Mine Detection Using Scattering Parameters and an Artificial Neural Network

Gregory L. Plett, *Student Member, IEEE,* Takeshi Doi, and Don Torrieri, *Senior Member, IEEE*

Abstract—***The detection and disposal of anti-personnel land mines is one of the most difficult and intractable problems faced in ground conflict. This paper presents detection methods which use a separated-aperture microwave sensor and an artificial neural-network pattern classifier. Several data-specific preprocessing methods are developed to enhance neural-network learning. In addition, a generalized Karhunen-Loeve transform and the eigenspace separation transform are used to perform data reduction and reduce network complexity. Highly favorable results have been obtained using the above methods in conjunction with a feedforward neural network.***

Index Terms—**Anti-personnel land-mine detection, neural-network applications, pattern classification, feature extraction, dimensionality reduction, statistical outlier removal, microwave sensor**

I. Introduction

ANTI-PERSONNEL land-mine detection is a problem of significant military, economic and humanitarian concern. Not only do land mines pose a deterrent to military activity during conflict, they also remain lethal long after that conflict is over. The statistics are staggering; eighty percent of the victims of anti-personnel mines are civilians, many of whom are children [1]. The number of such mines currently in place is estimated at over 100 million, and is increasing at the rate of one half to one million mines per year. There is great motivation to research new counter-mine systems. Current anti-personnel land mines cost between \$3–\$25 per mine and clearance methods cost between \$300–\$1000 per mine removed [2]. A mine detection system which will operate with a high degree of accuracy while being simple enough to deploy at low cost is needed.

One of the problems faced by the countermine system is that newer mines contain very little metal, and thus it is not feasible to use conventional metal detectors to locate them. A more sophisticated approach is presented here, where neural-networks are trained to detect simulated land mines using data collected from a mine lane facility located at Ft. Belvoir, VA. This data consists of a set of measurements made with a *separated aperture sensor* operating in the *waveguide near cutoff* mode [3]. Previous work has shown that neural-network pattern classifiers can successfully detect anti-tank mines [3], [4], [5], and the present work addresses the much more difficult problem of detecting smaller anti-personnel mines.

Emphasis is placed on the processing of sensor data for use with a neural-network classifer in order to maximize detection rates while minimizing false-alarm rates. The structure and type of data used as input to the network is explained and justified. Three methods which mitigate certain undesirable and uncontrollable properties of the training data set are discussed. These procedures permit better neural-network learning and improved generalization. In addition, a generalized Karhunen-Loeve transform and the eigenvalue separation transform are explored as effective preprocessing methods to perform data reduction and increase mine detection. Finally, the performance of artificial neural-network mine detectors incorporating these methods are summarized.

Manuscript received June 25, 1996; revised January 22, 1997. This work was supported by ARMY Contract DAAK70–92–0003.

G.L. Plett and T. Doi are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9510.

D. Torrieri is with the Army Research Laboratories, 2800 Powder Mill Road, Adelphi, MD 20783.

Publisher Item Identifier S 1045–9227(97)05239–9.

II. Sensor Description and Data Acquisition

An experimental apparatus at Ft. Belvoir was used to obtain the training and test data from a simulated mine lane. The fundamental operating principles of the separated aperture sensor are described in references [3] and [6]. A block diagram of the measurement system is shown in Fig. 1. A Hewlett Packard 8753A Network Analyzer capable of generating sinusoidal signals between 300 kHz and 3 GHz was attached, through an HP 85046B *S*-parameter test set, to a separated aperture sensor. The sensor, located (nominally) 2” above the ground, consisted of a transmitting aperture or horn which injected microwaves into the soil, and a receiving aperture which provided a return signal to the Hewlett Packard equipment. For the data set used in the current work, the resonant frequency of the horns was 1 GHz. Figure 2 shows a pictorial diagram of the sensor. The horns look like metal troughs with the dipole antennae running longitudinally.

Both the signal detected by the “receiving” horn and the return signal picked up by the transmitting aperture were sensed. The HP *S*-parameter test set automatically recorded the complex-valued scattering parameters or *S*-parameters of the two-port network formed by the microwave horns and the earth waveguide. The tests used continuous-wave sine waves, and standing waves were detected and used to compute the *S*-parameters. The horns were connected directly to the *S*-parameter test set, and the measurements were repeated five times and averaged to reduce noise.

The energy radiated from a dipole antenna is transmitted in both electric and magnetic waves. It has been found that the leakage mode which propagates across the ground surface is a transverse electric mode wave [7]; the detection mode, however, is largely transmitted through the soil over the target. There is some energy absorbed by the target, with the strongest absorbed field component parallel to the dipoles.

The mine lane facility at Ft. Belvoir was 790.5” long and 40.5” wide. Data was collected every 1.5” across the length and width of the lane, thus forming a 527 row by 27 column grid

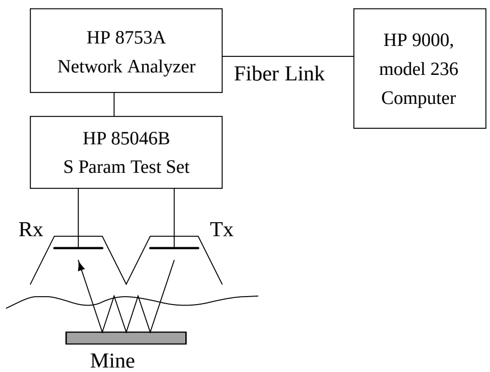

Fig. 1. Block diagram of the measurement setup.

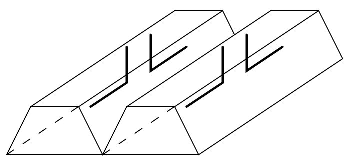

Fig. 2. Diagram of the microwave horns and antennae.

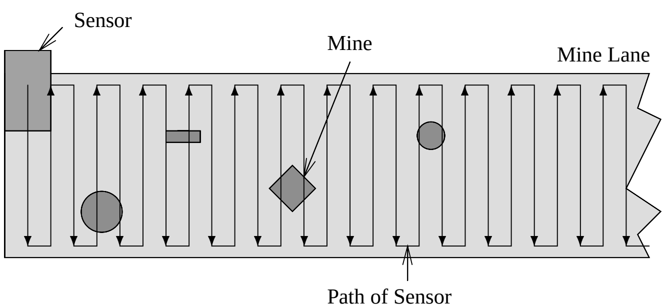

Fig. 3. The movement of the sensor over the mine lane during data acquisition.

of measurements. At each point, all S-parameters were measured (S_{11},S_{12},S_{21},S_{22}). Eleven different frequencies in 40 MHz intervals were used ranging from 800 MHz to 1200 MHz, inclusive. At each point, the magnitude and phase component of each S-parameter was measured for a total of 88 ($4\times11\times2$) measurements at each location in the grid. Figure 3 shows how the sensor head was scanned over the mine lane.

Previous data sets were acquired using a wheeled cart which housed the data collection equipment. The more recent data was collected by a system hung on an overhead rail. The measuring system was controlled at a distance by an HP 9000 computer via optical fiber (HP-IB) link. The computer was able to move the sensor head laterally with a worm gear control and record all measurements automatically. To move the sensor longitudinally (from row to row), a human operator moved the apparatus the required 1.5” along the overhead rail. In addition, the new data set was measured over a mine lane of sandy soil which included five different mine types:

1. 6” diameter × 3” wood disk.
2. 5” diameter × 3” hard plastic disk.
3. 3.5” diameter × 3” nylon disk.
4. 6” × 6” × 3” wood block.
5. 4” × 2” “butterfly” shaped plastic and metal target.

There were a total of 114 targets: 15, 30, 12, 15 and 42 of each mine type, respectively.

III. Neural Networks for Pattern Recognition

When considering the implementation of any pattern recognition system (PRS), careful thought must be given to several issues:

1. Which “features” from the original data should be used as input to the PRS?
2. What pattern recognition methodology should be used?
3. What should the outputs of the PRS be?
4. What is the basis for evaluating the performance of a particular system, and comparing the behavior of two competing systems?

The third question can be answered based on the objective at hand. Ultimately, it would be desirable to be able to both detect mines and discriminate between mine types; however, it is far more important that the network is able to correctly detect mines in the first place. Consequently, for this work, the output of the PRS was constrained to be binary. A positive output for a given input pattern indicated that a mine was thought to be present at that spatial location, and a negative output indicated that a mine was considered absent.

This decided, the fourth question can now be addressed. Simply, we wish to maximize the detection rate of mines. However, care must be taken to ensure that the PRS is not simply “memorizing” mine patterns but is also able to generalize to patterns not previously seen. Hence, we wish to maximize the generalization ability of the PRS.

To accomplish this goal and to train and test the PRS, the mine lane was divided into five approximately equally spaced sections. Data from the first, third and fifth sections were used for training, and the second and fourth sections were used only for testing. This resulted in 8343 training patterns and 5886 test patterns. The large number of test patterns allowed an accurate measurement of the PRS’s generalization ability.

A further performance consideration is the rate of false positives (false alarms) generated by the PRS. An overly high false-positive rate will cause the human operator to distrust its results, rendering the device useless. Our composite performance criteria is then: (1) Maximize the true positive rate in the testing region. Should two results be equal in this regard, (2) Minimize the false-positive rate in the testing region. A third, qualitative, consideration is the complexity of the PRS. If two competing systems have otherwise equivalent performance, the simpler system is preferred.

Deciding which pattern recognition methodology to use is a more difficult issue. Statistical approaches abound, as do different sorts of neural and hybrid systems [8], [9], [10]. Several of these methodologies were investigated, with the conclusion that the neural-network based system is the least complex, and yields the best performance (Specific results are deferred until section IX).

It is assumed that the reader is familiar with neural networks. If not, the tutorial paper [11] is an excellent place to start. In short, the following are three valuable attributes of neural networks which are not generally shared by other computing systems:

- Neural networks are not *programmed*. Rather, they are *trained* to perform highly refined tasks for which no precise rules exist. One such task is pattern recognition (e.g., the detection of mines in a minefield), in situations where no empirical or theoretical rules are known.
- Neural networks can generalize, that is, deliver accurate responses for inputs that were not presented during training. This is a crucially important feature since no system can be trained on all possible mine and background patterns.
- Neural networks can be continuously trained, keeping up with changing input environments. This is important if the system must work in different settings. (e.g., sandy soil versus loamy soil).

The final question to consider when constructing a pattern recognition system is "Which features from the data should be used as input to the PRS?" It is interesting that this is even an issue since, for example, the *data processing inequality* from information theory [12] states that "no clever manipulation of the data can improve the inferences that can be made from the data." While this is true, the correct choice of features will minimize the amount of useful information lost, while maximizing the attenuation of irrelevant information. It becomes critical to wisely choose features which will minimize the complexity of the PRS, and maximize its speed and detection capability. Hence, much of this paper is dedicated to answering this question.

IV. Minimal Window

In the initial research, each training and test pattern applied to the neural network was obtained by using a square spatial window of data from many nearby points. These spatial windows ranged in size from 5×5=25 points to 13×13=169 points. At each point in a window, only the S_{11} parameter for each of a number of frequencies was included in the pattern. The decision as to whether or not a mine was actually present at the center of the window was based on data collected over the entire window area.

Although the use of spatial windows yielded good results, their application to a practical mine detector presents great difficulties because of their inherent operational requirements. Consequently, the spatial windows were abandoned. Instead, each pattern applied to the neural network was obtained from data measured at a single point. This new approach could be interpreted as the use of a minimal (1×1) window. The principal advantages of a minimal window are:

1. Easier and faster measurements,
2. The possibility of mine detection near natural or man-made obstacles encountered in the field,
3. The possibility of collecting larger sets of training and test patterns by including points near the edges of the mine lanes,
4. A reduction in the complexity of the neural network, and,
5. Better accommodation of irregular terrain.

A minimal window requires that each input pattern applied to the neural network be derived from measurements taken from a single location. These measurements potentially include the magnitudes and phases of all four scattering parameters at a number of frequencies. However, extensive simulation trials have indicated that little or no performance improvement is obtained by including the phase measurements or by including measurements at more than seven frequencies. Thus, each input pattern can be reduced to only 28 components—the magnitudes of the four scattering parameters at the seven central frequencies.

V. Data Flattening

One significant deterrent to automated mine detection was found to be caused by large, step-like jumps in magnitude which periodically occur between successive rows of a data set[1]. Figure 4 shows one example jump in the 800 MHz S_{11} data between rows 414 and 415.

The dramatic instantaneous change in signal level between rows leads us to postulate that a period of time had elapsed between measurements along one row and the next, and that ambient conditions changed during that interval. This conclusion is supported by the fact that the data-collection system automatically recorded data along a given row, but that operator intervention was required to move the apparatus from row to row. Since no jumps *within* a given row were observed, the phenomenon must be related to elapsed time between row measurements. It is possible that the jumps are caused by a change in the ambient temperature or humidity of the mine lane from one day to the next. At the beginning of the day, the conditions would be significantly different from the evening before, and a jump would occur between rows in the data. In addition to jumps, there is a significant downward slope in the data, which means that even a short lunch break could be the cause of a jump in the data (as evidenced by some jumps "down"). Unfortunately, not all of these jumps are detectable or removable.

The jumps are artificial in the sense that they would not occur in a field system searching for mines. Furthermore, the jumps decrease the network's performance since many different signal levels artificially correspond to mine patterns. Additionally, the "minimal window" paradigm prohibits the use of adjacent spatial data as network input to help resolve the problem. Overall, a much simpler network can be used to detect mines if the jumps are removed.

A preprocessing function attempts to eliminate the jumps by globally processing each data set. First, the mean and standard deviation of both the magnitude and phase components of each data set are computed. Next, the average magnitude and phase value for each row of data are computed. If the difference in the average of either the magnitude or phase between two successive rows is greater than one standard deviation of the data set, a jump is hypothesized to have occurred at this location. However, it is possible that this change in level might be due to a high-variance change in the "true" data. To guard against this possibility, once a jump is postulated to have occurred, it is verified by checking that all the columns change value in the same direction across the jump. It is unlikely that all columns would change significantly in the same direction if no artificial jump were present.

[1] The term "data set" here refers to 527×27 magnitude and phase measurements made for one S parameter and frequency combination.

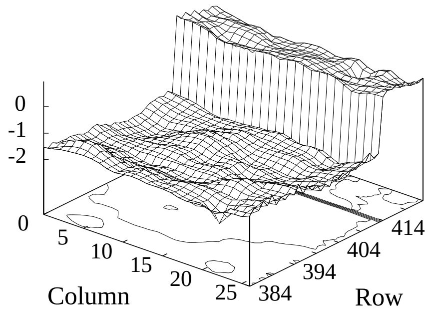

Fig. 4. Unflattened data.

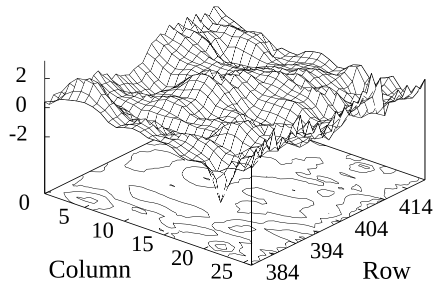

Fig. 5. Flattened data.

The ramps, and thus the jumps at the ramp edges, are removed as follows: for each interval of data between two jumps, a linear regression calculation of the best least squares fit of a line to the average row magnitudes is made. This is the approximation of the "ramp" experienced during measurement. This line is extended over all of the columns to form a plane, and is subtracted from the data.

For example, the 800 MHz S_{11} data file has original average row magnitudes as displayed in Fig. 6. Many jumps are clearly evident; consider the one between rows 414 and 415 (the highlighted section between the arrows corresponds to the three-dimensional region and jump drawn in Fig. 4). When the jumps are removed using this method, the new average row magnitudes are as shown in Fig. 7. Looking at the data in three-dimensional format (Fig. 5) we see that the jump in consideration has been removed.

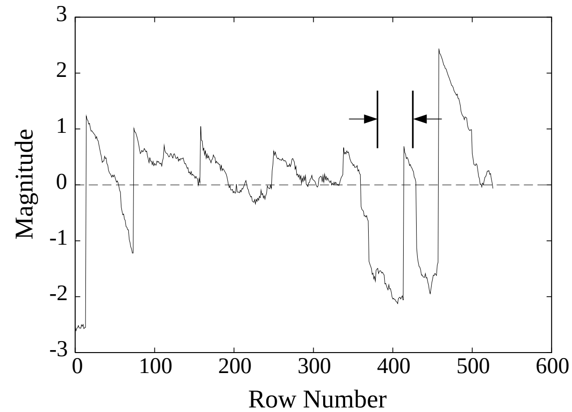

Fig. 6. Unflattened average row magnitudes.

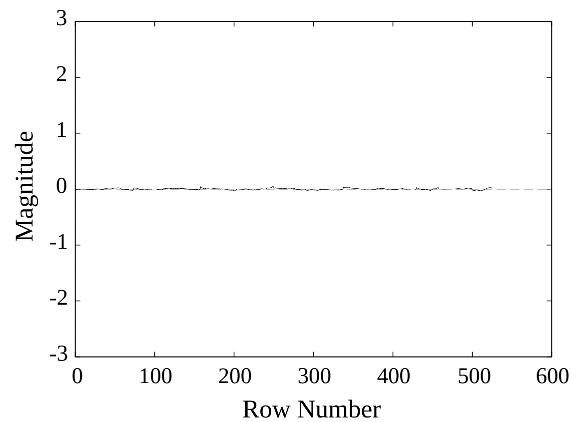

Fig. 7. Flattened average row magnitudes.

VI. Outlier Removal and Balanced Data Sets

Supervised training of a neural network requires the association of a desired response with each training and test pattern. An assignment of a desired response was possible for each location in the mine lane because the size, orientation, and center positions of the mines were known. At each location used to measure a pattern, a positive response was assigned if a mine overlapped the 1.5"×1.5" grid square with that location at its center. A negative desired response was assigned if the grid square was not overlapped by a mine. A training or test pattern with a positive desired response is called a mine pattern, and one with a negative desired response is called a background pattern.

The generalization capability of neural networks can be impaired by overfitting noisy training patterns. The separating surface generated by a neural-network classifier may exhibit excessive fluctuations rather than the smooth boundary needed for good generalization. To remove from the training set those statistical outliers which are the cause of overfitting, two approaches may be considered. In the first approach, either vector quantization or a nearest-neighbor mathematical clustering technique is used to find anomalous patterns and then remove them from the training set. Alternately, an analysis can be done to ascertain which physical conditions might cause certain patterns to be outliers. Both methods were pursued in this work, with the finding that the much simpler second method consistently produced equal or better results. Details of this latter approach are presented here.

Outliers are most likely to be generated when the sensor is placed near the edge of a mine, primarily because of the potential multipath effects. For example, at a mine boundary, a negative desired response may be assigned based on the sensor position, even though the training pattern is much more typical of a mine pattern or is much different from the other training patterns. If all the training patterns measured near mine edges are regarded as potential outliers and are removed from the training sets, one would expect improved neural-network generalization.

To eliminate potential outliers, a pattern measured at a specific grid point was retained in the training set only when the four nearest surrounding points had the same desired response. Along the edges and corners of the mine lane, a missing neighbor was excluded from the requirement. This method of remov-

```
 1:  ...........................
 2:  ...........................
 3:  ...........................
 4:  .................=----=....
 5:  ................=-1111-....
 6:  ................-11111-....
 7:  ................-11111-....
 8:  ................-11111-....
 9:  =---=...........=-1111-....
10:  -333-............=----=....
11:  -333-......................
12:  -333-......................
13:  =---=......................
```

Fig. 8. Section of mine lane showing points removed by the four-nearest-neighbor and eight-nearest-neighbor methods.

ing outliers is called the four-nearest-neighbor outlier-removal method. Additional neighbors can be included in the criterion, leading for example, to the eight-nearest-neighbor method.

An immediate problem resulting from the use of the nearest-neighbor methods was that the number of mine patterns was significantly reduced, impairing the generalization capability of the neural network. To prevent this problem, only background patterns were subjected to the nearest-neighbor methods. All mine patterns were retained in the training set. To illustrate this concept, Fig. 8 depicts the mine lane for rows 1–13. Points in the mine lane which correspond to a grid square overlapping some section of a mine are labeled with the numeral one or three, corresponding to the type of the mine. Points corresponding to a background pattern which would be removed by either the four-nearest-neighbor or eight-nearest-neighbor methods are noted with '–', and points corresponding to background patterns which would be removed only by the eight-nearest-neighbor method are noted with '='. Purely background points are noted with '.'.

We remark that the subsequent scarcity of training patterns measured in the vicinity of mine edges may cause a mine detector to be insensitive to the presence of a mine when the sensor is placed above the edge of a mine. However, in a practical implementation, the mine detector will probe the central part of the mine at some time during the detection process. Therefore, the exclusion of statistical outliers will not degrade a practical implementation.

One further comment relating to the nature of the data is that the backpropagation algorithm used for neural-network training converges very slowly, if at all, for binary classification problems in which the bulk of the training patterns belong to a single class. In the presence of such a data imbalance, one approach to accelerating the convergence is to modify the backpropagation algorithm [13]. A much simpler approach, which was adopted, is to artificially create a balanced training set of approximately equal numbers of mine patterns and background patterns. The balanced training set is created by replicating each pattern in the smaller class, the class of mine patterns, a sufficient number of times. As a result, the mine and background patterns are presented to the neural network with approximately the same frequency during the training phase. The use of a balanced training set together with the nearest-neighbor outlier-removal method produced a substantial improvement in the simulated performance of the mine detector relative to its performance using the original unbalanced set.

VII. The Karhunen-Loeve and Generalized Karhunen-Loeve Transforms

The Karhunen-Loeve (KL) transform [10] is a linear transformation of the form:

$$Y = A^T X. \tag{1}$$

where A is an orthonormal transformation matrix, and X is a k-dimensional input vector. The motivation behind using this transformation on the mine lane data is twofold. First, the pattern space may be transformed in such a way that a network is more capable of separating the mine patterns from the background patterns. Secondly, data reduction can be achieved; that is, we can compute $Y_m = A_m^T X$ where A_m contains the first $m \leq k$ columns of A, and Y_m has the resulting dimension m. Generally, better separation and lower dimensionality can be attained at the same time, and the data reduction is achieved without a significant loss of "energy" from the input data, and thus without a significant loss of performance. This data reduction will result in faster training and faster and more compact network realizations.

For the KL transform, the matrix A is set to be the modal matrix of X which can be found as follows: First, compute the covariance matrix for X:

$$\Sigma_X = E[(X - \overline{X})(X - \overline{X})^T]. \tag{2}$$

Then, compute the eigenvalues, λ_i, and eigenvectors, u_i, of Σ_X. Since Σ_X is symmetric and positive semi-definite, all the eigenvalues will be non-negative and real. Without loss of generality, we can order the eigenvalues such that:

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_k \geq 0. \tag{3}$$

Compose the matrix A by setting its columns equal to the ordered eigenvectors:

$$A = [u_1 u_2 \cdots u_k]. \tag{4}$$

If we delete columns from A in order to reduce the dimension of Y, (for example, if we retain the first m components and discard the rest) it can be shown that the resulting mean square error of a reconstruction is:

$$E[\|\epsilon\|^2] = \sum_{i=m+1}^{k} \lambda_i. \tag{5}$$

Clearly, if we are to minimize the error while discarding components from Y, we must discard those components with the smallest eigenvalues.

In practice the covariance matrix Σ_X was estimated by taking all N input vectors $X_i \in \{X_1 \ldots X_N\}$ from the training set and computing

$$\widehat{\Sigma}_X = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})^T, \tag{6}$$

where

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i. \tag{7}$$

The KL transform presents a method of compactly representing information while maximizing the preserved energy. However, we recognize that it is not the optimum way to process data as input to a pattern recognition system, since the data is inherently from two or more distinct classes, and this distinction is not exploited. Rather, we would like to maximize the average distance between patterns of the two classes. Toward this end, we have generalized the concept of the KL transform such that A contains the eigenvectors of the modal matrix of background patterns only. This transform then produces components of each pattern of the two classes in such a way that the components of one class have a larger average energy in certain directions than the components of the other class. In principle, the correlation matrix calculated from only the mine patterns could be used instead, but this approach was not adopted because of the relative scarcity of mine patterns and the consequent potential inaccuracy in the computations. Furthermore, the next section presents an alternate method which maximally separates the average distance between patterns of each class.

VIII. The Eigenspace Separation Transform

The eigenspace separation (ES) transform [14] allows one to reduce the neural-network size while enhancing its generalization accuracy as a binary classifier. The transform projects each training or test pattern into an orthogonal subspace and thereby reduces the dimensionality of the input vector applied to the neural network. The generalization improvement is possible because the orthogonal projection transforms the two classes of patterns in such a way that the vectors of one class tend to have an average length greater than those of the other class. The KL transform preserves information about the inputs in the sense of minimizing the mean square error. The ES transform preserves and makes more accessible the information that is most relevant to a classifier: information that allows separation of data clusters.

The calculation of the transformation matrix U proceeds as follows.

1. Compute the $k \times k$ matrix,

$$\widehat{M} = \frac{1}{N_1}\sum_{i=1}^{N_1} X_{1i}X_{1i}^T - \frac{1}{N_2}\sum_{i=1}^{N_2} X_{2i}X_{2i}^T, \tag{8}$$

   where X_{1i} is mine pattern i, N_1 is the number of mine patterns in the training set, X_{2i} is background pattern i, and N_2 is the number of background patterns in the training set.
2. Calculate the eigenvalues of $\widehat{M}$, λ_i, $i = 1, 2, \ldots, k$.
3. Calculate the sum of the positive eigenvalues

$$E_p = \sum_{\substack{i=1 \\ \lambda_i > 0}}^{k} \lambda_i, \tag{9}$$

   and the sum of the absolute values of the negative eigenvalues

$$E_n = \sum_{\substack{i=1 \\ \lambda_i \le 0}}^{k} |\lambda_i|. \tag{10}$$

4.(a) If $E_p > E_n$, then calculate all m orthonormal eigenvectors that are associated with the positive eigenvalues of $\widehat{M}$. Use these eigenvectors as the columns of the $k \times m$ matrix U.
(b) If $E_n > E_p$, then calculate all m orthonormal eigenvectors that are associated with the negative eigenvalues of $\widehat{M}$. Use these eigenvectors as the columns of the $k \times m$ matrix U.
(c) If $E_p = E_n$, then either subset of orthonormal eigenvectors can be used in constructing U. In general, choose the smaller subset if possible.

After the matrix U is constructed, $Y_i = U^T X_i$ is computed for each pattern X_i in the training and test data sets. Then each Y_i is applied to the neural network.

IX. Simulation Results

This section summarizes the performance achieved using a neural network for land-mine detection. Consideration is also given to statistical methods for pattern recognition. The effect of modifying some of the variable factors in this system are discussed. Some of the relevant result tables are given here and others are available in reference [15].

A. Neural Mine Detection

Various simulation experiments were conducted in order to design a neural network mine detection system. As mentioned previously in section III, the performance of competing results was judged primarily on the basis of maximizing the ability of the system to detect mines for patterns not previously seen or trained on. A second objective was to minimize the false-alarm rate, and a third to minimize complexity. In order to measure generalization ability, the mine lane was divided into five approximately equally spaced sections. Data from the first, third and fifth sections were used for training, and the second and fourth sections were used only for testing. This resulted in about 8343 training patterns and 5886 test patterns (the numbers varied slightly depending on whether statistical outliers were retained or removed). The large number of test patterns allowed an accurate measurement of generalization ability.

Different candidate neural network architectures were investigated. These included fully-connected feedforward networks and partially-connected or bottlenecked networks. The bottlenecked networks were studied in an attempt to reduce the number of weights in the network. Since the majority of the weights are in the first hidden layer of the network, each neuron in the first hidden layer was constrained to select its inputs from a small subset of the available inputs. Two partially-connected architectures were studied: in one configuration, the inputs were divided into four groups corresponding to the four S-parameters; in the other configuration, the inputs were divided into seven groups corresponding to the seven frequencies used in the measurements. It was found, however, that a fully connected network for which the input data was preprocessed using either of the transform methods (hence, reducing the dimension of the input data) provided superior performance. This is not surprising, since the transformations intelligently choose the “features” to be used, which may not necessarily be split by frequency or *S*-parameter. Hence, fully connected feedforward neural networks are considered to be the best choice for this work.

In addition to the interconnectivity, the size of the network was studied. It was found that a three-layer (two hidden layers of neurons and one output layer) network gave the best results. A two-layer network had difficulty modeling the complex relationship between the input signals and the classification output, and a four-layer network was unnecessarily complex. Through simulation, the optimum number of neurons for each layer was established. Network sizes ranging up to eighteen neurons in the first hidden layer, twelve in the second hidden layer and one output neuron were tested. The conclusion of this survey was that detection capability was not terribly sensitive to the specific architectures within this range, and that networks with eight neurons in the first hidden layer, three neurons in the second hidden layer, and one neuron in the output layer (8:3:1) consistently showed good results. This architecture is what was used in all results presented here.

At each spatial location in the mine lane, 88 measurements were taken and may be used as input to the mine detection system. However, making this large number of measurements was considered to be a hindrance to a practical detection system. Therefore, simulations were conducted to determine the most vital parameters for mine detection. For neural networks of comparable size (similar number of weights), it was determined that only seven frequencies (880 MHz to 1120 MHz in 40 MHz intervals) are necessary to obtain satisfactory results. Using additional frequencies added very little to performance, and using fewer frequencies started to significantly degrade the performance. The selection of frequencies is intuitively justified by noting that the resonant frequency of the measurement antennae was 1 GHz. Hence, those frequencies closest to 1 GHz contained the most information.

Preprocessing the input data before using a neural network proved to be vital for optimum performance. Table I reports simulation results when using input data without flattening, outlier removal or data balancing. An 8:3:1 network was used, with twenty-eight input parameters. The rows of the table report results using different transformations. Simulations performed for the first row used no input transformation and simulations for the second row used a generalized KL transformation, retaining all 28 of the 28 components. This produces a network with equal number of weights to the untransformed case. The third line shows simulations performed with a generalized KL transformation which retained 16 of the 28 components. This network is of equal size to the one used in the fourth line to simulate the ES transformation.

For each of these cases, twenty simulations were run, each having a different set of initial weights. The first four columns report the average of the best performance achieved in each of the twenty simulations, and the following four columns report the peak performance achieved over the entire twenty simulations. The average testing "true positive" (TP) and "false positive" (FP) columns are used as a basis for comparing results. We see in this table that detection rates of about 65%–67% may be achieved, with corresponding false-positive rates of about 1%.

These results are poor. However, by removing the underlying "ramp" and "jumps" in the raw data, mine detection increased an additional 5%–9%, as reported in Table II. A further 5%–9% improvement is achieved by deterministic four-nearest-neighbor and eight-nearest-neighbor outlier removal, as reported in Tables III and IV. By balancing the data, improvement by about 15% is achieved (Tables V and VI), although this improvement comes at the cost of a higher false-positive rate. Finally, Table VII presents results where both the magnitude and phase of the twenty-eight input components are used, resulting in a total of fifty-six input components. Performance is roughly the same, even though almost twice as many weights are required! For this reason, we have elected to not use phase information as network input.

A short comment is in order relating the deterministic outlier-removal methods with vector-quantizer (VQ) based statistical outlier-removal methods. A variety of VQ algorithms were used to generate the *centroids* of the training data. Each of these centroids was assigned a class based on a majority vote of the training vectors falling within that classifier bin. All patterns in the training set whose mine-type did not agree with the class of the bin in which it resided were considered to be statistical outliers, and removed from the training data. The averaged results for simulation runs with each of these VQ methods are presented in Table VIII. Emphasis is not placed on the particular algorithms here; the curious reader is referred to reference [16] for the LBG method; reference [17] for the LVQ1 and OLVQ1 methods; reference [18] for the FSCL method and reference [19] for the C&SL method. The primary conclusion is that simulation results tended to be roughly equivalent, but perhaps slightly poorer, when the VQ outlier-removal methods were used, as compared to when the deterministic nearest-neighbor methods were used. In addition, the added complexity of using these methods was not warranted. However, one intriguing insight provided by using VQ lends significant credence to the nearest-neighbor outlier-removal methods. Those points considered to be outliers by the VQ methods were plotted according to their position in the mine lane. These points were very highly correlated with those points removed by the nearest-neighbor methods, as illustrated in Fig. 9. The left hand column depicts with '.' the location of those vectors which were retained by the LBG VQ method, and with 'X' those vectors which were discarded. The right hand column shows the location of the mines, and those points which would be eliminated via the nearest neighbor methods. We see from this figure that many of the points discarded by the VQ method occur near the boundaries of the mines, where accurate determination of the desired response is difficult. This observation supports the use of the deterministic outlier-removal methods.

Additionally, both of the transform methods used in this study, the generalized KL transform and the ES transform, showed good results. Both were able to reduce the number of inputs to the network, and hence the network complexity, while retaining equivalent performance. While the results are nearly equivalent for the KL and ES transforms, the ES transform has the advantage that it pre-determines the number of coefficients to retain. Significant experimentation is otherwise required to find this value.

We have seen that an increase in mine detection can come at

TABLE I
MINE DETECTION PERFORMANCE COMPARING TRANSFORM PREPROCESSING METHODS WITH UNFLATTENED DATA, NO OUTLIER REMOVAL AND NO DATA BALANCING.

Transform	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Method	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
None [28/28]	82.0	0.5	67.2	1.3	83.3	0.7	75.0	1.7
KL [28/28]	81.1	0.4	67.3	1.2	87.9	0.6	83.3	2.7
KL [16/28]	80.5	0.6	65.8	1.0	80.3	0.6	75.0	0.8
ES [16/28]	81.1	0.4	65.4	1.3	86.4	0.6	72.9	1.8

TABLE II
MINE DETECTION PERFORMANCE COMPARING TRANSFORM PREPROCESSING METHODS WITH FLATTENED DATA, NO OUTLIER REMOVAL AND NO DATA BALANCING.

Transform	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Method	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
None [28/28]	82.7	0.6	72.8	1.4	84.8	0.6	77.1	2.4
KL [28/28]	81.8	0.4	73.6	1.7	84.8	0.1	81.2	2.8
KL [16/28]	81.1	0.5	72.6	1.5	80.3	0.4	79.2	1.7
ES [16/28]	82.3	0.6	74.6	1.4	81.8	0.6	77.1	1.1

TABLE III
MINE DETECTION PERFORMANCE COMPARING TRANSFORM PREPROCESSING METHODS WITH FLATTENED DATA, FOUR-NEAREST-NEIGHBOR OUTLIER REMOVAL AND NO DATA BALANCING.

Transform	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Method	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
None [28/28]	86.7	0.5	81.0	2.4	87.9	0.5	85.4	1.9
KL [28/28]	88.7	0.6	79.9	2.4	93.9	0.9	85.4	2.1
KL [16/28]	85.2	0.7	81.2	2.6	86.4	0.6	85.4	2.7
ES [16/28]	86.0	0.8	79.9	2.5	87.9	0.6	85.4	2.6

TABLE IV
MINE DETECTION PERFORMANCE COMPARING TRANSFORM PREPROCESSING METHODS WITH FLATTENED DATA, EIGHT-NEAREST-NEIGHBOR OUTLIER REMOVAL AND NO DATA BALANCING.

Transform	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Method	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
None [28/28]	88.0	0.5	81.1	2.6	93.9	1.0	89.6	3.6
KL [28/28]	88.0	0.5	81.0	2.5	90.9	0.5	85.4	2.4
KL [16/28]	86.2	0.8	82.7	3.1	92.4	1.1	91.7	3.8
ES [16/28]	88.7	0.9	82.5	2.8	89.4	0.6	89.6	3.2

TABLE V
MINE DETECTION PERFORMANCE COMPARING TRANSFORM PREPROCESSING METHODS WITH FLATTENED DATA, FOUR-NEAREST-NEIGHBOR OUTLIER REMOVAL AND DATA BALANCING.

Transform	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Method	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
None [28/28]	99.4	6.3	97.9	10.1	100.0	4.6	100.0	8.8
KL [28/28]	99.5	6.5	97.6	10.4	100.0	6.9	100.0	9.9
KL [16/28]	98.5	7.4	97.8	11.6	98.5	6.4	100.0	8.5
ES [16/28]	99.5	7.5	97.9	9.7	100.0	8.0	100.0	8.5

TABLE VI
MINE DETECTION PERFORMANCE COMPARING TRANSFORM PREPROCESSING METHODS WITH FLATTENED DATA, EIGHT-NEAREST-NEIGHBOR OUTLIER REMOVAL AND DATA BALANCING.

Transform	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Method	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
None [28/28]	99.7	6.4	99.4	10.5	100.0	5.5	100.0	7.5
KL [28/28]	99.8	6.4	99.6	10.7	98.5	4.2	100.0	8.0
KL [16/28]	98.9	8.2	98.5	12.1	98.5	8.8	100.0	11.8
ES [16/28]	99.7	7.9	98.8	10.5	100.0	7.9	100.0	8.8

TABLE VII
MINE DETECTION PERFORMANCE COMPARING TRANSFORM PREPROCESSING METHODS WITH FLATTENED DATA, FOUR-NEAREST-NEIGHBOR OUTLIER REMOVAL AND DATA BALANCING. BOTH MAGNITUDE AND PHASE DATA USED.

Transform	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Method	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
None [56/56]	99.5	3.6	98.2	10.8	100.0	3.3	100.0	7.9
KL [56/56]	99.8	4.2	98.5	11.4	100.0	1.6	100.0	8.8
KL [33/56]	99.4	5.8	99.1	11.8	100.0	4.6	100.0	7.9
ES [33/56]	99.7	5.7	99.2	11.8	100.0	2.8	100.0	8.5

```
69: ...........................   ......=---=................
70: ........X..................   .....=-111-=...............
71: ....XX....X................   .....-11111-...............
72: .....X.....................   .....-11111-...............
73: .....X.....................   .....-11111-...............
74: ......X..X............XX...   .....-11111-...............
75: ......XXX...............X..   .....=-----=........=----=.
76: .....................XX....   ....................-2222-.
77: ..............XX.....X.....   ............=---=...-2222-.
78: ...............X.....X...X.   ...........=-555-...-2222-.
79: ................X....X...X.   ...........-5555-...=-222-.
80: ................X.....XXX.X   ...........-555-=....=---=.
81: ............X..X........XX.   ..=--=.....=---=...........
82: .....X........X............   .=-22-=....................
83: ......X....................   .-2222-....................
84: .X.........................   .-2222-....................
85: ...........................   .-2222-....................
86: .....X.....................   .=-22-=........=-----=.....
87: ...................XXX.....   ..=--=.........-44444-.....
88: ...............X..X..X.....   ...............-44444-.....
```

Fig. 9. Sample VQ outlier location

TABLE VIII
MINE DETECTION PERFORMANCE COMPARING VECTOR QUANTIZER STATISTICAL OUTLIER-REMOVAL METHODS. FLATTENED DATA, AND DATA BALANCING USED.

VQ Algorithm	Averaged Results: Training % TP	Training % FP	Testing % TP	Testing % FP
LBG	99.8	4.7	97.7	8.8
LVQ1	99.7	4.2	96.9	8.7
FSCL	99.2	4.8	96.5	9.1
OLVQ1	99.5	3.9	95.8	8.1
C&SL	99.5	3.2	95.0	6.6

the expense of increased false-alarm rate. These two rates are not independent of each other; in fact, there is a strong correlation between them and a tradeoff must be made when maximizing the performance of the system. For the results reported thus far, emphasis was placed on maximizing the true-positive rate. In the best systems, the peak performance was perfect (100% detection was achieved in the test region). Although this best possible detection rate was achieved, there are cases when the false-positive rate may be undesirably high. Therefore, the question arises whether some of the detection performance can be sacrificed for a better false-positive rate. Some results are presented in Table IX to illustrate this concept.

In the nontransformed case using the maximum detection criterion, the peak performance achieved was 100.0% TP with 7.5% FP. An averaged measure over 20 different initial conditions resulted in 99.4% TP and 10.5% FP. However, if we use a minimum detection rate of 90% or 95% as the performance criteria[2], the results change slightly. Most notably, the false-positive rate decreases. It has not been determined what kind of performance level is desirable or acceptable for practical use. This may be best left to the user of the system.

B. Statistical Mine Detection

It is not our intention to deluge the reader with volumes of results, but a few more comparisons are instructive. Simulations of statistical classifiers have been done, including probabilistic neural networks (PNN), radial-basis-function networks (RBF) and single-nearest-neighbor classifiers (1NN). None of these systems were found to improve on the best neural-network results, and all have significantly more parameters, thus being considered more complex. The results presented here are for the single-nearest-neighbor classifier, which is the most basic.

A single-nearest-neighbor classifier compares the input pattern under test to all training patterns, and computes its class to be equal to the class of the training pattern with smallest Euclidean distance to it. Simulation results with this classifier are presented in Table X, along with references to the corresponding table of neural network results. Results for “KL [28/28]” or “KL [56/56]” are omitted since Euclidean distances are preserved by this transform and hence the results are identical to the untransformed case. Additionally, there was no equivalent concept to data balancing, so there is no comparison for those runs.

[2] For these new performance criteria, all neural network results with at least a 90% or 95% TP rate were examined and the one with the lowest FP rate was chosen as the best result.

TABLE X
MINE DETECTION PERFORMANCE FOR A SINGLE-NEAREST-NEIGHBOR STATISTICAL CLASSIFIER.

Transform Method	Training % TP	Training % FP	Testing % TP	Testing % FP
Simulations Corresponding to Table I				
None [28/28]	100.0	0.0	64.6	2.4
KL [16/28]	100.0	0.0	66.7	2.7
ES [16/28]	100.0	0.0	75.0	2.5
Simulations Corresponding to Table II				
None [28/28]	100.0	0.0	79.2	3.8
KL [16/28]	100.0	0.0	87.5	4.4
ES [16/28]	100.0	0.0	85.4	3.2
Simulations Corresponding to Table III				
None [28/28]	100.0	0.0	85.4	4.4
KL [16/28]	100.0	0.0	85.4	5.1
ES [16/28]	100.0	0.0	85.4	4.0
Simulations Corresponding to Table IV				
None [28/28]	100.0	0.0	85.4	4.4
KL [16/28]	100.0	0.0	91.7	5.3
ES [16/28]	100.0	0.0	89.5	4.1
Simulations Corresponding to Table VII				
None [56/56]	100.0	0.0	81.3	5.0
KL [33/56]	100.0	0.0	83.3	5.4
ES [33/56]	100.0	0.0	85.4	4.4

It is interesting to notice that for the unflattened data the ES transform greatly enhances the results. Recall that the goal of the ES transform was to maximize the distance between two classes. This is clearly helping the nearest-neighbor classifier. On the other hand, we did not notice such an increase in the neural-network results when using ES. This leads us to conclude that the neural network is performing a much more complicated functional mapping than nearest neighbor and is better able to adjust to whichever input format is used. We also observe that for all the flattened data cases, the ES transform is not able to produce such significant performance gains. This leads us to believe that the flattening procedure results in data which has approximately equally distributed norm, and hence is difficult to separate with a linear transform.

Furthermore, we note that the 8:3:1 neural network used for previous results required 167 or 283 adjustable weights (depending on whether sixteen or twenty-eight inputs are used, and including bias weights). A single-nearest-neighbor classifier with approximately 8300 training patterns and sixteen (or twenty-eight) values per pattern requires about 130,000 (or 230,000) weights! Clearly the neural network is a much simpler and more compact representation.

TABLE IX
MINE DETECTION AND FALSE-POSITIVE TRADEOFFS. (NONTRANSFORMED)

Performance	Averaged Results (20sims)				Peak Performance			
	Training		Testing		Training		Testing	
Criteria	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
Four-Nearest-Neighbor Outlier Removal								
Maximum Detection	99.4	6.3	97.9	10.1	100.0	4.6	100.0	8.8
95% Detection	99.5	5.6	96.1	8.4	98.5	7.2	97.9	7.1
90% Detection	99.1	5.1	92.7	7.3	98.5	7.2	97.9	7.1
Eight-Nearest-Neighbor Outlier Removal								
Maximum Detection	99.7	6.4	99.4	10.5	100.0	5.5	100.0	7.5
95% Detection	99.6	5.6	96.2	8.8	100.0	4.2	97.9	8.3
90% Detection	99.2	5.6	93.4	8.0	98.5	6.2	97.9	8.8

Finally, we comment on PNN and RBF. The performance of these two methods was found to be somewhere between that of single-nearest-neighbor and the best neural-network results. The PNN requires even more parameters than 1NN and the RBF fewer, but the latter is still more complex than the neural network. We conclude that neural networks are the better methodology to use.

X. CONCLUSIONS AND FUTURE DIRECTIONS

Future directions in this work include training networks to detect anti-personnel mines in multiple distinct or conglomerate soil types. It is also felt that the addition of an acoustic instrument to the measurement system, which would modulate the microwaves around the mines' resonant frequency, would provide an additional independent measurement at each spatial location, and would reduce the uncertainty of the network's output. Augmenting the sensor apparatus with temperature and humidity detectors might give more insight into the "jumps" and aid in their removal.

In conclusion, the work conducted has shown that a neural network, trained with broadband *S*-parameter data is able to detect buried anti-personnel land mines with very high accuracy. Furthermore, by varying the design threshold of the rate of desired true positives, the level of false alarms may be made arbitrarily low. Data preprocessing has turned out to be a very important step in the detection process. By "flattening" the data to remove irregularities produced by the measurement process, a very useful data set was obtained. From that point, either the KL or ES transformation was used on the data to reduce the dimensionality of the input data while rotating the underlying coordinate system such that a small network was more easily able to separate the categories of mine and background pattern. The removal of statistical outliers using the four-nearest-neighbor or eight-nearest-neighbor methods and balancing the data set further aided the network in its classification ability.

XI. ACKNOWLEDGMENTS

This work was financially supported by ARMY contract DAAK70-92-0003. We would also like to thank Dr. B. Widrow for his constant support and creative input.

REFERENCES

[1] Bloom, D., "Land mines keep wars from ever coming to an end", *The Christian Science Monitor*, vol. 87, no. 57, pp. 19, February 1995.

[2] Webster, D., "One leg, one life at a time", *The New York Times Magazine*, January 23, 1994.

[3] Azimi-Sadjadi, M., Poole, D., Sheedvash, S., Sherbondy, K., and Stricker, S., "Detection and classification of buried dielectric anomalies using a separated aperture sensor and a neural network discriminator", *IEEE Transactions on Instrumentation and Measurement*, vol. 41, no. 1, pp. 137–143, February 1992.

[4] Azimi-Sadjadi, M. and Stricker, S., "Detection and classification of buried dielectric anomalies using neural networks—Further results", *IEEE Transactions on Instrumentation and Measurement*, vol. 43, no. 1, pp. 34–39, February 1994.

[5] Lehr, M., "Adaptive multisource decision-making: Detecting land mines with neural networks using separated aperture sensor data collected at Fort Belvoir", Final Report, U.S. Army Contract DAAK70-89-K-0001, Stanford University, April 1992.

[6] Riggs, L. and Amazeen, C., "Research efforts with the waveguide beyond cutoff or separated aperture dielectric anomaly detection scheme", Internal Report #2497, U.S. Army Ft. Belvoir RD&E Center, Ft. Belvoir, VA, 1989.

[7] Clague, F., "Summary of experiments with the separated aperture technique of dielectric anomaly detection, final report", Report number SR-723-29-89, U.S. Army Ft. Belvoir RD&E Center, Ft. Belvoir, VA, 1989.

[8] Duda, R. O. and Hart, P. E., *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

[9] Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York, second edition, 1990.

[10] Schalkoff, R., *Pattern Recognition: Statistical, Structural and Neural Approaches*, Wiley, New York, 1992.

[11] Widrow, B. and Lehr, M., "30 years of adaptive neural networks: Perceptron, Madaline, and backpropagation", *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415–1442, September 1990.

[12] Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, Wiley series in telecommunications. Wiley, New York, 1991.

[13] Anand, R., Mehrotra, K., Mohan, C., and Ranka, S., "An improved algorithm for neural network classification of imbalanced training sets", *IEEE Transactions on Neural Networks*, vol. 4, no. 6, pp. 962–969, November 1993.

[14] Torrieri, D., "A linear transform that simplifies and improves neural network classifiers", in *Proceedings of ICNN—IEEE International Conference on Neural Networks*, Washington DC, June 1996, pp. 1738–1743.

[15] Doi, T. and Plett, G. L., "Adaptive neural networks for mine detection; Scientific and technical report", Final Report, U.S. Army Contract DAAK70-92-K-0003, Stanford University, 1995, (available as: http://www-isl.stanford.edu/people/glp/mines95.ps).

[16] Gersho, A. and Gray, R., *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[17] Kohonen, T. et al., "LVQ_PAK: The learning vector quantization program package v 2.1", Available via anonymous FTP to: cochlea.hut.fi (130.233.168.48), Oct 9, 1992.

[18] Ahalt, S. C., Krishnamurthy, A. K., Chen, P., and Melton, D. E., "Com-

petitive learning algorithms for vector quantization", *Neural Networks*, vol. 3, pp. 277–290, 1990.

[19] Ueda, N. and Nakano, R., "A competitive & selective learning method for designing optimal vector quantizers", in *Proceedings of ICNN—IEEE International Conference on Neural Networks*, March 1993, pp. 1444–1450.

Gregory L. Plett (S'97) was born in Ottawa, ON, in 1968. He received the B.Eng. degree in computer systems engineering with high distinction from Carleton University, Ottawa, in 1990, and the M.S. degree in electrical engineering from Stanford University, CA, in 1992. He is currently a Ph.D. candidate at Stanford University where he is researching aspects of adaptive control under the supervision of Professor Bernard Widrow.

Takeshi Doi was born in Chigasaki, Japan on January 7, 1967 and moved to Hinsdale, Illinois in 1969. He received the B.S. degree in bioengineering from the University of California at Berkeley in 1990 where he performed research under the supervision of Professor Andrew Neureuther on resist analysis and electromagnetic scattering in photolithography. He received the M.S. degree in electrical engineering from Stanford University, CA in 1992 and is currently pursuing his Ph.D. in electrical engineering under the supervision of Professor Bernard Widrow in the area of pattern recognition using artificial neural networks.

Don J. Torrieri (SM'90) received the B.S. degree from the Massachusetts Institute of Technology, Cambridge, the M.S. degree from the Polytechnic University, Farmingdale, NY, and the Ph.D. degree in electrophysics from the University of Maryland, College Park.

He analyzed electronic systems at the Naval Research Laboratory, Washington, DC, and since 1977, has been employed by the U.S. Army. His current research interests include the analysis and assessment of communication systems and electronic countermeasures.

Dr. Torrieri is the author of *Principles of Secure Communication Systems*, 2nd ed. (Dedham, MA: Artech House, 1992) and a contributer of *Acousto-Optic Signal Processing* (N.J. Berg and J.N. Lee, Eds., New York: Marcel Dekker, 1983).