FAST HYPERSPECTRAL ANOMALY DETECTION VIA SVDD

Amit Banerjee, Philippe Burlina
Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20723

Reuven Meth
SET Corporation
Greenbelt, MD 20770

ABSTRACT

We present a method for fast anomaly detection in hyperspectral imagery (HSI) based on the Support Vector Data Description (SVDD) algorithm. The SVDD is a single class, non-parametric approach for modeling the support of a distribution. A global SVDD anomaly detector is developed that utilizes the SVDD to model the distribution of the spectra of pixels randomly selected from the entire image. Experiments on Wide Area Airborne Mine Detection (WAAMD) hyperspectral data show improved Receiver Operating Characteristic (ROC) detection performance when compared to the local SVDD detector and other standard anomaly detectors (including RX and GMRF). Furthermore, one-second processing time using desktop computers on several 256×256×145 datacubes is achieved.

Index Terms— Hyperspectral Imaging, Anomaly Detection, Real-Time Hyperspectral Processing.

1. INTRODUCTION

Detecting anomalies in hyperspectral imagery entails the task of locating pixels with spectral signatures that deviate significantly from the background. While they can process images completely in-scene, anomaly detectors usually suffer from a high false-alarm rate due to simplifying assumptions imposed on the background signature distribution [1][2].

Most anomaly detectors use a two-step approach: 1) they characterize the background by estimating a model for the background spectra, and 2) they detect anomalies as pixels that deviate from the background model. There are two broad approaches to determining this support region: parametric or non-parametric.[1][2][4][5][6].

Parametric approaches require the selection, validation, and estimation of the probability density function (PDF) of the background spectra [9][10]. Several examples are based on the well-known Reed-Xiaoli (RX) detector [3], which assumes that the spectra are Gaussian distributed. However, it has been shown that the Gaussian and other unimodal distributions are not a good fit to the data and often lead to many false alarms. Furthermore, estimating a PDF for high dimensional HSI data is difficult, since the need for samples increases exponentially with the number of bands [12].

In this work, we use the SVDD, a support vector approach, to model the support region of the background distribution [1][7][8]. The SVDD is a non-parametric method that has several benefits, including: Non-Gaussian modeling: it can model arbitrarily-shaped and multi-modal distributions; Sparsity: fewer training samples are needed to characterize the background in high-dimensional spaces; Good Generalization: the method avoids overfitting and yields good generalization results [1]. Furthermore, we derive a global SVDD anomaly detector that can rapidly analyze the data cube.

The paper is organized as follows. After deriving the SVDD algorithm, we present a global version of the SVDD anomaly detector in Section 3. Section 4 discusses the computation issues regarding the detectors, and the experimental results are provided in Section 5. Finally, Section 6 concludes the paper.

2. DERIVING THE SVDD

The linear SVDD attempts to find a hypersphere that best describes the region of the feature space in which a set of data points lie [7][8]. For HSI, consider the spectra from a set of pixels in a hyperspectral image, denoted as $T = \{\mathbf{x}_i, i = 1,..,M\}$. We seek the minimum enclosing hypersphere $S = \{\mathbf{x} : \| \mathbf{x} - \mathbf{a} \|^2 < R^2\}$ that contains the training set T. This is a constrained optimization problem stated as:

$$\min(R) \text{ subject to } \mathbf{x_i} \in S, i = 1,..,M.$$

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

$$L(R, \mathbf{a}, \alpha_i) = R^2 - \sum_i \alpha_i \left\{ R^2 - \left(< \mathbf{x}_i, \mathbf{x}_i > -2 < \mathbf{a}, \mathbf{x}_i > + < \mathbf{a}, \mathbf{a} > \right) \right\}.$$

The first term in this equation is the radius which we aim to minimize. The second term constrains the hypersphere to contain each training point $\mathbf{x}_i$. Taking the partial derivatives of L with respect to R and $\mathbf{a}$ and setting them to 0 yields another expression for the Lagrangian function to be maximized with respect to the α_i, :

$$L = \sum_i \alpha_i \left(< \mathbf{x}_i, \mathbf{x}_i > \right) - \sum_{i,j} \alpha_i \alpha_j \left(< \mathbf{x}_i, \mathbf{x}_j > \right).$$

After optimizing L with respect to α_i, it is typical to discover a large fraction of the weights α_i, is equal to zero. The training examples with non-zero weights are called support vectors and lie on the boundary of the support region. Hence, the support vectors must lie on the boundary of the hypersphere. Therefore the SVDD yields a sparse representation of the density function's support expressed entirely in terms of the support vectors.

A hypersphere will not likely give a tight representation of the shape of the background density function in the original input space. To model arbitrarily complex distributions, one may resort to the nonlinear SVDD, which maps the data from the input space to a higher-dimensional feature space through the use of a mapping $\Phi(\mathbf{x})$. This models the support of the distribution as a minimum enclosing hypersphere in the feature space. This hypersphere corresponds to a tighter description for the support of the PDF in the original input space.

Proceeding as in the linear SVDD case, it can be shown, using the well-known "kernel trick," that the SVDD statistic is then expressed as:

$$SVDD(\mathbf{y}) = R^2 - K(\mathbf{y},\mathbf{y}) + 2\sum_i \alpha_i K(\mathbf{y},\mathbf{x}_i),$$

where K is a kernel function defined as $K(\mathbf{x},\mathbf{y}) = <\Phi(\mathbf{x}),\Phi(\mathbf{y})>$. We use the popular Gaussian radial basis function (RBF) for the kernel function, defined as $K(\mathbf{x},\mathbf{y}) = \exp(-\|\mathbf{x}-\mathbf{y}\|^2/\sigma^2)$. The SVDD function then simplifies to

$$SVDD(\mathbf{y}) = C + \sum_i \alpha_i K(\mathbf{y},\mathbf{x}_i),$$

where C is a constant offset, since $K(\mathbf{y},\mathbf{y}) = 1$. The SVDD-based decision rule is then expressed as $SVDD(\mathbf{y}) \geq t$ for some threshold t.

Finally, the RBF scale parameter σ^2 controls how well the SVDD generalizes to unseen data and the tightness of the support boundaries [7][8]. This allows the SVDD to model non-Gaussian, multi-modal distributions. An optimal value for σ^2 can be found using cross-validation or a minimax technique [13][1].

3. SVDD GLOBAL ANOMALY DETECTOR

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, able to achieve high detection performance and low false alarm rates with extremely rapid computational processing. Global background modeling also provides the capability for enhanced ROC performance due to: a) more accurate background characterization obtained in the presence of closely spaced targets which is corrupted when a local model is used, and b) the rejection of local anomalies which are not truly anomalous to the scene (e.g. isolate trees, shadows).

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

- Randomly select a set of N background pixels from the image as the training set
- Given the set of background spectra, estimate an optimal value for σ, the scale parameter of the RBF kernel, using a cross-validation or minimax method [1][13].
- 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.
- For each pixel in the image perform the decision test:
  - If $SVDD(\mathbf{y})$, 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 further reduce the computational expense of the algorithm, we take a closer look at the $SVDD(\mathbf{y})$ function. Expanding $SVDD(\mathbf{y})$ yields:

$$f(\mathbf{y}) = C + c(\mathbf{y})\sum_i \alpha_i C_i \exp\left(\frac{-2\mathbf{y}^{'}\mathbf{x}_i}{\sigma^2}\right),$$

where $C_i = \exp(\frac{-2\mathbf{x}_i^{'}\mathbf{x}_i}{\sigma^2})$ may be pre-computed at the beginning after the 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 computed at every pixel.

4. COMPUTATIONAL ANALYSIS

HSI anomaly detection algorithms can be divided into two primary categories: local background modeling and global background modeling. For local background modeling, the background is re-estimated in a hollow window for each pixel under test. This results in high computational cost due to the necessity of background estimation at each spatial location in the hypercube. For global background modeling, the background is estimated once for the entire hypercube. However, since the global background contains many materials, the modeling required is much more complex than that required locally.

The RX algorithm models the local background in the vicinity of each pixel as a unimodal Gaussian distribution, and has been the standard for HSI anomaly detection in recent years [3]. The computational complexity of this

approach stems from the inversion of the covariance matrix that is required for the background estimation which in general is of order $O(N_b^3)$ where N_b is the number of bands. (We note that we assume direct inversion implementation which is cubic. Efficient inversion methods are available which reduce complexity to $O(N^{2.376})$, but such methods are not often used. Any reduction in inversion complexity would similarly apply to other approaches where an inversion is used so there is no loss of generality in the comparison of SVDD to RX.) Due to the local nature of the estimation, this inversion is required at every pixel, so actual processing is of order $O(N_p N_b^3)$ where N_p is the number of spatial "pixels" in the hyperspectral image. By using a global background estimation for RX, N_p can be removed from the order of magnitude of the processing, however, the anomaly detector will not work well since the background is not globally unimodal. A variation of the RX algorithm, the Gaussian Mixture RX (GMRX) algorithm, may be used to provide a better fit to the background model by employing a mixture of Gaussians to estimate the multi-modal nature of the background. This process has order $O(N_m N_b^3)$ where N_m is the number of Gaussian components in the mixture. However, the number of required components is not known *a priori*, so in practice, the background model is estimated inaccurately resulting in poor detection performance.

The Gaussian Markov Random Field (GMRF) anomaly detector [11] uses characteristics of the GMRF to compute the covariance inverse indirectly resulting in processing which is linear with respect to band dimensionality. Computational complexity is derived in [11] where the dominant processing may be shown to be $O(N_b N_c^2)$ where N_c is the number of clutter pixels. Due to the fact that the background modeling is unimodal, this approach is not well suited for global background characterization.

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 $O(N_c^3)$. Since Support Vector Machines (SVMs) are effective with few exemplars, the background modeling is very robust to low training so N_c may be relatively low. In addition, since the background modeling 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 low false alarm rate with rapid computational processing. Computational complexity and associated characteristics affecting algorithm performance are summarized in Table 1.

	Processing Complexity	Clutter Model / Notes
RX	$O(N_p N_b^3)$	L, U, N
GMRX	$O(N_b^3)$	G, M, N / 1
GMRF	$O(N_b N_c^2)$	L, U, N
SVDD (local)	$O(N_p N_c^3)$	G, M, HDNP / 2
SVDD (global)	$O(N_c^3)$	G,M, HDNP / 2

Key: N_p, N_b, N_c: Number of pixels, spectral bands, clutter training samples
L: Local , G: Global, N: Normal/Gaussian, U: Unimodal, M: Multimodal
HDNP: High dimensional non-parametric
1: Requires mixture dimension *a priori*
2: N_c small (SVM effective with few exemplars)

Table 1: *Processing complexity and algorithm characteristics affecting detector performance.*

5. EXPERIMENTS

We compare the detectors using images taken from the Wide Area Airborne Mine Detection (WAAMD) dataset. The COMPact Airborne Spectral Sensor (COMPASS) sensor, providing 256 bands in the VIS/NIR/SWIR part of the spectrum (400-2350nm), was used to image the minefields. For the images used in this study, the sensor was flown at an altitude of 2000 feet, with a Ground Sample Distance (GSD) of approximately 8 inches. The images contain two types of mines whose sizes are approximately 2x2 pixels.

Global processing as employed by global SVDD results in enhanced false alarm suppression due to the rejection of candidates which 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. Figure 1 shows ROC curves depicting the probability of detection/False Alarm (ROC curve) performance of the RX, GMRF, local SVDD, and global SVDD algorithms on a sample 256×256×145 hypercube containing 38 targets. Due to the large disparity in performance between detectors, we also show a zoomed-in view of the upper left portion of the ROC plots in Figure 2 to provide a clear view of the performance for the global SVDD detector. Table 2 provides processing times for each of the algorithms.

We note that using global modeling for RX and GMRF is not a viable option due to the unimodal clutter characterization of these algorithms that does not enable global modeling of the clutter. For the case shown, the targets were on the order of 2×2 pixels (spatially) which adversely affects the assumptions of the Markov modeling

which in its current implementation, uses a minimum of a 3×3 Markov window.

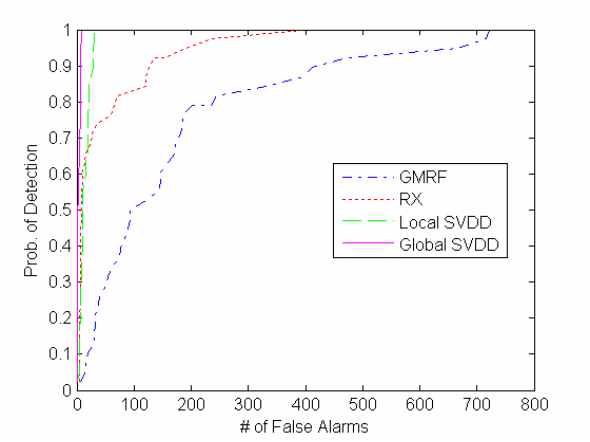

Figure 1: *ROC performance for the GMRF, RX, local SVDD and global SVDD detectors.*

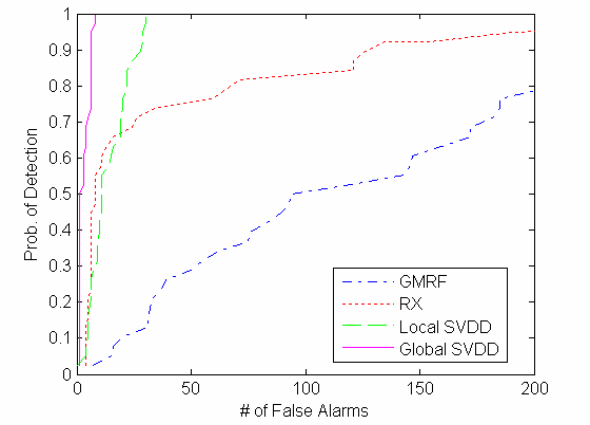

Figure 2: *Zoomed in view of upper left region of ROC plots in* Figure 1 *providing a clearer view of global SVDD performance.*

Algorithm	Runtime (s)
Global SVDD (C++)	1.2
Local SVDD (MATLAB)	43
GMRF (C++)	28
RX (C++)	506

Table 2: *Processing times for global SVDD, local SVDD, GMRF and RX anomaly detectors on a 256×256×145 hypercube.*

6. CONCLUSION

We have successfully developed and demonstrated an HSI anomaly detector that can effectively detect targets more rapidly and robustly than alternative detectors. The detector's strength lies in the SVDD formulation which provides automated, multi-modal, high-dimensional non-parametric modeling for the background clutter that is readily adaptable to global background characterization. Current throughput of one 256×256×145 hypercube per second on a PC platform indicates the viability of real-time anomaly detection using this method.

7. ACKNOWLEDGEMENTS

This work was supported in part by Air Force Research Laboratory contract FA8718-06-C-0061 under the supervision of Lt Michael Ciampa. The authors also thank M. Schatten of the U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate for providing the WAAMD data used in the experiments.

8. REFERENCES

[1] A. Banerjee, P. Burlina, C. Diehl "Support vector methods for anomaly detection in hyperspectral imagery," in IEEE Transactions on Geoscience and Remote Sensing, August 2006.

[2] Banerjee, P. Burlina, R. Chellappa, "Adaptive target detection in foliage-penetrating SAR images using alpha-stable models," in IEEE Trans. Image Proces., vol. 8 , pp 1823 – 1831, Dec., 1999.

[3] 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 Processing, vol. 6, pp. 143-156, January, 1997.

[4] J.A. Gualtieri and R.F. Cromp, "Support vector machines for hyperspectral remote sensing classification," 27th AIPR Workshop, Proc. of the SPIE, Vol. 3584, pp. 221-232, 1998.

[5] P. Gandhi and S. Kassam, "Analysis of CFAR processors in non-homogeneous backgrounds," IEEE Trans. Aerospace and Electronic Systems, vol. 24, 1988.

[6] H.M. Finn, "A CFAR design for a window spanning two clutter fields," IEEE Trans. Aerosp. Electr. Syst., vol. 22, 1986.

[7] D.M.J. Tax and R.P.W. Duin, "Data domain description using support vectors," in Proc. European Symposium on Artificial Neural Networks, M. Verleysen, editor, pp. 251-256, Brussels, Belgium, April, 1999.

[8] D.M.J Tax, A. Ypma, and R.P.W Duin, "Support vector data description applied to machine vibration analysis," In: M. Boasson, J. Kaandorp, J.Tonino, M. Vosselman (eds.),Proc. 5th Annual Conference of the Advanced School for Computing and Imaging (Heijen, NL, June 15-17), 1999, 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 Proc. Magazine, vol. 19, no. 1, pp. 58-69, Jan. 2002.

[10] 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 2000, Honolulu, HI, pp. 1597-1599, July, 2000.

[11] S. Schweizer, and J.M. Moura, "Efficient detection in hyperspectral imagery," IEEE Trans. Image Processing, vol. 10, no. 4, pp. 584-597, April, 2001.

[12] D. Landgrebe, "Hyperspectral image data analysis," IEEE Signal Processing Magazine, vol. 19, no. 1, pp. 17-28, January, 2002.

[13] V. N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.