

Copyright 2013 IEEE. To appear in the 1st IEEE Global Conference on Signal and Information Processing. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works, must be obtained from the IEEE. Contact: Manager, Copyrights and Permissions / IEEE Service Center / 445 Hoes Lane / P.O. Box 1331 / Piscataway, NJ 08855-1331, USA. Telephone: + Intl. 908-562-3966.

Spatio-Spectral Anomalous Change Detection in Hyperspectral Imagery

James Theiler

Los Alamos National Laboratory

Los Alamos, NM, USA 87545

jt@lanl.gov

Abstract—Because each pixel of a hyperspectral image contains so much information, many (successful) algorithms treat those pixels as independent samples, despite the evident spatial structure in the imagery. One way to exploit this structure is to incorporate spatial processing into pixel-wise anomalous change detection algorithms. But if this is done in the most straightforward way, a contaminated cross-covariance is produced. A spatial processing framework is proposed that avoids this contamination and enhances the performance of anomalous change detection algorithms in hyperspectral imagery.

Index Terms—hyperspectral, change detection, spatial filter, stacked filter, cross-covariance

I. INTRODUCTION

From two images of the same scene, we want to identify the interesting changes that occurred [1], [2], [3]. Because it is difficult to give a mathematical definition for *interesting*¹ the emphasis is on *anomalous* change. The anomalous changes are rare and unusual, and not like the *pervasive differences* which occur throughout the scene. These pervasive differences may be due to calibration, illumination, look angle, and even the choice of remote sensing platform. They can be caused by misregistration [4], [5], [6] of the images, or by diurnal and seasonal variations [7] in the scene. Because these differences *are* pervasive, their effects can be statistically characterized, just from the image pair. By contrast the anomalous changes are assumed to be relatively rare, and occur in only a small part of the image or image archive. Because the nature of the change is not known beforehand, algorithms for anomalous change detection are unsupervised.

When the images are hyperspectral, then *pixelwise* change detection algorithms are attractive. Because the pixels in a hyperspectral image can have hundreds of spectral channels (providing color information far beyond the usual red, green, and blue components that define a typical photographic image), it is possible to treat these pixels as independent samples and still obtain useful results. The current study is motivated by the conviction that adjacent pixels are *not* independent, and that the obvious spatial structure in real images can be exploited to produce better results. Indeed, it was observed in [8] that simple spatial smoothing of the images could improve change detection performance. The aim here is to provide a more principled framework for pursuing that observation.

¹For example, consider defining n to be the smallest uninteresting number. Now try to tell me that this n is not very interesting!

II. PIXELWISE CHANGE DETECTION

Given two co-registered images, let $\mathbf{x} \in \mathbb{R}^{d_x}$ be the spectrum (*e.g.*, reflectance or radiance at d_x different wavelengths) of a pixel in one image and $\mathbf{y} \in \mathbb{R}^{d_y}$ be that of the corresponding pixel in the other image. Here, d_x (resp. d_y) is the number of spectral channels in the \mathbf{x} (resp. \mathbf{y}) image. For pixelwise ACD algorithms, anomalousness depends only on the spectra at the individual pixels: $\mathcal{A}(\mathbf{x}, \mathbf{y})$. One computes anomalousness \mathcal{A} for every (\mathbf{x}, \mathbf{y}) in the co-registered image pair; the largest values are candidates for anomalous changes.

Perhaps the most straightforward change detection approach is to consider the difference image [9] $\mathbf{e} = \mathbf{y} - \mathbf{x}$; where there are anomalous changes, \mathbf{e} will be anomalously large. This approach is particularly vulnerable to pervasive differences, and does not even make sense when the two images are of different modalities (it certainly requires $d_x = d_y$). A more flexible approach, called the chronochrome [10], performs a linear transformation one of the images in order to minimize the least squares difference between them: that is, $\mathbf{e} = \mathbf{y} - L\mathbf{x}$, where L is chosen to minimize the average $\langle \mathbf{e}^T \mathbf{e} \rangle$ over the image pair. Instead of a linear regression, one might perform a nonlinear fit, for instance using a neural network [11]: $\mathbf{e} = \mathbf{y} - \mathcal{L}(\mathbf{x})$. Instead of minimizing simple least squares, one can minimize total least squares [12], which leads to a family of algorithms that are mathematically similar to “multivariate alteration detection” [13] and “covariance equalization” [14].

A distribution-based approach for ACD was suggested in [8], and further developed in [15], [16]. Here, we write $p(\mathbf{x}, \mathbf{y})$ as a probability density of the “non-anomalous” data (this is something that can be inferred from the imagery), and $p_a(\mathbf{x}, \mathbf{y})$ as a probability density of the “anomalously changed” pixel pairs. Whatever model we propose for p_a , we can write the optimal anomalous change detector using the likelihood ratio:

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \frac{p_a(\mathbf{x}, \mathbf{y})}{p(\mathbf{x}, \mathbf{y})}. \quad (1)$$

As suggested in [8], since we are interested in anomalous *changes*, not just anomalies, we can write $p_a(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$, where the $p(\mathbf{x})$ and $p(\mathbf{y})$ correspond to the separate distributions of \mathbf{x} and \mathbf{y} (which, again, can be inferred from those images separately). This leads to a mutual-information based change detector [15]:

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \log p(\mathbf{x}) + \log p(\mathbf{y}) - \log p(\mathbf{x}, \mathbf{y}). \quad (2)$$

A. Hyperbolic Anomalous Change Detection (HACD)

When the data are Gaussian, then we can express the distributions in terms of covariance matrices.² Write $X = \langle \mathbf{x}\mathbf{x}^T \rangle$ and $Y = \langle \mathbf{y}\mathbf{y}^T \rangle$ as the covariance matrices of the individual images (assuming means have been subtracted from the \mathbf{x} and \mathbf{y} images), and $C = \langle \mathbf{y}\mathbf{x}^T \rangle$ as the cross-covariance between the images. Then (2) leads to the hyperbolic anomalous change detector (HACD):

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T \end{bmatrix} Q \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad (3)$$

with

$$Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}^{-1}. \quad (4)$$

A useful way to re-interpret the components of the HACD algorithm is

$$Q = R_o^{-1} - R_1^{-1}, \quad (5)$$

where

$$R_i = \begin{bmatrix} X & C_i^T \\ C_i & Y \end{bmatrix}, \quad (6)$$

with $C_o = \langle \mathbf{y}\mathbf{x}^T \rangle$ and $C_1 = 0$. Here C_o is the cross-covariance associated with normal pixels and C_1 corresponds to an anomalous pixel. For normal pixels in co-registered images, \mathbf{y} and \mathbf{x} are correlated (they are, after all, measurements of the same position on the ground). But if there is an anomalous change at some location, then that correlation is expected to vanish.

III. SPATIALLY ENHANCED ACD

The pixelwise algorithms in the previous section ignore the evident spatial structure in hyperspectral images. There are a number of approaches for exploiting that structure (*e.g.*, [20] employs a Markov Random Field); the aim here is to leverage the pixelwise algorithms that have already been developed.

Given images a and b , consider applying a spatial filter to each image (needn't be the same spatial filter for both images), resulting in new images Fa and Fb . The purely spectral (non-spatial) way of applying change detection at pixel location (i, j) is to set $\mathbf{y} = a_{ij}$ and $\mathbf{x} = b_{ij}$ and apply the function $\mathcal{A}(\mathbf{x}, \mathbf{y})$. For spatially enhanced ACD, we'd like to use $\mathbf{y} = (Fa)_{ij}$ and $\mathbf{x} = (Fb)_{ij}$ and again simply apply the function $\mathcal{A}(\mathbf{x}, \mathbf{y})$. No muss. No fuss.

Early trials with this approach [8] showed that it *could* improve change detection ability, but more extensive experiments [21] found cases where even generic versions of this approach led to problematic results.

To clarify this situation, we will first consider a restricted class of spatial filters that depend on the spatial neighborhood of a pixel of interest, but do not depend on the pixel itself. If a is an image, and Sa is the filtered image, then we can

²In fact, a slightly larger class of distributions than Gaussian can be expressed in terms of covariance matrices. These are elliptically-contoured distributions; they have been advocated as better models of hyperspectral data [17], [18], [19] and have been observed to exhibit better anomalous change detection performance [16].

TABLE I
SCHEMES FOR SPATIAL PROCESSING

Scheme	Assignments
Standard spectral	$\mathbf{y} = a, \quad \mathbf{x} = b$
Low-pass (smoothing) filter	$\mathbf{y} = a + Sa, \quad \mathbf{x} = b + Sb$
High-pass (sharpening) filter	$\mathbf{y} = a - Sa, \quad \mathbf{x} = b - Sb$
Stacked filter	$\mathbf{y} = [a; Sa], \quad \mathbf{x} = [b; Sb]$
Proposed scheme	$\mathbf{y} = a, \quad \mathbf{x} = [b; Sb; Sa]$
Control: single image	$\mathbf{y} = a, \quad \mathbf{x} = Sa$

TABLE II
INDUCED CROSS-COVARIANCE

Scheme	$C = \langle \mathbf{y}\mathbf{x}^T \rangle$ when a is anomalous
Standard spectral	0
Low-pass (smoothing) filter	$(Sa)(b + Sb)^T$
High-pass (sharpening) filter	$(Sa)(b - Sb)^T$
Stacked filter	$[0 \ 0; (Sa)b^T \ (Sa)(Sb)^T]$
Proposed scheme	$[0; 0; 0]$
Control: single image	0

express the value of Sa at the (i, j) pixel as a function of pixels values in a :

$$(Sa)_{ij} = f(\{a_{k\ell} \mid (k, \ell) \in \mathcal{N}_{ij}\}), \quad (7)$$

with \mathcal{N}_{ij} an “exclusive neighborhood” of (i, j) – it includes pixels near (i, j) but does not include (i, j) itself. Although f can be any function, the experiments in this paper will consider the simple case where f is the average of its arguments. This convolution with an annulus-shaped kernel is by itself something of an unusual filter, but we observe that many filters of interest can be expressed in the form $Fa = a + \beta Sa$, with $\beta > 0$ a low-pass filter (a smoothing filter) and $\beta < 0$ a high-pass or sharpening filter (which has been shown to be effective in some target detection contexts [22]; see also [23]). We can furthermore consider more generic combinations of a and Sa by making a *stacked* filter: $Fa = [a; Sa]$. This notation corresponds to an image Fa with twice as many spectral channels as the image a ; the first set corresponding to a itself and the second set to Sa . These schemes are listed in Table I, along with one further suggestion, which constitutes the main innovation of this paper. The proposed scheme takes advantage of the fact that when a pixel a_{ij} is anomalous, it is uncorrelated with $(Sa)_{ij}$ as well as b_{ij} and $(Sb)_{ij}$. Thus, the partition of the data so that $\mathbf{y} = a$ and $\mathbf{x} = [b; Sb; Sa]$ means that \mathbf{y} and \mathbf{x} are uncorrelated when a is anomalous.

A. Induced cross-covariance

In [21], it was observed that spatial processing can induce cross-covariance even for anomalous pixels, leading to $C_1 \neq 0$. As an example, consider the stacked spatial scheme, with $\mathbf{y} = [a; Sa]$ and $\mathbf{x} = [b; Sb]$. We see that

$$C = \langle \mathbf{y}\mathbf{x}^T \rangle = \langle [ab^T \ a(Sb)^T; (Sa)b^T \ (Sa)(Sb)^T] \rangle. \quad (8)$$

For C_1 , we consider the case that a is an anomalous pixel, and in that case we expect $\langle ab^T \rangle = 0$ and $\langle a(Sb)^T \rangle = 0$, because the anomalous pixel in a is uncorrelated both with

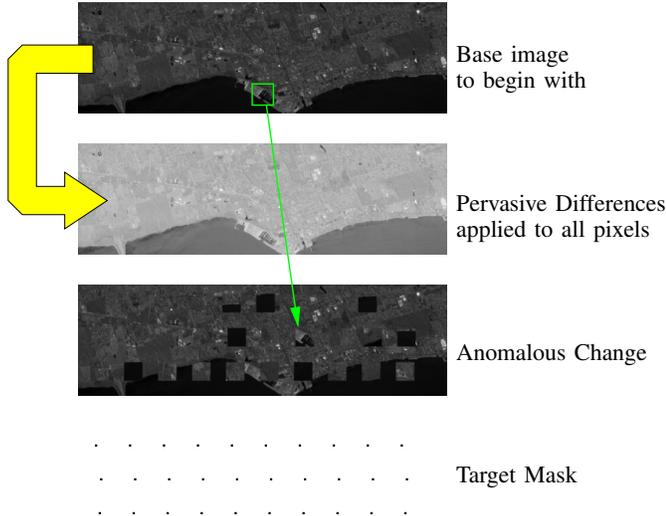


Fig. 1. Simulation framework for spatio-spectral ACD. The top two images correspond to pervasive differences; these are the two images from which $\mathcal{A}(\mathbf{x}, \mathbf{y})$ is trained, and they provide the false alarm versus threshold function. Applying $\mathcal{A}(\mathbf{x}, \mathbf{y})$ to the second and third images, and paying attention only to the pixels identified in the target mask, provides the detection rate versus threshold. Combining the two gives detection rate versus false alarm curves that are shown in Fig. 2 and Fig. 3.

the corresponding pixel b and with its neighborhood S_b . But even if a is anomalous, that doesn't mean its neighborhood S_a is anomalous, so we do not expect either $\langle (Sa)b^T \rangle$ or $\langle (Sa)(Sb)^T \rangle$ to be zero. So for the stacked spatial scheme, we have a nonzero induced cross-covariance $C_1 \neq 0$.

Because the pixel-wise $\mathcal{A}(\mathbf{x}, \mathbf{y})$ in (3) effectively *assumes* $C_1 = 0$, it is not well matched to the nonzero C_1 case. To the extent that the induced cross-covariance is limited, the anomalous change detector may still be effective, but it is not optimal. One solution is to estimate the nonzero C_1 and to modify the HACD algorithm accordingly [21], but a simpler alternative is to employ schemes for which the induced cross-covariance is zero.

Table II lists the induced cross-covariance for each of the schemes in Table I, treating the a pixel as anomalous. The one spatial scheme that induces no cross-covariance is the proposed scheme, with $\mathbf{y} = a$ and $\mathbf{x} = [b; Sb; Sa]$.

IV. EXPERIMENTS

To compare the performance of the different schemes, some experiments were performed using hyperspectral data from the AVIRIS sensor [24]; the experiments use a 150×500 pixel clip of data from flightline f960323t01p02_r04_sc01 [25] taken near Titusville, FL.

The framework for testing ACD algorithms is illustrated in Fig. 1. The first step is to create a second image by imposing a ‘‘pervasive difference’’ everywhere in the image. This simulates the normal and uninteresting changes that occur when two images of the same scene are taken under different conditions. For the experiments here, two kinds of pervasive difference image pairs were created. In the ‘Misreg’ case, the image is smoothed with a 3×3 kernel and misregistered by one pixel.

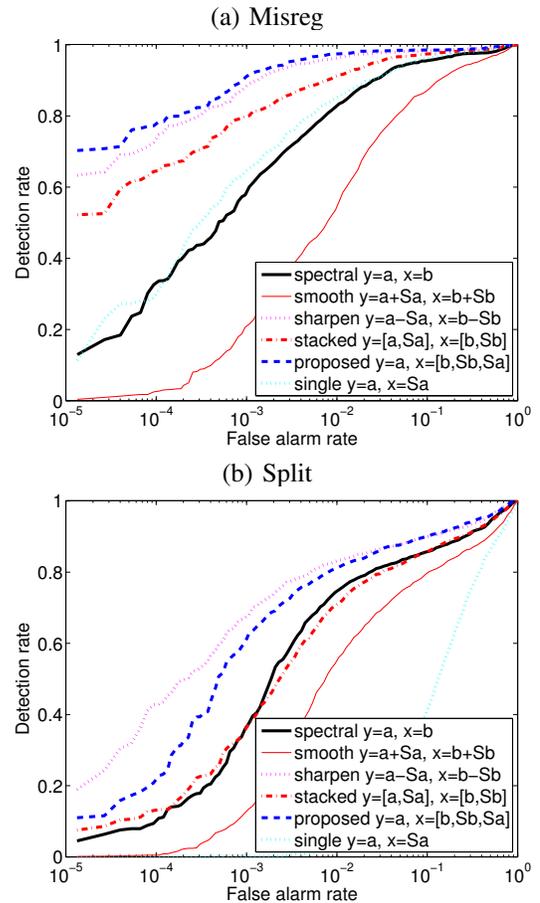


Fig. 2. ROC curves compare change detection performance for AVIRIS data using simulated subpixel anomalous changes and simulated pervasive differences. In (a), the image is smoothed and misregistered by one pixel; in (b), the pervasive difference images are taken by splitting the channels.

For the ‘Split’ case, the 224-channel hyperspectral image is split into two images, the first contains channels 1 to 112, and the second contains 113 to 224. In both cases, the resulting image pairs are subsequently reduced to 10 channels each, using canonical correlation analysis. (The utility of this last step, and details of its implementation, are described in [15].)

The second step in the framework is to create anomalous changes. Where the pervasive differences are applied to the whole image, the anomalous change effectively occurs at only one point on the image. For single pixel anomalies, we replace a random pixel with a random draw from some distribution. Because we want to produce anomalous *changes*, without being distracted by individual *anomalies*, the random draw is from the image itself. That is: a given pixel is replaced by a pixel from somewhere else in the image. For anomalies that are larger than a pixel, whole patches are moved. For subpixel changes, the given pixel is replaced with a linear combination of itself and the other randomly chosen pixel. For the experiments performed here, very small changes (0.25 pixels) were employed so that the performance difference between the different schemes would be more apparent. Although the anomalous changes are presumed to be rare, we can

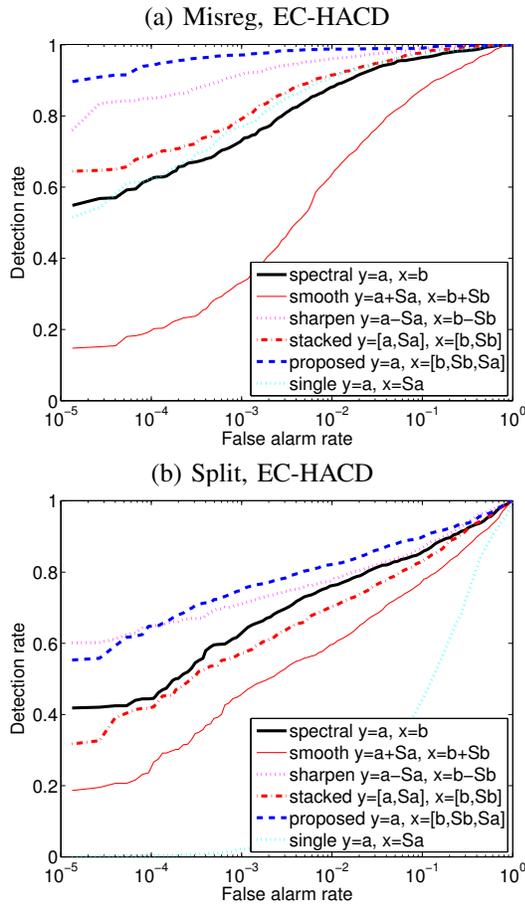


Fig. 3. Same as Fig. 2, but using the elliptically-contoured HACD [16], instead of the Gaussian HACD in (3).

simulate many of them in one trial by separating the individual anomalies so that they do not interfere with each other during spatial processing.

Receiver Operator Characteristic (ROC) curves are shown in Fig. 2. For our small anomalies, we observe that sharpening is better than smoothing; this general observation was noted earlier [21], where it was also observed that larger anomalies preferred smoothing to sharpening. The proposed scheme is better or competitive with the sharpening filter, even though there is no explicit sharpening. Like the proposed filter, the stacked filter is also adaptive, but it does not perform as well, presumably because of its nonzero induced cross-covariance.

The ‘single’ curve provides a control experiment; it attempts to detect the ‘change’ using only a single image. With $y = a$ and $x = Sa$, it infers whether a pixel is an anomalous change using only the local spatial context.

Comparing Fig. 3 to Fig. 2, we observe that the conclusions associated with the HACD algorithm apply also for the elliptically-contoured version (EC-HACD [16]), with the EC algorithm generally outperforming the Gaussian HACD.

ACKNOWLEDGEMENT

This work was supported by the Los Alamos Laboratory Directed Research and Development (LDRD) program.

REFERENCES

- [1] A. Schaum and E. Allman, ‘‘Advanced algorithms for autonomous hyperspectral change detection,’’ *IEEE Applied Imagery Pattern Recognition (AIPR) Workshop*, vol. 33, pp. 33–38, 2005.
- [2] M. T. Eismann, J. Meola, A. D. Stocker, S. G. Beaven, and A. P. Schaum, ‘‘Airborne hyperspectral detection of small changes,’’ *Applied Optics*, vol. 47, pp. F27–F45, 2008.
- [3] M. T. Eismann, J. Meola, and A. D. Stocker, ‘‘Automated hyperspectral target detection and change detection from an airborne platform: Progress and challenges,’’ *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4354–4357, 2010.
- [4] J. Meola and M. T. Eismann, ‘‘Image misregistration effects on hyperspectral change detection,’’ *Proc. SPIE*, vol. 6966, p. 69660Y, 2008.
- [5] J. Theiler, ‘‘Sensitivity of anomalous change detection to small misregistration errors,’’ *Proc. SPIE*, vol. 6966, p. 69660X, 2008.
- [6] J. Theiler and B. Wohlberg, ‘‘Local co-registration adjustment for anomalous change detection,’’ *IEEE Trans. Geoscience and Remote Sensing*, vol. 50, pp. 3107–3116, 2012.
- [7] M. T. Eismann, J. Meola, and R. Hardie, ‘‘Hyperspectral change detection in the presence of diurnal and seasonal variations,’’ *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, pp. 237–249, 2008.
- [8] J. Theiler and S. Perkins, ‘‘Proposed framework for anomalous change detection,’’ *ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection*, pp. 7–14, 2006.
- [9] L. Bruzzone and D. F. Prieto, ‘‘Automatic analysis of the difference image for unsupervised change detection,’’ *IEEE Trans. Geoscience and Remote Sensing*, vol. 38, pp. 1171–1182, 2000.
- [10] A. Schaum and A. Stocker, ‘‘Long-interval chronochrome target detection,’’ *Proc. International Symposium on Spectral Sensing Research (ISSSR)*, 1998.
- [11] C. Clifton, ‘‘Change detection in overhead imagery using neural networks,’’ *Applied Intelligence*, vol. 18, pp. 215–234, 2003.
- [12] J. Theiler and A. Matsekh, ‘‘Total least squares for anomalous change detection,’’ *Proc. SPIE*, vol. 7695, p. 76951H, 2010.
- [13] A. A. Nielsen, K. Conradsen, and J. J. Simpson, ‘‘Multivariate alteration detection (MAD) and MAF post-processing in multispectral bi-temporal image data: new approaches to change detection studies,’’ *Remote Sensing of Environment*, vol. 64, pp. 1–19, 1998.
- [14] A. Schaum and A. Stocker, ‘‘Hyperspectral change detection and supervised matched filtering based on covariance equalization,’’ *Proc. SPIE*, vol. 5425, pp. 77–90, 2004.
- [15] J. Theiler, ‘‘Quantitative comparison of quadratic covariance-based anomalous change detectors,’’ *Appl. Optics*, vol. 47, pp. F12–F26, 2008.
- [16] 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 Lett.*, vol. 7, pp. 271–275, 2010.
- [17] D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, ‘‘On the statistics of hyperspectral imaging data,’’ *Proc. SPIE*, vol. 4381, pp. 308–316, 2001.
- [18] D. B. Marden and D. Manolakis, ‘‘Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data,’’ *Proc. SPIE*, vol. 5425, pp. 558–572, 2004.
- [19] 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 International Geoscience and Remote Sensing Symposium (IGARSS)*, p. I:221, 2008.
- [20] T. Kasetkasem and P. K. Varshney, ‘‘An image change detection algorithm based on Markov random field models,’’ *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, pp. 1815–1823, 2002.
- [21] J. Theiler, N. R. Harvey, R. Porter, and B. Wohlberg, ‘‘Simulation framework for spatio-spectral anomalous change detection,’’ *Proc. SPIE*, vol. 7334, p. 73340P, 2009.
- [22] C. C. Borel and R. F. Tuttle, ‘‘Improving the detectability of small spectral targets through spatial filtering,’’ *Proc. SPIE*, vol. 7812, p. 78120K, 2010.
- [23] 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.
- [24] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, ‘‘The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS),’’ *Remote Sensing of Environment*, vol. 44, pp. 127–143, 1993.
- [25] AVIRIS Free Standard Data Products, Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA). <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>.