SONAR TARGET RECOGNITION

Sonar target recognition deals with identifying the source and nature of sounds by employing various signal-processing strategies. Target recognition includes detection (knowing something is out there), classification (knowing whether or not it is a target of interest), and identification (knowing the type of target). Sonar targets, such as submarines, surface ships, autonomous underwater vehicles, mines, and intruders, may be quiet or emit various sounds that can be exploited for passive sonar target recognition.

There are passive and active modes of sonar target recognition. In passive sonar operation, typical sound emissions exploited for target recognition are as follows (1):

1. *Transients.* Unintentional (dropping a tool, hull popping from a depth change, periscope cavity resonances, etc.) and intentional (low-probability-of-intercept signals for navigation and communication) signals with short time duration and wideband characteristics
2. *Machinery Noise.* Noise caused by the ship's machinery (propulsion and auxiliary)

3. *Propeller Noise.* Cavitation at or near the propeller and propeller-induced resonances over the external hull
4. *Hydrodynamic Noise.* Radiated flow noise, resonance excitation, and cavitation noise caused by the irregular flow of water past the moving vessel

While transients occur infrequently, the latter three types exist continuously. They collectively give rise to line-component (i.e., sinusoidal) and continuous spectra, which are known as passive narrowband (PNB) and passive broadband (PBB), respectively. Passive sonar processors perform signal processing on raw data generated by a number of passive sonar arrays mounted throughout the vessel, present both audio and video channels to sonar operators, and generate contact reports by comparing extracted signature parameters or features—harmonic lines characteristic of propeller types, transient characteristics, cavitation noise properties, and so on—with templates stored in the passive sonar database. Sonar operators listen to audio channels and watch displays before validating or correcting the processor-generated contact reports.

The second mode of sonar operation is active. Active sonar can be used to ensonify quiet targets. Echo patterns can give considerable insights into target structures, which can be useful for active target detection and classification. For instance, low-frequency sonars penetrate the body of the vessel, eliciting echoes caused by both specular reflection and the sound waves interacting with discontinuities in the body (2). High-frequency sonars are commonly used to image an unknown target after being cued by other long-range sensors. Mid-frequency sonars are used in tactical situations for target recognition by taking advantage of both specular echo patterns and moving target indication (MTI) based on Doppler after reverberation suppression (3). The operational concept of active sonar is very similar to that of radar. Active sonar processors perform beam forming, replica correlation, normalization, detection, localization, ping-to-ping tracking, and display formatting. Sonar operators differentiate underwater targets from background clutter using echo returns.

Since the end of the Cold War, there has been a proliferation of regional conflicts in which the US Navy must project power in littoral waters in order to maintain peace. This paradigm shift has forced the US Navy to focus on shallow-water sonar processing. The shallow-water environment is characterized in general by (1) a high level of the ambient noise, (2) complex propagation or multipath, and (3) a lot of clutter from merchant ships, marine biologics, and complex bottom topography. Furthermore, new quieter threats, such as diesel-electric submarines, are a major challenge to passive sonar target detection and recognition especially when coupled with the shallow-water environment. As a result, most advanced sonar processors rely on a combination of active processing and full-spectrum passive processing that takes advantage of every available signal bandwidth for improved sonar target-recognition performance. The use of an active sonar to compensate for poor passive detection performance of quieter threats in shallow water, however, can pose problems because of too many echo returns unless automatic detection and recognition algorithms reduce the number of returns to a manageable level for sonar operators.

The main objective of sonar automatic target recognition (ATR) is information management for sonar operators. Unfortunately, sonar ATR is confronted with many challenges in these situations. Active target echoes must compete with reverberation, clutter (any threshold-crossing detection cluster from nontarget events), and background ambient noise while passive signals must be detected in the presence of interfering sources encompassing biologics, background noise, and shipping traffic. Furthermore, environmental variation in shallow water can alter signal structures drastically, thus degrading target-recognition performance. These challenges must be overcome through a synergistic combination of beam forming, signal processing, image processing, detection, situationally adaptive classification, tracking, and multisensor fusion.

Sonar ATR is an interdisciplinary field that requires diverse knowledge in acoustics, propagation, digital signal processing, stochastic processes, image understanding, hardware and software tradeoffs, and human psychology. The foremost task here is to convert a large amount of raw data from multiple sensors into useful knowledge for situational awareness and human decision making. The challenge is to design a robust system that provides a high probability of correct recognition (P_{CR}) at low false-alarm rates (P_{FA}) in complex and nonstationary environments.

To design an effective sonar target-recognition system, we must explore a number of algorithms in the areas of signal projection or filtering, interference suppression, feature extraction, feature optimization, and pattern classification (4). The five crucial components of sonar target recognition are the following.

1. Signal sorting in various spaces, such as time, frequency, geometric space, and transformation space
2. Signal processing that takes advantage of the underlying physical mechanism by which target signatures are generated
3. Compact representation of signal attributes (features)
4. Design of a classifier that takes advantage of the underlying good-feature distribution
5. Performance quantification in terms of operationally meaningful criteria

In short, the key to achieving excellent target-recognition performance is an integrated and systematic approach that spans the entire spectrum of sonar processing in a mutually reinforcing manner.

In this context, we introduce an integrated sonar ATR paradigm that addresses the five components effectively as shown in Fig. 1. Data projection deals with representing signals as compactly as possible while preserving crucial signal attributes. Since we do not have the a priori knowledge about good features, we initially extract as many pertinent features as possible. Feature ranking involves finding features that add value to target recognition and deleting the ones that do not.

Classifiers estimate class-conditional probability density functions (pdfs) to map input features onto an output decision space. It is essential that this mapping algorithm be devoid of model-mismatch errors to achieve upper bounds in classification performance. The performance upper bounds in classification are conceptually similar to the Cramer-Rao lower

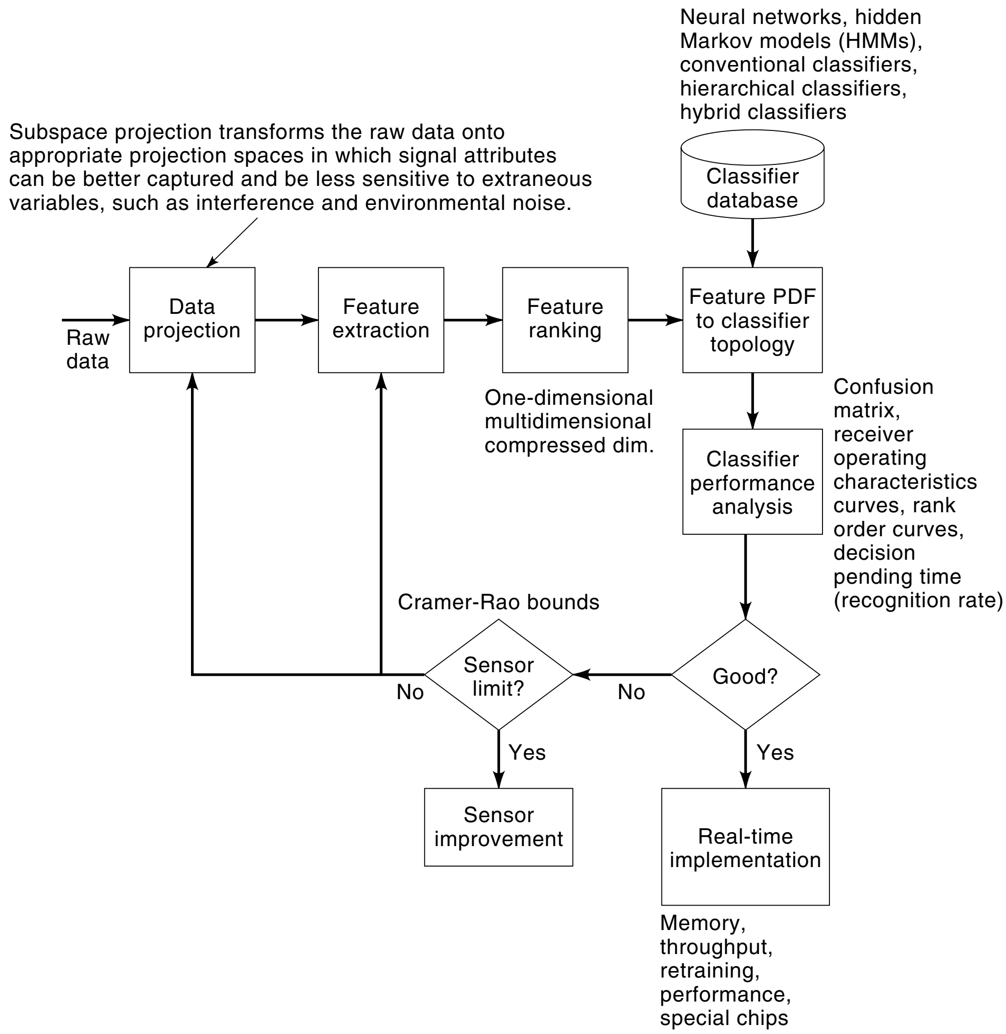

Figure 1. The integrated ATR paradigm combines signal filtering, feature optimization, and classification to achieve maximum sonar target-recognition performance.

bounds (CRLBs) in parameter estimation (5). Model-mismatch errors can occur if the classifier structure does not model the underlying good-feature pdf adequately. The CRLB concept allows us to assess whether poor performance is attributable to sensor limitation (sensors not providing enough useful information) or algorithm limitation (algorithms not capturing all the useful information in data).

This article is organized as follows. We first study how various aspects of signal transformation, signal classification, and data compression can be combined in order to extract the maximum amount of useful information present in sensor data. Next, we apply sonar target-recognition theories to challenging real-world problems—active sonar classification and passive full-spectrum processing for transient signal classification. Finally, we explore new, advanced concepts in sonar target recognition. Throughout this article, our focus is on the general framework of sonar target recognition so that the readers can appreciate the big picture on how sonar targets are recognized.

INTEGRATED SONAR ATR PROCESSING

In this section, we introduce the integrated sonar ATR processing and explain the role of each processing block within the system's context. Figure 2 depicts a general sonar-processing flowchart.

Joint time-space processing sorts multiple signals as a function of time of arrival (TOA), direction of arrival (DOA), and spectral band. That is, any separation in TOA, DOA, or frequency will be sufficient for signal deinterleaving. Beam forming handles DOA sorting while wideband pulses are used for TOA sorting in active sonar. Each separated signal will then be projected to appropriate transformation spaces. The main purposes of signal projection are data compression and energy compaction.

For example, a continuous wave (CW) time-domain signal can be projected onto the frequency domain by the Fourier transform. This signal-projection operation yields two related benefits: compression of the entire time-domain data into one frequency bin and signal-to-noise ratio (SNR) improvement by a factor of 10 log N_{FFT}, where N_{FFT} is the size of the fast Fourier transform (FFT). Not only does signal projection improve the probability of discriminating multiple sinusoids by virtue of data compression, but it enhances the algorithm robustness in parameter estimation thanks to the SNR gain. The key concept here is that multiple projection spaces be investigated as a function of signal characteristics to obtain orthogonal, mutually reinforcing information for improved detection and classification.

In general, most traditional detectors, such as a replica correlator or an m-out-of-n detector (m detections in n opportunities, where $M < N$ constitutes detection), rely on a single parameter—integrated energy after constant-false-alarm-rate (CFAR) processing—for detection (6). This approach is acceptable as long as the number of false returns that exceeds the detection threshold remains reasonable. Unfortunately, the number of false alarms can be rather significant in today's operating environments.

Instead of relying on the amplitude feature alone, we extract and fuse multiple signal attributes using a classifier. ATR can be performed in sequential steps, borrowing from the *divide-and-conquer* paradigm. In Fig. 2, we first perform target-versus-nontarget discrimination, followed by target identification. The latter processing itself can be broken into hierarchical steps depending on the complexity of target types (7). Furthermore, both static and dynamic features, coupled with integration of frame-based classification scores, can be used to improve the confidence level of target identification. Now, we discuss signal projection, feature optimization, and target recognition thoroughly.

Signal Projection and Feature Extraction

The main objective of signal projection is low-dimensional signal characterization, which naturally leads to *subspace filtering*. Figure 3 illustrates the basic concept of signal projection. Let $y = f(x)$, where x and y represent raw and projected data, respectively. The $f(\cdot)$ is a projection operator that transforms x and y in order to compactly represent x in y. The behavior of x is governed by the probability law derived from its components: target and clutter. That is, the probability law consists of two conditional pdfs, $P(x|\text{target})$ and $P(x|\text{clutter})$. In general, the overlap between the two class-conditional pdfs is quite high, rendering target recognition difficult in x.

Signal projection alleviates this problem by projecting x onto y in which both target and clutter components are captured with a much smaller set of parameters (dimension reduction or energy compaction) (5). More important, capturing target and clutter components in a reduced dimension improves the probability of separating target and clutter in y—subspace filtering. Therefore, the criteria for selection of projection algorithms are the amount of energy compaction and the extent to which various signals can be separated.

We present two examples to illustrate the effectiveness of signal-specific data projection. In adaptive interference suppression, the interference component can be modeled more efficiently in the projected vector space spanned by y. After in-

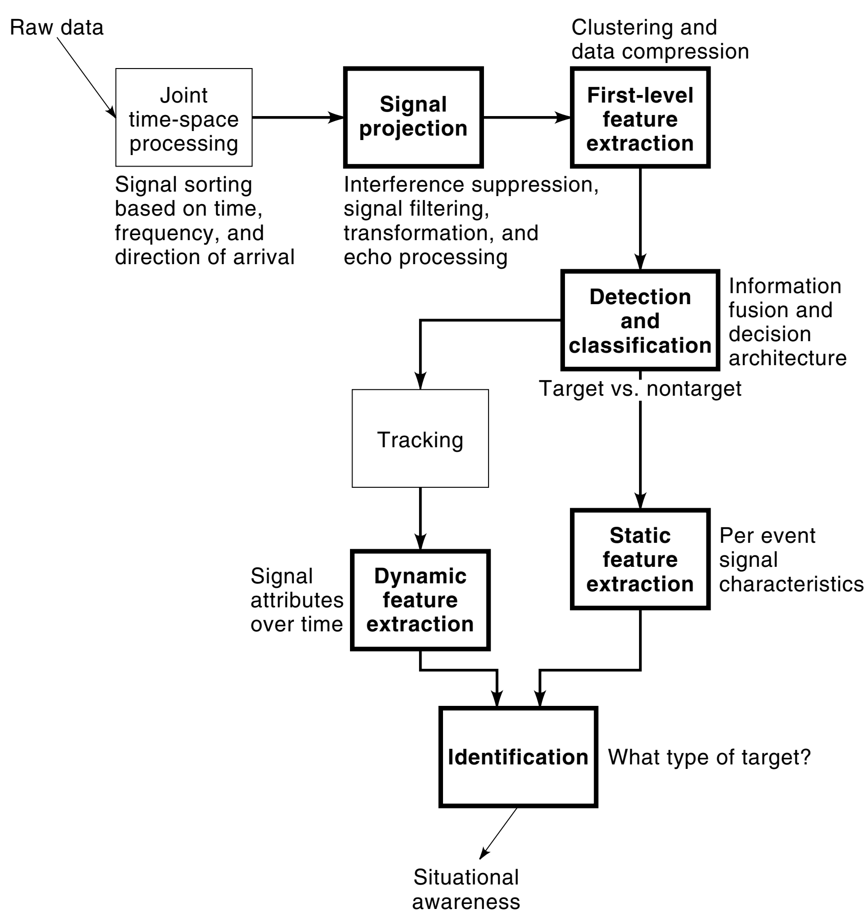

Figure 2. For high-performance sonar target recognition, many processing elements—beam forming, signal projection, tracking, and pattern recognition—must work in cooperation within the overall systems framework. In this article, we focus on the boldfaced blocks.

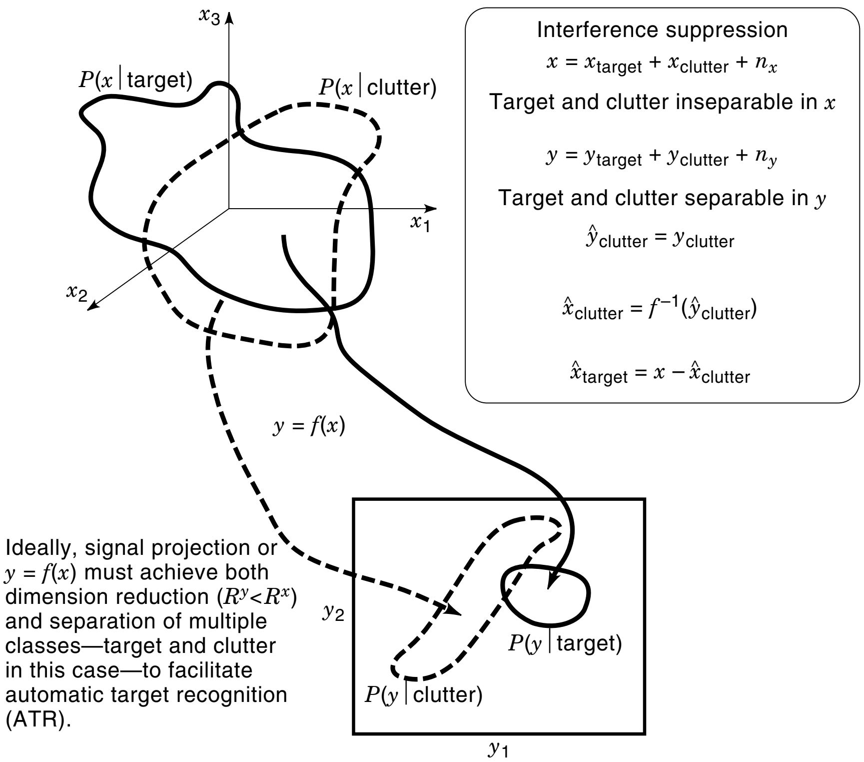

Figure 3. Conceptual framework of signal projection—dimension reduction and subspace filtering. In general, dimension reduction occurs when the number of basis functions in y for representing a signal is less than that in x. n_x and n_y refer to noise in x and y, respectively.

terference modeling, its structure in x can be estimated through reverse transform and coherently subtracted from the original time-series data as shown in Fig. 3. One such approach is the principal component inversion (PCI), where the interference structure is modeled as a linear combination of orthogonal basis vectors derived from a Toeplitz data matrix (8). This approach has been applied successfully to reverberation suppression for CW, hyperbolic frequency-modulated (HFM), and linear frequency-modulated (LFM) waveforms. Figure 4 shows the results of PCI on reverberation suppression for a CW waveform. Note that PCI was able to recover a low-Doppler target hidden in reverberation.

The second example deals with time-frequency representation of sonar transients. Although the short-time Fourier transform (STFT) is the most widely used time-frequency distribution function, Ghitza's ensemble interval histogram (EIH) deserves a special mention here because of the importance of aural processing in sonar target recognition. EIH is based on an auditory neural model (9) that consists of two parts: the preauditory part comprising a bank of cochlear

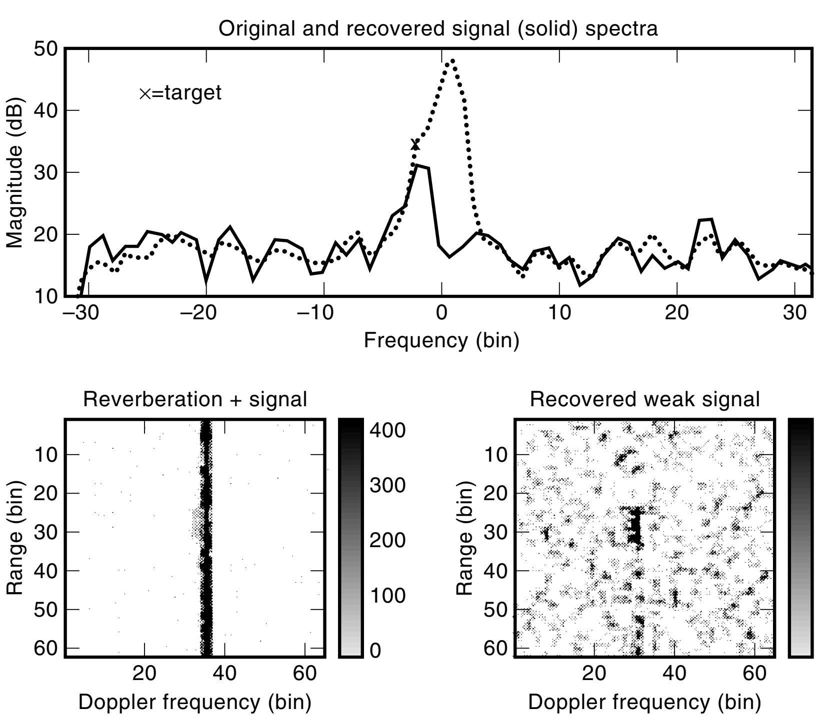

Figure 4. PCI estimates the interference structure using principal components and coherently subtracts it from the raw waveform to extract the weak signal.

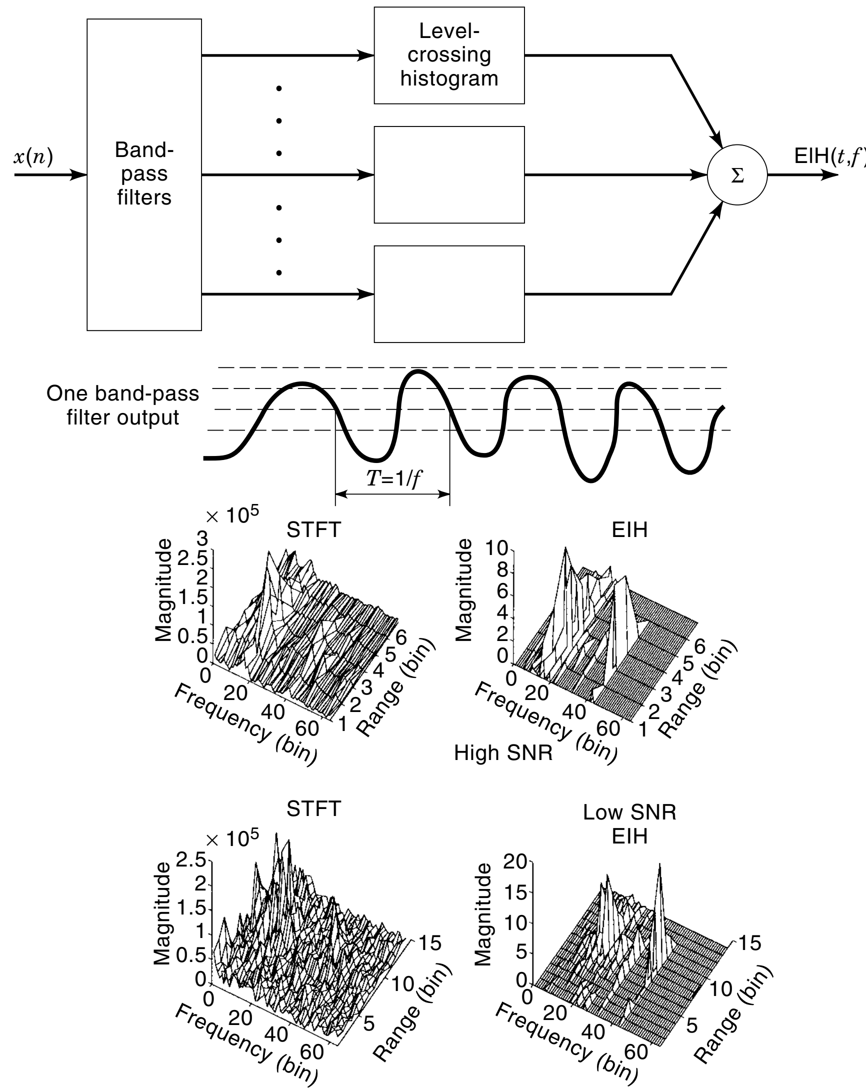

Figure 5. EIH is an auditory neural model that provides robust transient signal characterization, particularly at low SNR. This transient contains a dual-tone structure, which is preserved better with EIH than with STFT.

bandpass filters whose cutoff frequencies are logarithmically spaced for multispectral analysis and the postauditory part that performs spectral content estimation via multiple level-crossing detectors as shown in Fig. 5. Note that EIH captures the time-frequency characteristics of the transient with a dual-tone structure more accurately than STFT, particularly at low SNR.

After signal projection, features are extracted from each projection space. Feature extraction is a process by which signal attributes are computed from various projection spaces and fused in a compact vector format. Good features should possess the following desirable traits:

1. Large interclass mean separation and small intraclass variance
2. Insensitive to extraneous variables (little dependence on SNR)
3. Computationally inexpensive to measure
4. Uncorrelated with other features
5. Mathematically definable
6. Explainable in physical terms

Features can be broadly categorized into *static* and *dynamic* types. For very short events, we can extract static features that characterize the entire event period. For events with longer durations, it is often advantageous to compute key features at a fixed time interval so that their transition characteristics over time can be further exploited for signal discrimination. It is intuitive that a hybrid classifier that can accommodate both static and dynamic features usually outperforms classifiers that rely exclusively on either static or dynamic features alone.

Feature Optimization

Feature optimization is an integral part of sonar target recognition and involves feature normalization and ranking based on an appropriate criterion. Normalization is necessary to prevent numerical ill-conditioning. Feature ranking can be broadly categorized into two types (4):

1. Derive M features $y = [y_1 \cdots y_M]^t$ from the original N features ($M < N$) by applying an $M \times N$ linear transformation matrix A or a nonlinear mapping function $g(\cdot)$

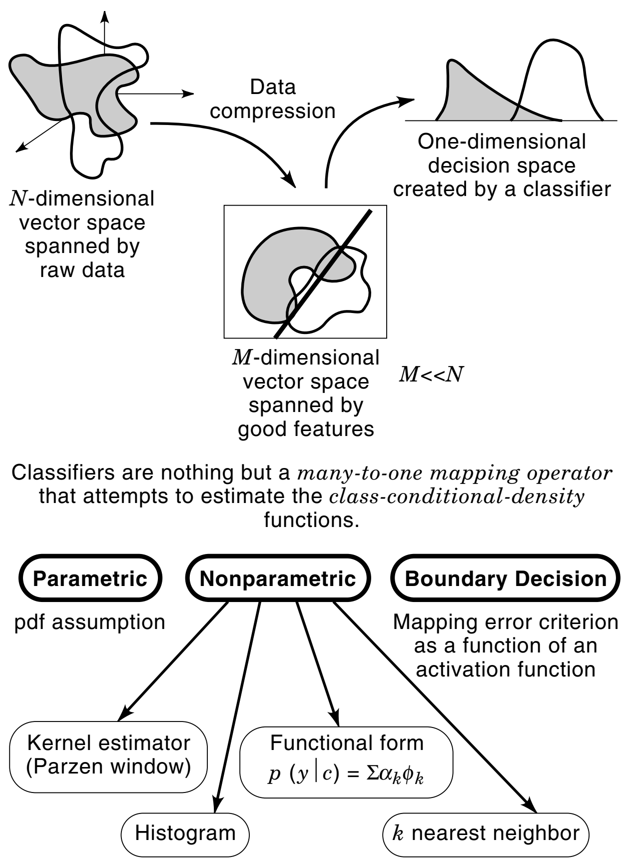

Figure 6. Classifiers map the vector space spanned by selected features onto a decision dimension.

to the original feature vector x such that

$$y = Ax \quad \text{or} \quad y = g(x) \tag{1}$$

2. Rank individual features according to their contribution to the overall recognition performance. This can be further divided into computationally efficient single-dimensional feature ranking, computationally expensive multidimensional feature ranking, and feature ranking in a compressed feature dimension as a compromise. The multidimensional ranking approach is equivalent to a combinatorial problem of finding the best M-feature subset out of the N original features. We will denote this method as a feature-subset selection approach.

Automatic Target Recognition—Mapping Features to Classifiers

The fundamental issue in classifier design is quantifying the extent to which a classifier captures all the useful information present in input features (training data) while remaining flexible to potential mismatch between training and test data. In order to achieve the performance of the optimal Bayes classifier, we need to approximate the class-conditional pdfs from the available training data and design a classifier architecture based on the estimated class-conditional pdfs. This approximation can take a form of parametric, nonparametric, and boundary-decision types. Figure 6 describes the relationship between feature extraction and classification succinctly.

In general, parametric classifiers make strong assumptions regarding the underlying class-conditional pdfs while nonparametric classifiers estimate class-conditional pdfs from the available training sonar data. On the other hand, boundary-decision classifiers construct linear or nonlinear boundaries that separate multiple classes (targets) according to some error-minimization criteria. The key concept here is that some classifiers do better than others for certain feature sets. Therefore, synergy between a classifier and a good-feature subset must be maximized whenever possible. For example, if class-conditional pdfs exhibit unimodal, Gaussian characteristics, a simple parametric classifier may suffice. In contrast, if class-conditional pdfs are multimodal and non-Gaussian, nonparametric classifiers with adaptive vector quantization would be preferred to parametric classifiers. In essence, a system designer must perform judicious trade-offs in the areas of target-recognition performance and computational requirements during training and actual sonar system operations as a function of the amount of available training data, anticipated feature-space perturbation by environmental variation, and the need for in situ adaptation.

REAL-WORLD EXPERIMENTS

In this section, we apply theories to two challenging, real-world problems. These examples illustrate how various signal-processing concepts in echo processing, filtering, and pattern recognition can be integrated to detect the presence of sonar targets.

Active Sonar Target Recognition

One of the most difficult challenges in active sonar processing is differentiating target returns from false returns. In impulsive-echo-range (IER) processing, an additional challenge is dealing with stochastic impulsive source variability. In order to resolve range ambiguities, impulsive sources are transmitted at a variable repetition rate in a multistatic environment. The goal of active sonar target recognition is to remove as much clutter as possible while maintaining an acceptable target-recognition performance for an eventual confirmation by sonar operators. In this section, we present an active target-echo recognition algorithm using an integrated pattern-recognition paradigm that spans a wide spectrum of signal and image processing—target physics, exploration of projection spaces, feature optimization, and mapping the decision architecture to the underlying good-feature distribution (4,10).

Projection-Space Investigation. In general, selection of a projection space is domain specific and largely motivated by inputs from experienced sonar operators and phenomenology. For example, operators often listen for distinct "metallic" sounds for aural discrimination. This observation implies that various speech-processing algorithms can be applicable to sonar target recognition. Moreover, energy detector and time-frequency distribution (TFD) outputs seem to provide a good operator aid for visual discrimination. The complex time-varying echo structures dictate the use of frame-based processing to capture time-dependent signal attributes. Transformation algorithms should be able to perform both noise (ambient noise and reverberation) suppression and separation of target and clutter components.

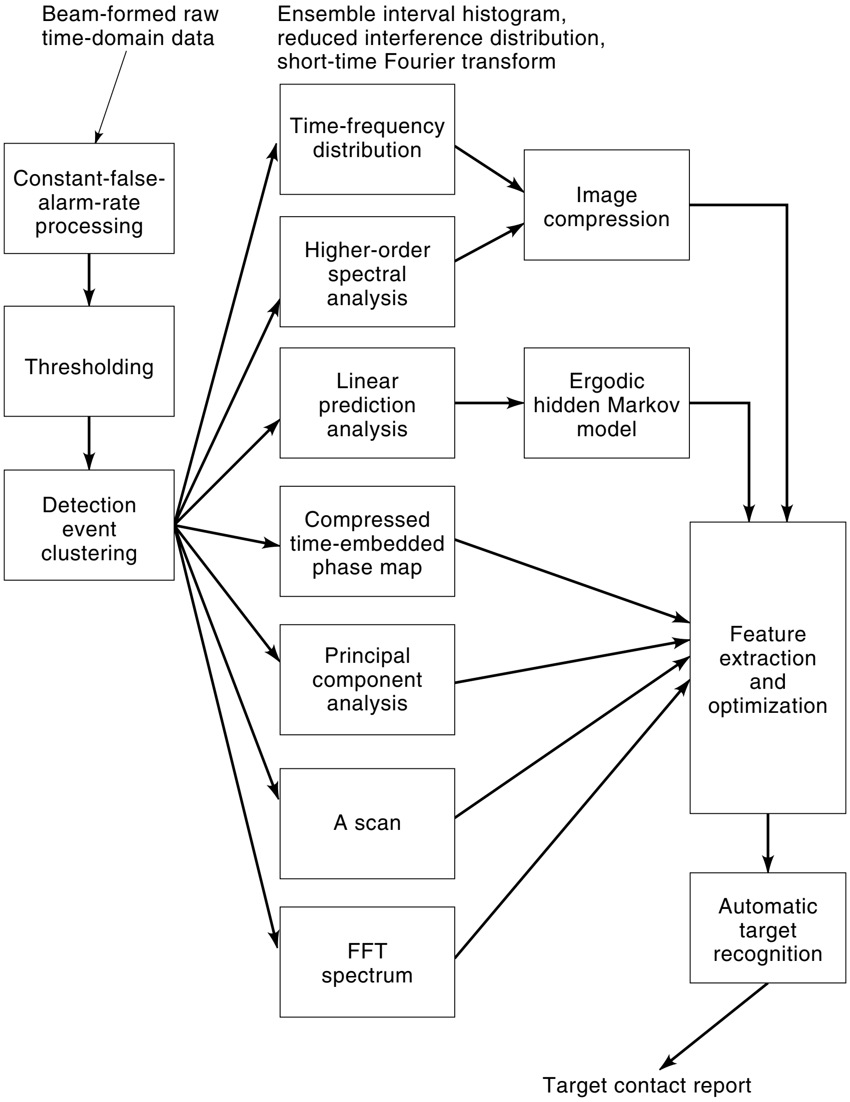

Figure 7. The overall processing flow chart.

Figure 7 depicts the overall processing strategy consisting of detection-cluster or *snippet* segmentation, feature extraction, feature optimization, fusion, and classification. First, we perform snippet segmentation based on CFAR detection-threshold crossing. Each segmented snippet is projected onto various projection spaces.

We extract features from seven projection spaces consisting of smoothed energy or A-scan output, FFT spectrum, TFD using STFT, the reduced interference distribution (RID) (11), and EIH, higher-order spectrum (HOS) (12), principal component analysis (PCA), a compressed phase map (13), and a speech-related processing domain using linear prediction, cepstral, and δ cepstral coefficients. Instead of extracting high-dimensional features from raw TFD and HOS projection spaces, we utilize an image coding algorithm to achieve further data compression (14). After feature extraction, we perform thorough feature analyses for feature optimization and ranking to select the optimal feature subset based on an appropriate class separability criterion. Finally, we evaluate the target-recognition performance using the selected feature subset and construct the best classifier topology. In essence, given the optimal feature subset, selection of the best classifier structure is equivalent to finding the best mapping function between input parameters (features) and desired outputs (class label—target or clutter). Now we describe projection spaces with good features in detail.

1. *Temporal Space.* Derived mainly from the energy detector and linear predictor outputs, temporal features provide clues on target extent and highlight structures (bow and stern planes, railings, and periscopes) as a function of aspect. For seamounts with a few distinct scatterers, the envelope structure is complex and asymmetrical as measured by shape skewness and kurtosis, while the cylindrical target at broadside yields a symmetrical, Gaussian envelope shape. Good features from this projection space are pulse width, rise and fall times, and amplitude and shape statistics.
2. *Time-Frequency Distribution with Image Compression.* Features from the TFD attempt to capture spectral and temporal variations associated with the highlight structure and secondary arrivals from helical and flexural waves (15). We explore the following three TFDs to as-

sess the impact of time-frequency resolution on active classification: STFT, RID, and EIH.

3. *Compressed Phase Map.* A phase map is a convenient way of representing time-embedded samples in a multidimensional state space and is quite effective in capturing dynamics of low-dimensional, deterministic signals. A typical example can be found in nonlinear dynamical system modeling (13). For this application, we capture transitional signal characteristics from sample-to-sample differences of the energy detector output. For returns from smooth-surface objects, sample-to-sample deviations of the differencer output are small and their trajectory follows a well-defined path with small fractal dimension. Fractal dimension provides information on how much of the state space is filled by the trajectory. On the other hand, returns from complex-scattering objects, such as seamounts and wrecks, exhibit large trajectory fluctuations, leading to a diffused phase map with large fractal dimension. The same concept of subspace filtering is used to capture desirable signal transitional characteristics efficiently. That is, we use the singular value decomposition (SVD) to project noisy points in the state space R^d onto a new space R^{N_r}, where d and N_r represent the original embedding dimension (the total number of consecutive time samples used in constructing the state space) and the reduced dimension representing the signal subspace, respectively. The computational procedures are explained below.

   a. Generate a differencer output as follows:

$$p(n) = \frac{x(n) - x(n-1)}{x(n)} \tag{2}$$

   where $x(n)$ is the normalized energy detector output.

   b. Construct a phase map matrix Φ of size $d \times K$ using time-delay embedding of the differencer output,

$$\begin{aligned} P_n &= \{p_n p_{n-1} \ldots p_{n-d+1}\}^t \\ \Phi &= \{P_1 P_2 \ldots P_n \ldots P_K\} \end{aligned} \tag{3}$$

   K is $N - d$, where N and d denote a total length of differencer output p_n and the embedding dimension, respectively.

   c. Perform the SVD on the covariance matrix $R = \Phi\Phi^t$. Estimate the matrix rank using the minimum description length (MDL) criterion (16) to obtain orthonormal projection operators:

$$d(k) = -(p = k)N_{\text{av}} \log_{10} \left(\frac{\displaystyle\sum_{i=k+1}^{p} \lambda_i^{1/p-k}}{\displaystyle\frac{1}{p-k} \sum_{i=k+1}^{p}} \right) + 0.5k(2p-k)\log_{10} N_{\text{av}} \tag{4}$$

   where N_{av} is the averaged sample size, p is the dimension of R, λ_i is the ith eigenvalue arranged in descending order of magnitude, and $k = 0, 1, \ldots, p-1$. The rank of R is equal to the value of k that minimizes $d(k)$,

$$N_r = \arg\min_k d(k) \tag{5}$$

   d. Use the estimated signal subspace projection operator to project a full-rank matrix Φ to the compressed phase space:

$$R = U \sum U^t \tag{6}$$

$$\Phi_r = U^t_{1:N_r} \Phi \tag{7}$$

   where $U_{1:N_r}$ and Φ_r denote a left singular matrix with a rank N_r and a compressed phase map, respectively.

4. *Speech-Processing Features.* The primary motivation for extracting speech-processing related features is that the eye (visual) and the ear (aural) process the same information in a somewhat different fashion. For example, the eye is capable of processing a large amount of information in a short time, but tends to be deficient in details. On the other hand, the ear has a much higher dynamic range and resolution and thus can better distinguish details but is slower than the eye. The main objective for applying frame-based speech processing to IER clutter reduction is to capture detailed acoustic transitional characteristics that cannot be captured adequately from the visual projection spaces. Echoes from objects with various structural properties—rib, air-filled cavity, solid filling (seamounts), chemical filling (mines)—can possess distinct sound characteristics, which can be compactly represented with linear prediction, cepstral, and δ cepstral coefficients. Linear predictive coding estimates spectral phase and amplitude variation over time while cepstral coefficients attempt to separate spectral envelope from the underlying harmonic structure. We use standard ergodic hidden Markov models (HMMs) to characterize both target and clutter echoes (17,18). We extract features from concatenated log-likelihood ratio scores as well as transition and observation statistics associated with each state (2).

Real-Data Analysis Results. In this section, we present our clutter-reduction performance results based on real-data analysis and compare our performance with that of the baseline processing that consists of CFAR detection and rule-based clutter rejection. For this analysis, we use segmented detection clusters from the shallow-water real-active-data set and ground truth information obtained during data reconstruction. After extracting features from the seven projection spaces, we perform a comprehensive feature analysis for feature pruning and optimization prior to classification performance analysis. We evaluate target-recognition performance using the top 10 to 15 features.

Borrowing from the *divide-and-conquer* paradigm, we perform hierarchical sequential pruning classification in two steps: primitive and fine classification (7). During the first stage of primitive classification, the pulse width is used to reject obvious false contacts. We use a conservative prescreening threshold to ensure that there is little risk of false dismissal of genuine target echoes. Not only is this approach computationally attractive due to the reduced number of detection clusters to process during the computationally intensive second stage, but it provides an additional benefit of not having to waste degrees of freedom on modeling obvious false contacts later in fine classification. For the second-stage fine classification, we derive clutter-reduction performance from an average of 64 independent runs to minimize performance bias caused by uneven class population.

We evaluate performances of multivariate Gaussian classifier (MVG), *k*-nearest-neighbor classifier (KNN), nearest-neighbor classifier (NNC), probabilistic neural network (PNN), and fast backpropagation neural network (FBPN) (4) to determine the most appropriate classifier architecture. Since the underlying multidimensional feature pdfs exhibit unimodal characteristics with reasonable class separation as shown in Fig. 8, MVG and PNN perform quite well while KNN and NNC perform poorly. (KNN and NNC are nonparametric classifiers that estimate class-conditional pdfs from a small fraction of training data. This procedure can backfire if class-conditional pdfs are unimodal.) Boundary-decision classifiers, such as FBPN, perform well initially as decision boundaries are relatively simple for a small decision dimension. Nevertheless, as the decision dimension increases, the class boundaries become more complex and FBPN's performance suffers. In summary, MVG and PNN provide the best performance because their mapping structures match the underlying good-feature pdfs.

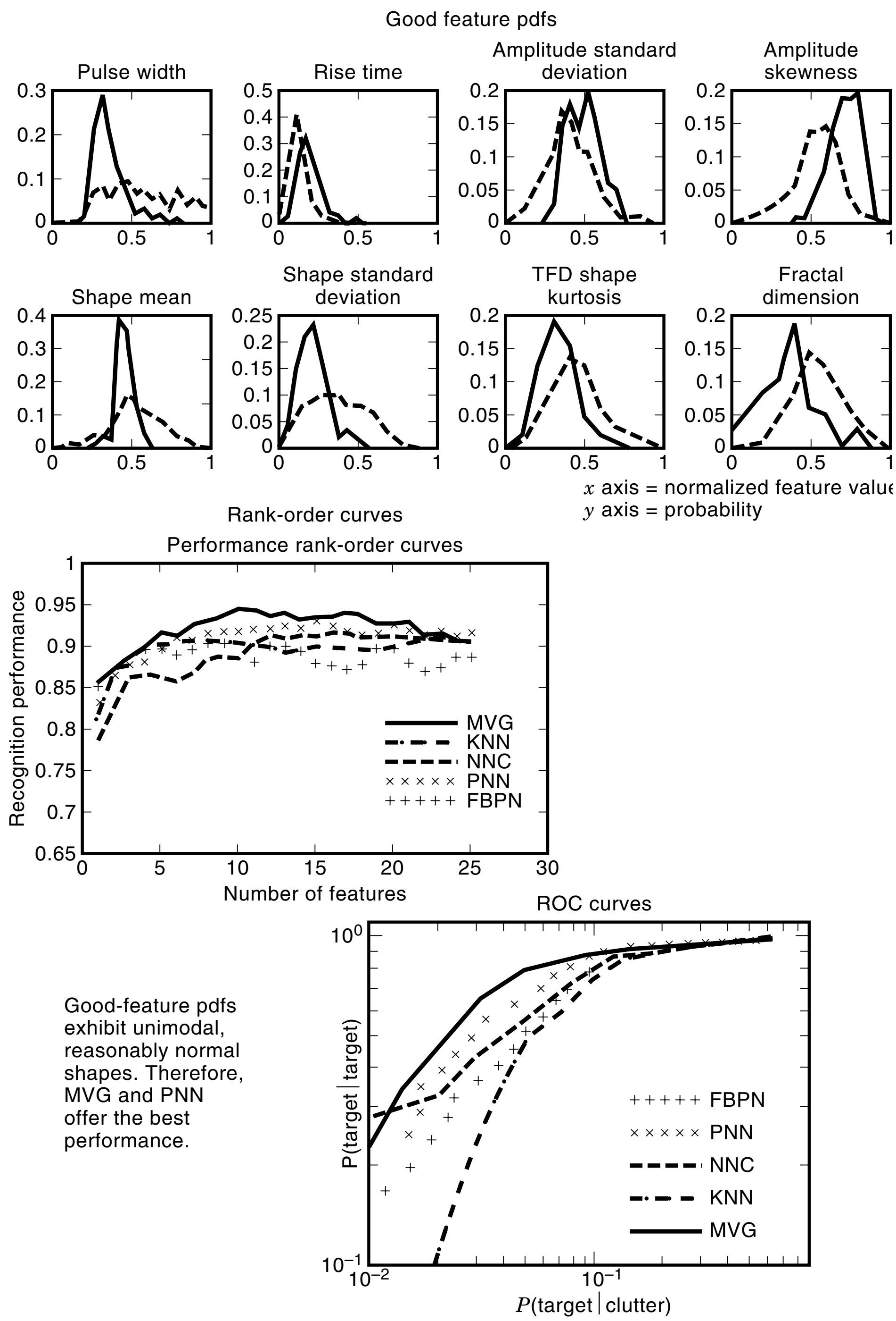

Figure 8. Performance rank-order curves are useful in determining an appropriate decision dimension in classification. Since good-feature pdfs (solid, target; dotted, clutter) seem unimodal and slightly non-Gaussian with some class overlap, PNN and MVG perform the best.

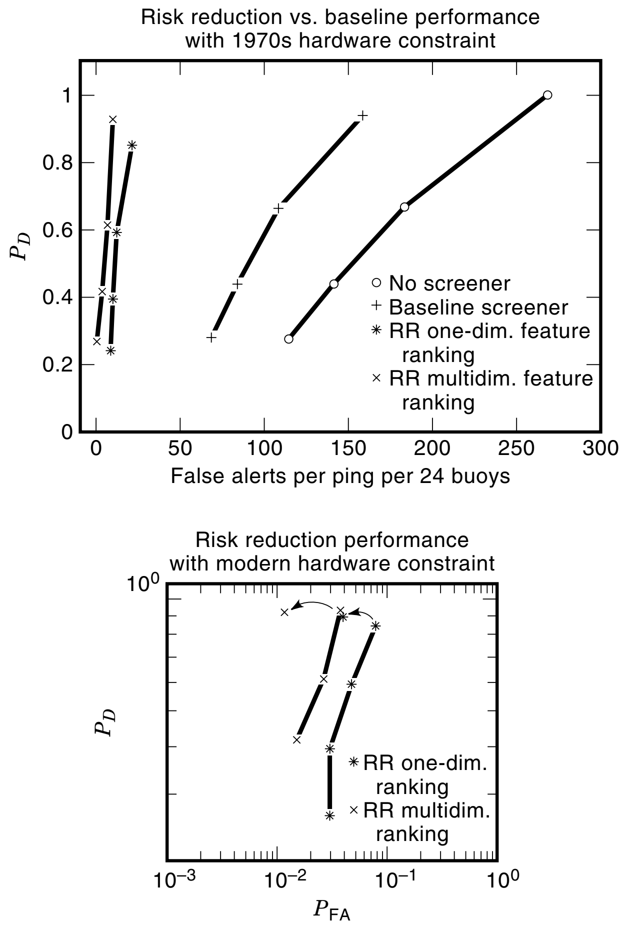

Figure 9. Classification ROC curves demonstrate clutter-reduction performance improvement with our sequential hierarchical classification approach at four different SNRs. The bottom figure shows the improved clutter-reduction performance with the modern hardware constraint at the lowest SNR only. RR stands for risk reduction. $P_D = P$(target/target). P_{FA} = (target/clutter). Arrows show performance improvement.

Figure 9 shows receiver operating characteristics (ROC) curves for the baseline and risk-reduction processing with the two computational resource constraints in an operationally meaningful format. For this analysis, we use both one-dimensional and multidimensional feature-ranking algorithms to assess the clutter-reduction performance. The motivation for using the computationally expensive multidimensional feature-ranking algorithm is that it enables us to derive the performance upper bounds for a given data set and a feature set. The baseline processing consists of a constant-false-alarm-rate normalizer, a short-time averager, and a threshold detector. The baseline rule-based screener uses pulse width and fall time for clutter rejection. We used the baseline performance as a benchmark with which our risk-reduction performance was compared. Operating points are derived from the echo returns after detection as a function of SNR.

Our real-data analysis results indicate that we can achieve maximum classification performance with approximately 10 to 15 features. Note that using the first risk-reduction algorithm with one-dimensional feature ranking based on the multimodal overlap measure (MOM) defined as

$$\mathrm{MOM}_i = \int_{y_i} \mathrm{Min}[P(y_i|\mathrm{target}),\ P(y_i|\mathrm{clutter})]\,dy_i \tag{8}$$

where y_i is the ith feature (the lower the MOM, the better the corresponding feature in differentiating the target from clutter), we were able to achieve over 90% false-alarm reduction from the baseline/no-screener approach. The bottom ROC curves show clutter-reduction performance comparison between the computationally inexpensive features (derived from the A-scan, FFT, and STFT outputs) and features extracted from the seven projection spaces in the traditional P_D-versus-P_{FA} format. With the top 15 features, we were able to achieve an additional 4.5% improvement in overall correct classification performance (88.6% to 93.1%) for snippets that exceed the lowest SNR threshold. This improved performance translates to a 5% increase in P(target|target) (P_D jumped from 0.85 to 0.90) and a 50% reduction (7.8% to 3.9%) in P(target|clutter).

Passive Sonar Target Recognition

In order to maximize recognition performance of passive target emissions, it is important that we understand and exploit the underlying signal microstructure. PBB acoustic signatures often exhibit a microstructure that has time-varying, low-dimensional characteristics if projected onto an appropriate transformation space. With this in mind, we investigate how our knowledge of signature characteristics can be reflected on the PBB algorithm design to enhance target-recognition performance in shallow water. For this analysis, we use SWell-EX1 and PBB data sets provided by the Naval Research and Development (NRaD) and the Office of Naval Research (ONR), respectively (19).

Our processing strategy is based on exploitation of any microstructure inherently present in the target signature by projecting raw data onto various projection spaces, identification of key parameters or "features" crucial in determining the presence of a signal, designing a classifier topology that best matches the underlying feature distribution, and thorough detection performance analysis and comparison with that of a traditional energy detector to quantify performance gains as a function of input SNR.

Technical Approach. Figure 10 depicts the PBB processing flowchart consisting of subspace projection, feature extraction, and classify-before-detect processing. We initially project raw data onto a time-frequency map using the STFT to capture time-varying striation patterns visible in the PBB target signature. The next step is to emphasize important target signature attributes with image compression and Viterbi line extraction.

Image compression takes advantage of transform coding and principal component filtering to emphasize desirable signal components while suppressing noise. The Viterbi line extractor works as an adaptive, variable-length line integrator that enhances the time-varying striation pattern present in the PBB signature. Figure 11 demonstrates the effectiveness of the Viterbi line extractor in recovering weak time-varying frequency lines.

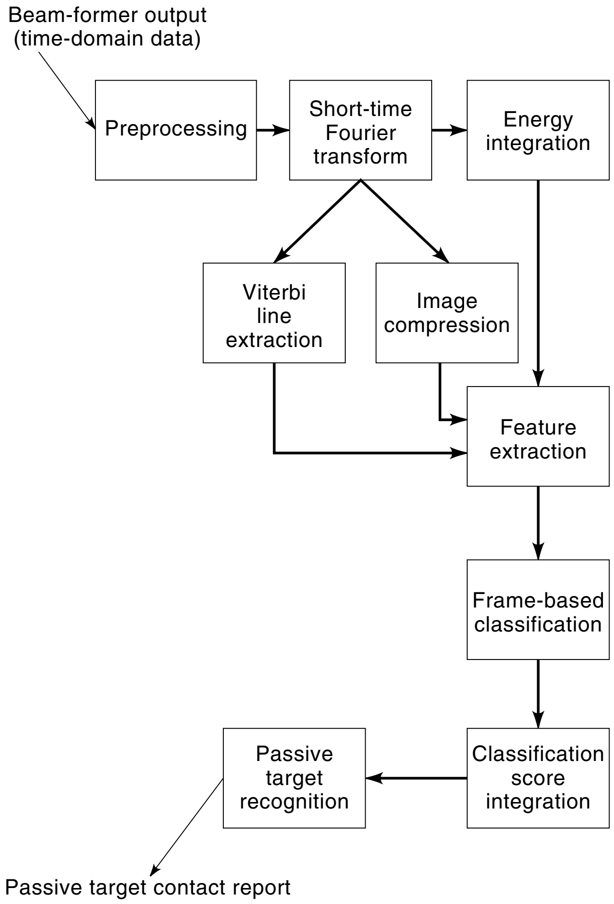

Figure 10. The PBB classify-before-detect flow chart.

The objective of the classify-before-detect processing is to utilize a more favorable decision space spanned by multiple, mutually reinforcing discriminatory features than the traditional amplitude decision space based on the integrated energy, particularly at low SNR. Finally, we compare the performance of our classify-before-detect algorithm with that of the conventional energy detector in terms of ROC curves and processing gain as a function of input SNR.

Real-Data Analysis Results. In this section, we present real-data analysis results. Figure 12 shows STFT spectrograms of the typical PBB target signature before and after various transformations: singular value decomposition (SVD), two-dimensional (2-D) discrete cosine transform (DCT), and compressed 2-D DCT. The signal that we are interested in detecting occupies the middle half of the spectrograms.

We initially extract a total of 64 features from the three projection spaces and perform thorough feature optimization and classification performance analysis using the Integrated Pattern-Recognition Toolbox. We achieve the maximum recognition performance using 8 to 10 features. We evaluate the extracted feature set with five classifiers that represent the three broad classifier categories: parametric, nonparametric, and boundary decision. Since good-feature pdfs are both non-Gaussian and multimodal, nonparametric classifiers based on vector quantization or k nearest neighbors outperform the others.

We quantify performances of the classify-before-detect algorithm in terms of the ROC curves and processing gain as a function of input SNR and compare them with those of the traditional energy detector. For performance evaluation of our algorithm, we use randomly partitioned, independent training and test data sets for algorithm tuning and cross validation. Figure 12 displays the ROC curve comparison of our

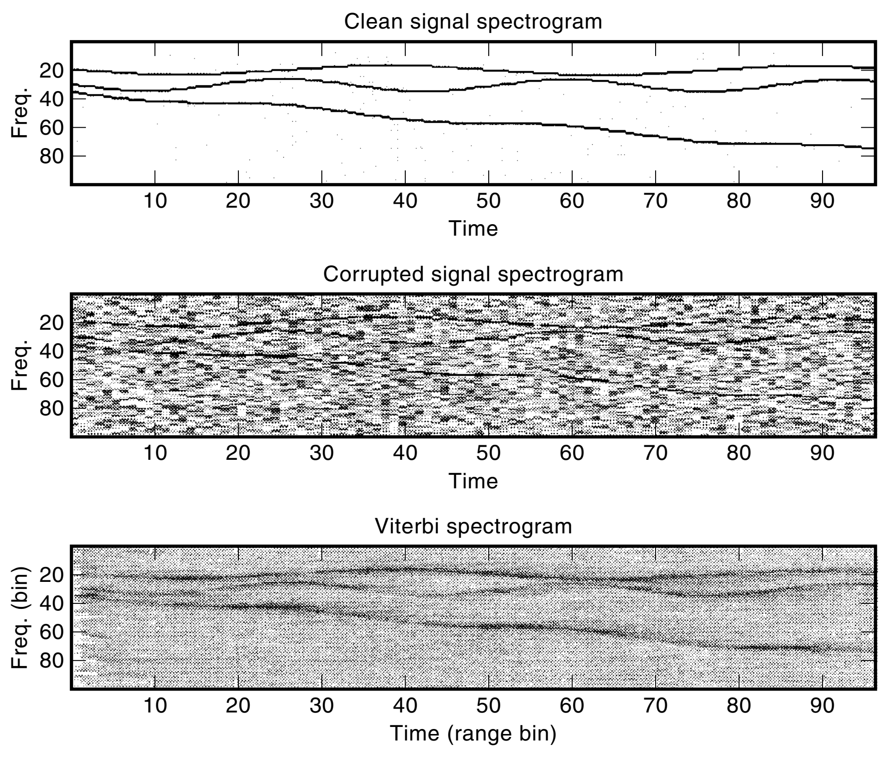

Figure 11. The Viterbi line extractor can effectively recover weak wandering frequency lines.

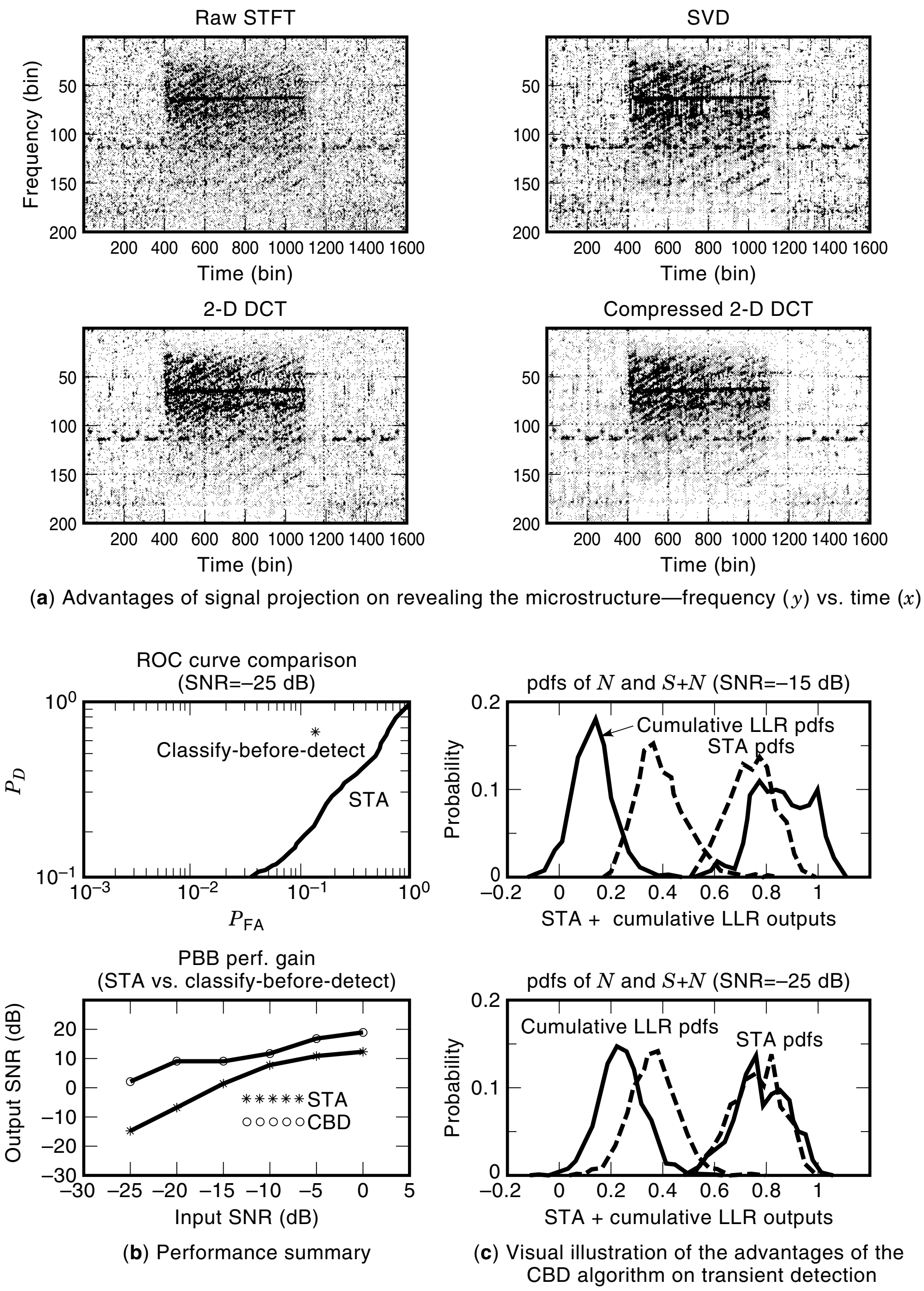

Figure 12. PBB acoustic signature and SWell-Ex1 ambient noise spectrograms and the CBD algorithm performance summary. N and S + N denote noise and signal + noise, respectively.

classify-before-detect algorithm with the energy detector. We also summarize and compare the processing gain of the two detectors.

Overall, we achieve an average of 10 dB additional detection performance improvement with the classify-before-detect approach over the traditional energy detector. The integration sizes for the short-term averager (STA) and the classify-before-detect processing are 10 and 5 frames, respectively. We deliberately compare the performance of our algorithm with the 5 frame integration to that of the STA with 10 frames to provide a slightly pessimistic performance comparison. That is, using the integration size of 10 for the classify-before-detect processing would have resulted in a higher processing gain. The input SNR is measured with respect to the full band while the output SNR is derived from the STA and cumulative log-likelihood ratio (LLR) pdf plots using the deflection index criterion. Note that output SNR in decibels is 10 $\log(\Delta\mu^2/2\sigma_s\sigma_n)$, where $\Delta\mu$ is the mean difference between the signal-plus-noise and noise-only pdfs. σ_s and σ_n denote standard deviations of signal-plus-noise and noise-only pdfs, respectively. Since the STA processing involves STFT, envelope detection, and two-dimensional integration (signal subband and time), the output SNR is not a simple function of the temporal integration size.

The advantage of the classify-before-detect algorithm can be better appreciated by a qualitative look at the pdf plots of the STA and classify-before-detect cumulative LLR outputs. Figure 12 shows the signal-plus-noise and noise-only pdfs of the two processing outputs at input SNRs of -15 and -25 dB. At -25 dB, the two pdfs at the STA output completely overlap, rendering detection in the amplitude space very difficult if not impossible. On the contrary, pdf plots derived from the cumulative LLR output show a good separation, indicating that a judicious selection of features combined with an appropriate classifier topology is crucial in achieving an additional detection performance improvement.

EMERGING TECHNOLOGIES IN SONAR TARGET RECOGNITION

The two key areas for future research are accurate quantification of classification performance upper bounds and situationally adaptive target recognition. In this section, we first explore the underlying concepts of data compression, class separability, and sufficient statistics in the context of estimating performance upper bounds in classification. Next, we provide insights into developing a reconfigurable feature-classifier architecture to accommodate environmental variability.

Classification Cramer-Rao Bounds

Let us make a suite of measurements y that can be described by the probability function $p_\theta(y)$, where θ parametrizes $p(y)$ and $p_\theta(y) = p(y|\theta)$. If $z = f(y)$, where the dimension of z is smaller than that of y and $p_\theta(y|z) = p(y|z)$, then we say that z captures all the useful information in y. Furthermore, z is more memory efficient than y since $f(\cdot)$ compresses y into a sufficient statistic (7,20).

Sufficient statistics are closely related to class separability. In general, optimality score J is measured by

$$J(y, h, z_\Omega) = \frac{1}{N_y} \int_{y=h(z_\Omega)} \mathrm{CS}[p_{\theta_1}(y|z_\Omega), \ldots, p_{\theta_{N_c}}(y|z_\Omega)]\, dy \quad (9)$$

where N_y is the dimension of y, z_Ω is the overlapped region (between two classes) in z that gets projected onto y via a mapping operator $h(\cdot)$ ($h(\cdot)$ is in essence $f^{-1}(\cdot)$ and a function of a classifier structure), and $\mathrm{CS}(\cdot)$ is a class separability function that measures the degree of feature space overlap between classes. In essence, a classifier performs the $f(\cdot)$ operation. Therefore, θ is equivalent to class label while y and z denote an input feature vector and a classification LLR score, respectively. In short, the degree of sufficient statistics can be measured by class separability in the multidimensional feature space Ω.

This concept can be reinforced with an interesting two-class, two-feature problem as shown in Fig. 13. In this case, we use the following two classifiers:

1. *Linear Fisher's Classifier (LFC).* This is a simple boundary-decision classifier that computes a weight vector ω that maximizes the Rayleigh quotient $\omega^t S_b \omega / \omega^t S_w \omega$, where ω is the first eigenvector of the following generalized eigenvalue problem.

$$S_b x = \lambda S_w x \quad (10)$$

where $\lambda_1 > \lambda_i, i > 1$. S_b and S_w refer to the interclass and within-class scatter covariance matrices, respectively. For a two-class problem, ω can be directly computed by

$$\omega = S_\omega^{-1}(\mu_1 - \mu_2) \quad (11)$$

where μ_i is the ith class mean vector. The LLR score can be approximated as $\omega^t y$, where y is an input test feature vector. Frequently, it is possible that the two classes may share the same mean vectors, but can be differentiated by the difference in the covariance matrices. In this case, we can use the generalized likelihood ratio test (GLRT) concept to derive the weight vector as the eigenvector of $R_1^{-1}R_2$ associated with the largest eigenvector, where R_i is the ith class covariance matrix. In short, depending on the estimate of $\Delta\mu$,

$$\omega = \begin{cases} S_\omega^{-1}(\mu_1 - \mu_2) & \Delta\mu > \gamma \\ \text{eigenvector of } R_1^{-1}R_2 & \text{otherwise} \end{cases} \quad (12)$$

A successive implementation of LFC coupled with token pruning (i.e., feature vectors or tokens that fall into separable regions are pruned so that the next stage LFC works with the remaining feature tokens—successive approximation of class-conditional pdfs) at each stage forms the backbone of a discriminant neural network (DNN) architecture (4).

2. *Multivariate Gaussian Classifier (MVG).* This is a parametric classifier that assumes that the multidimensional feature pdf can be characterized by its mean vector μ and covariance matrix R. Mathematically, it computes the Mahalanobis distance associated with each class and selects the class with the shortest distance:

$$d(i) = (y - \mu_i)^t R_i^{-1}(y - \mu_i) \quad (13)$$

$$i_y = \underset{1 \le i \le N_c}{\arg\min}\, d(i) \quad (14)$$

$$\mathrm{LLR}_{ij} = d(i) - d(j) \quad (15)$$

where i and N_c refer to the class index and the number of classes, respectively. i_y is the selected class label for an input test feature vector y.

For this problem, the two class-conditional pdfs—$p_{\theta_1}(y)$ and $p_{\theta_2}(y)$—are both normal with the same covariance matrix, but with different mean vectors. Naturally, MVG or LFC with ω that maximizes the Rayleigh quotient is the Bayes classifier.

In order to measure the extent to which MVG captures useful information present in the two input features, the following class separability function is used:

$$\mathrm{CS} = |p_{\theta_1}(y|z_\Omega) - p_{\theta_2}(y|z_\Omega)| \quad (16)$$

where z_Ω is the region in z with high class overlap. As expected for a class separability measure, $\mathrm{CS} \approx 0$ when $p_{\theta_1}(y|z_\Omega) \approx p_{\theta_2}(y|z_\Omega)$. The areas in z with relatively little class overlap are excluded since prediction errors in those regions are minimal. That is, we zero in on the area with most predic-

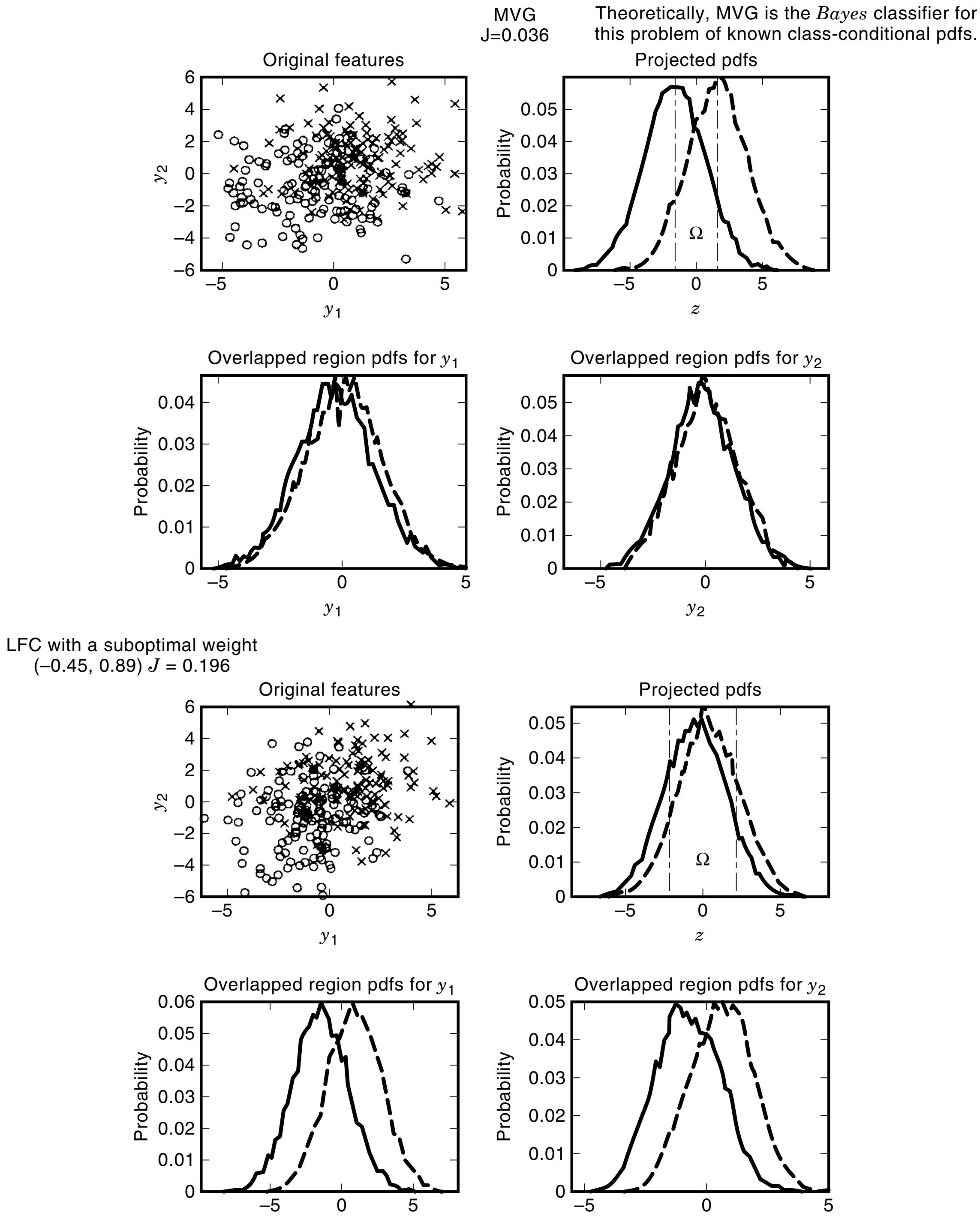

Figure 13. For a two-class problem with multivariate Gaussian pdfs, MVG is the Bayes classifier. MVG and LFC with a suboptimal weight vector of $[-0.45, 0.89]$ yield J of 0.036 and 0.196, respectively. The J score of zero means that the two class-conditional pdfs in y derived from the overlapped region in z (i.e., Ω) completely overlap—capturing all the useful information in the original feature space y.

tion errors to investigate the extent to which prediction performance can be further improved.

For comparison, LFC with a suboptimal weight vector ω of $[-0.45, 0.89]$ in $z = \omega^t y$ was implemented. As expected, MVG performs far superior to LFC as evidenced by a smaller amount of class overlap in z. More important, the optimality score J for MVG is much lower than that for LFC. Based on numerous experiments with a number of known and unknown class-conditional pdfs, J of less than 0.0375 implies that a classifier is in essence the Bayes classifier (21). That is, the correct classification performance of around 70% in this case cannot be further improved by changing the classifier architecture. Instead, we should concentrate on gathering additional input data to improve the information content.

Situationally Adaptive Target Recognition

Environmental robustness requires that target-recognition algorithms be insensitive to extraneous confusion factors. In real-time implementation, we employ the following strategies to mitigate the negative impacts of environmental variation on target-recognition performance:

1. Implement more features that absolutely necessary for automated feature subset selection as a function of environment.
2. Train classifiers adaptively by joint supervised and unsupervised learning (22). In essence, the original class-conditional pdfs are used as a starting point and as the system receives new data, it adaptively adjusts or estimates "slightly" different new class-conditional pdfs using a combination of self-organizing feature mapping and expectation-maximization algorithms.
3. If possible, collect and process new data with known ground truth.
4. Develop software toolboxes to facilitate rapid in situ algorithm optimization. Typical toolboxes deal with ground truthing and target-cluster segmentation, pattern recognition, and environmental prediction.

Nevertheless, it is imperative that we resort to a totally integrated approach that sequentially removes as much clutter as possible while accommodating environmental uncertainties. This approach would entail robust adaptive joint time-space filtering, matched-field processing, acoustic tomography, detection, reconfigurable feature extraction and classification (2), localization, and multiping- and multisensor-based fusion. Recent advances in acoustic communication permit in situ acoustic channel calibration that can be used to model the extent of target-signature distortion caused by rapid channel fluctuations (23). Furthermore, several university research teams are investigating how dolphins and bats use acoustic sonars to make fine discriminations between objects with small differences in material composition, shape, and interior in an adaptive fashion despite environmental variations (24,25). These research activities can shed insights into the processing architecture of the future sonar target-recognition system. As computing power doubles every 18 months according to Moore's law, we are bound to witness an integrated sonar target-recognition system that can adapt to changing environmental conditions to provide robust performance in four to seven years.

ACKNOWLEDGMENT

The authors would like to thank Dr. Weita Chang, Dr. Rick Wayland, Dr. Dick Heitmeyer, Tom Hayward, and Dr. Marshall Orr for their support. The research efforts mentioned here were supported by the Naval Air Warfare Center under Contract No. N62269-94-C-1179, the Office of Naval Research under Project No. RJ14C42, and the Naval Research Laboratory under Contract No. N00014-93-C-2246.

BIBLIOGRAPHY

1. R. J. Urick, *Principles of Underwater Sound for Engineers,* New York: McGraw-Hill, 1967.
2. D. Kil, F. Shin, and R. Fricke, LFA target echo characterization with hidden Markov models and classifiers, *J. Underwater Acoust.,* **41** (7): July 1995 (a special theme issue on multisensor fusion).
3. W. Chang and B. Bosworth, Performance comparison of neural network and conventional classifiers and significance of feature set for single ping active classification, *Naval Undersea Warfare Center (NUWC) TR Report No. 10743,* January 1995.
4. D. H. Kil and F. B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization,* Woodbury, NY: AIP Press, 1996.
5. L. L. Scharf, *Statistical Signal Processing,* Reading, MA: Addison-Wesley, 1991.
6. M. I. Skolnik, *Introduction to Radar Systems,* New York: McGraw-Hill, 1980.
7. D. Kil and F. Shin, A unified approach to hierarchical classification, *Proc. ICASSP,* VI, Atlanta, GA, May 1996, pp. 1549–1552.
8. D. W. Tufts, D. H. Kil, and R. R. Slater, Reverberation suppression and modeling, in D. D. Ellis, J. R. Preston, and H. G. Urban (eds.), *Ocean Reverberation,* Boston: Kluwer, 1993.
9. O. Ghitza, Auditory models and human performance in tasks related to speech coding and speech recognition, *IEEE Trans. Speech Audio Process,* **2** (II): 115–132, 1994.
10. D. Kil, F. Shin, and R. Wayland, Active impulsive echo discrimination in shallow water by mapping target physics-derived features to classifiers, *IEEE J. Oceanic Eng.,* **22**: 66–80, 1997.
11. J. Jeong and W. J. Williams, Kernel design for reduced interference distributions, *IEEE Trans. Signal Process,* **40**: 402–412, 1992.
12. C. L. Nikias and M. R. Gaghuverr, Bispectrum estimation: A digital signal processing framework, *Proc. IEEE,* **75**: 869–891, 1987.
13. C. Myers et al., Modeling chaotic systems with hidden Markov models, *Proc. ICASSP,* IV, San Francisco, CA, April 1992, pp. 565–568.
14. D. Kil and F. Shin, Reduced dimension image compression and its applications, *Proc. Int. Conf. Image,* III, Washington, D.C., October 1995, pp. 500–503.
15. C. N. Corrado, Jr., Mid-frequency acoustic backscattering from finite cylindrical shells and the influence of helical membrane waves, Ph.D. dissertation, MIT, Cambridge, MA, 1993.
16. M. Wax, Detection and estimation of superimposed signals, Ph.D. dissertation, Stanford University, Stanford, CA, 1985.
17. A. S. Weigend and N. A. Gershenfeld (eds.), *Time Series Prediction,* Reading, MA: Addison-Wesley, 1994.
18. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE,* **77**: 257–285, 1989.
19. F. B. Shin and D. H. Kil, Full-spectrum processing using a classify-before-detect paradigm, *J. Acoust. Soc. Am.,* **99**: 2188–2197, 1996.
20. E. Real, Feature extraction and sufficient statistics in detection and classification, *Proc. ICASSP,* VI, Atlanta, GA, May 1996, pp. 3049–3052.

21. D. Kil and F. Shin, Cramer-Rao bounds on stock price prediction, *J. Forecasting,* 1997 (a special issue on neural networks).

22. B. Shahshahani and D. Landgrebe, Classification of multi-spectral data by joint supervised-unsupervised learning, *TR-EE-94-1,* (Purdue University Technical Report), Purdue University, Lafayette, IN, January 1994.

23. M. Johnson, M. Grund, and D. Brady, Reducing the computational requirements of adaptive equalization in underwater acoustic communications, *Proc. Oceans,* III, San Diego, CA: October 1995, pp. 1405–1410.

24. A. Simmons, Biosonar acoustic imaging for target localization and classification by bats, *SPIE Conf. 3079,* Orlando, FL: April 1997, pp. 7–13.

25. N. P. Chotiros et al., Observation of buried object detection by a dolphin, *SPIE Conf. 3079,* Orlando, FL: April 1997, pp. 14–18.

DAVID H. KIL
FRANCES B. SHIN
Lockheed Martin