SEGMENTED REGRESSION FOR SPATIO-SPECTRAL BACKGROUND ESTIMATION

James Theiler and Amanda Ziemann

Space Data Science and Systems Group
Los Alamos National Laboratory
Los Alamos, NM 87545

ABSTRACT

We formulate hyperspectral target detection in terms of a local context by modeling the relationship of individual pixels with the annuli of pixels that surround them. A prediction of the center pixel in terms of the annulus pixels provides an estimate of the target-free pixel value, and this estimate can be used as a baseline against which a measurement of that pixel is compared. When the measurement is far from the baseline, that is evidence that the target-free hypothesis is incorrect – and that there is a target at that pixel. The predictor is adaptive to the image, and in this paper, we suggest making it more adaptive by segmenting the image into qualitatively different regions, and learning a new predictor for each region. We learn a new predictor for each segment, and attempt to optimize the segmentation so as to minimize the prediction error, overall. We apply this approach to some well-known hyperspectral datasets, and find that (as expected) the average prediction error is reduced and (less expected) that the "segments" that are discovered are spatially quite scattered, and (in the case with two segments) tend to group image pixels into edges and non-edges.

***Index Terms*—** Hyperspectral, Target detection, Anomaly detection, Background estimation, Machine learning, Regression

1. INTRODUCTION

To detect an anomaly or target at a given pixel location is to distinguish that pixel as inconsistent with some model for the background [1]. One very simple model is a multivariate Gaussian; for this model, the expected target-free value at *every* pixel in the image is given by the mean of that Gaussian. This is the basis of many standard target detection algorithms, such as the adaptive matched filter [2] and the adaptive coherence estimator [3]. Despite the simplicity of this model, it is often surprisingly effective [4]. Extensions of the Gaussian to, for instance, elliptically contoured distributions can lead to more flexible detectors [5–8].

This work was supported by the United States Department of Energy NA-22 project on Hyperspectral Advanced Research and Development for Solid Materials (HARD Solids).

To further extend these *global* models is to make a separate estimate at each pixel based on that pixel's *local* context. For instance, the well-known RX algorithm [9] computes a local mean from a moving window (an annulus, actually, since it doesn't include the pixel itself) centered on the pixel of interest. This kind of local background estimation has been applied to target detection as well, with very promising results [10–14]. A related approach applies image sharpening before applying a global detector [15, 16]; the effect of this filter is similar to the subtraction of a local mean. More recently, it has been suggested that instead of a local mean, a more general model be used to predict the center pixel as a function of the pixels in the annulus that surround it [17–19].

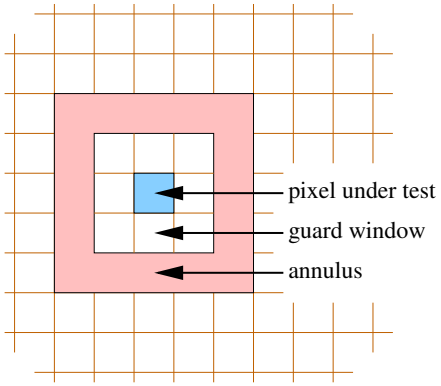

Fig. 1. A given pixel of interest is estimated by the pixels in the annulus that surround the pixel. This estimate can be done for every pixel in the image.

2. LOCAL REGRESSION

Consider a pixel y surrounded by an annulus of pixels $\mathbf{x}$ (Fig. 1). If y is target-free and corresponds to typical background, then we expect to be able to approximate it with $\hat{y} = f(\mathbf{x})$ for some function f that can be learned from the data. In particular, we choose f to minimize some measure of the average error that $\hat{y}$ makes in approximating y. The simplest way to do this (and this simple way is what we do

here) minimizes squared error:

$$\sum_{n=1}^{N} \|y_n - f(\mathbf{x}_n)\|^2, \tag{1}$$

where the samples are taken from the image: $(\mathbf{x}_1, y_1)$, ..., $(\mathbf{x}_N, y_N)$. Ideally, these samples are taken from parts of the image that are free of targets, but generally we tolerate a few target-containing pixels in exchange for having the samples be truly representative of the image statistics. We will call $f(\mathbf{x})$ a "local estimator," and think of it as a generalization of the concept of "local mean" – whereas local mean would be the average of the pixels represented in $\mathbf{x}$, this framework considers more general functions of those pixels.

Although Eq. (1) seems like the obvious thing to optimize, it has been pointed out [20] that there are times "when closer isn't better." Even though this is what was optimized in Ref. [17], that paper noted: "Although mean squared error provides one way to evaluate performance, we ultimately prefer an approach that more directly maps to target detection performance." In more recent work [18, 19], direct measures of performance were employed. But for the current work, in its current stage, we employ squared error while acknowledging its imperfection. (After all, "farther isn't better" either!)

Given the local estimator $f(\mathbf{x})$, we can define a global covariance:

$$R = (1/N) \sum_{n=1}^{N} (y_n - f(\mathbf{x}_n)) (y_n - f(\mathbf{x}_n))^{\mathsf{T}}, \tag{2}$$

and we can use this covariance to define both anomaly and target detectors. For instance, the anomalousness for a pixel y, given an annulus $\mathbf{x}$, will be given by

$$\mathcal{A}(\mathbf{x}, y) = (y - f(\mathbf{x}))^{\mathsf{T}} R^{-1} (y - f(\mathbf{x})). \tag{3}$$

We can also use a matched filter to detect an *additive* target $\mathbf{t}$ in this background; that detector would be given by

$$\mathcal{D}(\mathbf{t}, \mathbf{x}, y) = \mathbf{t}^{\mathsf{T}} R^{-1} (y - f(\mathbf{x})) = \mathbf{q}^{\mathsf{T}} (y - f(\mathbf{x})), \tag{4}$$

where $\mathbf{q} = R^{-1}\mathbf{t}$ is the matched filter coefficient vector. As an aside, we could also consider ACE instead of matched filter:

$$\mathcal{D}_{\text{ACE}}(\mathbf{t}, \mathbf{x}, y) = \frac{\mathbf{t}^{\mathsf{T}} R^{-1} (y - f(\mathbf{x}))}{\sqrt{(y - f(\mathbf{x}))^{\mathsf{T}} R^{-1} (y - f(\mathbf{x}))}}. \tag{5}$$

3. SEGMENTED REGRESSION

It makes sense to tailor $f(\mathbf{x})$ to different parts of the image; the best $f(\mathbf{x})$ for a deep jungle might be different from the best $f(\mathbf{x})$ for an arid desert. Rural scenes might be different from urban scenes, farmland different from forest, *etc*. A natural approach for dealing with this issue is to first segment the scene into these different kinds of areas, and then to fit $f(\mathbf{x})$ separately to each segment. This leaves open the (notoriously ill-defined) question of how to approach the segmentation in the first place.

Algorithm 1 Segmented regression

Require: Pixel values $y_1, \ldots, y_N$
Require: Annulus values $\mathbf{x}_1, \ldots, \mathbf{x}_N$
Require: Initial segmentation (could be random) $k_1, \ldots, k_N$ with $k_i \in \{1, \ldots, K\}$
repeat
  for $k \in 1, \ldots, K$ **do**
    Define $\mathcal{N}_k \leftarrow \{n \,:\, k_n = k\}$
    $f_k \leftarrow \text{argmin}_f \sum_{n \in \mathcal{N}_k} \|y_n - f(\mathbf{x}_n)\|^2$
  for $n \in 1, \ldots, N$ **do**
    $k_n \leftarrow \text{argmin}_k \|y_n - f_k(\mathbf{x}_n)\|$
until convergence
return $f_1, \ldots, f_K$

Our approach is to simultaneously optimize $f(\mathbf{x})$ *and* the segmentation. Imagine that we have segmented the image into K distinct categories, and let k_n be the category label for the n'th pixel. Let $f_k(\mathbf{x})$ correspond to the regression function that is optimized to the k'th segment. What we want to optimize is

$$\sum_n \|y_n - f_{k_n}(\mathbf{x}_n)\|^2. \tag{6}$$

The best choice for cluster label is the label that leads to the best fit:

$$k_n = \text{argmin}_k \, \|y_n - f_k(\mathbf{x}_n)\|^2, \tag{7}$$

and so Eq. (6) can be written

$$\sum_n \min_k \|y_n - f_k(\mathbf{x}_n)\|^2. \tag{8}$$

One "solution" to Eq. (8) is to treat it as a k-means problem. Starting with some initial segmentation of the image k_n, optimize the function f_k by fitting to the points n for which $k_n = k$, and do this for each k, as shown in Algorithm 1. Note that many of the tricks that can be used to break out of local minima in k-means (*e.g.*, splitting and merging schemes) would apply equally well in this context.

It bears remarking that this is *not* just a more complicated function $f(\mathbf{x})$; the choice of which k to use depends not only on $\mathbf{x}$ but on y as well.

This approach solves two problems with the segmented regression idea. One is that it provides a tangible criterion to the segmentation step, which is a perennial matter of contention with segmentation; two is that the optimization of the segmentation is simultaneous with the optimization of the f's.

Table 1. Comparison of RMS values with a single function $f(\mathbf{x})$ and with the segmented $f_1(\mathbf{x}), \ldots, f_K(\mathbf{x})$ with $K = 2$.

Dataset	Bands	Size (Pixels)	$K = 1$ RMS	$K = 2$ RMS
Omaha	8	1384×1243	126.3	96.4
Indian Pines	200	145×145	2106.6	1853.5
Cooke City	126	800×200	1633.1	1414.6
Reno	356	600×320	6268.6	5332.1

4. EXPERIMENTS

In this study, we considered four spectral images: an eight-band multispectral image of Omaha, Nebraska from the WorldView-2 satellite [21]; the venerable hyperspectral Indian Pines dataset [22]; the hyperspectral Cooke City, Montana dataset from the RIT Blind Test study [23]; and a hyperspectral image of Reno, Nevada from SpecTIR [24]. In all three cases, we used a small annulus (5×5 with a 3×3 guard window surrounding the single pixel of interest). Although there are 16 pixels in this annulus, we invoke the dihedral symmetry of the grid (*i.e.*, invariance to reflections and ninety-degree rotations) [18] to reduce the effective number to three. To further simplify the experiment, we take $f(\mathbf{x})$ to be linear in $\mathbf{x}$. In every case, we iterate ten times and observe informally that the solutions appear to be near convergence.

We see in Table 1 that segmented regression more accurately estimates the background (it has lower RMS) for all four of these images. The segmentation that minimizes RMS is seen (in Fig. 2) to be fairly scattered, but with a definite tendency to organize along the edges in an image.

5. FUTURE WORK

The results presented here are preliminary. So far, we have only used $K = 2$ segments. Although extension to more segments is warranted, we see already that the two segments are often qualitatively different, that the two models (f_1 and f_2) are quite distinct, and that the resultant RMS is substantially lower with $K = 2$ instead of $K = 1$.

We note that RMS is a doubly naive measure of performance. For one thing, we measure in-sample fitting error, and it is no surprise that that is smaller when the number of free parameters is larger. It is incumbent upon us, in future work, to demonstrate that the better RMS is statistically significant and that it extends to out-of-sample performance. Also, we know that smaller RMS does not always map directly to better performance [20], so future experiments will measure detection and false alarm rates for target and anomaly detection problems.

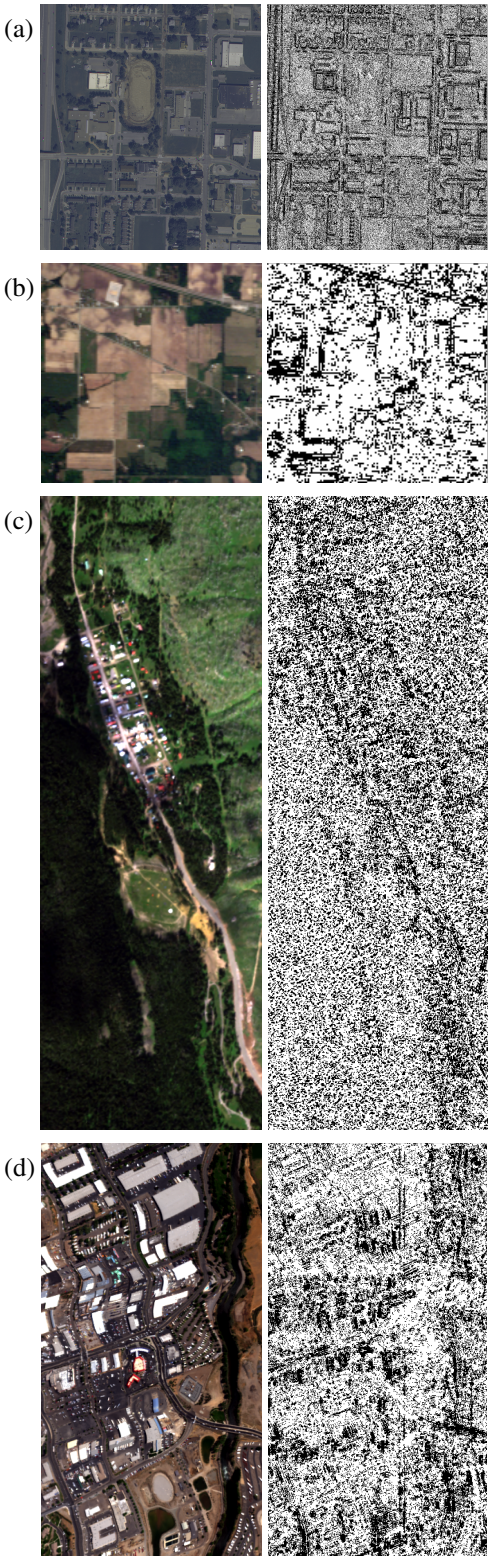

Fig. 2. RGB images (left) and pixel segmentations (right). Two segments ($K = 2$) are shown here, indicated with white pixels and with black pixels. Shown are: (a) Omaha, (b) Indian Pines, (c) Cooke City, and (d) Reno.

6. REFERENCES

[1] S. Matteoli, M. Diani, and J. Theiler, "An overview background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 2317–2336, 2014.

[2] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 10, pp. 853–863, 1974.

[3] L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," *Proc. Asilomar Conference on Signals, Systems, and Computers*, 1996.

[4] B. R. Foy, J. Theiler, and A. M. Fraser, "Unreasonable effectiveness of the adaptive matched filter," *Proc. MSS (Military Sensing Symposia) Passive Sensors Conference*, 2006.

[5] D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE*, vol. 4381, pp. 308–316, 2001.

[6] J. Theiler and B. R. Foy, "EC-GLRT: Detecting weak plumes in non-Gaussian hyperspectral clutter using an elliptically-contoured generalized likelihood ratio test," *Proc. IEEE IGARSS*, p. I:221, 2008.

[7] J. Theiler, C. Scovel, B. Wohlberg, and B. R. Foy, "Elliptically-contoured distributions for anomalous change detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, pp. 271–275, 2010.

[8] G. Grosklos and J. Theiler, "Ellipsoids for anomaly detection in remote sensing imagery," *Proc. SPIE*, vol. 9472, pp. 94720P, 2015.

[9] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1760–1770, 1990.

[10] Y. Cohen and S. R. Rotman, "Spatial-spectral filtering for the detection of point targets in multi- and hyperspectral data," *Proc. SPIE*, vol. 5806, pp. 47–55, 2005.

[11] C. E. Caefer, M. S. Stefanou, E. D. Nelson, A. P. Rizzuto, O. Raviv, and S. R. Rotman, "Analysis of false alarm distributions in the development and evaluation of hyperspectral point target detection algorithms," *Optical Engineering*, vol. 46, pp. 076402, 2007.

[12] C. E. Caefer, J. Silverman, O. Orthal, D. Antonelli, Y. Sharoni, and S. R. Rotman, "Improved covariance matrices for point target detection in hyperspectral data," *Optical Engineering*, vol. 47, pp. 076402, 2008.

[13] Y. Cohen, D. G. Blumberg, and S. R. Rotman, "Subpixel hyperspectral target detection using local spectral and spatial information," *J. Applied Remote Sensing*, vol. 6, pp. 063508, 2012.

[14] Y. Cohen, Y. August, D. G. Blumberg, and S. R. Rotman, "Evaluating sub-pixel target detection algorithms in hyper-spectral imagery," *J. Electrical and Computer Engineering*, vol. 2012, pp. 103286, 2012.

[15] C. C. Borel and R. F. Tuttle, "Improving the detectability of small spectral targets through spatial filtering," *Proc. SPIE*, vol. 7812, pp. 78120K, 2010.

[16] C. C. Borel, "Methods to find sub-pixel targets in hyperspectral data," *Proc. 3rd IEEE WHISPERS*, 2011.

[17] J. Theiler and B. Wohlberg, "Regression framework for background estimation in remote sensing imagery," *Proc. 5th IEEE WHISPERS*, 2013.

[18] J. Theiler, "Symmetrized regression for hyperspectral background estimation," *Proc. SPIE*, vol. 9472, pp. 94721G, 2015.

[19] J. Theiler and A. K. Ziemann, "Right spectrum in the wrong place: a framework for local hyperspectral anomaly detection," *Proc. Computational Imaging XIV*, 2016.

[20] N. Hasson, S. Asulin, S. R. Rotman, and D. Blumberg, "Evaluating backgrounds for subpixel target detection: when closer isn't better," *Proc. SPIE*, vol. 9472, pp. 94720R, 2015.

[21] N. T. Anderson and G. B. Marchisio, "Worldview-2 and the evolution of the DigitalGlobe remote sensing satellite constellation: introductory paper for the special session on WorldView-2," *Proc. SPIE*, vol. 8390, 2012.

[22] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons, 2003.

[23] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, pp. 915–918, 2008.

[24] SpecTIR Advanced Hyperspectral and Geospatial Solutions, "Free data samples," Online: `http://www.spectir.com/free-data-samples/`.