A Support Vector Method for Anomaly Detection in Hyperspectral Imagery

Amit Banerjee, *Member, IEEE*, Philippe Burlina, *Member, IEEE*, and Chris Diehl, *Member, IEEE*

Abstract—This paper presents a method for anomaly detection in hyperspectral images based on the support vector data description (SVDD), a kernel method for modeling the support of a distribution. Conventional anomaly-detection algorithms are based upon the popular Reed–Xiaoli detector. However, these algorithms typically suffer from large numbers of false alarms due to the assumptions that the local background is Gaussian and homogeneous. In practice, these assumptions are often violated, especially when the neighborhood of a pixel contains multiple types of terrain. To remove these assumptions, a novel anomaly detector that incorporates a nonparametric background model based on the SVDD is derived. Expanding on prior SVDD work, a geometric interpretation of the SVDD is used to propose a decision rule that utilizes a new test statistic and shares some of the properties of constant false-alarm rate detectors. Using receiver operating characteristic curves, the authors report results that demonstrate the improved performance and reduction in the false-alarm rate when using the SVDD-based detector on wide-area airborne mine detection (WAAMD) and hyperspectral digital imagery collection experiment (HYDICE) imagery.

Index Terms—Hyperspectral, support vector data description, target detection.

I. INTRODUCTION

RECENT advances in hyperspectral sensors with high spectral and spatial resolution have led to an increased interest in exploiting spectral imagery for target detection. Given the availability of spectral libraries for a wide range of materials, detection algorithms that exploit a known target signature have been widely investigated. It has been shown [19] that such algorithms are dependent on the degree of signal mismatch between the spectral libraries and the spectra observed in an image. Complications arising from: 1) accurate spectral calibration; 2) compensating for atmospheric effects to convert radiance spectra to the reflectance domain; or 3) the availability of reliable atmospheric data to convert reflectance values to the radiance spectra can lead to errors that hinder the performance of known signature detectors.

Such complications can be avoided by using an anomaly detector. Detecting anomalies in hyperspectral imagery entails the task of locating pixels with spectral signatures that deviate significantly from the local background. Anomaly detectors have the advantage of not requiring *a priori* knowledge of the target's spectral signature and, therefore, can process images completely in scene. However, anomaly detectors usually suffer from a high false-alarm rate (FAR) due to the simplifying assumptions imposed on the background signature distribution. In this paper, we present a new kernel-based approach that removes these assumptions to detect spectral anomalies while mitigating the FAR.

Lacking prior target signatures, anomaly-detection methods model the background and find pixels in the scene which are not well described by the background model. This is typically achieved by identifying the region of the given feature space that contains most of the background pixels. If the pixel under test falls in this region, it is labeled as part of the background; if it lies outside of the background's region of support,[1] it is detected as an outlier and declared as a target.

In general, there are two types of methods for computing the region of support for the background. One is to estimate the underlying probability density function (pdf) for the background signature and threshold the result. The form of the pdf resolves the shape, and the threshold value determines the size of the support region. The most commonly used models for the density function $f(\mathbf{x})$ are the local Gaussian model, the global Gaussian mixture model, and the global linear mixture model. The second approach is to estimate the region of support directly without assuming its shape.

The Reed–Xiaoli (RX) algorithm, which is the benchmark anomaly detector for hyperspectral imagery, uses the local Gaussian model [1]. With this approach, the background pixels in a local neighborhood around the pixel under test are assumed to be independent identically distributed Gaussian random variables. After estimating the background mean vector and covariance matrix, the Mahalanobis distance between the pixel under test and the background mean vector is compared to a threshold to detect an anomaly [1].

There are two issues with the RX algorithm that limit its performance. In many environments, it has been shown empirically that the local normal model provides an inadequate representation of the underlying distribution [9] leading to poor false alarm performance. This is especially true when the local background contains multiple classes of terrain. Using a goodness-of-fit test statistic for hyperspectral imagery based upon the Barringhaus, Henze, Epps, and Pully (BHEP) test, it has been shown that the normal model is not valid in most situations [10]. In addition, the RX algorithm is computationally intensive when operating on hyperspectral imagery. This is due to the need to estimate and invert large covariance matrices.

[1]The region of support for the random variable $\mathbf{x}$ can be represented by an indicator function $I(\mathbf{x})$, where $I(\mathbf{x}) = 1$ if the pdf of $\mathbf{x}$ is greater than or equal to a given threshold, and $I(\mathbf{x}) = 0$ if $f(\mathbf{x})$ is less than the threshold [17].

If the local background contains multiple types of terrain, the background cannot be properly modeled by a unimodal distribution. To more properly characterize nonhomogeneous backgrounds, researchers have employed the mixture of Gaussian models [11]. This approach models the background signature distribution as a linear combination of Gaussian distributions. The number of distributions, their weights, and the parameters of the normal pdfs are estimated using a stochastic expectation maximization (SEM) method [12]. Given the parameters of the mixture of Gaussian model, a generalized likelihood ratio test (GLRT) is applied to detect outliers. While this Gaussian mixture model provides an improved performance over the RX algorithm, it still requires the estimation and inversion of large covariance matrices and is further limited by the need to know or estimate *a priori* the number of classes of terrain in the image.

In [13], the use of Gauss–Markov random fields (GMRFs) is introduced for hyperspectral anomaly detection. The spectral and spatial correlations of background clutter in hyperspectral imagery are described by the GMRF. The estimated parameters of the GMRF are then used in the GLRT to detect the outliers. The principal advantages of this method are that it is computationally efficient, even for high-dimensional data, and uses spatial information that is usually ignored by most detectors. However, the model assumes that the background is locally homogeneous, which can lead to a poor performance when the targets are located along the clutter boundaries.

To perform the anomaly detection, the ultimate goal is to estimate the shape and size of the support region. Since the values of the pdf are ultimately not of interest, estimating the density function across the entire feature space is unnecessary. In keeping with Vapnik's [17] principle of avoiding more difficult estimation problems along the path to the desired solution, one can estimate the region of support directly for the background and avoid the problem of estimating the underlying pdf.

Recently, large-margin techniques, such as support vector machines (SVMs), have received considerable attention for classifying high-dimensional non-Gaussian data. SVMs yield good generalization performance on such problems by directly estimating a decision boundary with maximal separability. Motivated by the success of SVMs in classifying pixels in hyperspectral imagery [2], [3], we extend the SVM approach to detect spectral anomalies. For the particular problem of anomaly detection, several one-class classifiers have been derived in the literature, including the support vector data description (SVDD) [7], [8] and the ν-SVM [15] methods. These classifiers are able to directly estimate the support region for a given dataset. The SVDD is a technique that has been used in several domains, such as faulty-machine-part detection [8] and image retrieval [18]. In this paper, we develop an algorithm that utilizes the SVDD to detect spectral anomalies that lie outside the region of support for local background pixels.

For anomaly detection, the SVDD approach has the following benefits: *Nonparametric*: it is a data-driven method that avoids prior assumptions about the distribution of the data. *Sparsity*: fewer training samples are needed, and thus fewer pixels are needed to accurately characterize the background; *Good Generalization*: the method avoids overfitting and yields

good generalization results when compared to other classical methods [8], [17]; and *Use of kernels*: by exploiting the "kernel trick," the SVDD method is able to accurately model the support of nontrivial multimodal distributions. Therefore, the SVDD-based anomaly detector yields more powerful detection performance for targets embedded in non-Gaussian and nonhomogeneous backgrounds.

The recently introduced kernel RX algorithm [20] is a related technique for anomaly detection. It is a nonlinear version of the RX detector that uses kernels to model non-Gaussian distributions. The kernel RX detector has two key differences with the SVDD approach. While the SVDD avoids the problem of density estimation, the kernel RX assumes a Gaussian distribution in the feature space. This assumption is the equivalent to a Parzen estimate for the distribution in the original feature space [20]. The second difference is that the kernel RX requires the estimation and inversion of a large covariance matrix, a computational burden that the SVDD avoids.

This paper is organized as follows. Section II provides an overview of the derivation of the SVDD one-class classifier. The algorithm to detect hyperspectral anomalies using the SVDD is presented in Section III, along with a discussion of a new SVDD test statistic. Experimental results are provided in Section IV to evaluate the RX and SVDD anomaly detectors. Finally, concluding remarks are given in Section V.

II. DERIVING THE SVDD

In the following, we first introduce the linear SVDD, which models the support of the distribution as the minimal enclosing hypersphere containing the data in the original input space. Then, we generalize this result through the use of kernel functions to the nonlinear SVDD, which first maps the input space to a high-dimensional feature space and then estimates the minimum enclosing hypersphere in the feature space.

Before proceeding, we note that the one-class classifiers mentioned above, the SVDD and the ν-SVM, are related. While the SVDD computes an optimal hypersphere that contains the data, the ν-SVM estimates a large-margin hyperplane that separates the data and the origin of the space in which the data resides [16]. It has been shown [15], [16] that these methods are equivalent when the radial basis function (RBF) is chosen as the kernel function, which is the common practice.

A. Linear SVDD

To determine the minimum enclosing hypersphere $S = \{\mathbf{x} : \|\mathbf{x} - \mathbf{a}\|^2 < R^2\}$ that contains the training set $T = \{\mathbf{x}_i, \ i = 1, \ldots, M\}$, we must solve the following constrained optimization problem:

$$\min(R) \quad \text{subject to } \mathbf{x}_i \in S, \qquad i = 1, \ldots, M.$$

The center $\mathbf{a}$ and radius R of the minimum enclosing hypersphere can be found by optimizing the following Lagrangian:

$$L(R, \mathbf{a}, \alpha_i) = R^2 - \sum_i \alpha_i \left\{ R^2 - (\langle \mathbf{x}_i, \mathbf{x}_i \rangle - 2\langle \mathbf{a}, \mathbf{x}_i \rangle + \langle \mathbf{a}, \mathbf{a} \rangle) \right\}. \tag{1}$$

The first term in (1) is the radius, which we aim to minimize. The second term constrains each training point $\mathbf{x}_i$ to lie within the sphere with center $\mathbf{a}$ and radius R. The optimal solution must satisfy the Karush–Kuhn–Tucker (KKT) conditions. Taking the partial derivatives of L with respect to R and $\mathbf{a}$ and setting them to 0 yields

$$\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_i \alpha_i = 1 \qquad (2)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \Rightarrow \mathbf{a} = \frac{\sum_i \alpha_i \mathbf{x}_i}{\sum_i \alpha_i}. \qquad (3)$$

By combining (2) and (3), a simple expression for the center of the sphere is found

$$\mathbf{a} = \sum_i \alpha_i \mathbf{x}_i. \qquad (4)$$

The center of the sphere is the center of mass of all training points, with the weights equal to the Lagrange multipliers. Furthermore, the inequality $\|\mathbf{x}_i - \mathbf{a}\|^2 < R^2$ implies $\alpha_i \geq 0$ according to the nonnegativity constraint of the KKT. Substituting (2) and (4) into (1) yields another expression for the Lagrangian function to be maximized with respect to α_i

$$L = \sum_i \alpha_i \left(\langle \mathbf{x}_i, \mathbf{x}_i \rangle \right) - \sum_{i,j} \alpha_i \alpha_j \left(\langle \mathbf{x}_i, \mathbf{x}_j \rangle \right). \qquad (5)$$

After optimizing L with respect to α, it is typical to discover that a large fraction of the α_i is equal to zero. The training examples with nonzero α_i are called support objects[2] and lie on the boundary of the support region. This is a result of the complementary slackness condition of the KKT, which states that $(\|\mathbf{x}_i - \mathbf{a}\|^2 - R^2)\alpha_i = 0$ for optimality. Hence, the support objects must lie on the boundary of the hypersphere. Therefore, the SVDD yields a sparse representation of the support expressed entirely in terms of the support objects.

Once the minimum enclosing hypersphere has been found, outliers are identified by testing whether or not the test examples lie within or outside the hypersphere. When the decision rule $\|\mathbf{y} - \mathbf{a}\|^2 \geq R^2$ holds true for a test example $\mathbf{y}$, the example is labeled as an outlier. Expanding

$$\|\mathbf{y} - \mathbf{a}\|^2 = (\mathbf{y} - \mathbf{a})^T (\mathbf{y} - \mathbf{a})$$

$$= \left(\mathbf{y} - \sum_i \alpha_i \mathbf{x}_i \right)^T \left(\mathbf{y} - \sum_i \alpha_i \mathbf{x}_i \right)$$

$$= \langle \mathbf{y}, \mathbf{y} \rangle - 2 \sum_i \alpha_i (\langle \mathbf{y}, \mathbf{x}_i \rangle) + \sum_{i,j} \alpha_i \alpha_j \left(\langle \mathbf{x}_i, \mathbf{x}_j \rangle \right) \qquad (6)$$

the decision rule becomes

$$\langle \mathbf{y}, \mathbf{y} \rangle - 2 \sum_i \alpha_i \left(\langle \mathbf{y}, \mathbf{x}_i \rangle \right) + \sum_{i,j} \alpha_i \alpha_j \left(\langle \mathbf{x}_i, \mathbf{x}_j \rangle \right) \geq R^2. \qquad (7)$$

We refer to (4) as the linear SVDD decision rule.

[2]The terms support object and support vector are used interchangeably throughout the paper.

B. Kernel-Based SVDD

In most instances, a hypersphere does not provide a tight representation of the support of the data in the original input space. Additional flexibility is needed to model arbitrarily complex distributions. To address this, the nonlinear SVDD maps the data from the input space to a higher dimensional feature space through the use of a mapping $\Phi(\mathbf{x})$, and models the support of the distribution as a minimum enclosing hypersphere in the feature space. This hypersphere corresponds to a tighter boundary for the support region in the original input space.

Proceeding, as in the previous section, we now seek the smallest hypersphere in the induced feature space $S = \{\Phi(\mathbf{x}) : \|\Phi(\mathbf{x}) - \mathbf{c}\|^2 < R^2\}$ including the entire set of mapped training examples $T = \{\Phi(\mathbf{x}_i), \ i = 1, \ldots, M\}$. We therefore need to solve the following constrained optimization problem:

$$\min(R) \quad \text{subject to } \Phi(\mathbf{x}_i) \in S, \qquad i = 1, \ldots, M.$$

The corresponding Lagrangian is expressed as

$$L(R, \mathbf{a}, \alpha_i) = R^2$$
$$- \sum_i \alpha_i \left\{ R^2 - (\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_i) \rangle - 2 \langle \mathbf{c}, \Phi(\mathbf{x}_i) \rangle + \langle \mathbf{c}, \mathbf{c} \rangle) \right\} \quad (8)$$

with Lagrange multipliers α_i. Setting the partial derivatives of L with respect to R and $\mathbf{a}$ to zero, and substituting the results into L yields

$$L = \sum_i \alpha_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_i) \rangle - \sum_{i,j} \alpha_i \alpha_j \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \quad (9)$$

with $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$. This also gives an expression for the center of the sphere

$$\mathbf{c} = \sum_i \alpha_i \Phi(\mathbf{x}_i).$$

As above, the hypersphere center is the center of gravity of the support vectors given the optimal weights α_i. The corresponding decision rule to detect an anomaly for a test pixel $\mathbf{y}$ is $\text{SVDD}(\mathbf{y}) = \|\Phi(\mathbf{y}) - \mathbf{c}\|^2 \geq R^2$, which expands to

$$\text{SVDD}(\mathbf{y}) = \|\Phi(\mathbf{y}) - \mathbf{c}\|^2$$
$$= \left\| \Phi(\mathbf{y}) - \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}_i) \right\|^2$$
$$= \langle \Phi(\mathbf{y}), \Phi(\mathbf{y}) \rangle - 2 \sum_i \alpha_i \langle \Phi(\mathbf{y}), \Phi(\mathbf{x}) \rangle$$
$$+ \sum_{i,j} \alpha_i \alpha_j \langle \Phi(\mathbf{y}), \Phi(\mathbf{x}) \rangle. \qquad (10)$$

Note that the above expression consists of inner products of the mapping function Φ. Using the well-known "kernel trick," these inner products can be represented by a kernel function $K(\mathbf{x}, \mathbf{y})$. Kernel functions provide a computationally efficient technique to implicitly map the data into the induced feature space and compute the inner product. The only requirement imposed on the kernel function is that it satisfies Mercer's theorem [17]. Mercer's theorem states that in order for a kernel function
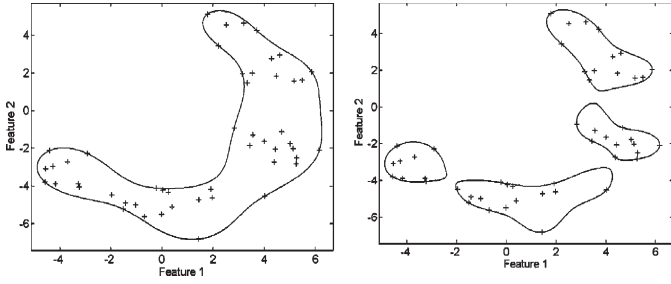
Fig. 1. SVDD decision boundary using $\sigma = 5$ (left) and $\sigma = 3$ (right). Note how using a smaller value for sigma leads to a tighter decision boundary.

Fig. 2. Example of a hollow window used to collect pixels for background estimation.

$K(\mathbf{x}, \mathbf{y}) : \mathbf{R}^N \times \mathbf{R}^N \to \mathbf{R}, (\mathbf{x}, \mathbf{y}) \mapsto z = K(\mathbf{x}, \mathbf{y})$ to admit an eigenfunction expansion of the form

$$K(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle = \sum_{k=1}^{\infty} b_k \phi(\mathbf{x}) \phi(\mathbf{y})$$

it must be positive semi-definite.

$$\int K(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0, \forall g(.) \in L_2.$$

Under these conditions, the SVDD statistic can be simply expressed as

$$\mathrm{SVDD}(\mathbf{y}) = K(\mathbf{y}, \mathbf{y}) - 2 \sum_i \alpha_i K(\mathbf{y}, \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \geq R^2.$$

Such continuous symmetric positive semi-definite kernel functions play the role of a dot product, thereby forming a new Hilbert space where the problem of determining the minimum enclosing hypersphere can be posed in a sensible fashion.

In this paper, we use the popular Gaussian RBF as the kernel function defined as

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2} \right).$$

The choice of the RBF simplifies the $\mathrm{SVDD}(\mathbf{y})$ test statistic to

$$\mathrm{SVDD}(\mathbf{y}) = 1 - 2 \sum_i \alpha_i K(\mathbf{y}, \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \geq R^2 \tag{11}$$

since $K(\mathbf{y}, \mathbf{y}) = 1$. The RBF has one free parameter, which is the scale parameter σ. This parameter affects the tightness of fit for the training data (and, therefore, boundary smoothness), as shown in Fig. 1. As explained in [7] and in Section III-A, it is a measure of how well the SVDD generalizes to unseen data. By varying the scale parameter of the RBF, the SVDD can determine multiple regions of support for a dataset. As can be seen Fig. 1, this allows the SVDD to model multimodal distributions. For anomaly detection, this implies that the SVDD can detect targets embedded in multiple types of clutter with fewer false alarms.
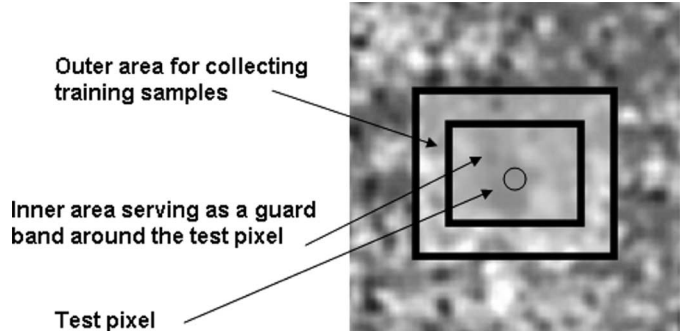
III. SVDD HYPERSPECTRAL ANOMALY-DETECTION ALGORITHM

A. Outline of Algorithm

Using the above derivation, the steps for the SVDD anomaly detector are similar to those of the RX algorithm. The SVDD statistic is computed at each pixel to determine if the pixel lies within the support region of the background pixels. This approach replaces the Mahalanobis distance used in the RX detector with the $\mathrm{SVDD}(\mathbf{y})$ measure. The steps for the SVDD-based anomaly detection are as follows.

1) Select the dimensions of the hollow background window.
2) Estimate the sigma, which is the scale parameter of the RBF kernel.
3) Given the window size and the scale parameter, for each pixel:
   a) sample the pixels from the local neighborhood using the background window;
   b) using these pixels, compute the SVDD parameters $(\mathbf{a}, \alpha_i, R)$ to model the region of support for the background clutter;
   c) decision.
      i) If $\mathrm{SVDD}(\mathbf{y})$, the SVDD test statistic for pixel $\mathbf{y}$ is less than the detection threshold T; the pixel is part of the background.
      ii) Otherwise, declare the pixel as a target.

Details for steps 1) and 2) of the algorithm are given below.

Dimensions for the Background Window: Local background samples around the pixel under test are used to estimate the background SVDD parameters. The samples are collected from a hollow rectangular window centered at the pixel. The window consists of two regions, the inner and outer window regions, as shown in Fig. 2. The dimensions of each region are based on the spatial resolution of the image and the expected size of the targets. The size of the inner area should accommodate the typical or largest expected target in the scene. The size of the outer window should be large enough to collect a sufficient number of training samples for the background estimation.

Estimating Sigma: Choosing an appropriate value for the RBF kernel scale parameter is critical for acceptable detection performance. As shown in [7], [8], and [19], and illustrated in Fig. 1, if the value of sigma is too small, the resulting decision boundary overfits the data and does not generalize well to the unseen data. If sigma is too large, the classifier underfits the data

and computes a possibly trivial decision boundary. We present a minimax approach for estimating a scale-parameter value that yields good generalization performance.

The performance of anomaly detectors is commonly characterized by two measures of error: the FAR (classifying a background pixel as an outlier), and the false acceptance rate (accepting a target pixel as part of the background). Since the target spectra are unknown, it is not possible to estimate the false acceptance rate from the training samples. Therefore, we use a method to estimate sigma that minimizes the FAR. Using an argument based on leave-one-out cross validation, the following upper bound on the FAR can be derived for the SVDD [7], [8], [17]:

$$P_{\text{fa}} \leq \text{E}\left[\frac{\#\text{SV}}{N}\right].$$

In this expression, P_{fa} is the FAR, $\#\text{SV}$ is the number of support vectors in the SVDD model, and N is the total number of training samples. Given that the expectation in the above expression is over the training sets of size N, we cannot estimate the expectation from a single training set. Therefore, we substitute the resulting number of support vectors for the SVDD model in place of the expectation to obtain an approximate upper bound. Through a resampling process, it would be possible to develop a lower variance estimate of the expectation. Yet at this point in our investigation, it is not clear that the extra computational cost is warranted.

To avoid the computational expense of estimating sigma for every SVDD model while scanning through the image, we have chosen instead to estimate a value of sigma that minimizes the average FAR across the entire image. Using the above result, one can derive the following approximate minimax estimate for sigma that minimizes an approximate upper bound on the average FAR

$$\hat{\sigma} = \min_{\sigma} \frac{1}{M} \sum_{i=1}^{M} P_{\text{fa}\_i}$$

$$\approx \min_{\sigma} \left\{ \frac{1}{M} \sum_{i=1}^{M} \text{E}\left(\frac{\#\text{SV}_i}{N}\right) \right\}$$

$$\approx \min_{\sigma} \left\{ \frac{1}{M} \sum_{i=1}^{M} \frac{\#\text{SV}_i}{N} \right\}$$

where M is the number of training sets, N is the number of examples in each training set, and $\#\text{SV}_i$ is the number of support vectors found for the ith training set.

The algorithm to obtain this global estimate for sigma is as follows: Generate multiple sets of training data by randomly selecting pixels from the image. For each set of training data, the SVDD decision boundary is determined using different values for sigma. For each value of sigma, the average fraction of support vectors is computed over all of the training sets. The sigma that produces the smallest average fraction of support vectors is the minimax estimate.

Fig. 3 provides an example of the minimax estimate for sigma described above. In this example, three sets of sixty training samples are randomly selected from a hyperspectral image. The SVDD classifier is trained on all three datasets with
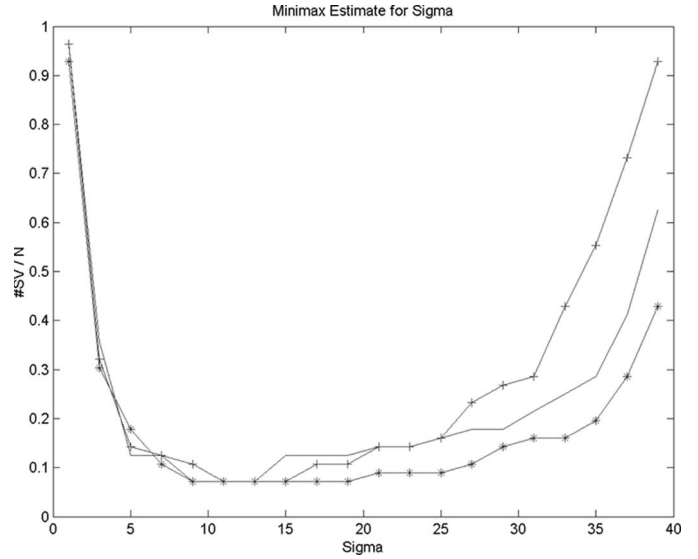

Fig. 3. Fraction of support vectors as a function of sigma for three SVDD models. From these curves, the optimal estimate for sigma that minimizes the upper bound for the FAR is approximately 13.

varying values for sigma. It can be seen that the number of support vectors is minimum for all three curves at $\sigma = 13$.

B. Normalized SVDD Test Statistic

Anomaly detectors compute a statistic that provides a measure of similarity to the background for each pixel in the image. The statistic can be used to directly compare the pixels and identify those that are more likely to be anomalies. This allows for the use of a single threshold for the entire image. By varying this global threshold, an empirical receiver operating characteristic (ROC) curve can be generated to evaluate the performance of the detector.

One class of such detectors is the constant FAR (CFAR) detectors [4]–[6]. The CFAR property is useful because it allows for the determination of a single detection threshold that: 1) generates a desired number of false alarms; 2) is independent of the estimated parameters; and 3) can be used for every pixel in the image.

The RX algorithm is an example of a CFAR detector [11]. Using the Mahalanobis distance, each pixel is normalized to yield a zero-mean Gaussian random variable with an identity covariance matrix. This allows the algorithm to compute a detection threshold via

$$\int_{\mathbf{x}: f(\mathbf{x}) < T} f(\mathbf{x})d\mathbf{x} = P_{\text{fa}}$$

where $f(\mathbf{x}) \sim N(\mathbf{0}, \mathbf{I})$, and P_{fa} is the desired FAR. Note that T is computed independently of the estimated background parameters. Furthermore, if the pixels are Gaussian distributed, the threshold will generate the desired number of false alarms.

For the SVDD, we can similarly derive a normalized test statistic. In contrast to the RX algorithm, we motivate the following normalization procedure geometrically and strictly avoid imposing assumptions about the underlying distribution. Recall from Section II-B that the nonlinear SVDD estimates

the minimum enclosing hypersphere centered at $\mathbf{a}$ with radius R in the feature space as the support region for the background pixels. From (10), the distance between the mapped example $\Phi(\mathbf{y})$ and the centroid of the hypersphere $\mathbf{c}$ is

$$\|\Phi(\mathbf{y}) - \mathbf{c}\|^2 = 1 - 2\sum_i \alpha_i K(\mathbf{x}_i, \mathbf{y}) + \sum_i \sum_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$$

which is the original SVDD statistic. Since the radius of the optimal hypersphere will vary from pixel to pixel, the distance to the centroid cannot be used to compare the similarities of multiple pixels to their respective backgrounds. Therefore, we utilize a normalized version of the SVDD test statistic

$$\begin{aligned} \mathrm{SVDD}_N(\mathbf{y}) &= \frac{\|\Phi(\mathbf{y}) - \mathbf{c}\|^2}{R^2} \\ &= \frac{1 - 2\sum_i \alpha_i K(\mathbf{x}_i, \mathbf{y}) + \sum_i \sum_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)}{R^2}. \end{aligned}$$
(12)

By dividing the original statistic by the squared radius, the $\mathrm{SVDD}_N(\mathbf{y})$ function effectively transforms the feature space so that the minimum enclosing hypersphere encompassing the data has unit radius. This normalization allows for direct comparison of SVDD function values for different pixels in the image. The detection threshold T can now be chosen independently of the estimated SVDD parameters.

Selecting the detection threshold analytically is an important consideration. CFAR detectors, such as the RX, allow for the estimation of a threshold based on a desired FAR. However, if the data are non-Gaussian, the CFAR properties no longer hold, and choosing an appropriate threshold based on the FAR is not possible. As the SVDD is a nonstatistical technique, it is difficult to compute the threshold analytically. Instead, its detection threshold can be determined empirically as follows. The SVDD detector is applied to multiple images of different clutter types imaged under various atmospheric conditions. The resulting ROC curves are stored and indexed by the clutter and atmospheric conditions. When applying the SVDD detector on a new image, the ROC curve most closely approximating the new operating conditions is used to determine the detection threshold. Given an acceptable range for the FAR, the optimal threshold can be inferred.

IV. Experimental Results

In this section, we evaluate the computational and detection performance of the RX, SVDD, and normalized SVDD algorithms. Examples of how the FAR is affected by the signal-to-noise ratio and the background-clutter statistics are given. An analysis of how algorithm speed varies with the number of spectral bands and the number of training samples is provided. The analysis offers some insight into selecting the appropriate anomaly detector for different situations.

Data Overview: We compare the detectors using images taken from the wide-area airborne mine detection (WAAMD) and hyperspectral digital imagery collection experiment (HYDICE) datasets. The WAAMD imagery is provided by the Night Vision Electronic Systems Directorate (NVESD). The compass sensor, providing 256 bands in the visible/near-infrared/shortwave-infrared (VIS/NIR/SWIR) part of the spectrum (400–2350 nm), was used to image the minefields. For the images used in this study, the sensor was flown at an altitude of 1000 and 2000 ft with a ground sample distance (GSD) of approximately 4 and 8 in, respectively. The images contain two types of mines of which the sizes are approximately 2×2 pixels.

The HYDICE sensor was used to acquire the Forest Radiance II data collect. The sensor provides 210 spectral bands in the VIS/NIR/SWIR wavelengths with approximately 1-m spatial resolution. We focus on two images that contain the targets from detection experiments 2, 3, and 4. There are over 20 target vehicles in the scenes, including many that are partially hidden along the tree lines. There are a handful of additional "transient" nonmilitary vehicles in the images that are not marked as ground truth but used in the evaluation.

Preprocessing: The detection algorithms are evaluated on both multispectral and hyperspectral versions of the imagery. For hyperspectral data, the only preprocessing step involved manually removing the water and CO_2 absorption bands. The number of remaining bands is 145 for the WAAMD data and 133 for the HYDICE imagery. To generate multispectral data, an additional step was taken to reduce the total number of bands to seven by averaging adjacent spectral bands. The spectrum was divided into seven regions where the averaging was performed; the regions correspond to red, green, blue, two NIR, and two SWIR regions.

For all of the detectors, a hollow window is used to estimate the local background model. The window dimensions are determined by the largest expected target size in the scene, which can be determined by the pixel resolution of the data. A hollow window is used to avoid incorporating target pixels in the estimation of the background statistics. For the RX detector, a mean vector and a covariance matrix are estimated for each pixel from the pixels in the hollow window. For the SVDD detectors, the samples from the background window are used to train the SVDD classifier at each pixel. The dimensions of the window are the same for all detectors. However, a larger window is used for hyperspectral data, since more training samples are required to estimate the covariance matrix for the RX algorithm. The actual window sizes used in the experiments are given below.

Performance Evaluation: The detectors are evaluated by comparing their ROC curves using groundtruthed imagery. The locations of the target pixels are represented by rectangular boxes. To generate the ROC curves, multiple thresholds are used for the RX, $\mathrm{SVDD}(\mathbf{y})$, and the $\mathrm{SVDD}_N(\mathbf{y})$ test statistics. After applying a threshold, the detected pixels are clustered via a connected components algorithm; each cluster is counted as one detection or false alarm. If any part of the cluster falls within the target box, it is considered to be a detection, otherwise, it is a false alarm. In addition, detections arising from bad lines in the image are removed manually.

In Figs. 4 –7, ROC curves are provided to evaluate the ability of the algorithms to detect mines using the WAAMD data. A total of 44 mines were placed in a highly cluttered dirt field
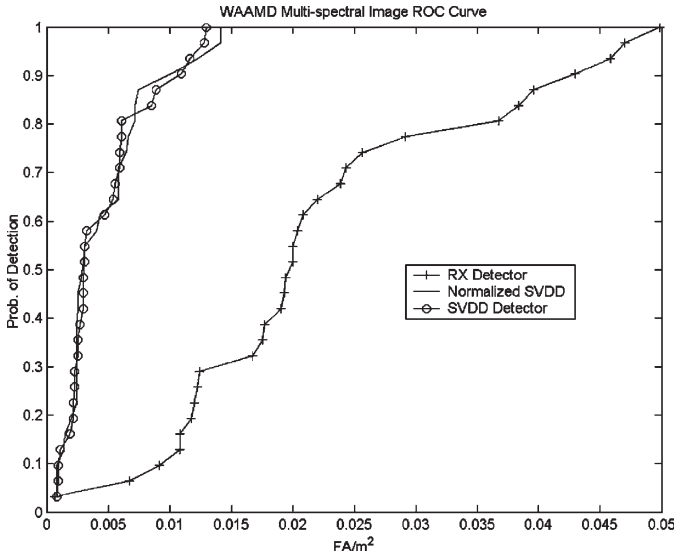
Fig. 4.   ROC curve comparing SVDD and RX detectors on WAAMD imagery using seven bands. The 1400×256 image contains 44 live mines in a dirt field and was taken at 1000-ft altitude.
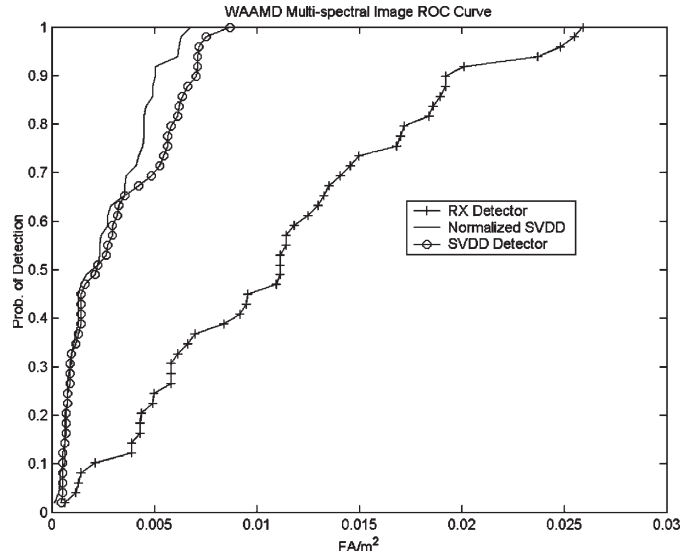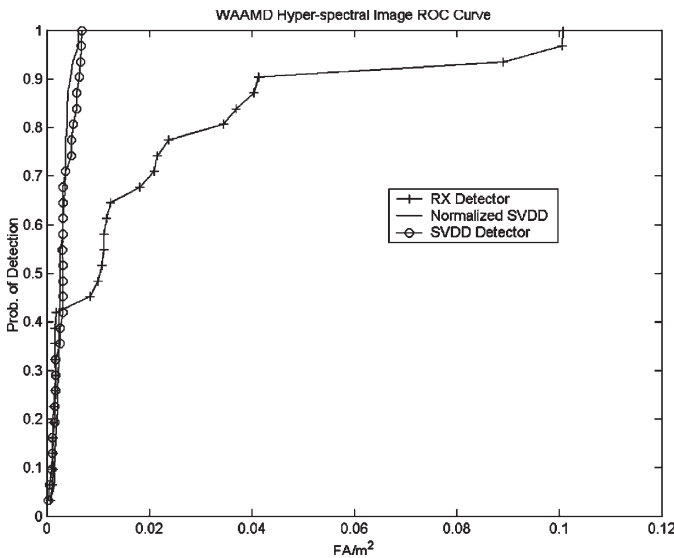

Fig. 6.   ROC curve comparing SVDD and RX detectors on WAAMD imagery using seven bands. The 1200×256 image contains 44 live mines in a dirt field and was taken at 2000-ft altitude.

Fig. 5.   ROC curve comparing SVDD and RX detectors on WAAMD imagery using 145 bands. The 1400×256 image contains 44 live mines in a dirt field and was taken at 1000-ft altitude.
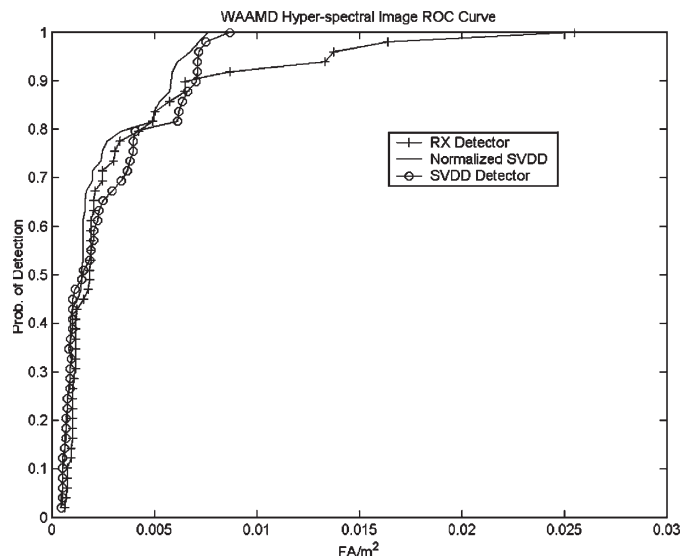

Fig. 7.   ROC curve comparing SVDD and RX detectors on WAAMD imagery using 145 bands. The 1200×256 image contains 44 live mines in a dirt field and was taken at 2000-ft altitude.

resulting in a nonhomogeneous background. The mine field was imaged with the compass sensor at altitudes of 1000 and 2000 ft, with a GSD of 4 and 8 in, respectively. The size of the 1000 and 2000-ft altitude images are 1400×256 and 1200×256 pixels, respectively. For multispectral processing, the dimensions of the inner and outer regions for the hollow background window are 7×7 and 13×13, respectively. For hyperspectral, the dimensions of the inner and outer regions for the hollow background window are 7×7 and 15×15, respectively.

The ROC curves comparing the detectors on the 1000-ft altitude image are shown in Figs. 4 and 5. In the multispectral example of Fig. 4, the SVDD detectors clearly outperform the RX algorithm throughout the curve. For 100% detection, the SVDD detectors reduce the FAR by a factor of five. In

the hyperspectral case (Fig. 5), the SVDD detectors generate approximately the same FAR as with the multispectral data (nearly 0.015 FA/m^2). However, the FAR of RX detector nearly doubles when processing the hyperspectral data. In the hyperspectral representation of the image, the FAR of the SVDD detectors is nearly an order of magnitude less than the RX algorithm. One explanation for the performance difference is the distribution of the background spectra. With a 4-in GSD, the rocks and variations in the dirt field are more pronounced, which makes the background highly cluttered and nonhomogeneous. Hence, the Gaussian assumption of the RX detector results in a high FAR. The SVDD is better able to model the non-Gaussian and possibly multimodal support of the background pixels and reduce the number false alarms.
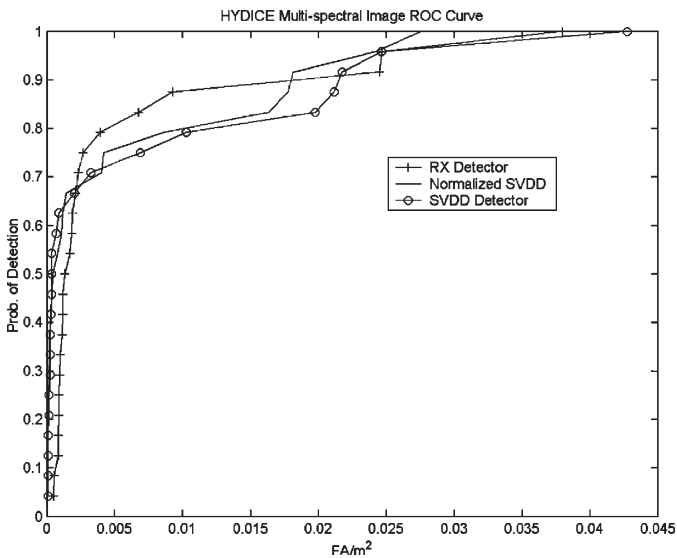
Fig. 8. ROC curve comparing SVDD and RX detectors on HYDICE imagery using seven bands. The 1280×320 image contains vehicles that are on roads and are partially camouflaged near the tree lines.
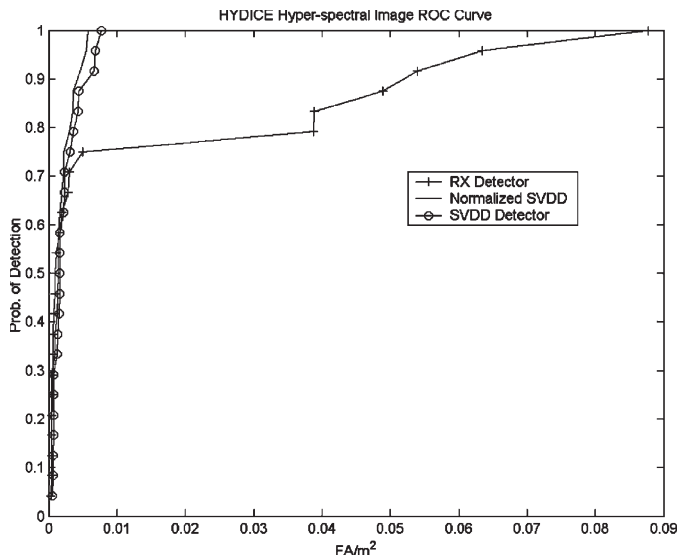

Fig. 9. ROC curve comparing SVDD and RX detectors on HYDICE imagery using 133 bands. The 1280×320 image contains vehicles that are on roads and are partially camouflaged near the tree lines.

From the curves in Figs. 6 and 7, similar results are observed for the image at 2000-ft altitude. Processing the multispectral data, the SVDD detectors outperform the RX detector throughout the ROC curve. For 100% detection, they generate half of the false alarms of the RX. When analyzing hyperspectral data, the ROC curves for the SVDD detectors remain consistent, yielding an FAR of 0.01 FA/m^2 for 100% detection. For this image, the SVDD reduces the FAR by a factor of nearly 2.5 instead of 10. One possible explanation for the difference is that the background for the higher altitude image is more homogeneous and, therefore, does not violate the unimodal assumption of the RX algorithm. It is also instructive to examine the signal-to-interference-noise ratio (SINR) for the targets in the two images. The SINR for each target is computed by using the target spectrum and its covariance matrix estimated from a local neighborhood of pixels. For the 1000-ft image, the range of SINR values is between 23 and 28 dB, for the multispectral data, and 20 to 26 dB for the hyperspectral data. For the 2000-ft image, the SINR ranges between 26 and 29 dB for both multi- and hyperspectral versions of the image. These numbers suggest that for lower SINR levels, the SVDD detectors offer significant improvement over the RX algorithm. As the SINR increases, the performance gain between the algorithms appears to lessen.

The algorithms are also evaluated using the Forest Radiance II dataset to detect vehicles in various degrees of concealment. The HYDICE sensor was used to image vehicles on dirt roads in a heavily forested region at 1-m GSD. The background consists of foliage, grass, and roads, and is smoothly varying and homogeneous. The size of the image used in this study is 1280×320 pixels. For multispectral processing, the dimensions of the inner and outer regions for the hollow background window are 10×10 and 13×13, respectively. For the hyperspectral, the dimensions of the inner and outer regions for the hollow background window are 10×10 and 17×17, respectively.

The ROC curves in Figs. 8 and 9 show that all of the detectors work equally well in the low false-alarm region. In this area of the curve, the vehicles are predominantly in the open, so they are easier to detect. Using the hyperspectral data (Fig. 9), there is a clear separation in the ability of the RX and SVDD algorithms to detect targets partially occluded along the tree lines. Since the local background of these targets contains multiple types of terrain, a multimodal representation for the background is required. In such cases, the SVDD is able to compute a more robust detection threshold than the RX detector. Furthermore, the SINR values for the targets in the multispectral image are above 30 dB. In the hyperspectral image, the SINR ranges from 27 to 33 dB. At such high signal-to-noise ratios, the targets are well separated from the background. Therefore, detectors, such as the RX that uses linear decision boundaries, perform similarly to nonlinear detectors, such as the SVDD (as seen in Fig. 8).

We summarize the results with two observations. First, the SVDD detectors are particularly useful for detecting low SINR targets. The spectra for these targets may not be linearly separable from the background, therefore, a nonlinear decision boundary is required. Second, the SVDD detectors work well with both multispectral and hyperspectral data. They are able to incorporate the extra information provided by the addition spectral bands to reduce the number of false alarms for 100% detection. In contrast, the performance of the RX detector appears to be sensitive to the number of bands. Among the reasons for the difference are: 1) the Gaussian assumption of the RX may be less valid for higher dimensional data, as explained in [14] and 2) robust estimation of the covariance matrix is more difficult as the number of bands increases.

Computational Complexity: The ROC curves in Figs. 4–9 illustrate the increase in detection performance offered by the SVDD approach. In addition to reducing the FAR, anomaly detectors should also be computationally efficient in order to
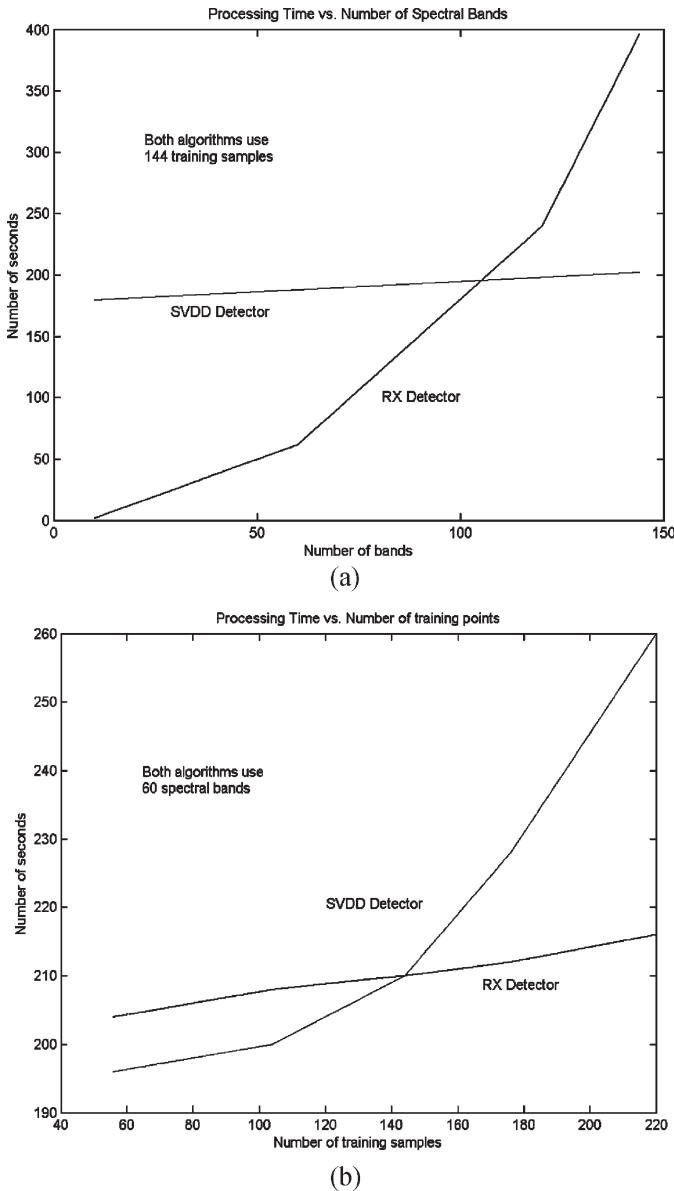
(a)

(b)

Fig. 10. Illustration of how the computational complexity of the RX and SVDD anomaly detectors varies with the number of spectral bands and the number of training samples.

process a datacube in near real time. Thus, we consider the computational complexity of the RX and SVDD algorithms.

The two critical factors that impact the speed of the detectors are the number of spectral bands and the number of pixels in the hollow window used to estimate the background model. For the RX detector, it has been shown [13] that its computational complexity is linear with respect to the number of pixels used to estimate the background statistics. Since it needs to estimate and invert a $B \times B$ covariance matrix (where B is the number of spectral bands), the number of floating-point operations per second for the RX detector is approximately quadratic with respect to the number of bands [13].

For the SVDD, the converse is true. By avoiding the need to invert large covariance matrices, the complexity of the SVDD increases linearly with the number of bands, as only dot products need to be computed. However, the speed of the SVDD

detectors scales with the training-set size [17]. They require the inversion of an $N \times N$ kernel matrix, where N is the number of background pixels used to train the classifier at each pixel. Hence, their complexity is exponential with respect to the size of the background window.

In Fig. 10, the processing times for both algorithms as a function of the number of spectral bands and number of training samples are shown. The algorithms are implemented in C++ on a 2.8-GHz Pentium machine, and the code has not been fully optimized. We measured the CPU time required to process a 256×100 pixel image. Fig. 10(a) shows that for a fixed number of training samples (144 local-background pixels), the processing time for the SVDD increases linearly with the number of spectral bands, while the RX increases quadratically. In Fig. 10(b), it can be seen that the converse is true. For a fixed number of bands (60 in this example), the SVDD increases quadratically with the number of training samples, while the RX increases linearly.

V. CONCLUSION

We note that other algorithms, such as the mixture of Gaussian anomaly detector and the subspace RX (SSRX), produce similar reductions in the FAR. However, the proposed SVDD approach for anomaly detection has several key differences.

1) It is nonparametric and has the ability to model non-Gaussian background clutter.
2) It can model nonhomogeneous backgrounds using mul-timodal distributions without making any prior assumptions (i.e., number of modes).
3) It avoids the need to invert large covariance matrices when processing hyperspectral data.

We have also used a geometric interpretation of the SVDD to derive normalized metric that is appropriate for anomaly detection in spectral imagery. ROC curves that compare the performance of the RX and the proposed detector show a marked reduction in the number of false alarms when using the SVDD. Further analysis of the background clutter and SINR for the targets to explain the differences in performance is also provided.

REFERENCES

[1] X. Lu, L. E. Hoff, I. S. Reed, M. Chen, and L. B. Stotts, "Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 143–156, Jan. 1997.

[2] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE—27th AIPR Workshop*, 1998, vol. 3584, pp. 221–232.

[3] J. A. Gualtieri, S. R. Chettri, R. F. Cromp, and L. F. Johnson, "Support vector machines classifiers as applied to AVIRIS data," in *Proc. Summaries 8th JPL Airborne Earth Sci. Workshop: JPL Publication 99-17*, Feb. 1999, pp. 217–227.

[4] P. Gandhi and S. Kassam, "Analysis of CFAR processors in non-homogeneous backgrounds," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 427–445, Jul. 1988.

[5] H. M. Finn, "A CFAR design for a window spanning two clutter fields," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 155–169, Mar. 1986.

[6] A. Banerjee, P. Burlina, and R. Chellappa, "Adaptive target detection in foliage-penetrating SAR images using alpha-stable models," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1823–1831, Dec. 1999.

[7] D. M. J. Tax and R. P. W. Duin, "Data domain description using support vectors," *Proc. Eur. Symp. Artif. Neural Netw.*, M. Verleysen, Ed., Brussels, Belgium, Apr. 1999, pp. 251–256.

[8] D. M. J. Tax, A. Ypma, and R. P. W. Duin, "Support vector data description applied to machine vibration analysis," *Proc. 5th Annu. Conf. Adv. School Comput. Imag.*, M. Boasson, J. Kaandorp, J. Tonino, and M. Vosselman, Eds., Heijen, The Netherlands, Jun. 15–17, 1999, pp. 398–405.

[9] D. Stein, S. Beaven, L. E. Hoff, E. Winter, A. Shaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[10] N. Henze and T. Wagner, "A new approach to the BHEP tests for multivariate normality," *J. Multivar. Anal.*, vol. 62, no. 1, pp. 1–23, 1997.

[11] S. G. Beaven, D. Stein, and L. E. Hoff, "Comparison of Gaussian mixture and linear mixture models for classification of hyperspectral data," in *Proc. IGARSS*, Honolulu, HI, Jul. 2000, pp. 1597–1599.

[12] A. D. Stocker and A. P. Schaum, "Application of stochastic mixing models to hyperspectral detection problems," *Proc. SPIE*, vol. 3071, pp. 47–60, 1997.

[13] S. Schweizer and J. M. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.

[14] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[15] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," Microsoft Research, Redmond, WA, TR MSR 99-87, 1999.

[16] B. Scholkopf, R. Williamson, A. J. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Proc. Neural Inf. Process. Syst.*, S. A. Solla, T. K. Leen, and K. R. Müller, Eds., 2000, pp. 582–588.

[17] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[18] C. Lai, D. M. J. Tax, R. P. W. Duin, E. Pekalska, and P. Paclik, "A study on combining image representations for image classification and retrieval," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 18, no. 5, pp. 867–890, 2004.

[19] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[20] H. Kwon and N. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.

Amit Banerjee (S'99–M'01) received the B.S. degree in electrical engineering from the University of Illinois, Urbana–Champaign, in 1994, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2000.

After serving as a faculty Research Assistant at the Center for Automation Research/Computer Vision Laboratory at the University of Maryland, he joined the Applied Physics Laboratory, Johns Hopkins University, Baltimore, MD, in July 2000, where he is currently a Senior Research Assistant. He has directed several research programs on developing methods for automated analysis of hyperspectral and video imagery for operational use. He has made contributions in the areas of non-Gaussian estimation and detection for remotely sensed hyperspectral and SAR data. His research interests also include computer vision and medical image analysis.

Philippe Burlina (S'92–M'93) received the B.S. degree in computer science from the Université de Technologie de Compiegne, Compiegne, France, in 1988, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 1994.

He is an Assistant Group Supervisor at the Applied Physics Laboratory, Research and Technology Development Center, Johns Hopkins University, Baltimore, MD. Prior to that, he held various industry R&D positions including, most recently, as a Director of software development at FileNet, which is a public company developing enterprise software platforms. His R&D expertise and interests include image/video exploitation, medical imaging, machine perception and learning, and content management/enterprise software platforms.

Chris Diehl (S'94–M'00) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 2000.

He is a Senior Research Scientist with the Research and Technology Development Center at the Applied Physics Laboratory (APL), Johns Hopkins University (JHU), Baltimore, MD, and an Assistant Research Professor with the JHU Department of Electrical and Computer Engineering. Since joining JHU/APL, he has led a number of research efforts developing and applying statistical machine-learning techniques to problems in video surveillance, hyperspectral anomaly detection, sonar classification, and biometric ID. His recent projects include work on sparse kernel machine learning for ultralow-power pattern classification in very large scale integration and entity resolution in informal online communications. His research interests include large-margin classification, kernel methods, incremental and active learning, and generalization-performance assessment.