
Discriminating Targets from Clutter

Daniel E. Kreithen, Shawn D. Halversen, and Gregory J. Owirka

■ The Lincoln Laboratory multistage target-detection algorithm for synthetic-aperture radar (SAR) imagery can be separated into three stages: the prescreener, the discriminator, and the classifier. In this article, we focus on the discrimination algorithm, which is a one-class, feature-based quadratic discriminator. An important element of the algorithm design is the choice of features. We examine fifteen features that are used in the discrimination algorithm—three features developed by Lincoln Laboratory, nine developed by the Environmental Research Institute of Michigan, two developed by Rockwell International Corporation, and one developed by Loral Defense Systems. The set of best features from this pool of fifteen was determined by a theoretical analysis, and was then verified by using real SAR data. Performance was evaluated for a number of different cases: for fully polarimetric data and HH polarization data and for 1-ft resolution data and 1-m resolution data. In all cases the theoretical performance analysis closely matched the real data performance. This closeness demonstrates a good understanding of the discrimination algorithm. In addition, we formulate a set of criteria for best feature choice that apply to quadratic discrimination algorithms in general.

LINCOLN LABORATORY HAS PROCURED a fully polarimetric, instrumentation-quality, high-resolution (1 ft by 1 ft), 35-GHz, millimeter-wave (MMW) synthetic-aperture radar (SAR), which has been used to gather imagery of targets of interest and clutter in a number of different locations and deployments. The radar, which is mounted in a Gulfstream G-1 aircraft, records data in-flight onto 24-track magnetic tapes. The tapes are then processed on the ground to form the SAR imagery. A recent *Lincoln Laboratory Journal* article by Leslie M. Novak et al. describes this radar system [1].

The Surveillance Systems group at Lincoln Laboratory has been developing algorithms to detect targets of interest in this SAR imagery. A block diagram of the algorithm suite is shown in Figure 1. The target-detection algorithm suite takes the form of a multistage algorithm. In theory, it is possible to construct a single algorithm that performs target detection in an optimal manner, and which exploits all of

the information present in a high-resolution SAR image. Unfortunately, it is often difficult to design algorithms using the single-algorithm approach, because high-resolution SAR imagery is difficult to model accurately and hence is poorly understood. The multistage approach becomes an attractive alternative, because of the reduction in required computational capability and the simplification in algorithm design.

The Lincoln Laboratory multistage algorithm has three separate stages, each of which performs easily identifiable functions. The first stage, which is called the *prescreener*, is a computationally simple algorithm whose function is to pass all targets and eliminate only obviously non-targetlike naturally occurring clutter. The second stage, called the *discriminator*, ideally eliminates all naturally occurring clutter that has been passed by the prescreener, and passes only man-made objects to the third stage, which is called the *classifier*. The classifier receives all man-made objects that have been passed by the discriminator and categorizes each

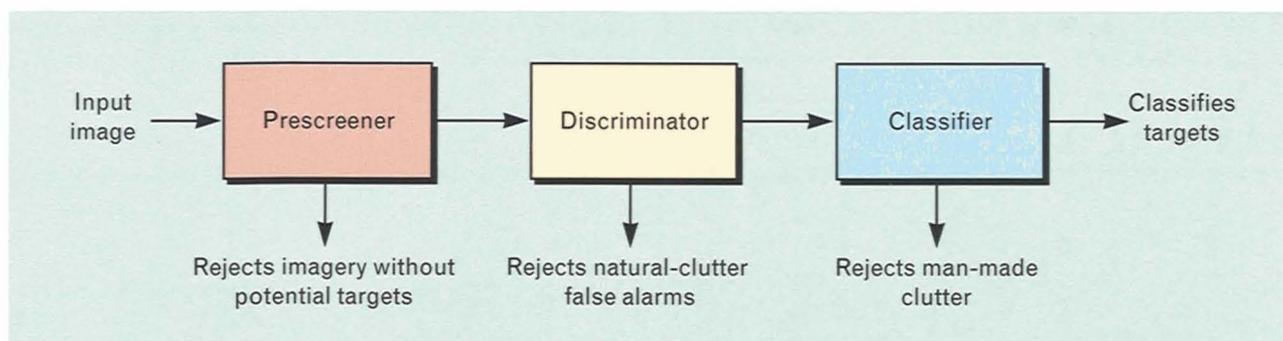


FIGURE 1. Block diagram of the multistage target-detection algorithm. This article concentrates on the discriminator stage.

one either as a target of interest (of which there can be a number of classes) or as an uninteresting man-made object.

In this article, we concentrate our attention on the second stage, the discriminator. The prescreener stage is covered elsewhere [2]; the classification stage is still under development.

Algorithm Description

The discrimination algorithm used in the Lincoln Laboratory automatic target-detection algorithm suite is centered around a one-class quadratic discriminator [3–5]. A one-class discriminator is trained only on a target-training set, and it assumes that the clutter false-alarm dataset (i.e., the set of false alarms passed by the prescreener stage) has unknown attributes in a feature space. Figure 2 illustrates the concept of this discrimination algorithm. For each region of interest, the algorithm produces a score that measures the distance from the candidate to the center of the target-training set (in a feature space). When the algorithm is properly trained, a lower value of this distance metric indicates a more targetlike candidate.

Key elements of the discrimination algorithm are the features used to compute the distance metric. We cover the discrimination features used in the Lincoln Laboratory target-detection algorithm suite in the next five sections of this article. Subsequent sections cover the discrimination algorithm itself in great detail.

Discrimination Features

A number of attributes that are present in the fully polarimetric, high-resolution SAR imagery can be exploited to discriminate between targets and clutter

false alarms. These attributes include size, shape, power, polarimetric properties, spatial distribution of reflected power, and dimensionality. Unfortunately, at present no method exists for developing discrimination features to exploit these attributes in any optimal fashion. The best that can be done is to design a feature that seems to exploit a specific attribute, and then test the feature on a variety of data to see if it separates targets from natural-clutter false alarms. If it does not separate targets from false alarms, the feature design is obviously poor; if it does separate them, then the feature *may* be a good one.

The other major criterion that a feature must satisfy is *orthogonality*. In simple terms, features used

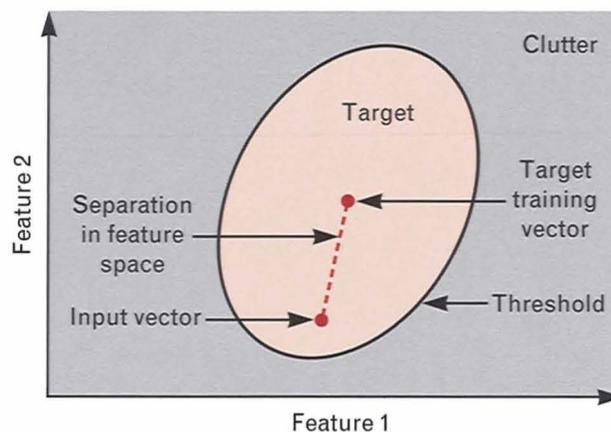


FIGURE 2. Conceptual diagram of a one-class discrimination algorithm. This diagram represents a two-dimensional feature space. If the separation in feature space is less than the threshold, then the region of interest from which the input vector is extracted is declared to be a target. Conversely, if the separation in feature space is greater than the threshold, the region of interest is declared to be clutter.

together in a discrimination algorithm must measure different attributes of the region of interest. For example, five different features that measure similar polarimetric properties of the candidate region of interest should not be used together in the same discrimination algorithm. In fact, using similar features is likely to make discrimination performance worse. This fact is a consequence of the likely occurrence that the target-training dataset will differ somewhat from the target-testing dataset. This phenomenon is covered more fully in the sidebar entitled "Adding Features Can Degrade Performance."

Another desirable property of the features in a discrimination algorithm is that they should be robust in a number of ways. Many feature algorithms require thresholds to be set to isolate the brightest scatterers, for example, or to isolate the scatterers with the most contrast. The feature values should not be too sensitive to the settings of these thresholds, because they may then work in one deployment situation but not in another similar situation. The features should also be somewhat robust to countermeasures; radar signatures of military vehicles are frequently altered by any number of methods. Some common methods include placing foliage and mud on the vehicle, adding metal parts to the vehicle, deploying camouflage netting around and on top of the vehicle, coating the target with radar-absorbing material, or simply opening the hatches of an armored target. An effective discrimination feature would ideally be insensitive to these methods and to other types of countermeasures.

We examined fifteen features for use in the Lincoln Laboratory target discrimination algorithm; three of these features were developed at Lincoln Laboratory [6], nine were developed at the Environmental Research Institute of Michigan (ERIM) of Ann Arbor, Michigan, two were developed by Rockwell International Corporation of El Segundo, California, and one was developed by Loral Defense Systems of Goodyear, Arizona. The non-Lincoln Laboratory features were developed under the Strategic Target Algorithm Research (STAR) contract, a yearlong research contract funded jointly by the Advanced Research Projects Agency (ARPA) and the United States Air Force. This contract was administered by Lincoln

Laboratory; the goal was to develop and test target detection algorithms by using a common dataset provided by Lincoln Laboratory. Each contractor's approach to the target-detection problem was somewhat different, but all used a number of features. We chose the most promising features for evaluation in this study.

Lincoln Laboratory Discrimination Features

The three Lincoln Laboratory discrimination features are standard deviation, fractal dimension, and weighted-rank fill ratio. They were developed by Leslie Novak, Michael Burl, and Gregory Owirka, all of Lincoln Laboratory. The features are computed from target-sized areas that are centered over the pixels identified by the prescener for further processing. The extent of the target-sized area is determined by the *a priori* knowledge of what type of target is being sought, and a box-spinning algorithm is used to determine target orientation [7].

The standard-deviation feature is computed from the typical estimator for the standard deviation. It uses the power (expressed in dB) of all the pixels in a target-sized box.

The fractal-dimension feature, which is illustrated in Figure 3, provides a measure of the spatial dimensionality of the potential target [6]. This feature estimates the Hausdorff dimension of the spatial distribution of the top N scatterers in the region of interest. For example, a straight line has a Hausdorff dimension of one, and a solid rectangle has a Hausdorff dimension of two. Various other space-filling objects with holes have a Hausdorff dimension that falls between one and two. An isolated point has a Hausdorff dimension of zero.

To compute the fractal-dimension feature, we threshold the region of interest by taking only the top N scatterers in terms of power. A binary image is created from these scatterers, and the minimum number n_1 of 1-pixel-by-1-pixel boxes ($d_1 = 1$) that cover all N scatterers is determined. This number, of course, is equal to the value N . Then the minimum number n_2 of 2-pixel-by-2-pixel boxes ($d_2 = 2$) that cover all N scatterers is determined. This number is less than or equal to N . If the spatial distribution of the scatterers is highly diffuse, the value of n_2 will be close to N ; if

ADDING FEATURES CAN DEGRADE PERFORMANCE

IN THE MAIN ARTICLE, a theoretical analysis of the one-class quadratic discriminator shows theoretical expressions for the probability of detection (P_d) and the probability of false alarm (P_{fa}) of the algorithm. We make the claim in the section on discrimination features that adding features does not necessarily improve discrimination performance. We show here that this is indeed the case by giving two examples; the first example shows discrimination performance with the set of five best features, and the second example shows discrimination performance with the same set of features in addition to seven other features.

The idea that adding features can degrade performance is, perhaps, counterintuitive. In fact, we cannot degrade performance by adding features if a few key conditions are met. These conditions are (1) the real data obey perfectly the multivariate Gaussian assumption made in the section entitled "Theoretical Analysis of the One-Class Quadratic Discrimination Algorithm," and (2) the target-training data and the target-testing data have exactly the same statistical distribution.

The multivariate Gaussian condition is difficult to verify for the real data, especially when a number of features are being consid-

ered (see the section entitled "Confirming the Gaussian Assumption"). Small departures from Gaussianity, however, are probably not the major cause of the phenomenon we are addressing here. Instead, the major cause of degraded performance is that the second condition is not being met.

Figure A shows performance curves for the 1-ft, polarimetric whitening filter (PWF) data using the features described in the section entitled "Best Features for Discrimination." This figure, which is the same diagram as Figure 17, also shows performance curves for the same dataset using

all twelve Lincoln Laboratory and ERIM STAR discrimination features. Notice the significant performance degradation that occurs when the seven extra features are added. Notes on interpreting this type of graph can be found in the sidebar entitled "Interpreting Plots of P_d versus FA/km^2 ."

Figure B is a notional diagram that illustrates the degradation phenomenon. The diagram is complicated but the explanation of it is relatively easy. There are three distinct sets of data displayed in the diagram: target training, target testing, and clutter false alarm. Each dataset is displayed for two features, which we call

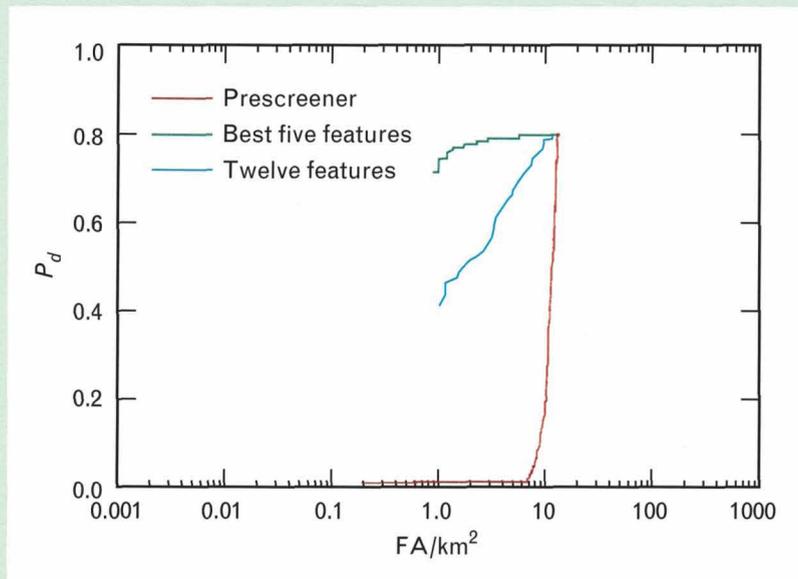


FIGURE A. Performance curves comparing discrimination performance for the five best features and for all twelve features. Performance degrades when more features are added.

Feature 1 and Feature 2. Imagine that the target-training data and the target-testing data are the same (training and testing is done on the same data, so ignore the red points on the diagram for the time being).

First, assume that the discrimination algorithm uses only Feature 1 (refer to only the green and black points plotted along the abscissa for this case). We see that the discriminator does a good job of separating targets from clutter by using Threshold A (three false alarms are called targets). The two key criteria are still obeyed. In this case the two key conditions are satisfied: the target-training data and target-testing data have ex-

actly the same statistical distribution, and the data obey (more or less) the Gaussian assumption (within each data type).

Now, while still assuming the target-training data and target-testing data are the same, we add Feature 2 (refer to only the green and black points in the middle of the graph). We can then draw Threshold B as an ellipse around the data, and do a still better job at separating targets from clutter (in this case one false alarm is called a target). The two key criteria are still obeyed.

If we assume now that the target-training data and the target-testing data are different datasets,

we remove the second key condition mentioned above. First, we assume that the discrimination algorithm uses only Feature 1 (refer to the green, red, and black points plotted along the abscissa). By drawing a threshold at the point labeled Threshold A, we see that the discrimination performance is equivalent to the previous case because the target-training data and the target-testing data for Feature 1 have similar characteristics.

Now, we add Feature 2. Performance is severely degraded because we cannot draw an ellipse (with the same orientation as Threshold B) around the center of mass of the green points that does not engulf large numbers of clutter false alarms (black points) while still engulfing the target-testing data (red points).

The important point is that any threshold ellipse must have the same orientation as the ellipse shown as Threshold B, because the orientation of the ellipse is determined by the statistical characteristics of the target-training data. This example is a particularly egregious illustration of the failure of the second key condition, because the target-training data and the target-testing data now have different statistical characteristics. The addition of Feature 2 obviously degrades performance severely. More subtle cases that significantly affect performance occur more frequently.

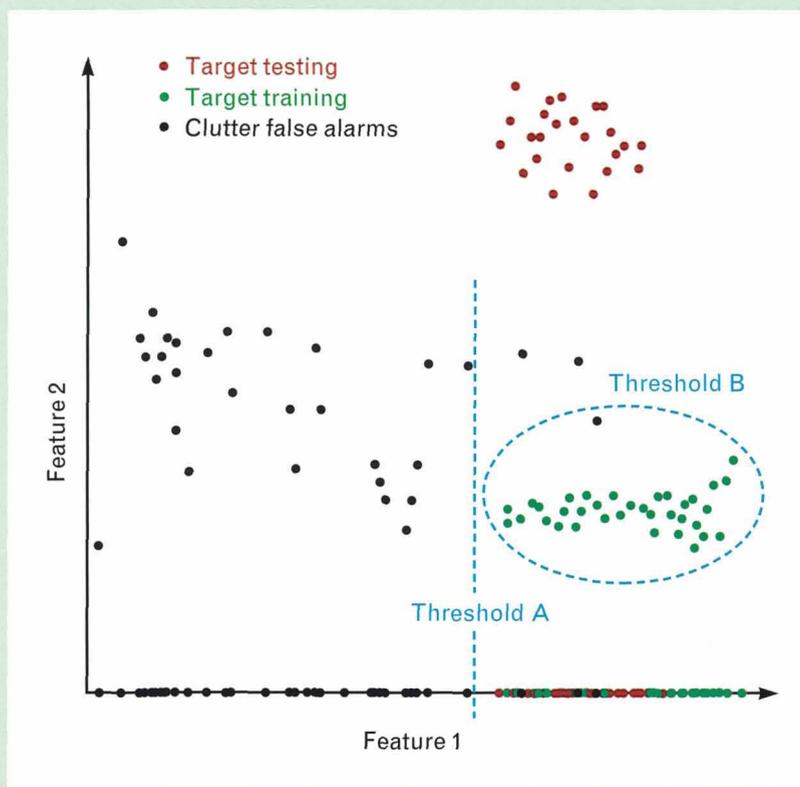


FIGURE B. Notional diagram of reason for performance degradation when features are added. This graph shows two features in a feature space.

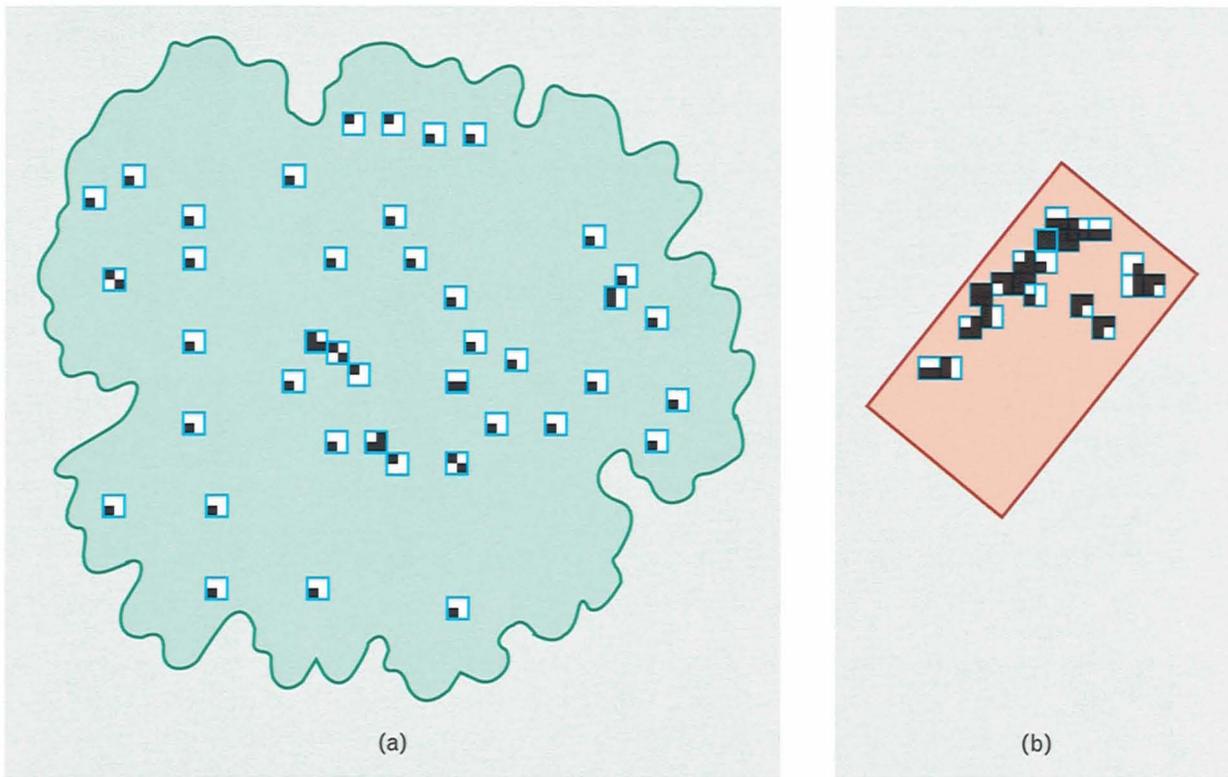


FIGURE 3. Calculation of the fractal-dimension feature, which measures the spatial bunching of the brightest pixels in a region of interest. (a) The brightest pixels for a tree tend to be widely separated, which requires a relatively large number of covering boxes and produces a low value for the fractal dimension. (b) The brightest pixels for a target tend to be closely bunched, which requires fewer covering boxes and produces a high value for the fractal dimension.

the scatterers are spatially bunched the value of n_2 will be considerably less than N . These values are determined for the two specific examples in Figure 3 and plotted in Figure 4, with the logarithm of n_1 and n_2 on the ordinate, and the logarithm of d_1 and d_2 on the abscissa. The negative slope of the line through the two points, which is given by

$$H_d = \frac{\log n_1 - \log n_2}{\log d_2 - \log d_1}, \quad (1)$$

is an estimate of the Hausdorff dimension of the region of interest. For the high-resolution data in this article, we used $N = 50$.

The weighted-rank fill-ratio feature is computed from the top N scatterers in the target-sized box. The feature is computed by totaling the power in the top N pixels within the target-sized box, and normalizing by the total power of all pixels in the box. This feature attempts to exploit the fact that power returns from

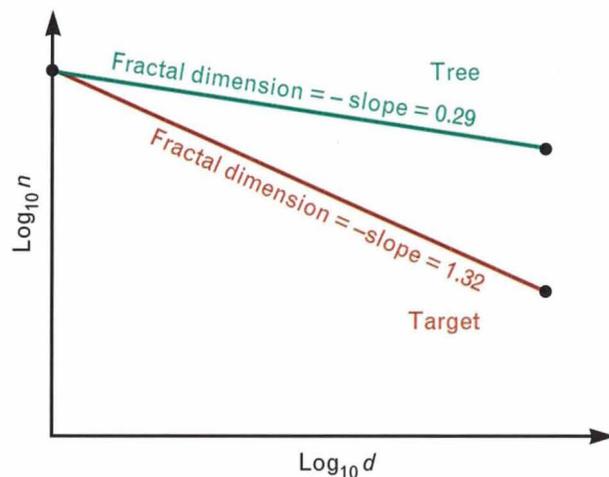


FIGURE 4. An estimate of the Hausdorff dimension of the tree and target in Figure 3. For both objects the total number of scatterers is 50; for the tree the minimum number of covering boxes is 41 and for the target the minimum number of covering boxes is 20. The negative slope of the line for each object is the estimate of the fractal-dimension feature. Targets tend to have higher fractal dimensions than natural clutter.

most targets tend to be concentrated in a few bright scatterers, whereas power returns from natural-clutter false alarms tend to be more diffuse. This feature measures a power-related property of the target-sized box, which makes this feature different from the fractal-dimension feature, which measures a spatial property of the entire region of interest.

ERIM Discrimination Features

The ERIM discrimination features were developed and provided to Lincoln Laboratory under the STAR contract mentioned above. They were modified for this study by altering the thresholds to account for a target dataset that was substantially different from the dataset used for the STAR contract. Instead of using a target-sized box as a preliminary step, as in the Lincoln Laboratory feature algorithms, the ERIM feature algorithms compute a target-shaped blob by performing morphological operations. These operations serve both as a method of grouping spatially related hits from the prescener and as a method of estimating the size, shape, and orientation of the supposed target.

There are three categories of ERIM discrimination features: size-related features, contrast-based features, and polarimetric features. Each of these three categories contains three features. The size-related features are mass, diameter, and square-normalized rotational inertia. The contrast-based features are maximum constant false-alarm rate (CFAR) statistic, mean CFAR statistic, and percent bright CFAR statistic. The polarimetric features are percent pure, percent pure even, and percent bright even. We describe each feature in detail in the following paragraphs.

The three size-related features utilize only the binary image created by the morphological operations. The mass feature is computed by counting the number of pixels in the morphological blob. The diameter is the length of the diagonal of the smallest rectangle (either horizontally oriented or vertically oriented) that encloses the blob. The square-normalized rotational inertia is the second mechanical moment of the blob around its center of mass, normalized by the inertia of an equal mass square.

The contrast-based features are determined by a CFAR algorithm. This algorithm can be described by

$$\frac{x - \hat{\mu}_c}{\hat{\sigma}_c}, \quad (2)$$

where x represents the test pixel, and $\hat{\mu}_c$ and $\hat{\sigma}_c$ are estimates of the local mean and local standard deviation, respectively, of the surrounding clutter. The estimates of the parameters from the surrounding clutter are accomplished by using the pixels in a window around the supposed target whose opening is large enough to exclude the target return. Figure 5 illustrates this window, and the opening is called the *guard area*.

The CFAR statistic given by Equation 2 is computed for each pixel to create a CFAR image. The maximum CFAR feature is the maximum value in the CFAR image contained within the target-shaped blob. This quantity is similar to the basic feature used in the prescener algorithm. The mean CFAR feature is the average of the CFAR image taken over the target-shaped blob. The percent bright CFAR feature is the percentage of pixels within the target-shaped blob that exceed a certain CFAR value.

The polarimetric discrimination features are based on a transformation of the linear polarization basis in which the Lincoln Laboratory MMW SAR gathers data to an even-bounce, odd-bounce basis described by the equations

$$\begin{aligned} E_{\text{odd}} &= \frac{|HH + VV|^2}{2} \\ &= 2|RL|^2 \end{aligned}$$

and

$$\begin{aligned} E_{\text{even}} &= \frac{|HH - VV|^2}{2} + 2|HV|^2 \\ &= |RR|^2 + |LL|^2. \end{aligned}$$

The odd-bounce channel given by the first equation corresponds to the radar return from a flat plate or a trihedral; the even-bounce channel corresponds to the radar return from a dihedral. Figure 6 illustrates examples of these reflectors, along with notional diagrams of how they reflect the radar energy. The use-

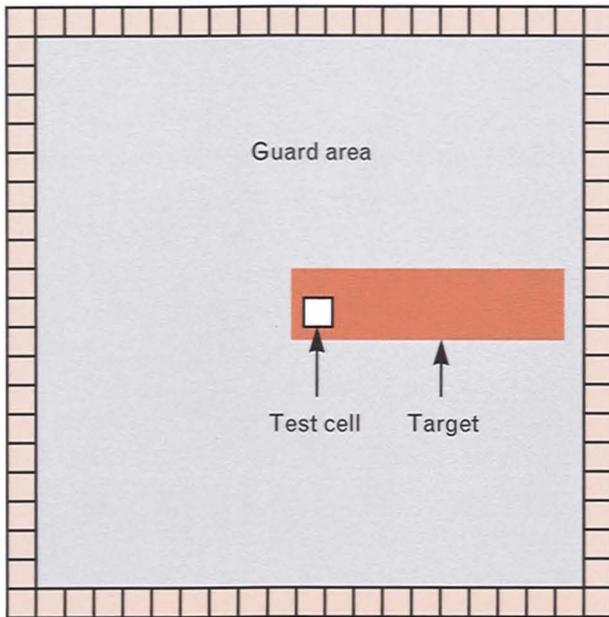


FIGURE 5. CFAR template, showing the pixel under test, and the surrounding window of pixels from which clutter estimates are computed. The test pixel and clutter window are separated by a guard area, which protects the clutter estimates from being corrupted by portions of the target return.

fulness of these polarimetric feature resides in the fact that few dihedral structures exist in natural clutter, but these structures are plentiful on most man-made targets. Natural clutter tends to exhibit more odd-bounce reflected energy than even-bounce reflected energy.

The ERIM polarimetric features are formed from the even-bounce and the odd-bounce images. The percent-pure feature is the fraction of pixels within the target-shaped blob for which at least a certain fraction of the scattered energy falls in either the even-bounce channel or the odd-bounce channel. Percent even is the fraction of pixels within the target-shaped blob for which at least a certain fraction of the scattered energy falls in the even-bounce channel. The percent-bright-even feature is the fraction of pixels that exceed a certain value in the CFAR image described above, and which are mainly even-bounce scatterers.

The main impetus for these features is that a man-made object exhibits approximately equal amounts of pure even-bounce energy and odd-bounce energy, whereas a natural-clutter false alarm is more likely to

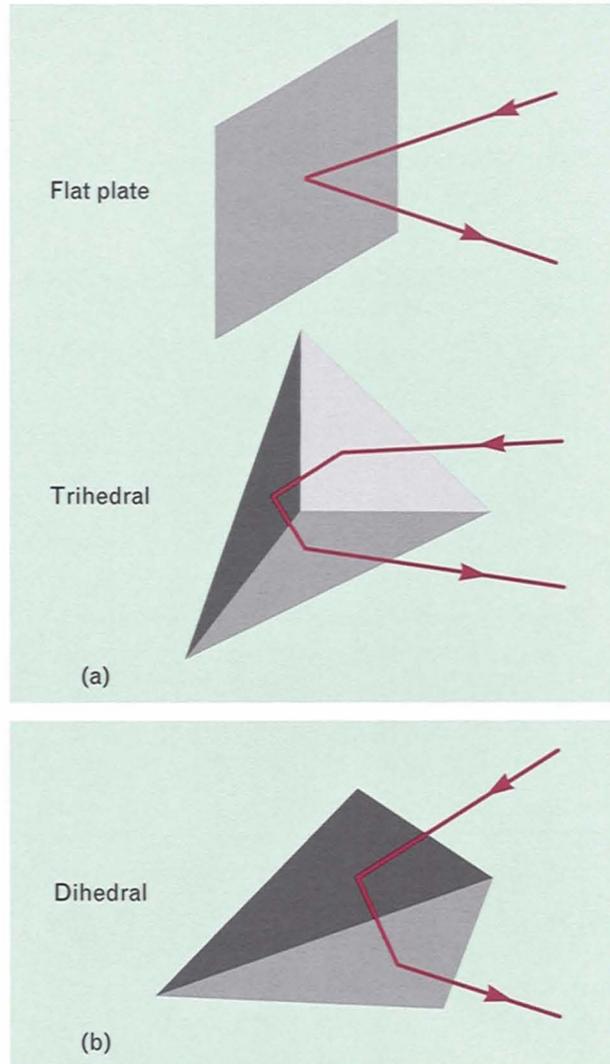


FIGURE 6. Reflection of radar signal from a variety of reflectors. (a) Odd-bounce reflectors include a flat plate and a trihedral. (b) Even-bounce reflectors include a dihedral. Radar backscatter from natural clutter is predominantly odd bounce, while backscatter from man-made objects is typically an equal mixture of even bounce and odd bounce.

exhibit large amounts of pure odd-bounce energy. Also, man-made objects are more likely to exhibit an equal mixture of even-bounce and odd-bounce energy than a natural-clutter false alarm.

Rockwell Discrimination Features

The Rockwell discrimination features were also developed and provided to Lincoln Laboratory under the STAR contract. Like the ERIM discrimination features, they were modified to account for the different type of target data used in the present study. These

features use the pixels in a target-sized box for feature calculation; the algorithm used to determine the orientation of the target-sized box is the same as that used for computing the orientation for the Lincoln Laboratory discrimination features described earlier. Some of the Rockwell discrimination features are similar to those already used by Lincoln Laboratory and ERIM. These similar features were not considered. Instead we concentrated on two other Rockwell features: (1) the polarimetric phase ratio feature, and (2) the specific-entropy feature.

The polarimetric phase ratio feature is another attempt to exploit differences in polarization between radar returns from targets and radar returns from clutter. The relative phase between the HH polarization channel and the VV polarization channel is used for this purpose. Only pixels in a target-sized box (the orientation of which is determined by the box-spinning algorithm described in Reference 7) are examined. In addition, to eliminate the low-return pixels that may have a random phase due to corruption by receiver noise, we use only the pixels that exceed a threshold in both the HH polarization channel and the VV polarization channel. This threshold, which is set to a percentage of the maximum power in a pixel in a given image, is set to a low value so the thresholding operation eliminates only the lowest return pixels whose phase is most likely to be corrupted by receiver noise.

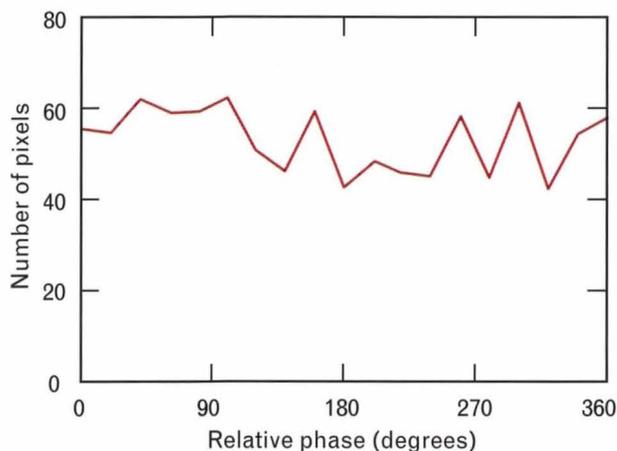


FIGURE 7. Histogram of relative phase between the HH polarization channel and the VV polarization channel. This type of plot is used in the calculation of the Rockwell polarimetric phase ratio feature.

The relative phase between the HH polarization channel and the VV polarization channel is then calculated for the remaining pixels within the target-sized box. The values are arranged in a histogram plot, such as the example shown in Figure 7, and the feature is calculated from this plot. The polarimetric phase ratio is defined as the number of pixels to fall within $\pm x^\circ$ of 180° relative phase on the histogram, divided by the number of pixels that fall within $\pm x^\circ$ of 0° relative phase on the histogram. We used a value of $x = 90^\circ$.

Specific entropy is the second Rockwell discrimination feature used in this study. Because of the complicated definition of this feature, it was not clear which step in the calculation provides the ability to separate targets from natural-clutter false alarms. To understand this feature better, we investigated it in considerable detail. A number of steps are involved in computing this feature:

1. Choose a threshold T that is set to the quantity corresponding to the 98th percentile of the surrounding clutter, and calculate a normalized amplitude by

$$a_i = \max(p_i - T, 0),$$

where a is the amplitude (in dB) above the threshold, p is the amplitude of the original pixel (in dB), i is the pixel tag number (of which there are m , which is the number of pixels in the target-sized box), and T is the value (in dB) of the threshold.

2. Normalize the amplitude by

$$f_i = \frac{a_i}{\sum_{i=1}^m a_i},$$

unless $a_i = 0$ for all $i = 1, \dots, m$, in which case the specific-entropy feature is set to zero.

3. Compute the specific-entropy feature by

$$\text{specific entropy} = - \sum_{i=1}^m \frac{f_i \log f_i}{\log m}.$$

The idea behind this feature is to exploit two supposed properties of a target: (1) the pixels exceeding the threshold T do not vary greatly in amplitude for a

target, but they do vary greatly for a natural-clutter false alarm; and (2) more pixels exceed the threshold T for a target than for a natural-clutter false alarm. The question that remains is which step in the feature calculation provides the separation between targets and natural-clutter false alarms.

We studied this problem by separating the calculation of the specific entropy into the steps described above, and then we calculated a feature based only on the operation in each separate step. To this end, we invented a simple *count* feature, which counts the number of pixels that exceeded the threshold T as it was calculated above, and normalizes this value by the total possible number of pixels in a target-sized box. This procedure was done for targets and for clutter false alarms, and the count feature was then plotted as a scatterplot versus the specific-entropy feature, as shown in Figure 8 for a sample target dataset.

In this plot, specific entropy is plotted on the

ordinate, and the logarithm of the count feature is plotted on the abscissa. If the two features are highly correlated, the plot shows the points falling along a straight line, which is indeed the case for this example. In fact, all the target and clutter false-alarm datasets that we examined showed similar scatterplots. For practical purposes, this scatterplot indicates that the count feature and the specific-entropy feature are equivalent. The extra steps given above in the calculation of the specific-entropy feature do little to increase the separation between targets and natural-clutter false alarms.

Loral Discrimination Feature

Loral Defense Systems was the third participant in the STAR contract. Many of their discrimination features substantially overlapped the features of the other two contractors and Lincoln Laboratory. For this reason, we examined only one Loral discrimina-

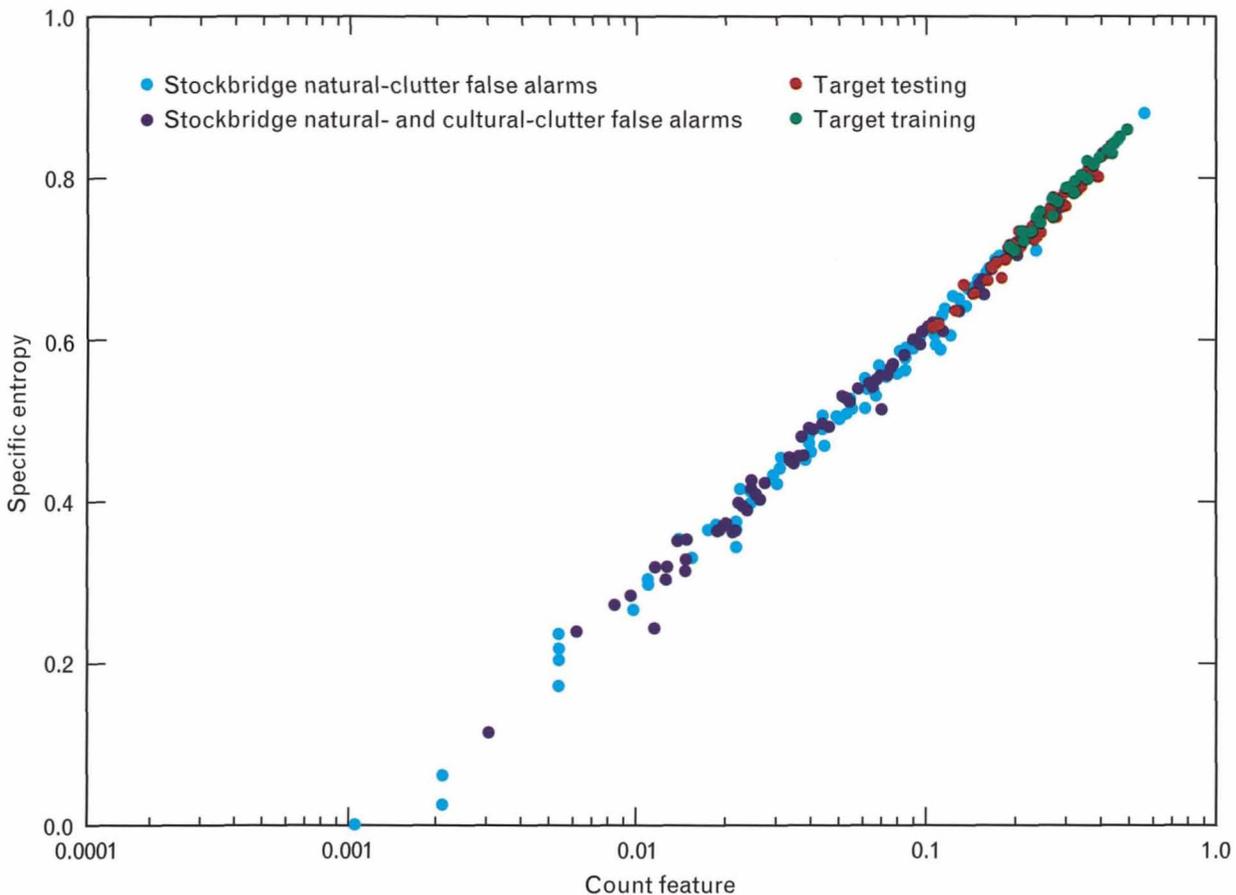


FIGURE 8. Scatterplot of the Rockwell specific-entropy feature versus count feature. Points falling on a straight line indicate a high correlation between the two features.

tion feature—the contiguousness feature.

To calculate this feature, we segment each image into three separate images based on the amplitude of individual pixels, as shown in Figure 9. All the pixels in an individual image are histogrammed and then divided into three categories. The lowest 25% of pixels are called *shadow* (Level 1), the middle 60% are called *background* (Level 2), and the top 15% are called the *target region* (Level 3). For the purposes of this study, we modified the procedure by applying the thresholding process only to each region of interest,

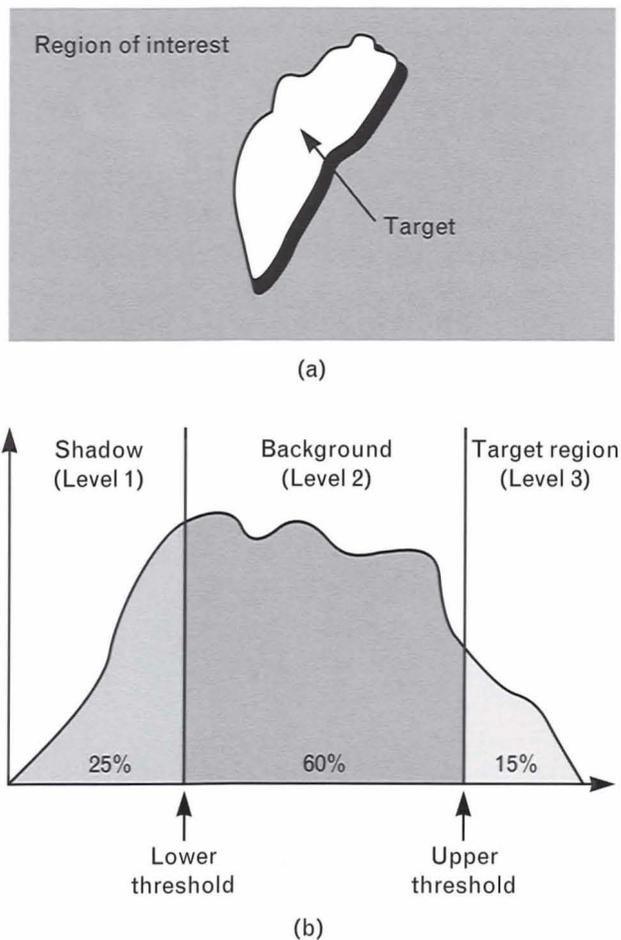


FIGURE 9. Concept of contiguousness feature. (a) The target within the region of interest has an irregular shape. The radar illuminates this shape from the top, which causes a shadow to extend downward in the image. (b) The two thresholds in the histogram of pixel power (in dB) divide the region of interest into three categories: shadow, background, and target region. The Loral contiguousness feature is computed by first forming six separate regions of interest based on these categories.

which consists of a target-sized area plus a border containing the surrounding clutter. Each region of interest was then segmented into the three categories by using only the clutter area surrounding each target candidate to determine the threshold levels.

The thresholding procedure effectively creates three regions of interest: one that contains the brightest pixels (called the Level 3 image), one that contains the dimmest pixels (called the Level 1 image), and one that contains the midlevel pixels (called the Level 2 image). The same thresholding procedure is performed on the CFAR image, which is derived in the same way as the CFAR image described in the section on ERIM discrimination features and determined by the expression in Equation 1. The contiguousness feature is determined by computing numbers from each of these six regions of interest.

The computation of the contiguousness feature is straightforward. Each contiguousness number is derived from only one image (i.e., the Level 1 CFAR image, the Level 2 CFAR image, the Level 3 CFAR image, the Level 1 image, the Level 2 image, or the Level 3 image). Each pixel included in each particular image is counted, and its immediate neighbors that appear in the same image are counted as well. The count is then normalized by the total number of possibilities that could have occurred (which is nine times the number of pixels in the image). The final number, which has a value between zero and one, is the contiguousness number for that image. This operation is done for every image, so the contiguousness feature gives six separate numbers for each region of interest.

Discrimination Algorithm

As mentioned at the beginning of the section on algorithm description, the Lincoln Laboratory discrimination algorithm is based on a one-class quadratic discrimination algorithm, the inputs of which are the feature vectors for each candidate region of interest. The algorithm is trained beforehand with representative target data only (no clutter data are used for the training, hence the one-class algorithm). Often these target data consist of images of targets with no countermeasures applied. The tests performed for this article use this training method.

The reasoning behind this type of training is that

predicting the use or modification of enemy targets is impossible, so training on exactly the same types of targets that will be encountered in a real situation is also impossible. Therefore, data gathered by using non-countermeasured targets seem a reasonable choice for a target-training dataset. Any realistic test of the discrimination algorithm, however, must include targets that have some countermeasures applied.

Theoretical Analysis of the One-Class Quadratic Discrimination Algorithm

The one-class quadratic discrimination algorithm used in the Lincoln Laboratory multistage target-detection algorithm can be described mathematically as

$$Z_i = \frac{1}{n} (\mathbf{X}_i - \hat{\mathbf{M}}_{tr})^T \hat{\mathbf{S}}_{tr}^{-1} (\mathbf{X}_i - \hat{\mathbf{M}}_{tr}), \quad (3)$$

for $i = 1, 2, \dots, k_t + k_c$,

where n is the number of features used in the discriminator, $\hat{\mathbf{M}}_{tr}$ and $\hat{\mathbf{S}}_{tr}$ are the estimates of the mean vector and variance-covariance matrix of the training target set, \mathbf{X}_i is a random vector representing the observed candidate features, and Z_i is a random variable representing the distance from the test point to the target-training class. The two variables k_t and k_c are the number of targets and the number of clutter false alarms, respectively, that the discriminator receives from the prescreeener stage of the multistage algorithm.

To analyze the discriminator given by Equation 3, we need to find the quantities

$$\text{Prob} \{ Z_i < K \mid i \text{ is target} \} = P_d \quad (4)$$

and

$$\text{Prob} \{ Z_i < K \mid i \text{ is clutter} \} = P_{fa}, \quad (5)$$

where K is the hard threshold. This analysis involves finding the probability distribution function (pdf) of Z_i for the target case and for the clutter false-alarm case, and then integrating the pdf according to Equations 4 and 5.

The distribution of Z_i for the target-training dataset is easy to calculate if the assumption is made that the estimates of \mathbf{M}_{tr} and \mathbf{S}_{tr} take on their true values. For tractability, we also assume that the features are multi-

variate Gaussian distributed. Then the Z_i values are chi-squared distributed [8], and the pdf can be written as

$$f_{tr}(z) = \frac{z^{\left(\frac{n}{2}-1\right)} \exp\left(\frac{-nz}{2}\right)}{\left(\frac{2}{n}\right)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)},$$

where $E(z) = 1$ and $\text{Var}(z) = 2/n$.

The distribution of Z_i under the target (non-training) and clutter false-alarm classes with these assumptions is more difficult to calculate. In each of the two cases, a different matrix \mathbf{A} must be found such that

$$\begin{aligned} \mathbf{A}_t^T \hat{\mathbf{S}}_{tr} \mathbf{A}_t &= \mathbf{I} \\ \mathbf{A}_t^T \mathbf{S}_t \mathbf{A}_t &= \mathbf{L}_t \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}_c^T \hat{\mathbf{S}}_{tr} \mathbf{A}_c &= \mathbf{I} \\ \mathbf{A}_c^T \mathbf{S}_c \mathbf{A}_c &= \mathbf{L}_c, \end{aligned}$$

where \mathbf{L}_t and \mathbf{L}_c are diagonal matrices. This operation, which is a simultaneous diagonalization, reduces the problem of evaluating Equations 4 and 5 to one of finding the distribution of

$$\frac{1}{n} \sum_{l=1}^n \lambda_{t,l} (X_l - \omega_{t,l})^2 \quad (6)$$

and

$$\frac{1}{n} \sum_{l=1}^n \lambda_{c,l} (X_l - \omega_{c,l})^2, \quad (7)$$

where

$$\begin{aligned} \lambda_{t,l} &= \text{diag}(\mathbf{L}_t)_l \\ \omega_{t,l} &= \mathbf{A}_t^T (\mathbf{M}_t - \hat{\mathbf{M}}_{tr})_l \end{aligned}$$

and

$$\begin{aligned} \lambda_{c,l} &= \text{diag}(\mathbf{L}_c)_l \\ \omega_{c,l} &= \mathbf{A}_c^T (\mathbf{M}_c - \hat{\mathbf{M}}_{tr})_l, \end{aligned}$$

and where the operator $\text{diag}(\cdot)$ indicates the extraction of the diagonal vector from the matrix argument. The quantity n is the number of features used in the discriminator.

Calculating the distribution of Equation 6 or Equation 7 without the multivariate Gaussian assumption would be difficult because the summation would then be over uncorrelated—but not necessarily independent—random variables. Once again, we make an assumption that the estimates of \mathbf{M}_{tr} and \mathbf{S}_{tr} take on their true values.

The characteristic function of the distribution of Equations 6 and 7 is given by

$$\phi_n(t) = \exp\left(\sum_{i=1}^n \frac{j t \omega_i^2}{1 - 2j t \lambda_i}\right) \prod_{b=1}^n (1 - 2j t \lambda_b)^{-\frac{1}{2}}.$$

In this equation, $j = \sqrt{-1}$. We can omit the target and clutter-false-alarm subscripts because the mathematics for the two cases is similar. This characteristic function can be inverted and integrated according to Equations 4 and 5 by using Fourier transform theory, so that

$$\begin{aligned} P_d(K) &= \int_0^{nK} f_t(z) dz \\ &= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \left[\frac{\Im[\phi_{t,n}(q)]}{q} \cos(nKq) \right. \\ &\quad \left. - \frac{\Re[\phi_{t,n}(q)]}{q} \sin(nKq) \right] dq \end{aligned}$$

and

$$\begin{aligned} P_{fa}(K) &= \int_0^{nK} f_c(z) dz \\ &= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \left[\frac{\Im[\phi_{c,n}(q)]}{q} \cos(nKq) \right. \\ &\quad \left. - \frac{\Re[\phi_{c,n}(q)]}{q} \sin(nKq) \right] dq. \end{aligned}$$

Confirming the Gaussian Assumption

We make a key assumption in the theoretical performance prediction for the discrimination algorithm given above—we assume that the datasets (target training, target testing, and clutter false alarm) are multivariate Gaussian distributed. There are two tests that are feasible to confirm this assumption; the first test is for the univariate case and the second test is for the bivariate case. Tests for higher dimensions exist, but they can become complicated and difficult to interpret [9] (or they rely on making some other crucial assumption, which can be difficult to check). We chose to perform the univariate and bivariate tests, for which we give the results here. We also created a scatterplot for a trivariate test.

All the tests described here were done as a check on algorithm performance and not as an end in themselves. We did not try to calculate exact quantitative measures for goodness of fit. An exact study would have added considerable complexity to our task, while providing little insight. Instead, our tests were done by using graphical techniques and the fits were performed by eye; only approximate Gaussianity can be ascertained by such techniques. The proof that the theory is an accurate predictor of performance is not contained in these tests, but rather in the comparison of real data results with theoretical results. This comparison is given in the section entitled “Real Data versus Theoretical Performance.”

The univariate test is straightforward. We plot each feature on Gaussian-scaled paper, and test it by examining if the cumulative density function is a straight line. In general, we found that most of the features for most of the datasets were adequately univariate Gaussian. In the few exceptional cases, the distributions were not far off, and the discrepancies were not significant in the final results. Figure 10 is an example of a univariate test with the fractal-dimension feature.

The bivariate case is tested by using scatterplots, which show data points of one feature versus another feature. For Gaussianity, these points should fall in an ellipsoidal bunch around the centroid of the data points. There should be more data points near the center of the ellipse, and fewer data points farther from the center of the ellipse. We could carefully and

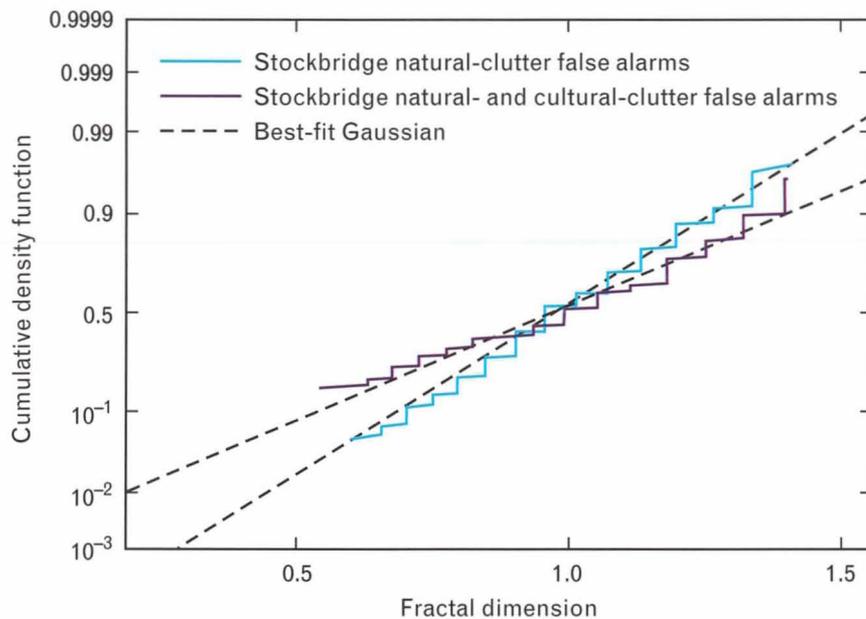


FIGURE 10. Univariate Gaussian test of the fractal-dimension feature. Any straight line on this graph represents a Gaussian curve.

quantitatively verify the percentage of points within a certain normalized radius of the center of mass, which would give a measure of how Gaussian the data were distributed. We did not, however, perform this quantitative verification; we merely observed how closely the data points were bunched around the center of mass.

These graphs are also useful for checking the correlation between two features; the more linear the data points are, the higher the correlation. If the data points fall in an ellipse that is horizontally or vertically oriented, then the data points are uncorrelated. This test is not just an interesting footnote; the section entitled “Feature Choice Guidelines” describes the importance of choosing features that are orthogonal (i.e., uncorrelated) for good discrimination performance. The scatterplots can also give additional insight into the ability of two features (taken simultaneously) to separate targets from clutter. Ideally, we would like the target-training dataset and the target-testing dataset to be coincident, and the clutter false-alarm dataset to be separated from the other two by a wide margin (measured both along the abscissa and the ordinate).

Figure 11 shows an example of a scatterplot for the fractal-dimension feature versus the weighted-rank fill-ratio feature. The target-training, target-testing,

natural-clutter false-alarm, and natural- and cultural-clutter false-alarm datasets are shown. The target-training and target-testing datasets seem to be reasonably elliptically distributed around their centers of mass, and they seem to be close to each other; both of these properties are desirable. The clutter false-alarm datasets seem to be somewhat less elliptical, but are reasonably well separated from both target datasets.

The three-dimensional scatterplot shown in Figure 12 illustrates all three Lincoln Laboratory discrimination features for the target-training, the target-testing, and the natural-clutter false-alarm datasets described in the section entitled “Data Used.” There are two things to be noticed about this figure. First, the figure clearly shows the separation between targets and natural-clutter false alarms, and it shows that the clutter false alarms intermingled with the target datasets tend to be those created by man-made objects (i.e., cultural clutter). Second, the figure helps confirm the approximate Gaussianity of the target datasets and the natural-clutter false-alarm dataset.

Notice the distribution of the red points (the target-training dataset) in the figure. If these red points are Gaussian distributed, they should form an ellipsoidal pattern around the center of the red point cloud with greater density of points toward the center. Likewise, the dark blue points (the target-testing

dataset), the green points (the natural-clutter false-alarm dataset), and the light blue points (the cultural-clutter false-alarm dataset) in the figure should be distributed in a similar manner for Gaussianity to hold.

Figure 12 shows that the red points and the dark blue points are distributed in an approximately ellipsoidal pattern around their respective centers. The green points, however, are less ellipsoidal in their distribution, and the light blue points are clearly non-ellipsoidal. As we demonstrate in a later section, the minor deviation of the green points (i.e., the natural-clutter false alarms) from Gaussianity does not greatly affect the agreement between the theory and the real data. The lack of Gaussianity in the light blue points (i.e., the cultural-clutter false alarms) is not critical because the discriminator is designed to eliminate the

natural-clutter false alarms and pass the cultural-clutter false alarms to the classification algorithm.

The goal of the discrimination algorithm, as stated earlier in this article, is to reject false alarms caused by natural clutter. For most of this article, we do not distinguish between clutter false alarms caused by natural clutter and clutter false alarms caused by cultural clutter, because it is impossible to know, in any kind of realistic scenario, which type of clutter false alarm a given region of interest is (or even if the region of interest is a clutter false alarm or a legitimate target).

In Figure 12 we separate the two types of false alarms for analysis purposes. For the discrimination algorithm to perform well, the targets must be separated from the natural-clutter false alarms. Figure 12 shows that the targets are indeed separated from the natural-clutter false alarms but not from the cultural-

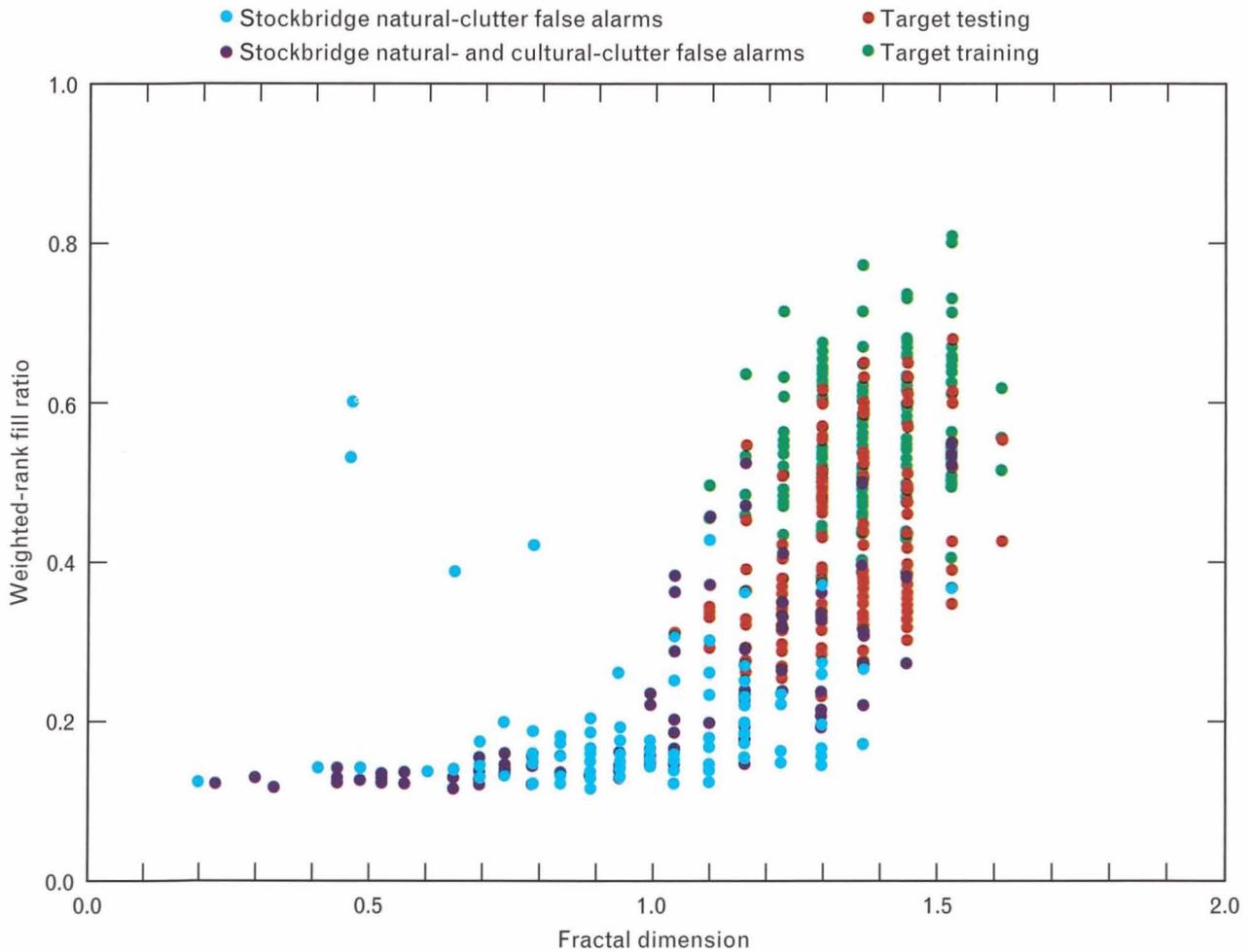


FIGURE 11. Scatterplot of fractal-dimension feature versus weighted-rank fill-ratio feature. For good discrimination performance with these two features, the target datasets should be separate from the false-alarm datasets.

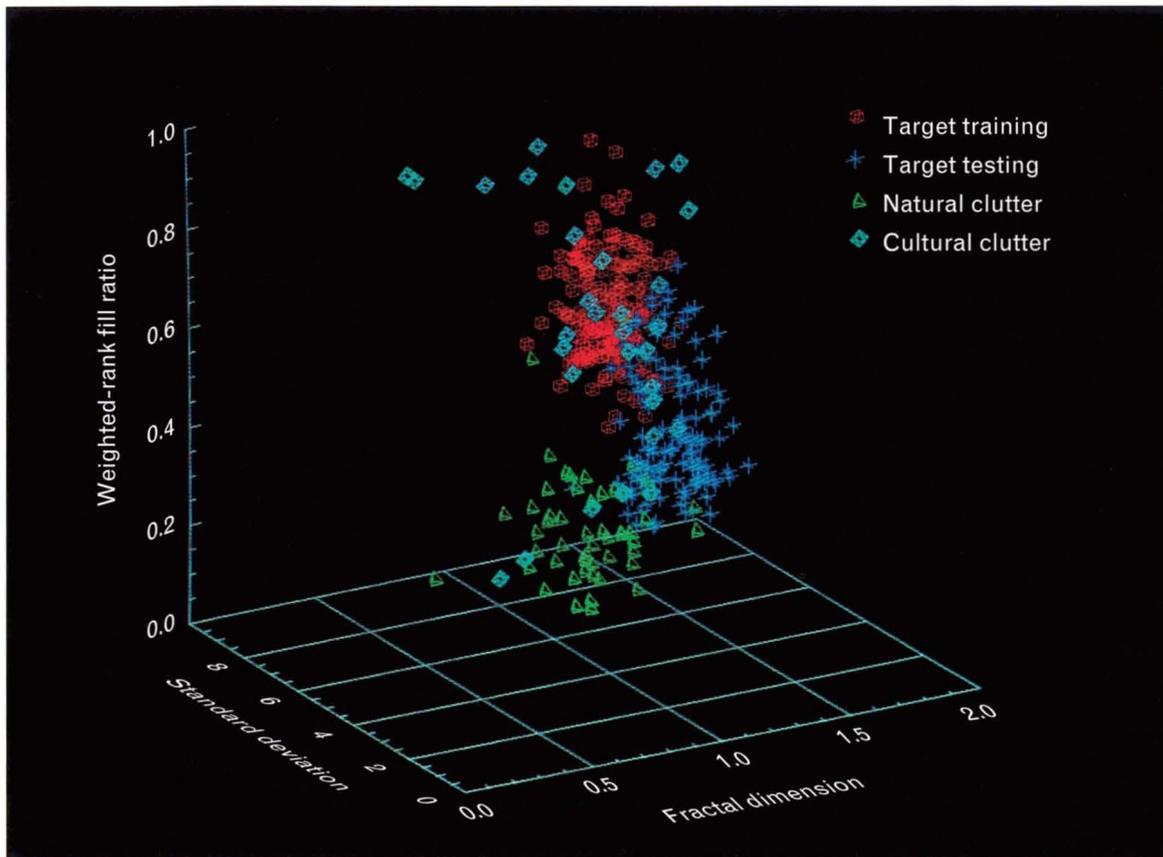


FIGURE 12. Three-dimensional scatterplot of the three Lincoln Laboratory discrimination features: fractal dimension, weighted-rank fill ratio, and standard deviation.

clutter false alarms. In other words, the discrimination algorithm does well in the role for which it was designed.

Goals

An important goal for this study is to choose the best set of features from the discrimination feature list given above in the section on discrimination features. As we already stated, these features are standard deviation, fractal dimension, weighted-rank fill ratio, mass, diameter, square-normalized rotational inertia, maximum CFAR statistic, mean CFAR statistic, percent bright CFAR statistic, percent pure, percent pure even, percent bright even, polarimetric phase ratio, specific entropy, and contiguousness. What determines a best set of these features? Initially we do not know whether the best set contains some combination of three features or four features or more, or all fifteen features, or even if the best set is composed of certain of these features and not others.

An additional goal is that the features chosen must be a robust set of features. Ideally, the features should work equally well, regardless of the target deployment, the countermeasures used, or the type of clutter being imaged. This invariance to data is impossible to achieve; a more realistic goal is for the same features to be part of the best feature set, regardless of data. As a test of this goal, we examine a number of different datasets.

Another goal is to understand the operation of the discrimination algorithm. We would like the theoretical expressions for the performance of the one-class quadratic discrimination algorithm to predict the behavior of the discrimination algorithm accurately. If the theory is accurate, then we can predict the performance of the algorithm for different combinations of features. A strategy for feature selection would then be to derive the parameters needed for the theory from the real data, compute the theoretical results for all combinations of features, and choose the best set

of features based on these predictions. We shall see that this strategy is a reasonable one.

We would also like to examine discrimination results for different resolutions and different polarizations. The Lincoln Laboratory MMW SAR has a resolution of 1 ft by 1 ft, but we can easily construct lower-resolution, single-look radar imagery from the data at hand. Additionally, the Lincoln Laboratory MMW SAR is fully polarimetric (fully polarimetric SAR data allow synthesis of any polarization or combination of polarizations). Many radar sensors are not fully polarimetric, and use only a single polarization. The most common polarization used is HH; therefore, we use HH data as well as fully polarimetric data in this study. We expect the best feature set to change, depending on our choice of resolution and polarization.

Choosing a Feature Set

The method we use to choose the best feature set is straightforward. First, the data are prescreened by using a simple two-parameter CFAR algorithm [2, 10]. This stage is designed to eliminate (with a minimum of computation) only the most obviously non-targetlike clutter. The prescreening algorithm operates on imagery that has already been reduced to a resolution of 1 m by 1 m. The resolution was reduced by taking a noncoherent average of each 4-pixel-by-4 pixel non-overlapping box (a pixel has a nominal resolution of approximately 0.23 m). This method of resolution reduction has two advantages: (1) it reduces the amount of data we need to process, and (2) it reduces the speckle that is present in the high-resolution SAR imagery (the article by Leslie M. Novak et al. in this issue gives an explanation of speckle in SAR imagery).

In the prescreener for this study we use a threshold value that allows the detection of 80% of the targets. This percentage was chosen for consistency among datasets; it was also chosen by considering the number of clutter chips that are passed to the discrimination stage. A higher probability of detection in the prescreener stage necessarily increases the number of clutter false alarms passed to the discriminator. Computation time and storage limitations preclude using a higher percentage value for the prescreener prob-

ability of detection. The data used in the prescreener algorithm were also processed by using the polarimetric whitening filter (PWF) [1], which combines the HH, HV, and VV polarization channels together in a manner that optimally decreases speckle. The HH polarization results use only the HH polarization SAR data, and hence do not use the PWF imagery for the prescreener algorithm.

The candidates identified by the prescreener (either on targets or on clutter false alarms) are then grouped spatially. The grouping algorithm is a simple one; all hits within a target-sized area are grouped into a single detection. This grouping operation exploits some of the spatial information inherent in the proximity of prescreener hits.

The discrimination algorithm is run on all the regions of interest selected by the prescreener and the grouping algorithm. First, all the features described in the section on discrimination features are computed for all regions of interest. The features are computed for the following four combinations of data: (1) 1-ft resolution and PWF polarization, (2) 1-m resolution and PWF polarization, (3) 1-ft resolution and HH polarization, and (4) 1-m resolution and HH polarization. The features were originally tuned (in terms of the thresholds used in the feature calculations themselves) for the 1-ft resolution, PWF case. The features are used without modification for this case as well as for the 1-ft resolution, HH polarization case.

Naturally, the polarimetric features cannot be calculated for the HH polarization case because the polarimetric features use polarizations other than HH. We therefore use a reduced set of features. For the 1-m resolution cases, we retune the features by computing them for a range of thresholds, and we choose the threshold that provides the best separation between targets and clutter false alarms for all datasets. This retuning is done separately for the PWF case and the HH polarization case. Therefore, these features are intended to give best-case results. Any use of these tuned features in other datasets can only approach the results shown in the article in general. Certainly, the 1-m resolution tests provide a better indication of the performance of the discrimination algorithm than is likely to be obtained in a real situation.

The parameters necessary for a theoretical evalua-

INTERPRETING PLOTS OF P_D VERSUS FA/KM^2

THE METHOD OF EVALUATION for the discrimination algorithm described in this paper involves plotting a curve that shows the probability of detection (P_d) versus the number of false alarms per square kilometer (FA/km^2). The measure of FA/km^2 scales directly to the probability of false alarm, which was theoretically derived for the discrimination algorithm in the section entitled "Theoretical Analysis of the One-Class Quadratic Discrimination Algorithm." Such curves are often referred to as receiver operating-characteristic (ROC) curves.

Figure A gives an example of a simple ROC curve (in red). Better performance is indicated in these types of plots by a curve moving upward and leftward. A plot such as this one might be used to evaluate the prescreener stage or the discrimination stage separately. A more complicated plot is necessary to evaluate the combination of the prescreener and discrimination stages.

Figure A also shows an example of a plot (blue and red) that might be used to evaluate both the prescreener and discrimination stages combined. Notice that the original ROC has grown a

number of extra lines, or *tails*. These extra lines (in blue) represent the improved performance provided by the discrimination stage of the multistage target-detection algorithm. Each extra line meets the curve of the original prescreener stage at a certain point. Each discrimination line emanating from these points describes the operating characteristic of the discrimination algorithm

for the particular set of inputs being provided by the prescreener at that operating point.

As the threshold of the prescreener is varied, the set of inputs provided to the discrimination algorithm varies as well. The evaluation criterion for performance in an ROC curve works here as well; the line moving upward and leftward indicates better performance.

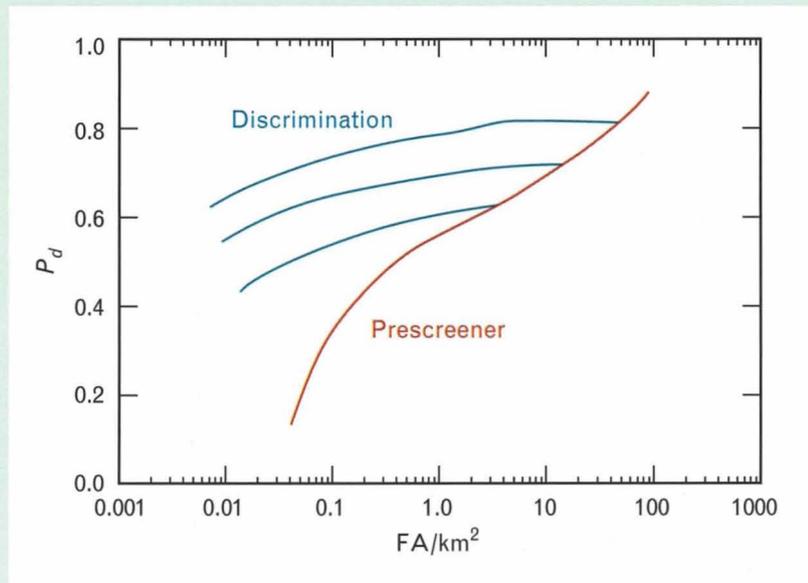


FIGURE A. Example of a P_d versus FA/km^2 curve, which is also known as a receiver operating-characteristic curve, for a multistage target-detection algorithm. The additional lines represent the performance of the discrimination stage of the algorithm. Three of these performance lines are shown; in fact, an infinite number of them are possible, because their intersections with the prescreener curve are dictated by the level at which the prescreener stage is operated.

tion of the discrimination performance are computed in each case from a target region-of-interest dataset and a clutter false-alarm region-of-interest dataset. Additionally, the assumptions necessary for the theory

to hold are checked in most cases. These checks are more fully detailed in the section entitled "Confirming the Gaussian Assumption." All combinations of the discrimination features are tested by using the

theory. Theoretical P_d versus P_{fa} plots are produced for each combination, and the best combinations are chosen for further analysis. The best combinations of features are given in the section entitled “Best Features for Discrimination.” The results we get by using the real data are then generated for the short list of good combinations. The real-data results are then compared with the theoretical results; these comparisons are given in the section entitled “Real Data versus Theoretical Performance.”

Performance evaluation is done by plotting the probability of detection P_d versus the number of false alarms per square kilometer (FA/km^2). The measure of false alarms per square kilometer is merely a rescaling of the probability of false alarm (P_{fa}) into an operationally meaningful measure. This rescaling is performed to remove the effect of sensor resolution (because a higher resolution image inherently gives more opportunities for false alarms to occur, the same P_{fa} value at different resolutions means different numbers of false alarms per square kilometer). The inter-

pretation of plots of P_d versus FA/km^2 is reviewed in the sidebar entitled “Interpreting Plots of P_d versus FA/km^2 .”

Data Used

All the target data used in this study were gathered with the Lincoln Laboratory MMW SAR in Stockbridge, New York. The targets consisted of two datasets of the same targets in different deployment conditions. The first dataset, which we use for discrimination algorithm training, is called the target-training dataset. The second dataset, which we use for discrimination algorithm testing, is called the target-testing dataset. There are three distinct clutter datasets; two gathered at Stockbridge, New York, and a smaller clutter dataset gathered in Concord, Massachusetts.

The first clutter dataset, which consists of mostly natural clutter, is called the Stockbridge natural-clutter dataset. Figure 13 is an example of this dataset; it shows a river with treelined banks (the river is the dark area curving through the middle of the image).



FIGURE 13. SAR image of natural clutter in Stockbridge, New York. The sensor is flying parallel to the top of the image, and the shadows extend downward in the image. Areas of high radar return are colored in bright yellow; areas of low radar return are in dark colors. The dark band in the middle of the image is a river with trees lining each bank. The smooth green areas are open fields.

The remainder of the image is an open field. Freshly plowed furrows in the open field can also be seen. The radar illuminates the area from the top of the image; therefore, the shadows cast by the trees point downward in the image. The Stockbridge natural-clutter dataset also includes some man-made objects (which are impossible to avoid entirely in the Stockbridge area), including the farmhouse shown in an earlier article by Novak [1]. The clutter in this dataset is considered to be relatively benign.

The second clutter dataset is called the Stockbridge natural- and cultural-clutter dataset. This dataset was gathered from a different area of the same Stockbridge collection site; it includes a farm-supply store that is shown in Figure 14 both as a SAR image and in an aerial photograph. The clutter in this dataset is considered to be moderately difficult.

The third clutter dataset is a small dataset gathered in Concord, Massachusetts, which is a few miles from Lincoln Laboratory. This dataset, which we refer to as the Concord dataset, consists entirely of man-made clutter, and is considered to be a very difficult dataset.

Figure 15 shows an example of imagery from this dataset.

Best Features for Discrimination

The method used to determine the best features for discrimination is fully described in the earlier section entitled “Choosing a Feature Set.” For two cases—the 1-ft resolution, PWF data of the Stockbridge natural-clutter dataset and the Stockbridge natural- and cultural-clutter dataset—we found the best features to be those given in Table 1. For the case of the 1-ft resolution, PWF, Concord man-made clutter, the feature set reduced to those features given in Table 2.

As stated earlier, we did not attempt to pick the best features for the 1-ft resolution, HH-polarization case. Instead we evaluated performance with the same features as the best-case features for the PWF data.

For the 1-m resolution, PWF, Stockbridge natural-clutter, and the natural- and cultural-clutter case, we found the best features to be those given in Table 3. The “optional” qualifier given in the table means that the feature does not increase or decrease any perfor-



FIGURE 14. (a) An optical photograph and (b) a SAR image of a farm-supply store in Stockbridge, New York. This store is an example of a man-made clutter discrete. The store parking lot is in the bottom of each image. Although the photograph and the SAR image were taken at different times, passenger cars can be seen in the parking lot in both images. The bright spots in the middle right area of the SAR image are caused by various metallic objects in the yard of the supply store.

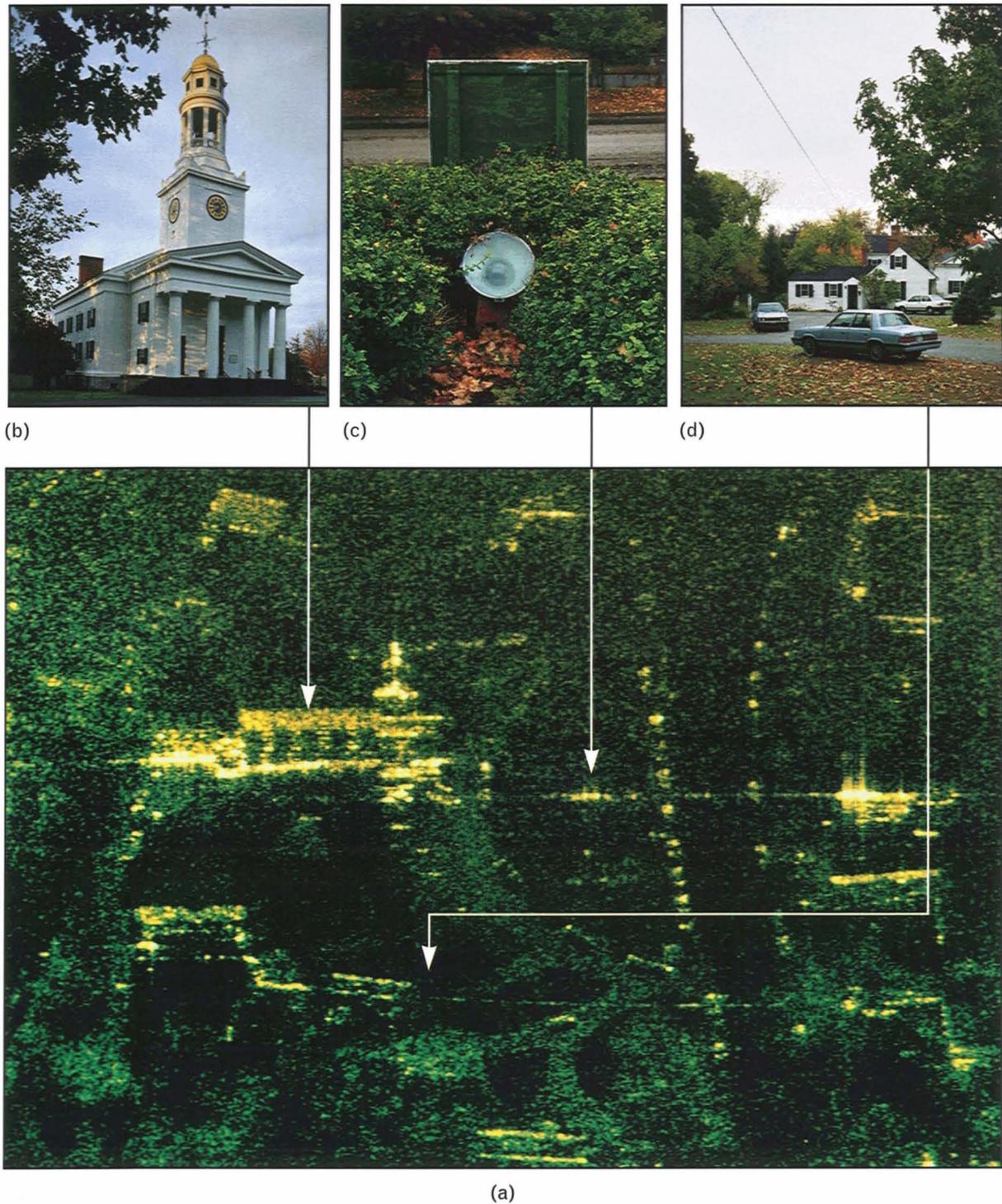


FIGURE 15. (a) A SAR image of man-made clutter in Concord, Massachusetts. The other three photographs illustrate specific objects visible in the SAR image: (b) the church and steeple, (c) a spotlight that illuminates the church at night, and (d) a house and a telephone wire suspended overhead. Note the bright columns along the side of the church in the SAR image. These columns clearly correspond in number and placement to the areas between the windows of the church in the optical photograph. Also notice the bright circular feature—the clock—on the church steeple.

Table 1. Best Features for 1-ft, PWF, Natural-Clutter Dataset and Natural- and Cultural-Clutter Dataset

<i>Feature</i>	<i>Description</i>
Fractal dimension	Lincoln Laboratory
Weighted-rank fill ratio	Lincoln Laboratory
Diameter	ERIM size
Mean CFAR or percent bright CFAR	ERIM CFAR
Percent pure	ERIM polarimetric

Table 2. Best Features for 1-ft, PWF, Man-Made Clutter Dataset

<i>Feature</i>	<i>Description</i>
Fractal dimension	Lincoln Laboratory
Percent bright CFAR	ERIM CFAR
Percent pure	ERIM polarimetric

Table 3. Best Features for 1-m, PWF, Natural-Clutter Dataset and Natural- and Cultural-Clutter Dataset

<i>Feature</i>	<i>Description</i>
Fractal dimension	Lincoln Laboratory
Diameter	ERIM size
Percent bright even	ERIM polarimetric
Percent pure	ERIM polarimetric
Mean CFAR	ERIM CFAR (optional)

mance ability with these datasets, but it could add or subtract a certain amount of robustness for other datasets. The best features for the 1-m resolution, HH-polarization case are given in Table 4. For this case, however, the discrimination algorithm provides

Table 4. Best Features for 1-m, HH, Natural-Clutter Dataset and Natural- and Cultural-Clutter Dataset

<i>Feature</i>	<i>Description</i>
Fractal dimension	Lincoln Laboratory
Diameter	ERIM size
Mean CFAR	ERIM CFAR (optional)

little or no performance gain over the prescreener alone.

Feature Choice Guidelines

An examination of the list of best features from the previous section, along with the scatterplots shown in the section entitled “Confirming the Gaussian Assumption,” reveals some interesting and important guidelines for choosing the best features. There are two general criteria for feature choice for this discrimination algorithm—*separation* and *orthogonality*. The separation criterion is the common-sense consideration that the feature must adequately separate the target training (and target testing) dataset from the clutter false-alarm dataset. The orthogonality criterion is less intuitive, and can be summarized by the idea that different features used in the discrimination algorithm must measure different attributes of the region of interest.

Unfortunately, we cannot easily predict exactly which attribute of a region of interest a feature measures. Sometimes two features that beforehand would seem to be highly correlated ultimately exhibit a low degree of correlation. We show an example of this type of behavior later in this section.

The best features listed in Table 1 are a good example of the orthogonality criterion. We see that the table includes two of the three Lincoln Laboratory discrimination features, which is not surprising because the three Lincoln Laboratory features were designed with orthogonality in mind. The first feature (fractal dimension) exploits the spatial relationship of the top N scatterers in the region of interest, while the second feature (weighted-rank fill ratio) exploits the

distribution of reflected power among all the scatterers on the target. Clearly, these two features were designed to measure different characteristics of the region of interest.

The three other features included as best features in the case shown in Table 1 all come from the ERIM discrimination features. Interestingly, the three chosen features each come from a different subset of features; the first comes from the ERIM size features, the second comes from the ERIM CFAR features, and the third comes from the ERIM polarimetric features. Even if the ERIM features were not designed with the orthogonality criterion in mind, we find it interesting that the choice of best features naturally selects one feature from each category.

A subset of the features listed in Table 1 works best in the man-made clutter dataset, as shown in Table 2.

The orthogonality criterion holds here as well, except that the two features not included, which were included in Table 1, no longer provide reasonable separation between targets and clutter false alarms.

For the case of 1-m resolution, there is an apparent exception to the two criteria given above in the best feature choices. Notice that Table 3 contains two ERIM polarimetric features. Figure 16 shows a scatterplot of these two features (percent bright even and percent pure) for the 1-m resolution datasets. From the scatterplot we can see that these two features are, in fact, uncorrelated and are therefore orthogonal in some meaningful sense. Apparently, in the 1-m resolution dataset the thresholding involved in calculating the percent-bright-even feature causes this feature to measure something other than the polarimetric properties of the region of interest. The

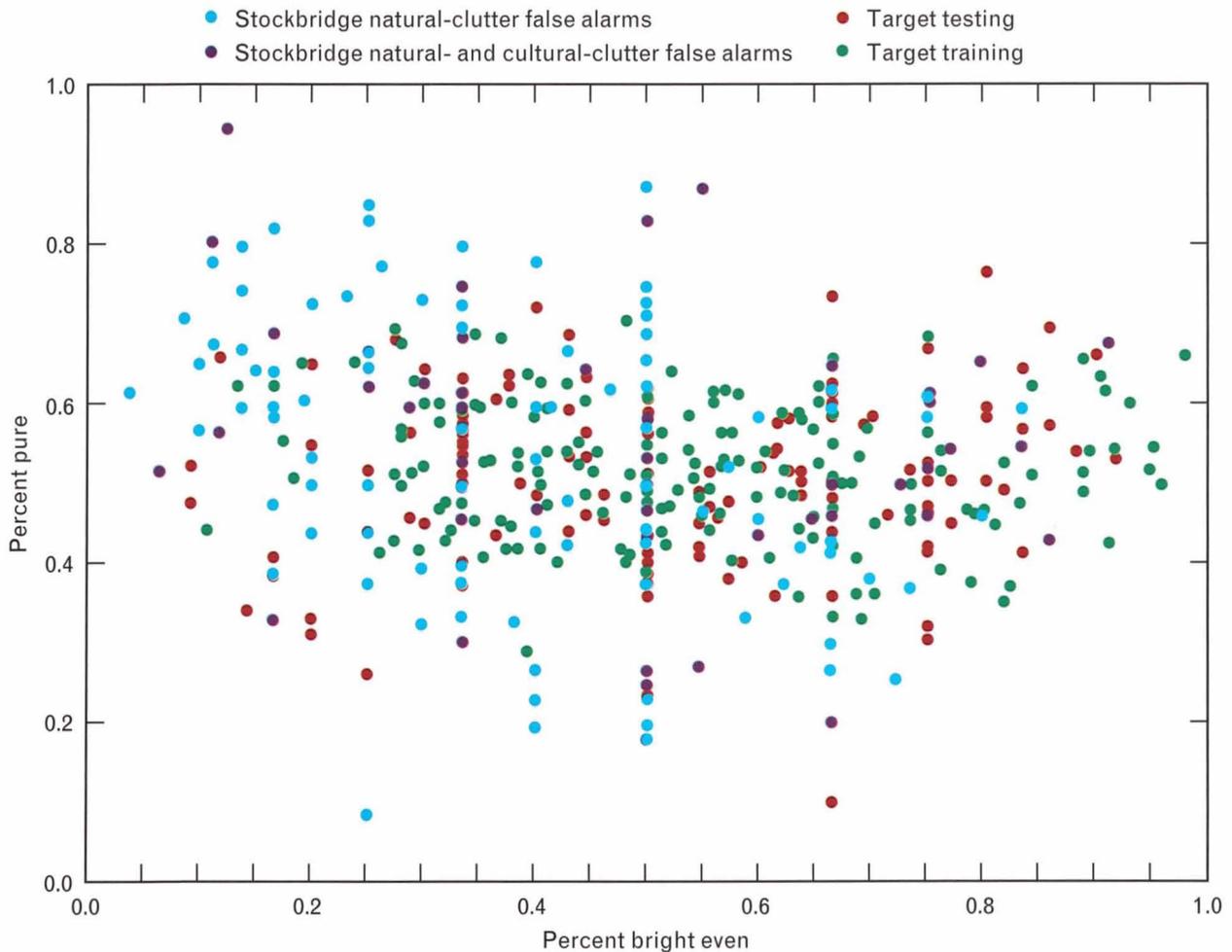


FIGURE 16. Scatterplot of percent bright even feature versus percent pure feature for 1-m-by-1-m resolution data. These two features are uncorrelated because the data points do not fall along a straight line.

five features chosen in Table 3 for the 1-m resolution case obey both the separation and orthogonality criteria given above. The same holds true for the best features in the HH-polarization case, which are given in Table 4.

Real Data versus Theoretical Performance

This section gives the real-data prescreener and discrimination results in the form of plots of P_d versus FA/km^2 . These plots are explained in the sidebar entitled "Interpreting Plots of P_d versus FA/km^2 ." We also plot on the same graphs the predictions computed from the theoretical analysis given in the section entitled "Theoretical Analysis of the One-Class Quadratic Discrimination Algorithm." In all cases, the theory and real data coincide closely. This fact demonstrates that the one-class quadratic discrimination algorithm is well understood as it is implemented in the Lincoln Laboratory multistage target-detection algorithm.

Figure 17 gives the combined prescreener and discrimination results for the 1-ft resolution, PWF data for the Stockbridge natural-clutter dataset, while Figure 18 gives the prescreener and discrimination results for the Stockbridge natural- and cultural-clutter dataset, and Figure 19 gives the prescreener and discrimination results for the Concord man-made-clutter dataset. Figures 20 and 21 show the prescreener and discrimination results for the Stockbridge natural-clutter dataset and the Stockbridge natural- and cultural-clutter dataset, respectively. Both results are for 1-ft resolution, HH-polarization data.

The remaining results are for the 1-m resolution case. Figures 22 and 23 show the prescreener and discrimination results for the Stockbridge natural-clutter dataset and the Stockbridge natural and cultural-clutter dataset, respectively, for PWF data. Figures 24 and 25 show the prescreener and discrimination results for the same two datasets for the HH-polarization case.

Polarization Comparisons

We can compare the discrimination results from the PWF data and the HH-polarization data for the same cases to draw a conclusion regarding the advantage of using a fully polarimetric radar versus using the more

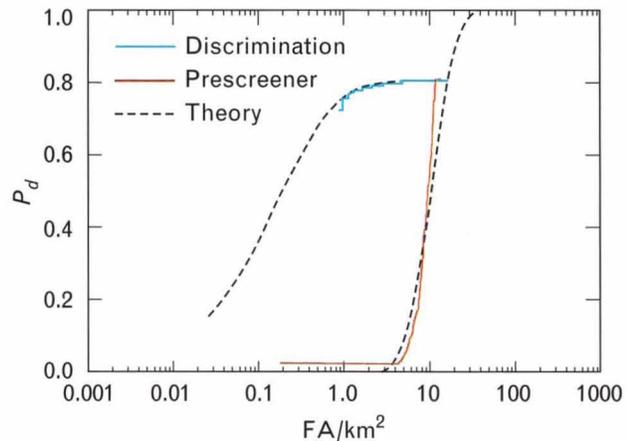


FIGURE 17. Comparison of real data and theoretical results for the 1-ft-by-1-ft resolution, PWF, Stockbridge natural-clutter case.

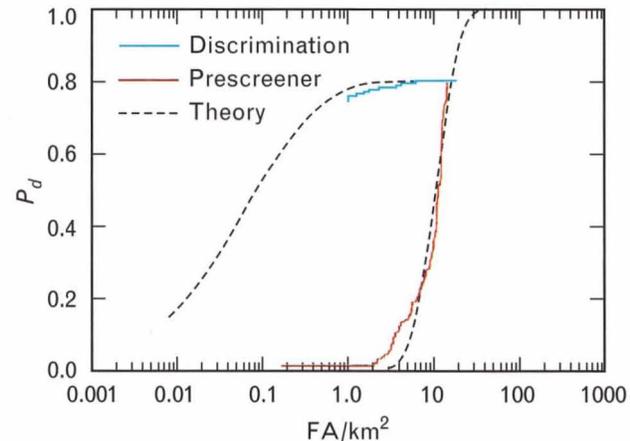


FIGURE 18. Comparison of real data and theoretical results for the 1-ft-by-1-ft resolution, PWF, Stockbridge natural- and cultural-clutter case.

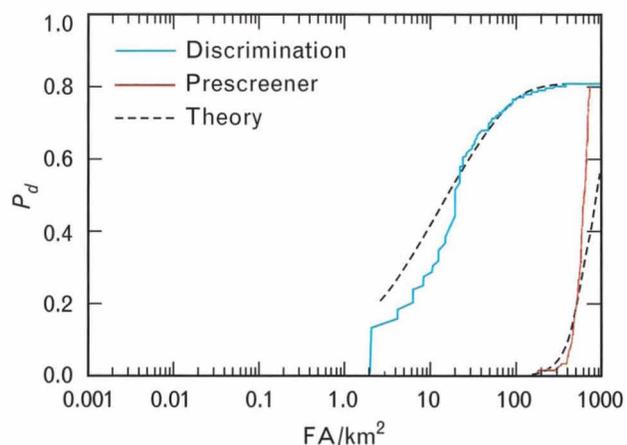


FIGURE 19. Comparison of real data and theoretical results for the 1-ft-by-1-ft resolution, PWF, Concord man-made clutter case.

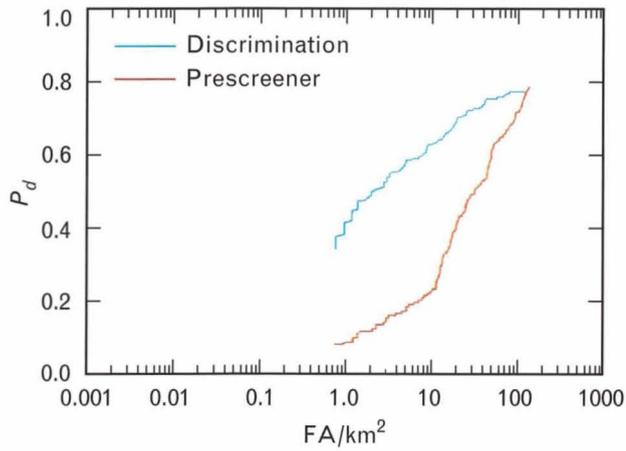


FIGURE 20. Comparison of real data and theoretical results for the 1-ft-by-1-ft resolution, HH-polarization, Stockbridge natural-clutter case.

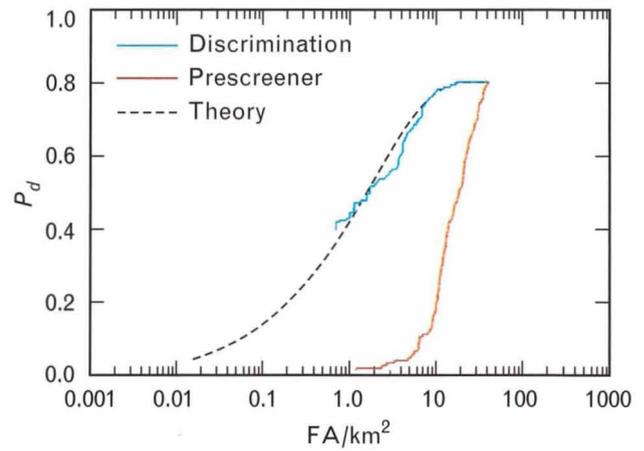


FIGURE 23. Comparison of real data and theoretical results for the 1-m-by-1-m resolution, PWF, Stockbridge natural- and cultural-clutter case.

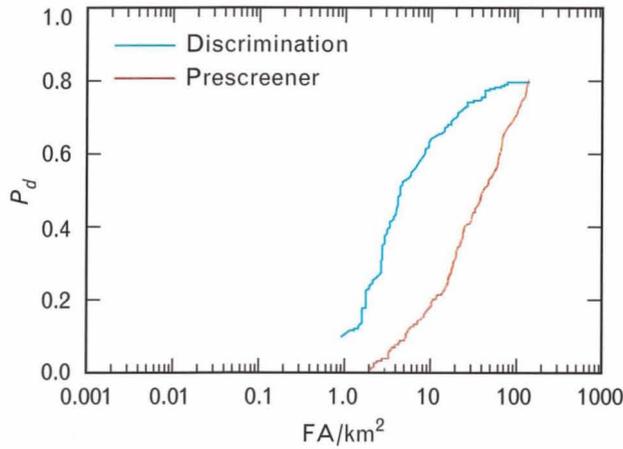


FIGURE 21. Comparison of real data and theoretical results for the 1-ft-by-1-ft resolution, HH-polarization, Stockbridge natural- and cultural-clutter case.

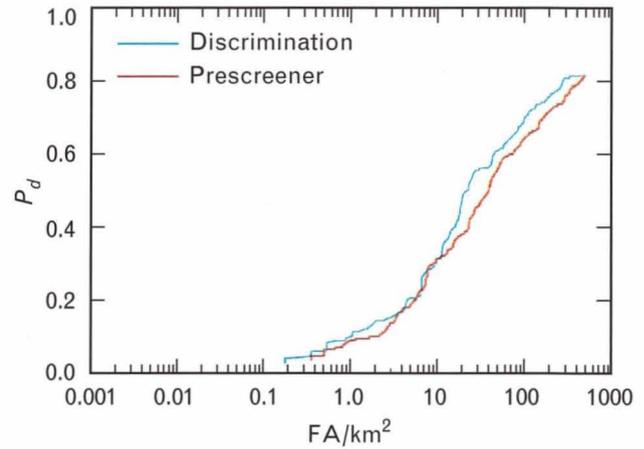


FIGURE 24. Comparison of real data and theoretical results for the 1-m-by-1-m resolution, HH-polarization, Stockbridge natural-clutter case.

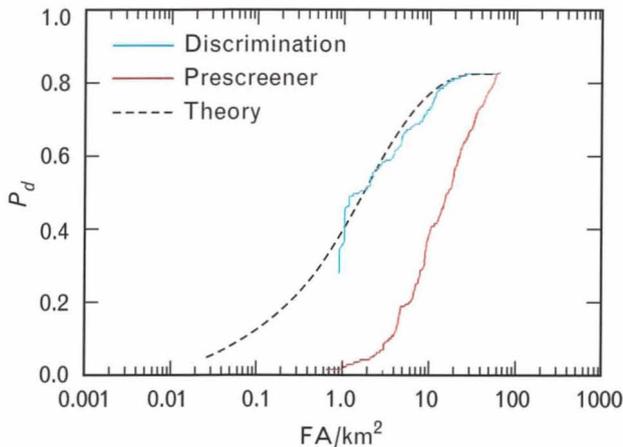


FIGURE 22. Comparison of real data and theoretical results for the 1-m-by-1-m resolution, PWF, Stockbridge natural-clutter case.

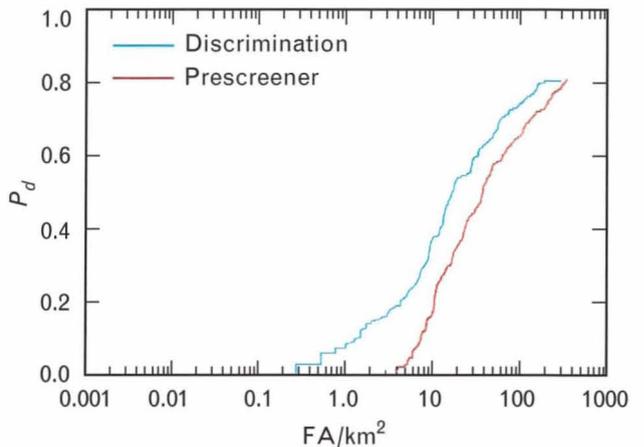


FIGURE 25. Comparison of real data and theoretical results for the 1-m-by-1-m resolution, HH-polarization, Stockbridge natural- and cultural-clutter case.

common HH single-polarization radar. We see that the performance increase is significant (a reduction of one to two orders of magnitude in false-alarm rate for equal probabilities of detection) in both the 1-ft resolution and the 1-m resolution cases. The performance difference is more pronounced for the higher-resolution data.

In general, during the course of this study, we noticed that the combination of higher-resolution data and fully polarimetric data provided a significant increase in performance. Either capability alone is not nearly as effective as the two capabilities together for the discrimination features we have studied in this article. In other words, building radars with both higher resolution and with fully polarimetric capability makes sense.

In the 1-m resolution case, the difference between the PWF data and the HH-polarization data is clear. Using the HH data alone, the discrimination algorithm provides little or no performance improvement over using the prescreeener algorithm alone. The features for the 1-m, HH-polarization case were tuned specifically for these datasets, so this result should be considered a best case. Clearly, there is no point in using the discrimination algorithm with these features for the 1-m resolution, HH-polarization dataset, because it provides little benefit and it requires additional computational capacity.

Resolution Comparisons

We can also compare the results from the 1-ft resolution case with the results from the 1-m resolution case. We see that the higher-resolution data allows a performance increase of more than an order of magnitude in terms of the false-alarm rate for a given probability of detection. This performance increase is approximately constant over the different cases given in Figures 17 to 25.

Conclusion

In this article, we discuss and evaluate the discrimination algorithm used in the Lincoln Laboratory multistage target-detection algorithm. This one-class quadratic discriminator uses features calculated from SAR imagery. The discrimination algorithm uses candidate regions of interest identified by the prescreeener, and

ideally eliminates all natural-clutter false alarms from further consideration, passing only targets and man-made clutter false alarms to the classification algorithm.

Fifteen discrimination features were evaluated for this study; three of the features were developed by Lincoln Laboratory and the remainder were developed by the three STAR contractors. These features were modified to account for the different types of data used in this study, and the best set of features was chosen for a number of different datasets and a number of different types of data. The best features remained constant from the natural-clutter dataset to the natural- and cultural-clutter dataset, which was a surprising and pleasing result. For best performance, we needed to select different feature sets for PWF and HH-polarization data, as well as for 1-ft and 1-m resolution data, which was not a surprising result.

We evaluated the features by using a theoretical expression that accurately predicted the real-data performance of the discrimination algorithm. This accuracy reflects a good understanding of how the discrimination algorithm functions as a part of the Lincoln Laboratory multistage target-detection algorithm for SAR data.

Acknowledgments

The authors would like to thank Dr. Leslie Novak for the help and encouragement given throughout the course of this study.

REFERENCES

1. L.M. Novak, M.C. Burl, R.D. Chaney, and G.J. Owirka, "Optimal Processing of Polarimetric Synthetic-Aperture Radar Imagery," *Linc. Lab. J.* **3**, 273 (1990).
2. W.W. Irving, G.J. Owirka, and L.M. Novak, "Adaptive Processing of POL-SAR Imagery," *Proc. IEEE 24th Asilomar Conf. on Circuits, Systems, and Computers, Pacific Grove, CA, November 5-7 1990*, p. 388.
3. K. Fukunaga, R.R. Hayes, and L.M. Novak, "The Acquisition Probability for a Minimum Distance One-Class Classifier," *IEEE Trans. Aerosp. Electron. Syst.* **23**, 493 (1987).
4. L.M. Novak, "On the Sensitivity of Bayes and Fisher Classifiers in Radar Target Detection," *Proc. IEEE 18th Asilomar Conf. on Circuits, Systems, and Computers, Pacific Grove, CA, November 5-7, 1984*, p. 367.
5. D.E. Kreithen and S.D. Halversen, "A Theoretical Analysis of a Ranking Discrimination Algorithm," *Proc. IEEE 26th Asilomar Conf. on Circuits, Systems, and Computers, Pacific Grove, CA, October 26-28, 1992*, p. 431.
6. M.C. Burl, G.J. Owirka, and L.M. Novak, "Texture Discrimination in Synthetic Aperture Radar Imagery," *Proc. IEEE 23rd Asilomar Conf. on Circuits, Systems, and Computers, Pacific Grove, CA, October 30-November 1, 1989*, p. 399.
7. S.D. Halversen, "Calculating the Orientation of a Rectangular Target in SAR Imagery," *Proc. IEEE 1992 National Aerospace and Electronics Conf. (NAECON '92), Dayton, Ohio, 18-22 May 1992*, p. 260.
8. N.L. Johnson and S. Kotz, *Continuous Univariate Distributions 2* (Wiley, New York, 1970).
9. R.R. Parenti and E.W. Tung, "A Statistical Analysis of the Multiple Target Multiple Shot Target Acquisition Problem," *Project Report TT-43*, MIT Lincoln Laboratory, 28 Jan. 1981.
10. G.B. Goldstein, "False-Alarm Regulation in Log-Normal and Weibull Clutter," *IEEE Trans. Aerosp. Electron. Syst.* **9**, 84 (1973).



DANIEL E. KREITHEN is a staff member in the Surveillance Systems group. His research speciality is in the detection of stationary targets in synthetic-aperture radar imagery. He received a Sc.B. degree in engineering from Brown University, and an M.S.E. degree in electrical engineering from Princeton University. Dan has been at Lincoln Laboratory since 1988.



SHAWN D. HALVERSEN is an associate staff member in the Surveillance Systems group. His research speciality is in the detection and discrimination of stationary targets. He received a B.A. degree in mathematics and a B.S. degree in physics from the University of West Florida, and an M.A. degree in mathematics from the University of Wisconsin. He has been at Lincoln Laboratory since 1990.



GREGORY J. OWIRKA is an assistant staff member in the Surveillance Systems group. He received a B.S. degree (cum laude) in applied mathematics from Southeastern Massachusetts University, and he is currently working on an M.S. degree in electrical engineering at Northeastern University. Greg's current research interests are in automatic target recognition. He has been at Lincoln Laboratory since 1987.