

8

One-class SVMs for hyperspectral anomaly detection

Amit Banerjee, Philippe Burlina and Chris Diehl

Applied Physics Laboratory, The Johns Hopkins University, USA

A support vector framework for hyperspectral anomaly detection is developed in this chapter. Conventional methods for detecting anomalies in hyperspectral images are based on the popular Reed–Xiaoli (RX) detector. However, these algorithms typically suffer from a large numbers of false alarms, due to the assumptions that the background is Gaussian and homogeneous. In practice, these assumptions are often violated, especially when the neighbourhood of a pixel contains multiple types of terrain.

To remove these assumptions, a novel anomaly detector is proposed and derived that incorporates a nonparametric background model based on the Support Vector Data Description (SVDD). The SVDD is a one-class support vector classifier that can model the support of a distribution. Expanding on prior work, a geometric interpretation of the SVDD is developed to propose a decision rule that utilizes a new test statistic and shares some of the properties of constant false-alarm rate (CFAR) detectors. Two versions of the algorithm are presented to detect either local or global anomalies and an analysis of their computational requirements is provided. Using receiver operating characteristic (ROC) curves, the improved performance and reduction in the false alarm rate when using the SVDD-based detector are demonstrated on Wide Area Airborne Mine Detection (WAAMD) and Hyperspectral Digital Imagery Collection Experiment (HYDICE) hyperspectral imagery.

8.1 Introduction

Recent advances in hyperspectral sensors with high spectral and spatial resolution have led to 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 (Scharf and Friedlander 1994) 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 (a) accurate spectral calibration, (b) compensating for atmospheric effects to convert radiance spectra to the reflectance domain, or (c) 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 due to simplifying assumptions imposed on the background signature distribution. This chapter presents a new kernel-based approach that removes these assumptions in order to detect spectral anomalies while mitigating the false alarm rate.

Lacking prior target signatures, anomaly detection methods model the background and find pixels in the scene that 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 labelled as part of the background; if it lies outside of the background's region of support, it is detected as an outlier and declared 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 making assumptions about the distribution.

The RX algorithm, which is the benchmark anomaly detector for hyperspectral imagery, uses the local Gaussian model (Lu *et al.* 1997). With this approach, the background pixels in a local neighbourhood 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 with a threshold to detect an anomaly (Lu *et al.* 1997).

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 (Stein *et al.* 2002) 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 (Henze and Wagner 1997). 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.

If the local background contains multiple types of terrain, the background cannot be properly modelled by a unimodal distribution. To more properly characterize non-homogeneous backgrounds, researchers have employed the mixture of Gaussians model (Beaven *et al.* 2000).

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 (Stocker 1999). 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 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 Schweizer and Moura (2001), 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 a GMRF. The estimated parameters of the GMRF are then used in a GLRT to detect 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 poor performance when targets are located along clutter boundaries.

To perform 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 (Vapnik 1998) 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.

Large margin techniques such as support vector machines (SVMs) have received considerable attention for classification tasks with 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 (Gualtieri *et al.* 1999; Gualtieri and Cromp 1998) this work seeks to 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) (Tax *et al.* 1999; Tax and Duin 1999) and the ν -SVM (Scholkopf *et al.* 1999) methods. These classifiers are able to estimate directly 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 (Tax *et al.* 1999) and image retrieval (Lai *et al.* 2004). In the following, the SVDD is utilized to detect spectral anomalies that lie outside the region of support for the background pixels.

For anomaly detection, the SVDD approach has the following benefits: *Non-parametric*: 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 with other classical methods (Tax *et al.* 1999; Vapnik 1998); *Use of kernels*: by exploiting the ‘kernel trick’, the SVDD method is able to model accurately the support of non-trivial, multi-modal distributions. Therefore, the SVDD-based anomaly detector yields a more powerful detection performance for targets embedded in non-Gaussian and non-homogeneous backgrounds.

The recently introduced kernel RX algorithm (Kwon and Nasrabadi 2006) 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 from

the SVDD approach. While the SVDD avoids the problem of density estimation, the kernel RX assumes a Gaussian distribution in the kernel space. This assumption is the equivalent to a Parzen estimate for the distribution in the original input space (Kwon and Nasrabadi 2006). 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 chapter is organized as follows: Section 8.2 provides an overview of the derivation of the SVDD one-class classifier. A discussion on optimizing the SVDD function is given in Section 8.3. The algorithm to detect hyperspectral anomalies using the SVDD is presented in Section 8.4, along with a discussion of a new SVDD test statistic. Experimental results are provided in Section 8.5 to evaluate the RX and SVDD anomaly detectors. Finally, concluding remarks are given in Section 8.6.

8.2 Deriving the SVDD

Following Tax and Duin (1999), this section describes the derivation of the SVDD. The linear SVDD that models the support of the distribution as the minimal enclosing hypersphere containing the data in the original input space is introduced. Then it is generalized 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, it should be noted 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 (Scholkopf *et al.* 2000). It has been shown (Scholkopf *et al.* 1999, 2000) that these methods are equivalent when the radial basis function (RBF) is chosen as the kernel function, as is common practice.

8.2.1 The linear SVDD

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

$$\min(R) \text{ subject to } \mathbf{x}_i \in S, i = 1, \dots, M. \quad (8.1)$$

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

$$L(R, \mathbf{a}, \alpha) = R^2 - \sum_i \alpha_i [R^2 - \langle \mathbf{x}_i, \mathbf{x}_i \rangle - 2\langle \mathbf{a}, \mathbf{x}_i \rangle + \langle \mathbf{a}, \mathbf{a} \rangle]. \quad (8.2)$$

The first term in (8.2) is the radius that is to be minimized. The second term constrains each training point \mathbf{x}_i to lie within the sphere with centre \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:

$$\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_i \alpha_i = 1 \quad (8.3)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \Rightarrow \mathbf{a} = \frac{\sum_i \alpha_i \mathbf{x}_i}{\sum_i \alpha_i}. \quad (8.4)$$

Combining Equations (8.3) and (8.4), a simple expression for the centre of the sphere is found:

$$\mathbf{a} = \sum_i \alpha_i \mathbf{x}_i. \quad (8.5)$$

The centre of the sphere is the centre of mass of all training points, with the weights equal to the Lagrange multipliers. Furthermore, the inequality $\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2$ implies $\alpha_i \geq 0$, according to the non-negativity constraint of the KKT. Substituting Equations (8.3) and (8.5) into (8.2) yields another expression for the Lagrangian function to be maximized with respect to the α_i :

$$L = \sum_i \alpha_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle - \sum_{i,j} \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \quad (8.6)$$

After optimizing L with respect to α_i , it is typical to discover that a large fraction of the α_i are equal to zero. The training examples with nonzero α_i are called support vectors and lie on the boundary of the support region. This is a result of the complementary slackness condition of the KKT that states that $(\|\mathbf{x}_i - \mathbf{a}\|^2 - R^2)\alpha_i = 0$ for optimality. Hence, the support vectors 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 vectors.

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 $\|\mathbf{y} - \mathbf{a}\|^2 > R^2$ holds true for a test example \mathbf{y} , the example is labelled as an outlier. Expanding

$$\|\mathbf{y} - \mathbf{a}\|^2 = (\mathbf{y} - \mathbf{a})^\top (\mathbf{y} - \mathbf{a}) \quad (8.7)$$

$$= \left(\mathbf{y} - \sum_i \alpha_i \mathbf{x}_i \right)^\top \left(\mathbf{y} - \sum_i \alpha_i \mathbf{x}_i \right) \quad (8.8)$$

$$= \langle \mathbf{y}, \mathbf{y} \rangle - 2 \sum_i \alpha_i \langle \mathbf{y}, \mathbf{x}_i \rangle + \sum_{i,j} \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \quad (8.9)$$

the decision rule becomes:

$$\langle \mathbf{y}, \mathbf{y} \rangle - 2 \sum_i \alpha_i \langle \mathbf{y}, \mathbf{x}_i \rangle + \sum_{i,j} \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle > R^2, \quad (8.10)$$

which is referred to as the linear SVDD decision rule.

8.2.2 The 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 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, the algorithm seeks the smallest hypersphere in the induced feature space $S = \{\boldsymbol{\phi}(\mathbf{x}) : \|\boldsymbol{\phi}(\mathbf{x}) - \mathbf{c}\|^2 \leq R^2\}$ including the entire set of mapped training examples $T = \{\boldsymbol{\phi}(\mathbf{x}_i), i = 1, \dots, M\}$. Therefore, the following constrained optimization problem needs to be solved:

$$\min(R) \text{ subject to } \boldsymbol{\phi}(\mathbf{x}_i) \in S, i = 1, \dots, M. \quad (8.11)$$

The corresponding Lagrangian is expressed as:

$$L(R, \mathbf{a}, \alpha) = R^2 - \sum_i \alpha_i \left(R^2 - \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_i) \rangle - 2\langle \mathbf{c}, \boldsymbol{\phi}(\mathbf{x}_i) \rangle + \langle \mathbf{c}, \mathbf{c} \rangle \right) \quad (8.12)$$

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:

$$L = \sum_i \alpha_i \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_i) \rangle - \sum_{i,j} \alpha_i \alpha_j \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle \quad (8.13)$$

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

$$\mathbf{c} = \sum_i \alpha_i \boldsymbol{\phi}(\mathbf{x}_i). \quad (8.14)$$

As above, the hypersphere centre is the centre 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 $\text{SVDD}(\mathbf{y}) = \|\boldsymbol{\phi}(\mathbf{y}) - \mathbf{c}\|^2 > R^2$, which expands to

$$\begin{aligned} \text{SVDD}(\mathbf{y}) &= \|\boldsymbol{\phi}(\mathbf{y}) - \mathbf{c}\|^2 \\ &= \|\boldsymbol{\phi}(\mathbf{y}) - \sum_{i=1}^N \alpha_i \boldsymbol{\phi}(\mathbf{x}_i)\|^2 \\ &= \langle \boldsymbol{\phi}(\mathbf{y}), \boldsymbol{\phi}(\mathbf{y}) \rangle - 2 \sum_i \alpha_i \langle \boldsymbol{\phi}(\mathbf{y}), \boldsymbol{\phi}(\mathbf{x}_i) \rangle + \sum_{i,j} \alpha_i \alpha_j \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle. \end{aligned}$$

Note that the above expression consists of inner products of the mapping function $\boldsymbol{\phi}$. 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 map implicitly the data into the induced feature space and compute the inner product. The only requirement imposed on the kernel function is that it satisfy Mercer’s theorem (Vapnik 1998). Mercer’s theorem states that in order for a kernel function $K(\mathbf{x}, \mathbf{y}) : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}, (\mathbf{x}, \mathbf{y}) \mapsto z = K(\mathbf{x}, \mathbf{y})$ to admit

an eigenfunction expansion of the form

$$K(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{y}_i) \rangle = \sum_{k=1}^{\infty} b_k \phi(\mathbf{x}) \phi(\mathbf{y}), \tag{8.15}$$

it must be positive and semi-definite:

$$\int K(x, y)g(x)g(y)dx dy \geq 0 \quad \forall g(\cdot) \in L_2. \tag{8.16}$$

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

$$SVDD(\mathbf{y}) = K(\mathbf{y}, \mathbf{y}) - 2 \sum_i \alpha_i K(\mathbf{y}, \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) > R^2. \tag{8.17}$$

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.

This work employs the popular Gaussian radial basis function (RBF) as the kernel function, defined as:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right). \tag{8.18}$$

The choice of the RBF simplifies the SVDD(y) test statistic to:

$$SVDD(\mathbf{y}) = 1 - 2 \sum_i \alpha_i K(\mathbf{y}, \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) > R^2, \tag{8.19}$$

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 Figure 8.1 (Tax and Duin 1999). By varying the scale parameter of the RBF, the

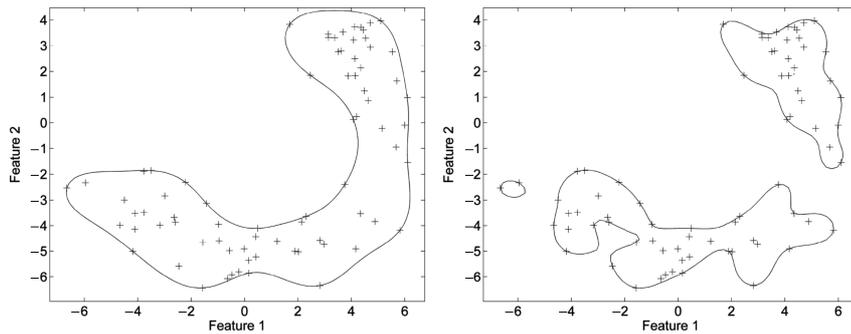


Figure 8.1 SVDD decision boundary using $\sigma = 5$ (left) and $\sigma = 3$ (right). Note how using a smaller value for sigma leads to a tighter decision boundary. (Modified from Banerjee, A., Burlina, P. and Diehl, C.P. (2006) A support vector method for anomaly detection in hyperspectral imagery, *IEEE Trans. on Geoscience and Remote Sensing*, 44(8), 2282–2291)

SVDD can determine multiple regions of support for a dataset. As can be seen Figure 8.1, this allows the SVDD to model multi-modal distributions. For anomaly detection, this implies that the SVDD can detect targets embedded in multiple types of clutter with fewer false alarms.

8.3 SVDD function optimization

Finding the SVDD decision boundary requires the following optimization:

$$\min_{\sigma} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \text{ subject to } 0 \leq \alpha_i \leq 1, \sum \alpha_i = 1. \quad (8.20)$$

Traditional support vector algorithms use standard quadratic programming (QP) techniques such as active set and conjugate gradient methods to minimize the function. These methods have a complexity of $\mathcal{O}(N^3)$, where N is the number of training samples. However, given a large number of training samples, standard QP techniques cannot solve the problem efficiently. They involve a matrix whose size scales quadratically with the number of active constraints (non-zero α_i s) and therefore require a lot of memory and a large number of iterations to arrive at a solution. For hyperspectral anomaly detection, this is a critical issue since many training pixels are needed to characterize the statistics of high-dimensional spectra.

To address this issue, a ‘chunking’ method was introduced in Vapnik (1982) to solve the dense SVM QP problem. The chunking algorithm uses the fact that the value of the quadratic function is the same if you remove the rows and columns of the matrix that correspond to zero Lagrange multipliers. Therefore, the large QP problem can be broken down into a series of smaller QP problems, whose ultimate goal is to identify and remove all of the nonzero Lagrange multipliers. Chunking reduces the size of the matrix from the number of training examples squared to approximately the number of nonzero Lagrange multipliers squared. However, chunking still cannot handle large-scale training problems, since even this reduced matrix cannot fit into memory.

In Osuna *et al.* (1997), a new class of QP algorithms was proposed by Osuna *et al.* for SVMs. They demonstrate that the large QP problem can be broken down into a series of smaller QP sub-problems. Sequential Minimal Optimization (SMO) is an example of such a fast method to find the solution for a very large QP optimization problem. SMO breaks this large QP problem into a series of QP problems involving only two examples. These small QP problems are solved analytically, which avoids using a time-consuming numerical QP optimization as an inner loop. The amount of memory required for SMO is linear in N , which allows SMO to handle very large training sets.

The results presented in Section 8.5 utilizes the SVDD implementation found in the NTU SVM software package (Chang and Lin 2001). This implementation uses the SMO minimization method to find the optimal hypersphere. To illustrate the increase in speed when using the SMO method, the SVDD algorithm is run using both the standard QP and SMO methods on the same data with the same value for sigma while varying the number of training samples. Plots of the run-times as a function of the number of samples are given for both methods in Figure 8.2. The plots show that the runtime for the standard QP increases exponentially with the number of training sample, while SMO increases linearly. Furthermore, SMO is an order-of-magnitude faster. This is a key benefit to using SMO for local hyperspectral anomaly detection, since the optimal SVDD decision boundary is computed for every pixel in the image.

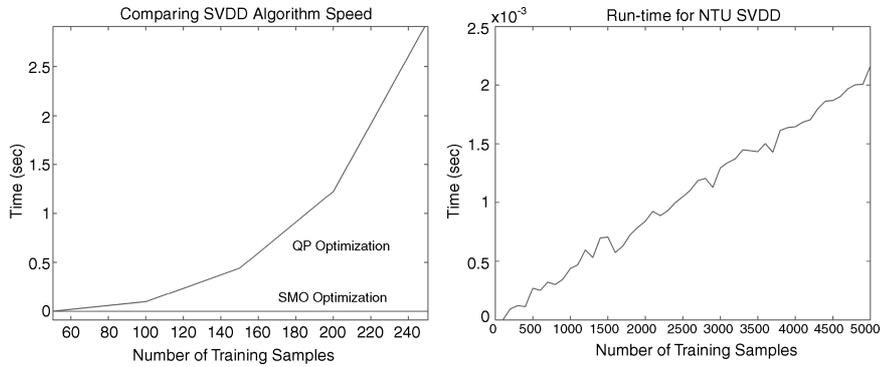


Figure 8.2 Comparing the speed of QP and SMO optimization methods. (Left) The runtime of the standard QP solver increases exponentially with the number of training samples, while SMO increases linearly. (Right) SMO runtime increases linearly with the number of samples.

Finally, it should be noted that incremental learning/update techniques found in the machine learning literature can be used to further improve the speed of local anomaly detectors (Syed *et al.* 1999). The local background of neighbouring pixels greatly overlap and share a large percentage of the same training data. Instead of retraining the SVDD at each pixel, incremental learning methods can efficiently update the SVDD model by considering only the new background spectra in the local neighbourhood.

8.4 SVDD algorithms for hyperspectral anomaly detection

This section presents local and global anomaly detection algorithms based upon the SVDD. A local anomaly detector finds pixels in the image that have different spectra from the clutter background in the local neighbourhood. A global anomaly detector finds pixels whose spectra are different from the other spectra in the entire image. The RX algorithm can also be used either as a local or global anomaly detector, depending on how the background samples are collected.

Both approaches have complementary advantages and disadvantages. Local anomaly detectors can use simpler models to describe the local background. However, they are computationally intensive and are susceptible to false alarms that are isolated spectral anomalies. For global background modelling, the background is estimated once for the entire hypercube. Since the global background contains many materials, the modelling required is much more complex than that required locally. They are often faster than the local anomaly detectors since the background clutter model is not computed for each pixel. Furthermore, global methods also provide the capability for enhanced ROC performance due to: (a) more accurate background characterization in the presence of closely spaced targets that is corrupted when a local model is used, and (b) the rejection of local anomalies that are not truly anomalous in the scene (e.g. isolated trees and shadows).

8.4.1 Outline of algorithms

The multi-modal background characterization of SVDD is virtually automatic and does not require any a priori knowledge of background characteristics (e.g. the number of components

to be used in a mixture model). As a result, SVDD is well suited for global background characterization and is able to achieve high detection performance and low false alarm rates with extremely rapid computational processing.

The steps for the global SVDD anomaly detector are as follows.

1. Randomly select a set of N background pixels from the image as the training set.
2. Given the set of background spectra, estimate an optimal value for σ , the scale parameter of the RBF kernel.
3. Using the spectra from the training pixels, estimate the SVDD parameters (\mathbf{a} , α_i , R) to model the region of support for the background clutter.
4. For each pixel in the image perform the decision test:
 - If $\text{SVDD}(\mathbf{y})$, the SVDD test statistic for pixel \mathbf{y} , is less than the detection threshold t , the pixel is part of the background.
 - Else, declare the pixel as an anomaly.

Note that since this is a global method, only the last step needs to be performed for each pixel, as the training is done only once for the entire image. To reduce the computational expense of the algorithm further, a closer look at the $\text{SVDD}(\mathbf{y})$ (8.19) yields:

$$\text{SVDD}(\mathbf{y}) = \kappa - c(\mathbf{y}) \sum_i \alpha_i C_i \exp\left(\frac{-2\mathbf{y}^\top \mathbf{x}_i}{\sigma^2}\right), \quad (8.21)$$

where $\kappa = 1 + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$, $c(\mathbf{y}) = \exp(-\mathbf{y}^\top \mathbf{y} / \sigma^2)$ and $C_i = \exp(-\mathbf{x}_i^\top \mathbf{x}_i / \sigma^2)$ may be pre-computed at the beginning after training when the support vectors \mathbf{x}_i have been found. Therefore, only the inner-product in the exponential term on the right-hand side of the above equation needs to be computed at every pixel.

The local SVDD anomaly detector is similar to 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 $\text{SVDD}(\mathbf{y})$ measure. The steps for the local SVDD based anomaly detector are as follows.

1. Select the dimensions of the hollow background window.
2. Estimate σ , the scale parameter of the RBF kernel.
3. Given the window size and scale parameter, for each pixel:
 - Sample pixels from the local neighbourhood using the background window.
 - Using these pixels, compute the SVDD parameters (\mathbf{a} , α_i , R) to model the region of support for the background clutter.
 - Decision:
 - If $\text{SVDD}(\mathbf{y})$, the SVDD test statistic for pixel \mathbf{y} is less than the detection threshold T , the pixel is part of the background.
 - Else, declare the pixel as a target.

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

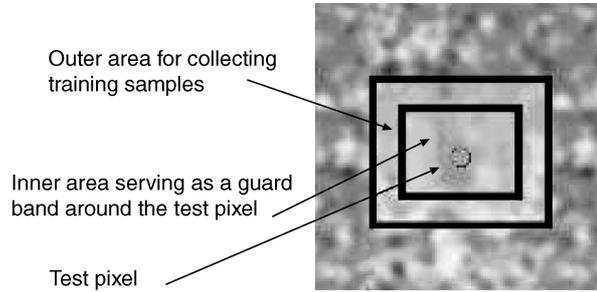


Figure 8.3 Example of a hollow window used to collect pixels for background estimation.

8.4.2 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 centred at the pixel. The window consists of two regions, the inner and outer window regions, as shown in Figure 8.3.

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 background estimation.

8.4.3 Estimating sigma

Choosing an appropriate value for the RBF kernel scale parameter sigma is critical for acceptable detection performance. Typically, the performance of anomaly detectors is commonly characterized by two measures of error: the false alarm rate (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 directly.

To reliably estimate sigma, a Neyman–Pearson framework is developed based on two observations. First, it has been shown (Tax *et al.* 1999; Scholkopf *et al.* 2000; Tax and Duin 1999) that an upper bound on the false alarm rate for the SVDD is:

$$P_{FA} \leq \mathbb{E} \left[\frac{\#SV}{N} \right]. \quad (8.22)$$

In this expression, P_{fa} is the false alarm rate, $\#SV$ is the number of support vectors in the SVDD model and N is the total number of training samples.

The second observation is that anomaly detectors implicitly assume that the target space is simply the complement of the support region of the background training samples. Without knowledge of the target spectra and their distribution, anomaly detectors attempt to model the support region of the background spectra. They consider any pixels whose spectra lie outside the support region as targets.

Analogous to the Neyman–Pearson framework, an estimate for sigma that seeks to maximize the probability of detection while maintaining a desired false alarm rate can be defined as follows:

$$\hat{\sigma} = \max_{\sigma} (P_D) \text{ such that } P_{FA} \leq \tau, \quad (8.23)$$

where P_D is the probability of detection and τ is the desired false alarm rate. Utilizing the two observations above,

$$\hat{\sigma} = \max_{\sigma} \frac{1}{C} \int_{R^c} \mathbf{1}(\mathbf{x}) d\mathbf{x} \text{ such that } \mathbb{E} \left[\frac{\#SV}{N} \right] \leq \tau \quad (8.24)$$

$$= \max_{\sigma} \text{volume}(R^c) \text{ such that } \mathbb{E} \left[\frac{\#SV}{N} \right] \leq \tau \quad (8.25)$$

$$= \min_{\sigma} \text{volume}(R) \text{ such that } \mathbb{E} \left[\frac{\#SV}{N} \right] \leq \tau, \quad (8.26)$$

where $\mathbf{1}(\cdot)$ is the indicator function and R is the support region for the background pixels, and C is a normalizing constant. The probability of detection is approximated as the volume of the feature space that lies outside of the support region (R^c). Thus, the estimator seeks the value of sigma that (a) minimizes the size of the support region for the background spectra and (b) generates an average fraction of support vectors that is less than the desired false alarm rate.

As shown in Tax *et al.* (1999), Scholkopf *et al.* (2000) and Tax and Duin (1999), the choice of sigma affects the size of the support region and the number of support vectors. Figure 8.1 illustrates the SVDD support regions for a set of training data with two different values for sigma. It shows that if the value of sigma is small, the resulting support region is small and the decision boundary overfits the data. If sigma is big, the support region is large and the classifier underfits the data with a possibly trivial decision boundary. Since sigma is proportional to the volume of the support region, the estimate for sigma can be expressed as:

$$\hat{\sigma} = \min_{\sigma} \sigma \text{ such that } \mathbb{E} \left[\frac{\#SV}{N} \right] \leq \tau. \quad (8.27)$$

Furthermore, sigma is inversely proportional to the number of support vectors (Scholkopf *et al.* 2000; Tax and Duin 1999). Hence, the choice for sigma should be small enough to minimize the volume of the support region while large enough to maintain the desired false alarm rate.

Given the expectation in the above expression is over training sets of size N , the expectation from a single training set cannot be estimated. Therefore the resulting number of support vectors for the SVDD model is substituted in place of the expectation to obtain an approximate upper bound.

To avoid the computational expense of estimating sigma for every pixel, one can choose to estimate a value of sigma that minimizes the average false alarm rate across the entire image, i.e.

$$\frac{1}{M} \sum_{i=1}^M P_{FA_i} \approx \frac{1}{M} \sum_{i=1}^M \mathbb{E} \left[\frac{\#SV_i}{N} \right] \approx \frac{1}{M} \sum_{i=1}^M \frac{\#SV_i}{N} \leq \tau, \quad (8.28)$$

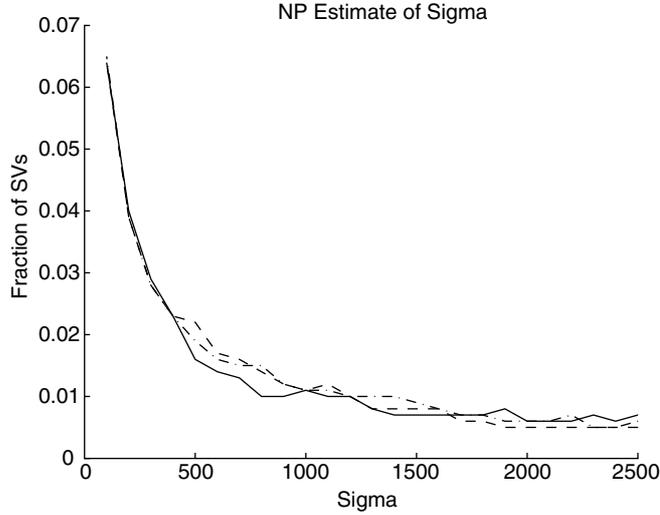


Figure 8.4 Fraction of support vectors as a function of sigma for three SVDD models. Sigma can be estimated using a Neyman–Pearson criterion.

where M is the number of training sets, N is the number of examples in each training set, P_{FA_i} is the i th training set, and $\#SV_i$ is the number of support vectors found for the i th training set. This approximation results in the following estimate for sigma:

$$\hat{\sigma} = \min_{\sigma} \sigma \text{ such that } \frac{1}{M} \sum_{i=1}^M \frac{\#SV_i}{N} \leq \tau. \quad (8.29)$$

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 smallest sigma such that the average fraction of support vectors is less than τ is the estimated value.

Figure 8.4 provides an example of the σ estimate as described above. In this example, three sets of 1000 training samples are randomly selected from a hyperspectral image. The SVDD classifier is trained on all three datasets with varying values for sigma. In this example, the smallest value for sigma that satisfies the desired false alarm rate of 0.01 is 1400, which is used as the estimate.

8.4.4 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 compare the pixels directly and identify those that are more likely to be anomalies. This allows for the use of a single threshold for the entire image, which is a key consideration for local anomaly detectors. By varying this

global threshold, an empirical receiver operating characteristic (ROC) curve can be generated to evaluate the detector's performance.

One class of such detectors is the constant false alarm rate (CFAR) detectors. The CFAR property is useful because it allows for the determination of a single detection threshold that (a) generates a desired number of false alarms, (b) is independent of the estimated parameters, and (c) can be used for every pixel in the image.

The RX algorithm is an example of a CFAR detector (Beaven *et al.* 2000). 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

$$\int_{\mathbf{x}: f(\mathbf{x}) < T} f(\mathbf{x}) d\mathbf{x} = P_{FA}, \quad (8.30)$$

where $f(\mathbf{x}) \sim N(\mathbf{0}, I)$, and P_{FA} is the desired false alarm rate. 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, a normalized test statistic can be similarly derived. In contrast to the RX algorithm, the following normalization procedure is motivated geometrically and strictly avoids imposing assumptions about the underlying distribution. Recall from Section 8.2 that the nonlinear SVDD estimates the minimum enclosing hypersphere centred at \mathbf{a} with radius R in the feature space as the support region for the background pixels. From Equation (8.11), the distance between the mapped example $\phi(\mathbf{y})$ and the centroid of the hypersphere \mathbf{c} is:

$$\|\phi(\mathbf{y}) - \mathbf{c}\|^2 = 1 - 2 \sum_i \alpha_i K(\mathbf{y}, \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j), \quad (8.31)$$

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. Hence, a normalized version of the SVDD test statistic is given as:

$$\text{SVDD}_N(\mathbf{y}) = \frac{\|\phi(\mathbf{y}) - \mathbf{c}\|^2}{R^2} = \frac{1 - 2 \sum_i \alpha_i K(\mathbf{y}, \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)}{R^2}. \quad (8.32)$$

By dividing the original statistic by the squared radius, the $\text{SVDD}_N(\mathbf{y})$ 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 the 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 false alarm rate. However, if the data is non-Gaussian, the CFAR properties of the RX detector no longer hold. As the SVDD avoids estimating a PDF, it is difficult to compute the threshold analytically. Instead, its detection threshold can be determined empirically by using nonparametric approaches for data modelling. One example is to use extreme value theory (EVT) (Pickands 1975), a branch of statistics dealing with extreme deviations from the mean. EVT methods

attempt to model the tails of the distribution empirically and determine a detection threshold based upon a desired false alarm rate. As this is an emerging area of study, its usefulness will be examined in future work.

8.5 Experimental results

In this section, the computational and detection performance of the SVDD anomaly detectors are evaluated. The first set of experiments compares the local anomaly detectors, in particular the RX, SVDD and normalized SVDD algorithms. Examples of how the false alarm rate 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.

The second set of experiments compares these local anomaly detectors with the global SVDD algorithm. The computational complexity of the algorithms is discussed, and their detection capabilities are analysed using ROC curves.

Data overview

The detectors are compared 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 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 feet, with a GSD of approximately 4 and 8 inches, respectively. The images contain two types of mines whose 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 metre spatial resolution. Two images that contain the targets from detection experiments 2, 3 and 4 are considered. There are over 20 target vehicles in the scenes, including many that are partially hidden along tree lines. There are a handful of additional ‘transient’ non-military vehicles in the images that are not marked as ground truth, but used in the evaluation.

Pre-processing

The detection algorithms are evaluated on both multi-spectral and hyperspectral versions of the imagery. For hyperspectral data, the only pre-processing step involved manually removing the water and CO₂ absorption bands. The number of remaining bands is 145 for the WAAMD data and 133 for the HYDICE imagery. To generate multi-spectral data, an additional step was taken 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 near IR and two short-wave IR 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 covariance matrix is 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 of local anomaly detectors

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, SVDD(\mathbf{y}) and the $SVDD_N(\mathbf{y})$ 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, else it is a false alarm. In addition, detections arising from bad lines in the image are removed manually.

In Figures 8.5–8.8, 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, resulting in a non-homogeneous background. The mine field was imaged with the COMPASS sensor at altitudes of 1000 and 2000 feet, with a GSD of 4 and 8 inches, respectively. The size of the 1000 and 2000 ft. altitude images are 1400×256 pixels and 1200×256 pixels, respectively. For multi-spectral 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.

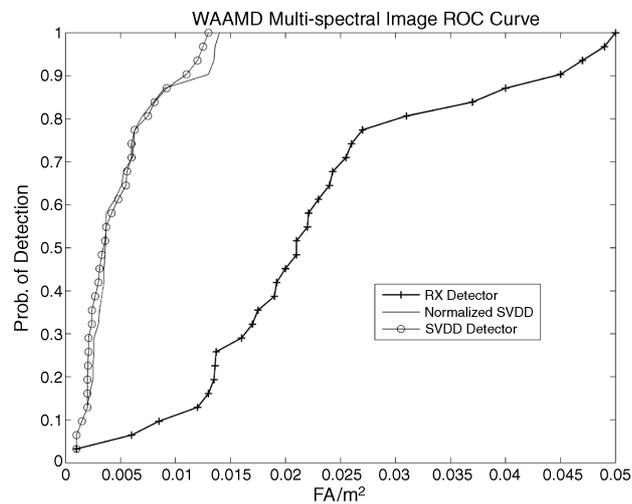


Figure 8.5 ROC curve comparing SVDD and RX detectors on WAAMD imagery using seven bands. The 1400×256 image contains 44 live mines in a dirt field and was taken at 1000 ft altitude.

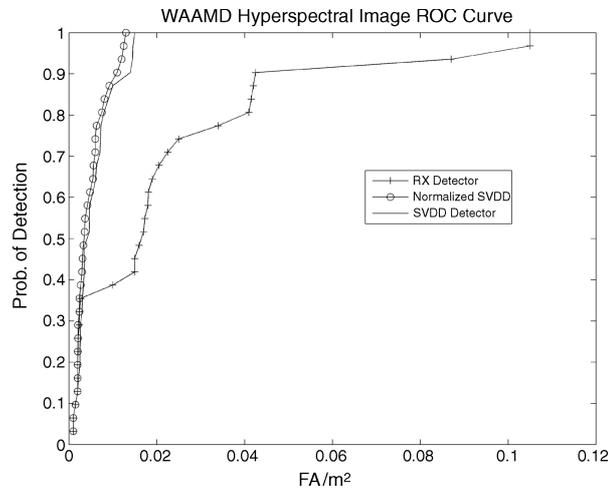


Figure 8.6 ROC curve comparing SVDD and RX detectors on WAAMD imagery using 145 bands. The 1400×256 image contains 44 live mines in a dirt field and was taken at 1000 ft altitude.

The ROC curves comparing the detectors on the 1000 ft. altitude image are shown in Figures 8.5 and 8.6. In the multi-spectral example of Figure 8.5, the SVDD detectors clearly outperform the RX algorithm throughout the curve. For 100% detection, the SVDD detectors reduce the false alarm rate by a factor of five. In the hyperspectral case (Figure 8.6), the SVDD detectors generate approximately the same false alarm rate as with the multi-spectral

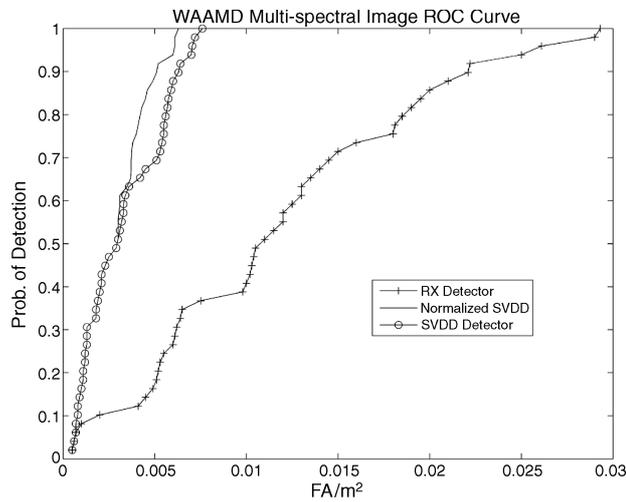


Figure 8.7 ROC curve comparing SVDD and RX detectors on WAAMD imagery using seven bands. The 1200×256 image contains 44 live mines in a dirt field and was taken at 2000 feet altitude.

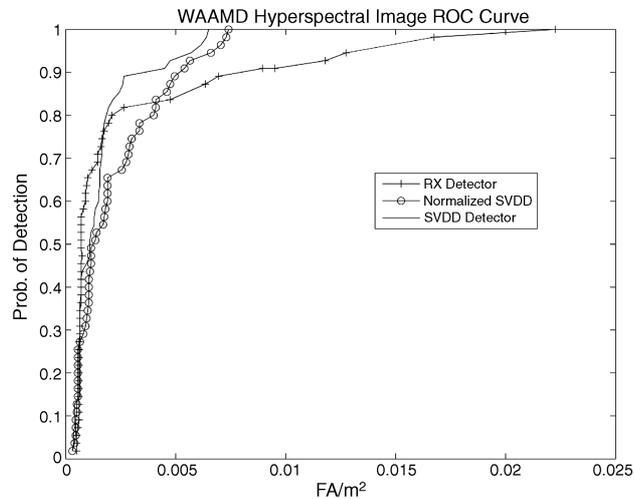


Figure 8.8 ROC curve comparing SVDD and RX detectors on WAAMD imagery using 145 bands. The 1200×256 image contains 44 live mines in a dirt field and was taken at 2000 feet altitude.

data (nearly 0.015 FA/m^2). However, the false alarm rate of the RX detector nearly doubles when processing hyperspectral data. In the hyperspectral representation of the image, the false alarm rate 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 inch GSD, the rocks and variations in the dirt field are more pronounced, making the background highly cluttered and non-homogeneous. Hence, the Gaussian assumption of the RX detector results in a high false alarm rate. The SVDD is better able to model the non-Gaussian and possibly multi-modal support of the background pixels and reduce the number false alarms.

From the curves in Figures 8.7 and 8.8, similar results are observed for the image at 2000 feet altitude. Processing the multi-spectral 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 analysing hyperspectral data, the ROC curves for the SVDD detectors remain consistent, yielding a false alarm rate of 0.01 FA/m^2 for 100% detection. For this image, the SVDD reduces the false alarm rate 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 neighbourhood of pixels. For the 1000 feet image, the range of SINR values is between 23 dB and 28 dB for multi-spectral data, and 20 to 26 dB for hyperspectral data. For the 2000 feet image, the SINR ranges between 26 dB 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.

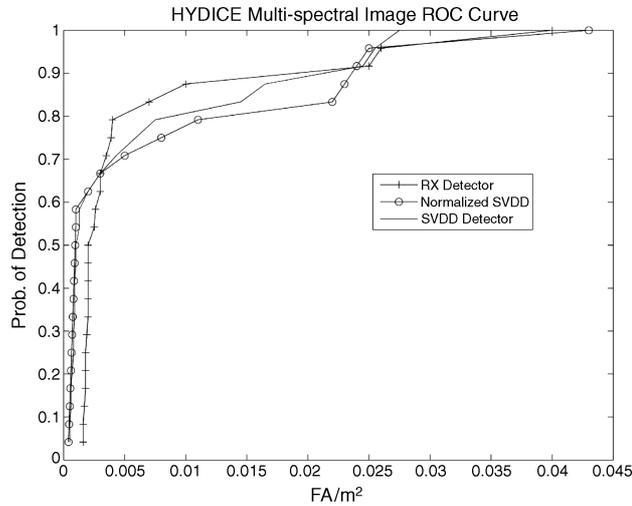


Figure 8.9 ROC curve comparing SVDD and RX detectors on HYDICE multi-spectral imagery. The 1280×320 image contains vehicles that are on roads and are partially camouflaged near tree lines.

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 1m 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 multi-spectral processing, the dimensions of the inner and outer regions for the hollow background window are 10×10 and 13×13 , respectively. For 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 Figures 8.9 and 8.10 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 hyperspectral data (Figure 8.10), there is clear separation in the ability of the RX and SVDD algorithms to detect targets partially occluded along tree lines. Since the local background of these targets contains multiple types of terrain, a multi-modal 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 multi-spectral image are above 30 dB. In the hyperspectral image, the SINR ranges from 27 dB 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 use linear decision boundaries perform similarly to nonlinear detectors, such as the SVDD (as seen in Figure 8.9).

The results are summarized 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, so a nonlinear decision boundary is required. Second, the SVDD detectors work well with both multi-spectral 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

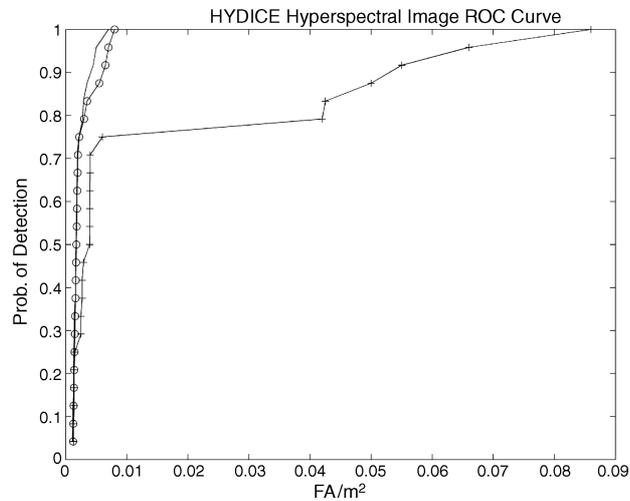


Figure 8.10 ROC curve comparing SVDD and RX detectors on HYDICE hyperspectral imagery. The 1280×320 image contains vehicles that are on roads and are partially camouflaged near tree lines.

appears to be sensitive to the number of bands. Among the reasons for the difference are (a) the Gaussian assumption of the RX may be less valid for higher dimensional data, as explained in Landgrebe (2002), and (b) robust estimation of the covariance matrix is more difficult as the number of bands increases.

Computational complexity

The ROC curves in Figures 8.5–8.10 illustrate the increase in detection performance offered by the SVDD approach. In addition to reducing the false alarm rate, anomaly detectors should also be computationally efficient in order to process a datacube in near real-time. Thus, the computational complexity of the RX and SVDD algorithms are considered.

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 (Schweizer 2001) 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 (FLOPS) for the RX detector is approximately quadratic with respect to the number of bands (Schweizer and Moura 2001).

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 (Lai *et al.* 2004). 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 Figure 8.11, the processing times for both algorithms as function of the number of spectral bands and number of training samples are shown. The algorithms are implemented

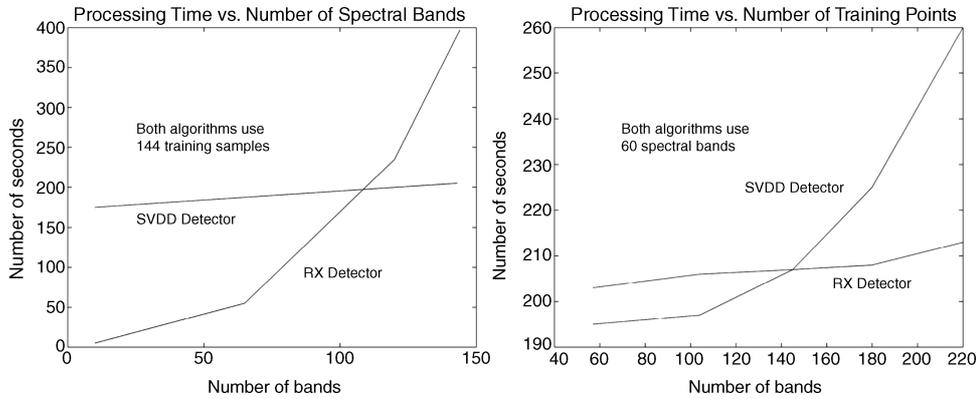


Figure 8.11 Illustration of how the computational complexity of the RX and SVDD anomaly detectors vary with the number of spectral bands and the number of training samples. Note that the speed of the SVDD in this example is based on the use of standard QP optimization.

in C++ on a 2.8 GHz Pentium machine, and the code has not been fully optimized. The CPU time required to process a 256×100 pixel image are measured. Figure 8.11(left) 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 Figure 8.11(right), 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.

Performance evaluation of global anomaly detectors

Finally, the performance of the global SVDD algorithm is compared with the local SVDD and RX anomaly detectors using the WAAMD imagery. Figure 8.12 shows their ROC

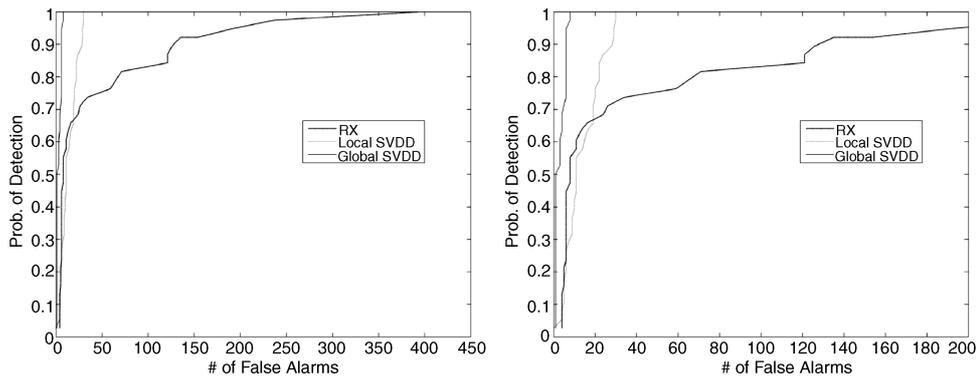


Figure 8.12 (Left) ROC performance for the RX, local SVDD and global SVDD detectors. (Right) Zoomed in view of upper left region of ROC plots providing a clearer view of global SVDD performance.

Table 8.1 Processing complexity and algorithm characteristics affecting detector performance

	Processing complexity	Clutter model/Notes
RX	$\mathcal{O}(N_p N_b^3)$	Local, Unimodal, Normal
SVDD (local)	$\mathcal{O}(N_p N_c^3)$	Local, Multimodal
SVDD (global)	$\mathcal{O}(N_c^3)$	Global, Multimodal

Key: N_p , N_b , N_c : Number of pixels, spectral bands, clutter training samples

performance on the WAAMD hyperspectral image captured at 2000 foot altitude containing 38 targets.

The global SVDD results in enhanced false alarm suppression due to the rejection of candidates that are locally anomalous but not globally anomalous (e.g. isolated trees). In addition, it provides enhanced detection due to enhanced background characterization that is not forced to include target spectra when targets are in close proximity to each other.

The computational complexity of the RX method stems from the inversion of the covariance matrix that is required for the background estimation, which in general is of order $\mathcal{O}(N_b^3)$ where N_b is the number of bands. Owing to the local nature of the estimation, this inversion is required at every pixel, so the actual processing time is of order $\mathcal{O}(N_p N_b^3)$, where N_p is the number of spatial pixels in the hyperspectral image. By using a global background estimate for RX, N_p can be removed from the order of magnitude of the processing, and thereby achieve rapid processing of the image. However, the anomaly detector will not work well since the background is generally not globally unimodal.

The SVDD approach does not require a covariance matrix inverse computation and is also linear with respect to spectral dimensionality. Processing is commensurate with the inversion of the system matrix kernel, which is of order $\mathcal{O}(N_c^3)$. Since SVMs are effective with few exemplars, the background modelling is very robust to low training so N_c may be relatively low. In addition, since the background modelling is multi-modal, automatic and does not require any a priori knowledge of background characteristics (e.g. the number of components to be used in a mixture model), global background characterization may be reliably performed. As a result, SVDD is able to achieve high detection performance and a low false alarm rate with rapid computational processing. Computational complexity and associated characteristics affecting algorithm performance are summarized in Table 8.1.

8.6 Conclusions

A geometric interpretation of the SVDD is used to derive a normalized metric that is appropriate for anomaly detection in spectral imagery. ROC curves comparing 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.

Other algorithms, such as the mixture of Gaussian anomaly detector and the sub-space RX (SSRX) produce similar reductions in the false alarm rate. However, the proposed SVDD approach for anomaly detection has several key differences.

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

References

- Banerjee, A., Burlina, P. and Chellappa, R. (1999) Adaptive target detection in foliage-penetrating SAR images using alpha-stable models. *IEEE Trans. Image Processing*, **8**, 1823–1831.
- Beaven, S.G., Stein, D. and Hoff, L.E. (2000) Comparison of Gaussian mixture and linear mixture models for classification of hyperspectral data. *IEEE Proc. International Geoscience and Remote Sensing Symposium*, 1597–1599.
- Chang, C-C. and Lin, C-J. (2001) LIBSVM : a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Comaniciu D. and Meer, P. (1997) Robust analysis of feature spaces: colour image segmentation. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 750–755.
- Finn, H.M. (1986) A CFAR design for a window spanning two clutter fields. *IEEE Trans. on Aerospace and Electronic Systems*, **22**.
- Fukunaga, K. and Hostetler, L.D. (1975) The estimation of the gradient of a density function. *IEEE Trans. Information Theory*, **21**, 32–40.
- Gandhi, P. and Kassam, S. (1998) Analysis of CFAR processors in non-homogeneous backgrounds. *IEEE Trans. Aerospace and Electronic Systems*, **24**.
- Gualtieri, J.A. and Cromp, R.F. (1998) Support vector machines for hyperspectral remote sensing classification. *27th AIPR Workshop, Proceedings of the SPIE*, **3584**, 221–232.
- Gualtieri, J.A., Chettri, S.R., Cromp, R.F. and Johnson, L.F. (1999) *Support Vector Machines Classifiers as Applied to AVIRIS Data, Summaries of the Eighth JPL Airborne Earth Science Workshop*: **99-17**, 217–227.
- Henze, N. and Wagner, T. (1997) A new approach to the BHEP tests for multivariate normality. *Journal Multivariate Analysis*, **62**, 1–23.
- Kwon, H. and Nasrabadi, N. (2006) Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery. *IEEE Trans. Geoscience Remote Sensing*, **43**.
- Lai, C., Tax, D.M.J., Duin, R.P.W., Pekalska, E. and Paclik, P. (2004) A study on combining image representations for image classification and retrieval. *International Journal of Pattern Recognition and Artificial Intelligence*, **18**, 867–890.
- Landgrebe, D. (2002) Hyperspectral image data analysis. *IEEE Signal Processing Magazine*, **19**, 17–28.
- Lu, X., Hoff, L.E., Reed, I.S., Chen, M. and Stotts, L.B. (1997) Automatic target detection and recognition in multiband imagery: a unified ML detection and estimation approach. *IEEE Trans. Image Processing*, **6**, 143–156.
- Osuna, E., Freund, R. and Girosi, F. (1997) Improved training algorithm for support vector machines. *Proc. IEEE Neural Networks in Signal Processing*, pp. 276–285.
- Pickands, J. (1975) Statistical inference using extreme order statistics, *Annals of Statistics*, Vol. 3, pp. 119–131.
- Scharf, L.L. and Friedlander, B. (1994) Matched subspace detectors. *IEEE Trans. Signal Proc.*, **42**, 2146–2157.

- Scholkopf, B., Platt, J., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. (1999) Estimating the support of a high-dimensional distribution. *Tech Report Microsoft Research*, 99.
- Scholkopf, B., Williamson, R., Smola, A.J., Shawe-Taylor, J. and Platt, J. (2000) Support vector method for novelty detection. *Neural Information Processing Systems*, 582–588.
- Schweizer, S. and Moura, J.M. (2001) Efficient detection in hyperspectral imagery. *IEEE Trans. Image Processing*, **10**, 584–597.
- Stein, D., Beaven, S., Hoff, L.E., Winter, E., Shaum, A. and Stocker, A.D. (2002) Anomaly detection from hyperspectral imagery. *IEEE Signal Processing Magazine*, **19**, 58–69.
- Stocker, A.D. (1999) Stochastic expectation maximization (SEM) algorithm. *Proc. DARPA Adaptive Spectral Reconnaissance Algorithm Workshop*.
- Syed, N., Liu, H. and Sung, K. (1999) Incremental learning with support vector machines. In *Proceedings of the Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence (IJCAI-99)*, Stockholm, Sweden 1999.
- Tax, D.M.J. and Duin, R.P.W. (1999) Data domain description using support vectors. *Proc. European Symposium on Artificial Neural Networks*, 251–256.
- Tax, D.M.J., Ypma, A. and Duin, R.P.W. (1999). Support vector data description applied to machine vibration analysis. *Proc. 5th Annual Conference of the Advanced School for Computing and Imaging*, 398–405.
- Vapnik, V. *Estimation of Dependences Based on Empirical Data*. Springer-Verlag, 1982.
- V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.