Proposed Framework for Anomalous Change Detection

James Theiler JT@LANL.GOV
Simon Perkins S.PERKINS@LANL.GOV
Los Alamos National Laboratory, Los Alamos, NM 87545 USA

Abstract

For the anomalous change detection problem, you have a pair of images, taken of the same scene, but at different times and typically under different viewing conditions. You are looking for interesting differences between the two images. There will be some differences that are pervasive, perhaps due to overall contrast, brightness or focus differences, or maybe due to atmospheric or even seasonal changes – but there may also be changes that occur in only a few pixels. These rare changes are potentially indicative of something truly changing in the scene, and the idea is to use anomaly detection to find them. But you want to identify the *changes* that are unusual. You do not want to be confounded by ususual pixels that are "similarly unusual" in both images. We propose a machine learning framework for identifying these anomalous changes.

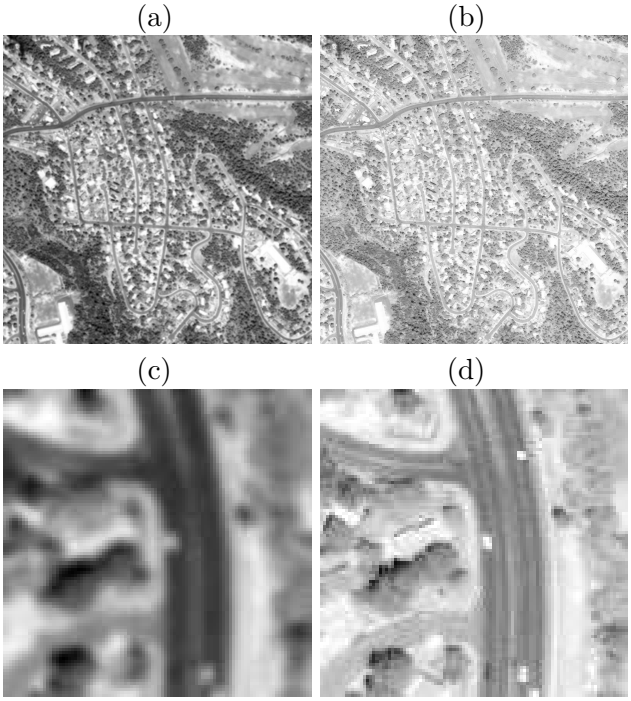

Figure 1. **(a,b)** What is changed? The right image is brighter, and sharper, but has less contrast. Every pixel in the image has changed, though mostly in predictable ways. **(c,d)** Close-up of the images in (a,b); as well as changes in brightness, focus, and contrast, the bottom image exhibits an actual change – an extra car has been (artificially) added to the scene.

1. Introduction

Change detection in imagery is of broad general interest (Radke et al., 2005), but it is especially useful in remote sensing. Using pictures taken from satellite or airborne platforms, accurate maps can be made of what is on the ground. These maps are valuable, but they are not static: new highways are built, new agricultural fields are planted, new crop circles mysteriously appear, new housing is constructed, old warehouses are bulldozed, wetlands are encroached, forests are burned. And often it is what has most recently changed that is most keenly of interest.

Change detection is important in other imagery too. Using pictures of the sky taken from ground-based telescopes, changes can indicate asteroids, flare stars, potentially even new astronomical phenomena. In practice, these changes usually in-

Appearing in Workshop on Machine Learning Algorithms for Surveillance and Event Detection at the *23rd International Conference on Machine Learning* (ICML), 2006.

dicate terrestrial phenomena and image artifacts that need to be filtered out (Vestrand et al., 2004).

While changes of a known character (is there a new star in the sky? has the marijuana been harvested yet?) might best be detected with manually written code, there is considerable interest in "anomalous change detection." The image analyst wants to know if anything *interesting* has happened in the scene. If the green fields have all turned brown, then that is a real change, but (particularly if it is observed over a sizeable fraction of the images) it is not necessarily an interesting one. We do not have a good mathematical definition of "interesting," but changes that are rare or unusual are good candidates. It therefore makes sense to bring the tools of anomaly detection to bear on this aspect of the change detection problem.

The anomalous change detection problem is illustrated in Fig. 1; we show two images – both artificially derived from the same initial IKONOS (Space Imaging, Inc., 1999) scene – for which there are both broad overall differences and one small anomalous difference. The aim is to identify this anomalous change.

There are two distinctions to make here. One is that anomalous change is different from (and rarer than) changes that occur over the whole scene; two is that the anomaly of interest is an anomalous *change*, not an anomaly of the scene itself.

To make these distinctions, we will exploit the perspective of anomaly detection as a classification problem. While this is not a new perspective – *e.g.,* check out Fig. 14.3 in the book by Hastie et al. (2001) – the classification framework has recently been put on rigorous footing (Steinwart et al., 2005). In this framework, an explicit "anomaly" or "background" distribution is identified, and the problem is treated as a binary classification with one class specified by the data and the other by this explicit distribution. The choice of this background class is nominally flexible, but in practice a flat Lebesgue measure is nearly always used. For the anomalous change detection problem, we propose to employ a non-flat background distribution that is specified *in terms of* the distribution of the data. Sampling from this background distribution can be performed with a kind of scrambled resampling of the original data.

2. Proposed Framework

A tenet of machine learning is that the distributions of interest are never known, only sampled from. To describe our proposal for anomalous change detection, however, we will start with the simpler scenario of known distribution. Section 3 discusses the more realistic situation when only the data samples are available.

Consider (x, y) pairs with $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$. Here, x is the value of a pixel in the first image, and y is the value of the same pixel in the second image. The quantities d_x and d_y represent the number of spectral channels in the images. If both images are taken with the same sensor, then $d_x = d_y$, but that is not necessary for the general formulation of the anomalous change detection problem. We write $P(x, y)$ as the probability distribution from which data samples (x_i, y_i) are drawn. Our goal is to find a function $f(x, y)$ with the property that $f(x, y) > 0$ identifies (x, y) as an anomalous change. Usually, we define a family of functions $f_\alpha(x, y)$, which are constrained to have the property that

$$\int \mathcal{I}\{f_\alpha(x, y) > 0\}\, P(x, y)\, dx\, dy \leq \alpha. \qquad (1)$$

where $\mathcal{I}$ is the indicator function; it is one if its argument is true, and zero otherwise. This constraint ensures that $f_\alpha(x, y)$ will, on average, incorrectly identify no more than a fraction α of the normal data samples as anomalies. Thus, α bounds the false alarm rate. In the language of hypothesis testing, α is the "size" of the detector.

Subject to the constraint in Eq. 1, the best anomaly detectors are those that find the most anomalies. But if we don't know what anomalies are, how can we measure this? The traditional approach has been to effectively posit a uniform background of anomalies.

This leads to a minimum volume formulation for anomaly detection; the aim is to find a function $f_\alpha(x, y)$ which satisfies Eq. 1 and that minimizes

the volume

$$V = \int \mathcal{I}\{f_\alpha(x,y) < 0\}\, dx\, dy. \tag{2}$$

For smooth $P(x,y)$, it's not hard to show that when the volume is minimized, the $f(x,y) = 0$ boundary occurs at a contour of constant density $P(x,y) = \rho$. Anomalies are interpreted as data that occur in low-density areas of phase-space, and where the density is low (specifically $P(x,y) < \rho$), we will have $f(x,y) > 0$, and the sample (x,y) will be considered anomalous. The detection of density level curves from finite samples was addressed in a PAC context by Ben-David and Lindenbaum (1997).

A more general approach, however, is to define a "background distribution" $Q(x,y)$ and minimize the volume with respect to this distribution:

$$V_Q = \int \mathcal{I}\{f_\alpha(x,y) < 0\}\, Q(x,y)\, dx\, dy. \tag{3}$$

A number of authors have remarked that $Q(x,y)$ need not be a flat distribution, but few specific suggestions for non-flat distributions have been made. In this paper, we argue that a nonuniform $Q(x,y)$ can be used to tease out unusual *changes* in an image, without the confound of otherwise unusual *pixels*. Specifically, we propose the product $Q(x,y) = P(x)P(y)$, with

$$P(x) = \int P(x,y)\, dy, \tag{4}$$

$$P(y) = \int P(x,y)\, dx. \tag{5}$$

This is the distribution that would be exhibited by the data if x and y were independent. By using this as the background, we obtain anomalies whose (x,y) dependency is unusual compared to the dependency encoded in $P(x,y)$. The anomaly detector we propose consists of level curves of the ratio

$$\frac{P(x,y)}{P(x)P(y)}. \tag{6}$$

This has the flavor of anomaly detection as contours of $P(x,y)$, but in this case we consider $P(x,y)$ *with respect to* $Q(x,y) = P(x)P(y)$, rather than an implicit choice of $Q(x,y) = 1$, corresponding to a uniform background.

2.1. Gaussian Distributions

To make the anomalous change detection framework a little more concrete, and to connect the approach with some existing methods that have proved successful for hyperspectral data, we will consider the case in which the measured data are drawn from Gaussian distributions. We emphasize that the proposed framework does not in any way depend on an assumption of Gaussianity.

We presume that the points (x,y) are drawn from a $d_x + d_y$ dimensional Gaussian distribution with mean zero[1] and covariance given by the matrix

$$\begin{bmatrix} X & C^T \\ C & Y \end{bmatrix} \tag{7}$$

where superscript T indicates transpose, and

$$X = \left\langle xx^T \right\rangle, \tag{8}$$

$$Y = \left\langle yy^T \right\rangle, \tag{9}$$

$$C = \left\langle yx^T \right\rangle. \tag{10}$$

We can "derive" these terms from a simple model. Suppose $s \in \mathbb{R}^d$ represents the spectrum at a location on the ground, and we will assume s is drawn from a Gaussian distribution with mean zero and covariance K_s. When we observe this ground location, by taking a picture of it, from an airborne camera say, we obtain the pixel value x, which we assume to have two components: one due to s and one due to intrinsic variability (*e.g.,* noise in the sensor). We write $x = L_x s + n_x$, where L_x is a $d \times d_x$ matrix that encapsulates the (assumed to be) linear relationship between what is on the ground and what is seen in the image, and the noise n_x has a Gaussian distribution with mean zero and covariance N_x. At some time later, a possibly different camera takes an image of the scene s and produces at that location a pixel whose value y is given by $y = L_y s + n_y$. We can write for this model that

$$\begin{aligned} X &= \left\langle xx^T \right\rangle \\ &= \left\langle (L_x s + n_x)(L_x s + n_x)^T \right\rangle \\ &= L_x K_s L_x^T + N_x \end{aligned} \tag{11}$$

[1] In practice, we subtract the average pixel value from each pixel to produce an image with zero mean.

where the cross-terms that involve averages of the products of s and n_x vanish, since the scene s is assumed to uncorrelated to noise n_x. Similarly,

$$Y = L_y K_s L_y^T + N_y, \quad (12)$$

$$C = L_y K_s L_x^T. \quad (13)$$

In this simple model, all transformations are linear and all distributions are Gaussian. These assumptions are not entirely realistic (and, again, are not at all required by the proposed framework), but the linear Gaussian picture provides a simple illustration of the basic ideas behind different approaches for anomalous change detection.

The normal points are drawn from the distribution $P(x, y)$, which is Gaussian with mean zero and covariance given by the matrix in Eq. 7. The goal is to identify boundaries in the combined (x, y) space (that is, in $\mathbb{R}^{d_x+d_y}$) which enclose most of these normal points, but outside of which anomalous changes will be found. In what follows, we will describe several approaches for drawing these boundaries.

2.1.1. Straight Anomaly Detection

This approach treats the anomalous change detection problem as one of standard anomaly detection in the combined (x, y) space. Contours of constant $P(x, y)$ are optimal in the sense of enclosing the maximum fraction of the normal data with the minimum volume. Fig. 2(a) illustrates this anomaly detector for a simple gaussian case.

We have not seen this approach advocated for the change detection problem, and for good reason. Straight anomaly detection does not adequately concentrate on the *change* in going from x to y; it finds all anomalies, but we want to restrict our attention to the anomalous changes.

2.1.2. Chronochrome

Another approach that is used in hyperspectral remote sensing goes by the name "chronochrome" (Schaum & Stocker, 1998). The basic idea is to fit a linear predictor[2] for y given x. In particular,

$$y \approx \hat{y} = Lx. \quad (14)$$

Here, L is fit using least-squares. Defining an error matrix

$$E = \left\langle (y - Lx)(y - Lx)^T \right\rangle, \quad (15)$$

one can show that the trace of E is minimized when $L = CX^{-1}$. The essence of the change detection algorithm is to find changes between y and $\hat{y} = Lx$. Write

$$\varepsilon = y - CX^{-1}x \quad (16)$$

and compute the covariance $\left\langle \varepsilon\varepsilon^T \right\rangle = Y - CX^{-1}C^T$. The largest changes ε, in the sense of Mahalanobis distance, are given by the largest values of $\varepsilon^T \left\langle \varepsilon\varepsilon^T \right\rangle \varepsilon$, or

$$(y - CX^{-1}x)^T (Y - CX^{-1}C^T)^{-1} (y - CX^{-1}x). \quad (17)$$

The comparison of this quantity to an appropriate threshold defines the chronochrome anomaly detector. The boundary that separates anomalous changes from normal is an elliptical cylinder centered on the axis defined by $y = CX^{-1}x$.

An idiosyncracy of the chronochrome is that it is asymmetric with respect to x and y. If one instead looks for a linear map such that $x \approx \hat{x} = L'y$, and minimizes $|x - L'y|^2$, then one obtains a different chronochrome detector: namely,

$$(x - C^T Y^{-1} y)^T (X - C^T Y^{-1} C)^{-1} (x - C^T Y^{-1} y), \quad (18)$$

which leads to a different cylinder, whose axis is specified by $x = C^T Y^{-1} y$.

Fig. 2(b) shows both the standard chronochrome and the chronochrome that is obtained when the role of x and y are swapped.

[2] For the more general non-Gaussian case, one may prefer to fit a nonlinear predictor; *e.g.,* Clifton (2003) uses neural networks to learn a nonlinear relationship betweeen x and y and then argues that "substantial differences between the expected and actual values represent an unusual change."

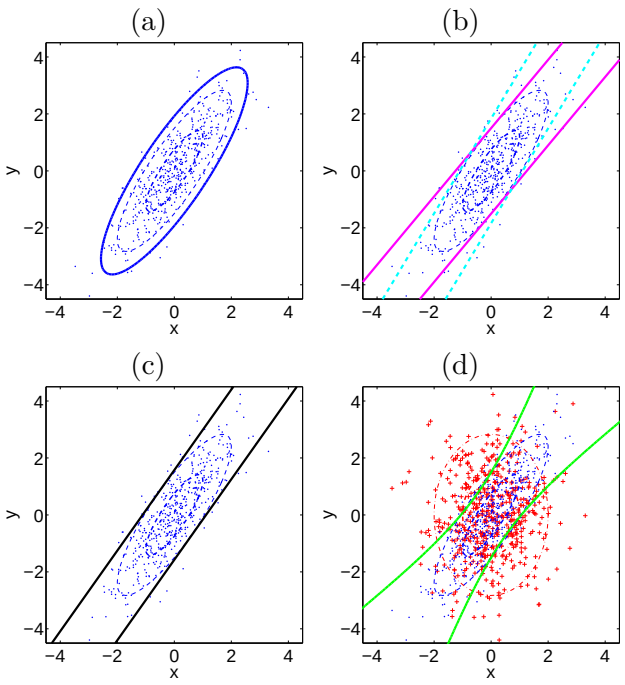

Figure 2. Experiment illustrating boundaries for different anomaly detectors for the simple case of Gaussian distributions, linear relationships, and $d_x = d_y = 1$. The data are shown as $N = 500$ points sampled from the normal distribution $P(x, y)$, parameterized by the covariance matrix in Eq. 7 with $X = 1$, $Y = 2$, and $C = 1.2$. In these scatterplots, each point corresponds to a pixel in a (hypothetical) scene, with x corresponding to the value at that pixel in the first image, and y is the value at the same pixel in the second image. **(a)** The dashed elliptical lines correspond to one and two sigmas. The heavy solid line corresponds to an elliptical boundary that encloses 95% of the normal data. Points inside the boundary would be identified as normal, and points outside as anomalous. **(b)** Using the same data, two variants of the chronochrome are shown. The solid lines correspond to the standard variant, based on minimizing $|y - Lx|^2$; the heavy dashed lines illustrate the variant based on minimizing $|x - L'y|^2$. That the solid and dashed lines do not conicide illustrates the asymmetry between x and y in the chronochrome. Points between the parallel lines are classified as normal changes; points outside are anomalous changes. **(c)** The covariance equalization detector is shown as parallel lines that are parallel to the long axis of the ellipse. Among all anomaly detectors based on parallel lines (in this example), this is the detector with the smallest distance between the lines. **(d)** The solid hyperbolic lines show the anomaly detector based on the framework proposed here. Also shown is the "background" distribution $P(x)P(y)$, shown with dash-dotted ellipses for the one and two sigma contours, and with a sample of $N = 500$ points shown as plus signs. The hyperbolic lines optimally separate the normal (dots) from the background (pluses) data, but still constrained to enclose 95% of the normal data.

2.1.3. Covariance Equalization

More recently, Schaum and Stocker (2004) introduced a "covariance equalization" approach as an approximation to the chronochrome that avoids directly computing C. In this approximation, the data is modelled by $y \approx \hat{y} = L^* x$, where now L^* is given by the expression $L^* = Y^{1/2} R X^{-1/2}$; this does not involve C, but does employ an orthonormal matrix R that is not formally specified. For convenience, and because it works well in practice (for the $d_x = d_y$ case), Schaum and Stocker (2004) recommend $R = I$; that is: $L = Y^{1/2} X^{-1/2}$. Although this is introduced as an approximation to the "optimal" chronochrome, it has a nice property, not shared by the chronochrome, of being symmetrical with respect to x and y. It can also be implemented by separately whitening the data in the x and y images, and then doing a simple subtraction. That is,

$$\varepsilon = X^{-1/2} x - Y^{-1/2} y \tag{19}$$

and the anomalies are pixels for which $\varepsilon^T \left\langle \varepsilon \varepsilon^T \right\rangle \varepsilon$ is large. See Fig. 2(c).

2.1.4. Proposed Framework

Finally, we describe how the proposed anomalous change detection framework would be applied to the Gaussian model. Here, $P(x, y)$ is Gaussian with covariance given by Eq. 7, and $P(x)$ and $P(y)$ are Gaussian with covariances given by X and Y, respectively. In this framework, anomalous changes are given by level curves of Eq. 6; up to translation and scale factors, the negative logarithm of Eq. 6 is the quadratic function

$$\begin{bmatrix} x^T & y^T \end{bmatrix} K \begin{bmatrix} x \\ y \end{bmatrix} \tag{20}$$

where the matrix K is given by

$$K = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}^{-1}. \tag{21}$$

Fig. 2(d) illustrates what these boundaries would be for the simple case of single-band images ($d_x = d_y = 1$) with $X = 1$, $Y = 2$, and $C = 1.2$. Note that K is not a positive-definite matrix; so the contours in Eq. 20 are hyperbolic, not elliptical.

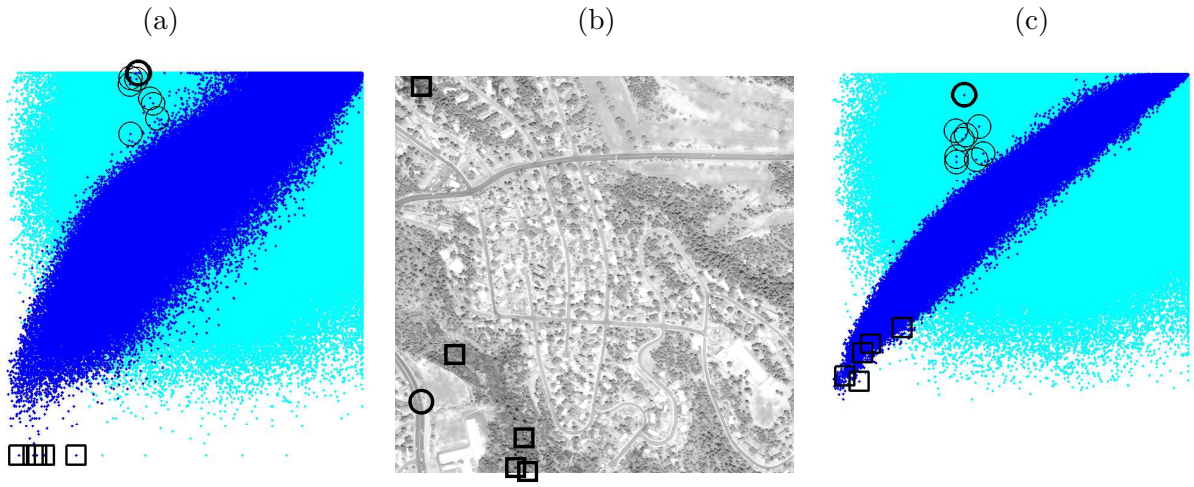

Figure 3. **(a)** Scatterplot of x versus y for the image data, with x corresponding to Fig. 1(a), and y corresponding to Fig. 1(b). The dark points correspond to the actual data, and the light points (cyan in the color version of this plot) corresponding to the resampled data. The circled points correspond to the nine pixels that were artificially modified; the darkest circle is the center of the 3×3 modification. The squares indicate points that might be identified as anomalies, or outliers, in the distribution of actual data, but which do not represent particularly anomalous changes. **(a)** The image from Fig. 1(b), indicating with a circle the pixels that have actually been altered, and with squares the five "outlier" pixels identified in panel (a). These are the darkest pixels in the scene and appear as shadows in some of the wooded areas. **(c)** Similar to (a), but obtained after smoothing the images in panel (b). The effect of smoothing is to make the images more "alike" (so the dark diagonal band is narrower) and thereby to make the most anomalous changes more evident.

3. Scrambled resampling approach

In practice, we do not know the underlying distribution $P(x, y)$, but instead have a set of samples $\{(x_1, y_1), (x_2, y_2), \ldots\}$ which we presume to be drawn from this distribution. Our aim is to find anomalous pixel pairs among this set.

A resampling approach for anomaly detection in multispectral imagery was introduced previously (Theiler & Cai, 2003). The idea is extended here to the anomalous change detection problem.

Our approach is to identify a normal and a background class, and to produce samples from each class. The normal class is specified by the original data: $\{(x_1, y_1), (x_2, y_2), \ldots\}$. The background class is specified by an artificially generated dataset that is obtained from a resampling of the original data in which the alignment of x and y is shuffled. The resampling treats x and y independently, and produces a dataset that looks like $\{(x_i, y_j), \ldots\}$, where i and j are usually unequal. This resampled dataset corresponds to samples that could have been drawn from the distribution $Q(x, y) = P(x)P(y)$.

This is illustrated in Fig. 3(a,b) for the image example shown at the beginning of this paper.

With samples from two classes, we can use virtually any standard binary classification scheme to train a discriminating function $f(x, y)$ so that $f(x, y) > 0$ implies (x, y) is in the anomaly class. Pixels i for which $f(x_i, y_i)$ is numerically large are identified as the most anomalous changes.

We remark that background data, generated by random sampling from an underlying distribution, was used in the DLD-SVM (Density Level Detection Support Vector Machine) algorithm of Steinwart et al. (2005). A similar sampling was employed by Tax and Duin (2002) to distinguish solutions obtained from different parameters of the SVDD (Support Vector Domain Description) algorithm, introduced in Tax and Duin (1999). But in both of these cases, the sampling was from a flat distribution.

4. Another example: looking at the sky

The images of star fields in Fig. 4(a,b) are taken roughly a minute apart; finding changes at this

short timescale is a very challenging problem, but is just becoming feasable (Vestrand et al., 2004). Naive change detection would seek those pixels that had most changed between the two images; that is, one computes a direct difference of the images and identifies the pixels with the largest absolute values. The top 100 such pixels are identified in Fig. 4(c). Closer inspection of these locations indicates a slight misregistration between the two images. Because of a small rotational error between the two images, the largest differences occur at the locations of stars away from the center of the image. An alternative approach follows the framework suggested here: two scatterplots are generated, one for the data and one for the resampled data, and binary classification is used to distinguish the two datasets. For practical reasons, with this example, we simplified the problem by only considering data for which a brightening occurred: $y_i > x_i$. A support vector machine with a radial basis kernel (Chang & Lin, 2001) was employed, and a threshold was adjusted to produce the 100 "most anomalous" pixels. As shown in Fig. 4(d), these anomalous changes are associated with a satellite streak.

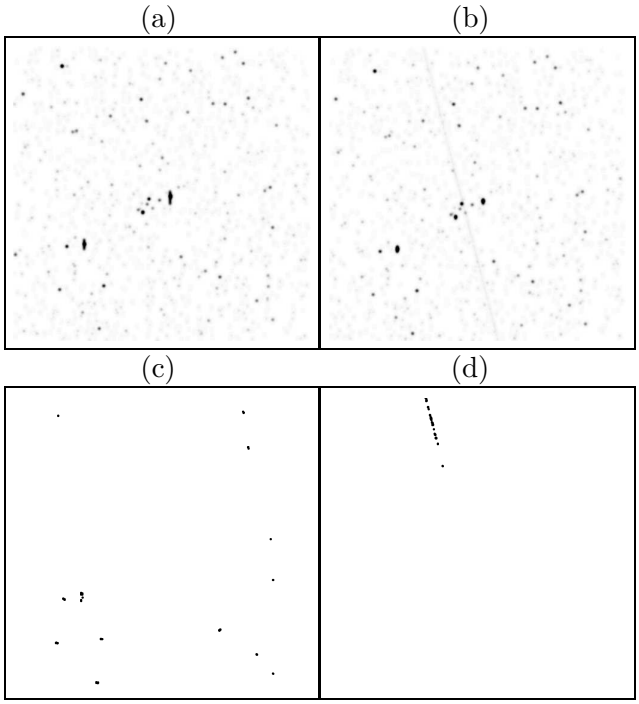

Figure 4. **(a,b)** Two images of a starfield taken several seconds apart. To help with the visibility on the page, we have inverted the colors (so black spots correspond to bright stars), and smoothed the image. Although in this case, the seeing is nearly identical in the two images, the registration isn't exact. More sophisticated registration can be done, but we were partly interested in whether the resampling approach could overcome the effects of misregistration. **(c,d)** Locations of the 100 most anomalous pixels. Panel **(c)** indicates the 100 pixels for with the largest absolute difference in brightness; panel **(d)** is based on a binary classification, using a support vector machine with a gaussian kernel, between the "normal" data class and the resampled "anomaly" class.

It is tempting to say that the satellite streak is more interesting than the image misregistration (and therefore the proposed framework produces "better" anomalies), but in both cases the anomalies point to aspects of the image analysis that need to be addressed. In a production setting, one will need to include better image co-registration *and* explicit testing for airplane, meteor, and satellite streaks. When this is done, anomalous change detection can be applied anew, and it will undoubtedly find new artifacts that need to be identified and filtered out. This is, and Sisyphus would agree, valuable progress.

5. Some image-specific issues

The proposed formulation does not explicitly require that the data be organized as pixels in images, but imagery provides both a natural setting and a compelling application for the change detection problem.

As illustrated by the star field example above, co-registration of the two images is important. In practice, the co-registration is never exact, but the effect of small misregistration can often be ameliorated by some smoothing of the images. Smoothing can also be useful if the anomalies of interest are of a characteristic size – the effect will be to supress anomalous changes whose spatial extent is smaller than the smoothing kernel.

Smoothing is just one example of a spatio-spectral operator that can be applied to modify imagery. To the extent that these global operators (by global, we refer to the application of the operator over the whole image) can lead to pairs of images that are more nearly identical, that

seems likely the improve the ability to identify small local anomalous changes. This effect is illustrated in Fig. 3(c). We can imagine a scenario in which spatio-spectral operators are systematically applied to the images with the goal of producing more nearly "equalized" images. We have previously investigated the use of a genetic algorithm for exploring these operators in regression (Theiler et al., 1999) and classification tasks (Harvey et al., 2002), but feature extraction for anomaly detection is a new challenge.

A practical difficulty that arises with star fields is that the statistics are dominated by the large dark areas between the stars. This can bog down the numerics for one thing. But it can also lead to a large dynamic range in the density ratio $P(x, y)/P(x)P(y)$ in the areas of the image that are of least interest.

Acknowledgements

We are grateful to the anonymous reviewers for useful comments which led to improvements in this manuscript. This work was supported by the Los Alamos Laboratory Directed Research and Development (LDRD) program.

References

Ben-David, S., & Lindenbaum, M. (1997). Learning distributions by their density levels: A paradigm for learning without a teacher. *J. Computer and System Sciences*, *55*, 171–182.

Chang, C.-C., & Lin, C.-J. (2001). LIBSVM: a library for support vector machines. `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Clifton, C. (2003). Change detection in overhead imagery using neural networks. *Applied Intelligence*, *18*, 215–234.

Harvey, N. R., Theiler, J., Brumby, S. P., Perkins, S., Szymanski, J. J., Bloch, J. J., Porter, R. B., Galassi, M., & Young, A. C. (2002). Comparison of GENIE and conventional supervised classifiers for multispectral image feature extraction. *IEEE Trans. Geosci. and Remote Sens.*, *40*, 393–404.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *Elements of statistical learning: Data mining, inference, and prediction.* New York: Springer-Verlag.

Radke, R. J., Andra, S., Al-Kofahi, O., & Roysam, B. (2005). Image change detection algorithms: A systematic survey. *IEEE Trans. Image Processing*, *14*, 294–307.

Schaum, A., & Stocker, A. (1998). Long-interval chronochrome target detection. *Proc. 1997 International Symposium on Spectral Sensing Research.*

Schaum, A., & Stocker, A. (2004). Hyperspectral change detection and supervised matched filtering based on covariance equalization. *Proc. SPIE*, *5425*, 77–90.

Space Imaging, Inc. (1999). `http://www.spaceimaging.com/products/ikonos`.

Steinwart, I., Hush, D., & Scovel, C. (2005). Density level detection is classification. In L. K. Saul, Y. Weiss and L. Bottou (Eds.), *Advances in neural information processing systems 17*, 1337–1344. Cambridge, MA: MIT Press.

Tax, D., & Duin, R. (1999). Data domain description by support vectors. *Proc. ESANN99* (pp. 251–256). Brussels: D. Facto Press.

Tax, D., & Duin, R. (2002). Uniform object generation for optimizing one-class classifiers. *J. Machine Learning Res.*, *2*, 155–173.

Theiler, J., & Cai, D. M. (2003). Resampling approach for anomaly detection in multispectral images. *Proc. SPIE*, *5093*, 230–240.

Theiler, J., Harvey, N. R., Brumby, S. P., Szymanski, J. J., Alferink, S., Perkins, S. J., Porter, R. B., & Bloch, J. J. (1999). Evolving retrieval algorithms with a genetic programming scheme. *Proc. SPIE*, *3753*, 416–425.

Vestrand, W. T., Theiler, J., & Wozniak, P. R. (2004). Unsolved problems in observational astronomy II. Focus on rapid response - mining the sky with "thinking" telescopes. *Astronomische Nachrichten*, *325*, 477–482.