A Neural Network Filter to Detect Small Targets in High Clutter Backgrounds

Mukul V. Shirvaikar and Mohan M. Trivedi

***Abstract*— The detection of objects in high-resolution aerial imagery has proven to be a difficult task. In our application, the amount of image clutter is extremely high. Under these conditions, detection based on low-level image cues tends to perform poorly. Neural network techniques have been proposed in object detection applications due to proven robust performance characteristics. A neural network filter was designed and trained to detect targets in thermal infrared images. The feature extraction stage was eliminated and raw gray levels were utilized as input to the network. Two fundamentally different approaches were used to design the training sets. In the first approach, actual image data were utilized for training. In the second case, a model-based approach was adopted to design the training set vectors. The training set consisted of object and background data. The neuron transfer function was modified to improve network convergence and speed and the backpropagation training algorithm was used to train the network. The neural network filter was tested extensively on real image data. Receiver Operating Characteristic (ROC) curves were determined in each case. The detection and false alarm rates were excellent for the neural network filters. Their overall performance was much superior to that of the size-matched contrast-box filter, especially in the images with higher amounts of visual clutter.**

I. Introduction: The ATR Problem

Object detection in aerial imagery is a basic task in military reconnaissance. It has proven to be a task with a high degree of difficulty due to the nature of the imagery [1]. We deal with targets in high-resolution aerial imagery acquired by infrared sensors (see Fig. 1). In spite of low sensor noise, detection performance is mainly affected by

- small target signature,
- high background clutter,
- continuous variations in background,
- huge amounts of image data, and
- high processing speed required.

Efficient approaches to detection are based on multistage analysis of data in various domains such as the spectral, spatial, and topographic domains. The main thrust was to progressively narrow the *"focus of attention"* so as to avoid processing all the image data [2], [3].

Automatic Target Recognition (ATR) systems have proved to be very vital and successful in military applications. ATR systems can substantially reduce operator loads in real-lfe scenarios. The studies by Peters [4], Schachter [5], and Bhanu [6] describe the basic structure of ATR algorithms. Fig. 2 shows the basic configuration for an ATR system. There

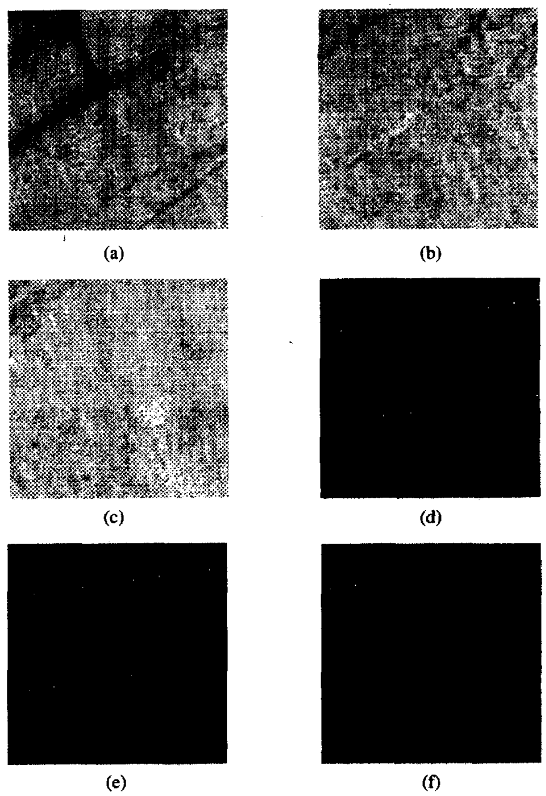


Fig. 1. Typical high-resolution thermal infrared images (a), (b), and (c) with object signatures and object location map (d), (e), and (f). The images are characterized by high clutter, small targets, varying background, and huge amounts of data (each image has 512×512 pixels).

may also exist a preprocessing component which improves the quality of the initial image. The preprocessor generally uses one or more local filters or histogram equalization or linear expansion or contrast stretching to reduce noise and increase the contrast between the targets and the background. It may also estimate target size based on range information. The detector finds and locates those regions in the image which are most likely to contain the targets. Further image analysis separates the potential target from the background by examining the features of the location passed to it by the detector. It is supposed to reject the clutter or false alarms and select the potential targets. The ATR systems use different degrees of *a priori* information to recognize the targets of interest, and this has a direct impact on the robustness and generality of the algorithms. As evaluated for their effectiveness by Schachter [5], the generality of most of the complete ATR systems is suspect, i.e., they perform poorly when confronted by test images not used in the training phase.

Manuscript received October 13, 1992; revised March 4, 1993 and August 3, 1993. This work was supported by the U.S. Army Belvoir Research, Development and Engineering Center under Grant DAAK70-89-K-0003.

The authors are with the Computer Vision and Robotics Research Laboratory, Electrical and Computer Engineering Department, The University of Tennessee, Knoxville, TN 37996-2100 USA.

IEEE Log Number 9212565.

1045-9227/95$04.00 © 1995 IEEE

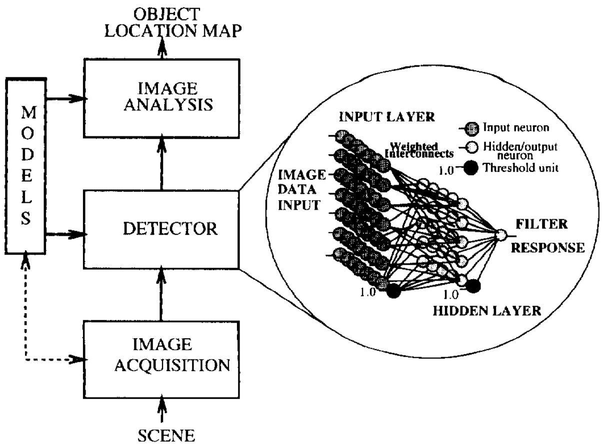


Fig. 2. The basic ATR system configuration, with the neural network filter and its architecture.

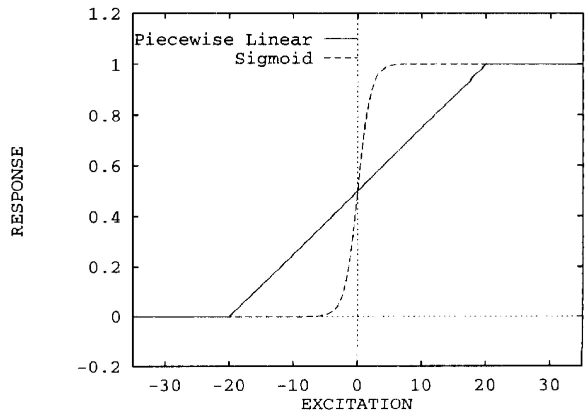


Fig. 3. The sigmoid (dashed) and modified piecewise linear (solid) neuron transfer function.

Often expressed is the opinion that the detection approaches developed so far do not meet the requirements of robustness, as would be desired. Typical detection rates range from 60 to 90% and false alarm rates are high. The human visual system is astonishingly resilient to these conditions especially after being trained. A trained eye is found to perform consistently better than most detection systems designed to date. This is the case for both target detection as well as false alarm rejection. This can be partially explained by the fact that most automatic detection systems have yet to reach the competence level of the human visual system. The visual system can reach intelligent conclusions even from low-level features, unlike current automatic systems. This can be attributed to the significantly larger processing power of the brain. The limited success of present computational architectures and techniques has led to research in the application of new ideas to detection technology. Some of these are: *parallel processing, multisensor fusion, fractals,* and *neural networks* [7].

It has been hypothesized that computational architectures similar to the brain might be the solution to grasping the higher level vision cues or features. Roth [8] has put forth a convincing argument for the utilization of neural networks in detection systems. The argument is based on the characteristics exhibited by neural networks, namely: 1) learning capabilities, 2) adaptability, and 3) graceful degradation. It has been shown that neural networks perform nearly as well as parametric optimal detectors for detecting noisy signals [9]. Neural network backpropagation training algorithms have been trained to differentiate between surface and submarine targets using acoustic signals emitted by them [10], and also as a part of other object recognition schemes [11]. Roth [12] designed a neural net for the extraction of weak targets in high clutter environments. In order to maintain reasonable false alarm rates (FAR), a constant FAR (CFAR) detector selects high thresholds, due to which weak targets are missed completely. Neural net simulation of feedforward and graded-response Hopfield nets were shown to implement the optimum postdetection target track receiver, and substantial signal-to-noise gain was achieved.

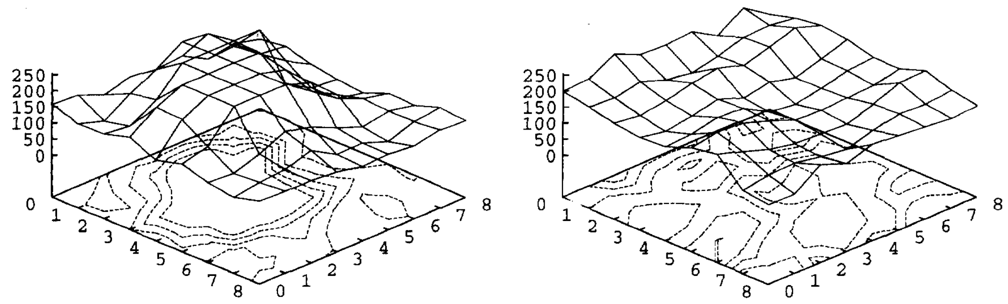
TABLE I
A Comparison of the BP Training Convergence Characteristics for the Sigmoid and the Modified Piecewise Linear Neuron Characteristics

CONVERGENCE FOR	SIGMOID CHARACTERISTIC	PIECEWISE LINEAR CHARACTERISTIC
ex-OR PROBLEM	Converged in 68 iterations	Converged in 11 iterations
IMAGE-BASED TRAINING	Failed to converge	Converged in 58 iterations
MODEL-BASED TRAINING	Failed to converge	Converged in 49 iterations

The remainder of the paper describes the design and training of a neural network filter, for the detection of weak targets in high-resolution thermal infrared imagery, followed by experimental results and conclusions.

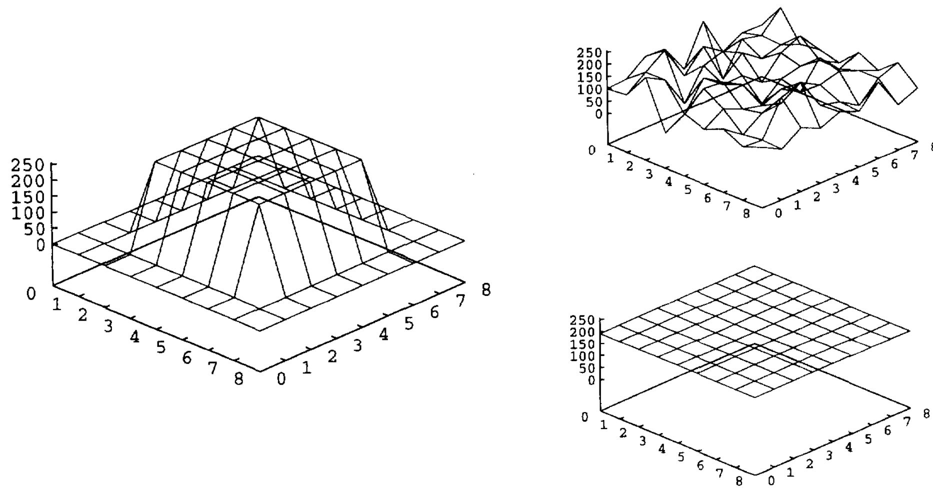
II. The Neural Network Filter

The neural network filter consists of a feedforward neural network with two layers. Neurons in any layer are connected only to neurons in the next layer. The input to the neural network consists of raw gray level values. Fig. 2 shows a sketch of the neural network architecture. Unlike previous ATR approaches [12], in which the entire image was the input to the neural net, we use the neural network like a moving window transform. Operation of the neural network filter over an image is similar to the operation of a spatial domain filter. The neural net as a filter has recently been applied to scene segmentation [13], [14] and wafer inspection [15], but has not been applied in the ATR domain. The neural network filter is convolved with the image to produce output at each pixel. The neuron output is scaled across the gray level range. The neural network filtering thus produces a gray level response map filtered image. The filter response is supposed to be high for target pixels and low



(a) (b)

Fig. 4. The image data samples for network training (a) object, (b) background. The training sample set consisted of seven object and seven background samples from image 1(a).



(a) (b)

Fig. 5. The model-based samples for network training (a) object, (b) background. Two of the twelve background models used are shown.

for background pixels. The filtered image can be thresholded to obtain the intermediate object location map. False alarm rates can be controlled by threshold selection strategies, low thresholds being favored at the detector stage so as not to preclude any targets from subsequent stages. The requirements for the detection stage are a high detection rate with a low false alarm rate.

A. The Piecewise Linear Neuron Characteristic

In the examples considered in this paper, the input stage of the neural network filter contains 81 nodes (an array of 9×9). The major difference between the neural network filter and other nets lies in the neuron transfer functions for the hidden and subsequent layers. Fig. 3 shows the modified neuron transfer function utilized to implement the network [16], [17]. The commonly used sigmoid characteristic is stretched and approximated in a piecewise linear fashion to obtain the desired transfer function. The excitation for each hidden layer neuron is a weighted combination of the input layer neurons. The response of each hidden layer neuron can be compared to the response of a nonlinear spatial domain filter.

Neural networks consisting of two input and two hidden layer neurons were trained for the classic ex-OR problem. The network with the modified neurons performed significantly better. Table I shows a comparison of the backpropagation (BP) training convergence characteristics for the sigmoid and the modified piecewise linear neuron characteristics. The neural network filter failed to converge with a sigmoid characteristic, for both image-based training and model-based training, even after 1000 iterations in each case. On the basis of these observations, the major advantages the modified transfer function provided, could be summarized as follows:

1) improved convergence properties for training,
2) significant reduction in training time required, and
3) reduction in computation time during operation.

The reduction in computation time can be attributed to the lower mathematical complexity of the piecewise linear neuron characteristic.

B. A Novel Model-Based Training Methodology

The backpropagation training algorithm was utilized to train the network. Network weights were initialized to small random values. The traditional algorithm was used with modifications [16] made in the error computation for backpropagation, due to the modified neuron transfer function. Network interconnection weights w_{ij} are modified at each step using the delta values for the entire network, with the following equation:

$$\Delta w_{ij}(n) = L\delta_j y_i + M\Delta w_{ij}(n-1) \qquad (1)$$

where L is the learning rate, M the momentum factor, and n is the step number, and

$$\delta_j = \rho(t_j - o_j)$$
$$\delta_j = \rho\left(\sum \delta_k w_{kj}\right)$$

are the errors or deltas for the output nodes and hidden nodes, respectively, ρ is the slope of the linear characteristic ($\rho = 0.025$ was used), and o_j and t_j are the actual and observed target outputs for the neural network filter.

Two approaches were used to compile the training sample set for the network training algorithm. In the first case, actual image data were utilized with the ground truth information. Seven object and seven background samples were chosen from the image, the background samples being chosen randomly. Fig. 4 shows plots of the image data samples for the objects and background.

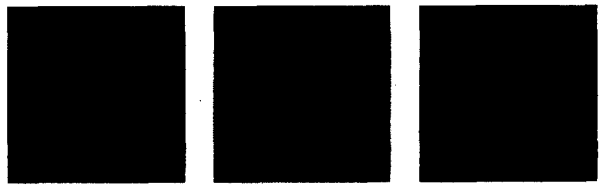
It is often difficult to obtain actual ground truth information to train the network. A model-based approach is more convenient in such situations. The second training set was designed using object and background models, samples of which are shown in Fig. 5. One object model and 12 background models to cover the various possibilities were included in the training set.

The training process was quick and took 58 iterations in the case with 14 image data training samples and 49 iterations for the case with 13 model-based samples. This was achieved with a learning rate of 5 and momentum factor of 0.9. One iteration consisted of presenting the entire sample set to the network once. Training thresholds of 2.5% were specified for both high and low output cases.

III. Experimental Results

Fig. 1 shows typical high-resolution images in the infrared spectrum along with ground truth data. As can be seen, the object signatures are small (2–3 line-pairs/object) compared to the image size (512–512 pixels). The neural network filter was tested on several images. Fig. 6(a) and (b) shows the detection results for the image in Fig. 1(a) with the image data training samples and the model-based training samples, respectively. The detection results were compared to the contrast box filter results for the same image set [2]. Fig. 6(c) shows the results of contrast box filter analysis, for the image in Fig. 1(a). The same tests were conducted on other test images in Fig. 1, but the binarized and filtered images for them are not shown here.

Experiments were conducted to test the sensitivity analysis of the filter response to threshold selection. Each test image had three filter responses. The false alarm rates were determined at varying detection rates for each filter response for each image. These Receiver Operating Characteristics (ROC)



(a) (b) (c)

Fig. 6. The filter responses: (a) the neural filter trained with model-based data, (b) neural filter trained with image data, and (c) the size-matched contrast box filter. The results shown are for the image in Fig. 1(a).

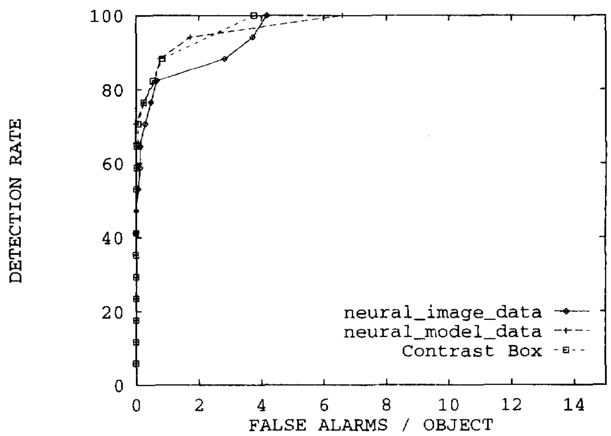


Fig. 7. ROC plots for the image in Fig. 1(a). Relative performance of the neural and contrast-box filters is graphically demonstrated. In this case, the performance of the neural and contrast box filters is comparable.

curves are shown in Figs. 7–9. The ROC plots give us a true picture of the performance characteristics of each filter and serve as an effective metric to compare them [18]. The detection rates at different false alarm rates for the various experiments are tabulated in Table II.

The test images 1 (a), (b) and (c), are arranged in increasing order of visual clutter. The detector filters being the first stage of the ATR system, high detection rates are desirable even at the expense of an unfavorable false alarm rate. The false alarm rates at high detection rates thus give us a measure of the filter performance. Low false alarm rates in the 85–100% detection rate range are an indication of good detector performance. As seen from Fig. 7, the performance of the image data trained neural network filter compares favorably with the contrast box filter. Figs. 8 and 9 display the true power of the neural filter trained with model-based data. Its performance is far superior to the other two filters, which is significant, *since it is for the images with higher amounts of visual clutter*. The neural filter trained with actual image data performed intermediate to the other filters in each case. This result can be attributed to the fact that the actual image data used for training purposes were far from ideal, due to the high amount of image clutter.

IV. Conclusions

The detection of target signatures in surveillance imagery is hampered by several conditions such as small target size,

TABLE II
PERFORMANCE COMPARISON FOR THE NEURAL NETWORK AND CONTRAST-BOX FILTERS

	FILTER	CONTRAST BOX	NEURAL (IMAGE DATA TRAINED)	NEURAL (MODEL DATA TRAINED)
IMAGE	DETECTION RATE P_D (%)	FALSE ALARMS PER OBJECT	FALSE ALARMS PER OBJECT	FALSE ALARMS PER OBJECT
1(a)	47.06	0.00	0.00	0.00
	52.94	0.00	0.06	0.00
	58.82	0.00	0.12	0.00
	64.71	0.00	0.12	0.00
	70.59	0.06	0.29	0.00
	76.47	0.24	0.47	0.24
	82.35	0.53	0.65	0.59
	88.24	0.82	2.82	0.76
	94.12	2.29	3.71	1.71
	100.00	3.76	4.18	6.59
1(b)	43.75	0.12	0.12	0.25
	50.00	0.19	0.12	0.25
	56.25	0.19	0.19	0.25
	62.50	0.25	0.19	0.25
	68.75	0.25	0.38	0.31
	75.00	0.31	1.00	0.94
	81.25	1.19	1.75	1.75
	87.50	2.06	3.75	2.00
	93.75	3.88	9.25	5.44
	100.00	17.88	12.56	6.19
1(c)	0.00	0.00	0.14	0.14
	14.29	0.00	0.14	0.29
	28.57	0.14	1.00	0.71
	42.86	1.14	1.00	1.00
	57.14	1.29	1.43	1.29
	71.43	1.86	2.29	1.71
	85.71	2.57	7.86	2.43
	100.00	16.57	8.43	2.43

background clutter, and unpredictable background variations. The human visual system performs very competently at such tasks due to its high processing power which allows it to use high level vision cues for decision making. An argument for the application of neural networks to automatic detection has been made, based on the characteristics of neural systems.

A neural network filter was trained and tested for the detection of small targets in high-resolution aerial thermal imagery. The neural network filters were trained using two types of samples: actual image data and target/background models. A modified neural network transfer function was utilized to implement the network, and the backpropagation algorithm was used to train the filter. For detection, the neural network filter is convolved with the entire image, and the convolution response map is thresholded to classify the different regions in the input image as a target or a background. ROC curves were used to compare filter performances. The performance of the neural network filters was superior to the contrast box filter in images with a high amount of visual clutter. The neural filter trained with model-based data performed better than the one trained with actual image data. The neural network filter operation can be compared to the use of multiple spatial domain filters, each contributing to the final response. This provides superior detection performance as our results illustrate. The results form a good argument for the use of neural filters in ATR applications, especially if they can be implemented in hardware in the future.

The adaptability of the neural network is very useful during training when the weights of the interconnections between the input layer and the hidden layer neurons are modified *in parallel* to achieve optimal decision surfaces. Thus, the parallelism of neural networks can be utilized to synthesize multiple spatial filters in parallel. Most of the current networks are artificial or simulated on computers. Much research has been done in the neural network area in recent years and its applicability is burgeoning [19]. Headway has been made in the implementation of neurons and interconnections in hardware. Neural network hardware technology is in nascent stage and the initial VLSI implementations have only been in existence for a short time. *Real neural computers can be*

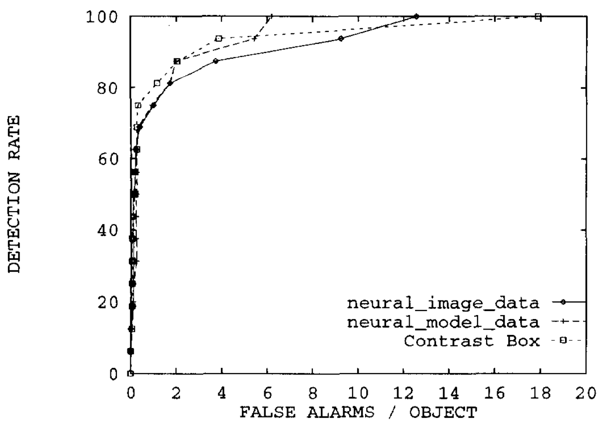


Fig. 8. ROC plots for the image in Fig. 1(b). Relative performance of the neural and contrast-box filters is graphically demonstrated. The contrast-box filter has an unacceptable high false alarm rate at high detection rates.

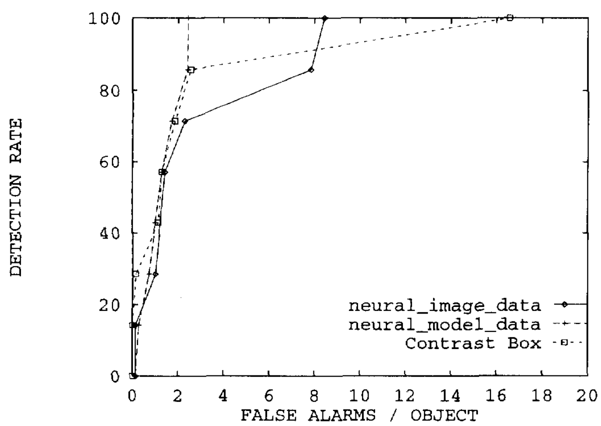


Fig. 9. ROC plots for the image in Fig. 1(c). Relative performance of the neural and contrast-box filters is graphically demonstrated. The neural filter trained with model-based data performs much better than the other two filters.

anticipated to be in operation sometime in the foreseeable future. The scope for development is promising, and this area should be noted as an important alternative for future ATR systems.

Further research could be conducted to optimize the training vector set, exhaustive model generation schemes, strategies for adaptability to new backgrounds, and continuous-learning neural networks.

ACKNOWLEDGMENT

We sincerely appreciate the valuable assistance provided by G. Maksymonko and P. McConnell. We thank the reviewers for their comments, which helped us to improve the quality of the manuscript.

REFERENCES

[1] M. V. Shirvaikar and M. M. Trivedi, "Developing texture-based image clutter measures for object detection," *Opt. Eng.*, vol. 31, pp. 2628–2639, Dec. 1992.

[2] ——, "Design and evaluation of a multistage object detection approach," in *Proc. Appl. Artif. Intell. VII Conf., SPIE,* Orlando, FL, Apr. 1990, pp. 14–22.

[3] M. M. Trivedi, C. Chen, and D. Cress, "Object detection by step-wise analysis of spectral, spatial and topographic features," *Comput. Vis., Graph. Image Processing,* vol. 51, pp. 235–255, 1990.

[4] R. A. Peters, II, "Image complexity measurement for predicting target detectability," Ph.D. dissertation, Dep. Elec. Comput. Eng., Univ. Arizona, 1988.

[5] B. J. Schachter, "A survey and evaluation of FLIR target detection/segmentation algorithms," in *Proc. Image Understanding Workshop,* Sept. 1982, pp. 49–57.

[6] B. Bhanu, "Automatic target recognition: State of the art survey," *IEEE Trans. Aerosp. Electron. Syst.,* vol. AES-22, pp. 364–379, July 1986.

[7] F. A. Sadjadi, "Automatic object recognition: Critical issues and current approaches," in *Proc. Automat. Object Recogn. Conf. SPIE,* Orlando, FL, Apr. 1991, pp. 303–313.

[8] M. W. Roth, "Survey of neural network technology for automatic target recognition," *IEEE Trans. Neural Networks,* vol. 1, pp. 28–43, Mar. 1990.

[9] C. F. Bas and R. J. Marks, "The layered perceptron versus the Neyman–Pearson optimal detection," in *Proc. Int. Joint Conf. Neural Networks,* Singapore, Nov. 1991, 18–20, pp. 1486–1489.

[10] R. H. Baran and J. P. Coughlin, "Neural network for passive acoustic discrimination between surface and submarine targets," in *Proc. Automat. Object Recogn. Conf. SPIE,* Orlando, FL, Apr. 1991, pp. 164–176.

[11] C. Mehanian and S. J. Rak, "Bidirectional log-polar mapping for invariant object recognition," in *Proc. Automat. Object Recogn. Conf. SPIE,* Orlando, FL, Apr. 1991, pp. 192–199.

[12] M. W. Roth, "Neural networks for extraction of weak targets in high clutter environments," *IEEE Trans. Syst., Man, Cybern.,* vol. 19, pp. 1210–1217, Sept./Oct. 1989.

[13] H. A. Malki, "Image segmentation using multilayer neural network," in *Proc. Int. Joint Conf. Neural Networks,* Baltimore, MD, June 1992, vol. 4, pp. 354–360.

[14] E. Viennet and F. F. Soulie, "Multiresolution scene segmentation using MLPs," in *Proc. Int. Joint Conf. Neural Networks,* Baltimore, MD, June 1992, pp. v.3, 55–59.

[15] D. I. Sikka, "Two dimensional curve shape primitives for detecting line defects in silicon wafers," in *Proc. Int. Joint Conf. Neural Networks,* Baltimore, MD, June 1992, pp. v.3, 591–596.

[16] M. V. Shirvaikar, "A Neural network system for character recognition," Master's thesis, Univ. Maine, 1988.

[17] M. T. Musavi, A. Rajavelu, and M. V. Shirvaikar, "A neural network approach to character recognition," *Neural Networks,* vol. 2, pp. 387–393, Sept. 1989.

[18] R. C. Eberhart and R. W. Dobbins, *Neural Network PC Tools: A Practical Guide.* New York: Academic, 1990.

[19] C. G. Y. Lau and B. Widrow, Special Issue on Neural Networks, *Proc. IEEE,* vol. 78, Sept. 1990.