

Underwater Target Classification Using Wavelet Packets and Neural Networks

Mahmood R. Azimi-Sadjadi, *Senior Member, IEEE*, De Yao, Qiang Huang, and Gerald J. Dobeck

Abstract—In this paper, a new subband-based classification scheme is developed for classifying underwater mines and mine-like targets from the acoustic backscattered signals. The system consists of a feature extractor using wavelet packets in conjunction with linear predictive coding (LPC), a feature selection scheme, and a backpropagation neural-network classifier. The data set used for this study consists of the backscattered signals from six different objects: two mine-like targets and four nontargets for several aspect angles. Simulation results on ten different noisy realizations and for signal-to-noise ratio (SNR) of 12 dB are presented. The receiver operating characteristic (ROC) curve of the classifier generated based on these results demonstrated excellent classification performance of the system. The generalization ability of the trained network was demonstrated by computing the error and classification rate statistics on a large data set. A multiaspect fusion scheme was also adopted in order to further improve the classification performance.

Index Terms—Feature extraction, linear predictive coding, neural network, underwater target classification, wavelet packets.

I. INTRODUCTION

THE PROBLEM of detection and classification of underwater targets from the acoustic backscattered signals has attracted a lot of attention in recent years [1]–[19]. This involves discrimination between targets and nontarget objects as well as the characterization of background clutter. Several factors contribute to make the detection and classification processes a very complex problem. These include: nonrepeatability and variation of the target signature for different aspect angles and range of the target from the sonar, competing clutter caused by biological sources in the water column, surface and bottom reverberation effects, and lack of any *a priori* knowledge about the shape and geometry of the targets. Consequently, efficient and robust detection and classification schemes are needed to solve this complex problem.

Various signal processing schemes [1]–[8] have been applied to extract signatures of submerged targets from narrow band sonar data mainly for detection purposes. In [1], a joint time-frequency analysis of the impulse response of a spherical shell has

been studied for the characterization of the surface waves using Wigner–Ville distribution. The wavelet transform using a five-cycle cosine modulated Gaussian wavelet approximation was applied to the impulse response of spherical shell of differing thickness to examine resonance characteristics of the elastic target [2]. In [3], a wavelet-based method that uses an artificial neural network to adaptively compute discriminatory information on a target in the form of locations, size, and weights of Gaussian patches in the time-scale is described. De Billy [4] used the short time Fourier transform (STFT) to determine the resonance spectrum of submerged elastic cylindrical wires. In [5]–[7], different adaptive and spectral-based schemes for the isolation of target specular reflections in noisy backscattered signals were developed. The targets can be of arbitrary shapes and moreover no underlying model assumptions about the target and signal statistics are required. The effectiveness of these schemes was tested on several narrow band backscattered data sets from a mine target and a chunk of cement. A novel time delay and signal estimation scheme was developed in [8] that uses a discrete wavelet transform (DWT) and two different adaptation mechanisms for detection of weak target signals in highly nonstationary and cluttered environments. In [9], a multichannel extension of this detection scheme was developed for multiple aspects.

Classification of underwater targets from the acoustic backscattered signals has recently attracted a lot of attention as well. The pioneer paper by Gorman and Sejnowski [10] was perhaps the first paper which reported the application of neural networks to this area. They used simple spectral features as the input to the neural-network classifier in order to distinguish a cylindrical target from a rock with similar shape. The method in [11] uses the resonant scattering property of underwater objects that is dependent on the object size, shape, structure, and composition. A signal processing scheme called “G-Transform,” which consists of three sequential fast Fourier transform (FFT) of the backscattered signals, was developed [12] to represent the resonance or the modulation on the frequency spectrum of the backscattered signal. When used in conjunction with neural networks, the scheme was able to successfully detect and identify targets in shallow water. The generalized target description method in [13] is based on generating a parametric model for the target impulse response which models a range-distributed target as an array of point scatters. The model is a linear filter composed of the sum of differentiated and integrated delta functions of various orders. Wavelet transform was then used to extract the model parameters as the features for the classification of a mine and a rock from the backscattered signals. In [14], several methods were

Manuscript received September 4, 1998; revised June 28, 1999 and September 24, 1999. This work was supported by the Office of Naval Research (ONR321TS) and the Biosonar Program under Contract N00014-99-1-0166. The data and technical support were provided by the NSWC, Coastal Systems Station, Panama City, FL.

M. R. Azimi-Sadjadi, D. Yao, and Q. Huang are with the Signal/Image Processing Laboratory, Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: azimi@engr.colostate.edu).

G. J. Dobeck is with the Signal/Image Processing, NSWC, Dahlgren Division, Coastal Systems Station, Panama City, FL 32407-7001 USA.

Publisher Item Identifier S 1045-9227(00)01700-8.

tested for finding a good data representation scheme via optimal decomposition using wavelet basis functions or dictionaries. The classification results of different representation schemes were compared on two data sets. They reported 81.4% correct classification rate for 18.6% false alarm rate using single aspect data and on the same 40-KHz data set used in this paper. In [15], the standard matched filtering method and the spectrogram correlation and transformation (SCAT) algorithms were fused into one classification system. The SCAT algorithm [16] is a sonar receiver model for the big brown bat and is composed of three major components. The first cochlear block implements the encoding of frequency modulated sonar transmission in a spectrogram-like format, while the other two blocks, namely the spectral correlation block and spectral transformation block, process in parallel the temporal and spectral information in the sonar echoes. The ROC curve for this combined system [15] indicated approximately 75% classification accuracy for 10% false alarm rate on the same 40-KHz data set. The classification accuracy is significantly improved when the classification results of multiple aspects were fused together. In [17], Carpenter and Streilein used the preprocessed and matched filtered data of [15] and developed an ARTMAP-based fusion classification network. They reported a correct classification rate of 93.1% by fusing three 30° separation. In this scheme, the choice of the vigilance parameter for the ART network determines the granularity degree of the clusters. This choice becomes crucial when several noisy realizations of the backscattered data are to be used.

In this paper, a new wavelet-based classification scheme [18], [19] is developed to discriminate mine-like and non-mine-like objects from the acoustic backscattered signals used in [14], [15], and [17]. Wavelet packet is used to decompose the acoustic backscattered signals into several subbands. A fourth-order linear predictive coding (LPC) model is fitted to each subband signal and the coefficients are extracted and used as features. To choose the best set of features which adequately represent the data, a feature selection scheme is also adopted. This scheme not only provides a pertinent set of features but also reduces the dimensionality of the feature space. This reduced feature set is then used to train and test a two-layer backpropagation neural network. The classification performance is demonstrated on several data sets that contain synthesized reverberation with 12-dB signal-to-reverberation ratio (SRR). The results show excellent correct classification rate of over 92% at a very low false alarm rate of less than 8% for the single aspect classification case. The generalization capability of the neural-network classifier is also demonstrated by performing statistical analysis on a large number of test cases. Results on an expanded test data set, where each backscattered data was corrupted with ten different reverberation realizations, demonstrated excellent robustness. Further improvement in the performance is obtained by fusing backscattered data from multiple aspects.

II. ACOUSTIC BACKSCATTERED DATA AND PREPROCESSING

The data set for this study was provided by the Coastal Systems Station (CSS) in Panama City, FL. It was collected in a

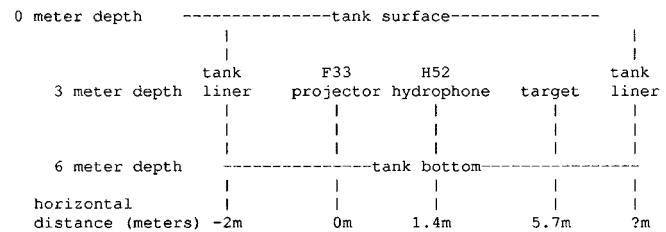


Fig. 1. Experimental setup.

controlled environment, namely, a water tank which is about 6 m deep. The experimental setup is shown in Fig. 1.

There are six objects: two mine-like targets and four non-targets. The mine-like targets are: a blunt-bullet-shape metallic object filled with an explosive simulant, and a truncated-cone-shape plastic object filled with an explosive simulant. The non-mine-like targets are: a 50-gallon water filled drum, an irregular shaped limestone rock; a somewhat smooth granite rock and a wooden log that was saturated in water. In the sequel, these six objects will be referred to as objects 1–6, respectively. Each object was suspended into the tank and as viewed from the top, rotated counterclockwise about the vertical axis. The cylindrical objects were suspended horizontally in the tank. Each object was insonified at 72 aspect angles from 0 to 355° with 5° increments. The transmit signal was a linear frequency modulated (LFM) pulse that was 1 ms in duration and covered the frequency band of 20–60 KHz.

A good database plays a critical role in proving out a classification methodology. It was recognized that the acoustic backscattered database, as it stands, is limited in its usefulness for developing and comparing various classification algorithms. However, as will be discussed next, these limitations can be removed by 1) appropriate signal scaling; 2) adding realistic noise; and 3) removing certain artifacts. These modifications result in an effective database that can be used to study a very challenging classification problem.

The classification process is usually too computationally intensive to apply to backscattered data collected over a large area. Therefore, some type of detection process is usually used to screen the data for potential targets. In this application, it is envisioned that a detection screener would send to the classification stage only the data that appears to come from objects that have mine-like acoustic target strengths. Consequently, target strength would no longer be a significant classification feature. The challenging problem for the classifier is to identify features that will eliminate the false targets that have target strengths similar to the mine. Accordingly, a procedure was adopted to scale the signals in the existing database so that the 90-percentile target strength of each object was the same.

In backscatter classification, the most difficult type of noise to overcome is bottom, surface, and volume reverberation. By linear acoustics, which describes the dominant physics of this problem, reverberation and the target signature are in the same frequency band as the transmit signal. This makes reverberation difficult to remove without distorting the target's signature. Therefore, an important aspect of the classifier's evaluation is its robustness against reverberation. Consequently, a procedure was developed for synthesizing and adding reverberation-like

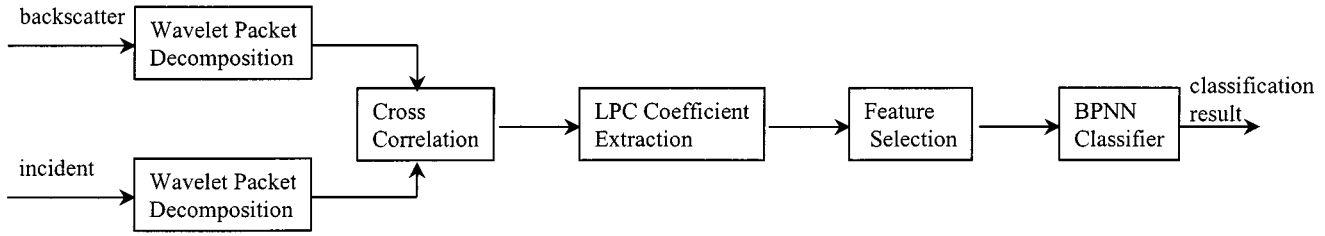


Fig. 2. Wavelet packets-based classification systems.

noise to the backscattered signals. Furthermore, to better evaluate the robustness, the classifier was evaluated on an expanded database formed by adding multiple realizations of reverberation to each backscattered signal.

Due to the experimental setup in the small test pool, the presence of the secondary reflections or multipath in the backscattered signals is inevitable. These reflections normally arrive at the hydrophone after the main target return. There is, however, another artifact caused by the reflections between the projector and the hydrophone that can temporally overlap with the main return depending on the aspect angle. Thus, if care is not taken, these artifacts can be inadvertently taken as features to discriminate between target types. This may result in classification performance that seems very good, but is actually highly biased and unreliable. Additionally, the data acquisition electronics time-gates and captures each segment of the backscattered data for recording. It so happened that the time where the dominant target signature starts in the recorded backscattered signal is somewhat unique between targets and can be used to discriminate between some of them.

A procedure was developed to remove the effects of the above-mentioned artifacts prior to the feature extraction process. This procedure mimics a detection process that uses windowing in the matched-filter domain and then performs inverse filtering to recover the clean backscattered signal. This method exploits the fact that the artifacts are more separated from the main target return in the matched-filtered domain than in the time domain. This is described in Step 3) of the following preprocessing procedure.

The entire preprocessing procedure consist of the following steps that are implemented in order:

- Step 1) *Scaling*: Generate the “scaled” signal from the raw acoustic backscattered data to compensate for biases and amplifier and filter gain settings used in the recording of the data.
- Step 2) *Downsampling*: The original sampling rate, 2 MHz, is significantly higher than the signal bandwidth. The backscattered signals were down-sampled by a factor of four. This makes the signals of size 2048 points without losing important information.
- Step 3) *Multipath and Artifact Removal Procedure*: The process involves cross-correlating the backscattered signal at every aspect with the incident and finding the location of the peak x in the matched filter output. A window covering the range of $[x - \text{left}; x + \text{right}]$ is then extracted. In this study

we used $\text{left} = 111$ and $\text{right} = 400$ to extract a window of size 512 points. This segmented signal was then zero-padded to its original size and inverse-filtered using

$$H(k) = \frac{X(k)}{|X(k)|^2 + e} \quad (1)$$

where $X(k)$ is the DFT of the incident signal and $e = 0.0025 \text{Max}(|X(k)|^2)$ is added to avoid singularity problem of the inverse filter. The result of this procedure is a clean backscattered signal free of artifact and multipath effects.

- Step 4) Scale each target so that each has a similar target strength. Divide each target’s 72 backscattered signals by SRA where the SRA is the signal reference amplitude, which is the largest amplitude that is less than 90% of the peak amplitude of all aspects for the target of interest [15].
- Step 5) *Reverberation Effects*: Add synthesized reverberation to all the scaled backscattered signals. The synthesized reverberation is generated [15] by convolving white noise realizations with the incident signal, and then scaling to the desired SRR using

$$\text{SRR (dB)} = 20 \log_{10} \left(\frac{1.0}{1.414 * \text{rms of reverberation}} \right). \quad (2)$$

An SRR of 12 dB was used in this paper as this corresponds to typical operating condition.

III. FEATURE EXTRACTION AND CLASSIFICATION

The block diagram of the high-level processing system is shown in Fig. 2. As can be seen, there are several subsystems that perform wavelet packets decomposition, cross-correlation process, LPC coefficients extraction in each subband, feature selection/reduction and neural-network-based classification. These are briefly described in the following sections.

A. Wavelet Packet-Based Feature Extraction

Wavelet packets are closely related to subband decomposition using filter banks [20]. The idea is that given a signal, a pair of low-pass and high-pass filters is used to yield two sequences capturing different frequency subband features of the original signal. These sequences are then subsampled by a factor of two. This forms one level of wavelet decomposition. This process can be repeated to partition the frequency spectrum into smaller frequency bands for resolving different subtle features

while localizing the temporal information. Wavelet packets provide a signal-dependent subband decomposition using a filter bank with a rich menu of orthonormal basis functions suited to the temporal-spectral signal properties.

Due to the variations in the characteristics of different objects, e.g., elasticity, geometrical shape, and size, frequency characteristics of the backscattered signals for these objects tend to be different. Using wavelet packet decomposition, each subband extracts certain tonal information of the signal depending on the frequency range and content of the backscattered signal. This is similar to the filter bank model proposed in [16] for the pre-processing and encoding of information in the cochlea of bats. The filter bank in [16] consists of several band-pass filters each tuned to pick up certain frequency information.

Unlike the discrete wavelet transform [20] which gives octave band decomposition of the frequency axis, the wavelet packet decomposition tree can have any arbitrary structure depending on the signal properties. The full tree which divides the frequency axis into equal-bandwidth components obviously yields a linear division of the spectrum similar to the short-time Fourier transform (STFT) [20]. Generally, a criterion such as energy or entropy is selected and used to construct the best tree structure for the decomposition of a particular class of signals. In this paper, based on the reason that target information typically resides in the same frequency band as the transmit signal, only those subbands within the transmit signal bandwidth are selected. To avoid phase distortion that might be caused as result of using nonlinear phase filters in the filter bank while at the same time ensuring the orthogonality of the signal representation, fourth-order Symlet wavelet [20] is used in each level. This wavelet has a nearly symmetrical shape and at the same time offers orthogonality property. The incident signal is also decomposed using the same tree structure.

In each selected subband, the cross-correlation between the incident signal and the backscattered signal is performed as shown in Fig. 2. Then, the LPC scheme [21] commonly used for speech coding and recognition applications is applied to the matched filtered result in each subband. A fourth-order linear autoregressive (AR) model

$$a_0y(n) = x(n) - \sum_{i=1}^4 a_iy(n-i) \quad (3)$$

is employed for this signal representation and the corresponding coefficients a_i 's are used as features for classification.

B. Feature Selection and Reduction

From the statistical decision theory, it is well known that the probability of classification error decreases when additional measurements are taken into consideration. However, this is true for infinite sample sets for which the estimation errors of the system parameters can be ignored. Bellman's curse of dimensionality [22] shows that the classification performance will not be improved as more features are added. That is, more features would correspond to more classification system parameters to be estimated, and thus in the case of only finite training samples, this leads to increased estimation error. Due to the estima-

tion error, the system can be so finely tuned to the training set that it lacks generalization capability.

A feature selection scheme is, therefore, needed to choose the most useful information from the complete feature space to form a feature vector in a lower dimensional space, and remove any redundant and irrelevant information that may have detrimental effects on the classification performance.

To select an appropriate set of features, a criterion function can be used to provide the discriminatory power of the individual features. Two criterion functions based on the Bhattacharyya distance measure are considered here.

- 1) *Correlated-Features Assumption*: The candidate features are assumed to be correlated. In this case, the criterion function [23] is defined by

$$D_B(i, j) = \frac{1}{8}(\mu_j - \mu_i)^t \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mu_j - \mu_i) + \frac{1}{2} \ln \frac{\left| \frac{1}{2}(\Sigma_i + \Sigma_j) \right|}{(|\Sigma_i| |\Sigma_j|)^{1/2}} \quad (4)$$

where μ_i and μ_j are the mean vectors for classes i and j and Σ_i , Σ_j are the corresponding covariance matrices. Various schemes, such as sequential-based methods [23] and genetic algorithms [24], can then be applied to provide efficient feature selection.

- 2) *Uncorrelated-Features Assumption*: The candidate features are assumed to be uncorrelated. In this case, the Fisher discriminant function [22] is used to evaluate the distance between every two classes for each single feature, i.e.,

$$D_B(i, j) = \frac{|\mu_i - \mu_j|^2}{\sigma_i^2 + \sigma_j^2} \quad (5)$$

where σ_i^2 is the variance of the features in class i .

In this paper, this scheme is adopted because of its simplicity and better numerical stability since the computation of the inverse of the covariance matrices in (3) may lead to numerical problems, especially when the matrices are estimated based on a limited data set.

The criterion function (4) is computed for each feature in the complete feature set which contains 30 LPC coefficients. Those features for which D_B values are relatively larger than the others are selected to form the reduced feature vector for the subsequent classification. This process led to 22 LPC features which possess high discriminatory ability.

C. Neural-Network Classifier

A two-layer backpropagation neural network (BPNN) [25] with structures 22-42-2 is used for classification using the reduced feature vectors. Classification decision is based upon discriminating mine-like (objects 1 and 2) from nonmine-like (objects 3–6) objects. The neural network used fast backpropagation learning with momentum term $\alpha = 0.9$ and an adaptive learning rate. Momentum decreases backpropagation's sensitivity to small details in the error surface in order to prevent the network from getting stuck in shallow minima, and the adaptive

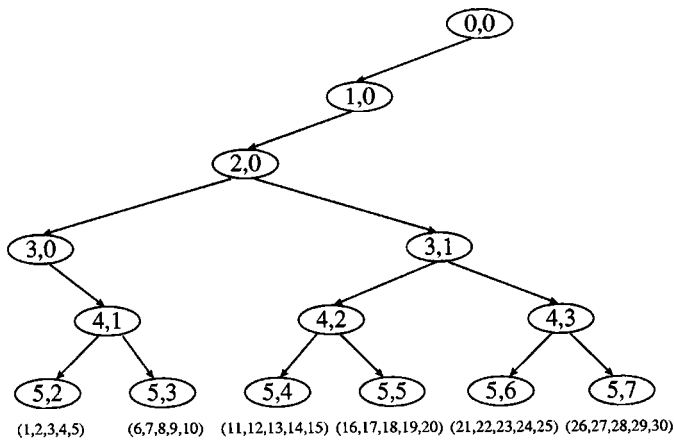


Fig. 3. Pruned tree structure of wavelet packet decomposition.

learning rate attempts to keep the learning step size as large as possible while keeping the learning stable.

IV. TEST RESULTS

A. ROC Curves of the Classifier

The backscattered signals were decomposed using a five level full-tree wavelet packets decomposition which corresponds to a STFT with total of $2^6 - 1$ subbands. Sym4 filter [20] which provides orthogonality and nearly symmetrical impulse response was chosen for this decomposition. This guarantees a nearly linear phase characteristic which is very important for the speech and acoustic signal processing applications. The subbands at level 5 that reside within the bandwidth of the transmit signal, labeled along the frequency axis from (left-to-right) [5, 2] to [5, 7] were chosen. The “pruned tree structure” in Fig. 3 shows these six selected subbands. The fourth-order LPC coefficients of the matched filtered signals at these subbands were calculated to form the complete feature vector with 30 elements.

The whole data set was divided into two sets, one for training and one for testing. Even aspect angles were used for training, and odd aspect angles were used for testing. The training and testing data sets were generated based on the preprocessed data by adding synthesized reverberation as described before. Thus, we have a total of 216 training samples and 216 testing samples for each realization. The feature reduction scheme in Section III-B was then applied and the values of the Fisher discriminant functions for the features were calculated based upon the training data set. This resulted in the selection of 22 features (out of 30) with high discriminatory ability. These correspond to LPC coefficients a_1, a_2, a_3, a_4 in subband [5, 2], a_1, a_2, a_3, a_4 in [5, 3], a_3, a_4 in [5, 4], a_0, a_2, a_3 in [5, 5], a_0, a_1, a_2, a_3, a_4 in [5, 6], and a_0, a_1, a_2, a_3 in [5, 7].

These features were used to form the reduced feature vector which is then used as the input to the two-layer BPNN classifier. Several BPNN structures were tested with different number of hidden layer neurons and the optimal performance was achieved for the network with 22-42-2 structure.

The receiver operating characteristic (ROC) curve [26] is a good measure when performance of different schemes needs

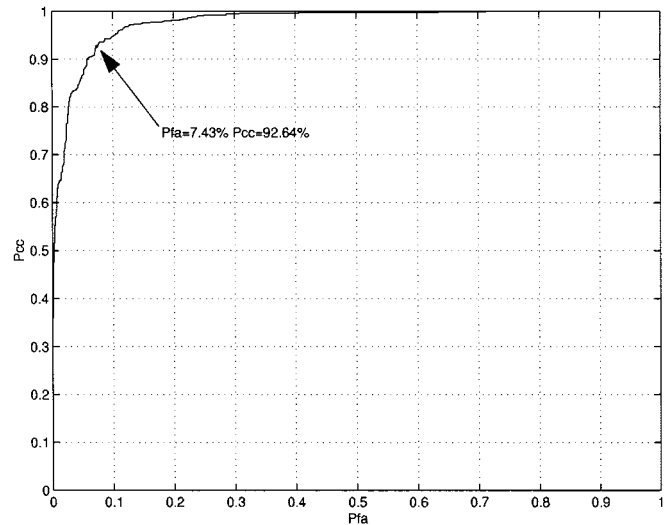


Fig. 4. ROC curve for the testing data (ten realizations).

TABLE I
CONFUSION MATRIX, TEN REALIZATIONS
(2160 SAMPLES—720 TARGETS, 1440 NONTARGETS)

True Class	Chosen class	
	mine-like targets	non-targets
mine-like targets	667(92.64%)	53(7.43%)
non-targets	106(7.43%)	1334(92.64%)

to be compared. The ROC curve is the plot of the conditional probability of the correct classification versus conditional probability of false-alarm responses. The result [26] stated that the area under the ROC curve gives the probability of the correct classification. To test the performance of the classifier, more testing data were generated by adding different realizations of synthesized reverberation to the odd aspect angle backscattered signals. Fig. 4 shows the ROC curve obtained based on 2160 cases or ten different noisy realizations. As can be observed from this ROC curve, the classifier provides excellent classification results based upon only the data of single aspect of the sonar. As a comparison, our proposed system provides over 92% correct classification for less than 8% false alarm rate which is an improvement over results reported elsewhere [14], [15], [17]–[19]. To further improve the performance of the proposed system, the classification results of multiple aspects of the sonar will be fused together in Section IV-D.

The confusion matrix generated based on the classification results on these ten realizations (2160 samples) is given in Table I. The threshold for the decision making is the one which makes the correct classification rate (P_{cc}) = 1; false alarm rate (P_{fa}), i.e., the point where misclassification rate is equal to the false-alarm rate. This corresponds to the “knee” of the ROC curve marked on Fig. 4.

As in [17], in order to observe the classification performance for each individual target, a circular diagram was generated to show the classification error locations versus aspect angles for all the six objects. Fig. 5 shows the classification error locations and the associated frequencies represented in different gray levels for ten noisy realizations. Consequently, the number of

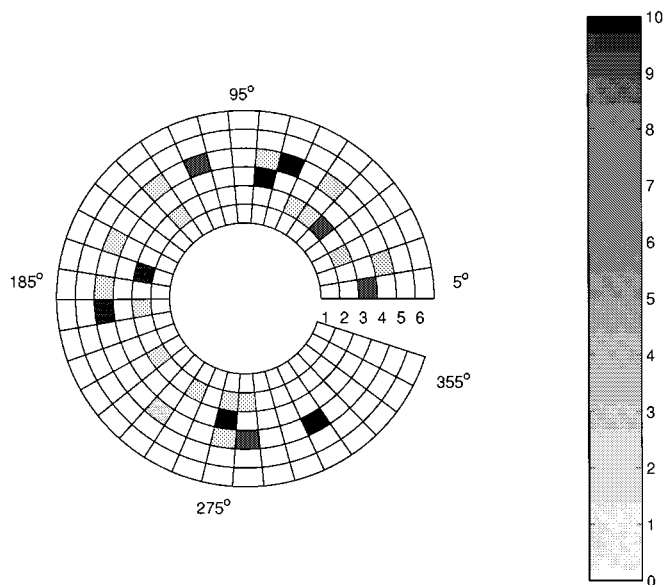


Fig. 5. Error locations versus aspect angles.

gray levels representing the frequency of classification error at a particular aspect angle varies from zero to ten. Note that every ring corresponds to a particular object and every grid on the ring represents an aspect angle. In the counter-clockwise direction, the aspect angles start from 5 to 355° with 10° step since there are 36 aspect angles for every object in the test set. The classification decision was made based on the threshold corresponding to the “knee” of the ROC curve. This plot indicates perfect classification rate for the cylindrical mine target (object 1) that is classified in all cases without any misclassifications. This is also true for nonmine objects 5 and 6 (i.e., granite rock and wooden log). Most of the misclassifications happen for the second target and the limestone rock owing to their irregular shape and similarity of the features.

The robustness of the selected features and generalization capability of the neural-network classifier have been carefully studied next since they are critical issues for a classification system.

B. Robustness of the Selected Features

To get a simple view on the statistical distribution of the selected features, the mean and variance values of these features were computed for the two classes, i.e., targets and nontargets, and for both the training and testing sets (ten realizations). In order to show the statistical distribution for each feature independent of the statistics of the other features, a normalization was performed by dividing each feature by the maximum value for that particular feature. This maximum value was determined based upon the entire data set which included target and nontarget cases in both training and testing sets. The plot in Fig. 6 shows the mean and variance for each normalized feature computed for target (circle) and nontarget (asterisk) cases and for the training (left) and testing (right) sets. This plot clearly indicates that mean values of the training and testing feature sets are very close to each other. Additionally, the mean values for target and nontarget classes are nicely separated, while the range of their

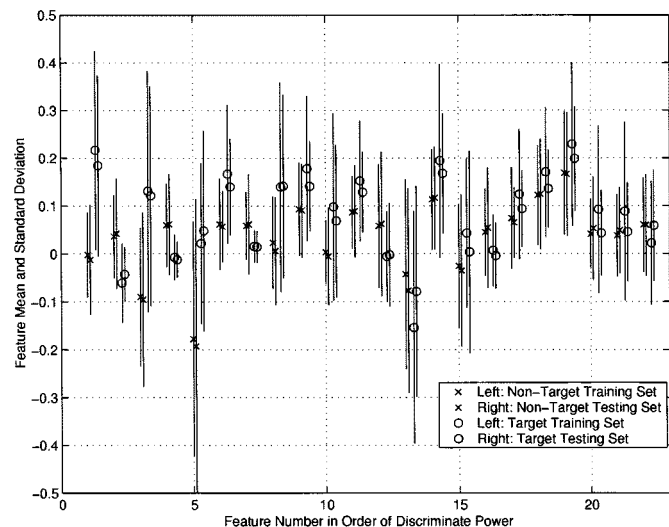


Fig. 6. Mean and standard deviation of every selected feature.

distributions do not overlap significantly for most of the principal features.

Based on the above analysis, we know that the selected features are well mapped with the two classes. The other aspect that may affect the classification performance is the classifier which utilizes the discriminant power of the features and the learning process to make the classification decision. The robustness and generalization of our classifier are studied next.

C. Generalization and Robustness of the Classifier

The same training set was used to train the network with different random initial weights and in order to test the consistency in the network’s performance. The initial weights were uniformly distributed over a small range of values which were determined based on the minimum and maximum values of the training set data. The network training was repeated for ten trials and Fig. 7 shows the corresponding classification performance for each testing trial based on one realization only. With the exception of two trials which failed to converge, the network’s performance was stable and the classification rate for 10% false alarm rate varied with only 10% range as shown in Fig. 7.

More testing cases were also generated to test the generalization ability of the classifier. Fig. 8 shows the classification result on another testing set of ten realizations. Again the system provided around 92% correct classification for only 8% false alarm rate. This demonstrates good generalization ability of the classifier.

We further studied the error rate and classification rate statistics for a larger number of samples. For a given set of n input pattern vectors, the number of misclassifications m can be viewed as a random variable resulting in a randomly varying empirical error rate $\epsilon_{\text{empirical}} = \frac{m}{n}$. In [27], irrespective of the pattern source and the type of the classifier, it is shown that the random variable $\{m\}$ follows a binomial distribution whose probability density function can be given by

$$\Pr(\epsilon_{\text{empirical}}|n, \epsilon) n \binom{n}{n\epsilon_{\text{empirical}}} \epsilon^{n\epsilon_{\text{empirical}}} \cdot (1 - \epsilon)^{n(1-\epsilon_{\text{empirical}})} \quad (6)$$

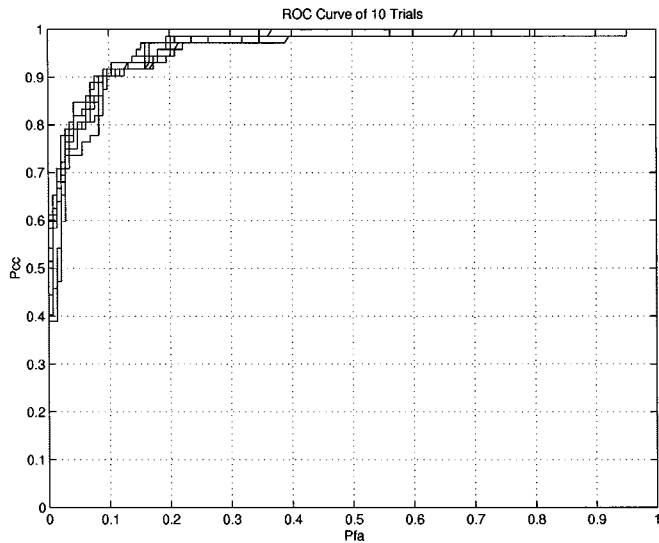


Fig. 7. ROC curves of several different trials.

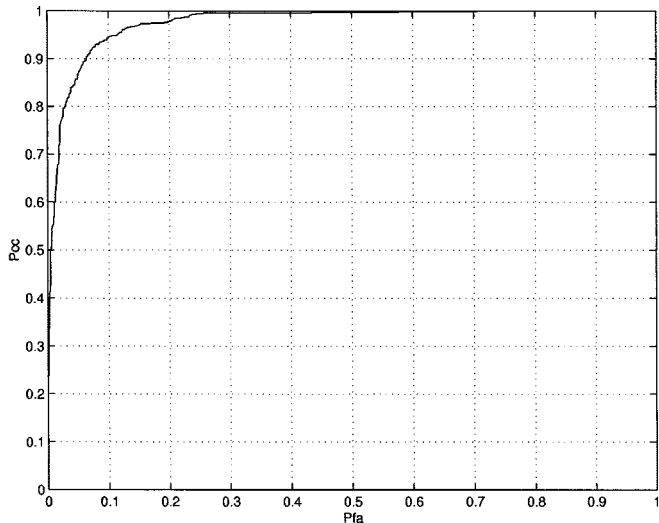


Fig. 8. ROC curve—new testing data (ten realizations, SRR = 12 dB).

where ϵ is the true error rate. Thus, we can evaluate the mean and variance of the empirical error rate $\epsilon_{\text{empirical}}$, i.e., $E\{\epsilon_{\text{empirical}}\} = \epsilon$ and $\text{var}\{\epsilon_{\text{empirical}}\} = (1/n)\epsilon(1 - \epsilon)$. As n increases, the variance vanishes and the empirical error rate converges to the true error rate, ϵ . The required value for n can be obtained [27] from $n_{\text{required}} = (1 - \epsilon)/(\tau^2\epsilon)$, where τ is the tolerance.

To estimate the error rate of the neural-network classifier, 50 different Monte Carlo trials were performed. For each trial, a different testing set with different reverberation realizations was used. As before, each set contained 216 backscattered signals corresponding to odd aspect angles and a different 12 dB synthesized reverberation sequence. Based on the above analysis and the expected error rate of the classifier, the number of cases is large enough to provide an accurate estimate of the error rate. Fig. 9(a) shows the histogram of 50 individual error measurements for the 50 trials under the assumption that false positive and negative errors have the same weighting, i.e., the false alarm

and misclassification errors are summed up to a total classification error. The classification decision is made based on the threshold at the “knee” point of the ROC curve. The legend provides the mean and standard deviation of the error rates.

Fig. 9(b)–(d) shows the histograms of the 50 individual correct classification rate measurements under false alarm rates of 5, 10, and 15%, respectively. The legend gives the mean and standard deviation of the correct classification rates. The larger standard deviation at low false alarm rates is obviously attributed to the sharp slope of the ROC curve at these points. These figures further demonstrate the robustness and generalization characteristics of the proposed classification system.

D. Multiaspect Fusion Results

The classification results of the classifier based on the single aspect were fused together in order to further improve the overall classification performance. This is due to the fact that even though some target and nontarget cases may not be correctly classified at a specific aspect, it is very likely that they will be correctly classified at the neighboring aspects. Thus, the combined results of several aspects may lead to some improvements in the classification performance. In this study, similar to [17], we have used sequences of three aspects separated by certain degrees. The reason is that in actual mine-hunting scenarios an object is declared as a potential mine if strong indications exist in three to four sonar pings. The 3-aspect sequences are chosen so that overlap between them is allowed, e.g., $\{5, 35, 65\}$, $\{35, 65, 95\}$, \dots , $\{355, 25, 55\}$ for 30° aspect separation interval. A study was also carried out to examine the effects of varying the amount of aspect separation interval on the classification performance of the fusion system. The performance plot was generated when this separation interval was varied from 10° to 40° in 10° steps and the optimum case was selected. For each aspect, the two outputs of the classifier for target and nontarget are applied to the fusion system where the final decision is made based upon the classification results of the three aspect angles. Both linear and nonlinear fusion schemes were studied. The linear fusion scheme is based on the least squares (LS) method, while the nonlinear fusion scheme utilizes a backpropagation neural network with inherent nonlinearity.

1) *Linear Fusion Scheme*: Let us define the output matrices of the neural-network classifier (recall the network has two output nodes) corresponding to the three aspects and for all the training data set with N samples by Y_1 , Y_2 , and Y_3 (Y_i are $N \times 2$ and $Y = [Y_1, Y_2, Y_3]$ be the complete data matrix of size $N \times 6$. This is applied as an input to the fusion system. The weight vector of the linear fusion system is denoted by $W = [w_1, \dots, w_6]^T$ and the desired output vector of size $N \times 1$ by T . Now, given the input data matrix Y and the desired output vector T , to obtain the optimal weight vector, we solve the following least square problem:

$$W_{\text{opt}} = \arg \min_W \|Y^T W - T\|^2 \quad (7)$$

which gives [28]

$$W_{\text{opt}} = (Y^T Y)^{-1} Y^T T. \quad (8)$$

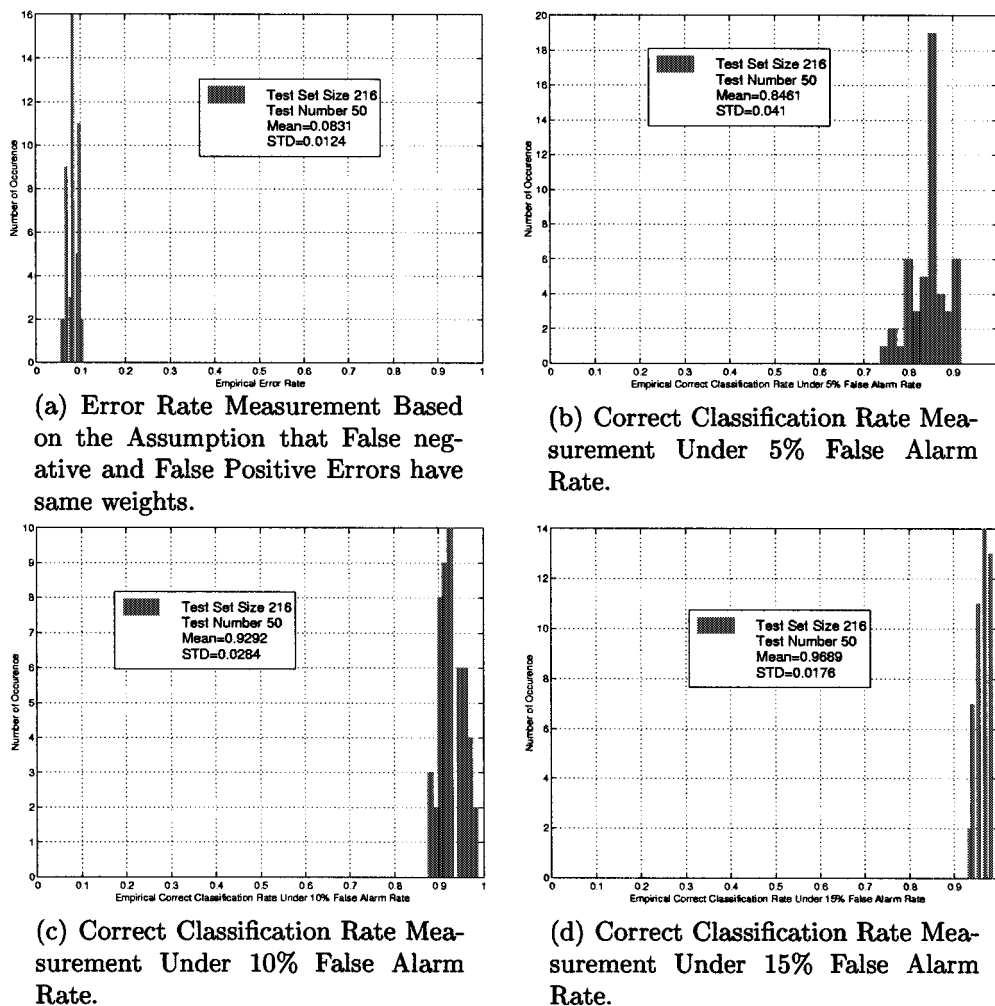


Fig. 9. Error and correct classification rates measurements.

The single-aspect classification results of two different realizations based on the even aspect angle data set with different synthesized reverberation realizations were arranged in 3-aspect sequence form, e.g., $\{0, 30, 60\}, \dots, \{350, 20, 50\}$ for 30° separation case. This data set (432 samples) was then used to find the optimal weight vector W_{opt} . Once the training is completed and the optimal weight vector is obtained, the final classification result, O , for a testing data matrix Ψ , can be obtained using

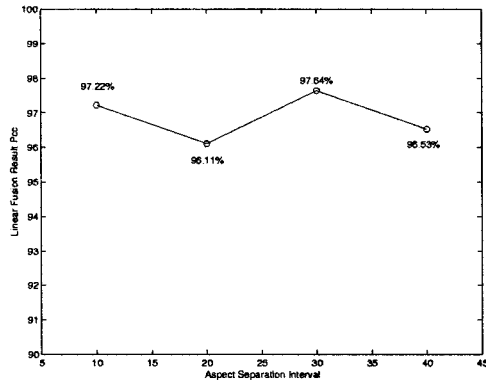
$$O = \Psi W_{opt}. \quad (9)$$

The same testing data set (ten realizations) used in Section IV-A was then applied to this linear fusion system. As shown in the plot in Fig. 10(a) the best performance for the linear fusion system was achieved for 30° aspect separation interval. Fig. 11(a) shows the corresponding ROC curve of the final classification performance of this 3-aspect linear fusion system. The P_{cc} and P_{fa} at the knee of this ROC curve are found to be 97.64, 2.36%, respectively, which indicate about 5% improvements over those of the single-aspect case.

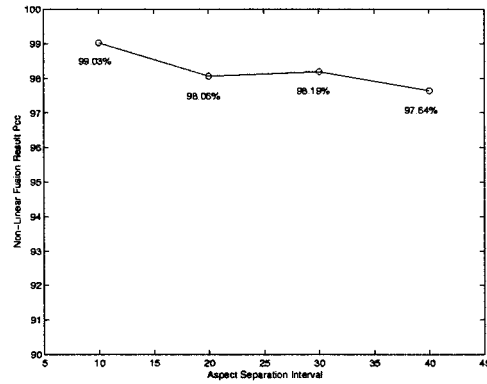
2) *Nonlinear Fusion Scheme*: A two layer backpropagation neural network was designed to perform the nonlinear 3-aspect fusion. The network had six input nodes corresponding to the classification outputs of the three aspects, four hidden layer

nodes, and two output nodes for final decision. The same validation set used in the linear fusion system which contained 432 samples was used to train the neural network. The network reaches the sum squared error (SSE) goal of ten within 200 epochs. The network was then tested on the same testing data set used in Section IV-A. In this case, the best performance was obtained for 10° separation interval as shown in Fig. 10(b). Fig. 11(b) shows the corresponding ROC curve of the 3-aspect fusion results for this two-layer fusion network. The P_{cc} and P_{fa} at the knee of this ROC curve are found to be 99.03, and 0.97%, respectively, which exhibit about 7% improvements over those of the single-aspect case.

Finally, Fig. 12 shows the classification error locations versus aspect angles for this nonlinear fusion system and for ten noisy realizations. In this case, the gray levels indicate errors of the system as a result of fusing the current aspect angle and its two previous aspects which lag 10 and 20° , respectively. Again the classification decision is made based on the threshold at the knee of the corresponding ROC curve. Comparing to the single aspect results in Fig. 5, it is evident that the nonlinear multispect fusion system significantly improved the overall classification performance. More specifically, substantial improvements are obtained for objects 2 and 4. In addition, objects 1, 3, 5, and

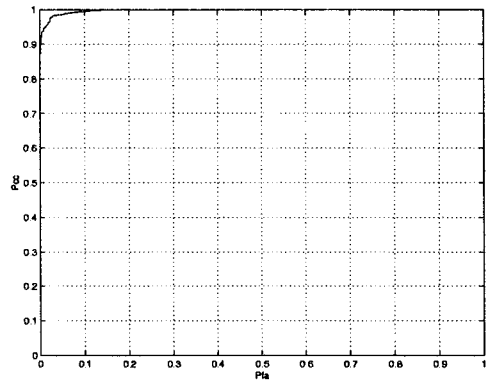


(a) Pcc for Linear 3-aspect Fusion System with different aspect interval.

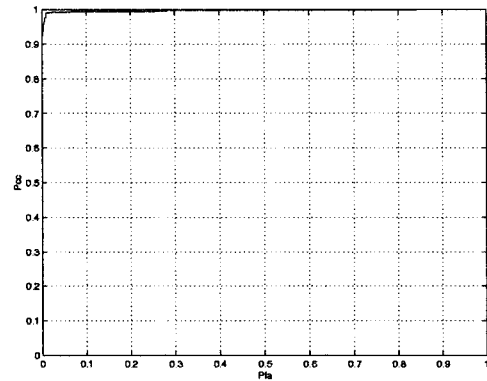


(b) Pcc for Non-linear 3-aspect Fusion System with different aspect interval.

Fig. 10. P_{cc} of the 3-aspect fusion results with different aspect interval on the testing data (ten realizations).



(a) ROC Curve of Linear 3-aspect Fusion System.



(b) ROC Curve of Non-linear 3-aspect Fusion System.

Fig. 11. ROC curve of the 3-aspect fusion results on the testing data (ten realizations).

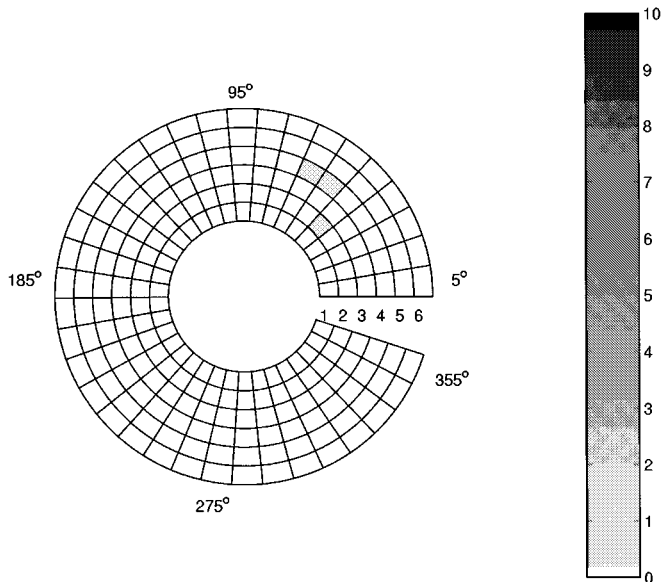


Fig. 12. Error locations versus aspect angles—after the nonlinear 3-aspect fusion.

6 are now correctly classified at all aspect angles in the testing data.

In summary, all these results point to this important observation that the multispect fusion indeed improves the classification performance and that the nonlinear fusion scheme outperforms the linear one.

V. CONCLUSION

A wavelet packets-based scheme for classification of underwater targets from the acoustic backscattered signals was developed in this paper. Wavelet packets were used to yield a “multi-look” analysis of the backscattered signals in various subbands. The fourth-order LPC coefficients of the signals at different subbands were used as the features for classification. In addition, a statistical-based feature selection scheme was applied to choose the most pertinent features and at the same time reduce the dimensionality of the feature space. A backpropagation neural-network classifier was then used to perform the discrimination between targets and nontargets based upon a reduced set of features. The ROC curve generated based on the results of

ten different realizations indicated excellent performance of the classifier for this application. In addition, the performance analysis of the neural-network classifier demonstrated the robustness and good generalization capability of the system. Finally, multispect fusion was performed to show that great improvement of the classification performance of the system can be obtained.

REFERENCES

- [1] J. P. Sessarego *et al.*, "Time-frequency analysis of signals related to scattering problems in acoustics—Part I: Wigner–Ville analysis of echoes scattered by spherical shell," in *Wavelets*, New York: Springer-Verlag, 1987.
- [2] J. Wilbur and S. G. Kargl, "Application of wavelets to acoustic resonance-elastic targets surrounded by biologics," in *Proc. ICASSP'93*, Minneapolis, MN, Apr. 1993, pp. IV.492–IV.495.
- [3] B. Telfer, H. Szu, and G. Dobeck, "Adaptive wavelet classification of acoustic backscatter," *Proc. SPIE*, vol. 2242, pp. 661–668, 1994.
- [4] M. De Billy, "Determination of the resonance spectrum of elastic bodies via the use of short pulses and Fourier transform theory," *J. Acoust. Soc. Amer.*, vol. 79, no. 2, Feb. 1986.
- [5] M. R. Azimi-Sadjadi, J. Wilbur, and G. Dobeck, "Isolation of resonances in acoustic backscatter from elastic targets using adaptive estimation schemes," *IEEE J. Oceanic Eng.*, vol. 20, pp. 346–353, Oct. 1995.
- [6] M. Hasan and M. R. Azimi-Sadjadi, "A modified block FTF adaptive algorithm with applications to underwater target detection," *IEEE Trans. Signal Processing*, vol. 44, pp. 2172–2185, Sept. 1996.
- [7] M. Hasan, M. R. Azimi-Sadjadi, and G. Dobeck, "Multiple time delay estimation using new spectral estimation schemes," *IEEE Trans. Signal Processing*, vol. 46, pp. 1580–1590, June 1998.
- [8] M. R. Azimi-Sadjadi, S. Charleston, J. Wilbur, and G. Dobeck, "A new time delay estimation in subbands for resolving multiple specular reflections," *IEEE Trans. Signal Processing*, vol. 46, pp. 3398–3403, Dec. 1998.
- [9] C. Yuan, M. R. Azimi-Sadjadi, J. Wilbur, and G. Dobeck, "Underwater target detection using multichannel subband adaptive filtering and high order correlation schemes," *IEEE J. Oceanic Eng.*, vol. 25, pp. 192–205, Jan. 2000.
- [10] R. P. Gorman and T. J. Sejnowski, "Learned classification of sonar targets using a massively parallel network," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 7, July 1988.
- [11] B. A. Telfer, H. H. Szu, and G. Dobeck, "Adaptive time-frequency classification of acoustic backscatter," in *Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr.*, Orlando, FL, Apr. 1995.
- [12] G. Goo and W. W. L. Au, "Detection and identification of buried objects in shallow water," in *Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr.*, vol. 2765, Orlando, FL, Apr. 1996, pp. 201–214.
- [13] P. H. Carter and G. Dobeck, "Classification of acoustic backscatter using the generalized target description," in *Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr.*, vol. 2765, Orlando, FL, Apr. 1996, pp. 190–200.
- [14] N. Intrator, Q. Q. Huynh, and G. Dobeck, "Feature extraction from acoustic backscattered signals using wavelet dictionaries," in *Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr.*, vol. 3079, Orlando, FL, Apr. 1997, pp. 183–190.
- [15] L. L. Burton and H. Lai, "Active sonar target imaging and classification system," in *Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr.*, vol. 3079, Orlando, FL, Apr. 1997, pp. 19–33.
- [16] P. A. Saillant, J. A. Simmons, S. P. Dear, and T. A. McMullen, "A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram and transformation receiver," *J. Acoust. Soc. Amer.*, vol. 94, pp. 2691–2712, 1993.
- [17] G. A. Carpenter and W. W. Streilein, "ARTMAP-FTR: A neural network for fusion target recognition with application to sonar classification," in *Proc. 1998 SPIE Int. Symp. Aerospace/Defense Sensing Contr.*, Orlando, FL, Apr. 1998, pp. 342–356.
- [18] Q. Huang, M. R. Azimi-Sadjadi, and G. Dobeck, "Underwater target classification using wavelet packets and neural networks," in *Proc. 1998 IEEE Int. Conf. Neural Networks*, vol. 1, Anchorage, AK, May 1998, pp. 177–182.
- [19] M. R. Azimi-Sadjadi, Q. Huang, and G. Dobeck, "Underwater target classification using multispect fusion and neural networks," in *Proc. 1998 SPIE Int. Symp. Aerospace/Defense Sensing Contr.*, Orlando, FL, Apr. 1998, pp. 334–341.

- [20] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [21] L. Rabiner *et al.*, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [22] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [23] P. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [24] C. H. Chen, L. F. Pau, and P. S. P. Wang, *Handbook of Pattern Recognition and Computer Vision*. Singapore: World, 1993.
- [25] S. Haykin, *Neural Network*. New York: McMillan, 1994.
- [26] M. L. Meistrell, "Evaluation of neural-network performance by receiver operating characteristic (ROC) analysis: Examples from biotechnology domain," *Comput. Methods Programs Biomed.*, vol. 32, pp. 73–80, 1990.
- [27] J. Schurmann, *Pattern Classification, A Unified View of Statistical and Neural Approaches*. New York: Wiley, 1996.
- [28] J. A. Benediktsson, J. R. Sveinsson, O. K. Ersoy, and P. H. Swain, "Parallel consensual neural network," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 54–64, Jan. 1997.



Mahamood R. Azimi-Sadjadi (S'81–M'81–SM'89) received the M.S. and Ph.D. degrees from the Imperial College of Science and Technology, University of London, U.K., in 1978 and 1982, respectively, both in electrical engineering with specialization in digital signal/image processing.

He is currently a Full Professor at the Electrical and Computer Engineering Department at Colorado State University (CSU), Fort Collins. He is also serving as the Director of the Multisensory Computing Laboratory (MUSCL) at CSU. His main areas of interest include digital signal and image processing, target detection, classification and tracking, adaptive filtering and system identification, and neural networks. His research efforts in these areas resulted in more than 100 journal and refereed conference publications. He is the coauthor of the book *Digital Filtering in One and Two Dimensions*, (London, U.K.: Plenum, 1989).

Dr. Azimi-Sadjadi is the recipient of the 1999 the ABELL Teaching Award, 1993 ASEE-Navy Senior Faculty Fellowship Award, 1991 CSU Dean's Council Award, and 1984 DOW Chemical Outstanding Young Faculty Award. He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS.



De Yao received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, July 1992. He is currently a Graduate Research Assistant in the Electrical and Computer Engineering Department of Colorado State University, Fort Collins, CO, working toward the M.S. degree.

He worked as an Electronics Engineer in ShanDong Telecommunications Inc., ShanDong, China, from August 1992 to April 1996. From April 1996 until January 1998, he was with Goldtron Telecommunications Pte. Ltd., Singapore, as a Hardware Engineer. His research interests include digital signal processing, classification, and communication techniques. He is presently working at Air Net Communications Corporation, Melbourne, FL.



Qiang Huang received the B.Sc. degree in electronics engineering from Peking University, China, in 1995 and the M.Sc. degree in electrical and computer engineering from Colorado State University, Fort Collins, CO, in 1998.

From 1997 to 1998, he was a graduate Research Assistant at Digital Image and Signal Processing Laboratory, Colorado State University, where he was engaged in research of target classification and neural networks. He is currently working at Cisco Systems, Inc.



Gerald J. Dobeck received the B.S. degree in physics from the University of Massachusetts, Amherst, in 1970 and the M.S. and Ph.D. degrees in electrical engineering from the University of South Florida, Tampa, FL, in 1973 and 1976, respectively.

Since 1976, he has been employed at the Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL. His current research interests include automatic detection and classification of targets in cluttered environments from synthetic/real aperture sonar imagery, the

echo structure of acoustic returns, underwater electro-optic imagery, and gradiometer/magnetometer signals. He is the Project Leader of the Sensor Signal and Image Processing Project under the Office of Naval Research 6.2 Mine Countermeasures program. He is the Technical Leader on the development of automated detection and classification algorithms of naval sea mines from sonar imagery, broadband acoustic backscatter data, and magnetic gradiometer data. He has authored or coauthored more than 60 technical reports and papers.

Dr. Dobeck received the 1981 and 1996 Commanding Officer/Executive Director Award for Science and Technology. He is a Reviewer for IEEE, ASME, SPIE, and the *Journal of Underwater Acoustics* and has been the Session Chair at past IEEE and SPIE conferences.