

Detection and Classification of Buried Dielectric Anomalies Using Neural Networks—Further Results

Mahmood R. Azimi-Sadjadi, *Senior Member, IEEE*, and Scott A. Stricker, *Member, IEEE*

Abstract—The development of a neural network-based detection and classification system for use with buried dielectric anomalies is the main focus of this paper. Several methods of data representation are developed to study their effects on the trainability and generalization capabilities of the neural networks. The method of Karhonen-Loeve (KL) transform is used to extract energy dependent features and to reduce the dimensionality of the weight space of the original data set. To extract the shape-dependent features of the data, another data preprocessing method known as Zernike moments is also studied for its use in the detector/classifier system. The effects of different neural network paradigms, architectural variations, and selection of proper training data on detection and classification rates are studied. Simulation results for nylon and wood targets indicate superior performance when compared to conventional schemes.

I. INTRODUCTION

THE development of a target detector system capable of detecting various types of dielectric anomalies, such as landmines, under different environmental conditions presents many problems [1], [2]. Targets have diverse sizes and compositions. The dimensions can range from a few inches to over a foot in diameter, and they may be composed of metal, plastic, wood, or a combination of these. The target emplacement environment can also be variable. Soil properties vary with location and moisture conditions. Rocks, roots, soil roughness, and soil inhomogeneities appear as targets to many sensors. The occurrence of anomalies with responses which closely resemble those of the targets may lead to high false positive rates. All these problems contribute to an overall system with unacceptable detection rates and high false alarm rates which may lead to serious injury or death to personnel and costly damage to equipment. Thus, an improvement over the capabilities of existing methods can lead to a corresponding increase in the survival of vital resources. In an optimal detector system, the detection rates should be high, and the occurrence of false positives should remain low in order to avoid delays in a military environment where speed is crucial.

Conventional target detection schemes such as thresholding and correlation matching [2], [3] have only been partially successful and have been shown to produce high false alarm rates. The performance of thresholding is directly related to

the threshold value used and the response of the sensor over a target. In [3] thresholding was performed on three sets of data, namely raw magnitude, ASUM, and TAD. The average of sums, ASUM, is an algorithm that calculates the weighted average of signal-background difference, weighted by the standard deviation of the background clutter. The target amplitude detector, TAD, computes the signal-to-background ratio at each spatial location, and then the magnitudes of the signal at a particular location are averaged over all the available frequencies of data. The method of correlation matching is based on calculating the cross-correlations between the data sets and a template in order to determine the amount of similarity between them. This method was also shown [3] to produce good detection rates but unacceptable false-alarm rates. These results and others presented in [3] indicate that neither of these methods can provide robust detection nor can they be used for classification. This severely limits the effectiveness of these methods for real-world applications. In [3] the application of neural networks was first studied for this problem. The results indicated substantial improvements over the conventional schemes.

In this paper the goal is to further study the use of neural networks and develop a system which offers better detection and classification rates and lower false-alarm rates than the previous systems. Specific emphasis is placed on the development of various data representation and pre-processing schemes and network structures and paradigms for automatic detection and classification of mine analogs [4]. The organization of this paper is as follows. Section II gives a discussion on various feature extraction schemes which are used in this paper. These include KL transform [3], [5] and Zernike moments [6], [7]. In Section III, the preprocessed data obtained using these methods are used to train and test several back-propagation networks (BPN) for detection and classification of targets based upon the extracted features. Section IV deals with the application of the Kohonen self-organized network [9] to target classification. The final Section gives the conclusions and discussions on this work.

II. FEATURE EXTRACTION METHODS

One of the most critical aspects involved with target detection and classification is the proper choice of a data representation scheme which impacts not only the size but also the speed and accuracy of the system. A scheme is usually chosen such that it extracts the salient features and reduces the dimensions of the data. In addition, it is desirable to orthogonalize the input data in order to provide a good set of independent features for

Manuscript received October 14, 1992; revised March 19, 1993. This work was supported by the U.S. Army Belvoir RD&E Center under contract No. DAAL03-86-D-0001.

M. R. Azimi-Sadjadi is with the Department of Electrical Engineering, Colorado State University, Fort Collins, CO 80523.

S. A. Stricker is with the Advanced Development Group, Exabyte Corporation, Boulder, CO 80301

IEEE Log Number 9216491.

training and testing of the neural network. In this section the data reduction and feature extraction are performed using two feature extraction methods, namely the 2-D KL transform or principal components [3], [5] and Zernike moments [6]–[7]. Simulation results on detection and classification of two types of targets when the feature sets are used as inputs to the neural network detector/classifier are presented in the subsequent sections.

A. 2-D KL Transform

The method of KL transform or principal components [3], [5] is a data reduction and feature extraction scheme which is based upon the statistical properties of the data. This is accomplished by projecting the data set along the directions of the individual orthonormal eigenvectors of the covariance matrix of the stationary vector stochastic process from which the samples are generated. If the first few eigenvalues of the covariance matrix contain most of the signal energy, the dimensionality of the data can be greatly reduced without losing much information by only retaining those components along with the principal eigenvectors. In addition to its optimality in data reduction, which is not shared by other feature extraction schemes, it provides salient features of the data that are decorrelated. This property is ideally suited for detection and classification purposes, as these decorrelated features or components can be used to train the detector/classifier effectively.

The target or anomaly returns within a window can be considered as a stationary 2-D complex random image. Let $\{x(m, n)\}$ be a 2-D random image of an ensemble set of M images each of size $(N \times N)$ with zero mean and covariance function $r_x(k, l)$ which can be computed by finding the average over all available M images using

$$r_x(k, l) \cong \frac{1}{MN^2} \sum_{i=1}^M \sum_{m=1}^{N-k} \sum_{n=1}^{N-l} x_i(m, n) x_i^*(m-k, n-l), \quad (1)$$

where $\{x_i(m, n)\}$ represents the i^{th} image array in the ensemble set.

Having determined the covariances, the eigenvectors and eigenvalues of the covariance matrix can be computed and used to form the transformation matrix [3], [5]. To accomplish data reduction, the magnitudes of the eigenvalues will be observed in order to determine how many would be sufficient to represent the image data accurately. Since the eigenvalues represent the energy of the signal, the process of finding the significant number of eigenvalues will be based on the number needed to represent 90% (or more) of the total energy of the signal [3]. This percentage was chosen so that the original signal could be accurately represented; while at the same time, a significant reduction in data could be obtained. The reduced transformation matrix, Ψ_r , can then be obtained by using the first L eigenvectors corresponding to the L most significant eigenvalues [3], [5]. This matrix will be used to produce the reduced transformed image \mathbf{Y}_r using the relationship

$$\mathbf{Y}_r = \Psi_r^* \mathbf{X}, \quad (2)$$

where \mathbf{X} is the row-ordered vector of the original image with size $K \times 1$. This transformation produces the reduced image \mathbf{Y}_r of size $L \times 1$. This decorrelated image [5] can now be used to train a neural network, the architecture of which will only require L inputs. This gives an immense reduction in size compared to a network that uses the entire image which corresponds to K inputs.

B. Zernike Moments

In [7], Zernike introduced a set of complex polynomials which form a complete orthogonal set over the interior of the unit circle, i.e. $x^2 + y^2 < 1$, where x and y are the spatial coordinates. Let the set of these polynomials be denoted by $\{V_{nm}(x, y)\}$. The form of these polynomials as described in [6] is

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta), \quad (3)$$

where n is a positive integer or zero, m is a positive or a negative integer satisfying the constraints $n - |m|$ even and $|m| \leq n$, ρ is the length of the vector from origin to (x, y) pixel, θ is the angle between vector ρ and x -axis in counter-clockwise direction, and $R_{nm}(\rho)$ is the radial polynomial defined as

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} \frac{(-1)^s (n-s)! \rho^{n-2s}}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!}. \quad (4)$$

Note that $R_{n,-m}(P) = R_{n,m}(P)$. This development assumes that the region of support of the original $M \times M$ discrete image, $f_d(x, y)$, is first mapped onto a square defined by the region $-1 \leq x \leq 1$, $-1 \leq y \leq 1$. This implies that the coordinate locations will no longer be integers but will have real values in the $[-1, +1]$ range. The Zernike moment of order n with repetition m for a function $f_d(x, y)$ can then be expressed as

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y g(x, y) V_{nm}^*(\rho, \theta), \quad x^2 + y^2 \leq 1, \quad (5)$$

where $g(x, y)$ is the mapped version of $f_d(x, y)$. To compute the Zernike moments of a given image, the center of the image is taken as the origin, and pixel coordinates are mapped to the range of the unit circle. Those pixels falling outside the unit circle are not used in the computation. It can be shown [6] that the magnitudes of the Zernike moments are rotation, scaling, and translation invariant; thus one can concentrate on $|A_{nm}|$ with $m \geq 0$ as far as the Zernike features are concerned. It is also shown [6] that A_{00} and A_{11} are constant for all normalized images; therefore they will not be used to classify images.

III. NEURAL NETWORK DETECTOR

The data used for this research was collected using the separated aperture sensor [2] which operates in the microwave band of frequencies starting from 600 MHz to 1 GHz. The data sets for frequencies 728, 760, 792, 824, and 856 MHz were used. The sensor scans each mine lane taking measurements at 27 positions horizontally (1.5" apart) for a total scan width

of 39". At each horizontal position, the complex scattering parameter was measured and stored. Once a row was scanned, the cart was moved forward 1.5", and the process was repeated until the entire 61' lane was covered. The buried targets were of identical shape and size and were made of nylon or wood. The data set for nylon consisted of 15 target and 12 nontarget windows, while those of the wood contained 12 target and 12 nontarget windows for each specified operating frequency.

The detection and classification are performed using a three-layer BPN [8]. In these networks, the choice of training data set can have a drastic effect on the performance. Empirical studies indicate that the best performance may be achieved by including targets with strong, medium, and weak returns in the training data set. The following sections present the results of neural networks which were trained with selected data from different feature extraction schemes.

A. KL Results

The "generalized KL transformation" matrix which was used to reduce the target or non-target images was formed from only nylon and wood target images (i.e. no nontargets). To preserve more than 90% of the total signal energy, the 50 (out of 225) eigenvectors corresponding to the 50 most significant eigenvalues were used to generate the transformation matrix. All the images were then reduced by this transformation matrix. The transformed images were used to train and test a two-layer network. The network had 50 inputs, 24 hidden layer nodes and three output nodes which were used to indicate nylon targets, nontargets, and wood targets with desired values of {1, 0, 0}, {0, 1, 0} and {0, 0, 1}, respectively. The training data consisted of 4 selected target and 4 nontarget images for both nylon and wood mine lanes at each frequency. The training took 30,000 training iterations to converge. The performance of the BPN is characterized by three parameters: 1) *detection rate*: the percentage of time that the network correctly recognizes a target; 2) *classification rate*: the percentage of time that the network both detects and correctly identifies the target type as either nylon or wood; 3) *false alarm rate*: the percentage of time that the network incorrectly detects a background (or nontarget anomaly) as a target. The detection, classification and false-alarm rates obtained using the KL data set for the resonant frequency (792 MHz) are, respectively, 93.3%, 71.4% and 0% for nylon data and 91.7%, 90.9% and 0% for the wood data. These results indicate that it is possible to obtain detection and false-alarm rates that surpass those of [3].

To extend this experiment, a generalized KL transformation matrix was formed with five randomly selected targets from each of the five-frequencies. This matrix consisted of the 50 most significant eigenvectors (90% of the total energy) and was used to reduce all the images in the five-frequency data set. Again, the training data consisted of the four selected images in each data set. The results of testing this five-frequency neural network with architecture 50-24-3 are shown in Table I.

These results show good detection and false-alarm rates for both nylon and wood targets and perfect classification for wood, but very poor classification for nylon. This poor

TABLE I
PERFORMANCE RESULTS FOR FIVE-FREQUENCY
GENERALIZED TARGET KL TRANSFORMATION MATRIX

Frequency	Nylon			Wood		
	Detection (%)	Classification (%)	False-Alarm (%)	Detection (%)	Classification (%)	False-Alarm (%)
728 MHz	86.7	0	0	88.3	100	0
760 MHz	86.7	0	0	88.3	100	0
792 MHz	93.3	14.3	0	91.7	100	0
824 MHz	93.3	7.1	0	91.7	100	0
856 MHz	100	6.7	0	100	100	0
Overall Total	92.0	5.8	0	90	100	0

classification rate for nylon targets could be blamed on several factors. It could be caused by the data set that was used to train the network, or it could be caused by a loss of generalization capability that may have occurred as a result of combining targets from five-frequencies to form the KL transformation matrix.

In order to further analyze the results of Table I, the locations of the targets that were not detected by the five-frequency network were tabulated to see if the same targets were being missed in all the frequencies of data. It was noted that a couple of targets that were not detected from the images produced at the lower frequencies were indeed detected from those formed at the higher frequencies. This prompted the development of a composite parallel neural network structure consisting of five single-frequency networks. The idea behind this network is that some single-frequency networks may be able to detect targets that others could not; thus by combining the outputs of all the single frequency networks, superior performance rates could result. Note that the input images, X_1 , are reduced using their associated generalized KL transformation matrix formed with four nylon and four wood targets for each of the five frequencies. The outputs of the individual networks are then combined to form average output values. The results of this composite network are shown in Table II. From these results, it is seen that the individual networks performed well, especially the 792 and 824 MHz networks with the 824 MHz network giving the best results when considering both the classification and detection rates. The performances of these networks, when compared to the individual networks shown in Table I, show several cases of improved results for both nylon and wood. The combined results of the composite network are given in the last row of this table.

It should be noted that there were several KL networks that provided 100% detection for both nylon and wood targets with 0% false-alarm rates but showed poor classification results. This fact led to the idea of using the KL networks for detection only and developing some other method to facilitate the classification of the targets once they are detected. This can be conceptualized in a system block diagram as shown in Fig. 1. With this in mind, the next section will explore how Zernike moments and possibly the KL transform could be

TABLE II
PERFORMANCE RESULTS FOR INDIVIDUAL NETWORKS
USING GENERALIZED KL MATRIX

Frequency	Nylon			Wood		
	Detection (%)	Classification (%)	False-Alarm (%)	Detection (%)	Classification (%)	False-Alarm (%)
728 MHz	73.3	63.6	16.7	91.7	100	8.3
760 MHz	66.7	70.0	0	91.7	90.9	8.3
792 MHz	86.7	69.2	0	91.7	90.9	0
824 MHz	100	66.7	0	91.7	90.9	0
856 MHz	100	66.7	25.0	100	91.7	8.3
Composite	80	66.7	0	91.7	100	0

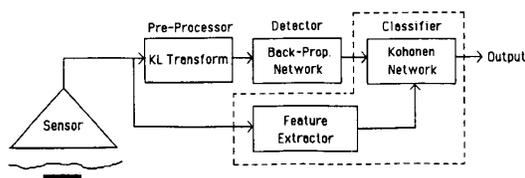


Fig. 1. Block diagram of overall detector/classifier neural network system. The system blocks inside the dashed line only operate when a target is detected.

used as the feature extractor for the classifier subsystem of the block diagram in Fig. 1. The classifier is a self-organizing map developed by Kohonen [9] which has the ability to effectively create spatially organized representations of various features of input signals. This type of network has been shown in [10] to produce better results than BPN especially for high-dimension data with overlapping class boundaries. Since the landmine classification task involves the heavily overlapping classes of nylon and wood targets, it is expected that some improvements in the classification rates can be achieved when self-organized networks are used.

IV. SELF-ORGANIZED NEURAL NETWORK CLASSIFIER

The self-organizing map is a paradigm developed by Kohonen [9], [10] that creates a vector quantizer by adjusting weights from common inputs to cells of the network through an unsupervised learning process. The cells then become tuned to various input patterns or classes, and their locations tend to become ordered in a meaningful coordinate system. The spatial location of a cell in the network will then correspond to a particular group of input signal patterns or a cluster. The clusters or classes formed will approximate the probability density function of the input vectors. Kohonen [9] observed that if the maps are to be used for pattern classification, their accuracy can be greatly increased if the cells are fine tuned using supervised learning principles known as Learning Vector Quantization (LVQ) methods. There are two different types of LVQ schemes [9], [10] that can be applied to the landmine pattern classification task.

The LVQ1 algorithm [9] is used to adjust the weight vector, \mathbf{m}_i , of an image, \mathbf{X}_i , such that they are pulled

away from classification boundaries. This is accomplished with an iterative supervised learning which results in an arrangement of \mathbf{m}_i that closely resembles that of a Bayes classifier. The LVQ2 algorithm [9] can also be used to fine tune the results of a self-organized map by positioning the weight vectors of an image according to its two nearest neighbors. A windowing operation is performed to move the midplane of the neighboring weight vectors toward the crossing surface of the class distributions, thus causing the midplane to asymptotically coincide with the Bayes decision boundary.

A. Results of KL Data

The data which was obtained by using the KL transform, as described in Section III, was used to train and test several LVQ networks. Both the LVQ1 and LVQ2 algorithms were used to fine tune the networks. The LVQ simulations were implemented using the recommended number of training iterations which was equal to 40 times the number of training vectors. The default initial value of $\alpha = 0.3$ was used for the adaptive gain. This value was monotonically decreased to zero during the training process. There are two factors that should be checked to monitor the training process [11]. First, the medians of the shortest distances between initial codebook vectors for different classes should not differ by more than a factor of two. Second, to ensure a good approximation, these medians should be somewhat smaller than the standard deviations of the input samples in all the respective classes. For example, the medians of the shortest distances for the nylon and wood KL data were 0.054 and 0.049 respectively. The standard deviations of the KL input vectors were 0.078 for nylon and 0.059 for wood. A check was also performed to see if the number of training iterations was adequate for network convergence. This was done by setting the number of training iterations equal to 500 times the number of training vectors. The simulation results for this network turned out to be identical to those trained with the default value of 40 times the number of training vectors. This shows that the networks were being fully trained, as indicated by the two factors above, when the default values were used.

Table III shows the simulation results for the classification of nylon and wood targets by single-frequency networks that were trained with the selected data from the five different frequencies, as described in Section III. Note that the '*' in Table III indicates that the network would not converge for this data set. This means that during the initialization process, the training samples were found to be unacceptable representations of their specific class identifiers. This table also shows results for a five-frequency KL network that was trained with four nylon and four wood targets from each of the five frequencies of data. The single-frequency LVQ networks showed generally fair classification rates for wood targets but poor rates for nylon. The five-frequency network, however, performed better overall for both nylon and wood. It can be deduced from these results that more training data can lead to better classification rates. The single-frequency networks only had four nylon and four wood targets for a total of eight input vectors to use for training data; whereas, the five-frequency network had four

TABLE III
LVQ SIMULATION RESULTS FOR KL NETWORKS

Performance (Classification %)				
Frequency	Single Frequency		Five Frequency	
	Nylon	Wood	Nylon	Wood
728 MHz	80.0	58.3	66.7	100
760 MHz	53.3	75.0	80.0	100
792 MHz	66.7	58.3	80.0	91.7
824 MHz	53.3	83.3	80.0	91.7
856 MHz	*	*	80.0	91.7

nylon and four wood targets from each of the five frequencies giving a total of 40 input training vectors.

The classification results of Table III can be compared with those of Tables I and II in Section III. This comparison of classification rates for the KL data using BPN and LVQ networks shows that the classification rates for nylon targets were greatly increased by using the LVQ networks, while the trade-off of a slight loss in accuracy for wood classification was observed in a few cases.

B. Zernike Moments Results

A set of 47 Zernike moments starting with $A_{2,0}$ and ending with $A_{12,12}$ was calculated from (5) and used to train and test LVQ networks. It was believed that this set of data would produce good classification rates since Zernike moments are orthogonal and thus are well suited for training an LVQ network. Table IV shows results for single-frequency LVQ networks and a five frequency LVQ network. The single-frequency networks performed exceptionally well for wood but poorly for nylon. The five-frequency LVQ network outperformed the single-frequency networks and produced several cases of acceptable classification rates for both nylon and wood targets. This result again points to the belief that more training data may lead to better results as previously discussed with the KL LVQ networks. The results of five-frequency LVQ networks which have classification rates of greater than 80% for both nylon and wood targets at a specific frequency would be considered as acceptable for use as the classifier block of the overall detector/classifier system.

C. Results of Raw Data

The final set of data that was used to train and test the LVQ networks is the raw (unprocessed) magnitude data that was obtained directly from the separated aperture sensor. This data represents the magnitude (in dB) of the scattering parameter [3] between the sensor transmitter and receiver. A set of single frequency networks and a five-frequency network were trained with the raw data, and the classification results are shown in Table V. For the single-frequency networks there were three cases for which the LVQ networks would not train, and furthermore there were not any cases where the classification for both nylon and wood was greater than 80% simultaneously.

TABLE IV
LVQ SIMULATION RESULTS FOR ZERNIKE MOMENT NETWORKS

Performance (Classification %)				
Frequency	Single Frequency		Five Frequency	
	Nylon	Wood	Nylon	Wood
728 MHz	46.7	100	66.7	83.3
760 MHz	73.3	100	86.7	75.0
792 MHz	66.7	91.7	80.0	83.3
824 MHz	60.0	83.3	73.3	83.3
856 MHz	66.7	100	86.7	91.7

TABLE V
LVQ SIMULATION RESULTS FOR RAW DATA NETWORKS

Performance (Classification %)				
Frequency	Single Frequency		Five Frequency	
	Nylon	Wood	Nylon	Wood
728 MHz	86.7	58.3	60.0	91.7
760 MHz	80.0	41.7	86.7	100
792 MHz	80.0	33.3	86.7	100
824 MHz	*	*	93.3	91.7
856 MHz	73.3	58.3	93.3	91.7

However, the five-frequency LVQ network had four cases where the classification rates were greater than 80% for both nylon and wood targets. This leads to the observation that once again the five-frequency network performed better than the single-frequency networks. This reinforces the previous statements about the correlation between the amount of training data and the performance of the LVQ networks. It may be surprising to note that the five frequency LVQ network trained with raw (unprocessed) data performed better, for some cases, than many of the LVQ networks which were trained with the preprocessed data sets presented earlier. However, this fact brings up the trade-offs between the dimensionality of the network and its performance which must be considered before a final statement about the best data representation scheme can be made. The training vectors for raw data contained 225 elements; this is more than four times the number of elements needed to represent the images with the KL or Zernike data. This leads to larger architectures and significantly longer training times for the raw-data LVQ networks.

V. CONCLUSION

The central focus of this paper was the application of signal-processing methods and neural network technology to the detection and classification of landmines. It was shown that the neural network-based approach provided substantial improvements in detection, classification, and false-alarm rates when compared to conventional pattern recognition schemes

[3]. These improvements in performance were directly related to the use of neural networks in conjunction with several different signal-processing schemes which were implemented to extract features from the landmine images.

The KL transform was used to perform data reduction and to extract the energy dependent features of the images. It was shown that the KL transform can pack the energy of the target and nontarget images into the most significant eigenvalues. Other advantages of the KL transform include its invariance to translation and rotation of the input data, and even more importantly, the decorrelation property of the KL transform values which led to improvements in the training of neural networks. Results of BPN and LVQ networks trained with KL data showed outstanding performance for detection, classification, and false-alarm rates. Another feature extraction method, namely Zernike moments, was studied to determine its effectiveness in providing appropriate data for the classification of nylon and wood targets. This technique showed some promising classification results when used to train a given frequency of LVQ network. A set of raw (unprocessed) data was also used to train and test LVQ networks. A five-frequency LVQ network trained with raw data was shown to produce classification results that were better than many of the networks which had been trained with preprocessed data. However, there is a big trade-off in network size which in turn leads to much longer training times for the raw-data networks.

Finally, special neural network models [12] could be investigated for future work to perform translational, rotational and magnitude invariance pattern classification.

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," U.S. Army Ft. Belvoir RD&E Center, Ft. Belvoir, VA, Report No. 2294, pp. 27-32.
- [2] L. S. Riggs and C. A. Amazeen, "Research efforts with the waveguide beyond cutoff or separated aperture dielectric anomaly detection scheme," U.S. Army Ft. Belvoir RD&E Center, Ft. Belvoir, VA, Report No. 2497, 1989.
- [3] M. R. Azimi-Sadjadi, D. E. Poole, S. Sheedvash, K. D. Sherbondy and S. Stricker, "Detection and classification of buried dielectric anomalies using a separated aperture sensor and a neural network discriminator," *IEEE Trans. Instrum. Meas.*, Vol. 41, pp. 137-143, February 1992.
- [4] S. A. Stricker, "Detection and classification of buried dielectric anomalies using neural networks," M.S. thesis, Colorado State University, Fort Collins, CO, Mar. 1992.
- [5] A. K. Jain, *Fundamentals of Digital Image Processing*, NJ: Prentice Hall, 1989.
- [6] A. Khotanzad and Jiin-Her Lu, "Classification of invariant image representations using a neural network," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-38, no. 6, pp. 1028-1038, June 1990.
- [7] F. Zernike, *Physica*, vol. 1, p. 689, 1934.
- [8] D. C. Plant, S. J. Nowlan and G. E. Hinton, "Experiments on learning by back-propagation," CMU Internal Report, June 1986.
- [9] T. Kohonen, "The self-organizing map," *Proc. of the IEEE*, vol. 78, no. 9, Sept. 1990.
- [10] T. Kohonen, G. Barna and R. Chrisley, "Statistical pattern recognition with neural networks: benchmarking studies," *Proc. of Inter. Joint Conf. on Neural Networks*, pp. 1.61-1.68, July 1988.
- [11] T. Kohonen, J. Kangas, J. Laaksonen and K. Torkkola, "LVQ-PAK: The learning vector quantization program package version 1.1," Helsinki University of Technology, Laboratory of Computer and Information Science, Dec. 1991.
- [12] M. B. Reid, L. Spirkovska and E. Ochoa, "Simultaneous position, scale and rotation invariant pattern classification using third-order neural networks," *Trans. Neural Networks*, vol. 1, pp. 154-159, July 1989.



Mahmood R. Azimi-Sadjadi (S'81-M'81-SM'89) was born in Tehran, Iran in 1952. He received the B.S. degree from the University of Tehran, Iran, in 1977, the M.Sc. and Ph.D. degrees from the Imperial College, University of London, England, in 1978 and 1982, respectively, all in electrical engineering.

He served as an assistant professor in the Department of Electrical and Computer Engineering, University of Michigan-Dearborn. Since July 1986 he has been with the Department of Electrical Engineering, Colorado State University where is is an Associate Professor. His areas of interest are digital signal/image processing, multidimensional system theory and analysis, adaptive filtering, system identification and neural networks. He is the co-author of the book, *Digital Filtering in One and Two Dimensions*, published in 1989.

He is the recipient of the 1990 Battelle Faculty Fellowship Award and the 1984 DOW Chemical Outstanding Young Faculty Award of the American Society for Engineering Education.



Scott A. Stricker (S'88-M'92) was born in Great Bend, KS, in 1967. He received the B.S. (Magna Cum Laude) and M.S. degrees, both in electrical engineering, from Colorado State University, Fort Collins, CO in 1990 and 1992, respectively.

In 1992 he joined the Advanced Development Group at Exabyte Corp. in Boulder, CO. Since joining Exabyte, he has worked as a servo control engineer on the company's future generation of 8 mm helical scan tape storage products. His current research interests include control systems, signal

processing, and neural networks.