

Detection of Ephemeral Changes in Sequences of Images

James Theiler
Space and Remote Sensing Sciences
Los Alamos National Laboratory
Los Alamos, NM 87544
Email: jt@lanl.gov

Steven M. Adler-Golden
Spectral Sciences, Inc.
4 Fourth Avenue
Burlington, MA 01803-3304

Abstract—The formalism of anomalous change detection, which was developed for finding unusual changes in pairs of images, is extended to sequences of more than two images. Extended algorithms based on RX, Chronochrome, and Hyper are presented for identifying the most anomalously changing pixels in a sequence of co-registered images. Experimental comparisons are performed both on real data with real anomalies and on real data with simulated anomalies.

I. INTRODUCTION

While the identification of “interesting” features in a single image is an almost hopelessly open-ended task, the detection of interesting *changes* in a pair of co-registered images is a more feasible undertaking. The change detection problem is nonetheless confounded by pervasive differences (in illumination, calibration, registration, etc.) that are inevitable between images, but the *anomalous change detection* (ACD) paradigm [1]–[6] treats these differences as something that can be learned from the images themselves, and those changes that do not fit the pervasive pattern are identified as anomalous. It is these anomalous changes that are candidates for interesting features; ultimately, a human analyst decides what is *truly* interesting, but the algorithm’s job is to identify a short list of candidates for the analyst to investigate.

A recently developed machine learning framework [7] extended this change detection methodology to arbitrary data distributions, and even for Gaussian distributions was shown to exhibit improved performance [8]. But the ACD algorithms in Refs. [1]–[8], considered only the problem of finding pairwise changes; when more than two images are available, there is both a greater challenge (the change has more places to hide) and a greater opportunity (multiple correlations among all the images) for producing an effective change detector. In this situation, an interesting change might occur in just one of the images, and then disappear. Or it might linger. Or it might come and go. As the number of images increases, so too do the combinatorial possibilities. An “ephemeral” change detection algorithm should be sensitive to all of these possibilities,

An RX-based anomaly detector for multiple images has previously been proposed [9]. In this paper, we extend the machine learning approach [7], [8] to multiple images, enabling detection of an ephemeral change in one or a few of the images by exploiting the information in all of the images.

II. PAIRWISE ANOMALOUS CHANGE DETECTION

Let $\mathbf{x}_1 \in \mathfrak{R}^{d_1}$ and $\mathbf{x}_2 \in \mathfrak{R}^{d_2}$ be corresponding pixels in image 1 and 2, and define the stacked vector

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \quad (1)$$

We remark that if $d_1 = d_2 = 1$, then the images are single-band images (*i.e.*, “black and white”) but that the formalism holds as well for pairs of images that are multispectral ($d > 1$) or hyperspectral ($d \gg 1$, typically $d > 100$). Furthermore, there is no requirement that $d_1 = d_2$ or even $d_1 \approx d_2$.

An anomalous change detector is a scalar function $\mathcal{A}(\mathbf{z})$ which specifies the “anomalousness” of the change that is exhibited by the pixel going from \mathbf{x}_1 to \mathbf{x}_2 .

For an important class of ACD algorithms, this anomalousness is quadratic in \mathbf{z} , and can be expressed in terms of covariances and cross-covariances of the various images [8]. The first step for these algorithms is to subtract from each pixel the mean value over all the pixels in the image; for these mean-subtracted pixels, we have $\langle \mathbf{x} \rangle = 0$ (and therefore $\langle \mathbf{z} \rangle = 0$), and the expression for anomalousness is given by

$$\mathcal{A}(\mathbf{z}) = \mathbf{z}^T Q \mathbf{z}, \quad (2)$$

where quadratic coefficients in the matrix Q are functions of

$$C_{ij} = \langle \mathbf{x}_i \mathbf{x}_j^T \rangle, \quad (3)$$

the cross-covariance matrix between the i th and j th images.

A. RX

Write $P(\mathbf{x}_1, \mathbf{x}_2)$ as the joint probability distribution associated with the pixels in the first and second images. The “rare” pixel pairs are the values of $\mathbf{x}_1, \mathbf{x}_2$ where $P(\mathbf{x}_1, \mathbf{x}_2)$ is small. For a Gaussian distribution, low probability density is associated with large Mahalanobis distance to the centroid, which leads to the RX detector [10]. Here, we consider the covariance of the stacked vector \mathbf{z} , which is given by

$$\langle \mathbf{z} \mathbf{z}^T \rangle = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \quad (4)$$

and then the anomalousness measure is given by $A(\mathbf{z}) = \mathbf{z}^T Q_{\text{RX}} \mathbf{z}$, where

$$Q_{\text{RX}} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1}. \quad (5)$$

B. Hyper

While RX seeks anomalous values of the stacked vector \mathbf{z} , the machine learning framework [7] moderates that by the individual anomalousness of \mathbf{x}_1 and \mathbf{x}_2 . The idea is to emphasize not unusual *values* of \mathbf{x}_1 or \mathbf{x}_2 , but unusual *relationships* between \mathbf{x}_1 and \mathbf{x}_2 . This is achieved by comparing the joint distribution $P(\mathbf{x}_1, \mathbf{x}_2)$ to the product of the marginal distributions: $P(\mathbf{x}_1)P(\mathbf{x}_2)$. The ratio of these two distributions, in the Gaussian case, leads to an expression for anomalousness that produces hyperbolic boundaries; specifically $A(\mathbf{z}) = \mathbf{z}^T Q_{\text{Hyper}} \mathbf{z}$, where

$$Q_{\text{Hyper}} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1} - \begin{bmatrix} C_{11}^{-1} & 0 \\ 0 & C_{22}^{-1} \end{bmatrix}. \quad (6)$$

C. Chronochrome

In the chronochrome detector [1], [4], a least-squares linear fit is made of image 2 based on image 1. A linear operator (a matrix) L estimates

$$\mathbf{x}_2 \approx \hat{\mathbf{x}}_2 = L\mathbf{x}_1 \quad (7)$$

and L is chosen so that the residual

$$\mathbf{e} = \mathbf{x}_2 - \hat{\mathbf{x}}_2 = \mathbf{x}_2 - L\mathbf{x}_1 \quad (8)$$

has minimum variance. This variance is given by the trace of the dispersion matrix $E = \langle \mathbf{e}\mathbf{e}^T \rangle$, and one can show that this is minimized when $L = C_{21}C_{11}^{-1}$. The anomalousness measure is given by Mahalanobis distance from the centroid of the distribution of residuals; specifically,

$$\begin{aligned} A(\mathbf{x}_1, \mathbf{x}_2) &= \mathbf{e}^T E^{-1} \mathbf{e} \\ &= (\mathbf{x}_2 - C_{21}C_{11}^{-1}\mathbf{x}_1)^T \\ &\quad \times [C_{22} - C_{21}C_{11}^{-1}C_{12}]^{-1} \\ &\quad \times (\mathbf{x}_2 - C_{21}C_{11}^{-1}\mathbf{x}_1). \end{aligned} \quad (9)$$

A more compact representation is given by $A(\mathbf{z}) = \mathbf{z}^T Q_{\text{CC}} \mathbf{z}$, where [8]

$$Q_{\text{CC}} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1} - \begin{bmatrix} C_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \quad (10)$$

We remark that there is an asymmetry in the definition of the chronochrome; in this derivation, we fit \mathbf{x}_2 as a function of \mathbf{x}_1 . Had we chosen to fit $\mathbf{x}_1 \approx L'\mathbf{x}_2$, a different, and not equivalent, chronochrome detector would be obtained.

A symmetrized chronochrome can be obtained by averaging the two chronochromes; that is:

$$Q_{\text{CCsym}} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1} - \frac{1}{2} \begin{bmatrix} C_{11}^{-1} & 0 \\ 0 & C_{22}^{-1} \end{bmatrix}. \quad (11)$$

We remark that these are just a few examples (out of a larger set that was surveyed in Ref. [8]) of quadratic covariance-based ACD algorithms. The other algorithms have a similar

flavor, but use different expressions for Q . We also remark that the three detectors defined in Eqs. (5,6,11) are related by a simple expression:

$$Q_{\text{CCsym}} = \frac{1}{2}(Q_{\text{RX}} + Q_{\text{Hyper}}). \quad (12)$$

III. EXTENSION TO MULTIPLE IMAGES

Following the notation in the previous section, we will write \mathbf{x}_n for a pixel in the n th image, and consider the stacked vector

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}. \quad (13)$$

With this notation, we can extend the quadratic detectors described in the previous section so that they can be applied to more than two images at a time.

A. Extended RX

The extension of RX to the multiple-image change detection problem was described in Ref. [9] and was called ‘‘temporal whitening.’’ As with RX for two images, the anomalousness measure is given by $A(\mathbf{z}) = \mathbf{z}^T Q_{\text{RX}} \mathbf{z}$, where now

$$Q_{\text{RX}} = \langle \mathbf{z}\mathbf{z}^T \rangle^{-1} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}^{-1}, \quad (14)$$

and C_{ij} is defined in Eq. (3) as the cross-covariance between the i th and j th images. In the more general, *i.e.*, non-Gaussian, case, the anomalies are identified as those pixels that occur at low values of the joint distribution $P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$.

B. Extended Hyper

The extension of the machine learning framework to more than two images leads to a detector of small values of the ratio of the joint distribution to the product of marginals; that is:

$$\frac{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)}{P(\mathbf{x}_1)P(\mathbf{x}_2) \cdots P(\mathbf{x}_n)}. \quad (15)$$

For Gaussian distributions, this leads to $A(\mathbf{z}) = \mathbf{z}^T Q_{\text{Hyper}} \mathbf{z}$, where Q_{Hyper} can be expressed as

$$Q_{\text{Hyper}} = Q_{\text{RX}} - Q_{\Delta} \quad (16)$$

and

$$Q_{\Delta} = \begin{bmatrix} C_{11}^{-1} & 0 & \cdots & 0 \\ 0 & C_{22}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_{nn}^{-1} \end{bmatrix}. \quad (17)$$

C. Extended Chronochrome

The description of the pairwise chronochrome in Section II-C noted the ambiguity of the change $\mathbf{x}_1 \rightarrow \mathbf{x}_2$ versus $\mathbf{x}_2 \rightarrow \mathbf{x}_1$. For more than two images, this ambiguity becomes combinatorial. Let I and J denote nonempty disjoint subsets of $\{1, \dots, n\}$. Write \mathbf{x}_I as the stacked vector composed of the pixel elements \mathbf{x}_i with $i \in I$ (and similarly for J). Then we can approximate \mathbf{x}_J using a linear transform of \mathbf{x}_I ; in particular, we can find a matrix L_{JI} so that the residual

$$\mathbf{e} = \mathbf{x}_J - \hat{\mathbf{x}}_J = \mathbf{x}_J - L_{JI}\mathbf{x}_I \quad (18)$$

has minimum variance. We can write this matrix as $L_{JI} = C_{JI}C_{II}^{-1}$, where C_{JI} is composed of blocks given by C_{ji} where $j \in J$ and $i \in I$, and C_{II} is composed of blocks given by $C_{ii'}$ for $i \in I$ and $i' \in I$. As with the pairwise chronochrome, the anomalousness of the pixel is given by $\mathbf{e}^T \langle \mathbf{e}\mathbf{e}^T \rangle^{-1} \mathbf{e}$.

Since the anomalousness depends only on the images that are in the union $I \cup J$, it is convenient to write it in terms of the modified stacked vector

$$\hat{\mathbf{z}} = \begin{bmatrix} \mathbf{x}_I \\ \mathbf{x}_J \end{bmatrix}, \quad (19)$$

in which case the anomalousness is given by

$$A(\hat{\mathbf{z}}) = \hat{\mathbf{z}}^T \left(\begin{bmatrix} C_{II} & C_{IJ} \\ C_{JI} & C_{JJ} \end{bmatrix}^{-1} - \begin{bmatrix} C_{II}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right) \hat{\mathbf{z}}. \quad (20)$$

If we write P_{IJ} as the permutation/projection operator that achieves $\hat{\mathbf{z}} = P_{IJ}\mathbf{z}$, then we can write the quadratic coefficient matrix for the extended chronochrome

$$Q_{IJ} = P_{IJ}^T \left(\begin{bmatrix} C_{II} & C_{IJ} \\ C_{JI} & C_{JJ} \end{bmatrix}^{-1} - \begin{bmatrix} C_{II}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right) P_{IJ}. \quad (21)$$

Two problems with this detector are that it is asymmetrical, and that there are exponentially many (roughly 3^n) ways to choose I and J . Both of these can be addressed by symmetrizing the detector. There are many ways this might be done, but we will describe two natural approaches.

1) *Symmetric Extended Chronochrome: CC-I:* Perhaps the easiest way to symmetrize the extended chronochrome is to follow the expression in Eq. (11), and write

$$Q_{CC-I} = Q_{RX} - \frac{1}{n} Q_{\Delta} \quad (22)$$

$$= \left(1 - \frac{1}{n}\right) Q_{RX} + \frac{1}{n} Q_{\text{Hyper}}. \quad (23)$$

This is a simple and easy-to-compute expression that is based on averaging n extended chronochromes. The i th chronochrome in this scheme is given by $I = \{i\}$, and $J = \{1, \dots, i-1, i+1, \dots, n\}$, and corresponds to fitting all but the i th image as a linear function of the i th image.

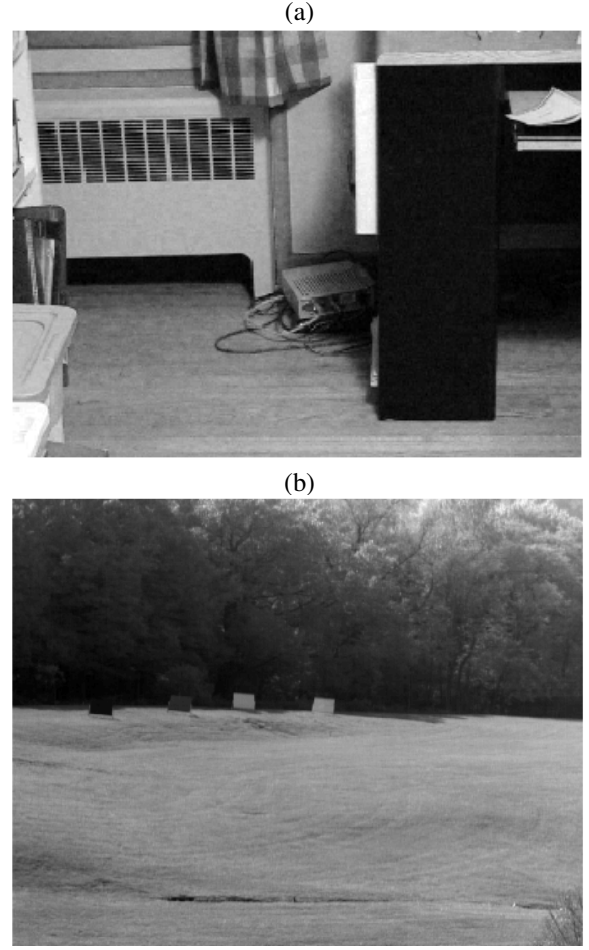


Fig. 1. (a) First frame of a short video clip that exhibits both pervasive differences (misregistration caused by moving the camera) and actual change (a blinking light). (b) Broadband image from a sequence of five hyperspectral data cubes, all of the same scene, and one of which includes as actual changes two tarps that have been placed on the grass (not seen in this image).

2) *Symmetric Extended Chronochrome: CC-II:* A somewhat more complicated but intuitively more appealing scheme reverses the role of I and J in CC-I. To be more specific, write

$$I_i = \{1, \dots, i-1, i+1, \dots, n\}, \text{ and} \quad (24)$$

$$J_i = \{i\}. \quad (25)$$

The i th chronochrome is a fit of the i th image as a linear function of the other images; the quadratic coefficient matrix for that is given by $Q_i = Q_{I_i J_i}$ where Q_{IJ} is given in Eq. (21). The symmetrized detector is then taken as the average over these chronochrome detectors; *i.e.*,

$$Q_{CC-II} = \frac{1}{n} \sum_{i=1}^n Q_i \quad (26)$$

$$= Q_{RX} - \frac{1}{n} \sum_{i=1}^n P_{I_i J_i}^T \begin{bmatrix} C_{I_i I_i}^{-1} & 0 \\ 0 & 0 \end{bmatrix} P_{I_i J_i}. \quad (27)$$

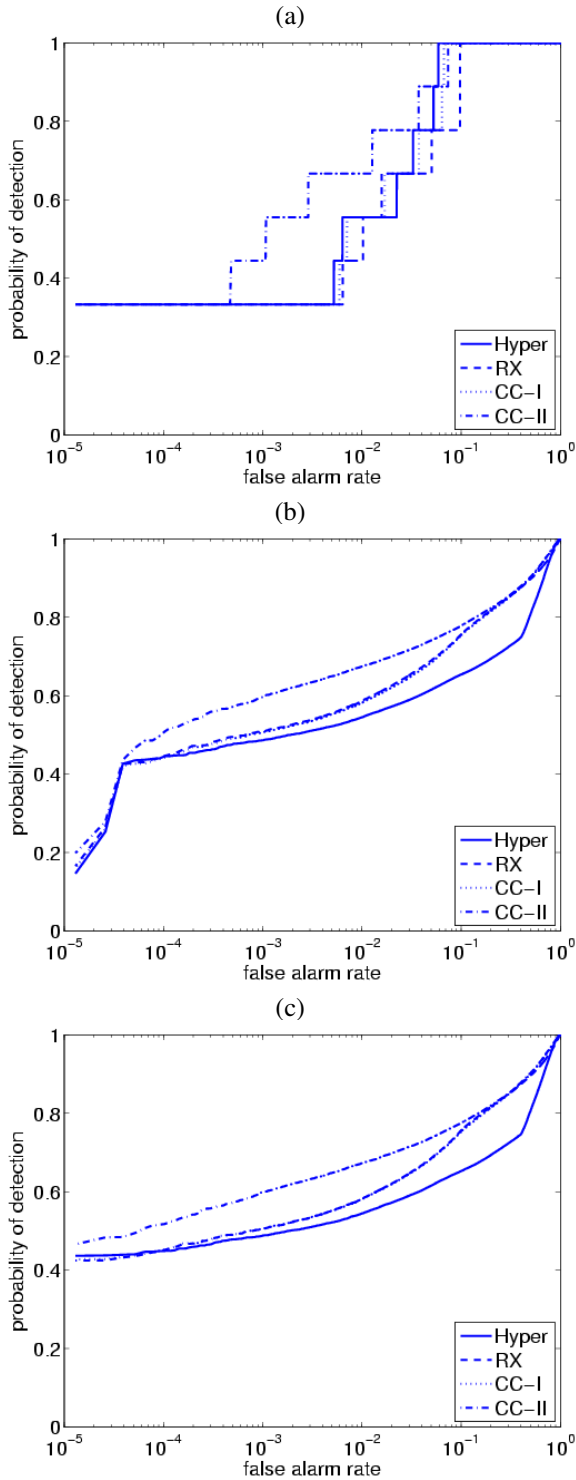


Fig. 2. ROC curves compare the extended anomalous change detectors provided by RX, Hyper, CC-I, and CC-II, described in Sections III-A, III-B, III-C1, and III-C2, respectively, applied to the dataset shown in Fig. 1(a). (a) Applied directly to the dataset, using the known anomalous change (the blinking light), the performance of the first three of these algorithms is nearly identical, while CC-II appears to perform better. (b) Artificial anomalous changes are injected into the 16th image in the frame. (c) Again, the anomalous changes are simulated, but in this case, the blinking light is numerically suppressed. In both of the simulated cases, RX is slightly better than Hyper in the high false alarm rate regime, but when the false alarm rate is low, the two methods are nearly identical. We see that CC-I performs almost identically to RX, and that CC-II provides the best performance.

IV. NUMERICAL EXPERIMENTS

We will apply and compare these algorithms on a variety of datasets, using both real and simulated anomalies. In particular, we have low-resolution single-band video [9], higher-resolution RGB video, and a short sequence of ten-band imagery that is obtained as the first ten principal components of what was initially 124-band hyperspectral imagery [5], [6].

A. Single-band video

In Ref. [9], a thirty-frame video clip (240×320 pixels) was taken of an office scene [Fig. 1(a)]. There is considerable frame-to-frame jitter (pervasive differences) in the images, and the actual temporal change is a slowly blinking light (about six blink cycles in the thirty frames). A “target” area of size 3×3 describes the actual blinking light, and an 11×11 buffer around this area separates it from the rest of the image which is considered off-target.

Based on this markup, ROC curves are computed for different ephemeral change detectors. Although Hyper often outperforms RX in detecting change between two images [8], we see in Fig. 2(a) for the dataset in Fig. 1(a) that the extended Hyper and RX algorithms produce almost identical performance. Similar performance is also observed for CC-I, the simple symmetrized chronochrome, but somewhat better performance is observed for CC-II.

B. RGB video

The single-band imagery described above was initially acquired as higher resolution (480×640 pixels) color video, and in this study we used that original data as well. Fig. 3(a) compares Hyper, RX, CC-I and CC-II, and as with the single band imagery, finds that Hyper, RX, and CC-I perform similarly, while CC-II is somewhat better.

C. Multispectral image sequence

A remarkably extensive change detection data set was acquired by Eismann *et al.* [5], [6], using a 124-channel hyperspectral imager viewing the same scene over the course of many months [Fig. 1(b)]. In a few images, anomalous changes are created by placing objects (small tarps) in the scene. The data sequence used here runs from August to November, and the third of five images includes the known change. We used the first ten principal components of the August image to define the projection of the 124 images to a ten channel multispectral dataset. It is on this dataset that the algorithms are compared in Fig. 4(a). In this case, the Hyper algorithm significantly outperformed the RX and chronochrome algorithms.

D. Simulation framework

Although the ultimate test of an algorithm’s utility is its ability to detect real anomalies in real images, this test is difficult to employ in practice. A detector that is better at finding one anomaly might be worse at finding another, so any statistically reliable comparison of detectors will require a lot of anomalies. But our algorithms are built around the

premise that real anomalies are rare. So although we will also compare algorithms by looking at real anomalies, an important part of our testing uses simulated anomalies.

Since we are seeking anomalous *changes*, as opposed to anomalies in general, our simulation framework uses a scheme that employs normal pixels. For two images, it has been proposed that this can be achieved by scrambling the pixels in one of the images [8]. That is: the normal changes are given by the data in its original form. The anomalous changes are given by all of the pixels in an image pair that is the same data as the original data, but with the pixels in one of the images scrambled. Here, we extend this to sequences of multiple images by choosing one of the images and scrambling the pixels in that dataset. Initially, we made n sets of these simulated anomaly image sequences, and the performance was given by the average over all of these. We found, however, for the examples we considered in this study, that by scrambling only the middle image (*e.g.*, 16th frame in a 30 frame sequence), we obtained similar results.

In Fig. 2(b,c), Fig. 3(b), and Fig. 4, ROC curves are computed for simulated anomalies. By and large, the results obtained for the real anomalies are echoed in these more extensive (but arguably less realistic) tests.

V. ANALYSIS

A. The case of highly correlated images

It is often the case in sequences of images, particularly of high-cadence video sequences, that there is high correlation between the images. In this high correlation limit, we find that RX and Hyper produce virtually identical detectors. A simple model shows why this is the behavior that would be expected.

In comparing the relative values of anomalousness for Hyper and RX, we consider the ratio

$$\frac{A_{\text{RX}}(\mathbf{z}) - A_{\text{Hyper}}(\mathbf{z})}{A_{\text{RX}}(\mathbf{z})} = \frac{\mathbf{z}^T Q_{\Delta} \mathbf{z}}{\mathbf{z}^T Q_{\text{RX}} \mathbf{z}}. \quad (28)$$

When the ratio is small, RX and Hyper are approximately equal; this will happen when the eigenvalues of Q_{Δ} are small compared to the eigenvalues of Q_{RX} . Fig. 5 shows what those eigenvalues are for two specific datasets.

To help make this comparison, first whiten the images. That is, apply the linear transformation $\tilde{\mathbf{x}}_i = C_{ii}^{-1/2} \mathbf{x}_i$ to the i th image. These transformations will not affect the behavior or the RX or Hyper detectors [8], and will produce new covariances and cross-covariances $\tilde{C}_{ij} = \langle \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_j^T \rangle$. In particular, we have that the whitened covariances $\tilde{C}_{ii} = I$ are identity matrices, and this leads to the result that the whitened difference matrix is itself the identity:

$$\tilde{Q}_{\Delta} = \begin{bmatrix} \tilde{C}_{11}^{-1} & 0 & \cdots & 0 \\ 0 & \tilde{C}_{22}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{C}_{nn}^{-1} \end{bmatrix} = I. \quad (29)$$

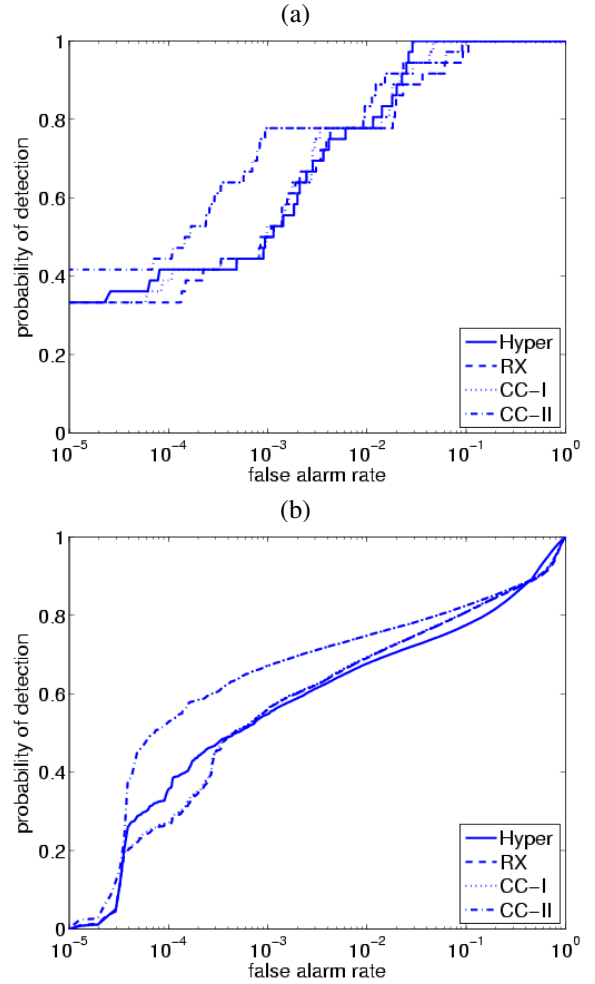


Fig. 3. ROC curves compare anomalous change detectors on a higher resolution RGB video clip of the blinking light. Two cases are considered: (a) the blinking light is the anomalous change; (b) anomalous changes are simulated.

Thus the ratio in Eq. (28) can be written

$$\frac{\mathbf{z}^T Q_{\Delta} \mathbf{z}}{\mathbf{z}^T Q_{\text{RX}} \mathbf{z}} = \frac{\tilde{\mathbf{z}}^T \tilde{Q}_{\Delta} \tilde{\mathbf{z}}}{\tilde{\mathbf{z}}^T \tilde{Q}_{\text{RX}} \tilde{\mathbf{z}}} = \frac{\tilde{\mathbf{z}}^T \tilde{\mathbf{z}}}{\tilde{\mathbf{z}}^T \tilde{Q}_{\text{RX}} \tilde{\mathbf{z}}} \quad (30)$$

and this ratio depends only on the eigenvalues of \tilde{Q}_{RX} .

In our simple model, we will presume that the images are highly correlated with each other, and in particular, we will write $C_{ij} = (1 - \epsilon)I$ for $i \neq j$, with $0 < \epsilon \ll 1$. This simple model considers only the case where all of the images have the same number of channels: $d = d_1 = d_2 = \cdots = d_n$. This leads to an expression for \tilde{Q}_{RX} given by

$$\tilde{Q}_{\text{RX}} = \begin{bmatrix} \tilde{C}_{11} & \tilde{C}_{12} & \cdots & \tilde{C}_{1n} \\ \tilde{C}_{21} & \tilde{C}_{22} & \cdots & \tilde{C}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{C}_{n1} & \tilde{C}_{n2} & \cdots & \tilde{C}_{nn} \end{bmatrix}^{-1}$$

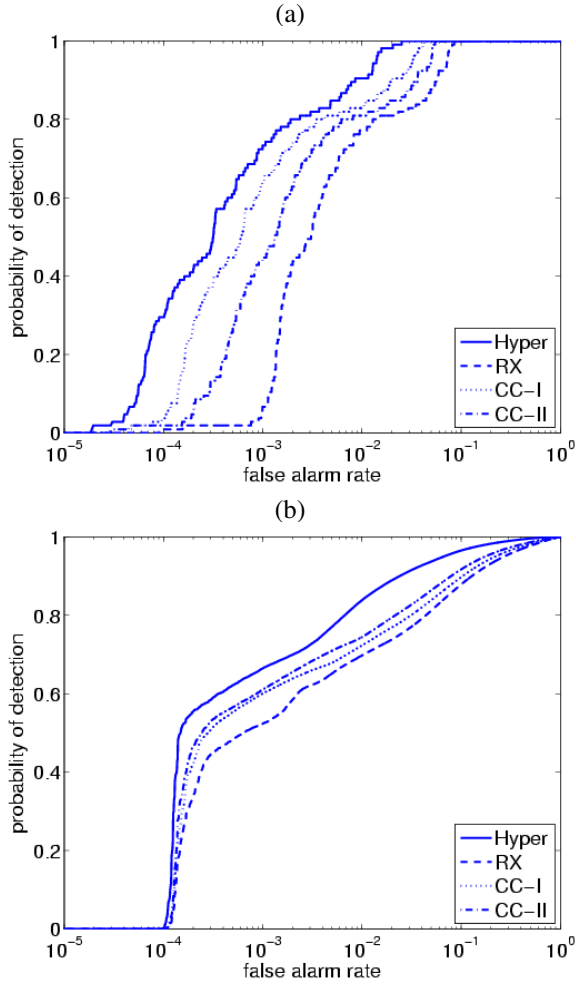


Fig. 4. ROC curves compare anomalous change detectors applied to a sequence of five multispectral (ten channel) images. For this multispectral imagery, the hyperbolic algorithm provides a decided performance advantage, both for the case (a) where the actual anomalous change is used, and for the case (b) with simulated anomalous changes. For these higher-dimensional less highly correlated images, the Hyper detector is more effective at detecting both the real and the simulated anomalous changes.

$$\begin{aligned}
 &= \begin{bmatrix} I_{11} & (1-\epsilon)I_{12} & \cdots & (1-\epsilon)I_{1n} \\ (1-\epsilon)I_{21} & I_{22} & \cdots & (1-\epsilon)I_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (1-\epsilon)I_{n1} & (1-\epsilon)I_{n2} & \cdots & I_{nn} \end{bmatrix}^{-1} \\
 &= \begin{bmatrix} (1-\epsilon) & & & \\ & \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1n} \\ I_{21} & I_{22} & \cdots & I_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n1} & I_{n2} & \cdots & I_{nn} \end{bmatrix} & & \\ & & & & + \epsilon I \end{bmatrix}^{-1}. \quad (31)
 \end{aligned}$$

Now, the matrix of all I 's has a total of nd eigenvalues, d of which have value n , and $(n-1)d$ of which have value zero. Thus, \tilde{Q}_{RX} has d small eigenvalues $1/[(1-\epsilon)n+\epsilon]$ and $(n-1)d$ large eigenvalues $1/\epsilon$. Most of the eigenvalues are much larger than one, so for directions \mathbf{z} with significant components in any of these $(n-1)d$ directions, the ratio in Eq. (30) (and therefore in Eq. (28) as well) will be small, and the RX and

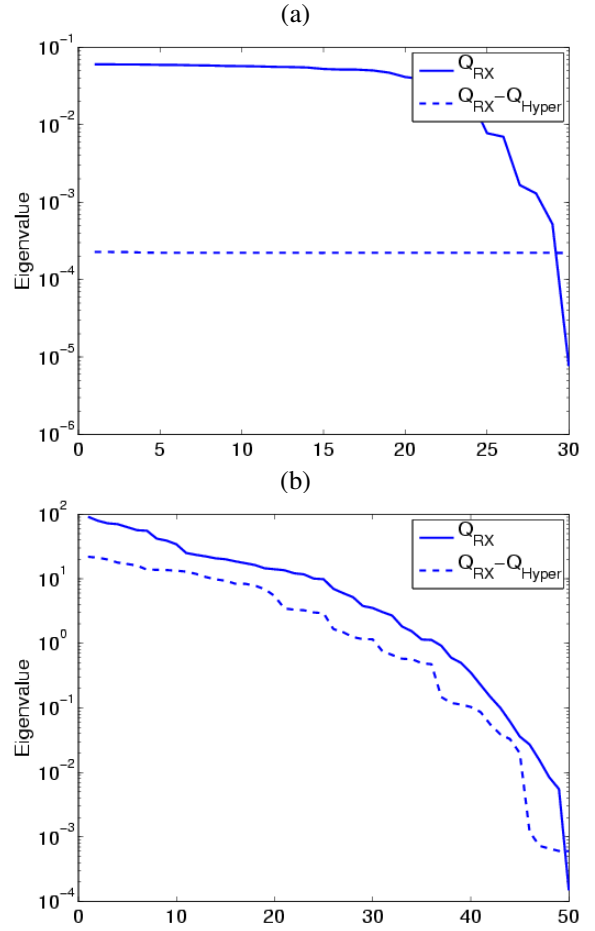


Fig. 5. Eigenvalues of the matrix Q_{RX} and the difference matrix $Q_{\Delta} = Q_{RX} - Q_{Hyper}$. When the eigenvalues of Q_{RX} are much larger than the eigenvalues of the difference matrix, then RX and Hyper will provide nearly identical performance. (a) In the blinking light data, there is high correlation between the images, and the eigenvalues of the difference matrix are much smaller than the eigenvalues of Q_{RX} . (b) In the Eismann-Meola data, the eigenvalue difference is not as large, and Fig. 4 shows that RX and Hyper give considerably different results for these data.

Hyper detectors will be nearly identical. For those directions where the ratio in Eq. (28) is not small, we will have that $\tilde{\mathbf{z}}^T \tilde{Q}_{RX} \tilde{\mathbf{z}}$ will be small; that is to say the difference between RX and Hyper is significant only when the anomalousness itself is small. The more anomalous is the change given by \mathbf{z} , the more alike will be the anomalousness measures of RX and Hyper.

For a sequence of highly correlated images, therefore, there is little difference in which pixels RX will find anomalously changed, versus which pixels Hyper will find anomalously changed. We remark that this argument holds for the CC-I detector as well, since it is a linear combination of RX and Hyper, but not for the CC-II detector.

B. Dimension Reduction

For a sequence of multispectral images, it can be advantageous, from both a statistical and a computational point of view, to transform the data into a sequence of lower-dimensional images.

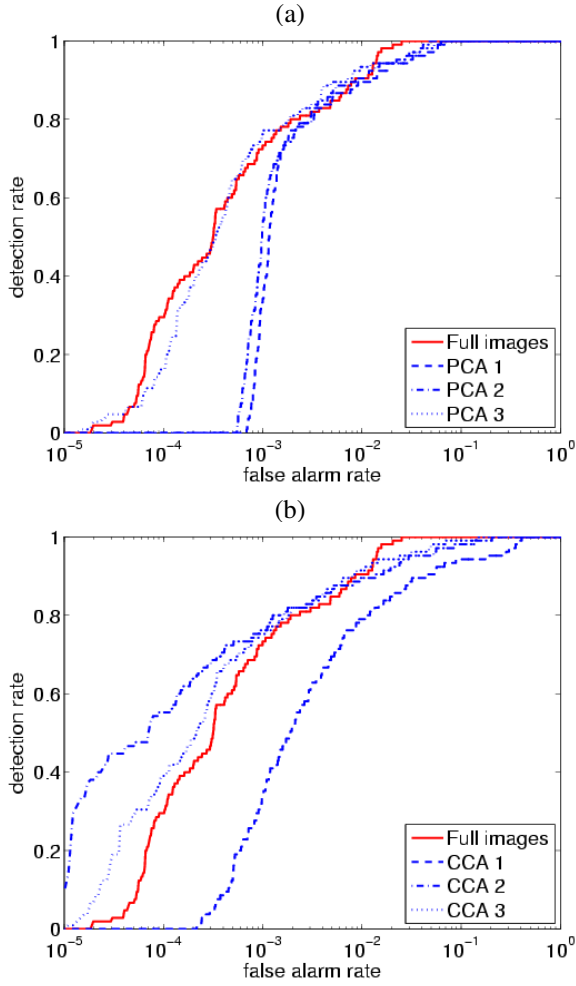


Fig. 6. ROC curves comparing change detection performance as a function of dimension reduction. There are five ten-channel “full images” in the dataset, and dimension reduction is performed using (a) principal components analysis (PCA), and (b) canonical components analysis (CCA). The utility of CCA for dimension reduction is evident in these curves. The first CCA component did not perform even as well as the first principal component, but for two components, CCA was a substantial improvement. In both cases (CCA and PCA), the $d = 3$ case performed about the same as with the full ten-dimensional dataset.

The simplest way to do this is by applying principal components analysis (PCA) to each of the images, and keeping only the top few components. An attractive alternative, however, for two or more images, is to employ canonical components analysis (CCA) instead. Although CCA is traditionally defined for two sets of variables (*e.g.*, two images) [11], it is possible to generalize it to multiple variables (*e.g.*, multiple images) [12]. Where PCA optimizes the variance in the individual images, CCA optimizes the correlation between the images. This can be advantageous because the more correlated images in a sequence are, the more that anomalous deviations from that correlation will stand out.

For two images, \mathbf{x}_1 and \mathbf{x}_2 , the first CCA component is a pair of vectors \mathbf{a}_1 and \mathbf{a}_2 which maximize the correlation between the scalar images $\mathbf{a}_1^T \mathbf{x}_1$ and $\mathbf{a}_2^T \mathbf{x}_2$. Specifically, the aim is to maximize $\langle (\mathbf{a}_1^T \mathbf{x}_1)(\mathbf{a}_2^T \mathbf{x}_2) \rangle = \mathbf{a}_1^T \langle \mathbf{x}_1 \mathbf{x}_2^T \rangle \mathbf{a}_2 =$

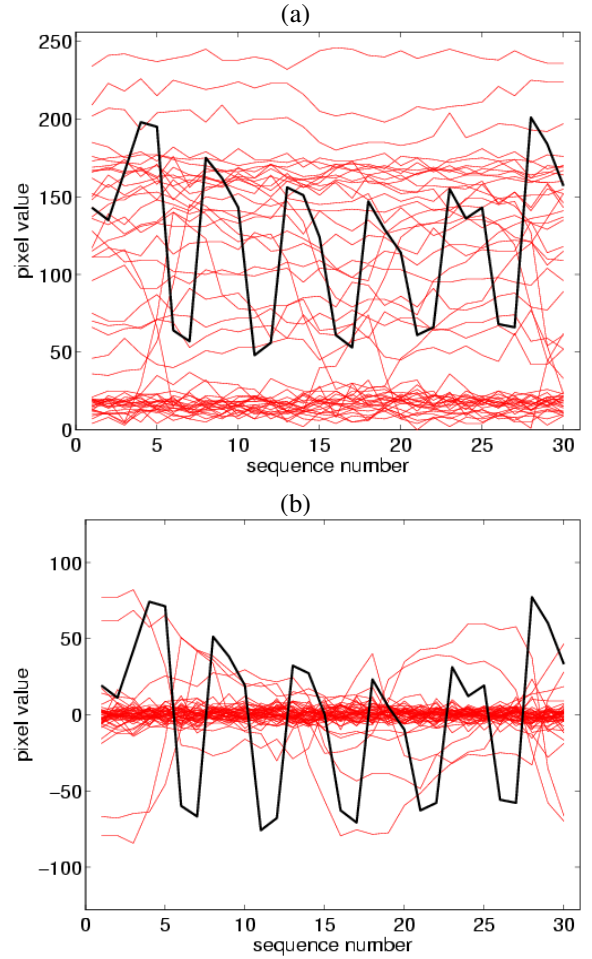


Fig. 7. Time sequences from the blinking light movie in Fig. 1(a). Panel (a) shows the time history of fifty randomly chosen pixels in the image as thin lines, and the history of the pixel corresponding to the blinking light as the heavy line. Panel (b) shows what these time sequences look like after pixel-wise subtraction.

$\mathbf{a}_1^T C_{12} \mathbf{a}_2$ subject to the constraints $\langle (\mathbf{a}_1^T \mathbf{x}_1)^2 \rangle = \mathbf{a}_1^T C_{11} \mathbf{a}_1 = 1$ and $\langle (\mathbf{a}_2^T \mathbf{x}_2)^2 \rangle = \mathbf{a}_2^T C_{22} \mathbf{a}_2 = 1$. An alternative formulation of this optimization is in terms of the generalized eigenvector problem:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} = \lambda \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix}. \quad (32)$$

This alternative formulation provides a natural generalization to sequences of more than two images:

$$Q_{RX}^{-1} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} = \lambda Q_{\Delta}^{-1} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}. \quad (33)$$

where Q_{RX} is defined in Eq. (14), and Q_{Δ} is defined in Eq. (17). The k generalized eigenvectors with the largest eigenvalues provide a projection of the images to a sequence of k -channel images. Fig. 6 compares the first three PCA and first three CCA components to the full ten components in the multispectral dataset shown in Fig. 1(b).

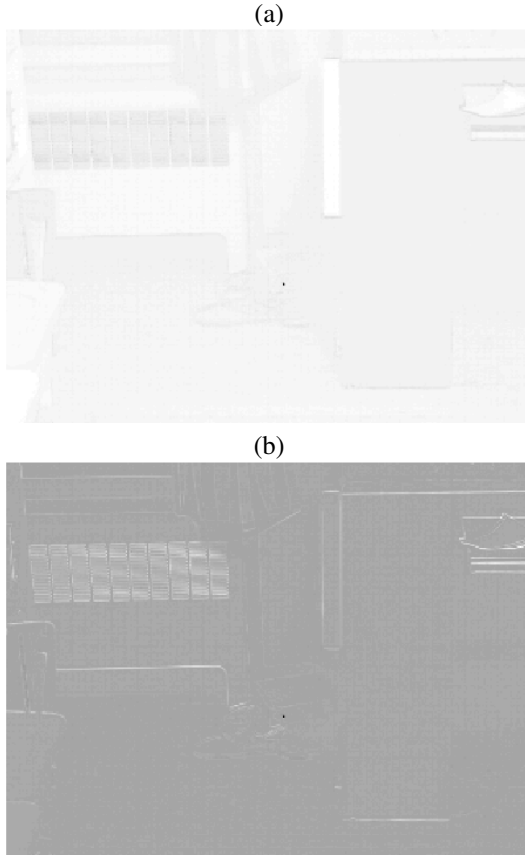


Fig. 8. Anomalousness image for the blinking light movie clip, based on (a) image-wise mean subtraction, and (b) pixel-wise mean subtraction.

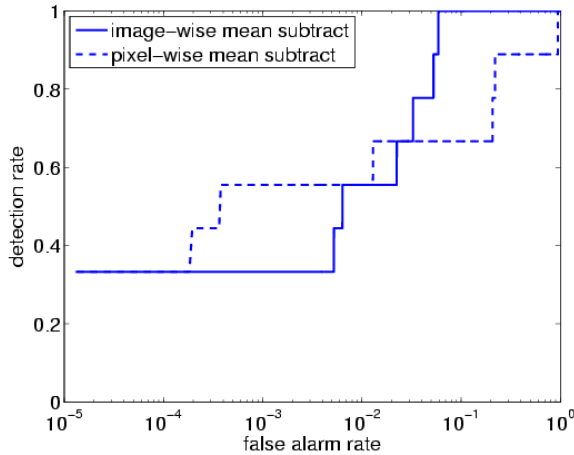


Fig. 9. ROC curves for the blinking light movie clip, using both image-based, and pixel-based mean subtraction.

C. Pixel-wise mean subtraction

In the approaches described so far, the first step is an image-wise mean subtraction. The mean value $\langle \mathbf{x} \rangle$ is computed over all the pixels in the image and this mean value is subtracted from each of the pixels in the image.

Fig. 7 illustrates a more aggressive mean subtraction. Here, the mean value for each pixel is computed from the corre-

sponding pixel in each of the images. This pixel-wise mean subtraction is performed in addition to the image-wise mean subtraction, and it only makes sense if each image in the sequence has the same number of channels. With only a few images, pixel-wise mean subtraction would likely throw away valuable information, but for a larger number, it may lead to better performance. Note that the pixel-wise mean-subtracted covariance matrix is singular, so a pseudo-inverse is needed for Q_{RX}^{-1} . Fig. 8 shows that the anomalousness image looks qualitatively different under the two schemes. Fig. 9 shows that the difference is not just cosmetic; the ROC curves differ as well.

VI. CONCLUSION

Multi-image extensions to the RX, Chronochrome, and Hyper algorithms are introduced and compared. All of these algorithms can be expressed as quadratic functions of a “stacked” pixel \mathbf{z} , defined in Eq. (13). Although Hyper is seen to perform better on a few frames of multispectral imagery, its performance on sequences of low-dimensional highly correlated images (as from a video sequence) is nearly identical to that of the simpler RX algorithm. A symmetrized variant of the extended chronochrome is found to outperform RX and Hyper in this video regime.

ACKNOWLEDGMENTS

We are grateful to Joseph Meola and Michael Eismann for generously providing their hyperspectral change data.

Work at Los Alamos was funded by the Laboratory Directed Research and Development (LDRD) program.

REFERENCES

- [1] A. Schaum and A. Stocker, “Long-interval chronochrome target detection,” *Proc. 1997 International Symposium on Spectral Sensing Research*, 1998.
- [2] C. Clifton, “Change detection in overhead imagery using neural networks,” *Applied Intelligence*, vol. 18, pp. 215–234, 2003.
- [3] A. Schaum and A. Stocker, “Joint subspace detection in hyperspectral sensing,” *Proc. MSS (Military Sensing Symposium) on Camouflage, Concealment, and Deception*, 2003.
- [4] A. Schaum and E. Allman, “Advanced algorithms for autonomous hyperspectral change detection,” in *33rd Applied Imagery Pattern Recognition (AIPR) Workshop: Emerging technologies and applications for imagery pattern recognition*. IEEE Computer Society Press, 2005, pp. 33–38.
- [5] 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.
- [6] J. Meola and M. T. Eismann, “Image misregistration effects on hyperspectral change detection,” *Proc. SPIE*, vol. 6966, p. 69660Y, 2008.
- [7] 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.
- [8] J. Theiler, “Quantitative comparison of quadratic covariance-based anomalous change detectors,” *Applied Optics*, vol. 47, pp. F12–F26, 2008.
- [9] S. M. Adler-Golden, S. C. Richtsmeier, and R. Shroll, “Suppression of subpixel sensor jitter fluctuations using temporal whitening,” *Proc. SPIE*, vol. 6969, p. 69691D, 2008.
- [10] 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.
- [11] H. Hotelling, “Relations between two sets of variables,” *Biometrika*, vol. 28, pp. 321–327, 1936.
- [12] J. R. Kettenring, “Canonical analysis of several sets of variables,” *Biometrika*, vol. 58, pp. 433–451, 1971.