

# Subpixel Anomalous Change Detection in Remote Sensing Imagery

James Theiler

Space and Remote Sensing Sciences

Los Alamos National Laboratory

Los Alamos, NM 87545

Email: jt@lanl.gov

**Abstract**—A machine-learning framework for anomalous change detection is extended to the situation in which the anomalous change is smaller than a pixel. Although the existing framework can be applied to (and does have power against) the subpixel case, it is possible to optimize that framework for the subpixel case when the size of the anomalous change is known. The limit of infinitesimally small anomaly turns out to be well-defined, and provides a new parameter-free anomalous change detector which is effective over a range of subpixel anomalies, and continues to have reasonable power against the full-pixel case.

## I. INTRODUCTION

Change detection in imagery is quite useful generally [1], but it has particular value in the remote sensing context. The aim of change detection is to find pixels in pairs of co-registered images that correspond to real changes on the ground. Differences that are due to variations in the environment (illumination, atmospheric distortion, *etc.*) or the sensor (focus, calibration, *etc.*) are generally of less interest. These less interesting differences are often *pervasive*, with the effect visible over the whole image. The more interesting changes, on the other hand, are often *anomalous*, and involve only a few pixels in the image. Refs. [2], [3] have argued that the interesting changes are the anomalous changes, and Ref. [4] proposed a framework that built on the machine learning formalism for anomaly detection, but recast the problem in terms of binary classification: pervasive differences versus anomalous changes. This paper will take that same point of view, but will consider the more extreme case that the anomalous changes are smaller than a pixel.

In Section II, the anomalous change detection framework will be described, and in Section III, that framework will be extended to the subpixel case. Section IV will describe the results of some experiments that compare full-pixel and subpixel detectors to each other and to a standard linear detector.

## II. ANOMALOUS CHANGE DETECTION

### A. Notation

Given two images of the same scene, let  $\mathbf{x}_i \in \mathcal{R}^{d_x}$  correspond to the spectra at the  $i$ th pixel of the first image (the “x-image”), and  $\mathbf{y}_i \in \mathcal{R}^{d_y}$  be the spectrum at the corresponding pixel in the second image (the “y-image”). Here,  $d_x$  is the

number of spectral channels in the image: for example,  $d_x = 1$  for panchromatic imagery,  $d_x = 3$  for RGB imagery, and  $d_x$  can be a hundred or more for hyperspectral imagery. Some change detection algorithms require that  $d_x = d_y$  (*i.e.*, that both images have the same number of spectral channels), but none of the methods described here have that requirement.

The anomalous change detection problem asks: for what pixels  $i$  is the change from  $\mathbf{x}_i$  to  $\mathbf{y}_i$  most atypical, compared to the bulk of the changes from  $\mathbf{x}$  to  $\mathbf{y}$  that occur over an image or image archive?

Following the usual machine-learning paradigm, we treat pixels as independent data samples,<sup>1</sup> drawn from a parent distribution whose density is given by  $P(\mathbf{x}, \mathbf{y})$ . The actual distribution  $P(\mathbf{x}, \mathbf{y})$  is not known, but inferences about it are made from the observed data (the vast bulk of which is assumed to exhibit only the pervasive differences between the two images).

### B. Anomaly detection as binary classification

The notion of anomaly detection as binary classification has been described previously [7], [8], [9], [10]. It is a useful paradigm, and that formalism has been extended to the problem of anomalous change detection [4], [11].

Write  $P^*(\mathbf{x}, \mathbf{y})$  to indicate the distribution of anomalous changes. If this distribution were known, *and* if the pervasive distribution  $P(\mathbf{x}, \mathbf{y})$  were known, then the Bayes-optimal distinction between pervasive differences and anomalous changes would be given by the likelihood ratio

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{P(\mathbf{x}, \mathbf{y})}{P^*(\mathbf{x}, \mathbf{y})}. \quad (1)$$

The various anomalous change detection algorithms that will be described here amount to different models for  $P^*(\mathbf{x}, \mathbf{y})$ .

This approach may initially seem counter-intuitive: it is the nature of anomalies that we do not know what they are; so how can we be expected to write an explicit model for them? But this is a probabilistic model, and it is the nature of probability that it can express our ignorance.

<sup>1</sup>While the use of spatial information, for instance via Markov random field models [5], [6], can improve change detection performance, the approach here will be to concentrate on the spectral information in the pixels. For small (especially subpixel) anomalous changes, the utility of spatial information is somewhat limited, and in any case, the presumption will be made that schemes to exploit spatial information could be applied to any of the spectral methods that are compared in this paper.

1) *Straight anomaly detection*: What might be considered “total ignorance” is the notion that anomalies are distributed uniformly over a region whose support encompasses the support of the distribution  $P(\mathbf{x}, \mathbf{y})$ . In fact, this is the usual assumption for anomaly detection [7], [8], [9], [10], and leads to the problem of density level detection. Here,  $P^*(\mathbf{x}, \mathbf{y}) = U(\mathbf{x}, \mathbf{y})$ , where  $U(\mathbf{x}, \mathbf{y}) = \text{constant}$  is a uniform distribution. And then  $\mathcal{L}(\mathbf{x}, \mathbf{y})$ , as given in Eq. (1), will have contours that correspond to the contours of  $P(\mathbf{x}, \mathbf{y})$ .

The problem with straight anomaly detection is that it does not provide a sense of “change.” A pixel that is unusual in one image, and similarly unusual in the second image, does not particularly indicate an unusual change, but would be out on the tail of the  $P(\mathbf{x}, \mathbf{y})$  distribution, and would be flagged as an anomaly.

2) *Generalized chronochrome*: The generalized chronochrome (introduced in Ref. [11], but based on the linear chronochrome developed by Schaum and Stocker [2]) considers contours not of the full distribution  $P(\mathbf{x}, \mathbf{y})$ , but of the conditional distribution  $P(\mathbf{y}|\mathbf{x})$ . Since the  $x$ -image and  $y$ -image are both of the same scene,  $P(\mathbf{y}|\mathbf{x})$  describes the expected variation of  $\mathbf{y}$ , at a pixel whose value in the  $x$ -image is observed to be  $\mathbf{x}$ . Low values of the conditional distribution correspond to unusual changes.

Since  $P(\mathbf{y}|\mathbf{x}) = P(\mathbf{x}, \mathbf{y})/P(\mathbf{x})$ , we can write  $P^*(\mathbf{x}, \mathbf{y}) = P(\mathbf{x})U(\mathbf{y})$ , and obtain contours of  $\mathcal{L}(\mathbf{x}, \mathbf{y})$  as given in Eq. (1).

It bears remarking that there is an asymmetry in the generalized (as well as the linear) chronochrome: if we consider expected changes in  $\mathbf{x}$  given  $\mathbf{y}$ , then we obtain  $P^*(\mathbf{x}, \mathbf{y}) = U(\mathbf{x})P(\mathbf{y})$  as the denominator in Eq. (1); different anomalous changes will be found depending on which of the two chronochromes are used.

3) *Anomalous change detection*: In the anomalous change framework proposed in Ref. [4], we take  $P^*(\mathbf{x}, \mathbf{y}) = P(\mathbf{x})P(\mathbf{y})$ . This is a distribution whose marginal distributions along the  $\mathbf{x}$  and  $\mathbf{y}$  directions match that of  $P(\mathbf{x}, \mathbf{y})$ , but for which  $\mathbf{x}$  and  $\mathbf{y}$  are treated as *independent* random variables. Thus, the straight “anomalousness” of a pixel in either of the individual images becomes irrelevant: it is anomalous relationships that are identified.

### C. When the distribution is Gaussian

A Gaussian  $P(\mathbf{x}, \mathbf{y})$  can be fully described in terms of its mean and covariance. Without loss of generality (and in keeping with usual practice), the mean is assumed to have been subtracted from each of the images. The covariance matrices can be written

$$X = \langle \mathbf{x}\mathbf{x}^T \rangle; \quad (2)$$

$$Y = \langle \mathbf{y}\mathbf{y}^T \rangle; \quad (3)$$

$$C = \langle \mathbf{y}\mathbf{x}^T \rangle. \quad (4)$$

Introducing  $\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$  as a vector in  $\mathcal{R}^{d_x+d_y}$  enables these equations to be summarized into a single covariance matrix:

$$K_o = \langle \mathbf{z}\mathbf{z}^T \rangle = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}. \quad (5)$$

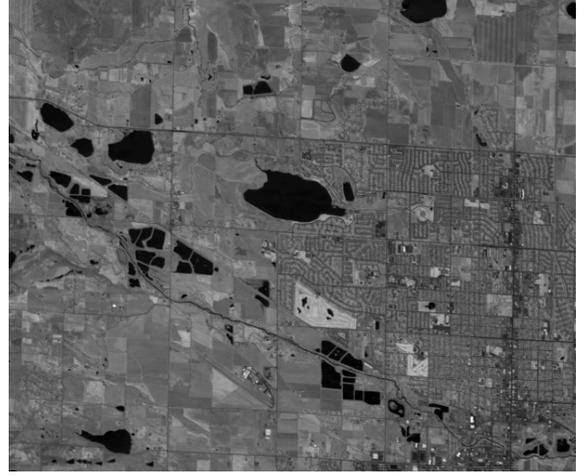


Fig. 1. Channel 75 (at an infrared wavelength of about  $1.06\mu\text{m}$ ) of a 224-channel AVIRIS image taken near Denver, Colorado. This image is of size  $614 \times 512$  pixels, with each pixel corresponding to approximately 15m on the ground.

And in the Gaussian case, the contours of the likelihood function  $\mathcal{L}(\mathbf{x}, \mathbf{y})$  are quadratic functions of  $\mathbf{x}$  and  $\mathbf{y}$ , which in general can be written by  $\mathbf{z}^T Q \mathbf{z}$  for some matrix  $Q$ .

For straight anomaly detection,  $Q = K_o^{-1}$ , and the likelihood contours follow the elliptical contours of  $\mathbf{z}^T K_o^{-1} \mathbf{z}$ .

When  $P(\mathbf{x}, \mathbf{y})$  is Gaussian, the generalized chronochrome becomes the standard chronochrome, and it can be shown that either  $Q = K_o^{-1} - \begin{bmatrix} X^{-1} & 0 \\ 0 & 0 \end{bmatrix}$  or  $Q = K_o^{-1} - \begin{bmatrix} 0 & 0 \\ 0 & Y^{-1} \end{bmatrix}$ , depending on which of the two chronochromes are used.

Finally, the anomalous change detection scheme developed in Ref. [4] leads to  $Q = K_o^{-1} - \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}^{-1}$ ; because the matrix  $Q$  has both positive and negative eigenvalues, the anomalous change boundaries are hyperbolic in  $\mathbf{z}$  space, and the algorithm is called “Hyper.”

## III. SUBPIXEL ANOMALOUS CHANGE

### A. Full-pixel anomalous change

Let  $(\mathbf{x}_1, \mathbf{y}_1)$  denote the  $(\mathbf{x}, \mathbf{y})$  values at a given location in the image. Let  $\mathbf{x}_2$  be the  $\mathbf{x}$  value at a different location, and  $\mathbf{y}_3$  the  $\mathbf{y}$  value at a third location. In general, a full-pixel anomalous change can be realized by sampling  $\mathbf{x}$  and  $\mathbf{y}$  values independently. Thus, while the point  $(\mathbf{x}_1, \mathbf{y}_1)$  corresponds to a “normal” change in the image, the point  $(\mathbf{x}_2, \mathbf{y}_3)$  is an anomalous change. That is:

$$\text{Normal: } \mathbf{z}_N = \begin{bmatrix} \mathbf{x}_N \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{bmatrix}; \quad (6)$$

$$\text{Anomalous: } \mathbf{z}_A = \begin{bmatrix} \mathbf{x}_A \\ \mathbf{y}_A \end{bmatrix} = \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{y}_3 \end{bmatrix}. \quad (7)$$

Here,  $\mathbf{z}_N$  is a typical point in the distribution  $P(\mathbf{x}, \mathbf{y})$ , while  $\mathbf{z}_A$  is a typical point in the distribution  $P^*(\mathbf{x}, \mathbf{y}) = P(\mathbf{x})P(\mathbf{y})$ .

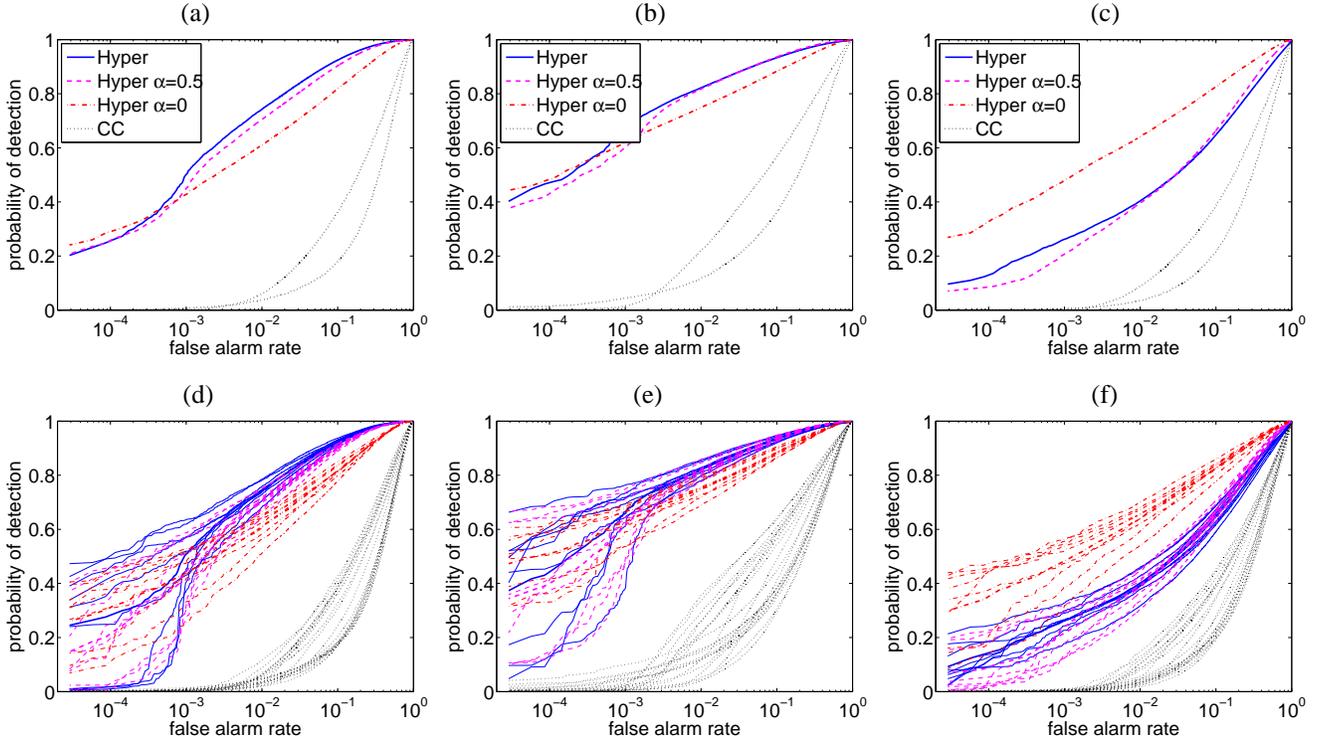


Fig. 2. ROC curves, on semilog axes, for five different anomalous change detectors: three variants of the Hyper algorithm, and two of the chronochrome (CC) algorithm. The top panels show the average of nine ROC curves, each corresponding to a different tile of the image. To illustrate the range of variation, the bottom panels show all nine of the ROC curves, as well as the average. The performance of the different algorithms is shown for: **(a,d)** full-pixel anomalous changes, **(b,e)** half-pixel anomalous changes, and **(c,f)** one-tenth-pixel anomalous changes.

### B. Subpixel anomalous change detection

For subpixel anomalies, with  $0 \leq \alpha \leq 1$  the fraction of the pixel that is covered by the anomaly, take linear combinations:

$$\begin{aligned} \mathbf{z}_N &= (1 - \alpha)\mathbf{z}_1 + \alpha\mathbf{z}_2 \\ &= (1 - \alpha) \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{bmatrix} + \alpha \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{y}_2 \end{bmatrix}; \end{aligned} \quad (8)$$

$$\begin{aligned} \mathbf{z}_A &= (1 - \alpha)\mathbf{z}_1 + \alpha\mathbf{z}_* \\ &= (1 - \alpha) \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{bmatrix} + \alpha \begin{bmatrix} \mathbf{x}_3 \\ \mathbf{y}_4 \end{bmatrix}. \end{aligned} \quad (9)$$

Note that the normal pixels values are also considered in terms of subpixel mixing. This is a necessary step because the naive approach – comparing the full-pixel  $\mathbf{z}_N$  from Eq. (6) with the subpixel  $\mathbf{z}_A$  in Eq. (9) – confounds the issues of normal-vs-anomaly with full-vs-subpixel. It in particular fails to account for the reduced variance that is produced by positive linear combinations; that effect leads to smaller overall variance in the distribution of  $\mathbf{z}_A$  values, which in turn leads to a paradoxical result (at least for small enough  $\alpha$ ) that identifies pixels near  $\mathbf{z} = 0$  as the anomalous changes.

The general distributions associated with these resamplings can be complicated, but for Gaussian distributions, the solution is straightforward. Here, the covariances are given by

$$\begin{aligned} K_N &= \langle \mathbf{z}_N \mathbf{z}_N^T \rangle = (1 + \alpha)^2 \langle \mathbf{z}_1 \mathbf{z}_1^T \rangle + \alpha^2 \langle \mathbf{z}_2 \mathbf{z}_2^T \rangle \\ &= [(1 + \alpha)^2 + \alpha^2] \langle \mathbf{z} \mathbf{z}^T \rangle \end{aligned}$$

$$= [(1 + \alpha)^2 + \alpha^2] \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}; \quad (10)$$

$$\begin{aligned} K_A &= \langle \mathbf{z}_A \mathbf{z}_A^T \rangle = (1 + \alpha)^2 \langle \mathbf{z} \mathbf{z}^T \rangle + \alpha^2 \langle \mathbf{z}_* \mathbf{z}_*^T \rangle \\ &= (1 + \alpha)^2 \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix} + \alpha^2 \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \end{aligned} \quad (11)$$

$$= [(1 - \alpha)^2 + \alpha^2] \begin{bmatrix} X & \theta C^T \\ \theta C & Y \end{bmatrix}, \quad (12)$$

where

$$\theta = \frac{(1 - \alpha)^2}{(1 - \alpha)^2 + \alpha^2}. \quad (13)$$

It follows that the contours that separate normal from anomalous changes will be given by constant values of  $\mathbf{z}^T Q \mathbf{z}$ , where

$$Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} X & \theta C^T \\ \theta C & Y \end{bmatrix}^{-1}. \quad (14)$$

When  $\alpha = 1$ , then  $\theta = 0$ , and this  $Q$  is the full-pixel anomalous change detection result from Ref. [4]. But by taking smaller values of  $\alpha$ , the anomalous change detector can be optimized for subpixel anomalies.

In the limit of vanishingly small subpixel anomaly (*i.e.*, as  $\alpha \rightarrow 0$ , so  $\theta \rightarrow 1$ ), we have that contours will be given by constant values of  $\mathbf{z}^T Q \mathbf{z}$ , where

$$Q = - \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} \begin{bmatrix} 0 & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1}. \quad (15)$$

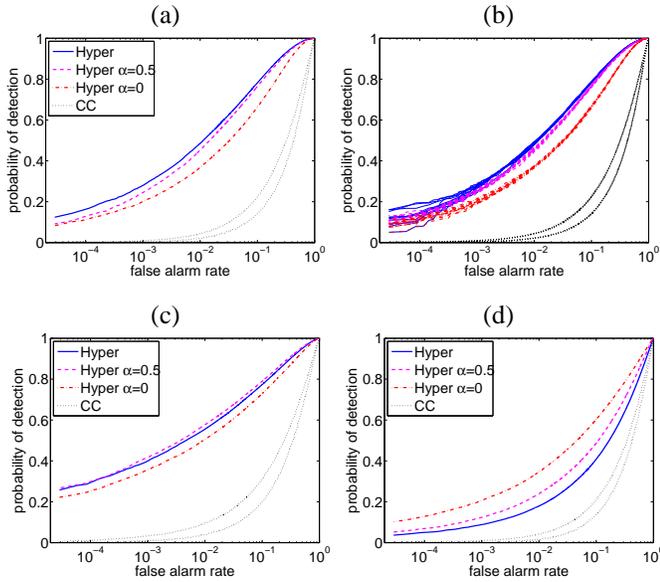


Fig. 3. ROC curves for the same five different anomalous change detectors shown in Fig. 2, but applied to Gaussian data with the same covariance as the real data. (a) Full-pixel anomalous changes; (b) full-pixel changes, showing all the ROC curves; (c) half-pixel; and (d) one-tenth pixel. Panel (b) shows that, unlike the real data, tile-to-tile variability of the Gaussian data is small.

Now, an actual change that occurs in a vanishingly small subpixel will be virtually undetectable. But the fact that this limit is well-defined provides a parameter-free detector for detecting small anomalous changes in imagery.

#### IV. RESULTS

To illustrate and test the optimized subpixel anomaly detector, an experiment was performed in which five anomalous change detectors were compared: two of these are the standard chronochrome detectors, and three are variants of the subpixel “Hyper” detector, given in Eq. (14). The three Hyper variants were: full-pixel ( $\alpha = 1$ ), half-pixel ( $\alpha = 0.5$ ), and the limiting case with  $\alpha \rightarrow 0$  which is given in Eq. (15).

An AVIRIS hyperspectral image (see Fig. 1) was cut into nine tiles, and for each tile, a noisy copy was made; this noise corresponds to the pervasive differences that would be observed in two images of the same scene. The noise was multiplicative and given by:  $y_i = x_i(1 + \epsilon g_i)$ , where  $g_i$  is a realization of unit variance Gaussian noise, and  $\epsilon$  is the noise level. Computing an anomalous measure  $\mathbf{z}^T \mathbf{Q} \mathbf{z}$  (where  $\mathbf{Q}$  depends on the anomalous change detection scheme) over the image pair provides an estimate of false-alarm rate versus anomalousness threshold.

Subpixel anomalous changes were then introduced into the image, and what is plotted in Fig. 2 and Fig. 3 is the fraction of those changes that are detected at the threshold that provides the given false alarm rate.

Since smaller noise levels  $\epsilon$  in the pervasive differences allow more sensitive detection of small anomalous changes, the three cases considered were: (a)  $\epsilon = 1.0$ ,  $\alpha = 1.0$ ; (b)  $\epsilon = 0.5$ ,  $\alpha = 0.5$ ; and (c)  $\epsilon = 0.1$ ,  $\alpha = 0.1$ .

These results are shown in the ROC curves of Fig. 2(a,b,c). For the full-pixel anomalous changes in Fig. 2(a), it is the full-pixel algorithm that exhibits the best performance; when the subpixel anomalous changes are very small (one-tenth pixel), then the limiting  $\alpha \rightarrow 0$  algorithm proved to work very well. There is considerably variability in this data, however, as illustrated in Fig. 2(d,e,f).

For the results shown in Fig. 3, the same computations were performed, but instead of using the real data, a simulated dataset was generated. The simulated data had the same mean and variance as the original data, but each pixel was an independent Gaussian variate. The results are essentially the same as for the real data, but with less ambiguity (as illustrated by in Fig. 3(b)).

Because the experiments are based on artificially introduced anomalies, they cannot be taken as definitive, but one of the clearest results from these tests – for real and for Gaussian data – is that both the full-pixel and subpixel anomalous change detection algorithm substantially outperformed the linear chronochrome algorithm.

In the situation of very small subpixel anomalies ( $\alpha = 0.1$ ), the limiting  $\alpha \rightarrow 0$  detector shows substantially better ability to detect these small changes than the other detectors that were considered – see Fig. 2(c,f) and Fig. 3(d). Interestingly, when the anomalous changes were larger, up to a full pixel, the  $\alpha \rightarrow 0$  detector was still competitive.

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, “Image change detection algorithms: A systematic survey,” *IEEE Trans. Image Processing*, vol. 14, pp. 294–307, 2005.
- [2] A. Schaum and A. Stocker, “Subclutter target detection using sequences of thermal infrared multispectral imagery,” *Proc. SPIE*, vol. 3071, pp. 12–22, 1997.
- [3] C. Clifton, “Change detection in overhead imagery using neural networks,” *Applied Intelligence*, vol. 18, pp. 215–234, 2003.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] S. Ben-David and M. Lindenbaum, “Learning distributions by their density levels: A paradigm for learning without a teacher,” *J. Computer and System Sciences*, vol. 55, pp. 171–182, 1997.
- [8] T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001, this anomaly detection approach is developed in Chapter 14.2.4, and neatly illustrated in Fig 14.3.
- [9] D. Tax and R. Duin, “Uniform object generation for optimizing one-class classifiers,” *J. Machine Learning Res.*, vol. 2, pp. 155–173, 2002.
- [10] I. Steinwart, D. Hush, and C. Scovel, “Density level detection is classification,” in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005, pp. 1337–1344.
- [11] J. Theiler and S. Perkins, “Resampling approach for anomalous change detection,” *Proc. SPIE*, vol. 6565, pp. 6565–65, 2007.