



## CONTRIBUTED ARTICLE

# A Modular Neural Network for Enhancement of Cross-polar Radar Targets

ANDREW M. UKRAINEC AND SIMON HAYKIN

McMaster University

(Received 1 October 1994; revised and accepted 1 May 1995)

**Abstract**—A polarimetric radar navigation system makes use of polarization rotating twist-grid retroreflectors in order to navigate a confined waterway, even in inclement weather or after dark. Despite the polarization diversity offered by such a radar target, depolarization allows significant cross-polar clutter to obscure the reflector return.

A novel modular neural network solution integrates an adaptive cross-polar interference canceller, a radial basis function network, and a conventional cell-averaging CFAR processor to successfully demonstrate the enhancement and detection of a polarization target. The modular solution outperforms any one of the aforementioned methods on their own. This is indicated subjectively through the display of the resultant processed images, and objectively by the estimates of target-to-clutter ratio and receiver operating curves.

A post-detection processor uses a priori information about the reflector location along the water-land boundary of the waterway. A fuzzy processor combines primary detection information with the output from a vision-based edge detector to effectively remove false alarms.

**Keywords**—Radar, Detection, Modular, Neural network, Navigation, Polarization, Image processing, Vision.

## 1. INTRODUCTION

### 1.1. Precise Navigation Problem

Inland confined waterways, such as St. Lawrence seaway and the Mississippi River, are of great importance in the transportation of goods by ship. The shipping community strongly desires to make best use of this resource. However, some factors exist that limit the time that the waterways may be used. During periods of low visibility, namely fog, heavy rain, or darkness, the ships cannot navigate by visual aids. The buoys which are used as visual aids in navigation are deployed at the beginning of the shipping season, and removed at the end of the season. To extend the season beyond the times when buoys are available, and to travel in times of poor visibility, another navigational aid is needed to

supplant, or replace, that of visual navigation by the ship's pilot. It goes without saying that such a system must be robust and provide an accuracy of navigation comparable to that of the ship's pilot. It was judged that a ship's pilot could navigate visually within an accuracy of approximately  $\pm 3$  m from the channel centerline and  $\pm 7$  m from the channel limits.

The use of a marine radar system was investigated as a means by which radar ranging to known targets could be used for triangulation to locate a ship within a waterway. The discriminants that are available to the radar systems designer for the identification of stationary targets are power, frequency, and polarization. In the context of noncoherent radar, polarimetric diversity is the only viable option. The polarization of an electromagnetic wave is defined as the direction of the electric field component. Most marine radars transmit with the electric field linearly polarized in the horizontal plane. A passive reflector target which is able to rotate the plane of polarization efficiently is obviously needed. It is known that a dihedral reflector mounted on a  $45^\circ$  angle from the horizontal has the desirable property of rotating a linearly polarized field through  $90^\circ$  efficiently. Early studies showed that the dihedral exhibited the polarimetric characteristics that could be exploited in navigation. Unfortunately, the dihedral only

---

Acknowledgements: Financial support provided by the Natural Sciences and Engineering Research Council (NSERC), Ottawa, for doing the work discussed herein is deeply appreciated. The authors are also grateful to an anonymous reviewer for many useful comments in an earlier version of the paper.

Requests for reprints should be sent to A. M. Ukrainec, Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada; E-mail; ukraivec@soma.crl.mcmaster.ca

exhibits its polarization rotating property over a very narrow azimuthal angle. Further investigations of reflector design resulted in the invention of the trihedral twist-grid reflector, which has the same polarization characteristics as the dihedral as well as a wide azimuthal response (Macikunas et al., 1988). Using the reflector in field experiments verified its desirable performance characteristics.

The novel polarimetric radar for accurate navigation (PRAN) system was thus invented at the Communications Research Laboratory, McMaster University. The system consists of a set of polarization-twisting reflectors situated along a confined waterway in known locations so that a ship with the proper radar equipment can ascertain its position with respect to the shoreline. Figure 1 shows a possible deployment of the system along a confined waterway. The interested reader is referred to Haykin (1992) which gives a complete historical perspective on the development of the PRAN system. The following summarizes the important properties of a trihedral twist-grid reflector:

- passive device;
- efficient rotation of polarization;
- high cross-polar radar cross-section;
- wide azimuthal angular response;

- low cost and simple construction;
- insensitive to alignment errors;
- robust.

Limitations in the marine radar system components, as well as multipath and the natural depolarization characteristics of the environment increase the clutter in the received radar returns and make the reflectors less visible than is desirable. Rather than using impractically large reflectors, or performing expensive upgrades to the radar system, the solution pursued herein involves processing the horizontally and vertically polarized returns jointly in some optimum fashion to increase the target-to-clutter ratio.

## 1.2. Signal Processing Solution

The phenomenon which causes a received wave to have a different polarization from the transmitted wave is called depolarization. The radar environment generally consists of targets, land clutter, and sea clutter whose back-scatter has different polarization properties. For example, an object may consist of various asymmetric objects with partially conductive surfaces, and this can cause depolarization of the incident wave. Diffuse scattering is the main cause for

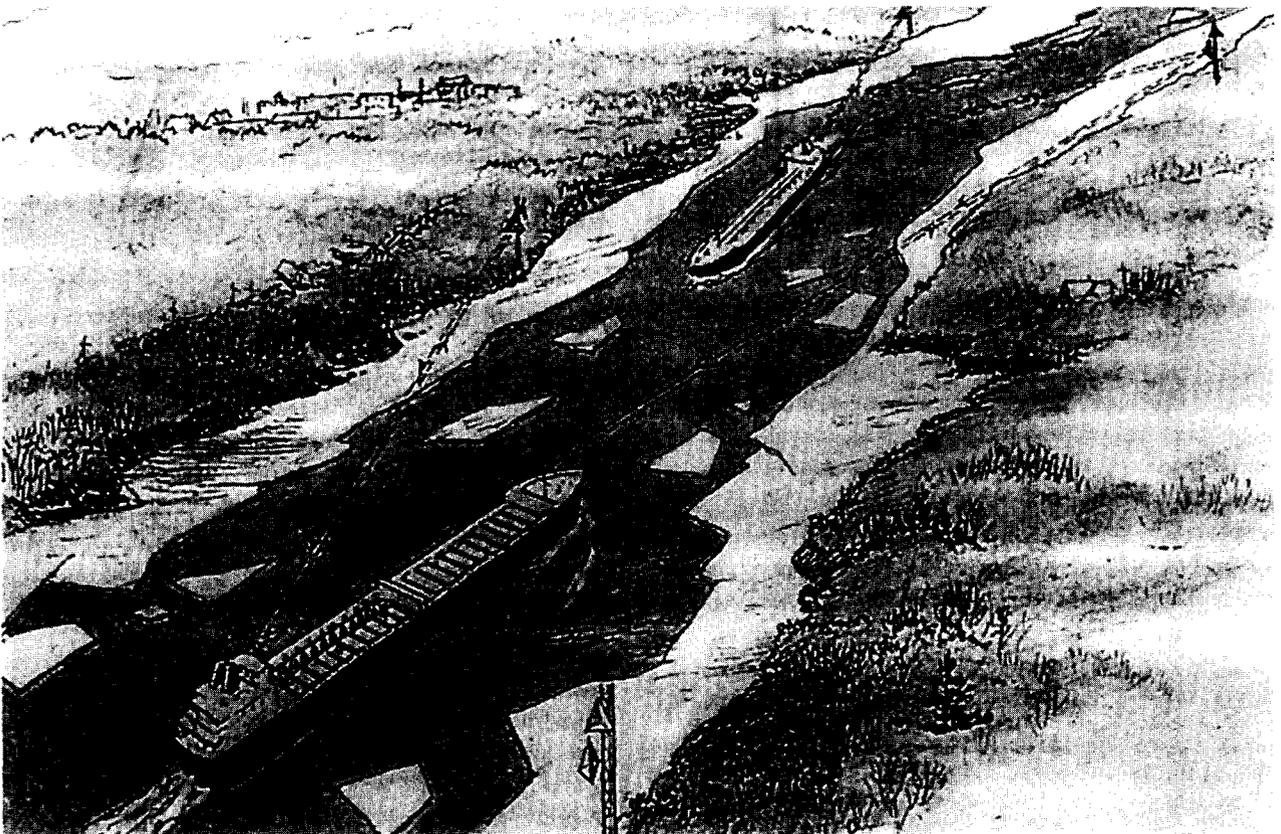


FIGURE 1. An artist's depiction of the polarimetric radar navigation system in operation.

depolarization in land and sea clutter, although man-made objects can cause small regions to have significant depolarization values. Due to the physical non-homogenous nature of various clutter regions, we therefore expect the depolarization values to vary over some range. Cross-polar returns for land clutter are quoted to be typically in the range of 3–10 dB less than the corresponding co-polar returns for linearly polarized X-band radar (Giuli, 1986). The corresponding range for sea clutter is 6–10 dB. For the problem at hand, the radar transmits only H-pol signals, and receives both H-pol and V-pol scattered energy non-coherently. In general, the clutter can be characterized as having a non-stationary and non-Gaussian joint probability density function. The signal processing solution must therefore address this important fact.

There is *a priori* information that an engineer can use to improve the visibility and hence the detection of the target. The targets are cooperative targets, and the location of the targets is controlled by the installers of the system. The placement of the reflector adjacent to a stationary undesirable point target with a large cross-polar response can therefore be avoided. The path from the ship's radar antenna to the reflector target must also be unobscured by natural objects (e.g., trees) or artificial objects (e.g., metal towers). The preferred location for these targets is therefore near the boundary between land and water.

Figure 2 is a block diagram of the radar signal processing system and the detection system. The radar signal processor accepts digitized radar video data from a dual-polarized scanning marine radar transceiver. The magnetron of the marine radar transmitter produces a narrow pulse (50 ns) of X-band signal at a regular rate determined by the pulse repetition interval (PRI = 294 ms). The parabolic antenna transmits the energy in a horizontal, linearly polarized fashion. The same antenna receives

reflected signals in both linear horizontal and vertical polarized planes. The radar receiver section has two logarithmic-response non-coherent sections that produce video sweeps from the received electromagnetic signal at each PRI interval for both horizontal (H) and vertical (V) polarizations. The rotational motion (28 rpm) of the scanning antenna results in video returns at regular azimuthal look directions, which map out a polar-coordinate image of the area surrounding the radar system. This video is normally presented to the operator on a plan position indicator (PPI) display. In an operational system, the operator could also choose to view the processed signal as well. The digitized video samples  $x_{HH}(n)$  and  $x_{HV}(n)$  are obtained from the dual-channel analog-to-digital (A/D) sampling system (30 MHz sampling rate, 8-bit quantization).

The samples are sent to the radar signal processing system, which processes HH-pol and HV-pol data in an optimum fashion. The radar signal processing system is a modular system, which makes use of adaptive and non-linear elements to address the non-stationary and non-Gaussian nature of the inhomogeneous clutter environment. One element of the signal processor is an adaptive cross-polar interference canceller, designed to cope with varying clutter regions. Another element is the mutual information network implemented using a radial basis function (RBF) network; it is trained using a mutual information principle in order to capture the non-Gaussian nature of the clutter. The outputs of these two elements are then combined together with the objective of maximum information preservation.

For automatic detection processing, the resultant image is passed to a conventional cell averaging-constant false alarm rate (CA-CFAR) processor. The enhanced image so produced (prior to thresholding) has the property of a constant false alarm rate. The thresholding stage maps the multivalued discrete

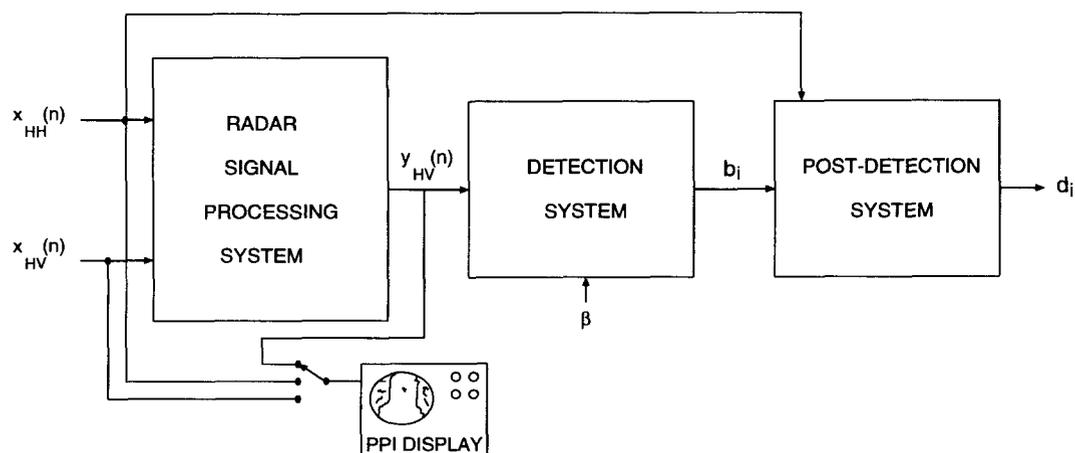


FIGURE 2. The signal processing and detection system.

image to a binary image, labeled  $b_i$  in Figure 2. The threshold  $\beta$  applied to the detection system determines the false alarm rate.

Finally, a post-detection processor uses *a priori* information about the waterway to remove false targets. Modern image processing techniques based on human vision are used to determine the land–water boundary. The post-detection processor uses fuzzy reasoning to only accept detections that occur in and around such a boundary, and to reject all other detections as false alarms. The final detected output,  $d_i$ , contains the desired reflector locations.

Throughout the development of the processing stages, an effort was made to consider solutions that are implementable in analog neural network architectures. Only simple elements need to be considered for use in the neural architecture, namely: delays, weights (both fixed and adaptive), multipliers, integrators, summers, threshold units, and Gaussian response units. Using these basic elements, structures such as adaptive filters, automatic gain controls, and fuzzy logic functions can be constructed.

## 2. MODULAR NEURAL NETWORK

A non-learning method is defined as a method that uses information other than that contained in the training samples (Hrycej, 1992). A non-learning modular neural network is defined as a super-structure that integrates trained neural sub-networks in a useful fashion, using a non-learning method. A non-learning modular neural network has the ability to integrate the desirable qualities of two learning-based networks to produce a result that is better than either network used on its own. As an additional benefit, the modular network should also be more robust in operation, since the network could continue to function even if there was a complete failure of one of the network sub-modules.

For a modular network to perform better than its parts, each sub-network must provide some independent information. Two signal processing methods are chosen which meet this criterion. The first is an adaptive cross-polar interference canceller (ACPIC); and the second, a mutual information-based network is trained to minimize the mutual information between its two outputs. This results in one output containing the dependent component of the data, and the other output containing the independent component of the data. This network is implemented in a radial basis function (RBF) network architecture, and is trained using unsupervised learning methodologies. Judging from a subjective inspection of the two radar images produced by these two methods, and from their operation, it is hypothesized that combining these methods should capture the desirable characteristics of both networks.

### 2.1. ACPIC Network

First, we investigate an adaptive signal processing methodology that can account for the non-stationary nature of the clutter process. The ACPIC is a linear network, capable of adapting to the cross-polar variations in the clutter environment.

The clutter statistics vary depending on whether it is sea clutter, various forms of land clutter, or strong point targets. For the purposes of this study, two typical areas were identified. One, called here the DOFASCO site (located by permission on the DOFASCO Company property), is in an industrial area. The target reflector is in an area that has many large metal buildings and machinery. The other site, called here the La Salle Park site (located by permission in La Salle Park conservation area), is a more benign area with park land and some residential land.

The signal processor is designed to track the statistical variations in clutter, removing that portion of the HH-pol return that is correlated with the HV-pol return. It operates by processing along each radar range sweep for each azimuth sampling of the scanning antenna, adapting in real-time to the changing polarization conditions of the radar clutter. The processing is independent from sweep to sweep, and therefore does not take advantage of correlations that may exist in the azimuthal direction. The result is an enhanced cross-polar target response. However, to be successful, the adaptive canceller must be tuned to the nature of the reflector, and must operate in a robust fashion.

**2.1.1. Principle of Adaptive Interference Cancellation.** A schematic diagram of the single-tap adaptive interference canceller is shown in Figure 3. The input is the sampled range series (or equivalently, time series) for a particular azimuth look direction. The canceller processes the HH-pol and HV-pol radar returns along the range dimension, as they are received, on a sweep-by-sweep basis, removing a weighted version of the HH-pol return (interference signal) from the HV-pol return (desired signal).

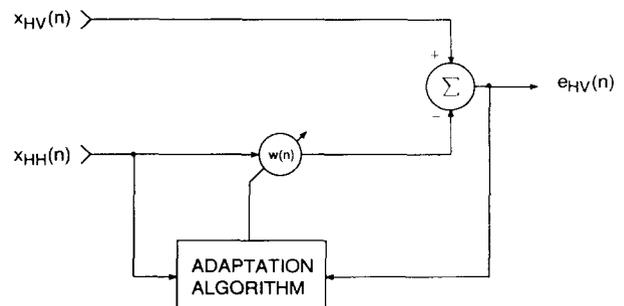


FIGURE 3. Discrete-time cross-polar interference canceller.

Various adaptive algorithms have been developed to adjust the tap weight in some kind of optimal fashion; however, we only consider the least-mean squares (LMS) adaptive algorithm here. For a detailed discussion of the operation of adaptive filtering algorithms and noise cancellation, the reader is referred to Widrow (1975), Widrow and Stearns, (1985) and Haykin (1991).

From Wiener filter theory, it is known that the optimum tap-weight solution under conditions of stationarity and a zero-mean Gaussian process is

$$w_{\text{opt}} = \frac{R_{x_{\text{HH}}, x_{\text{HV}}}(0)}{R_{x_{\text{HH}}, x_{\text{HH}}}(0)}, \quad (1)$$

where  $R_{x_{\text{HH}}, x_{\text{HV}}}(0)$  is the zero-lag cross-correlation between the two channels, and  $R_{x_{\text{HH}}, x_{\text{HH}}}(0)$  is the zero-lag autocorrelation of the HH-pol channel. When the statistics of the time series are stationary, the adaptive algorithm solution for the weight will approach this optimum value.

**2.1.2. The LMS Algorithm.** The desired response in the following equations is set to be the HV-pol range sweep, and the disturbance to be the HH-pol range sweep, as shown by the block diagram in Figure 3. The LMS update equations for the one-tap weight case are therefore

$$\hat{e}_{\text{HV}}(n) = x_{\text{HV}}(n) - \hat{w}(n) x_{\text{HH}}(n), \quad n = 0, \dots, N-1, \quad (2)$$

$$\hat{w}(n+1) = \hat{w}(n) + \mu_{\text{LMS}} x_{\text{HH}}(n) \hat{e}_{\text{HV}}(n), \quad w(0) = 0, \quad (3)$$

where  $\hat{e}_{\text{HV}}(n)$  is the estimated *a posteriori* error at time  $n$ ,  $x_{\text{HV}}(n)$  is the signal containing the desired response,  $x_{\text{HH}}(n)$  is the interference that is correlated to  $x_{\text{HV}}(n)$ ,  $\hat{w}(n)$  is the estimated tap weight,  $\mu_{\text{LMS}}$  is the step-size parameter, and  $N$  is the total number of samples. The weight update algorithm is stable for  $\mu_{\text{LMS}}$  in the range

$$0 < \mu_{\text{LMS}} < \frac{2}{\sigma_{x_{\text{HH}}}^2}, \quad (4)$$

where  $\sigma_{x_{\text{HH}}}^2$  is the variance of the input  $x_{\text{HH}}(n)$ .

**2.1.3. Transient Response.** In the LMS algorithm, the step-size parameter  $\mu_{\text{LMS}}$  determines the adaptive interference canceller transient performance in a non-stationary environment. By studying the impulse response behavior of the LMS algorithm, a reasonable choice for  $\mu_{\text{LMS}}$  can be made. Assuming that  $x_{\text{HH}}(n) = a$ , a constant, for  $n = 0, \dots, N-1$ , it can be shown (Ukrainec, 1994) that

$$\tau = -\frac{T}{\ln(1 - a^2 \mu_{\text{LMS}})}, \quad (5)$$

where  $\tau$  is the time constant for a decaying exponential, and  $T$  is the sample period.

Since the twist-grid reflector can be simplistically interpreted as causing a step-change in the cross-polar radar sweep, we can use this result to choose a reasonable  $\tau$  for the adaptive cross-polar clutter canceller. The pulse width of the radar transmitter fundamentally determines the length, or response duration of a point target. A good choice for  $\tau$  would be to allow the speed of adaptation to be as high as possible, but long enough so that the reflector response is not affected. Choosing a time constant that is too fast would cause the filter to start adapting to the reflector, and degrade the target-to-clutter ratio on the output. Choosing a time constant that is too long would impede the ability of the filter to adapt to changing clutter conditions and allow more clutter than necessary in the output.

## 2.2. Mutual Information Network

Gilbert stated that, "Information will be a measure of time or cost of a sort which is of particular use to the engineer in his role of designer of an experiment" (Gilbert, 1958). Here, the statistical measures of information theory are used as cost functions in the unsupervised learning of neural networks. The desire is to process the HH-pol and HV-pol signals jointly such that the mutual information between the outputs is minimized, under the constraint that the output variance stays equal to that of the input. The expectation is that the cross-polar reflector target response energy should be maximized in one of the outputs.

**2.2.1. Information-Theoretic Principles.** A few definitions are presented first. The differential entropy of a continuous random scalar variable  $X$  may be written as (Cover & Thomas, 1991)

$$H(X) = \int_S f(x) \log \left( \frac{1}{f(x)} \right) dx = - \int_S f(x) \log f(x) dx, \quad (6)$$

where  $f(x)$  is the probability density function (p.d.f) and  $S$  is defined as the support set, the set of values where  $f(x) > 0$ . The relative entropy, or as it is sometimes called, the Kullback-Leibler distance, is defined as

$$D(f_1 \| f_2) = \int_{S_1} f_1(x) \log \left( \frac{f_1(x)}{f_2(x)} \right) dx, \quad (7)$$

where the support set  $S_1$  of  $f_1(x)$  must contain the support set of  $f_2(x)$  for the measure to be finite. The measure can be thought of as an oriented measure of distance between two probability density functions. A special case exists when the relative entropy between the joint probability density function  $f_{XY}(x, y)$  and the product of its marginal respective probability density functions  $f_X(x)$  and  $f_Y(y)$  is considered. The result is the mutual information between two random variables, that is,

$$I(X; Y) = D(f_{XY}(x, y) \| f_X(x)f_Y(y)) \\ = \iint f_{XY}(x, y) \log \left( \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} \right) dx dy. \quad (8)$$

Equivalently, in terms of differential entropies,

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X), \quad (9)$$

where  $H(X)$  is the differential entropy of  $X$ , and  $H(X|Y)$  is the conditional entropy of  $X$  given  $y$ . The entropies  $H(Y)$  and  $H(Y|X)$  are similarly defined. In communications theory the mutual information is often used to measure the information capacity between the input and output of a noisy, band-limited channel.

The specific case of Gaussian p.d.f. is now considered. The differential entropy of a zero-mean, multivariate Gaussian distribution is found to be

$$H_G(X_1, X_2, \dots, X_d) = H_G(\mathbf{X}) = \frac{1}{2} \log(2\pi e)^d |\mathbf{R}|, \quad (10)$$

where  $\mathbf{R} = E(\mathbf{X}\mathbf{X}^T)$  is the autocorrelation matrix of dimension  $d \times d$ , and  $|\mathbf{R}|$  is the determinant of  $\mathbf{R}$ . This also gives the upper bound on the differential entropy of continuous variables in that

$$H(\mathbf{X}) \leq H_G(\mathbf{X}), \quad (11)$$

for a zero-mean  $\mathbf{X}$ , given the same autocorrelation matrix (Cover and Thomas, 1991). Assuming that the joint distribution is a bivariate Gaussian p.d.f., the mutual information is found to be equal to (Kullback, 1968)

$$I(X; Y) = -\frac{1}{2} \log(1 - \rho_{x,y}^2), \quad (12)$$

where  $\rho_{x,y}$  is the correlation coefficient between  $X$  and  $Y$ . It is defined in terms of the pertinent correlation functions  $R_{x,x}(0)$ ,  $R_{y,y}(0)$  and  $R_{xy}(0)$  for zero lag as follows:

$$\rho_{x,y} = \frac{R_{x,y}(0)}{\sqrt{R_{x,x}(0)R_{y,y}(0)}}. \quad (13)$$

In the Gaussian case, minimizing the mutual information is equivalent to driving the outputs to being statistically uncorrelated.

Some properties that are useful for computing information-related quantities in network architectures are now presented. The mutual information is equal to

$$I(X; Y) = 0, \quad (14)$$

if and only if  $X$  is independent of  $Y$ . The entropy of a random variable remains unchanged after translation,  $Y = X + k$ , so that

$$H(Y) = H(X + k) = H(X). \quad (15)$$

Under the linear matrix transformation  $\mathbf{Y} = \mathbf{W}\mathbf{X}$ , the entropy change is such that

$$H(\mathbf{Y}) = H(\mathbf{W}\mathbf{X}) = H(\mathbf{X}) + \log(|\mathbf{W}|), \quad (16)$$

where  $|\mathbf{W}|$  is the determinant of matrix  $\mathbf{W}$ . For any continuous matrix transformation of a multivariate random variable  $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ , the entropy is equal to (Walker & Akers, 1992)

$$H(\mathbf{Y}) = H(\mathbf{X}) - E[\log(|J_F(\mathbf{X})|)], \quad (17)$$

where  $J_F(\mathbf{X})$  is the Jacobian of the transformation.

**2.2.2. Unsupervised Learning of Neural Networks.** In the case of a linear network  $\mathbf{Y} = \mathbf{W}\mathbf{X}$ , finding the weight matrix that minimizes the mutual information at the outputs  $\mathbf{Y}$  is equivalent to finding an orthogonal transformation to decorrelate the output. This solution for a linear network is easily obtained and implemented, and is optimum in the case when a Gaussian p.d.f. describes the data. To find the solution for a nonlinear network is much more difficult. However, the advantage is that the nonlinear network has more degrees of freedom to find nonlinear mappings that satisfy the optimization conditions.

**2.2.3. Radial Basis Function Neural Network.** Radial basis function (RBF) neural networks have been successfully used by several researchers to solve difficult problems in signal processing (Broomhead & Lowe, 1988; Casdagli, 1989; Jones et al., 1989; Lowe, 1989; Lowe & Webb, 1989; Moody & Darken, 1989b; Saha & Keeler, 1990; Haykin & Ukrainec, 1993). The RBF network architecture used here is presented in Figure 4. The inputs connect to a nonlinear hidden layer. The hidden layer, in turn, is connected to the output by a linear layer. The hidden layer non-linear functions are of a type called radially

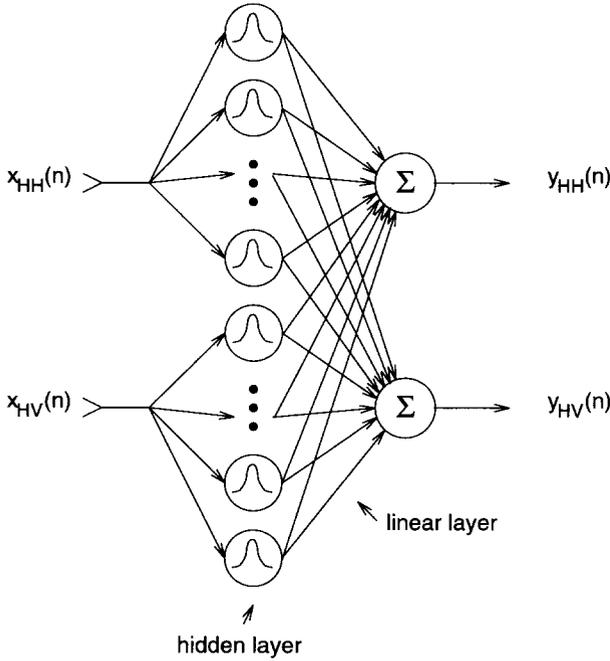


FIGURE 4. The mutual information (RBF) neural network architecture.

symmetric basis functions. These functions can be chosen to be of many possible forms. Here only the Gaussian form will be used. The non-linear functionals in the hidden layer are given by

$$\phi_j(\mathbf{x}) = e^{-1/2(\mathbf{x}-\mathbf{c}_j)^T \mathbf{S}_j (\mathbf{x}-\mathbf{c}_j)} = e^{-1/2\|\mathbf{x}-\mathbf{c}_j\|_{\mathbf{M}}^2} \quad (18)$$

where  $\mathbf{x} \stackrel{\text{def}}{=} (x_{\text{HH}}, x_{\text{HV}})^T$ ,  $\phi_j(\mathbf{x})$  is the  $j$ th radial basis function evaluated at the input vector  $\mathbf{x}$ ,  $\mathbf{c}_j$  is the  $j$ th RBF center, and  $\mathbf{S}_j$  is the  $j$ th multidimensional width, or spread. The term following the factor  $-1/2$  in the exponent is known as the Mahalanobis metric, or weighted Euclidean metric (hence the subscript  $\mathbf{M}$ ). The functional form of the RBF network is therefore given by

$$y_i = \sum_{j=1}^{N_h} w_{ji} \phi_j(\mathbf{x}) + w_{0i} \quad (19)$$

where  $y_i$  is the  $i$ th output function evaluated for the input vector  $\mathbf{x}$ . The output can therefore be written as the vector  $\mathbf{y} \stackrel{\text{def}}{=} (y_{\text{HH}}, y_{\text{HV}})^T$ . The parameter  $w_{ji}$  is the linear output weight connecting the  $j$ th hidden unit with the  $i$ th output. The hidden layer has a total of  $N_h$  radial basis functions. The weight  $w_0$  is the bias term. Given a set of input and output data vectors,  $\{\mathbf{x}(n), \mathbf{y}(n) | n = 0, \dots, N-1\}$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}(0)^T \\ \mathbf{x}(1)^T \\ \vdots \\ \mathbf{x}(N-1)^T \end{pmatrix},$$

$$\Phi = \begin{pmatrix} 1 & \phi_1(\mathbf{x}(0)) & \phi_2(\mathbf{x}(0)) & \dots & \phi_{N_h}(\mathbf{x}(0)) \\ 1 & \phi_1(\mathbf{x}(1)) & \phi_2(\mathbf{x}(1)) & \dots & \phi_{N_h}(\mathbf{x}(1)) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \phi_1(\mathbf{x}(N-1)) & \phi_2(\mathbf{x}(N-1)) & \dots & \phi_{N_h}(\mathbf{x}(N-1)) \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} w_{1,\text{HH}} & w_{1,\text{HV}} \\ w_{2,\text{HH}} & w_{2,\text{HV}} \\ \vdots & \vdots \\ w_{N_h,\text{HH}} & w_{N_h,\text{HV}} \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \mathbf{y}(0)^T \\ \mathbf{y}(1)^T \\ \vdots \\ \mathbf{y}(N-1)^T \end{pmatrix}.$$

Rewriting eqn 19 in matrix form,

$$\mathbf{Y} = \Phi(\mathbf{X})\mathbf{W}. \quad (20)$$

**2.2.4. Network Design Strategies.** The determination of the hidden layer parameters is a challenging task. Various procedures have been experimented with to learn the centers and widths (or spreads) of the hidden layer of RBF units.

Several non-adaptive strategies have been used to determine the RBF centers and spreads. The most straightforward choice for the location of the centers is to place them on an evenly spaced grid, spanning the input space. Unfortunately, a very large number of RBF units may be needed, since the number of units required grows exponentially with the dimensionality of the input space. As the dimensionality of the input grows, most of the input space becomes devoid of samples, and therefore a large percentage of the centers lie in an area where there are no data. Another more effective choice for the RBF centers is to set the centres equal to a random sampling of the input data. This strategy ensures that centers are located only in areas where there are data. It has been shown that as long as a sufficiently large number of centers are used, good prediction performance on a chaotic time series is achieved (Broomhead & Lowe, 1988). In either case, the RBF spreads are chosen using some heuristic method.

Supervised adaptation of the RBF centers, spreads, and output weights using optimization techniques have been used (Lowe, 1989; Moody & Darken, 1989b). This strategy can give a minimal RBF network configuration. Some of the disadvantages with using optimization techniques are considerable computational cost, poor scaling of learning as network complexity grows, and the presence of sub-optimal local minimum solutions. Low (1989) points out that the same final error performance can

be achieved with a network with a larger number of non-adaptive centers, with the same generalization performance.

The investigations of hybridized unsupervised/supervised training schemes have shown promise (Moody & Darken, 1989a, b; Nowlan, 1990b; Saha & Keeler, 1990). Some of the advantages are computational efficiency, good scaling of learning as network size grows, and faster convergence. The hybrid procedure consists of two stages of learning: an unsupervised clustering algorithm is used to determine the parameters of the hidden layer, followed by a supervised least-squares solution to the linear output weights. Moody and Darken (1989a, b) suggest the use of the  $k$ -means clustering algorithm to find suitable positions for the centers. As a result, a smaller number of RBF units are required. After clustering, heuristic methods are used to choose the spreads of the RBF units. The  $k$ -means algorithm is an approximate version of the maximum likelihood (ML) solution for determining the location of the means of a mixture density of component densities. The expectation maximization (EM) algorithm can be used to find the exact ML solution for the means and covariances of the density. A comparison of these two learning strategies on a classification problem was done by Nowlan (1990a, b), with the EM algorithm shown to be superior. Saha and Keeler also studied the use of the  $k$ -means clustering for the adjustment of RBF centers, and suggested an approach which they termed as *extended metric clustering* (Saha & Keeler, 1990), where clustering is done in an augmented input-output space. Once learning is complete, the cluster locations are projected back onto the input space, and used as the RBF unit centers. In recent studies by Ukrainec and Haykin (1991b) and Haykin and Ukrainec (1993) the hybrid training was applied successfully to signal processing problems. It was shown that a combination of EM training and extended metric clustering, named EMX clustering, gave the best overall performance in the example prediction and cancellation signal processing problems.

The clustering concept is used here to learn the hidden layer RBF parameters. The EM learning is performed in the two-dimensional input space  $[x_{HH}, x_{HV}]$ . After the learning process is complete, the center and spread parameters are projected down onto the  $x_{HH}$  and  $x_{HV}$  axes. This results in a one-dimensional RBF hidden layer. The linear layer then combines the localized representations to provide the desired mapping.

**2.2.5. Unsupervised Clustering Algorithm.** The expectation maximization (EM) algorithm is a general approach for iteratively computing the maximum-likelihood (ML) estimates of parameters of mixture

density problems. This algorithm has broad application for ML estimation from incomplete data (Dempster et al., 1977), mixture estimation (Redner & Walker, 1984), and unsupervised clustering (Duda & Hart, 1973). Here we will concentrate on the application of the EM algorithm for to learn the RBF centers and spreads through unsupervised clustering.

A mixture distribution of Gaussian component densities is given by (Duda & Hart, 1973)

$$p(\mathbf{x}(n) | \boldsymbol{\theta}) = \sum_{j=1}^{N_h} P(j) p(\mathbf{x}(n) | j, \boldsymbol{\theta}_j), \quad (21)$$

$$p(\mathbf{x}(n) | j, \boldsymbol{\theta}_j) = \frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} e^{-1/2(\mathbf{x}(n) - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}(n) - \boldsymbol{\mu}_j)}, \quad (22)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{N_h})$  is the vector of parameters (means and covariances) to be estimated,  $d$  is the dimensionality of the multivariate Gaussian density,  $\boldsymbol{\mu}_j$  is the mean, and  $\Sigma_j$  is the covariance. The *a priori* probabilities  $P(j)$  are called the mixing parameters.

The EM algorithm iteratively converges to a maximum of the likelihood function, yielding an estimate of the parameters of the component densities, as well as the mixing parameters  $P(j)$ . Although the algorithm is guaranteed to converge, there is no guarantee that it will converge to a global maximum. The update equations are given as follows (Duda & Hart, 1973):

$$\hat{P}(i) = \frac{1}{N} \sum_{n=0}^{N-1} \hat{P}(i | \mathbf{x}(n), \hat{\boldsymbol{\theta}}), \quad (23)$$

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_{n=0}^{N-1} \hat{P}(i | \mathbf{x}(n), \hat{\boldsymbol{\theta}}) \mathbf{x}(n)}{\sum_{n=0}^{N-1} \hat{P}(i | \mathbf{x}(n), \hat{\boldsymbol{\theta}})}, \quad (24)$$

$$\hat{\Sigma}_i = \frac{\sum_{k=0}^{N-1} \hat{P}(i | \mathbf{x}(n), \hat{\boldsymbol{\theta}}) (\mathbf{x}(n) - \hat{\boldsymbol{\mu}}_i) (\mathbf{x}(n) - \hat{\boldsymbol{\mu}}_i)^T}{\sum_{k=0}^{N-1} \hat{P}(i | \mathbf{x}(n), \hat{\boldsymbol{\theta}})}, \quad (25)$$

$$\hat{P}(i | \mathbf{x}(n), \hat{\boldsymbol{\theta}}) = \frac{\hat{P}(i) p(\mathbf{x}(n) | i, \hat{\boldsymbol{\theta}}_i)}{\sum_{j=1}^{N_h} \hat{P}(j) p(\mathbf{x}(n) | j, \hat{\boldsymbol{\theta}}_j)}. \quad (26)$$

An additional step was added to these standard update equations to ensure that the algorithm learned localized representations. The step

$$\text{if } \hat{\Sigma}_i^{jj} > \Sigma_i^{jj} \text{ set the diagonal element } \hat{\Sigma}_i^{jj} = \Sigma_i^{jj} \quad (27)$$

ensures that the spread is limited to a maximum given by  $\Sigma_1$ . Without this extra step, the algorithm may converge to solutions where one or more of the components of the mixture distribution span a large area of the sample space, overlapping other components. Although these are valid solutions, they are not desirable when the parameters are to be transferred to a RBF network that presupposes localized representations.

The equations describe a batch processing algorithm, where all the data are used for each iteration. An on-line version of the EM algorithm was suggested by Nowlan (1990a), where the density parameters can be continuously updated as new data become available. This may have application if the input space is slowly changing, and the representations need to be fine tuned. For the purposes of this study, the batch update algorithm is used exclusively.

It is evident that the RBF given in eqn (18) and the Gaussian component density in eqn (22) have almost the same form. It is hypothesized that the individually learned  $\hat{\mu}_j$  of the component densities should give a good location for the centers of the RBF units. Likewise, the estimated covariances can give the required spread of the RBF units.

The hybrid learning procedure is therefore given as follows:

1. choose the number of RBF units (and hence the number of component densities);
2. initialize the density parameters;
3. iterate the EM algorithm until convergence;
4. transplant the estimated parameters of the component densities into the RBF units, so that  $\mathbf{c}_j \leftarrow \hat{\mu}_j$ ,  $\mathbf{S}_j \leftarrow \alpha \hat{\Sigma}_j^{-1}$ , where  $0 < \alpha \leq 1$  (discussed below);
5. forward-propagate the input data to the output of the RBF hidden layer;
6. compute solution to linear weight layer.

The factor  $\alpha$  is introduced to increase the spreads in order to smooth the interpolation performance of the network. As shown by Ukrainec and Haykin (1991b), the performance increases as  $\alpha$  is decreased, although localization of response decreases.

**2.2.6. Minimum Mutual Information Learning.** Once the basic functions are trained using the unsupervised method described in the previous section, they are fixed and only the output layer of weights would need to be learned. The objective is to minimize the mutual information between the outputs, while keeping the output entropy fixed. The cost function is therefore

$$C(\mathbf{W}) = I(Y_{\text{HH}}; Y_{\text{HV}}) + \lambda |H(\mathbf{Y}) - H(\mathbf{X})|, \quad (28)$$

where  $Y_{\text{HH}}$  and  $Y_{\text{HV}}$  are random variables whose sample values are denoted by  $y_{\text{HH}}(n)$  and  $y_{\text{HV}}(n)$ , respectively. The constraint term explicitly ensures that the output entropy is constant, and equal to the total input entropy. The advantage in using a nonlinear network over the linear network is in the increased degrees of freedom in the mapping. The entropy of the output of the RBF network is found by combining the results of (16) and (17) so that

$$H(\mathbf{Y}) = H(\mathbf{X}) - E[\log(|J_{\mathbf{R}}(\mathbf{X})|)] + \log(|\mathbf{W}|), \quad (29)$$

where  $J_{\mathbf{R}}(\mathbf{X})$  is the Jacobian of the hidden layer transformation. The RBF network has an advantage here over other neural networks, such as the multilayer perceptron, in the sense that it has a set of fixed basis functions, or in other words, a non-adaptive hidden layer.

The numerical estimation of  $I(Y_{\text{HH}}; Y_{\text{HV}})$  requires either a *priori* assumed distribution model or a model-free estimate. A model-free estimate is possible but is computationally expensive [order  $N \log N$  (Fraser & Swinney, 1986)], and must be recomputed on every iteration of the optimization routine used to minimize the cost function. Previous researchers Becker and Hinton (1989) and Zemel and Hinton (1991) have used the Gaussian distribution model assumption when attempting to estimate mutual information. Preliminary studies done by Ukrainec and Haykin (1991a, 1992) have shown that it is possible to use the Gaussian-based mutual information measure given in (12) as an estimate of the mutual information. The advantage is that it is easy to compute. However, since the distribution is known to be non-Gaussian, it is also inaccurate. At best, this is an upper bound on the mutual information. The result in (11) indicates that for a given autocorrelation function, the differential entropy is upper bounded by the Gaussian distribution.

Using the Gaussian distribution assumptions of the (22) and (12) functions, the cost function of (28) is restated as

$$C(\mathbf{W}) = \frac{1}{2} \log(1 - \hat{\rho}_{y_{\text{HH}}, y_{\text{HV}}}^2) + \lambda \left( |\hat{\mathbf{R}}_{y_{\text{HH}}, y_{\text{HV}}}| - |\hat{\mathbf{R}}_{x_{\text{HH}}, x_{\text{HV}}}| + J_{\text{misc}} \right). \quad (30)$$

Under the Gaussian assumptions the entropy is proportional to the determinant of the autocorrelation function. The autocorrelation estimates of the input  $\mathbf{X}$  are given by  $\hat{\mathbf{R}}_{x_{\text{HH}}, x_{\text{HV}}} = \mathbf{X}\mathbf{X}^T$ , and similarly for the output  $\mathbf{Y}$ . These estimates are also used to compute the correlation coefficient  $\hat{\rho}_{y_{\text{HH}}, y_{\text{HV}}}^2$ . Through experimentation it was found that putting additional constraints on the output mean, variance, and skew

improved convergence and helped to avoid undesirable local minima. The term  $J_{misc}$  contains the sum of the additional constraints, such that

$$J_{misc} = |\hat{\mu}_{yHH}| + |\hat{\mu}_{yHV}| + |\hat{\sigma}_{yHH} - \hat{\sigma}_{yHV}| + |\hat{\gamma}_{yHH}| + |\hat{\gamma}_{yHV}|. \quad (31)$$

The estimate of the means

$$|\hat{\mu}_{yHH}| + |\hat{\mu}_{yHV}| \quad (32)$$

is intended to ensure a zero-mean output. An equal output variance term,

$$|\hat{\sigma}_{yHH} - \hat{\sigma}_{yHV}| \quad (33)$$

is introduced to encourage a circularly symmetric distribution. Finally, the third-order moment, or skew, is constrained; the additional penalty term is

$$|\hat{\gamma}_{yHH}| + |\hat{\gamma}_{yHV}|. \quad (34)$$

The quantity is normalized for the Gaussian distribution so that zero skew is equal to the skew of a Gaussian distribution, which has maximum entropy. In summary, constraints are introduced on the moments of the output so as to force the output to approximate a Gaussian p.d.f. A constrained optimization routine is used to minimize the cost function.

### 2.3. Modular Network Design Strategy

Figure 5 is a diagram of the modular network. After the data are processed by the ACPIC and RBF network, each resulting output is normalized with

respect to its own estimated histogram. This is done to make the data ranges approximately the same. The normalization is such that

$$\sum_{i=1}^{N_{bins}} \hat{n}(i) = N_{total}, \quad (35)$$

$$\sum_{i=1}^{N_b} \hat{n}(i)/N_{total} \simeq 0.02, \quad (36)$$

$$\sum_{i=N_w}^{N_{bins}} \hat{n}(i)/N_{total} \simeq 0.02, \quad (37)$$

where  $\hat{n}(i)$  is the estimated histogram,  $N_{bins}$  is the total number of histogram bins,  $N_b$  is the lower 2% bin (black), and  $N_w$  is the higher 2% bin (white) bin. Based on the  $N_b$  and the  $N_w$  values, the data are scaled to the [0,1] interval. The normalized data are then averaged, and the result processed by the CFAR algorithm. The un-informed choice for combining the normalized results from the individual networks is to sum them equally.

### 2.4. CFAR Processing and Detection

A constant false alarm rate (CFAR) processor is commonly used in radar systems. It prevents saturation of the detector due to increases in clutter or noise by adapting the detection threshold in step with the changing clutter or noise conditions. Likewise, in the case that the clutter or noise decreases, it lowers the threshold thereby increasing the detectability of weaker targets which otherwise would be missed. The operation of CFAR systems has been widely studied (Goldstein, 1973; Skolnik, 1980; Minkler & Minkler, 1990; Nitzberg, 1992). The basic idea is to estimate a sufficient statistic of the

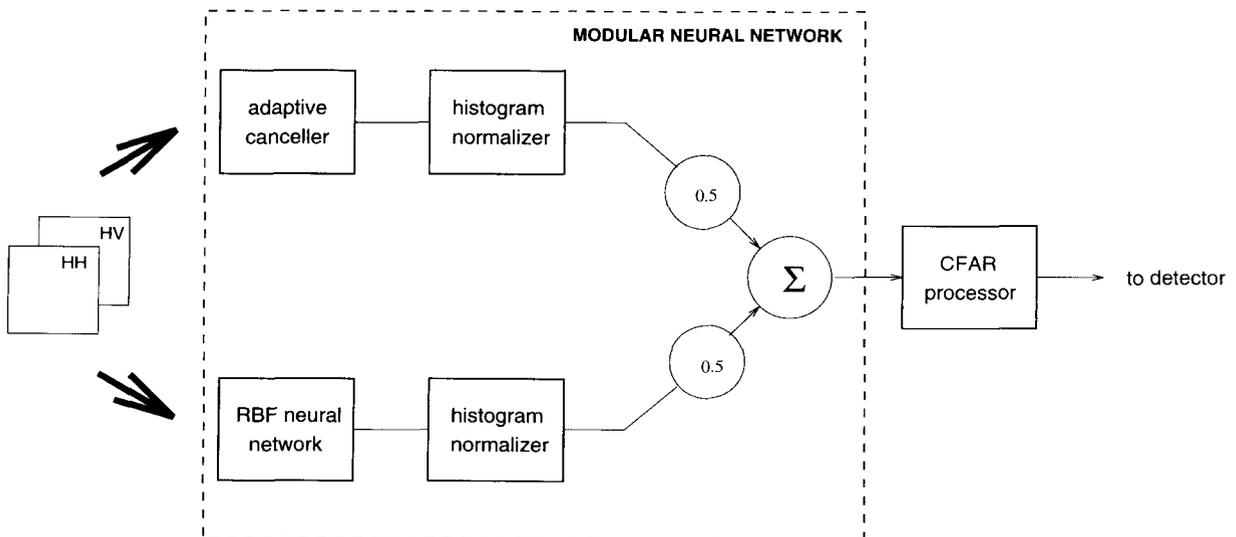


FIGURE 5. Modular neural network

clutter, and to use this estimate to control a detection threshold so that a constant false alarm rate results.

The CFAR processor is now derived. For the purposes of this study, the clutter process is assumed to be Rayleigh distributed. The Rayleigh probability density function is given by

$$p(x) = \frac{2x}{\sigma^2} \exp\left(-\frac{x^2}{\sigma^2}\right), \quad x > 0, \quad (38)$$

where  $x$  is the voltage amplitude, and  $\sigma^2$  is the variance. The ideal logarithmic receiver is described by the function

$$y = a \log(bx), \quad (39)$$

where  $a$  and  $b$  are scale factors. Under these assumptions, Croney (1956) shows that the theoretical variance of the output is

$$\sigma_y^2 = \frac{a^2 \pi^2}{24}, \quad (40)$$

which is independent of the variance of the input signal. The logarithmic receiver therefore has a CFAR-like operation, in the sense that clutter described by a Rayleigh distribution results in a constant variance in the output. The mean level of the clutter, however, is a function of the input power, and can be removed either by using a high-pass filter, or by using averaging to estimate the mean level and subtracting it. The cell-averaging CFAR (CA-CFAR) model assumes that the clutter in the neighbourhood of a cell under test is a stationary statistical process, with independent samples, and is representative of the clutter in the test cell. In practice, these statistical assumptions are often not consistent with the operating environment, resulting in a loss in performance.

To implement the desired cell-averaging operation, two-dimensional target masks and clutter masks are used. Two masks are defined: a  $23 \times 3$  pixel mask for the target, and a  $69 \times 9$  pixel mask for the surrounding clutter, as shown in Figure 6. The target mask size was chosen to reflect the approximate size of the reflector target, and the clutter mask size was chosen to be large enough so as to provide a fair estimate of the clutter power, while at the same time being small enough so as to be in a stationary region of clutter. This is a trade-off in CA-CFAR processing. The two masks are co-located, centered on the same pixel, the clutter mask having zero response where the target mask coincides with it. The pixels in the mask are all equally weighted, although this need not be the case. For example, if there is *a priori* information about the clutter p.d.f. character-

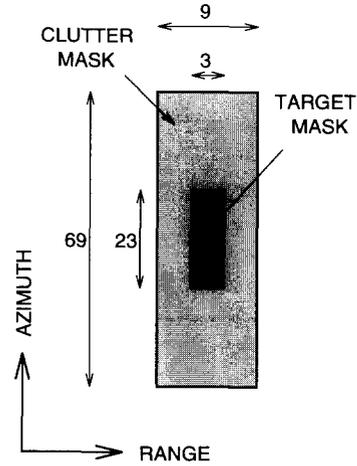


FIGURE 6. Target and clutter masks used in CFAR processing.

istics, or location of the target within the clutter, a particular weighting configuration could have a better performance. No such assumptions are made here. The masks are convolved with the image; the clutter-convolved result is subtracted from the target-convolved result. For a particular location in the image, the function can be expressed as

$$y = \frac{1}{N_t} \sum_i^{N_t} t_i - \frac{1}{N_c} \sum_i^{N_c} c_i, \quad (41)$$

where  $y$  is the output,  $t_i$  is the set of target pixels,  $N_t$  is the number of target pixels in the mask,  $c_i$  are the set of surrounding clutter pixels, and  $N_c$  is the number of clutter pixel in the mask. For the purposes of this study, no *a priori* information is used about the clutter orientation or distribution. The resultant image formed by the CA-CFAR processor possesses improved target visibility.

The detector stage after the CA-CFAR processing performs a thresholding function, which produces a binary result such that

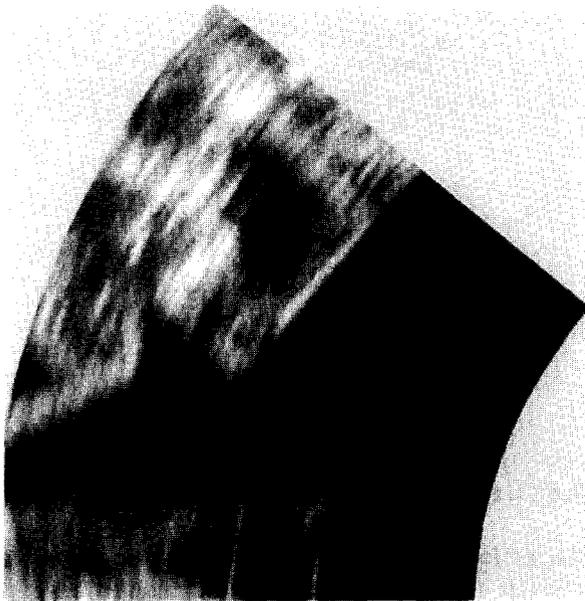
$$\begin{aligned} y > \beta, & \quad b_i = 1, \text{ target present} \\ y \leq \beta, & \quad b_i = 0, \text{ target absent,} \end{aligned} \quad (42)$$

for the  $i$ th pixel in the image. The parameter  $\beta$  controls the false alarm rate.

### 3. PERFORMANCE EVALUATIONS

#### 3.1. Traditional Processing

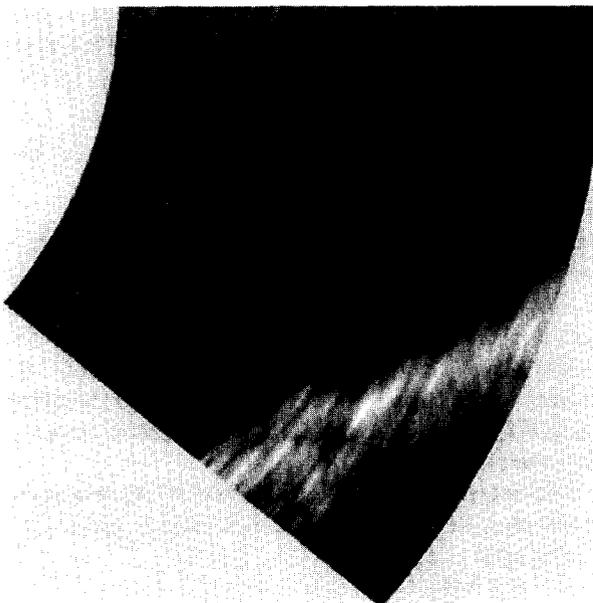
In the case of traditional processing techniques, only HV-pol image data are used as input to enhance the reflector target visibility. This is done by sending the HV-pol image to a CFAR processor. The traditional processing results provide a benchmark by which to



DOFASCO



DOFASCO



La Salle Park

FIGURE 7. The HH-pol sub-images of interest.



La Salle Park

FIGURE 8. The HV-pol sub-images with the locations of the reflectors marked.

gauge the performance of the more sophisticated joint HH-pol and HV-pol processing techniques.

In total, 28 scans (equivalent to 1 min worth) of radar data were recovered from a recording dated 23 November, 1987, and preprocessed. This same data set was used throughout the study. The example sub-images shown in Figures 7 and 8 are the basis for the visual and numerical comparison of the various joint signal processing techniques. Table 1 contains the position of the reflectors, the calculated radar cross-section (RCS), and the radar cell sizes. It is assumed that a ship's tracking algorithm window would be initialized to an area of similar (or smaller) size, in

order to acquire and track the reflector target position as the ship navigates the channel. The sub-images chosen are  $800 \times 420$  samples (azimuth samples  $\times$  range samples). This corresponds to a physical area of approximately  $2100 \times 2100 = 4,410,000\text{m}^2$ . Although the data are presented in scan-converted form (cartesian coordinates), the processing is done in the B-scan domain (range-azimuth coordinates).

The logarithmic receiver is characterized by injecting a test signal from a pulsed X-band; signal generator. In this way that instantaneous video

**TABLE 1**  
Experimental Parameters of the Radar Reflectors

	DOFASCO	La Salle Park
Range to reflector	2300 m	3100 m
Number of reflectors	2	1
RCS of reflector	4846 m <sup>2</sup>	458 m <sup>2</sup>
Reflector cell size (range)	7.5 m	7.5 m
Reflector cell size (azimuth)	2.7 m	4.0 m

voltage can be related to the input power, which is expressed in dBm ( $1 \text{ dBm} = 10 \log_{10} W + 30$ , where  $W$  denotes the power level in watts). The digitized samples are then calibrated and range normalized using the radar range equation for a point target. The resultant RCS estimate is therefore equal to

$$\sigma_R = P_r \frac{(4\pi)^3 \cdot R^4}{P_t \cdot G_a^2 \cdot \lambda_X^2}, \quad (43)$$

where  $R$  is the range in meters,  $P_t$  is the transmitted power,  $P_r$  is the received power,  $G_a$  is the one-way gain of the antenna, and  $\lambda_X$  is the wavelength of the transmitted signal pulse. The absolute calibration of the system requires the measurement of a calibration sphere. In the absence of this, only a relative measurement is possible. Defining the RCS of the target relative to a calibration sphere,

$$\sigma_r = \frac{\sigma_R}{\sigma_o}, \quad (44)$$

where  $\sigma_o$  is the RCS of the calibration sphere in m<sup>2</sup>, and  $\sigma_r$  is the RCS of the received signal relative to the calibration sphere, in m<sup>2</sup>/m<sup>2</sup>. The radar equation in (43) can be rewritten in logarithmic terms, so that

$$\sigma_r = P_r - P_t + 40 \log_{10}(R) + 30 \log_{10}(4\pi) - 20 \log_{10}(\lambda_X) \quad (45)$$

$$-20 \log_{10}(G_a) - L_{\text{misc}}(\text{dBm} \cdot \text{m}^2/\text{m}^2), \quad (46)$$

where  $L_{\text{misc}}$  contains the correction required for the calibration sphere and miscellaneous losses in the system that haven't been accounted for (e.g., insertion loss of rotary joint and feeds, VSWR mismatch). Since our interest is only in the relative visibility of the target, absolute calibration is unnecessary. Therefore, without loss of generality, it is assumed that  $L_{\text{misc}} = 0 \text{ dB}$  for the purposes of this paper.

The target-to-clutter ratio (TCR) estimate is given by the ratio of target power to clutter power. The TCR is given by

$$\text{TCR} = \hat{\mu}_t - \hat{\mu}_c \text{ dBm}, \quad (47)$$

where  $\hat{\mu}_t$  is the estimated mean clutter response, and

$\hat{\mu}_c$  is the estimated mean clutter response. Unfortunately, the TCR estimate does not provide a good measure of target enhancement or visibility. Various signal processing methods can scale the logarithmic data, resulting in a meaningless estimate of the TCR. For example, simply by scaling the data by a factor of 2, the TCR value would also double, without any true enhancement to the target. To overcome this problem, a dimensionless measure based on the TCR is proposed as a figure of merit by which to judge the various processing methods. The normalized TCR (NTCR) is given as

$$\text{NTCR} = \frac{\hat{\mu}_t - \hat{\mu}_c}{\hat{\sigma}_c}, \quad (48)$$

where  $\hat{\sigma}_c$  is the estimated standard deviation of the clutter power in dBm. If we assume that the p.d.f.s are Gaussian, the NTCR is equivalent to the threshold used to calculate the probability of false alarm ( $P_{\text{FA}}$ ) for this model, where

$$P_{\text{FA}} = Q(\text{NTCR}) = \frac{1}{2} \text{erfc}\left(\frac{\text{NTCR}}{\sqrt{2}}\right), \quad (49)$$

where the  $Q$  function is itself defined by

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy, \quad (50)$$

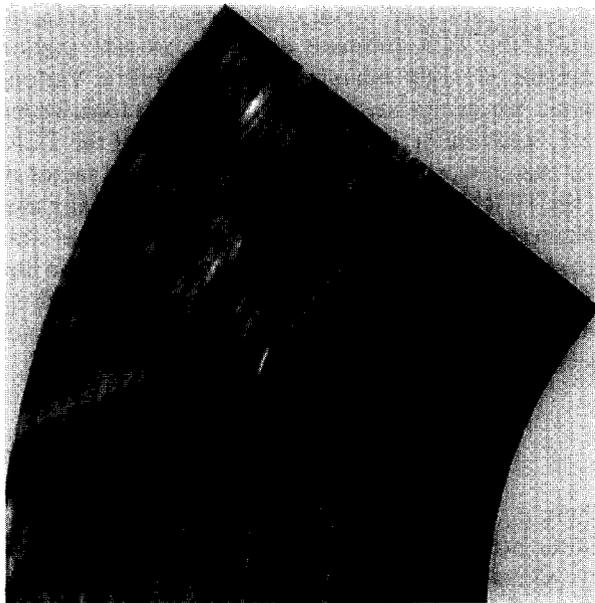
and  $\text{erfc}(\cdot)$  is the complementary error function (Wozencraft & Jacobs, 1965).

The target and clutter estimates are summarized in Table 2 for the HV-pol sub-images. The clutter patches were chosen to provide representative areas of clutter from both sites. The clutter patch within the DOFASCO site is approximately 572 m in azimuth and 100 m in range, and the corresponding La Salle Park clutter patch is approximately 528 m in azimuth by 100 m in range. The clutter response within the patch areas is averaged, and then averaged over all the scans to estimate mean clutter level. The mean of the peak reflector target value is estimated by averaging the peak response within the reflector target cell area over the 28 scans.

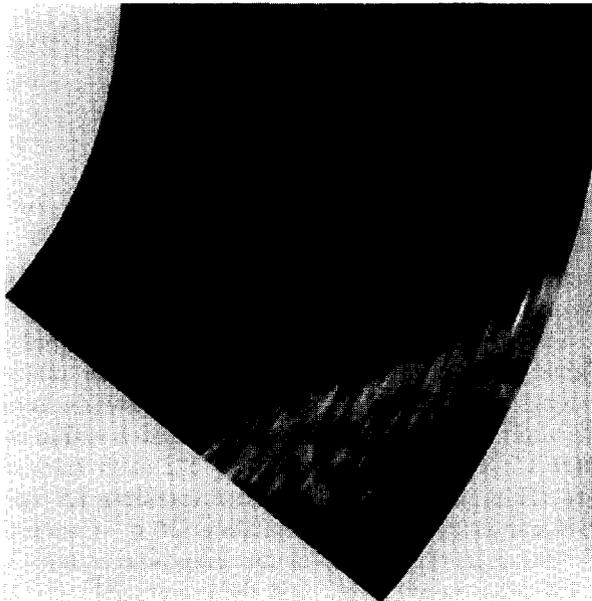
The resultant image formed by the traditional CAFAR processor possesses an enhanced visibility as demonstrated by the example sub-images shown in Figure 9. To improve the contrast of the printed half-

**TABLE 2**  
Estimated Parameters of Unprocessed HV-pol Sub-images

	DOFASCO	La Salle Park
Target ( $\hat{\mu}_t, \hat{\sigma}_t$ )	(16.4, 0.3) dBm	(19.4, 0.6) dBm
Clutter ( $\hat{\mu}_c, \hat{\sigma}_c$ )	(2.2, 1.2) dBm	(-3.2, 3.9) dBm
TCR	14.2 dBm	22.6 dBm
Normalized TCR	12	6



DOFASCO



La Salle Park

FIGURE 9. Sub-images of the CFAR processing, prior to detection.

tone sub-images, the lower 8% of the image values are mapped to black for all CFAR processed images (as estimated from the estimated histogram).

The enhanced target visibility should be reflected in higher NTCR values. Comparing results, the normalized TCR values in Table 3 show an improvement over the corresponding NTCR values for the unprocessed HV-pol images in Table 2.

### 3.2. ACPIC Processing

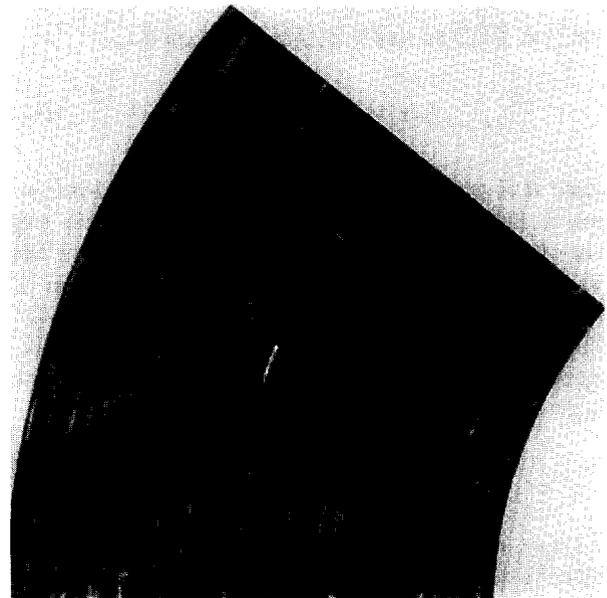
A reasonable setting for the adaptation constant  $\mu$

TABLE 3

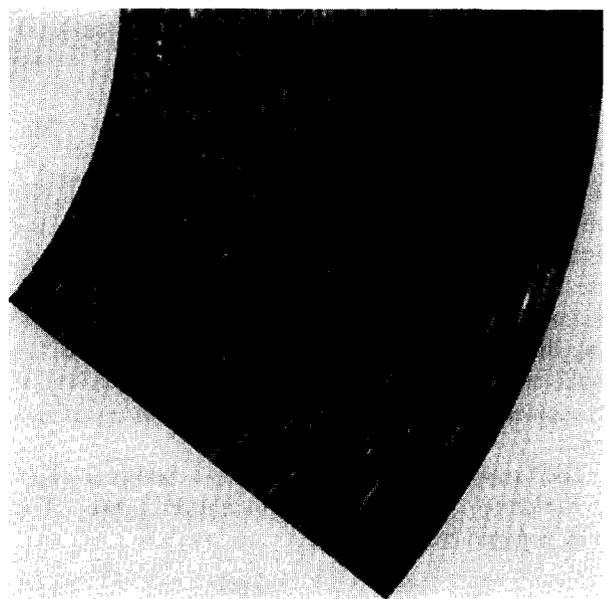
Estimated Parameters for CA-CFAR Processed HV-pol Sub-images

	DOFASCO	La Salle Park
Target ( $\hat{\mu}_t, \sigma_t$ )	(22.9, 0.3) dBm	(24.0, 0.9) dBm
Clutter ( $\hat{\mu}_c, \hat{\sigma}_c$ )	(2.8, 1.2) dBm	(-2.0, 3.7) dBm
TCR	20.1 dBm	26.0 dBm
Normalized TCR	16	7

was experimentally determined to be  $10^{-4}$ . This is confirmed by the transient response analysis. Using (5) and assuming  $a = -30$  dBm, the time constant of



DOFASCO



La Salle Park

FIGURE 10. Sub-images of ACPIC processed data.

the adaptive canceller is found to be approximately  $10.6T$ , where  $T$  is the sample interval. The target response in range is approximately  $2T - -3T$ ; so therefore the time constant meets the requirement of being short enough to adapt quickly to changing clutter conditions, and yet long enough for the target to not be filtered out.

The sub-images shown in Figure 10 are the result of processing with the adaptive interference canceller algorithm, followed by CA-CFAR. As is easily observed from the images, both the DO-FASCO and La Salle Park reflectors show greatly improved visibility. Table 4 summarizes the factor of improvement for the adaptive cross-polar clutter canceller output. The NTCR values exceed that of the CA-CFAR only processed images by nearly 20, indicating that the correlation between HH-pol and HV-pol radar sweeps is high.

A performance limitation of the canceller is generally observable along the edge of the sea and land clutter boundary. It takes a finite period of time (proportional to the time constant of the adaptive algorithm) for the canceller to respond to a different clutter region, in this case, land clutter. During this time an increase in clutter power can appear at the output.

### 3.3. Mutual Information Network Processing

First, a mixture density is learned using the EM algorithm, as described in Section 2.2.3. A  $200 \times 200$  pixel clutter region is used for training. The scatter plot of the combined DOFASCO and La Salle Park input data are shown in Figure 12, with the learned ellipsoidal standard deviations of the components of the mixture of Gaussian densities superimposed. The number of components in the density used to model the clutter is chosen to be seven, with two extra units introduced to represent the reflectors. These extra units were introduced manually in order to have a response from the target as well as the clutter. The manual introduction was necessary since the number of target samples is under-represented in the available data; a statistical learning method gives the target samples little weight, leaving the targets unrepresented. The locations of the centers is approximated from scatter plot of the target samples (see Figure 11), and set equal to (13,18) dBm and (6,14) dBm.

TABLE 4

TCR Estimate for ACPIC/CA-CFAR Processed Sub-images

	DOFASCO	La Salle Park
Target ( $\hat{\mu}_t, \hat{\sigma}_t$ )	(20.8, 2.2) dBm	(25.8, 2.6) dBm
Clutter ( $\hat{\mu}_c, \hat{\sigma}_c$ )	(-0.4, 0.6) dBm	(1.3, 0.8) dBm
TCR	21.2 dBm	24.5 dBm
Normalized TCR	35	29

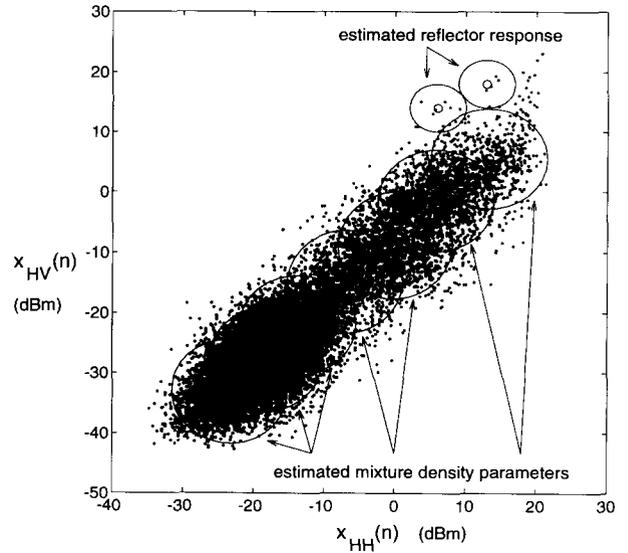


FIGURE 11. Scatter plot of training data. The components of the estimated mixture densities are superimposed.

The covariances are chosen to be circular, equal to  $\Sigma = \text{diag}(4, 4)$  dBm. Having adequately modeled the clutter and targets, the parameters are transferred to the RBF network, as described in Section 2.2.5. The parameters of the nine two-dimensional mixture component densities are projected onto the  $x_{HH}$  and  $x_{HV}$  axes, resulting in parameters for 18 one-dimensional RBF units. To improve the interpolation quality of the network, the factor  $\alpha = 0.01$  is applied to the clutter spread parameters.

The next step is the mutual information training to learn to the weights. A sub-sampled data set from DOFASCO and La Salle Park regions is used for this purpose. The data set is the same as the one shown in the scatter plot, with the same area coverage as the

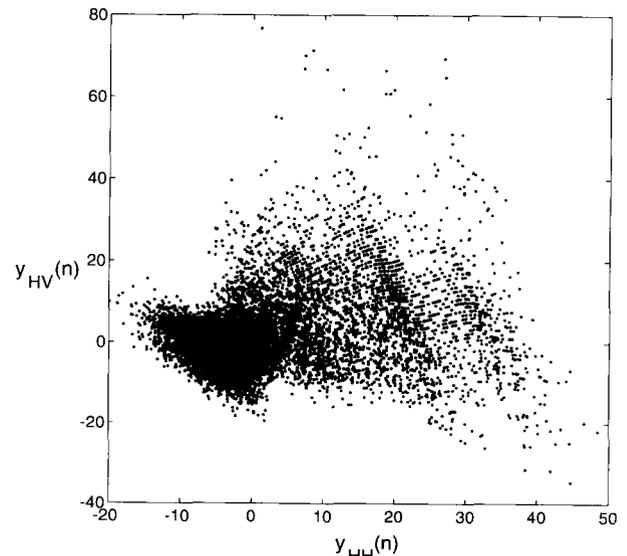


FIGURE 12. Scatter plot of output of the RBF network.

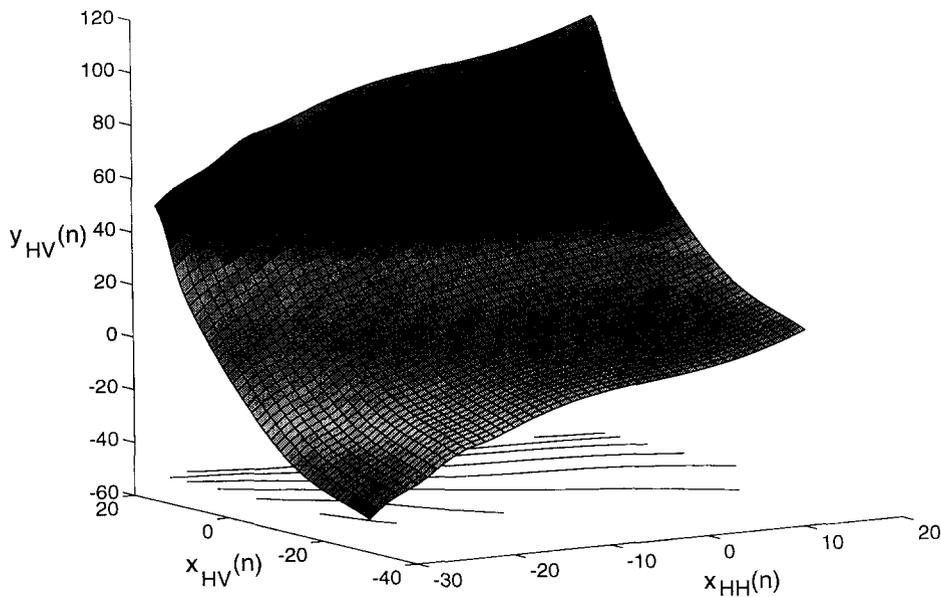


FIGURE 13. Surface plot of nonlinear transformation learned by the RBF network.

sub-images, except sub-sampled in range by a factor of four and in azimuth by a factor of 10. The weights are initialized to approximate the same mapping as that of linear network by solving for the RBF weights using least-squares fitting. A constrained optimization routine is used to minimize the cost function given by (30), subject to the constraints. After convergence, the residual mutual information is estimated to be equal to approximately 0.036 bits. The scatter plot of the output of the network is shown in Figure 12. As can be observed, the data distribution is more clustered around a single point, rather than distributed over a large range as in the scatter plot of the input data. Figure 13 shows the resultant non-linear mapping learned by the RBF network. The inputs are along the  $x$ - and  $y$ -axes, and the  $z$ -axis height is the  $y_{HV}$  output.

The output of the network is processed by the CFAR algorithm, as in the ACPIC case. The resultant sub-images in Figure 14 show much enhanced target visibility, in both DOFASCO and La Salle Park areas. The clutter is generally well suppressed. However, some false targets are visible as well. The statistics in Table 5 verify the higher visibility of the target. The normalized TCR values are the highest of all the individual methods considered thus far, for both DOFASCO and La Salle Park areas.

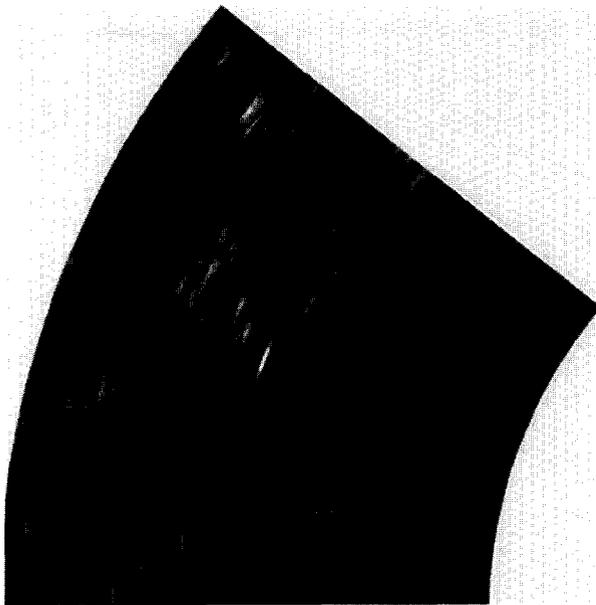
### 3.4. Modular Neural Network Processor

The sub-images in Figure 15 show an example of the resultant output from the modular network. The images retain desirable characteristics of both the individual processing techniques, suppressing both average and peak clutter while enhancing the target.

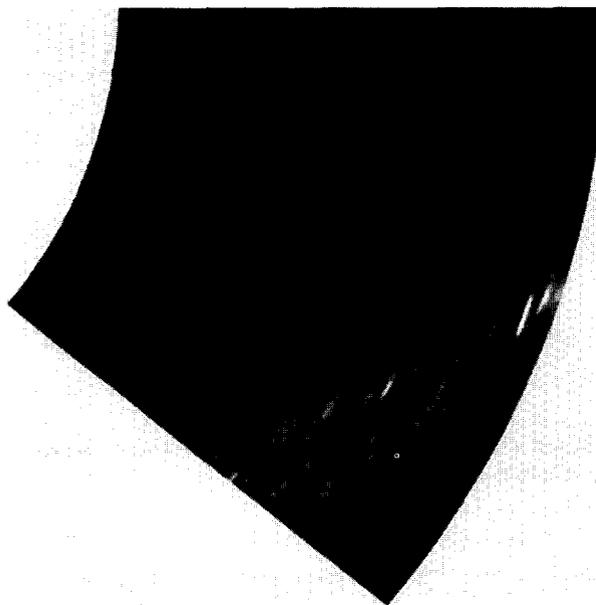
The estimated normalized TCR in Table 6 shows an improvement for the La Salle Park area, while the value for the DOFASCO area stayed essentially unchanged. It is likely that no further improvement is possible in the DOFASCO area in terms of the normalized TCR measure.

The normalized TCR was introduced as a measure that is useful for quantifying the visibility of the target against the average background clutter. A new measure is now introduced, namely the estimate of the receiver operating characteristic (ROC), which plots the estimated probability of detection against the estimated probability of false alarm for a particular processor. The ROC is indicative of the suppression of the peak clutter phenomenon, which drives the false alarm rate. A thorough statistical analysis for estimating the ROC requires many independent data sets containing target and clutter. Unfortunately, this experiment limits the analysis to a single scene averaged over 28 scans. Proceeding with this limitation in mind, the ROCs are plotted for the adaptive interference canceller network in Figure 16, for the RBF network in Figure 17, and for the modular network shown in Figure 18. The dotted lines show the 90% confidence intervals for the estimated probability of detection, and false alarm. The number of target samples is low; accordingly the graph to the left of the dotted line is regarded as not statistically significant. However, the overall trends indicated in the graph still provide useful information.

As can be observed from these plots, the ROC for the modular network shows an improvement for the DOFASCO area, yet no improvement is noted for the La Salle Park area. This is in contrast to the NTCR estimates which indicated improvement in the La Salle Park area but not in the DOFASCO area. The



DOFASCO



La Salle Park

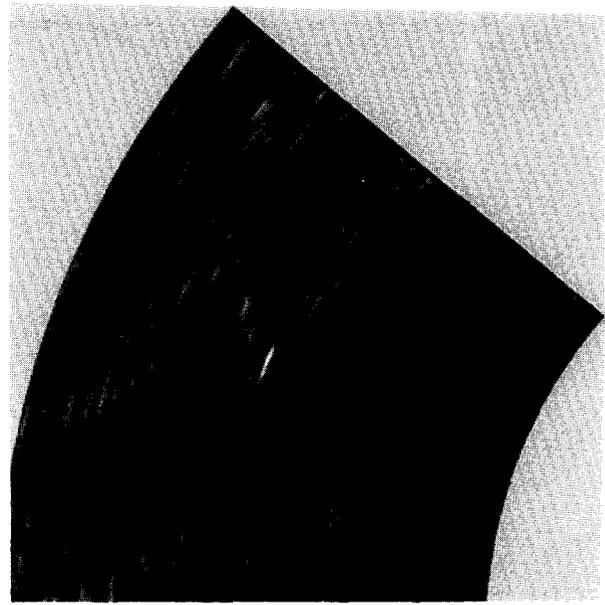
FIGURE 14. Sub-images of the RBF network-processed data.

average visibility is increased in the case of the La Salle Park area, and the peak clutter that is responsible for false alarms is reduced in the DOFASCO area (as suggested by the ROC plots).

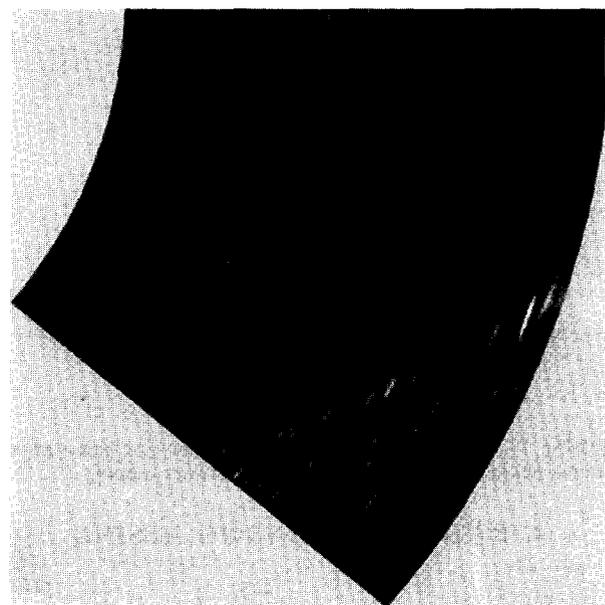
TABLE 5

Estimated Parameters of RBF Network-processed Sub-images

	DOFASCO	La Salle Park
Target ( $\hat{\mu}_t, \hat{\sigma}_t$ )	(84.4, 4.09)	(93.1, 6.22)
Clutter ( $\hat{\mu}_c, \hat{\sigma}_c$ )	(11.1, 1.58)	(10.7, 2.74)
TCR	80.4	82.4
Normalized TCR	46	30



DOFASCO



La Salle Park

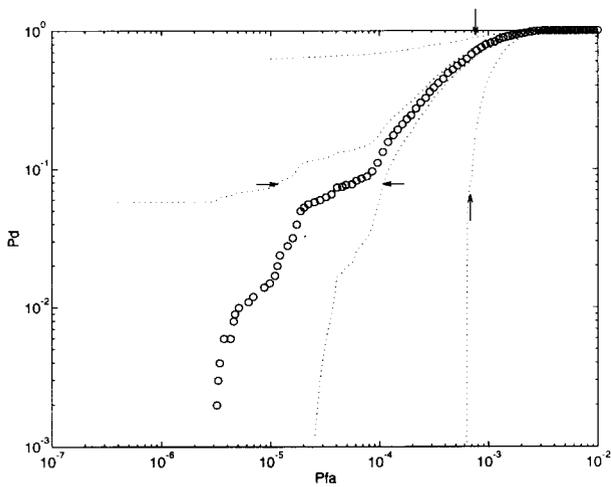
FIGURE 15. Sub-images of modular network-processed data.

This shows that the modular network is able to integrate the performance aspects of both networks to give an improved overall result.

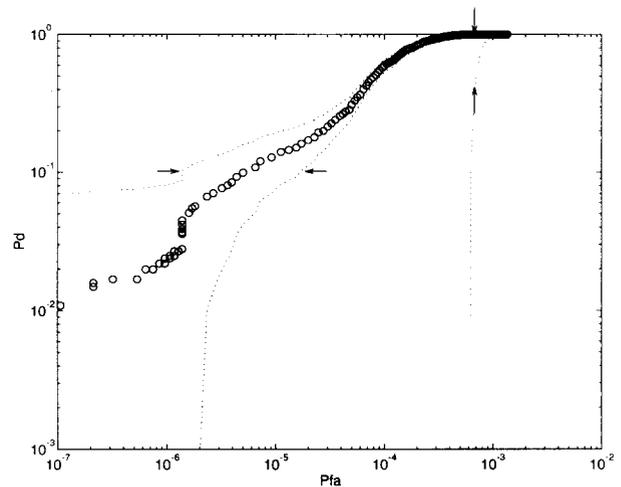
TABLE 6

Estimated Parameters of Modular Network Processed Sub-images

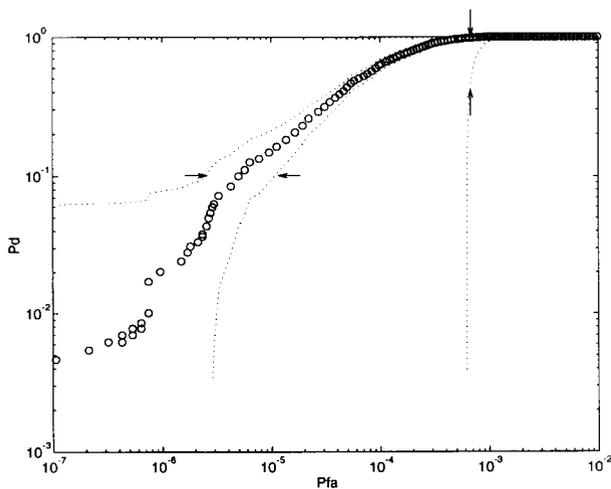
	DOFASCO	La Salle Park
Target ( $\hat{\mu}_t, \hat{\sigma}_t$ )	(2.06, 0.10)	(2.68, 0.18)
Clutter ( $\hat{\mu}_c, \hat{\sigma}_c$ )	(0.522, 0.034)	(0.556, 0.067)
TCR	1.54	2.12
Normalized TCR	46	32



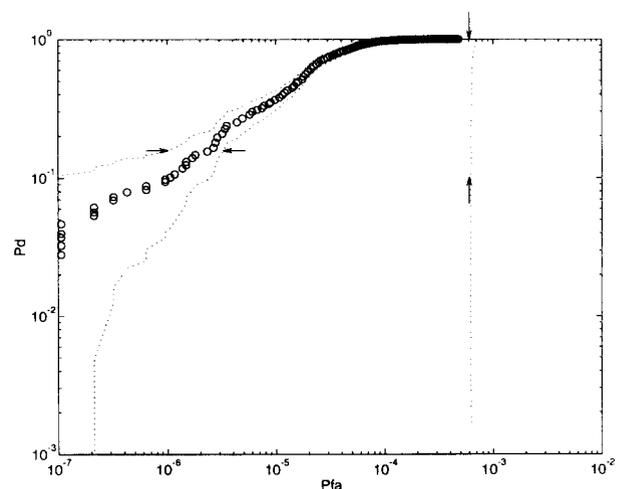
DOFASCO



DOFASCO



La Salle Park



La Salle Park

FIGURE 16. Estimated receiver operating characteristic curves for the adaptive interference canceller network output.

FIGURE 17. Estimated receiver operating characteristic curves for the RBF network output.

#### 4. POST-DETECTION PROCESSING

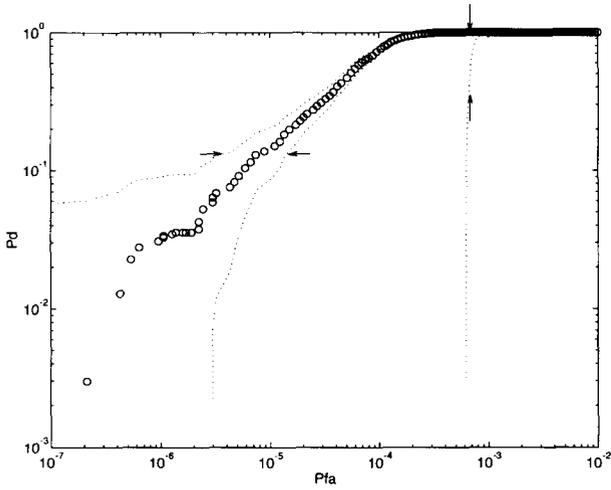
The radar environment contains various forms of clutter which may have high cross-polar radar returns. This type of clutter causes false alarms, degrading the overall performance of the detection system. Up to this point, the processing has not made use of any *a priori* information about the target location. *A priori* information about the target position will now be used in a post-detection processor to reduce the false alarm rate to a very small number.

One observation can be readily made about the location of a reflector target: the reflector must be visible to ships navigating the confined waterway, and should therefore be located near an unobstructed water-land interface. Although other choices further inland are possible, natural and man-made obstructions are more likely to limit target visibility, and

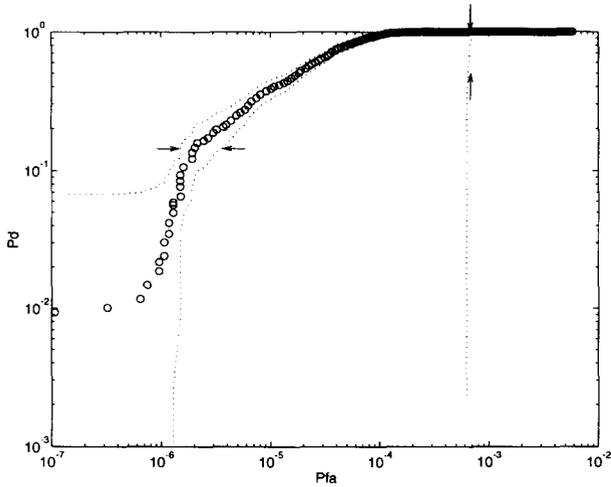
therefore make these locations a poor choice in general. It is therefore assumed that the reflector will always be located at a preferred location near the water-land interface, and detections that are distant from this boundary can be discounted as likely to be false alarms.

Another observation about the reflector target position is that the locations chosen for the reflectors would not be adjacent to an area containing clutter with high cross-polar radar returns. It is perceived to be an easy task to ensure that the reflectors are located in an area where cross-polar clutter is minimal, since this is under the control of the installers of the navigational system.

The post-detection processor described is designed to combine *a priori* location information along with the primary detection results, thus giving an improved final detection result. In this manner, false alarms can be virtually eliminated. Figure 19 is a



DOFASCO



La Salle Park

FIGURE 18. Estimated receiver operating characteristic curves for the modular network output.

block diagram of the entire signal processing, detection and post-detection system.

The output of the HH-pol and HV-pol modular network processing stage is presented to the detection system. After CFAR processing, the output of the threshold detector is a binary map of the location of the detections. If the threshold is set sufficiently low, both targets and false alarms are included in the set of detections. The water-land interface of interest is determined automatically using a vision-based edge detection algorithm, known as the CARTOON algorithm (Richards et al., 1988). The algorithm is tuned to the radar environment, accepting the HH-pol radar image as input in B-scan form (range-azimuth coordinates) and producing an edge map at the output.

#### 4.1. The CARTOON algorithm

A block diagram of the CARTOON algorithm is shown

in Figure 20. The basis of the algorithm is the detection of zero-crossings of an image after being filtered by the Laplacian operator

$$\nabla^2 G(x, y), G(x, y) = e^{-1/2(x^2/\sigma^2+y^2/\sigma^2)}. \quad (51)$$

The location of the zero-crossings correspond to edges in the image. This Laplacian operator is discussed in detail by Marr (1982). In general, the operator is chosen to be circularly symmetric. The algorithm was generalized in this particular application to use elliptically symmetric Gaussian operators, since the resolution of the radar image is generally different in range and in azimuth. The Laplacian of the Gaussian operator (normalized) is therefore

$$\nabla^2 G_n(r) = \left(1 - \frac{r^2}{2}\right) e^{-1/2r^2}, r^2 = \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}. \quad (52)$$

The CARTOON algorithm makes use of the fine and coarse operator masks to achieve a robust edge map, as indicated in Figure 20 by  $\nabla^2 G_f$  for the fine mask and  $\nabla^2 G_c$  for the coarse mask. After filtering with the masks, the positive and negative bitmaps are created by thresholding the images at 50% of their range between minimum and maximum values. The resultant bitmaps are ANDed together, and then smoothed by a Gaussian mask. The smoothed bitmaps are then thresholded at 6% of their range between minimum and maximum values. These bitmaps are ANDed together to form the resultant binary edge map. Only edges that are common to both fine and coarse filtered bitmaps appear in the final output.

The parameters of the CARTOON algorithm need to be adjusted to match the resolution of the radar system, and the scale of the desired edge map. Two masks, a fine mask and a coarse mask, need to be chosen so as to be fine enough to preserve the desired detail in the edges, yet coarse enough so that noise and small objects do not get recognized as edges. The scale was purposefully chosen so that small objects, such as ships, do not get detected as edges. For the fine mask, parameters  $\sigma_x = 6$  and  $\sigma_y = 24$ , and for the coarse mask, parameters  $\sigma_x = 12$  and  $\sigma_y = 48$  are used. The Gaussian smoothing mask used is chosen to be half the size of the fine mask, so that  $\sigma_x = 3$  and  $\sigma_y = 12$  for the operator. The thresholds are determined through experimentation. The first set of thresholds (that produce the fine and coarse edge maps) is set to 48% of the range between minimum and maximum values of the respective images that are presented to the threshold function. The second set of thresholds (that produce the positive and negative edge maps) is set to 6% of the range between

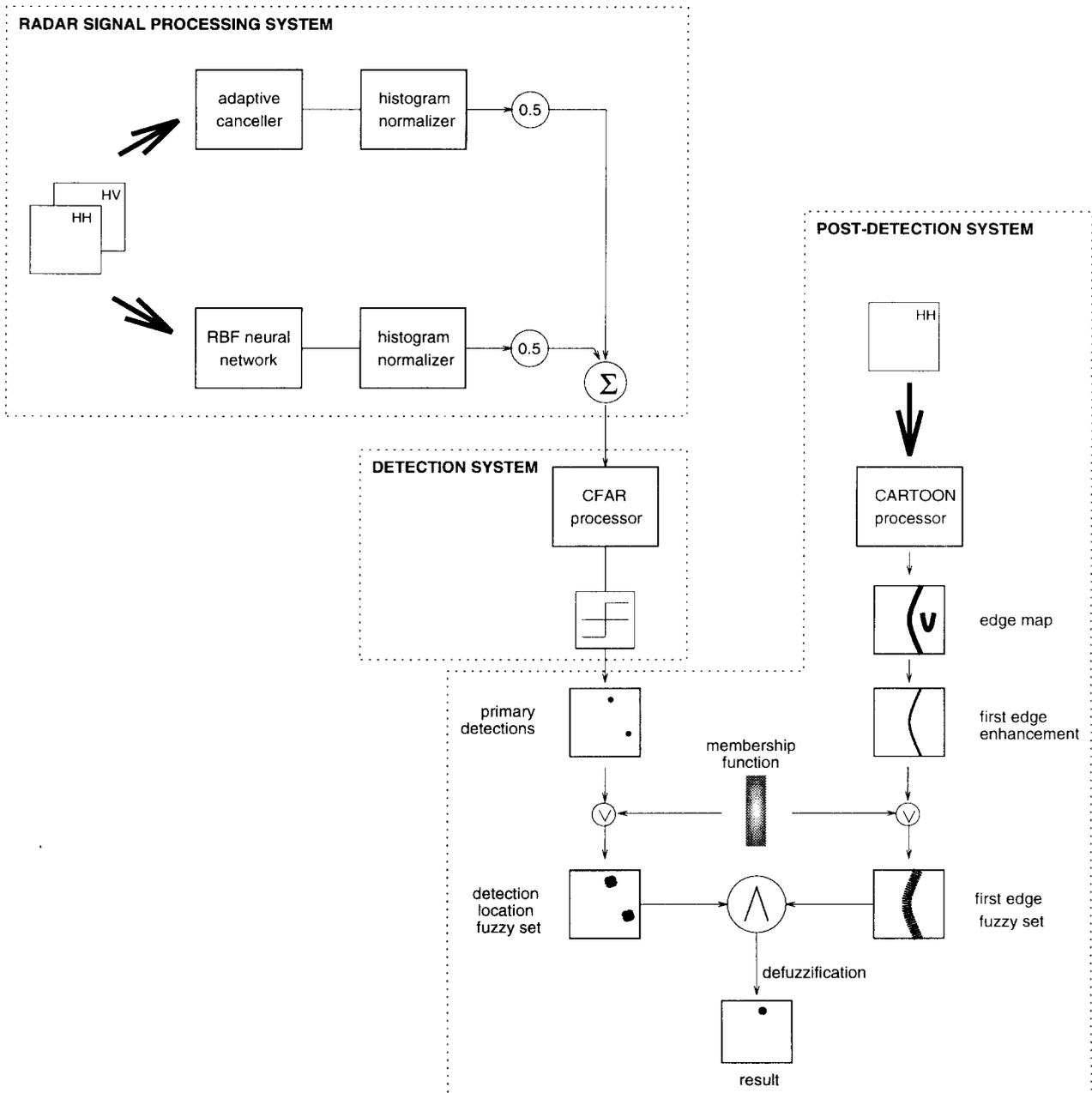


FIGURE 19. Block diagram of modular signal processing and detection system.

minimum and maximum values of the respective images that are presented to the threshold function.

Figure 21 shows the results for the scenes of interest. Note the false edge artifacts found in the La Salle Park image. These are due to the finite support of the image, and do not pose a serious problem. They can simply be ignored since the likelihood of false alarms as a result of sea clutter in a confined waterway is extremely low. In practice, we need only choose an image slightly larger than that required to avoid these artifacts.

Since we are only interested in the first water-land interface, the edge map produced by the CARTOON algorithm is presented to a “first-edge” detector, which

responds to the first edge it finds along each radar sweep; the resulting edge map is shown in Figure 22.

#### 4.2. Fuzzy Detection Processing

Fuzzy set theory is next used to combine the two pieces of information, specifically, that of detection locations and “first-edge” locations. The intersection of these two fuzzy sets effectively combines the detections with the edge information, resulting in a greatly reduced false alarm rate.

The crisp set of primary detections is fuzzified by choosing a Gaussian form for the membership function of the location of the detection. In a similar

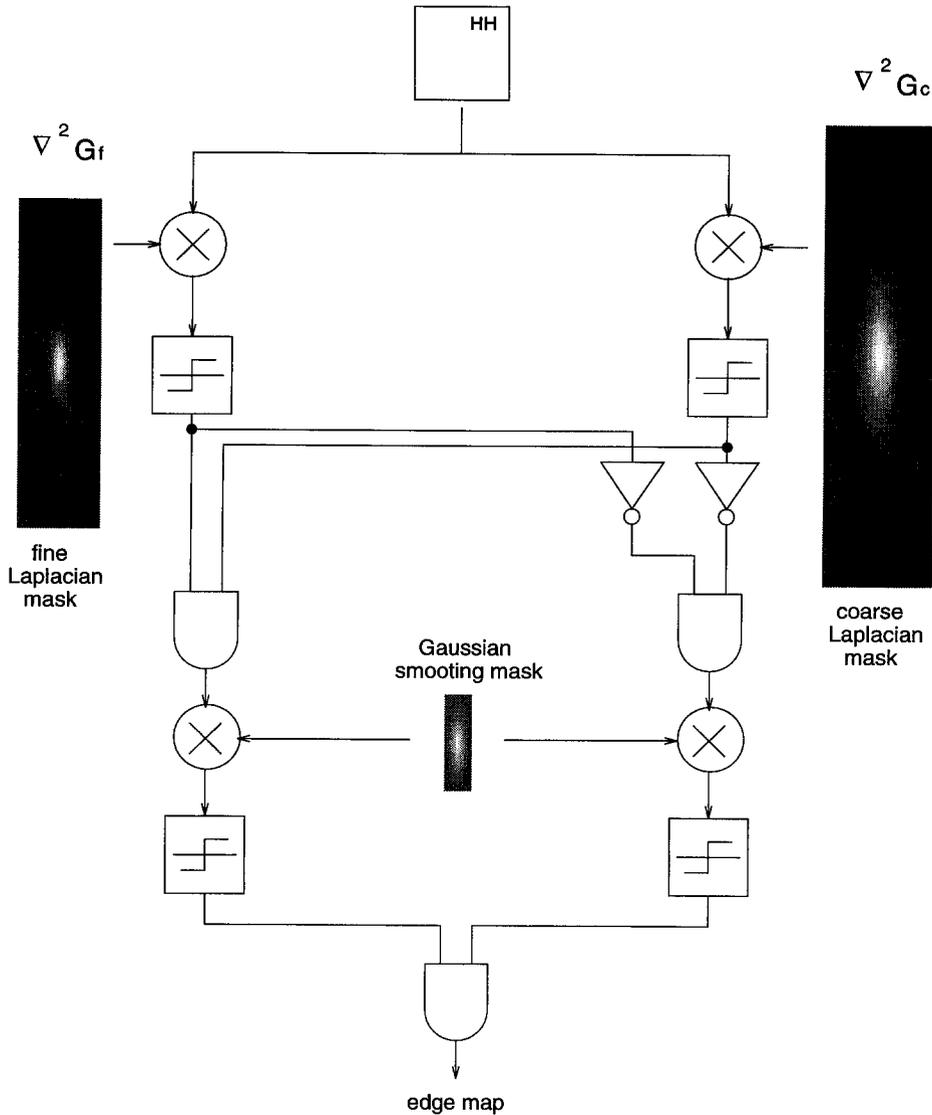


FIGURE 20. Block diagram of the CARTOON algorithm.

fashion, a Gaussian form for the membership function is used to specify the degree of “edginess” at a particular location in the radar image. This is done by superimposing a Gaussian function at each pixel in the corresponding detection and edge images. The two-dimensional Gaussian membership function is centered on each white pixel. Since adjacent pixels tend to cause the membership functions to overlap, the fuzzy union operator is used to combine these subsets. The resultant membership function is defined as

$$\mu_G(x, y) = \cup_i \mu_F(x - x_i, y - y_i), \quad (53)$$

where  $x_i$  and  $y_i$  are the locations of the non-zero pixels in the radar images,  $\mu_F$  is the membership function for the individual pixels, and the fuzzy union operator  $\cup$  is defined as

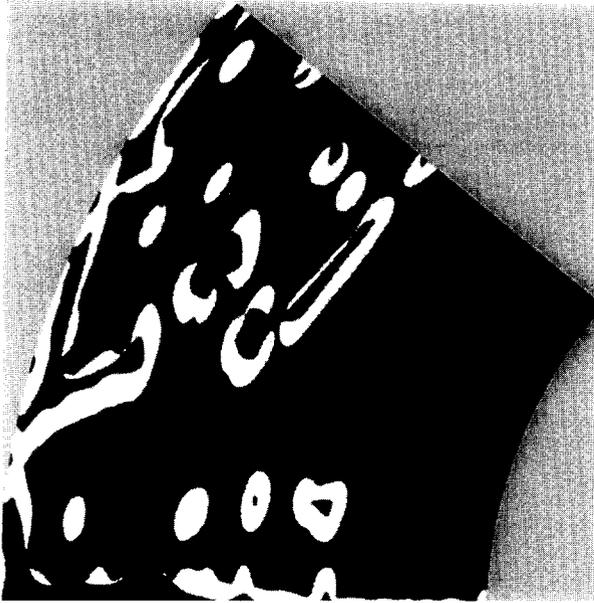
$$\cup_i \mu_F(x_i, y_i) \max_i \mu_F(x_i, y_i). \quad (54)$$

The operator finds the maximum value over the set of overlapping membership functions at each image location  $(x, y)$  (Terano et al., 1992).

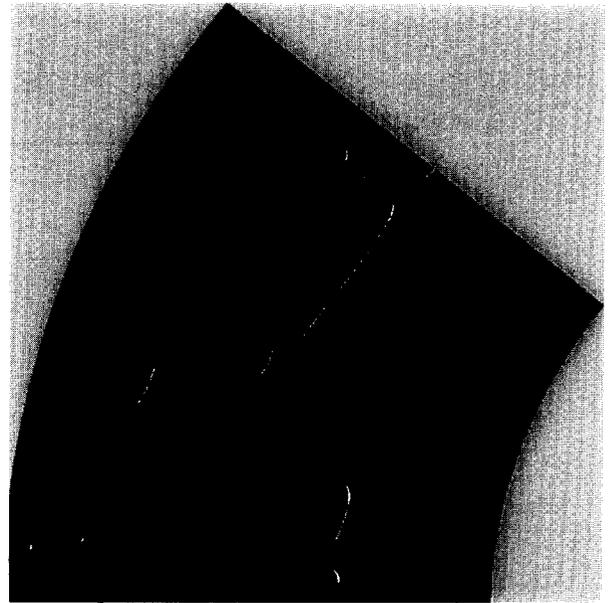
We are interested in the case where a detection is in the vicinity of an edge, or in other words, when a target and an edge are present together. It follows that the desired resultant set can be found by taking the intersection of the fuzzy set of detections with the fuzzy set of image edginess, such that

$$\mu_H(x, y) = \mu_{Gd}(x, y) \wedge \mu_{Ge}(x, y), \quad (55)$$

where  $\mu_{Gd}$  is the fuzzy set of detections, and  $\mu_{Ge}$  is the fuzzy set of edges. This is equivalent to determining the minimum value between the two fuzzy sets at each  $(x, y)$  location (Terano et al., 1992). Finally, since we are interested in a binary result, we need to defuzzify the set  $\mu_H(x, y)$ . A threshold is chosen in order to make the set of post-detections crisp, or in other



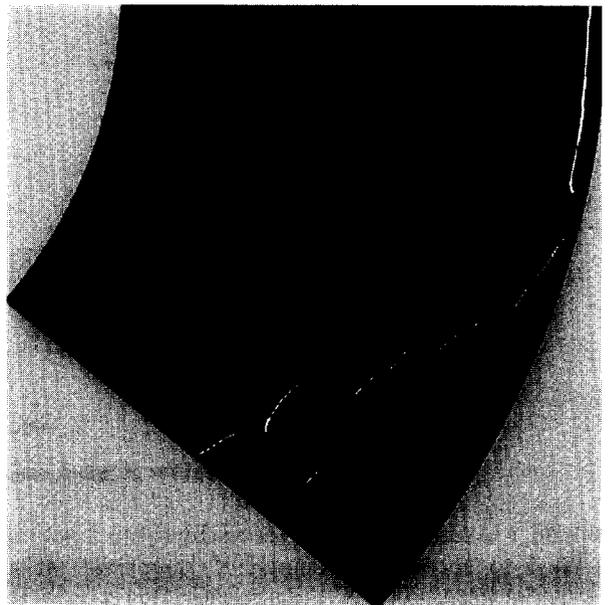
Dofasco



Dofasco



La Salle Park



La Salle Park

**FIGURE 21.** Binary edge maps produced by the CARTOON algorithm.

**FIGURE 22.** Result after radially processed "first-edge" detection

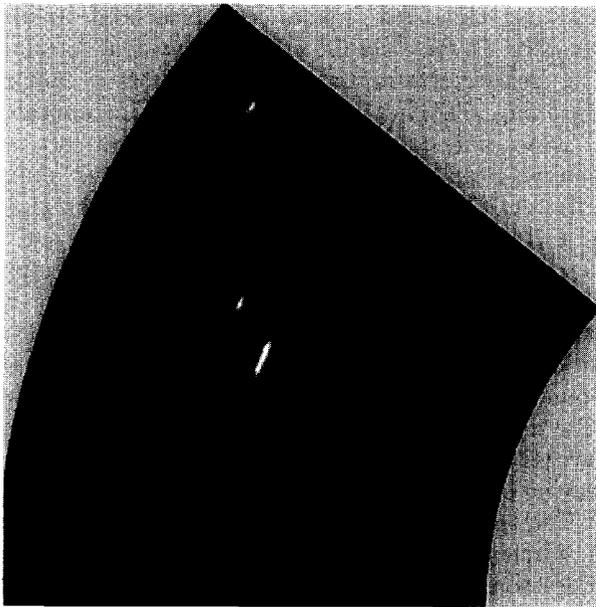
words, binary valued. The final binary result contains only detections that are near edges, which should be our reflector targets. Since edges (as defined in this section) generally occupy a small percentage of the total area of interest, this post-detection processor greatly reduces the probability of false alarm.

#### 4.3. Example of Post-Detection Processing

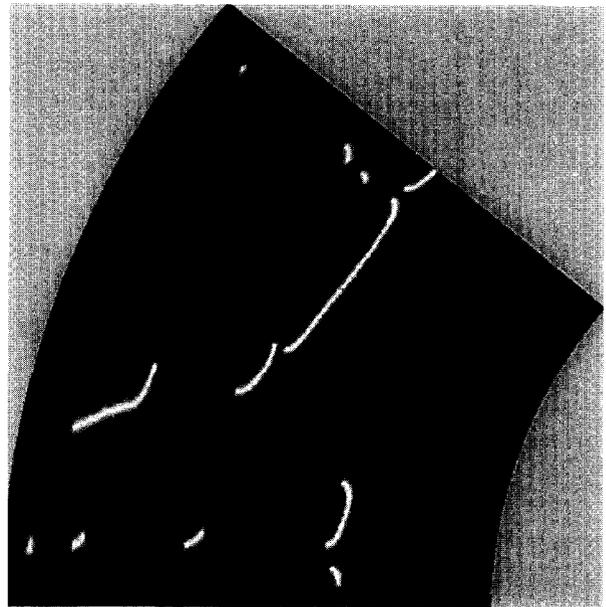
To demonstrate the operation of the post-detection processor, a threshold level was chosen for the primary detection stage such that the false alarms

would be present in the output. The fuzzified detections are shown in Figure 23, and the fuzzified "first-edges" in Figure 24. The Gaussian membership function was chosen to have the same parameters as that used for the smoothing function in the CARTOON algorithm. The final post-detection results are shown in Figure 25, where they are superimposed on a reduced intensity HH-pol image as a guide to the location of the detections. The false alarms are present after primary detections have been effectively removed, leaving only the targets of interest.

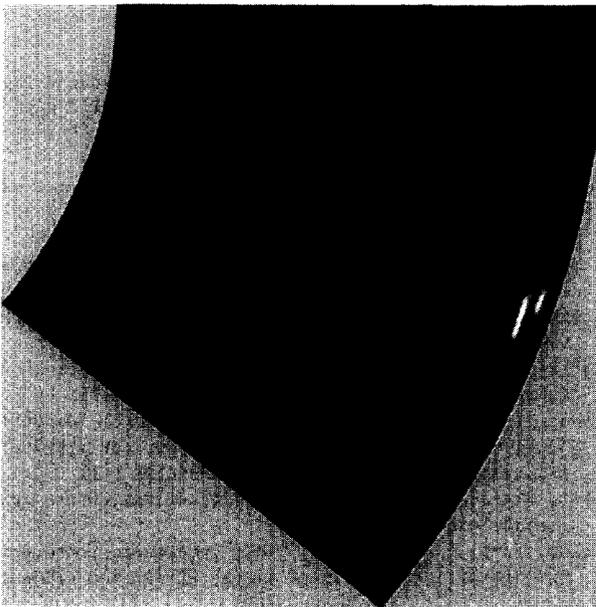
Only the "first-edge" areas are considered suitable



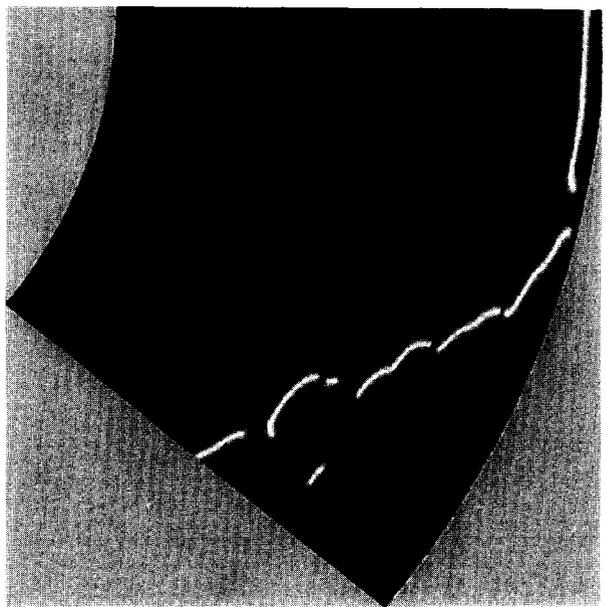
Dofasco



Dofasco



La Salle Park



La Salle Park

**FIGURE 23. Gaussian membership function applied to primary detections.**

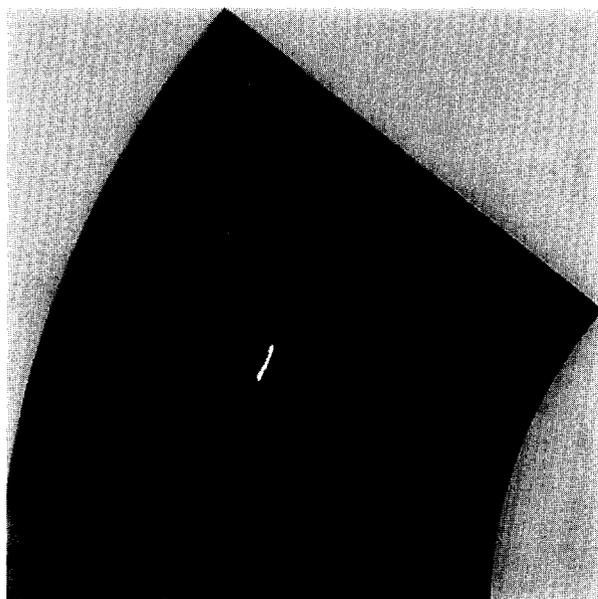
**FIGURE 24. Gaussian membership function applied to "first-edge" image.**

locations for the reflector targets. As long as this assumption holds true, it is obvious that the false alarm rate is reduced by the factor of first edge area over the total area of interest. Since the "first-edge" areas are estimated to be approximately 10% of the total sub-image land clutter area, the false alarm is reduced by a factor of 10 or more.

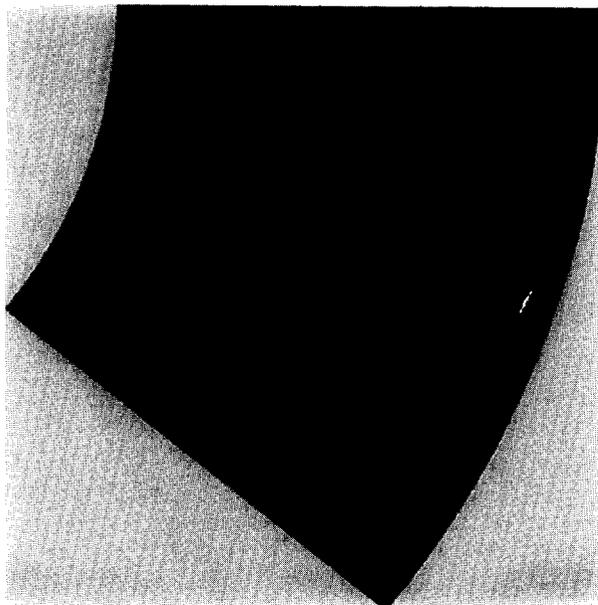
**5. SUMMARY**

The estimated normalized target-to-clutter ratio for the various signal processing methods examined in

this paper are summarized in Table 7. Each of the processing methods has a distinct character. The CA-CFAR processor uses a two-dimensional local estimate of clutter to achieve a constant false alarm rate performance at the output. The ACPI processor reduces clutter in a non-stationary clutter environment. The RBF network uses non-linear mapping to overcome the non-Gaussian nature of clutter statistics and provides a superior target enhancement. By integrating the desirable characteristics of these sub-network solutions into a modular network structure, a final average (and peak) clutter



Dofasco



La Salle Park

**FIGURE 25.** Final post-detection result. The original HH-pol image is superimposed as a visual aid.

suppression performance is achieved that is better than any one of the methods used by itself. The

**TABLE 7**

**Summary of NTCR performance of various signal processors**

Type of processor	NTCR	
	DOFASCO	La Salle Park
CA-CFAR	16	7
ACPIC/CA-CFAR	35	29
RBF net/CA-CFAR	46	30
Modular/CA-CFAR	46	32

resulting enhanced images are presented to the operator, as well as to the primary threshold detector. The post-detection stage strives to mimic human reasoning, albeit in a primitive fashion. It incorporates *a priori* information about the context of the waterway and the likely location of the radar reflectors. The solution involves a novel combination of a vision-based image processing technique and a fuzzy processor. The vision-based edge detection algorithm, CARTOON, is generalized to operate at radar resolutions. The fuzzy set reasoning combines the edge information corresponding to the water-land boundary with primary detection results to remove false targets. A demonstration using example images successfully eliminates false targets from the image, leaving only the desired reflector target locations.

The entire signal processing and detection system is implementable with relatively simple analog processing elements, in the spirit of a modular neural network architecture.

## REFERENCES

- Becker, S., & Hinton, G. E. (1989). Spatial coherence as an internal teacher for a neural network. Technical Report CRG-TR-89-7, Department of Computer Science, University of Toronto, Toronto, Ont.
- Broomhead, D. S., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321-355.
- Casdagli, M. (1989). Nonlinear prediction of chaotic time series. *Physica D*, 35, 335-356.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley.
- Cronley, J. (1956). Clutter on radar displays—reduction by use of logarithmic receivers. *Wireless Engineer*, 33(4), 83-95.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Proceedings of the Royal Statistical Society*, 39, 1-38.
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley.
- Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2), 1134-1140.
- Gilbert, E. N. (1958). An outline of information theory. *American Statistician*, 12, 13-19.
- Giuli, D. (1986). Polarization diversity in radars. *Proceedings of the IEEE*, 74(2), 245-269.
- Goldstein, G. B. (1973). False-alarm regulation in log-normal and weibull clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 9(1), 84-92.
- Haykin, S. (1991). *Adaptive filter theory* (2nd edn). Englewood Cliffs, NJ: Prentice Hall.
- Haykin, S. (1992). Polarimetric radar for accurate navigation. *Canadian Journal of Electrical and Computer Engineering*, 17(3), 130-135.
- Haykin, S., & Ukrainec, A. (1993). Neural networks for adaptive signal processing. In Kalouptsidis, N., & Theodoridis, S. (Eds), *Adaptive system identification and signal processing algorithms*. New York: Prentice Hall.
- Hrycej, T. (1992). *Modular learning in neural networks*. New York: John Wiley.
- Jones, R. D., Lee, Y. C., Barnes, C. W., Flake, G. W., Lee, K.,

- Lewis, P. S., & Qian, S. (1989). Function approximation and time series prediction with neural networks. Technical report, Los Alamos National Laboratory, Los Alamos, New Mexico.
- Kullback, S. (1968). *Information theory and statistics*. New York: Dover Publications.
- Lowe, D. (1989). Adaptive radial basis function nonlinearities, and the problem of generalization. In *First IEE International Conference on Artificial Neural Networks*, pp. 171–175.
- Lowe, D., & Webb, A. (1989). Adaptive networks, dynamical systems, and the predictive analysis of time series. In *First IEE International Conference on Artificial Neural Networks*, pp. 95–99.
- Macikunas, A., Haykin, S., & Greenlay, T. (1988). Trihedral radar reflector. U.S. Patent No. 4,843,396.
- Marr, D. (1982). *Vision*. New York: W. H. Freeman.
- Minkler, G., & Minkler, J. (1990). *CFAR*. Baltimore, MD: Magellan Book Company.
- Moody, J., & Darken, C. (1989a). Learning with localized receptive fields. In D. Touretzky, G. Hinton, & Sejnowski, T. (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 133–143), Morgan Kaufmann: San Mateo, CA.
- Moody, J., & Darken, C. J. (1989b). Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1(3), 281–294.
- Nitzberg, R. (1992). *Adaptive signal processing for radar*. Norwood, MA: Artech House.
- Nowlan, S. J. (1990a). Max likelihood competition in RBF networks. Technical Report CRG-TR-90-2, Connectionist Research Group, University of Toronto, Toronto, Canada.
- Nowlan, S. J. (1990b). Maximum likelihood competitive learning. In D. S. Touretzky (Ed.), *Advances in neural information processing systems 2* (pp. 574–582). IEEE, Morgan Kaufmann.
- Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, 26(2), 195–239.
- Richards, W., Nishihara, H., & Dawson, B. (1988). Cartoon: A biologically motivated edge detection algorithm. In W. Richards (Ed.), *Natural computation* (Chapter 4, pp. 55–69). Cambridge, MA: MIT Press.
- Saha, A., & Keeler, J. (1990). Algorithms for better representation and faster learning in radial basis function networks. Technical Report ACT-NN-028-90, Microelectronics and Computer Technology Corporation, Austin, TX.
- Skolnik, M. I. (1980). *Introduction to radar systems* (2nd edn). New York: McGraw-Hill.
- Terano, T., Asai, K., & Sugeno, M. (1992). *Fuzzy systems theory and its applications*. San Diego, CA: Academic Press.
- Ukrainec, A. M. (1994). Enhancement of cooperative cross-polar radar targets. PhD thesis, McMaster University, Hamilton, Ontario, Canada.
- Ukrainec, A., & Haykin, S. (1991a). Application of unsupervised neural networks to the enhancement of polarization targets in dual-polarized radar images. In *25th Asilomar Conference on Signals, Systems & Computers*. IEEE Computer Society.
- Ukrainec, A., & Haykin, S. (1991b). Signal processing with radial basis function networks using expectation maximization algorithm clustering. In *SPIE 36th International Symposium on Optical and Optoelectronic Applied Science and Engineering*.
- Ukrainec, A., & Haykin, S. (1992). Enhancement of radar images using mutual information based unsupervised neural network. In *Canadian Conference on Electrical and Computer Engineering* (pp. MA6.9.1–MA6.9.4). Toronto, Canada.
- Walker, M. R., & Akers, L. A. (1992). Information-theoretic analysis of finite register effects in neural networks. In *International Joint Conference on Neural Networks* (Vol. II, pp. 666–671) Piscataway, NJ: IEEE.
- Widrow, B. (1975). Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12), 1692–1716.
- Widrow, B., & Stearns, S. (1985). *Adaptive signal processing*. Englewood Cliffs, NJ: Prentice-Hall.
- Wozencraft, J. M., & Jacobs, I. M. (1965). *Principles of communication engineering*. New York: John Wiley.
- Zemel, R. S., & Hinton, G. E. (1991). Discovering viewpoint-invariant relationships that characterize objects. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems 3*. San Mateo, CA: Morgan Kaufmann Publishers.

## NOMENCLATURE

$\beta$	detection threshold
$x_{HH}(n)$	HH-pol sampled video return
$y_{HV}(n)$	processed HV-pol sampled video
$b_i$	primary detection $i$
$d_i$	final detection $i$
$w_{opt}$	optimum tap-weight
$R_{x_{HH}, x_{HV}}(0)$	statistical correlation between $x_{HH}$ and $x_{HV}$ at lag 0
$n$	sample index
$N$	total number of samples
$e_{HV}(n)$	output of interference canceller, at sample $n$
$\hat{w}(n)$	estimated tap weight
$\mu_{LMS}$	step-size parameter of LMS algorithm
$\sigma_{x_{HH}}^2$	statistical variance of a process $x_{HH}$
$a, b$	constants
$\tau$	time constant
$T$	sample period
$f(x)$	probability density function (p.d.f.)
$S$	support set
$H(X)$	entropy of $X$
$D(f_1    f_2)$	Kullback–Leibler distance
$I(X; Y)$	mutual information
$\mathbf{R}$	autocorrelation matrix
$ \mathbf{R} $	determinant of a matrix
$\rho_{x, y}$	correlation coefficient between $x$ , $y$
$R_{x, y}(0)$	correlation function between $x$ , $y$ at lag 0
$\mu$	statistical mean of a process
$\mathbf{W}$	weight matrix
$\phi_j(x)$	radial basis function $j$
$\mathbf{x}$	data vector
$c_j$	basis function center $j$
$S_j$	basis function spread $j$
$w_{ji}$	weight connecting $j$ th hidden unit to $i$ th output
$N_h$	number of hidden layer units
$p(\mathbf{x})$	p.d.f. of vector $\mathbf{x}$
$\theta_j$	vector of parameters
$P(j)$	<i>a priori</i> probability $j$
$d$	dimensionality of multivariate density
$\Sigma_i$	covariance matrix $j$
$\mu_j$	vector mean $j$
$\alpha$	spread factor
$C(W)$	cost function
$\lambda$	Lagrangian factor
$E[\cdot]$	expectation

$J_{\text{misc}}$	miscellaneous cost terms	$\lambda_X$	wavelength
$\hat{\gamma}$	estimated skew parameter	$\sigma_R$	radar cross-section of target
$\hat{n}(i)$	estimated histogram at bin $i$	$\sigma_o$	radar cross-section of calibration sphere
$N_{\text{bins}}$	number of histogram bins	$\sigma_r$	relative radar cross-section
$N_{\text{total}}$	total number of image data points	$L_{\text{misc}}$	miscellaneous losses
$N_b$	bin number of 2% black level	$\hat{\mu}_t$	estimated mean target response
$N_w$	bin number of 2% white level	$\hat{\mu}_c$	estimated mean clutter response
$t_i$	target pixels	$\hat{\sigma}_c$	estimated standard deviation clutter response
$c_i$	clutter pixels	$Q(x)$	$Q$ -function
$N_t$	number of target pixels	$\nabla^2 G(x, y)$	Laplacian operator
$N_c$	number of clutter pixels	$\mu_G(x, y)$	fuzzy membership function
$R$	range	$\vee_i$	fuzzy union operator over $i$
$G_a$	one-way gain of antenna	$\wedge$	fuzzy intersection operator
$P_r$	received power		
$P_t$	transmitted power		