wavelet Transform Feature Extraction

Fig. 3 shows a system overview of the DWT feature extraction methodology. The $1 \times (p+1)$ DWT feature vector $\bar{F} = [F_j]$ is computed as

$$F_j = \sqrt{\frac{1}{K}\sum_{k=1}^{K}|W_{j,k}|^2} \qquad (10)$$

where K is the number of coefficients at the decomposition level j, and p is the maximum number of decomposition levels. The feature vector is length $(p+1)$ due to the p levels of detail coefficients and one level of final approximation coefficients. Therefore, $(p+1)$ energy features represent each hyperspectral curve. The feature vector length varied depending on the choice of mother wavelet. For example, when using a Haar wavelet the feature vector is 1×8, and when using a Daubechies-10 wavelet the feature vector is 1×3.

When the wavelet basis is orthogonal, as is the case with all mother wavelets used for the DWT features in this study, the energy features represent a partitioning of the energy in the original hyperspectral curve. Thus,

$$\sqrt{\frac{1}{N}\sum_{\lambda=1}^{N}|f(\lambda)|^2} = \sum_{j=1}^{p+1}\sqrt{\frac{1}{K}\sum_{k=1}^{K}|W_{j,k}|^2} \qquad (11)$$

where N is the number of wavelengths or spectral channels in the hyperspectral curve [16]. That is, the $1 \times (p+1)$ DWT feature vector provides information about how the energy of the hyperspectral curve is partitioned according to scale.

Since the Gaussian derivative mother wavelets do not satisfy the MRA criteria, they cannot be implemented with the DWT dyadic filter tree. Therefore, other wavelet functions were investigated for the DWT features. These standard, commonly used mother wavelets are listed in Table I. All 36 mother wavelets are analyzed for their ability to detect whether or not a constituent band is present in a hyperspectral curve.

C. Performance Evaluation

For this study, LDA is used to reduce the wavelet-based feature vector to a feature scalar. The performance of these features are evaluated using receiver operating characteristic (ROC) curves [17]. The area under the ROC curve A_z can vary between 0.5 and 1.0, indicating, respectively, that the system does not separate the two classes (i.e., the two classes' features' probability density functions (PDF) are equivalent, which results in zero accuracy) and that the system perfectly separates the two classes, i.e., the two PDFs do not overlap, which results in 100% accuracy.

Further, the performance of the entire automated detection system is evaluated using ML classification accuracies [18]. The classifier is trained on 500 hyperspectral curves (250 with and 250 without targets) and tested on 500 hyperspectral curves (250 with and 250 without targets). The training data and testing

TABLE I
AREA UNDER ROC CURVE A_z FOR THE DM WITH A 3% AMPLITUDE CONSTITUENT BANDWITH VARYING WIDTH CENTERED AT 750 NM

Wavelet	ROI = "all"	$\alpha = 10$	$\alpha = 7$	$\alpha = 5$	$\alpha = 4$
Haar	0.609	0.646	0.639	0.773	0.713
Daub-2	0.573	0.625	0.608	0.675	0.675
Daub-3	0.639	0.626	0.685	0.602	0.602
Daub-4	0.579	0.555	0.570	0.588	0.588
Daub-5	0.560	0.562	0.561	0.571	0.571
Daub-6	0.561	0.562	0.555	0.572	0.572
Daub-7	0.579	0.549	0.554	0.554	0.554
Daub-8	0.569	0.557	0.556	0.561	0.561
Daub-9	0.557	0.568	0.570	0.569	0.569
Daub-10	0.562	0.556	0.554	0.569	0.569
Bior-1.3	0.621	0.611	0.607	0.620	0.620
Bior-1.5	0.553	0.564	0.555	0.569	0.569
Bior-2.2	0.614	0.564	0.615	0.548	0.548
Bior-2.4	0.577	0.540	0.569	0.535	0.535
Bior-2.6	0.587	0.585	0.586	0.586	0.586
Bior-2.8	0.567	0.534	0.552	0.553	0.553
Bior-3.1	0.690	0.713	0.707	0.712	0.712
Bior-3.3	0.568	0.567	0.577	0.579	0.579
Bior-3.5	0.565	0.547	0.550	0.555	0.555
Bior-3.7	0.560	0.535	0.542	0.535	0.535
Bior-3.9	0.564	0.563	0.560	0.530	0.530
Bior-4.4	0.572	0.537	0.567	0.545	0.545
Bior-5.5	0.572	0.567	0.576	0.575	0.575
Bior-6.8	0.563	0.537	0.549	0.551	0.551
Coiflet-1	0.613	0.558	0.579	0.628	0.628
Coiflet-2	0.573	0.573	0.573	0.570	0.570
Coiflet-3	0.567	0.556	0.521	0.515	0.515
Coiflet-4	0.566	0.565	0.566	0.553	0.553
Coiflet-5	0.549	0.556	0.555	0.555	0.555
Symlet-2	0.573	0.625	0.608	0.675	0.675
Symlet-3	0.639	0.626	0.685	0.602	0.602
Symlet-4	0.568	0.555	0.574	0.539	0.539
Symlet-5	0.561	0.567	0.572	0.570	0.570
Symlet-6	0.570	0.552	0.565	0.560	0.560
Symlet-7	0.568	0.563	0.563	0.562	0.562
Symlet-8	0.563	0.547	0.536	0.542	0.542

data are mutually exclusive. The ML classifier is a supervised method. It should be noted, however, that the use of training data does not require knowledge of the hyperspectral signal's noise or characteristics of the absorption band, such as amplitude, width, or location. This is critical since in many applications one can obtain training data where the target is present but does not have information about the target's true effects on the hyperspectral curve.

D. Study Data

In order to test the automated detection system, we need a database of hyperspectral curves for which the "ground truth" is known. For this reason, a database of hyperspectral digital image collection experiment (HYDICE) curves was collected and a known target was added to half of the curves. The additive target is a Gaussian function used to model a weak, subpixel target or constituent band. In order to test the robustness of the automated detection system, several databases were constructed. Some databases contain a positive amplitude anomaly, which models a target with higher reflectivity than surrounding materials in the scene, and some databases contain a negative amplitude anomaly, which models a target with lower reflectivity than surrounding materials in the scene. Also, several databases are constructed where the amplitude of the target constituent band is varied. Four amplitude cases

are investigated where the anomaly is 10%, 5%, 3%, and 1% of the amplitude of an average curve at that location. Thus, we can test the ability of the automated system to detect relatively strong versus weak targets.

The database used for this project was created from a single HYDICE data cube. The HYDICE data cube contained 100×100 pixels in each spectral band. The HYDICE cube contained 210 spectral bands that were about 10nm wide over the solar reflective portion of the spectrum from 400 nm to 2400 nm [18]. Fig. 4 shows the urban scene contained in the HYDICE data cube. The scene contained several man-made objects such as roads, parking lots, cars, and buildings, and some natural objects such as trees, grass, and soil.

The database developed for this project consisted of various data matrices (DM). Each DM contains 1000 randomly selected HYDICE curves where the pixel coordinates, (x, y), are determined using a two-dimensional (2-D) uniform distribution. A synthesized target (constituent band) is inserted into 500 curves of each DM. The constituent band is an additive Gaussian function. For this study, the detection of aerosols was of particular interest. Consequently, three center wavelengths were investigated: $\lambda_0 = 750$, 788, and 874 nm. The amplitude of the constituent band is a percentage of the mean value of all the DMs curves at that particular wavelength

$$A = \left[\frac{1}{1,000} \sum_{n=1}^{1,000} x_n(\lambda_0) \right] \times \text{percentage} \qquad (12)$$

where $x_n(\lambda_0)$ is the amplitude of the nth curve in the data matrix, and the percentage is 0.10, 0.05, 0.03, or 0.01%. The four varying amplitudes are used to provide a range of difficulty for the automated system. To provide insight into the effects of the constituent band's width, the standard deviation (σ) of the additive Gaussian, some DMs have constituent bands with a constant width, and some DMs have constituent bands with a width that fluctuates proportionally with the target's amplitude. For example, a constituent band with maximum amplitude of 10% has a width of $\sigma = \sqrt{10}$ HYDICE spectral bands ($\approx \sqrt{100}$ nm) while one with maximum amplitude of 3% has a width of $\sigma = \sqrt{3}$ spectral bands ($\approx \sqrt{30}$ nm).

The average SNR for each DM is computed by

$$SNR_{average} = \frac{1}{1000} \sum_{n=1}^{1,000} 10 \log_{10} \left(\frac{\sigma^2_{t\,arg\,et}}{\sigma^2_n} \right) \qquad (13)$$

where σ^2_{target} is the variance of the target signature, and σ^2_n is the variance of the nth curve in the data matrix. Table II shows the SNRs for several DMs used in this study. The SNRs show just how insignificant the target is compared to the HYDICE curves. Note that while the target's width has little impact on the average SNR, the SNR is clearly proportional to the amplitude of the target.

V. RESULTS

Table III shows the area under the ROC curve when using various Gaussian derivative mother wavelets for the DM with

Fig. 4. Nonuniform scene in HYDICE data cube used to generate hyperspectral curve database.

TABLE II
SNRs FOR ALL DATA MATRICES WITH THE FEATURE EXTRACTION ROI BEING THE ENTIRE HYPERSPECTRAL CURVE. DM1: POSITIVE AMPLITUDE TARGET WITH $\sigma = \sqrt{7}$ AND $\lambda_0 = 750$ nm. DM2: NEGATIVE AMPLITUDE TARGET WITH $\sigma = \sqrt{7}$ AND $\lambda_0 = 750$ nm. DM3: NEGATIVE AMPLITUDE TARGET WITH VARYING WIDTH AND $\lambda_0 = 750$ nm. DM4: POSITIVE AMPLITUDE TARGET WITH $\sigma = \sqrt{7}$ AND $\lambda_0 = 788$ nm. DM5: POSITIVE AMPLITUDE TARGET WITH $\sigma = \sqrt{7}$ AND $\lambda_0 = 874$ nm

	Signal-to-Noise Ratio (dB)				
Target Amplitude	DM1	DM2	DM3	DM4	DM5
10%	-52.10	-53.10	-51.40	-52.00	-46.30
5%	-64.70	-65.30	-66.50	-64.80	-59.40
3%	-74.20	-74.10	-77.50	-72.70	-68.30
1%	-92.80	-92.90	-98.20	-92.60	-87.70

a negative 3% amplitude constituent band with $\sigma = \sqrt{3}$ and $\lambda_0 = 750$ nm. Note that the $\Psi_7(\lambda)$ mother wavelet, i.e., the seventh derivative of a Gaussian, typically outperformed the other mother wavelets. The authors found this result to be generally consistent, regardless of the feature extraction ROI or the target's amplitude, variance, or center wavelength.

Fig. 5 shows the ML classification accuracies when using the $\psi_7(\lambda)$ mother wavelet. The results are shown for three sets of DMs: i) positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm, ii) negative amplitude constituent band with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm, and iii) negative amplitude constituent band with varying width and $\lambda_0 = 750$ nm. Note the trends in the results. The accuracies increase for increasing target amplitude and typically increase when the feature extraction ROI decreases from all of the hyperspectral curve to $\alpha = 10$ spectral bands (210 nm wide ROI). Fig. 6 shows the ML classification accuracies for the same three sets of DMs when using the $\psi_7(\lambda)$ mother wavelet with $\alpha = 10$. Note that classification accuracies were greater than 90% for constituent bands with amplitude of 3% or higher, and the classification accuracies were greater than 70% when the amplitude was 1%. These results are extremely impressive considering the average SNR ranged from -19.5 to -30.5 dB for the 3% amplitude DMs and ranged from -38.4 to -49.8 dB for the 1% amplitude DMs.

TABLE III

AREA UNDER ROC CURVE A_z WHEN USING CWT FEATURES FOR TARGET WITH NEGATIVE AMPLITUDE OF 3% AND $\sigma = \sqrt{3}$ LOCATED AT $\lambda_0 = 750$ nm

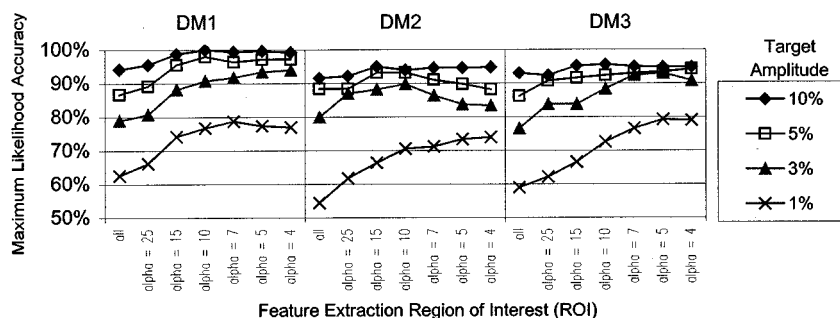|	ROI=all	α=25	α=15	α=10	α=7	α=5	α=4
$\psi_2(\lambda)$	0.681	0.739	0.839	0.905	0.865	0.926	0.964
$\psi_3(\lambda)$	0.746	0.805	0.888	0.920	0.949	0.941	0.938
$\psi_4(\lambda)$	0.751	0.813	0.840	0.891	0.931	0.956	0.981
$\psi_5(\lambda)$	0.773	0.810	0.937	0.930	0.931	0.972	0.914
$\psi_6(\lambda)$	0.796	0.801	0.895	0.897	0.939	0.975	0.986
$\psi_7(\lambda)$	0.791	0.877	0.905	0.913	0.963	0.985	0.958

Fig. 5. ML classification accuracy using CWT features with the $\psi_7(\lambda)$ mother wavelet. DMI: positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM2: negative amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM3: negative amplitude target with varying width and $\lambda_0 = 750$ nm.
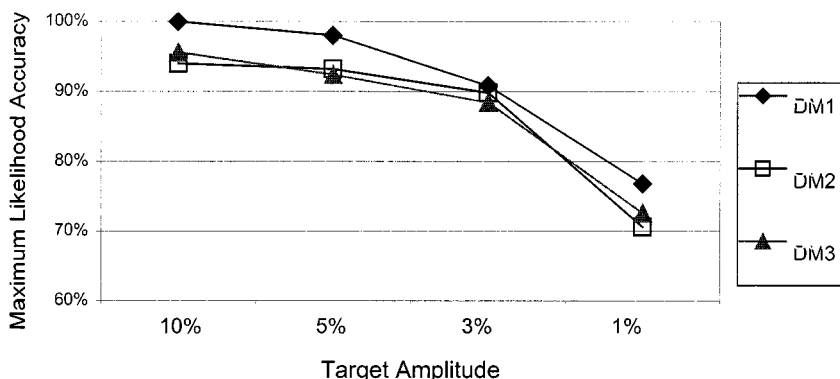

Fig. 6. ML classification accuracy using CWT features with the $\psi_7(\lambda)$ mother wavelet and $\alpha = 10$. DM1: positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM2: negative amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM3: negative amplitude target with varying width and $\lambda_0 = 750$ nm.

For the DWT method, 36 standard and commonly used mother wavelets were investigated. Table I provides the ROC results for all 36 mother wavelets for the DM with a negative 3% amplitude constituent band with varying width centered at 750 nm. Notice that certain mother wavelets outperformed the others regardless of the feature extraction ROI. These results were representative of all of the DMs investigated for this paper. Typically, the best performing mother wavelets were the Haar, Daubechies-2, Biorthogonal-1.3, Biorthogonal-3.1, Coiflets-1, and Symlets-2, regardless of the feature extraction ROI or the target's amplitude, variance, or center wavelength. These results may seem surprising since the mother wavelets did not particularly match the shape of the Gaussian absorption band. However, note that the individual wavelet coefficients were not used as features. Instead, the scalar partitioning of the signal's energy was used as the feature. Another interesting

point is that the six best mother wavelets were those who had the shortest impulse responses in the decomposition filters (see Fig. 3). That is, the lower order filters actually outperformed the higher order filters.

The six best mother wavelets were investigated more fully. Fig. 7 shows the ML accuracies for the DWT method when using the Haar mother wavelet. Note the trends in the results. The accuracies tend to increase for increasing target amplitude and tend to increase when the feature extraction ROI decreases from all of the hyperspectral curve to $\alpha = 7$. Fig. 8 shows the ML classification accuracies for the DWT method when using the Haar mother wavelet and a feature extraction ROI of $\alpha = 7$.

Fig. 9 shows a comparison between the CWT and the DWT methods when the feature extraction ROI was the entire hyperspectral curve, which corresponded to the case where no *a priori*
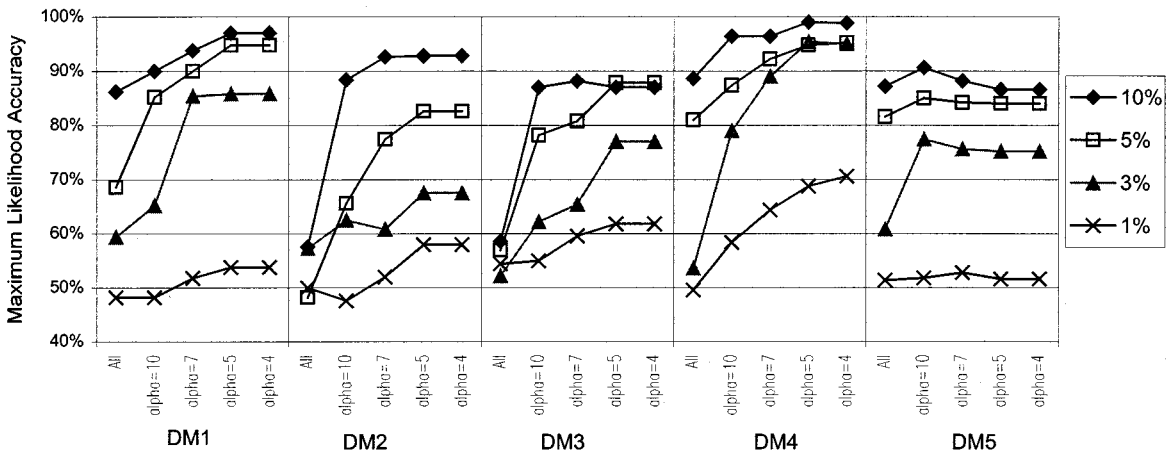
Fig. 7. ML classification accuracy using DWT features with the Haar mother wavelet. DM1: positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM2: negative amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM3: negative amplitude target with varying width and $\lambda_0 = 750$ nm. DM4: positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 788$ nm. DM5: positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 874$ nm.
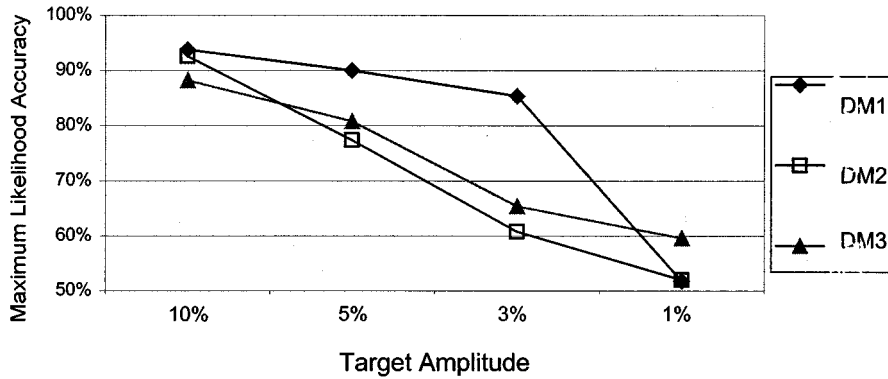

Fig. 8. ML classification accuracy using DWT features with the Haar mother wavelet and $\alpha = 7$. DM1: positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM2: negative amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM3: negative amplitude target with varying width and $\lambda_0 = 750$ nm.
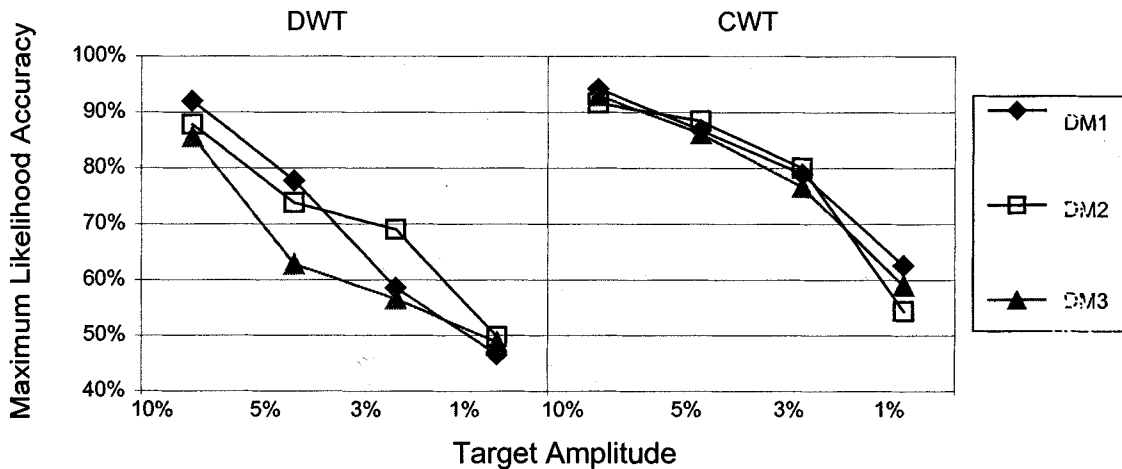

Fig. 9. ML classification accuracies with the feature extraction region of interest being the entire hyperspectral curve, using Haar and $\psi_7(\lambda)$ mother wavelet. DM1: positive amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM2: negative amplitude target with $\sigma = \sqrt{7}$ and $\lambda_0 = 750$ nm. DM3: negative amplitude target with varying width and $\lambda_0 = 750$ nm.

information was available about the possible location of the constituent absorption band. This is the worst case scenario: hyperspectral curves are indiscriminately processed in order to detect the subpixel target. Note the outstanding capability of the CWT method. Even when the anomaly amplitude was decreased to 3%, the CWT accuracies remained around 80%, whereas the DWT accuracies for the 3% amplitude target decline to around 60 to 70%.

VI. CONCLUSIONS

Wavelet-based feature extraction methods were developed and tested for detecting subpixel targets, specifically constituent absorption bands, in hyperspectral curves. The CWT was used to extract multiresolution features from the hyperspectral curve. Since a choice exists for the mother wavelet, derivatives of Gaussian functions were constructed and used for the wavelet function. These mother wavelets were used based on two motivating factors. The Gaussian-derivative wavelet decomposition is equivalent to i) Gaussian smoothing followed by a derivative operation at various window lengths and ii) correlating the hyperspectral curve with a Gaussian-derivative function. At each scale of the wavelet decomposition, the energy within a region of interest was computed and used to form a feature vector. The feature vector was then used in a statistical classification system to automatically determine whether or not a target, a constituent absorption band, was present. In general, the seventh derivative Gaussian mother wavelet performed best, and the classification accuracies of the CWT system increased with an increase in amplitude of the target and with a decrease in the size of the region of interest. Even when the amplitude of the target was only 3% as compared to the background hyperspectral curve ($SNR \approx -70$ dB) and the feature extraction region of interest was 21 spectral bands wide ($\alpha = 10$ or 210 nm), the classification accuracies were $\approx 90\%$.

In order to utilize the dyadic filter tree implementation, which is associated with commonly used fast algorithms, the DWT was also investigated. Since the Gaussian-derivative mother wavelets do not satisfy certain criteria for being used with the dyadic filter tree, other mother wavelets were utilized. Thirty-six standard, commonly used mother wavelets were investigated, and six of those consistently outperformed the others. The best six mother wavelets were the Haar, Daubechies-2, Biorthogonal 1.3, Biorthogonal-3.1, Coiflets-1, and Symlets-2. Of these six, it could not be determined that any one mother wavelet produced superior classification accuracies consistently. Thus, when considering these six, the Haar mother wavelet is a suitable choice because of its simplicity. It is quite surprising that the Haar mother wavelet performed so that it is discontinuous, and the additive target is a smooth Gaussian function. Also, it is interesting to note that the Haar wavelet function can be viewed as a low order approximation of the first derivative of a Gaussian function. Furthermore, the DWT with Haar mother wavelet can be viewed as a nonsmoothed, first-order derivative implemented as a finite approximation at varying window lengths.

Like the CWT system, the classification accuracies of the DWT system generally increased with an increase in amplitude of the target and with a decrease in the size of the region of interest. Typically, a significant increase in classification accuracies was obtained by decreasing the feature extraction region of interest to 21 spectral bands. Even when the amplitude of the target was only 5% as compared to the background hyperspectral curve ($SNR \approx -60$ dB) and the feature extraction region of interest was 21 bands wide ($\alpha = 10$ or 210 nm), the classification accuracies were typically 75% to 90%.

In general, the CWT system was found to provide superior classification accuracies as compared to the DWT system. The is due to the fact that the CWT method is tailored to match the behavior of more conventional methods [1], [2]. However, the DWT system has the advantage of lower computational requirements. Also, the mother wavelets investigated for the DWT system were not optimized. They were simply drawn from a pool of commonly used mother wavelets. One cannot say that the DWT results are optimum since there are infinitely many mother wavelets from which to choose. Thus, for this application, as well as many others, there is a great need for DWT implementations where the mother wavelet is optimized for the specific task at hand. However, even with the suboptimum DWT implementation, the target detection accuracies are very promising considering the relatively weak amplitude of the subpixel target.

REFERENCES

[1] R. L. Huguenin and J. L. Jones, "Intelligent information extraction from reflectance spectra: Absorption band positions," *J. Geophys. Res.*, vol. 91, no. B9, pp. 9585–9598, Aug. 1986.

[2] F. Tsai and W. Philpot, "Derivative analysis of hyperspectral data," *Remote Sens. Environ.*, vol. 66, pp. 41–51, Oct. 1998.

[3] Savitzky and N. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, pp. 1627–1638, July 1964.

[4] T. H. Demetriades-Shah, M. D. Steven, and A. C. Clark, "High resolution derivative spectra in remote sensing," *Remote Sens. Environ.*, vol. 33, pp. 55–64, July 1990.

[5] M. A. Piech and K. R. Piech, "Symbolic representation of hyperspectral data," *Appl. Opt.*, vol. 26, no. 18, pp. 4018–4026, Sept. 1987.

[6] ——, "Fingerprints and fractal terrain," *Math. Geol.*, vol. 22, pp. 457–485, May 1990.

[7] W. D. Philpot, "The derivative ratio algorithm: Avoiding atmospheric effects in remote sensing," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, pp. 350–357, May 1991.

[8] S. A. Werness, S. C. Wei, and R. Carpinella, "Experiments with wavelets for compression of SAR data," *IEEE Trans Geosci. Remote Sensing*, vol. 32, pp. 197–201, May 1994.

[9] P. Luigi-Dragotti, G. Poggi, and A. R. P. Ragozini, "Compression of multispectral images by three-dimensional SPIHT algorithm," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 416–428, Jan 2000.

[10] S. Fukuda and H. Hirosawa, "A wavelet-based texture feature set applied to classification of multifrequency polarimetric SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2282–2286, Sept. 1999.

[11] K. K. Simhadri, S. S. Iyengar, R. J. Holyer, M. Lybanon, and J. M. Zachary Jr, "Wavelet-based feature extraction from oceanographic images," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 767–778, May 1998.

[12] L. Zheng, A. K. Chan, S. Liu, W. Smith, and R. J. Holyer, "Directional clutter removal of aerial digital images using X-ray wavelet transform and Markov random field," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2181–2191, Sept. 1999.

[13] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1204–1211, May 1999.

[14] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[15] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[16] S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," *IEEE Trans Pattern Anal. Mach. Intell.*, vol. 21, pp. 83–88, Jan. 1999.

[17] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristics (ROC) curve," *Diag. Radiol.*, vol. 143, pp. 29–36, Apr. 1982.

[18] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. San Diego, CA: Academic, 1997.

Lori Mann Bruce (S'91–M'96) received the B.S.E. and Ph.D. degrees from the University of Alabama, Huntsville, in 1991 and 1996, respectively, and the M.S. degree from the Georgia Institute of Technology, Atlanta, in 1992, all in electrical engineering.

She has served as a Member of Technical Staff, U.S. Army Strategic Defense Command, and from 1996 to 2000, she was an Assistant Professor in the Electrical and Computer Engineering Department, Howard R. Hughes College of Engineering, University of Nevada, Las Vegas. Currently, she is an Assistant Professor in the Department of Electrical and Computer Engineering, Mississippi State University, where she is affiliated with the Engineering Research Center and the Remote Sensing Technology Center. Her research interests include wavelet transforms and automated pattern recognition in hyperspectral remote sensing and mammography.

Dr. Bruce is a member of Eta Kappa Nu, Phi Kappa Phi, and Tau Beta Pi.

Sara Larsen (S'01) received the B.S. in electrical engineering from the University of Nevada, Las Vegas, in May of 2001.

She is currently a Graduate Student in the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge. Her research interests at MIT are signal processing applications and neural computation. As an undergraduate, she worked as a Research Assistant under the guidance of Dr. L. M. Bruce. Her areas of specialization include digital signal processing, in particular using wavelets for feature extraction and signal enhancement for applications in patch clamp analysis, digital mammography, and remote sensing.

Ms. Larsen is a member of Tau Beta Pi.

Cliff Morgan (S'01) received the B.S. and M.S. degrees in electrical engineering from the University of Nevada, Las Vegas, in 1999 and 2000, respectively.

He is a Hardware Design Engineer with Hamel Davidson International Corporation, Durham, NC. While at the University of Nevada, Las Vegas, he was a Research Assistant, focusing his work on applying wavelet theory and statistical methods to the problem of target detection within images and signals. He served as an Aircraft Weapons Specialist and as a Digital Communication Technician in the United States Air Force from 1975 to 1995. His research interests include wavelet transforms and target detection in hyperspectral remote sensing.

Dr. Morgan is a member of Tau Beta Pi.