OPTIMAL KERNEL BANDWIDTH ESTIMATION FOR HYPERSPECTRAL KERNEL-BASED ANOMALY DETECTION

Heesung Kwon*, Senior member, IEEE, Prudhvi Gurram
Army Research Laboratory, ATTN: RDRL-SES-E
2800 Powder Mill Rd., Adelphi, MD 20783

Abstract

A kernel-based anomaly detection technique called Kernel RX [1] developed by one of the authors is being growingly used as a prescreening tool that detects anomalous pixel spectra in hyperspectral images. Targets of interest are then identified among the prescreened anomalous spectra based on reference spectral information using supervised classification/detection techniques [2, 3]. Unlike the regular RX detection [4] that basically assumes that given data can be characterized well by the second-order statistics, the Kernel RX algorithm can exploit higher-order statistics by implicitly mapping the data in a measurement space into a high dimensional, possibly infinite, feature space. Various nonlinear functions associated with given kernel functions are used for nonlinear mapping. Kernel RX then measures the normalized Euclidean distance in the high dimensional feature space, exploiting underlying nonlinear data structures of the given hyperspectral images. The Gaussian radial basis function (RBF) kernel $k(\mathbf{x}, \mathbf{y}) = exp(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{\sigma})$ has been widely used for kernel-based detection methods mainly because the data mapped by the associated mapping function is not spread all over the feature space but generally is confined to compact subspaces.

In applying the Kernel RX defined in Eq. (21) in [1], using an optimal bandwidth σ of the Gaussian RBF kernel is critical to maximize probability of detection (PD) at a given probability of false alarm (PFA). The Gaussian RBF kernel with narrow bandwidths (i.e., small σ values) can generally better detect anomalous objects buried in background clutter noise than the one with large bandwidths, but this comes at the expense of usually very high false alarm rates. On the other hand, the Gaussian RBF kernel with overly wide bandwidths (i.e., large σ values) smooths out subtle spectral differences between the background clutter and anomalies, driving down PD at a given PFA. Therefore, an optimum value of σ has to be estimated to provide optimal detection results given hyperspectral images. In [1], the parameter σ was manually tuned resulting in sub-optimal detection results for the given hyperspectral image data set.

In this paper, we propose a fully automated approach to estimate an optimum σ that maximizes the Kernel-RX cost function (i.e., the Kernel RX output – Eq. (21) in [1]) calculated based on a small portion of the spectra obtained from the given hyperspectral images. Cross validation is used to estimate the cost function based on a small subset of background spectra. No real spectra from targets or anomalous regions were used during the cross validation process because in anomaly detection, prior spectral information about the targets of interest is not available. As a result, only a relatively small number of randomly selected background spectra are used to estimate σ given the hyperspectral image. In randomly sampling the background spectra, it is assumed that anomalies occupy only a fraction of the entire image and hence the majority of the randomly selected samples are from the background. For the purpose of the cross validation, anomalous objects are simulated using only the background spectra by generating heavy-tailed noise samples and adding them to the validation data. A Gaussian scale mixture model [5] was used to generate the heavy-tailed noise. In cross validation, the Kernel RX cost function is repeatedly calculated between a validation set (i.e. the simulated anomalous objects) and a training set (the rest of the background spectra). The σ which gives the best PD vs. PFA performance is selected as the optimal σ.

After estimating the optimal σ^*, to further improve detection performance a full diagonal matrix Σ of the Gaussian Mahalanobis kernel $k(\mathbf{x}, \mathbf{y}) = exp(-(\mathbf{x}-\mathbf{y})^T\Sigma^{-1}(\mathbf{x}-\mathbf{y}))$, where Σ is a diagonal matrix, is also estimated based on the optimal σ^*. The derivation of Σ is based on two assumptions i) the volumes of the Gaussian

RBF kernel and the Gaussian Mahalanobis kernel with the optimal σ^* and Σ, respectively, are the same and ii) Σ is in fact a scaled version of a diagonal matrix V where each diagonal element is the variance of each band of the given hyperspectral image. Thus $\Sigma = qV$, where $q > 0$ and $q \in R$. Since the volumes are the same, the determinants of the diagonal matrix U with all the diagonal elements equal to σ^* and the diagonal matrix Σ being estimated should be the same. The determinants of U and Σ are expressed as

$$\begin{aligned} |U| &= {\sigma^*}^N, \\ |\Sigma| &= |qV| = q^N \prod_i v_i, i = 1 \ldots N, \end{aligned} \tag{1}$$

where v_i are the variances of the spectral bands and N is the number of spectral bands. Since $|U| = |qV|$,

$$\begin{aligned} \sigma^* &= q(\prod_i v_i)^{1/N}, \\ q &= \frac{\sigma^*}{(\prod_i v_i)^{1/N}}. \end{aligned} \tag{2}$$

The scale factor q is easily calculated using Eq. 2, and the full diagonal bandwidth matrix is obtained by $\Sigma = qV$.

Simulation Results

The proposed method was applied to a Hyperspectral Digital Imagery Collection Experiment (HYDICE) image called the Forest Radiance I (FR-I) whose spectral range spans $0.4 - 2.5\mu m$. A HYDICE imaging sensor generates 210 bands across the whole spectral range, but we only use 80 bands by discarding water absorption and low SNR bands that do not contribute to the discrimination of anomalies against the backgrounds. FR-I includes 14 targets located across the lower part of the image. 500 background samples ($\approx 1\%$ of the entire pixels) were randomly selected from FR-I and used for 5-fold cross validation. A kernel matrix (Gram Matrix) was repeatedly generated from the training data (400 samples) and the Kernel-RX cost function was calculated using the remaining validation data (100 samples). The cross validation over the 500 samples yielded the optimal $\sigma^* = 20$. The scale factor q was then calculated using Eq. 2.

Figure 1 shows three detection results by the Kernel RX with Σ, $\sigma^* = 20$, and $\sigma = 1$. All the anomalies (targets of interest) were similarly detected by the three different bandwidths. However, the use of Σ resulted in the best false alarm performance followed by σ^*, and $\sigma = 1$ as quantitatively supported by the ROC curves shown in Figure 2: the red curve (the top one) corresponds to the Kernel-RX performance with Σ and the second one from the top with σ^*.

Index Terms: Kernel bandwidth, non-linear anomaly detection, cross validation, Gaussian RBF function, kernel RX.

References

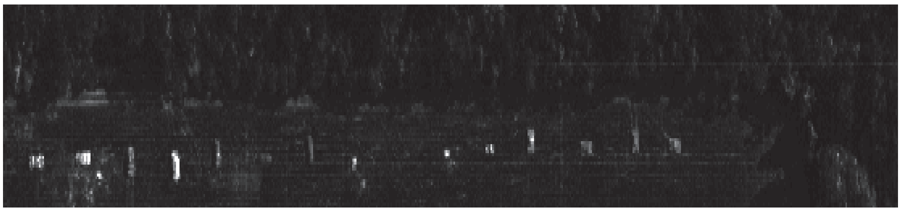
[1] H. Kwon and N. M. Nasrabadi, "Kernel rx-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.

[2] D. Manolakis and G Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

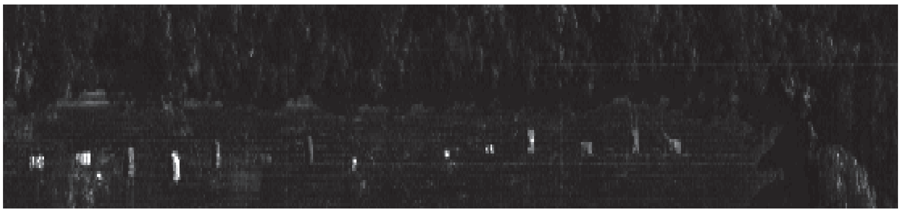
[3] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 2, pp. 178–194, 2006.

[4] I. S. Reed and X. Yu, "Adaptive multiband cfar detection of optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech and Signal Process.*, vol. 38, no. 10, pp. 1760–1770, 1990.

[5] Siwei Lyu and E. P. Simoncelli, "Statistical modeling of images with fields of gaussain scale mixtures," *Adv. Neural Iformation Processing Systems*, vol. 19, 2007.



(a)



(b)

(c)

Figure 1: Anomaly detection results using Kernel RX with (a) $\Sigma = qV$, (b) $\sigma^* = 20$ and (c) $\sigma = 1$. The optimal σ^* estimated for the hyperspectral image was 20.

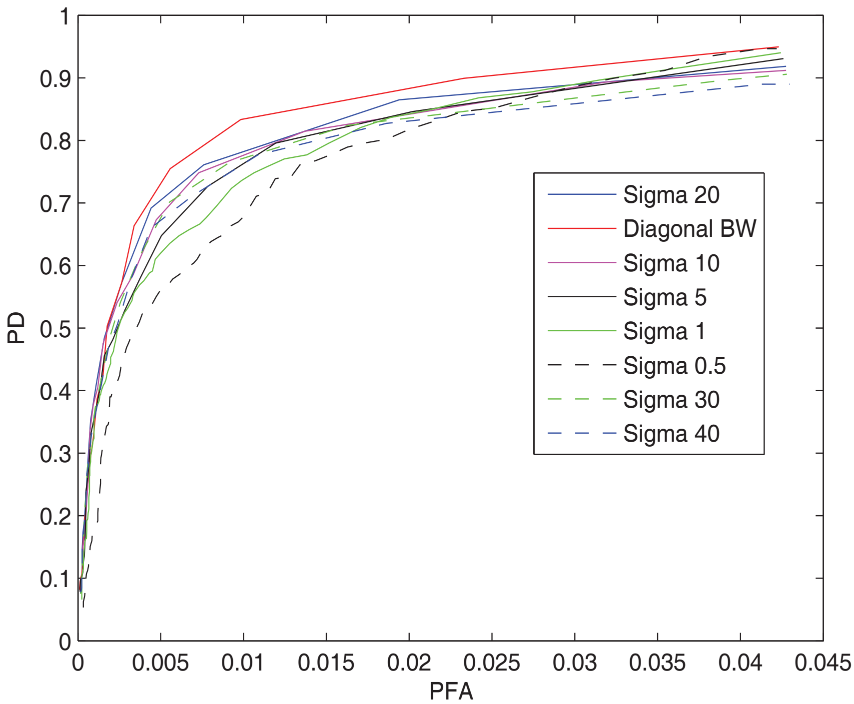


Figure 2: Receiver operating characteristic (ROC) curves of the Kernel RX performance on FR-I with various σ and Σ.