Detection and Classification of Buried Dielectric Anomalies Using a Separated Aperture Sensor and a Neural Network Discriminator

Mahmood R. Azimi-Sadjadi, David E. Poole, Sassan Sheedvash, Kelly D. Sherbondy, and Scott A. Stricker

Abstract—The problem of detection and classification of buried dielectric anomalies using a separated aperture microwave sensor and an artificial neural network discriminator was considered. Several methods for training and data representation were developed to study the trainability and generalization capabilities of the networks. The effect of the architectural variation on the network performance was also studied. The principal component method was used to reduce the volume of the data and also the dimension of the weight space. Simulation results on two types of targets were obtained which indicated superior detection and classification performance when compared with the conventional methods.

I. INTRODUCTION

THE development of a target detection system capable of detecting and classifying various types of buried dielectric anomalies, such as landmines, under different environmental conditions presents many problems. Historically, a number of target detection and classification schemes have been applied to this problem. However, in practice, they have been only partially successful and have been shown to produce high false-alarm rates. Some of the factors which inhibit the detection and classification are the diverse sizes and compositions of targets, variation of soil properties with location and moisture conditions, nonrepeatability of target signatures, competing clutter with similar responses as actual targets, and obscuring of targets in noise.

Over the years, several sensor systems such as separated aperture (or waveguide beyond cutoff), acoustics, seismic, and photon backscatter have been developed [1], [2] as means to solve this problem. However, each sensor has its own specific problems along with its advantages. The separated aperture sensor which operates in the microwave range can potentially offer a very good signal-to-clutter ratio when the sensor height from the ground surface is small. Acoustic sensors are ideally suited for real field implementations but have the problem of coupling

energy from the air into the ground because of reflections at the air–soil interface. In spite of its accuracy in detection, the photon backscatter sensor has physical limitations and usually requires high power. A detailed description of these sensors and a discussion on their principles of operation, benefits, and shortcomings were reported in [1].

The effort of this particular research is primarily focused on the development of a neural network-based [3], [4] detector and classifier for the separated aperture sensor which offers better accuracy and a lower false-alarm rate than previous systems. Specific emphasis is placed on the development of various data representation schemes, network training procedures, and network structures for automatic detection and classification of mine analogs. The data collected from the sensor which include the magnitude and phase components are presented to the neural networks in various formats to determine relative importance of these components as well as any effect of data reduction on the network performance. Data reduction is performed using the principal component (PC) scheme or Karhunen–Loeve (KL) transform [5], [6]. A number of network architectures are designed to take advantage of multiple or single frequencies and one or both of the magnitude and phase data components for training. The performance of the various neural network schemes is compared with the traditional methods of detection and classification, such as basic thresholding and correlation matching techniques.

II. THE SEPARATED APERTURE SENSOR

The sensor system used for data collection in this effort is formally called the "Separated Aperture Sensor" [2] which operates at a microwave range of frequencies. As shown in the simpified schematic diagram of this sensor in Fig. 1, the sensor is composed of a transmit and receive dipole pair separated by a metallic septum. Each dipole resides within a corner reflector. During operation, given a fixed input power, the output power measured at the receiving dipole is monitored. As the sensor head moves over the surface of the field, the received power varies. When the sensor head is over the uniform background (no target present), very little power is received. In contrast, there is a significant increase in the received power when
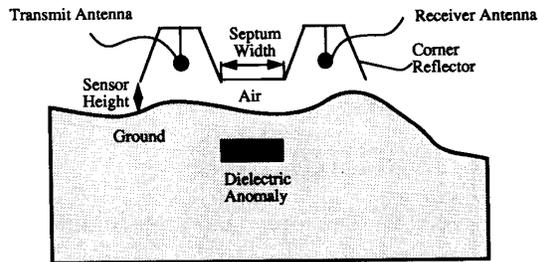
Fig. 1. Simplified schematic diagram of the Separated Aperature Sensor System.

Fig. 2. Threshold contour plots of magnitude data for nylon and wood at 792 MHz.

the sensor head is over a target. The metallic septum forms a waveguide with Earth's surface, and when the septum–Earth separation is small, this waveguide operates below cutoff. This results in an exponential attenuation of the direct and ground-reflected signals and consequently a substantial improvement in signal-to-clutter ratio. Experimental data collected under controlled conditions and comparison of signal-to-clutter ratios show the superiority of this sensor over all the other sensors ever attempted from both the detection and false-alarm rate rejection capabilities points of view [1]. What follows is a description of the fundamental operating principles of this sensor.

A. Sensor Operating Principles/Parameters

The fundamental operating principle of the separated aperture sensor is based upon the propagation of electromagnetic waves through the soil and their iteractions with a buried target. Owing to the potentially high energies associated with the direct coupling from the transmitter dipole to the receiver dipole and the reflection from the air–soil interface, the signal power could be very low. However, the key feature of the separated aperture sensor is its ability, under proper operating conditions, to substantially suppress the components corresponding to the direct coupling and ground-reflected energies.

A critical parameter is the height of the sensor head above the ground. If the height of the antenna is greater than one-half of the wavelength, the weak signal of the target tends to reside within or below the clutter. In short, the waveguide made between the sensor head and earth is lost, and the direct-coupling and ground-reflected energies begin to dominate the received energy, hence making it virtually impossible to detect the target response. Another parameter is the operating frequency, which should be chosen to minimize sensor height sensitivity and maximize depth penetration. Additional operating parameters, which play an important role, are the soil parameters of the earth in which the target is buried. The detection system is a dielectric anomaly detector, and one of the main factors for optimal detection is the difference in dielectric parameters between the soil and target. Specifically, if the dielectric parameters of the target are not significantly different from those of its surrounding soil, detection may not be possible, and due to the weakness of the received signal associated with the target, the system will be
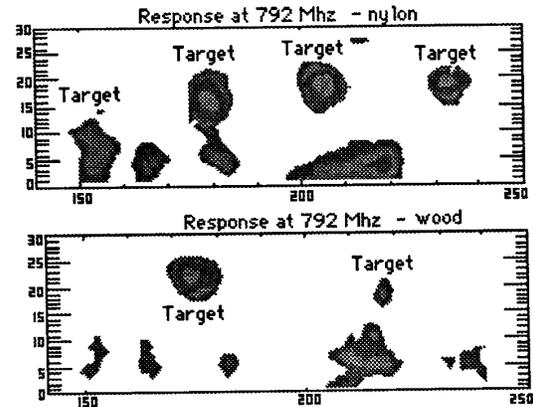
plagued with high false alarms generated by competing clutter. All of these factors must be considered when using this sensor.

B. Experimental Data Collection

The data were collected by a symmetrical-sensor head (resonant at 790 MHz) with a 6-in metallic septum over a band of frequencies from 600 MHz to 1 GHz. This band of frequencies is then sampled at multiple frequencies. All of the data was collected in dry loamy soil with a soil moisture content of 6% by weight [2]. The complex permittivity values vary substantially depending on the type of buried anomaly, type and moisture content of the soil and the operating frequency of the sensor [7], [8].

The targets that were used were of identical shape and size. They were in the shape of rectangular blocks, with the dimensions of 12" × 12" × 3", and were made of nylon and wood for the respective runs. The targets were buried at various positions and angles as opposed to the direction traversed by the sensor head throughout the soil. Fig. 2 shows a sample of amplitude returns (thresholded) for the nylon and wood sets. These contour plots show not only the responses of nylon and wood targets but also those of clutter. The total numbers of targets for each scan of nylon and wood lanes were 15 and 12, respectively. Data was collected every 1.5" in both directions, providing a matrix of data points corresponding to an area of approximately 39" by 750". At each position the scattering parameter, (S_{21}), was measured over a band of 51 frequencies from 600 MHz to 1 GHz.

III. DATA REPRESENTATION AND PREPROCESSING SCHEMES

One of the most critical aspects involved with target detection and classification is the choice of data representation which impacts the size, speed, and accuracy of the system. The data in these experiments may be described as being 2-D arrays of the magnitude and phase compo-

nents of the S_{21} signal at the receiver. A discussion of a number of data representation schemes follows.

A. Unprocessed Data

Each data set (nylon and wood), at each frequency, consists of two files: one representing the magnitude component of S_{21} at the receiver and the other representing the phase components of S_{21}. For use as input to the neutral networks, the phase data were converted to change in phase or delta-phase using the 2-D Robert gradient operator [5].

B. Preprocessed Data

In previous research efforts [9] two preprocessing schemes were applied to the data obtained using the separated aperture sensor. The first algorithm, which is referred to as "target amplitude descriptor" or TAD, computes the signal-to-background ratio at each spatial position. In this calculation the magnitude of the signal at a particular spatial position is obtained by averaging the magnitudes of all the signals at the same locations over all the frequencies. The second algorithm used with the separated aperture sensor is called the "average of sums" or ASUM. This algorithm calculates the weighted average of signal-background difference, weighted by the standard deviation of the background clutter [9].

C. Data Reduction Using the Principal Component Scheme

The KL transform is a method of principal components that can be used to reduce the size of a data set while closely preserving the quality of the original signal. It does this by packing the energy of a signal into a reduced dimension set of values. It has been shown that the KL transform is optimal in achieving accurate data compression of stochastic signals [5], [6]. In recent years, several neural network architectures [10], [11] have been developed which use a Hebbian learning rule to extract the principal components of a set of signals (or images) in real-time. This is particularly attractive to our application, since a composite architecture made of a principal component network as a preprocessor and a detector/classifier network can perform the entire operation in real-time.

The target or anomaly returns within a window can be considered as a 2-D complex random image. Let $\{x(i, j)\}$, $i, j \in [1, N]$, be a 2-D complex random image of an ensemble set of M images each of size $(N \times N)$ with zero mean and covariance function $r_x(m, n) := E[x(i, j) x^*(i - m, j - n)]$, where $E[\cdot]$ represents the expectation operator. Let the row-ordered arrangement of $x\{i, j\}$ be given by vector $\vec{X}$ as

$$\vec{X} := [x(1, 1)\, x(1, 2)\, \cdots\, x(1, N)\, x(2, 1)\, \cdots$$

$$x(2, N)\, \cdots\, x(N, 1)\, \cdots\, x(N, N)]^T. \tag{1}$$

The 2-D KL transform of the image $\vec{X}$ is defined as

$$\vec{Y} := \Psi^{*T}\vec{X} \tag{2a}$$

and the inverse transform is

$$\vec{X} = \Psi\vec{Y}. \tag{2b}$$

The transformation matrix Ψ contains the complete set of eigenvectors of the doubly Toeplitz [5] covariance matrix $R := E[\vec{X}\vec{X}^{*T}]$ in its columns and satisfies

$$\Psi^{*T}R\Psi = \Lambda \tag{3}$$

where Λ is a diagonal matrix containing the eigenvalues, λ_i, of matrix R. The eigenvalues represent the energy (variance) of the signal along the coordinates of the signal space. The interesting property of the KL transform is that the elements of the $\vec{Y}$ vector are decorrelated, which makes them ideal for neural network training. For this research, the covariance matrix R is formed from several images by computing the covariances in each image and finding the average over all the available images using

$$r_x(m, n) \approx \frac{1}{M(N - m)(N - n)} \sum_{k=1}^{M} \sum_{i=1}^{N-m} \sum_{j=1}^{N-n}$$

$$\cdot\, x_k(i + m, j + n)\, x_k^*(i, j),$$

$$\forall m, n \in [0, N - 1] \tag{4}$$

where $\{x_k(i, j)\}$ represents the kth image in the ensemble set.

IV. DETECTION AND CLASSIFICATION

The problem of detecting target signals in noise and/or clutter could be divided into two parts: (1) the detection of the signal in noise and/or clutter, and (2) the extraction of the information from the received signals, which is a problem of statistical parameter estimation. In most of the present target detection and classification systems the detection and the extraction of information are performed separately. For detection, methods such as matched filtering and correlation techniques are typically used [12]. Extraction of information for the target classification is normally accomplished by first estimating certain parameters attributed to each received signal and then using a standard classifier such as Bayes or maximum likelihood (ML) to separate the classes or patterns [13]. Using conventional parameter estimation techniques, accurate measurement of the parameters can only be made when the signal-to-noise/clutter ratio (SN/CR) is relatively large. In addition, the extracted features may not be distinctive enough to guarantee accurate and nonconflicting classifications. The objective of this research work is to develop more reliable and accurate schemes for target detection and classification using artificial neural networks.

Methods for detection of targets used with separated aperture in the past consisted primarily of basic thresholding techniques on the magnitude data. The following section presents the results of using the thresholding and correlation matching techniques.

A. Thresholding Results

Thresholding was performed on three sets of data: raw magnitude, ASUM, and TAD, with each set having two subsets, one each for nylon and wood targets. The performance of thresholding on each set is obviously directly related to the threshold value used and the response of the sensor over a target. As the threshold is lowered, the number of detections increases, and the percentage of correct detections approaches 100%, while at the same time the percentage of incorrect detections and/or number of false positives increases to a totally unacceptable level. The intention of the ASUM and TAD algorithms is to amplify the target response while minimizing the response of the background and nontarget anomalies. Even though a measure of success has been achieved, obviously the requirements of a high detection rate and a low occurrence of false positives are conflicting and difficult to achieve using this method. Table I provides a sample of results from use of the thresholding technique. Because SN/CR is so much lower in the wood data, the number of false positives becomes very high. In a mine detection scenario this would indicate a system incapable of performing accurate detection in diverse environments.

B. Correlation Matching Results

In basic terms, the correlation matching method provides cross-correlations between two data sets in order to determine the amount of similarity between them [12]. Two versions of the correlation matching scheme have been tested. The results of the first version were based upon the thresholded cross-correlation responses of targets in the respective data sets. The second version uses a background template and two target templates in an attempt to reduce false-alarm rates. In addition to the target correlations, this version computes the background correlations each time the target correlation threshold is exceeded. To be classified as a target, the computed correlations must be greater than the correlation threshold for a target and less than the correlation threshold for background. Table I presents the detection and false-alarm rates of both versions of the detector using data collected at 792 MH$_2$. As with the thresholding method, it was observed from the results in Table I that although lowering the thresholds improves the detection rate, it significantly increases the false-alarm rates. The detector showed much better results for nylon targets than it did for wood targets. The results in Table I might be somewhat misleading. The percentages in Table I were realized knowing beforehand the type of target being searched for. This allowed the use of the correct template. Theoretically, when the nylon target template is used on the wood data and vice-versa, the results are expected to show zero detection for both cases (i.e., no nylon targets in the wood data, etc.). However, in practice this was not the case, and more than 35% of the targets were detected incorrectly. This indicates that the correlation matching method can neither provide a robust detection, nor can it be used for classification pur-

TABLE I
RESULTS OF THRESHOLDING AND CORRELATION MATCHING METHODS

Conventional Methods		Performance	
		Nylon	Wood
Thresholding	792 MHz	100/29	92/69
	ASUM	93/22	100/76
	TAD	93/26	92/86
Correlation Matching	Version One @ 792 MHz	100/48	83/70
	Version Two @ 792 MHz	100/44	92/78
Note: All entries are of the form: Detection Rate (%) / False-Alarm Rate (%)			

poses. This severely limits the effectiveness of this method for target detection/classification applications.

C. Neural Network Approach

Neural networks offer potentially powerful, robust, and adaptive means of detecting and classifying targets in high cluttered background. In a supervised neural network such as multilayer back-propagation, the choice of training algorithm, network architecture, input signal representation, and training set, plays a dominant role in the generalization and training characterististics of these networks. The choice of input signal representation determines the size of the network, the dimensionality of the weight space, and the transient behavior of the learning. The network architecture is another important consideration for optimal training and generalization characteristics. It has been proven [14] that a three-layer perceptron neural network, with sigmoidal nonlinearity at nodes can approximate any arbitrary nonlinear function and generate any arbitrary complex decision region needed for detection and classification tasks. The choice of the training algorithm, on the other hand, determines the rate of convergence to a solution, time required to reach a solution and the optimality of the solution. If enough training samples and internal parameters are used, the input-output transformation may be defined to an arbitrary accuracy. In this case, the performance of the network can approach that of Bayes estimator which is optimal.

In this section emphasis is mainly focused on studying the neural network performance in relation to various data representation schemes and various network architectures. In most cases a feed-forward three-layer back-propagation network was employed to perform the detection/classification tasks. Three groups of experiments were conducted. What follows are descriptions for each group of experiments and analysis of the results.

Group 1–Magnitude Data Representation and Network Bottleneck

The purpose of this experiment is to study the effect of different magnitude data representations and different bottlenecks on network performance in terms of both detection and false-alarm rate characteristics. Two network architectures, 225-100-50-3 and 225-50-25-3, were considered. The desired output sequences for nylon targets, wood targets, and background are $(1, -1, -1)$, $(-1, 1, -1)$, and $(-1, -1, 1)$, respectively. The input

consisted of the magnitude of sensor data in windows of size 15 × 15 from the five frequencies. The mean for each window was first subtracted, and then the data was normalized. The networks were trained for six targets and six backgrounds from each data file, once for single frequency, once for multiple (five) frequencies and once for selected target windows. In the latter case, the strong returns were selected based upon their energy distribution. The frequencies that were used include 728, 760, 792, 824, and 856 MHz. The learning rate was chosen to be $\eta = 0.05$. The training data for multiple frequencies were presented to the network sequentially. To guarantee proper weight adjustment and prevent early saturation, the weights were initialized randomly in an uniform region $[-0.1, 0.1]$. The index of performance used to determine the trainability of the network is the Average Mean Squared Error (AMSE) at the output averaged over all the output units.

Once the network is trained, the generalization of the network is tested by using the training data and the testing data (the data from other parts of the lanes which the network had not seen before). The testing data at each frequency consisted of 15 targets and 12 background windows for nylon files and 12 targets and 12 background windows for wood files. As expected, the network with increased bottleneck (i.e., reduced number of nodes in the hidden layers) exhibited better generalization [15]. It must be pointed out that for a given network with fixed data representation, there exists a particular network architecture for which the generalization performance is optimal. A subsequent decrease in the number of nodes drastically degraded the performance of the network. The experiments conducted on a two-layer network indicated even more pronounced degradation in the performance, as a two-layer network is not capable of generating a complex decision region needed for this detection/classification problem. The training set chosen proved to be a key factor. The selected data training set, at a single frequency, performed substantially better than the others. The detection and false-alarm rates for this case and the multiple frequency case are given in the first and the second rows of Table II. Comparing the network performance for nylon and wood types of targets indicates that, in general, the magnitude data for nylon targets provide a better signal-to-noise ratio than wood in the same environment.

Group 2–Magnitude and Phase Data Representation and Network Architecture

The purpose of this experiment is to study the effect of magnitude, phase, and magnitude-phase data and also the network architecture on the performance in terms of detection and false-alarm rates. The training data consisted of both magnitude and phase components for each target and background window. Similar to the previous case, experiments using five frequencies were also included. In the first part of this experiment the unprocessed magnitude and phase data for six targets and six backgrounds at

TABLE II
RESULTS OF NEURAL NETWORKS METHODS

Neural Networks Methods			Performance	
Group/Case	Data Representation	Architecture	Nylon	Wood
1/1	Single Freq-Selected	225-50-25-3	87/3	84/7
1/2	Five Freq-Sequential	225-50-25-3	85/4	76/8
2/1	Magnitude & |Δ Phase| Sequential	2× 225-100-15-3	86/8	80/15
2/2	Single Freq-Magnitude & |Δ Phase|	5× 45-25-25-2	80/17	83/13
3/1	KL data, combined Ψ	26-13-5-3	100/0	83/8
3/2	KL data, target Ψ	26-13-5-3	100/0	83/8
Note: All entries are of the form: Classification Rate (%) / False-Alarm Rate (%)				

five frequencies were sequentially presented to the network. The network consisted of two 225-100-15-3 networks in parallel, one for magnitude and one for phase data. The testing results indicated that the combination of magnitude and phase data has a degrading effect when compared with the results of the magnitude alone case. The same experiment was repeated this time using the 2-D gradient [5] (2-D Roberts Gradient) of the phase data. The test results indicated some improvements in the detection and classification rates. This is largely attributed to the fact that the phase is relatively constant over the surface of the target and varying over the surface of non-target anomalies. Thus, the 2-D phase gradient data contain information about the location of the edges of the object. The results are given in the third row of Table II.

As can be seen, the inclusion of phase resulted in some degradation in the overall performance. It was first believed that this degradation may mainly be attributed to enlargement of the size of the network and substantial growth in the number of nodes which naturally leads to poor generalization capability and hence deterioration in the performance. Consequently, a network architecture was specifically designed to reduce the dimensionality of the weight space with the expectation to improve the generalization of the network. This architecture consisted of five 45-25-25-2 networks in parallel as shown in Fig. 3. The input data which was applied sequentially consisted of all five frequencies. The results are given in the fourth row of Table II. As can be seen, the reduction in the weight space dimensionality resulted in overall degradation in the performance. Consequently, the negative effects of adding phase data cannot be blamed entirely on the network size. After careful study of the phase data, it was observed that rather large numbers of discontinuity points exist in the phase data that are caused by improper conversion procedures and hence make the gradient phase data unreliable.

Group 3–Data Reduction

The purpose of this experiment is to study the effect of a data reduction scheme on network performance. The method of principal components, namely the 2-D KL transform, was used to reduce and decorrelate the training and testing data. This was achieved by evaluating the two-sided covariances of the magnitude data in each 15 × 15 window and averaging over all the available windows.
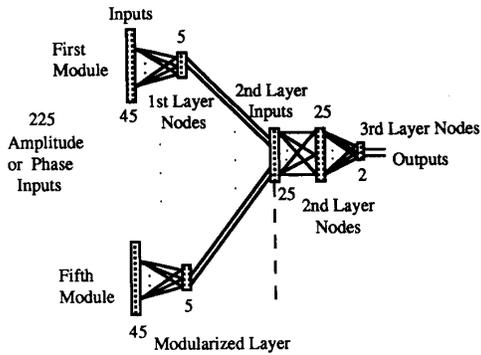
Fig. 3. The modular magnitude/phase only network architecture.

The covariances were used to form the double block Toeplitz covariance matrix R of the process. This matrix was then diagonalized using a unitary transformation to yield the eigenvalues associated with each type of image. Since these eigenvalues represent the energy of the signal, the process of finding the significant number of eigenvalues was based on the number needed to represent 90% of the total energy of the signal. That is, if K represents the total number of eigenvalues, and L denotes the number of significant ones, then we should have

$$\sum_{i=1}^{L} \lambda_i \geq 0.9 \sum_{k=1}^{K} \lambda_i. \tag{5}$$

Results of calculations show that $L = 26$ largest eigenvalues out of the set of $K = 225$ are sufficient to represent the images used in this study. Once the number of significant eigenvalues has been determined, the reduced transformation matrix Ψ, of size $(K \times L)$ can then be formed with the first L columns of Ψ associated with the first L largest eigenvalues. The reduced transformed image $\vec{Y}_r$ is given by

$$\vec{Y}_r = \Psi_r^{*T}\vec{X} \tag{6}$$

where $\vec{Y}_r$ is the reduced transformed image of size $(L \times 1)$. This decorrelated image can now be used to train a neural network, the architecture of which will only require L inputs. Note the immense reduction in size compared to the network that used the entire (15×15) image and had 225 inputs.

Two groups of experiments were conducted. In the first experiment, the reduced transformed images $\vec{Y}_r$ for each type of target and background were produced using their associated KL transformation matrices. A three-layer neural network with architecture 26-13-5-3 was trained with a data set consisting of 4 target and 4 background images for both nylon and wood cases at five frequencies. The testing results were outstanding and showed that detection and classification rates of 100% and false-alarm rates of 0% can be achieved by this network using KL transformed images as long as an image is transformed by its corresponding Ψ, matrix. It was found in further stud-

ies that if a target is transformed by a Ψ_r that was formed from backgrounds, the network could not correctly detect it. And similarly, if a background image is transformed by a Ψ_r that was formed from targets, the network will produce a false alarm.

With this in mind, a second set of experiments conducted to try to form a generalized KL transformation matrix that could be used to reduce any arbitrary image (target or background) without severely degrading the results of the neural networks. Two different generalized transformation matrices were formed. The first was formed from nylon target, wood target, and background images. The second consisted of only nylon and wood targets (no backgrounds). All images were then reduced by these transformation matrices, and a three-layer network was trained for each of the two sets of generalized reduced images. The results of these experiments for the resonant frequency (792 MHz) are shown in the fifth and sixth rows of Table II. These results indicate that it is possible to obtain detection and false-alarm rates that surpass those in the previous groups of experiments.

V. CONCLUSION

As can be expected with a large and complex problem the results are mixed, but some trends are evident. It is evident that the traditional techniques of thresholding and correlation matching are capable of providing high rates of detection, but they generally produce high false-alarm rates. The neural network approach showed comparable rates of detection and classification while reducing the false-alarm rate. The phase component of the signal can potentially convey useful information for target detection/classification. Interestingly, the locations in the data where the phase only and magnitude only networks made mistakes were at different locations; each was able to detect the targets the other missed. This could lead one to believe that coupling the two networks in some manner would produce superb results, but this was not the case. As can be seen from the results in Table II, there was an overall degradation in the performance when the phase data were included.

In general, several conclusions can be made in regard to the number of frequencies used, the network architecture, and the training and data representation. When examining the side frequencies, the tendency is to conclude that the targets show up more consistently in the side frequencies than anomalies do, and this enabled the network to discriminate better. The architecture and the method of training are also deterministic factors in the generalization capability of the network. Increasing the bottleneck, to a certain degree, has shown to improve the generalization. The choice of training scheme determines the rate of convergence and the optimality of the detection and classification. The choice of input data representation determines, in part, the size of the network, the dimensionality of the weight space and the transient behavior of the learning. The principal component method is found to be

very useful in extracting the salient features and decor-relating the data, which is an ideal characteristic for the training data. The results of the principal component method showed excellent detection and classification rates and extremely low false-alarm rates.

In summary, the use of neural networks for target detection/classification showed substantial improvements in performance over the previous methods, especially when used in conjunction with the principal component method.

REFERENCES

[1] R. V. Nolan, H. C. Egghart, L. Mittleman, R. L. Brooke, F. L. Roder, and D. L. Gravette, "MERADCOM MINE DETECTION: 1960-1980." Ft. Belvoir RD&E Center Rep. No. 2294, pp. 27-32, March 1980.

[2] L. S. Riggs and C. A. Amazeen, "Research efforts with the waveguid beyond cutoff or separated aperture dielectric anomaly detection scheme," Internal Rep. U.S. Army Ft. Belvoir RD&E Center, Fort Belvoir, VA 22060-5606, 1989.

[3] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Mag.*, vol. 4, no. 2, pp. 4-22, Apr. 1987.

[4] D. C. Plaut, S. J. Nowlan, and G. E. Hinton, "Experiments on learning by back propagation," CMU Internal Rep., No. CS-86-126, June 1986.

[5] A. K. Jain, "Fundamentals of digital image processing." Englewood Cliffs, NJ: Prentice Hall, 1989.

[6] V. R. Algazi and D. J. Sarkison," On the optimality of the Karhunen-Loeve expansion," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 319-321, March 1969.

[7] H. E. Bussey. "Dielectric measurements in a shielded open circuit coaxial line." *IEEE Trans. Instrum. Meas.*, vol. IM-29, pp. 120-124, June 1980.

[8] R. F. Harrington, *Time-Harmonic Electromagnetic Fields.* New York: McGraw-Hill Book Company, Inc., 1961, pp. 50-51.

[9] F. R. Clague. "Summary of experiments with the separated aperature technique of dielectric anomaly detection," Internal Rep., U.S. Army. Ft. Belvoir RD&E Center, Fort Belvoir, VA 2206-5606, 1989, pp. 1.3-1.4.

[10] E. Oja, "A simplified neuron model as a principal component analyzer," *J. Math. Biology*, vol. 15, pp. 267-273, 1982.

[11] S. Bannour and M. R. Azimi-Sadjadi, "Principal component extraction using recursive least squares learning method," in *Proc. of IEEE Joint Conference on Neural Networks*, pp. 2110-2115, Nov. 1991.

[12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory.* New York: Wiley, 1968.

[13] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis.* New York: Wiley, 1973.

[14] R. Hecht-Nielsen. "Theory of back-propagation neural networks," in *Proc. IEEE Conference on Neural Networks.* pp. 1-593-605, Jan. 1989.

[15] J. K. Kruschke, "Improving generalization in back-propagation networks with distributed bottlenecks," in *Proceedings of IEEE Conference on Neural Networks.* pp. 1-443-447, June 1989.