An Overview of Background Modeling for Detection of Targets and Anomalies in Hyperspectral Remotely Sensed Imagery

Stefania Matteoli, *Member, IEEE,* Marco Diani, *Member, IEEE,* and James Theiler

***Abstract*—This paper reviews well-known classic algorithms and more recent experimental approaches for distinguishing the weak signal of a target (either known or anomalous) from the cluttered background of a hyperspectral image. Making this distinction requires characterization of the targets and characterization of the backgrounds, and our emphasis in this review is on the backgrounds. We describe a variety of background modeling strategies—Gaussian and non-Gaussian, global and local, generative and discriminative, parametric and nonparametric, spectral and spatio-spectral—in the context of how they relate to the target and anomaly detection problems. We discuss the major issues addressed by these algorithms, and some of the tradeoffs made in choosing an effective algorithm for a given detection application. We identify connections among these algorithms and point out directions where innovative modeling strategies may be developed into detection algorithms that are more sensitive and reliable.**

***Index Terms*—Anomaly detection, background estimation, background modeling, hyperspectral imagery, target detection.**

I. Introduction

COMBINING imaging and high spectral resolution spectroscopy in a unique system, hyperspectral sensors provide a powerful means to discriminate the materials in a scene on the basis of their spectral signature. With the advances in hyperspectral sensing, the last two decades have seen a growing employment of hyperspectral target and anomaly detection in several different fields of real-life applications [30], [60], [61], [69], [78], [99], [100]. Methods to search the image for small rare objects/materials that have a specific spectral signature (target detection) or that are spectrally different from the surrounding background (anomaly detection) can be useful in environmental monitoring, to locate and identify hazardous materials; in defense applications, for mine detection and surveillance; and in public safety, for airborne search and rescue operations.

To detect small targets or anomalies in hyperspectral imagery, one of the key challenges is to characterize the background. Indeed, many detection algorithms proceed by estimating what the target-free background signal would be at a given pixel, then subtracting this background estimate from what is observed at that pixel, and finally performing further processing on the residual [60], [61], [63], [69], [78], [99]. Background data in hyperspectral imagery may be extremely cluttered, highly spatially variable, and intrinsically high-dimensional; and modeling this background is a rich and interesting task, in which more accurate models are rewarded with better detection performance. Classical approaches to background modeling, such as the multivariate normal (MVN) and the subspace models, have led to a variety of detection algorithms widely employed in the hyperspectral detection literature. The considerable mismatch that often occurs between classical, simple, easy to handle models and the complicated nature of real hyperspectral backgrounds has driven research efforts toward the development of more complex models capable of strengthening the weak links in the background modeling chain. On the one hand, many attempts have been pursued to make classical models more robust to specific limitations, while preserving their desirable properties. On the other hand, more recent research trends have involved tailored multimodal distributions, geometrical approaches, and nonlinear high dimensional spaces with the aim of providing increased adaptability to complex hyperspectral backgrounds.

In this paper, an overview of the target detection problem from the background characterization perspective is provided. In contrast to reviews that specifically focus on the detection algorithms [60], [61], [78], here the emphasis is on characterizing the background. Conventional background models are reviewed together with their main limitations and more complex background models specifically adopted to handle cluttered, spatially variable, nonhomogeneous backgrounds are surveyed.

The outline of the paper is as follows. First, conventional widely employed background models and their associated detection algorithms are reviewed in Section II, along with their major limitations. Section III provides an extensive overview of both well-known and more recent background models specifically adopted to handle some of these limitations. Finally, Section IV summarizes the conclusions of the work.

Manuscript received February 28, 2013; revised March 21, 2014; accepted March 28, 2014. Date of publication May 04, 2014; date of current version August 01, 2014. The work of S. Matteoli and M. Diani was supported by the Italian Ministry of University and Research (MIUR) under the framework of the PRIN project "Very High Spatial and Spectral Resolution Remote Sensing: A Novel Integrated Data Analysis System." The work of J. Theiler was supported by the United States Department of Energy's Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory.

S. Matteoli and M. Diani are with the Information Engineering Department, University of Pisa, Pisa 56126, Italy (e-mail: stefania.matteoli@iet.unipi.it; m.diani@iet.unipi.it).

J. Theiler is with Los Alamos National Laboratory, Los Alamos, NM 87545 USA (e-mail: jt@lanl.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2014.2315772

II. Classical Background Models and Related Anomaly and Target Detection Algorithms

In this section, we review classical background models that underlie detection algorithms widely employed in the literature. To this end, some basics of target detection and detector design are briefly reviewed in Section II-A. Then, Section II-B describes two specific background models: 1) the MVN model; and 2) the subspace model. Finally, some of the limitations of these background models are highlighted and discussed in Section II-C.

A. Anomaly and Target Detection Basics

As an organizing principle, target detection is treated as a binary hypothesis testing problem [43] that is performed at every pixel in the image. Let H_0 represent the null hypothesis that there is not a target at the pixel under test, and H_1 the alternative hypothesis that a target is present. Further, let $\boldsymbol{x}$ denote the d-dimensional vector corresponding to the hyperspectral measure of the pixel under test, where d is the number of spectral channels. We write

$$D(\boldsymbol{x}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{1}$$

to indicate that the detection statistic $D(\boldsymbol{x})$ is compared to a threshold η, with $D(\boldsymbol{x}) < \eta$ indicating the null hypothesis of no target, and $D(\boldsymbol{x}) > \eta$ indicating the alternative hypothesis that a target is present. The detection statistic $D(\boldsymbol{x})$ is often derived by exploiting a number N of background pixels, which are assumed to be free of the target signal. These background pixels may be characterized by means of a spectral variability model, for instance, by computing their second-order statistics (e.g., mean vector $\boldsymbol{\mu}$, correlation matrix $\boldsymbol{R}$, covariance matrix $\boldsymbol{C}$) or by evaluating the background probability density function (PDF) $p_b(\boldsymbol{x})$ in $\boldsymbol{x}$.

The background data can include all image pixels (i.e., *global* background) or just the pixels in the spatial neighborhood of the pixel under test (*local* background), thus leading to *global* and *local* detection algorithms, respectively. In the latter case, an annulus of pixels surrounding the pixel under test (see Fig. 1) is taken to characterize local background. For the local detector, choosing annulus size is a compromise between statistical precision in the estimates of the background properties (e.g., mean vector and correlation/covariance matrix), which favors a large annulus; and spatial nonstationarity of the background, which favors a small annulus. It can be effective to estimate the mean vector with a smaller annulus using the pixels physically close to the tested one, whereas the correlation/covariance matrix is estimated by using a larger annulus, because more pixels are required to provide a reliable estimate [2], [69]. This is consistent with the nonstationary model suggested in [41], which models an optical image as a nonstationary multivariate random process with a quickly spatially varying mean vector and a more slowly varying covariance matrix. Also shown in Fig. 1 is a guard window between the annulus and the pixel under test; this excludes from local background characterization the pixels that may contain the target signal, due to their closest proximity to the pixel under test [69], [78].

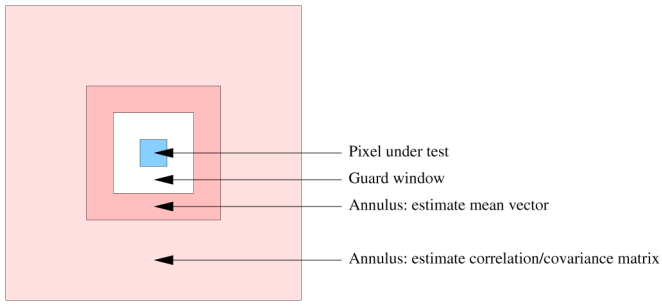


Fig. 1. Graphical representation of the annulus employed by local algorithms. The guard window is employed to avoid target pixels being included in local background characterization. Generally (see also Section II-B), a smaller annulus is used to estimate local background mean vector, whereas a larger one is adopted to estimate local background correlation/covariance matrix.

For the *anomaly detection* problem [69], [78], [99], we seek pixels in the image that are unusual compared to the rest of the pixels (i.e., the background). These pixels generally correspond to small rare objects that occupy a very small fraction of the area of the background in which they are embedded.

When a reference spectral signature of a given target of interest is available, the *target detection* problem can be addressed by searching the image for all pixels that contain the given target [60], [61]. This is generally performed by seeking the pixels whose spectrum shows a high degree of correlation with the target spectrum [37], [44], [47], [60], [61], [83].

In general, a model that provides a background PDF $p_b(\boldsymbol{x})$ will lead to an anomaly detector that is of the form of $D(\boldsymbol{x}) = g(1/p_b(\boldsymbol{x}))$, where $g(\cdot)$ is a monotonic function of its argument. One widely employed choice for $g(\cdot)$ is the logarithmic function, thus obtaining the so-called *background log-likelihood*

$$D(\boldsymbol{x}) = -log[p_b(\boldsymbol{x})]. \tag{2}$$

To derive a target detector based on $p_b(\boldsymbol{x})$, one needs a model for the target. One particularly simple but widely used example is the additive model, which posits that a target pixel will be of the form $\boldsymbol{x} + \gamma \boldsymbol{t}$, where $\boldsymbol{t}$ is the known target signature and γ an unknown scalar value (corresponding to the target strength or "abundance") [60]. One may then combine this background and target model by invoking the Generalized Likelihood Ratio Test (GLRT) to obtain the detector

$$D(\boldsymbol{x}) = g\left(\frac{\max_\gamma\{p_b(\boldsymbol{x} - \gamma \boldsymbol{t})\}}{p_b(\boldsymbol{x})}\right). \tag{3}$$

It is also possible to derive this kind of target detector with a more general approach than the GLRT [91].

B. Conventional Approaches to Background Modeling

As is evident from the general equations (2) and (3), both for the detection of anomalies and for specific targets, it crucially matters what model is adopted to characterize the background, and how that model is estimated. Different background models lead to different anomaly and target detection algorithms. In the following, we briefly outline two of the simplest and most widely used models for the background, namely the MVN model and the subspace model.

1) The MVN Model: The assumption that the background can be modeled with an MVN distribution has led to many useful detection algorithms [60], [69], and provides a starting point for many others (see Section III-A). Indeed, the mathematical tractability of the MVN has allowed straightforward derivation of detection algorithms as well as simple analytical evaluation of theoretical detection performance. Even for data that are not Gaussian, the MVN model can be effective for high dimensional data because the central limit theorem implies that many projections of the data will be approximately Gaussian. Also, the number of parameters in the MVN model scales only quadratically (not exponentially, as the curse of dimensionality would suggest, for more arbitrary distributional models) with the dimension, and this can provide a good compromise between under-fitting and over-fitting.

In the MVN background modeling approach, a multivariate Gaussian parametric model, depending on the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{C}$, is assumed for $p_b(\boldsymbol{x})$

$$p_b(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{C}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{4}$$

Among the various detection algorithms derived from the MVN model, the RX anomaly detection algorithm [81] is probably the most widely known, having become a standard against which other anomaly detectors are compared [60], [69], [78], [99]. The RX algorithm, as originally conceived by Reed and Yu [81], was born as a *local* detector. Based on the aforementioned nonstationary model for optical images [41], a local mean removal process [25], [66] is first performed by sliding an annulus (such as that shown in Fig. 1) over the image. After local mean removal, the covariance matrix (presumed to spatially vary more slowly than the mean vector) is locally estimated over the pixels enclosed within a larger annulus. Such pixels can reasonably be assumed to follow a (local) MVN distribution. The RX detector basically evaluates the Mahalanobis magnitude of the mean-subtracted pixel

$$D(\boldsymbol{x}) = \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{x}} \tag{5}$$

where $\tilde{\boldsymbol{x}}$ denotes that the local mean has been removed from $\boldsymbol{x}$ and $\tilde{\boldsymbol{R}}$ is the local background correlation matrix estimated over the mean-removed background data in the surrounding annulus. The RX detection statistic in (5) has constant false alarm rate (CFAR) even when the correlation matrix is estimated on the background data [81].

Although the name RX should strictly refer to the algorithm described above, this designation has been widely employed in the literature to indicate the Mahalanobis distance detector [60], [78], [99]

$$D(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \tag{6}$$

which evaluates the Mahalanobis distance of the pixel $\boldsymbol{x}$ from the estimated mean vector $\boldsymbol{\mu}$ of background data on the basis of the background covariance matrix estimate $\boldsymbol{C}$. The Mahalanobis distance detector has been applied both in global and local fashions.

For the additive target detection problem referred to in (3), the best detector is a matched filter (MF) [82]

$$D(\boldsymbol{x}) = \boldsymbol{t}^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \tag{7}$$

As with RX, it is possible to compute $\boldsymbol{\mu}$ and $\boldsymbol{C}$ locally, from an annulus surrounding the pixel of interest [2], [26].

Other well-known detectors have been derived following the MVN assumption and adopting the GLRT, including different variants of the MF [44], [82], [83] and the adaptive coherence estimator (ACE) detector [47].

2) Background Subspace Model: Subspace models of the background take a different point of view. Here it is assumed that the background is well modeled as lying in a low-dimensional subspace, but the background distribution within that subspace does not ultimately matter. For the target-free pixels, which are consistent with the null hypothesis, we can write

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{w} \tag{8}$$

where $\boldsymbol{B}$ defines the background subspace, $\boldsymbol{\beta}$ specifies the coefficients of the linear combination of vectors in $\boldsymbol{B}$, and $\boldsymbol{w}$ is a "lack-of-fit" noise term that should be of small magnitude if the model is accurate [58]. Although this is less prescriptive than models (such as MVN) that provide a specific $p_b(\boldsymbol{x})$, that lack of specificity provides a robustness advantage because the subspace model does not depend on details of that distribution within the subspace spanned by $\boldsymbol{B}$.

Within this subspace modeling perspective, also the target signal may be seen as lying in a subspace identified by $\boldsymbol{T}\boldsymbol{\gamma}$, where the columns of the matrix $\boldsymbol{T}$ span the subspace and $\boldsymbol{\gamma}$ is the unknown vector of coefficients identifying the direction of the target signature $\boldsymbol{t}$ within that subspace. The aforementioned additive model is a particular case in which the matrix $\boldsymbol{T}$ has a single column $\boldsymbol{t}$ and the direction of the target signature within the subspace is known. Therefore, the subspace modeling approach relies upon the following model for the hyperspectral pixel in the presence of the target

$$\boldsymbol{x} = \boldsymbol{T}\boldsymbol{\gamma} + \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{w}. \tag{9}$$

In general, the target is presumed to at least have some components that are orthogonal to the background subspace. By projecting out the background subspace and looking at the residuals onto the orthogonal subspace, the target signals may be enhanced with respect to the background. This basic idea was espoused in [37], and called orthogonal subspace projection (OSP). By building the projection matrix $\boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{B}\boldsymbol{B}^{\#})$, where $\boldsymbol{B}^{\#} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T$ is the pseudo-inverse of $\boldsymbol{B}$, so that $\boldsymbol{P}$ has the property that $\boldsymbol{P}\boldsymbol{B} = 0$, and by considering the simplified case of a target with known direction $\gamma\boldsymbol{t}$, from (9) we have that $\boldsymbol{P}\boldsymbol{x} = \gamma\boldsymbol{P}\boldsymbol{t} + \boldsymbol{P}\boldsymbol{w}$, and now the problem is separating the projected target signature from the projected noise term. In [37], where it is further assumed that $\boldsymbol{w}$ is a zero-mean Gaussian random vector with diagonal covariance matrix $\sigma^2\boldsymbol{I}_d$ (where $\boldsymbol{I}_d$ is the $d \times d$ identity matrix), this is best done with a filter that matches $\boldsymbol{P}\boldsymbol{t}$. The resulting OSP detector, derived also in [77], is

$$D(\boldsymbol{x}) = (\boldsymbol{P}\boldsymbol{t})^T\boldsymbol{P}\boldsymbol{x} = \boldsymbol{t}^T\boldsymbol{P}^T\boldsymbol{P}\boldsymbol{x} = \boldsymbol{t}^T\boldsymbol{P}\boldsymbol{x} = \boldsymbol{t}^T(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{B}^{\#})\boldsymbol{x}. \tag{10}$$

This detector simultaneously projects out the background and maximally enhances the target signal with respect to the residual background [37], [77].

OSP does not specify how $\boldsymbol{B}$ is to be chosen, and a number of approaches have been taken; an early, but simple and effective idea is to use the first few principal components of the background covariance matrix [38], [57].

The extension of this approach to targets that comprise a subspace (i.e., of possibly more than just one dimension) goes by the name of matched subspace detection [87], [88]. The matched subspace detectors may be derived by applying the GLRT with the models in (8) and (9).

The use of subspace modeling in anomaly detection is straightforward. On the basis of the anomaly-free model of (8), anomalousness is given by the magnitude of the projected vector in the residual subspace $\boldsymbol{P}\boldsymbol{x}$; i.e., $D(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}$.

In comparing the OSP detector and subspace anomaly detector with the MVN-based MF and RX anomaly detector, we see that $\boldsymbol{P}$ plays the role of $\boldsymbol{C}^{-1}$. Where the eigenvectors of $\boldsymbol{C}$ are large (i.e., in the directions where the background data have high variance), the corresponding eigenvectors of $\boldsymbol{C}^{-1}$ are small, and MVN-based detectors suppress the background in those directions. But in those directions, which are specified by $\boldsymbol{B}$, we have that $\boldsymbol{P}\boldsymbol{B} = 0$. Thus, where covariance-based detectors suppress, subspace methods project out entirely. The relationship between covariance-based and subspace-based detectors is also pointed out in [63], where it is further noted that if $\boldsymbol{B}_D$ is a matrix of the dominant (largest) eigenvectors of the covariance matrix $\boldsymbol{C}$, then $\boldsymbol{I} - \boldsymbol{B}_D\boldsymbol{B}_D^{\#}$ is the "principal component inversion approximation" of $\boldsymbol{C}^{-1}$.

An example of a hybrid detector that employs both schemes is the sub-space RX (SSRX) anomaly detector [92]. SSRX projects out the first few principal components, computes the covariance matrix on the remaining components, and employs RX on those components. A number of hybrid subspace/MVN algorithms are described in [22]; among these, the target constrained interference minimized filter (TCIMF) projects out directions defined by a matrix $\boldsymbol{B}$, similar to the OSP algorithm described above, but then employs an adaptive MF on the residuals.

Finally, it is worth mentioning that if non-negativity and sum-to-one physics-based constraints are imposed over the elements of the vector $\boldsymbol{\beta}$ in (8) and (9), the subspace model is called linear mixing model (LMM) [57] and the columns of the matrix $\boldsymbol{B}$ assume the physical meaning of *endmembers*.

C. Mismatch of Simple Models and Complex Backgrounds

Although there are many issues that affect the detection outcome (such as the spectral variability of real-world target spectral signatures, mismatch between library and in-scene spectral signatures, inaccuracies in the calibration and atmospheric compensation procedures, and inaccurate noise modeling [3], [60], [61]), one of the main causes that limit detection performance is inadequate background characterization, especially as regards detection of anomalies. A complex (and inadequately characterized) background provides a rich and unpredictable source of false alarms and affects target *detectability* as well. Background modeling is nontrivial. A good model needs to manage both background spatial variability (e.g., local, global, and mixture-based models) and spectral variability (e.g., probabilistic covariance-based models, or subspace-based models). Although conventional background models may seem effective on theoretical grounds, they cannot adapt to all the diverse situations that can occur in reality. This section will provide a brief snapshot of the major aspects that keep bringing further challenges in background modeling.

1) Non-Gaussian Behavior of the Background: Many conventional detection algorithms have assumed the background data arise from an MVN distribution. However, in real hyperspectral data, truly Gaussian behavior rarely occurs [21], [65], [74], [78]. This is especially the case for global detection algorithms, where the MVN model is adopted to explain the spectral variability of complex backgrounds encompassing several different materials and objects. A graphical example is provided in Fig. 2(a) and (b). Fig. 2(a) displays a true-color image of an urban area. The scene includes different types of terrain, water, roads, and buildings. Fig. 2(b) shows a scatterplot of the data in a simplified two-dimensional domain along two spectral channels. It is visually evident that the data distribution is far from Gaussian.

In fact, the MVN model has more effectively been employed to characterize background pixels in a (possibly) homogeneous local neighborhood around the test pixel within the framework of local algorithms. However, if the local neighborhood includes pixels belonging to different objects and materials, the local background is likely to follow a multi-modal non-Gaussian distribution. This situation is exemplified in Fig. 2(c) and (d).

Attempts to enforce the Gaussian assumption may be obtained by performing a local mean removal procedure (as in the RX algorithm as proposed by Reed and Yu [81]). In this case, local background pixels after local mean removal tend to follow a distribution that is closer to Gaussian and (mostly) mono-modal, as exemplified in Fig. 2(e) and (f).

Although MVN model often provides an incomplete representation of hyperspectral backgrounds, there are a number of algorithms that still use the Gaussian assumption and try to improve MVN model effectiveness. These algorithms are described in Section III-A. More recent research trends have definitely dropped the Gaussian assumption in favor of more complex multivariate models. These algorithms based on non-MVN models are described in Section III-B.

2) Contamination Due to the Target Signal and Outliers: Most detection algorithms require background second-order statistics (i.e., mean vector and covariance matrix) to be estimated by a set of data having the same distribution as the background. In operational situations, it may happen that this set of data is "contaminated" by the presence of pixels that strongly deviate from the background distribution such as outliers or pixels containing the target signal. This contamination leads to the corruption of the estimated second-order statistics. The covariance matrix is particularly susceptible to corruption due to contamination and this is generally coupled with a general degradation of the detection performance [53], [69], [71], [72], [98], [107]. Contamination effects may be particularly deleterious if the contaminating pixels contain the target signal, because the background characterization procedure is

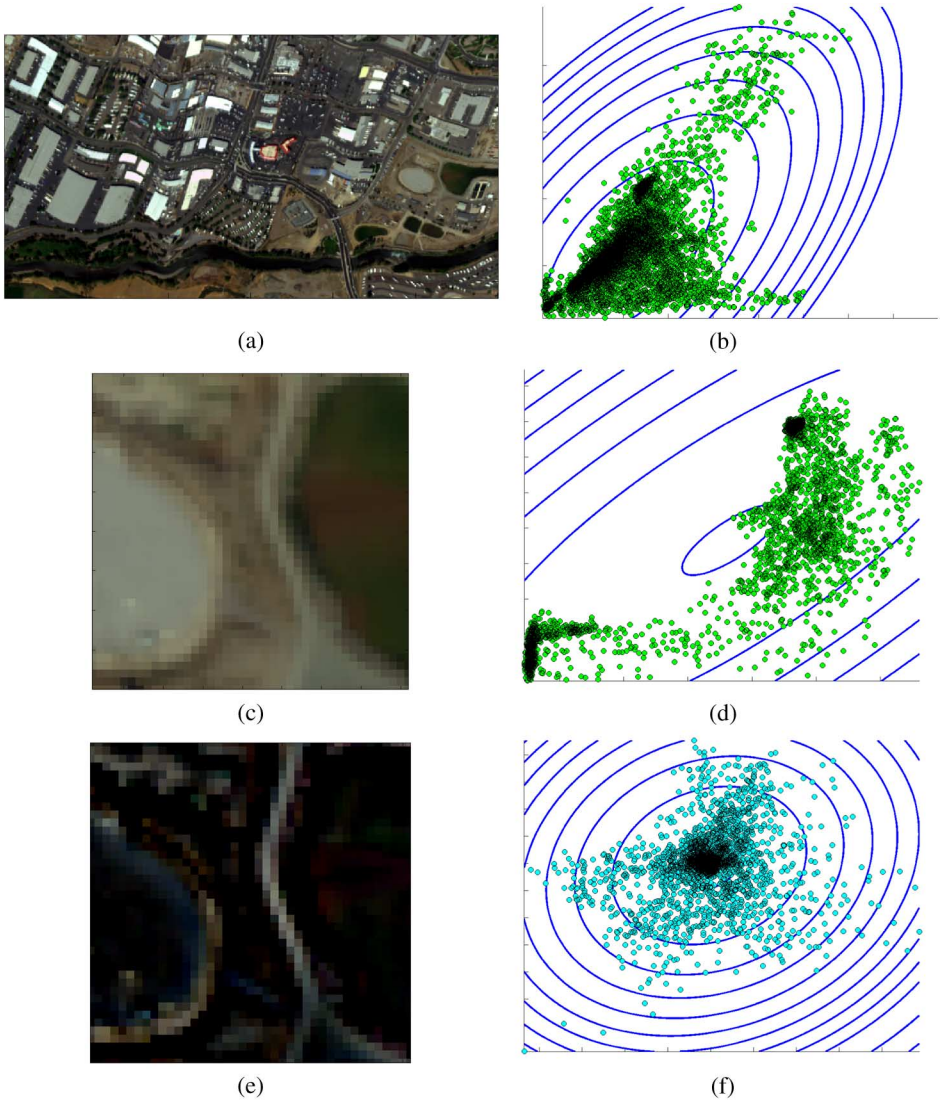


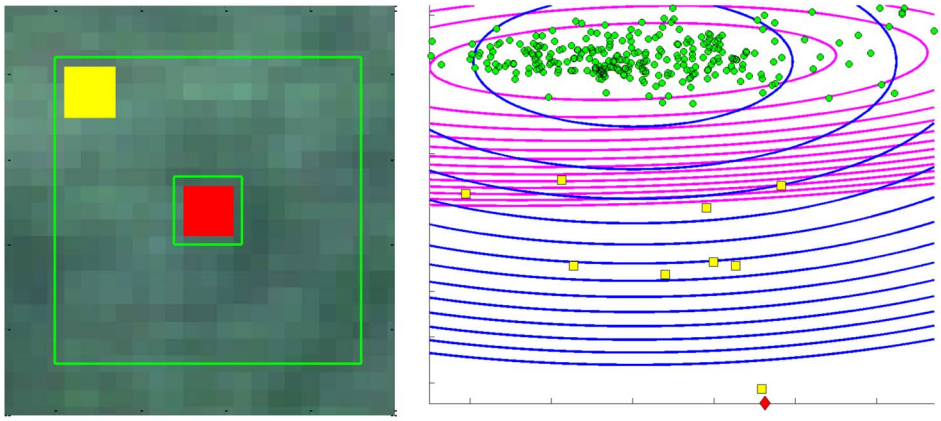
Fig. 2. Graphical example of a situation in which the background data are not well explained by the MVN model. (a) Spatial domain featuring a SpecTIR hyperspectral image acquired over the city of Reno [40]. A true-color representation is provided. (b) Spectral domain with respect to two spectral channels. The image pixels are represented as green circles. MVN contours are plotted in blue. (c) Spatial domain pertaining to a 20×20 subset of the image. (d) Spectral domain with respect to two spectral channels. The pixels belonging to the 20×20 image subset are represented as green circles. MVN contours are plotted in blue. (e) Spatial domain pertaining to the same 20×20 subset of the image after local mean removal. (f) Spectral domain with respect to two spectral channels. The pixels belonging to the 20×20 image subset after local mean removal are represented as cyan circles. MVN contours for the data after local mean removal are plotted in blue.

performed violating the key assumption under which adaptive detectors are developed, i.e., the availability of target-free training data.

Fig. 3 shows a graphical example of the impact of background contamination by the target signal on local RX detection. Fig. 3 (a) displays the situation of local RX applied to a target pixel, while the local background is homogeneous except for the presence of one contamination source bearing the target signal. Fig. 3(b) shows a scatterplot of the data in a simplified two-dimensional domain along the two principal axes. Although only a low percentage (about 3%) of background data is contaminated by the target signal, RX contours are strongly biased by the contaminated data and RX outcome at the target test pixel is expected to be much lower than the one that would be obtained in a contamination-free case.

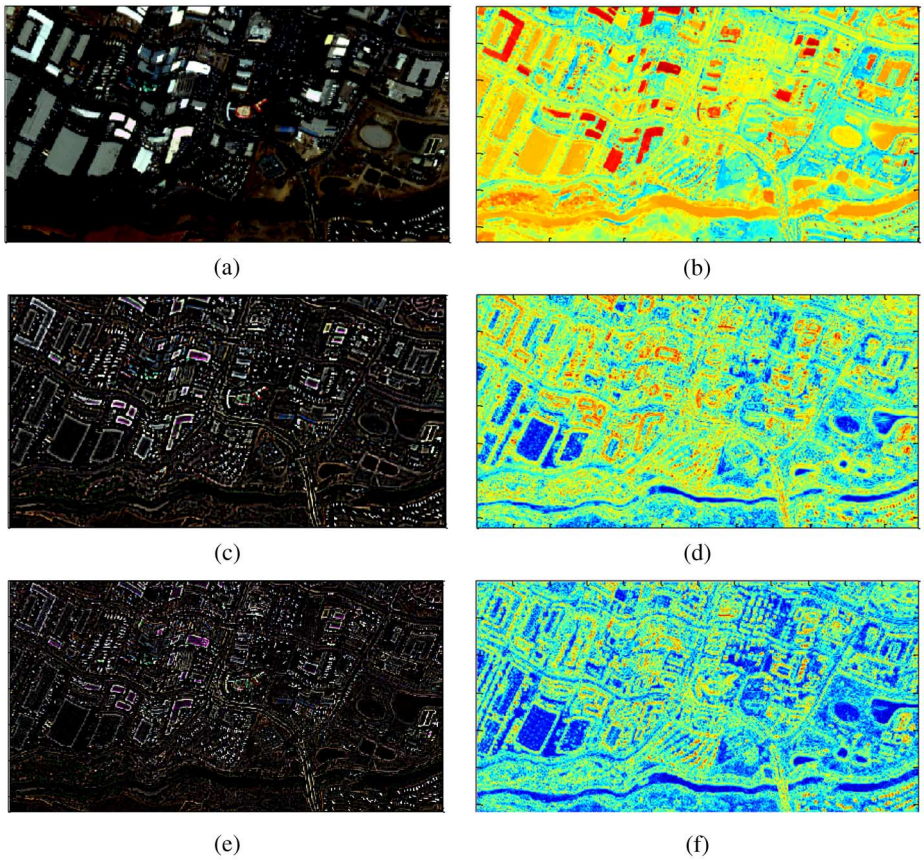
Algorithms specifically developed to deal with covariance corruption due to contamination are described in Section III-A2.

3) Edges and Spatial Structures: In contrast with global background characterization models, local models characterize the background in a small neighborhood of each pixel, thus adapting to the spatially varying statistics of the background. The result is that whereas global background characterization methods would provide the same output even if the positions of all the pixels were scrambled, local methods implicitly incorporate into the model that the pixels are laid out according to specific spatial patterns, textures, and structures. In Fig. 4, a graphical example highlights this distinction by considering the background mean vector estimator. The residual images obtained from the original image after mean vector removal may represent the input of a given detection algorithm, such as the Mahalanobis distance detector. As it is evident in the figure, local mean removal [Fig. 4(c)–(f)] provides stronger background suppression than global mean removal [Fig. 4(a), (b)]. Furthermore, the smaller the neighborhood



(a) (b)

Fig. 3. Graphical example of a situation in which the background data are contaminated by the target signal. (a) Spatial domain featuring the test pixel $\boldsymbol{x}$ belonging to a target (in red), a homogeneous local background (within the green annulus), and one contamination source (in yellow) bearing the same signal as the target. (b) Spectral domain with respect to the two principal axes. The pixels belonging to the local homogenous background are represented as green circles, the target pixel is denoted with a red diamond, and the pixels belonging to the contamination source are identified as yellow squares. RX contours for the contaminated background are plotted in blue. RX contours for the contamination-free case are plotted in magenta as reference. Contamination clearly biases RX contours and RX score in $\boldsymbol{x}$ is much lower than the contamination-free case.



(a) (b) (c) (d) (e) (f)

Fig. 4. Graphical example highlighting the impact of edges and structures on a simple background estimation model, using the image shown in Fig. 2(a). (a) Residual image (true-color representation) after removal of the global mean vector and (b) corresponding energy image. The major background structures are still clearly evident. (c) Residual image (true-color representation) after removal of the local mean vector using a 7×7 annulus and (d) corresponding energy image. The gross background structures are no longer evident. Edges and transitions between different background objects emerge, though, with respect to most of pixels where background is suppressed. (e) Residual image (true-color representation) after removal of the local mean vector using a 3×3 annulus and (f) corresponding energy image. Using a smaller annulus results in stronger background suppression. However, edges and transitions still are present in the residual image, though at a lesser extent.

used for evaluating the mean vector, the higher the capability to adapt to background variations. Nonetheless, Fig. 4(e) and (f) clearly shows that, even though a small 3×3 annulus is employed, residual background structures still manifest themselves in the residual image. These residual structures, mostly due to edges and transitions between different background objects, are likely to be source of false alarms after the detection step. Therefore, methods to explicitly

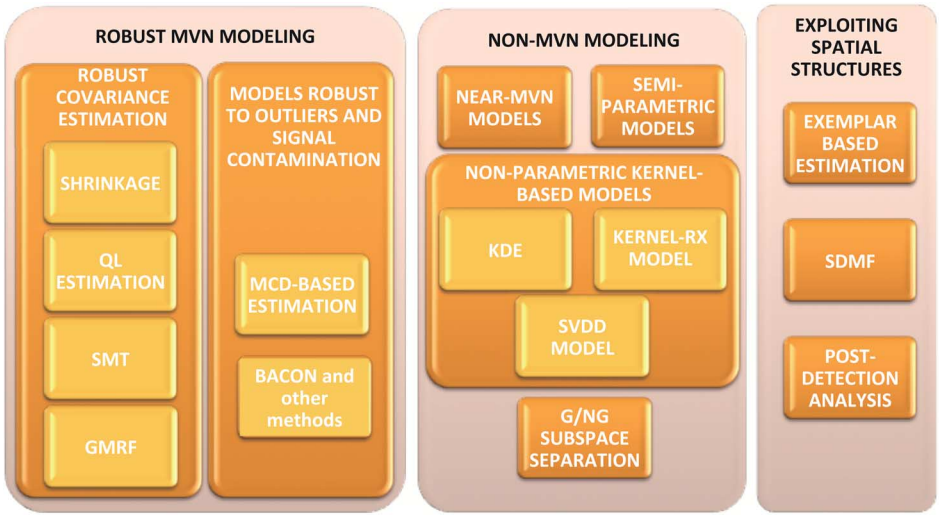


Fig. 5. Synoptic graphical summary table. The three main blocks identify the three main subsections of Section III, namely robust MVN modeling, non-MVN modeling, and exploitation of spatial structures. All algorithm names and acronyms in the table will be defined in the text.

account for background structures, either during or after the detection process, could be beneficial to the detection performance. Some possible approaches will be described in Section III-C.

III. Improved Background Modeling for Improved Detection Performance

This section will provide an overview of established as well as more recent attempts to improve background modeling in hyperspectral target detection applications. Section III-A will review robust MVN models, that is background models that still adopt the Gaussian assumption, although improving some of its limitations. Models that considerably depart from the Gaussian assumption in favor of an increased adaptability to complex nonhomogeneous backgrounds are presented in Section III-B. Section III-C describes approaches expressly designed to account for spatial structures and edges. All of these models are summarized in Fig. 5, which also provides a graphical outline for this section.

A. Robust MVN Background Modeling

Although the MVN model is clearly incomplete—as seen, for instance, in Fig. 2(b), (d), and (f)—it is an approximation that can still be useful. In Section III-B, we will describe various non-MVN models for the background, but here we will discuss approaches for making MVN more effective. In particular, although the MVN is a relatively simple model, estimation of the covariance matrix is not trivial, particularly for local models where the number of samples per estimate is small. Furthermore, the presence of outliers in the background data brings further challenges for reliable covariance matrix estimation.

1) MVN Models With Improved Covariance Matrix Estimation: Covariance matrix estimation is often a challenge. Although one can estimate the covariance matrix in a straightforward way, most detection algorithms actually use the inverse covariance matrix estimate, and the best estimate of the inverse covariance is not the same thing as the inverse of the best estimate of the covariance. The difficulty associated with covariance matrix estimation is increased for local algorithms, due to the small sample size. In [82], it is shown that the MF exhibits an average performance loss that scales like d/N where d is the number of channels and N is the number of independent samples used to compute the sample covariance matrix. It has been pointed out in [12] and [13], however, that this performance loss estimate is optimistic in the case that the covariance matrix arises from a distribution that is not a single Gaussian of the given covariance matrix. Similar performance scaling with d/N has been reported for RX-based anomaly detection [29].

a) Shrinkage: One of the most straightforward ways to improve the estimate of the inverse covariance matrix is by shrinkage [32], [39], [48], [79], [105]. If we write

$$\boldsymbol{S} = \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T \tag{11}$$

as the sample covariance matrix, then

$$\boldsymbol{C}_{shr} = (1-\alpha_{shr})\boldsymbol{S} + \alpha \boldsymbol{C}_o \tag{12}$$

is the shrinkage estimator. Here, $\boldsymbol{C}_o$ is an *underfit* (or *prior*) estimate of the covariance matrix; typical choices include an appropriately scaled identity matrix, or the diagonal elements of $\boldsymbol{S}$. Another possible choice for $\boldsymbol{C}_o$, when it is the local covariance matrix that is being estimated, is to use the global sample covariance matrix. The shrinkage parameter α_{shr} is typically much less than one, so the effect of shrinkage is a small perturbation on the sample covariance matrix $\boldsymbol{S}$; but especially when $\boldsymbol{S}$ is near-singular, the effect on the inverse matrix can be considerable.

b) Quasi Local Estimation: The idea of combining global with local estimators of covariance matrix also drives the "quasi-local" (QL) model [36]. The global sample covariance matrix $\mathbf{S}_G$ may be decomposed as $\mathbf{S}_G = \mathbf{V}_G\mathbf{D}\mathbf{V}_G^T$, where $\mathbf{V}_G$ is an orthogonal matrix and $\mathbf{D}$ is diagonal. If $\mathbf{S}_L$ is the local

sample covariance matrix, then $\mathbf{V}_G^T\mathbf{S}_L\mathbf{V}_G$ is the covariance matrix in the rotated coordinates that diagonalize $\mathbf{S}_G$. In general, $\mathbf{V}_G^T\mathbf{S}_L\mathbf{V}_G$ is not diagonal, but we can write

$$\boldsymbol{C}_{QL} = \mathbf{V}_G(diag(\mathbf{V}_G^{\mathrm{T}}\mathbf{S}_L\mathbf{V}_G))\mathbf{V}_G^T \tag{13}$$

where the $diag(\cdot)$ operator keeps the diagonal elements of a matrix and sets the off-diagonal values to zero. This estimator is readily invertible, since the diagonal elements of $\mathbf{V}_G^T\mathbf{S}_L\mathbf{V}_G$ will generically be strictly positive, and

$$\boldsymbol{C}_{QL}{}^{-1} = \mathbf{V}_G\Big(diag\big(\mathbf{V}_G^T\mathbf{S}_L\mathbf{V}_G\big)^{-1}\Big)\mathbf{V}_G^T. \tag{14}$$

In addition to providing a more stable estimator for the inverse covariance matrix, this scheme also provides an inexpensive way to compute the inverse; this can be important when local algorithms require a separate inverse covariance matrix at every pixel in an image.

One inflexibility with the QL estimator is that the "amount" of regularization is not adjustable; i.e., there is nothing like the parameter α_{shr} that is used for shrinkage schemes.

c) Sparse Matrix Transform: An extension of the QL estimator is given by the sparse matrix transform (SMT), first introduced in [20], and applied to remote sensing problems in [113]. The SMT scheme expresses a sample covariance matrix as a product of Givens rotations

$$\boldsymbol{S} = \boldsymbol{G}_1\boldsymbol{G}_2\ldots\boldsymbol{G}_K\boldsymbol{S}_K\boldsymbol{G}_K^T\ldots\boldsymbol{G}_2^T\boldsymbol{G}_1^T \tag{15}$$

where each Givens rotation is an orthogonal matrix that corresponds to a rotation about a single pair of coordinates

$$\boldsymbol{G} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos\theta & \cdots & -\sin\theta & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \sin\theta & \cdots & \cos\theta & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}. \tag{16}$$

The Givens rotations $\boldsymbol{G}_1, \boldsymbol{G}_2\ldots\boldsymbol{G}_K$ are chosen sequentially, using an algorithm described in [20], so that $\boldsymbol{S}_k = \boldsymbol{G}_k^T\boldsymbol{S}_{k-1}\boldsymbol{G}_k$ for $k = 1,\ldots,K$ are increasingly closer to diagonal. The approximation that SMT makes is to replace $\boldsymbol{S}_K$ with $diag(\boldsymbol{S}_K)$. In particular, if we write $\boldsymbol{Q}_K = \boldsymbol{G}_1\boldsymbol{G}_2\ldots\boldsymbol{G}_K$, then our estimate for the covariance becomes

$$\boldsymbol{C}_{SMT} = \boldsymbol{Q}_K\big(diag\big(\boldsymbol{Q}_K^T\boldsymbol{S}\boldsymbol{Q}_K\big)\big)\boldsymbol{Q}_K^T. \tag{17}$$

This formulation of SMT shares with the QL covariance matrix the notion of replacing an approximately diagonal matrix with its diagonal, but it has the advantage that there is a parameter K that can be used to adjust the quality of the approximation. It has been found in practice that better performance is obtained if SMT is combined with shrinkage; e.g.,

$$\boldsymbol{C}_{SMT}^{shr} = (1-\alpha_{shr})\boldsymbol{S} + \alpha\boldsymbol{C}_{SMT} \tag{18}$$

but it bears remarking that this estimator is not efficiently invertible as the straight $\boldsymbol{C}_{SMT}$ estimator.

This formulation of SMT, however, does not take advantage of a prior covariance estimator $\boldsymbol{C}_o$; an example of such a prior is the global covariance matrix $\boldsymbol{C}_G$. A suggestion for doing this [7], [36] follows the QL idea of working in the eigenspace defined by $\boldsymbol{C}_o$. Instead of the approximation $\mathbf{V}^{\mathrm{T}}\mathbf{S}\mathbf{V} \approx diag(\mathbf{V}^{\mathrm{T}}\mathbf{S}\mathbf{V})$ made in (13), we write

$$\boldsymbol{V}^T\boldsymbol{S}\boldsymbol{V} \approx \boldsymbol{Q}_K\big(\mathrm{diag}\big(\boldsymbol{Q}_K^T\boldsymbol{V}^T\boldsymbol{S}\boldsymbol{V}\boldsymbol{Q}_K\big)\big)\boldsymbol{Q}_K^T \tag{19}$$

where $\boldsymbol{Q}_K$ is the product of K givens rotations that approximately diagonalize $\boldsymbol{V}^T\boldsymbol{S}\boldsymbol{V}$, where $\boldsymbol{V}$ is the orthogonal matrix that diagonalizes $\boldsymbol{C}_o$ (i.e., $\boldsymbol{C}_o = \boldsymbol{V}^T\boldsymbol{D}\boldsymbol{V}$). This leads to the approximation

$$\boldsymbol{C}_{QL-SMT} = \boldsymbol{V}\boldsymbol{Q}_K\big(diag\big(\boldsymbol{Q}_K^T\boldsymbol{V}^T\boldsymbol{S}\boldsymbol{V}\boldsymbol{Q}_K\big)\big)\boldsymbol{Q}_K^T\boldsymbol{V}^T \tag{20}$$

which has the same flavor as QL covariance estimation, but uses $\boldsymbol{V}\boldsymbol{Q}_K$ in place of $\boldsymbol{V}$. This K provides an adjustable parameter; as K increases, $\boldsymbol{C}_{QL-SMT}$ more accurately approximates $\boldsymbol{S}$, but less robustly estimates the true underlying covariance matrix.

d) Gaussian Markov Random Field Model: In contrast to the approaches that employ sample covariance matrix as a starting point, a method suggested in [95] makes a direct parametric model of the inverse covariance matrix. This is obtained by characterizing both spectral and spatial covariance in terms of a local Gaussian Markov Random Field (GMRF). Even though the covariance matrix for a d-channel system will have $O(d^2)$ degrees of freedom, the GMRF model includes only four parameters, two of which relate to spatial variation. The authors point out that this simplicity leads to an anomaly detection algorithm faster than RX and allows for a smaller annulus size. We remark that the GMRF is one of the few models (another is described in [127]) for which the order of the spectral channels matters.

2) MVN Models Robust to Outliers: As noted in Section II-C2, the background data may be contaminated by the presence of pixels strongly deviating from the background distribution. These outlying pixels may be due to anomalies, to scarcely populated background classes whose signal content is considerably statistically different from the background signal of most pixels, or to the target signal.

In the case of target signal contamination, the "background" covariance matrix estimate inadvertently includes some target pixels. Thus, the estimated covariance matrix looks like $\tilde{\boldsymbol{C}} = \boldsymbol{C} + \varepsilon^2\boldsymbol{t}\boldsymbol{t}^T$, where $\boldsymbol{C}$ is the contamination-free covariance matrix estimate, $\boldsymbol{t}$ is the target signature, and ε^2 is associated with the contaminating signal energy and depends on the fraction of contaminated data.

The resulting contaminated MF is then $\tilde{\boldsymbol{q}} = (\boldsymbol{C} + \varepsilon^2\boldsymbol{t}\boldsymbol{t}^T)^{-1}\boldsymbol{t}$ instead of the correct $\boldsymbol{q} = \boldsymbol{C}^{-1}\boldsymbol{t}$, but it was noted in [107] that $\boldsymbol{q}$ and $\tilde{\boldsymbol{q}}$ are parallel, and that although the magnitude of $\boldsymbol{q}$ (and therefore the appropriate threshold η for detectors of the form $\boldsymbol{q}^T\boldsymbol{x} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$) is different for the two MFs, the performance is identical. Reference [107] did identify a more subtle kind of covariance contamination, based on a residual correlation of

plume and background in a given scene, and [111] showed how the effect of contamination is more pronounced when the target is not linearly additive.

When applied to local RX detection, the effect was found to be more pronounced [67], [71]. The RX detector in the presence of contamination resulted in $D(\boldsymbol{x}) = \boldsymbol{x}^T(\boldsymbol{C} + \varepsilon^2 \boldsymbol{t}\boldsymbol{t}^T)^{-1}\boldsymbol{x} = D_{RX}(\boldsymbol{x}) - \boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}$, where $\boldsymbol{M}$ is a positive semi-definite matrix depending on $\boldsymbol{C}^{-1}$, $\boldsymbol{t}$, and ε^2 (which includes the contaminating signal energy and the fraction of contaminated data). The study in [71] revealed that both detection and false alarm probabilities decrease monotonically with increasing contamination, with the average reduction of the detection probability being much larger than the average reduction of the false alarm probability. Contaminated RX performance degradation was quantitatively examined with respect to several parameters, such as the Contaminating Signal-to-Interference-plus-Noise Ratio $CSINR = \varepsilon^2 \boldsymbol{t}^T \boldsymbol{C}^{-1} \boldsymbol{t}$, which depends on the fraction of contaminated data and the strength of the contaminating signal with respect to background interference plus noise [71].

The covariance matrix estimate may also be corrupted by the presence of other outliers, i.e., pixel containing signals not necessarily associated with the target signal but that significantly deviate from the background distribution. This problem was examined in [35] for two specific detection algorithms. In [35], the effect of a single source of signal contamination was investigated with respect to the performance of Kelly's [44] detector and the Mean Level Adaptive Detector [35] when the contaminating signal was statistically independent of the target signal.

In the following, we briefly outline approaches specifically developed to mitigate covariance matrix corruption due to outliers. These approaches have turned out to be particularly useful to reduce the effect of signal contamination in anomaly detection applications.

a) Minimum Covariance Determinant-Based Covariance Matrix Estimation: Background covariance matrix susceptibility to corruption due to signal contamination has led to the development of robust-to-outlier detection methodologies [53], [70], [72], [98] aimed at making detection performance less sensitive to signal contamination although at the cost of increased architectural complexity.

In [72], the minimum covariance determinant (MCD) estimation technique [85] was embedded in the local RX detector, as a solution to mitigate local covariance matrix corruption. The resulting MCD-RX scheme estimates the covariance matrix over the N_0-dimensional subset ($N_0 < N$) of neighboring pixels characterized by the lowest covariance determinant (proportional to the squared hyperellipsoid volume), which is likely not to contain outliers [72], [85]. In [72] and [70], the MCD-RX detector was shown to considerably outperform RX for the detection of targets placed in close proximity. By recalling the graphical example depicted in Fig. 3, MCD-RX behavior is represented in Fig. 6 in the same simplified two-dimensional domain spanned by the two principal axes. Whereas RX contours are strongly biased due to the target pixels contaminating the background, MCD-RX provides contours that tightly follow the background data, thanks to the MCD robust estimation that allows the most outlying pixels to be recognized and neglected from background statistic evaluation.

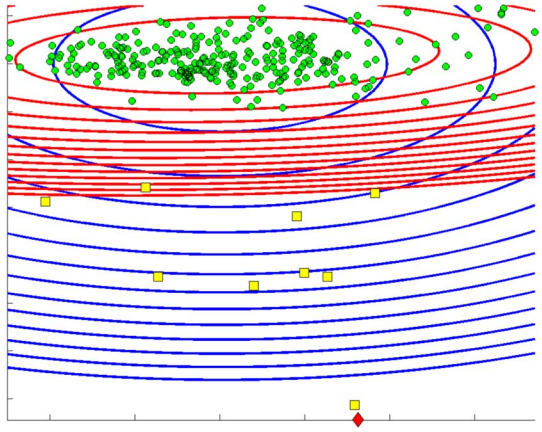


Fig. 6. Spectral domain with respect to the two principal axes of the situation depicted in Fig. 3(a). The pixels belonging to the local homogenous background are represented as green circles, the target pixel is denoted with a red diamond, and the pixels belonging to the contamination source are identified as yellow squares. RX contours for the contaminated background are plotted in blue. MCD-RX contours are plotted in red. By excluding the most outlying pixels (i.e., those belonging to the contamination source) from background statistics computation, MCD-RX provides more robust contours for the detection of the target.

Although promising, the use of MCD within the hyperspectral AD context has been generally restricted to perform *global* outlier detection on simplified low-dimensional data sets [53], [56], [98]. This was due to the high computational burden inherent in the MCD *iterative* structure [85], which unavoidably hinders a computationally efficient local application. A local algorithm that takes advantage of the MCD-RX approach and is also much more computationally efficient was developed in [70]. This algorithm, called Kurtosis-driven MCD-RX (K-MCD-RX), applies at each pixel a binary hypothesis test based on a statistic derived from the local sample kurtosis to decide if the pixel neighborhood contains significant outliers and, thus, deserves application of the more complex MCD-RX in lieu of RX. Theoretical criteria, based on the sample kurtosis distribution, are provided to automatically set the threshold for the kurtosis-based binary hypothesis test at each pixel [70]. In [68] and [70], experimental evidence was given of the capability of the kurtosis-based test of indicating the image regions most susceptible to covariance matrix corruption due to outliers. Also, K-MCD-RX was found to provide detection performance comparable to that of MCD-RX, even though robust MCD covariance estimation was applied to a low percentage of image pixels, with a considerable computational complexity reduction [68], [70].

b) Other Algorithms Featuring Robust Covariance Matrix Estimation: Other algorithms have been applied to robustly estimate background second-order statistics in hyperspectral imagery.

A standard approach for computing a robust covariance matrix is to de-weight samples based on their Mahalanobis distance from the mean vector [18]. Furthermore, simply removing the most anomalous data, and then re-computing the mean vector and covariance matrix from the remaining points, can lead to better hyperspectral target detection [8].

Another method is the blocked adaptive computationally efficient outlier nominator (BACON) developed in [14] and applied to robustly detect global anomalies in hyperspectral

images in [98]. It is worth mentioning that in the same work, a cluster-based version of BACON was implemented, following the idea in [21] and, at the same time, making the per-cluster statistic estimates robust to the presence of outliers.

It should be pointed out that the emphasis on "robust" covariance estimation may treat the periphery of the distribution as the source of contamination; but for anomaly detection (or for low false-alarm rate target detection), it is the periphery of the distribution that most requires accurate characterization. In [109], a variety of approaches were suggested, including an "anti-robust" estimator, for estimating a covariance matrix that incorporates robustness to the most outlying data while not de-emphasizing the data on the periphery. One of the approaches used a support vector machine in a manner similar to support vector data description (SVDD) [101], which is described below, but with a quadratic kernel. The emphasis on periphery makes these estimators more discriminative than generative.

Just as outliers can corrupt the covariance matrix estimate, they also corrupt the estimated principal components of the covariance matrix. This matters for subspace methods, and hybrid subspace methods, that employ the first few principal components to identify the large variance directions. There has been a lot of interest recently in the development of efficient algorithms for robustly estimating those principal components directly [19], [124].

B. Non-MVN Background Modeling

Although the MVN model can lead to useful algorithms for target and anomaly detection, it is ultimately an inadequate model because hyperspectral data are clearly (and by some measures, dramatically) non-Gaussian. Thus, there is an opportunity for algorithms to exploit this beyond-Gaussian structure to produce more sensitive target and anomaly detectors.

In the following, we will review more general background models that reject the Gaussian approximation entirely and employ more complicated multivariate distributions.

1) Near-MVN Models: A property of MVN models is that contours of the Gaussian distribution are ellipsoids, and the size, aspect, and orientation of those ellipsoids are given by a covariance matrix. Another property of MVN models is that the tails of the distribution have very low density, that scale like $e^{-\rho^2}$ as a function of radius ρ from the centroid. A variant of MVN models called elliptically-contoured (EC) models preserve the contour shape, and the parameterization with a covariance matrix, but permit distributions that have heavier tails than their Gaussian counterparts. Although these models have been used to model radar clutter for many decades, they were introduced to hyperspectral modeling by Manolakis *et al.* [45], [59], [62], [65]. A popular choice is to use the multivariate t-distribution for modeling the background PDF

$$p_b(\boldsymbol{x}) = c|\boldsymbol{C}|^{-\frac{1}{2}}\left[1+\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu-2}\right]^{-\frac{d+\nu}{2}} \tag{21}$$

where c is a constant (that depends on the parameter ν and the dimension d). Smaller values of the parameter ν correspond to heavier tailed distributions. The $\nu \to \infty$ limit produces the Gaussian distribution, and as $\nu \to 2$, the limit is a Laplacian distribution, for which the second and all higher moments diverge.

Because the contours of an EC distribution match those of the Gaussian distribution, standard anomaly detection algorithms (e.g., RX) based on EC models give exactly the same results as those based on the corresponding Gaussian model, but certain kinds of anomaly detection [74], [108] and anomalous change detection [112] have found productive uses for these fatter-than-Gaussian distributions.

But using EC in place of a Gaussian does lead to new target detection algorithms [90], [106], [122]. In particular, using GLRT for the multivariate-t with the additive target model leads to the EC-GLRT

$$D(\boldsymbol{x}) = \sqrt{\frac{(\nu-1)}{(\nu-2)+(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}\cdot\frac{\boldsymbol{t}^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\sqrt{\boldsymbol{t}^T\boldsymbol{C}^{-1}\boldsymbol{t}}} \tag{22}$$

which approaches the MF in the $\nu \to \infty$ limit, and looks like the ACE detector [47], [88] as $\nu \to 2$.

Instead of elliptical contours, the single band anomaly detection (SBAD) algorithm [46] effectively employs rectangular contours, since the detector is the maximum over detectors on each single band.

2) Semi-Parametric Models: Parametric models typically assume that the data are drawn from one specific parametric distribution (e.g., MVN, EC). Generally, parametric models can be reliably employed only to characterize background pixels in a homogeneous local neighborhood around the test pixel and have been more effectively used within local algorithms after local mean removal. When highly inhomogeneous backgrounds (e.g., global backgrounds, urban environments) have to be characterized, parametric models are no longer able to capture the complexity of the data. The presence of multiple materials within the scene suggests that *semi-parametric* distributions, such as the finite mixture models (FMMs), may provide a more accurate background characterization [21], [34], [69], [74], [99].

An FMM models the unknown background PDF $p_b(\boldsymbol{x})$ via a linear combination of PDFs of the same kind, thus accommodating multimodality

$$p_b(\boldsymbol{x}) = \sum_{j=1}^{J}\pi_j p_j(\boldsymbol{x};\boldsymbol{\theta}_j). \tag{23}$$

In (23), $p_j(\boldsymbol{x};\boldsymbol{\theta}_j)$ denotes the multivariate PDF of $\boldsymbol{x}$ pertinent to the jth mixture component, controlled by a given parameter vector $\boldsymbol{\theta}_j$, whereas $\{\pi_j\}_{j=1}^{J}$ are the mixing proportions (or weights), which are non-negative and sum to one.

The most widely employed FMM is the Gaussian mixture model (GMM), which has often been adopted to model global heterogeneous backgrounds in hyperspectral images [16], [21], [34], [69], [72], [74], [99]. Using the GMM means that each of the $p_j(\boldsymbol{x};\boldsymbol{\theta}_j)$ in (23) is an MVN PDF, where $\boldsymbol{\theta}_j$ is the vector specifying the mean vector and covariance matrix of the jth mixture component. With a sufficiently high number of mixture components J, any given PDF may be approximated, provided that the mixture parameters are adequately estimated [21], [99], [115].

A straightforward physical interpretation of hyperspectral background GMM modeling is that hyperspectral pixels of each

background class follow the MVN distribution. However, several experimental studies performed with real hyperspectral images [1], [64], [65], [74] have shown that mixtures of EC distributions may more adequately characterize the statistical behavior of multi-modal hyperspectral backgrounds with respect to a GMM. Specifically, by comparing the empirical and the theoretical cumulative distribution functions (CDFs) of the Mahalanobis distances computed between each pixel and the corresponding mixture component [1], [64], [65], [74], many background classes have been experimentally shown to exhibit heavy tails that are more accurately followed by an EC distribution.

In general, modeling $p_b(\boldsymbol{x})$ with an FMM means assuming that each pixel of the hyperspectral image originates from one mixture component, according to a given *a priori* probability. Thus, this is associated with segmenting the image into homogeneous clusters, each one corresponding to a subset of background pixels modeled by a single mixture PDF of the $\{p_j(\boldsymbol{x};\boldsymbol{\theta}_j)\}_{j=1}^{J}$. Whereas the segmentation procedure may be performed with simple clustering methods, such as the k-means [15], the expectation maximization (EM) method [15] takes into account the underlying distribution of the individual mixture components. Given the number of components J and an initial set of parameters, the EM algorithm iteratively seeks the maximum likelihood (ML) estimates of the FMM parameters [15]. Recently, a variational Bayesian method [115], originally employed with low-dimensional data sets, has been introduced in the context of FMM parameter estimation for hyperspectral images [74], [116]. The EM algorithm can be seen as a special case of this more general method [115]. By modeling some of the parameters with hidden variables characterized by proper prior distributions, the corresponding estimates assume physically meaningful values dictated by the corresponding prior distributions (e.g., the inverse covariance matrices of the components of a GMM should follow a Wishart distribution and the mixing proportions may be modeled as following a Dirichlet distribution, thus enforcing the non-negative and sum-to-one constraints). Casting the FMM parameter estimation in terms of Bayesian inference makes it possible to overcome some shortcomings of canonical EM, such as its intrinsic inability to automatically solve the model-order selection and the possibility of incurring in singular solutions associated with an unlimited likelihood function [74], [115].

Regardless of the specific clustering method used, detection may be performed following two different approaches. According to the well-known cluster-based approach [16], [21], [34], [72], a "cluster-conditional" test is applied to each pixel, for instance by computing a cluster-conditional Mahalanobis distance for FMM-based anomaly detection [16], [21], [34], [72]

$$D(\boldsymbol{x}) = \left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)^T \boldsymbol{C}_j^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right) \tag{24}$$

or a cluster-conditional MF for FMM-based target detection [16], [34]

$$D(\boldsymbol{x}) = \frac{\boldsymbol{t}^T \boldsymbol{C}_j^{-1}\boldsymbol{x}}{\sqrt{\boldsymbol{t}^T \boldsymbol{C}_j^{-1}\boldsymbol{t}}} \tag{25}$$

where $\boldsymbol{\mu}_j$ and $\boldsymbol{C}_j$ are the mean vector and covariance matrix of the MVN distribution modeling the jth mixture component, which may be chosen as the mixture component spectrally closest (in terms of Mahalanobis distance) to $\boldsymbol{x}$ [21].The approach of (24) has also been applied accounting for local information extracted from the spatial neighborhood of $\boldsymbol{x}$ [6], [16], [72], [96], [121], e.g., by selecting the jth mixture component as the spectrally closest one among those represented in the neighborhood of $\boldsymbol{x}$ [72].

In another approach, not only the parameters of the single mixture components but also the whole information associated to the FMM-based estimate of $p_b(\boldsymbol{x})$ is used. This is the case of the anomaly detector based on the FMM background log-likelihood [42], [69], [99]

$$D(\boldsymbol{x}) = -log\left[\sum_{j=1}^{J}\pi_j p_j(\boldsymbol{x};\boldsymbol{\theta}_j)\right]. \tag{26}$$

By recalling the graphical example depicted in Fig. 2, Fig. 7(a)–(d) displays the results of applying the GMM to model the images shown in Fig. 2(a) and (c). Specifically, Fig. 7(a) and (c) shows the images partitioned into a set of clusters by the EM algorithm. Fig. 7(b) and (d) represents the corresponding data scattered in simplified two-dimensional spectral domains together with GMM contours. In both global and local cases, the GMM is more effective than the MVN at following the inherently multimodal nature of the data.

3) Nonparametric Kernel-Based Models: For nonparametric models, the background data are not assumed to follow any specific distribution; instead, the background characterization is performed in a data-driven fashion. Nonparametric approaches employed in hyperspectral data processing include graph-based methods [5], [11], manifold methods [54], [55], and the kernel-based methods that we describe in this section.

a) Kernel Density Estimate of the Background Distribution: Kernel density estimation (KDE) is the most widely known approach to estimate an unknown PDF without assuming any fixed functional form for it [97]. Multivariate KDE has been used to estimate the background PDF of hyperspectral images within the framework of the background log-likelihood anomaly detector [73], [74], [75], [117], [118], [119]. Specifically, the general expression for multivariate KDE is as follows [97]:

$$p_b(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\left|\boldsymbol{H}(\boldsymbol{x},\boldsymbol{x}_n)\right|}\kappa\left[\boldsymbol{H}^{-1}(\boldsymbol{x},\boldsymbol{x}_n)(\boldsymbol{x}-\boldsymbol{x}_n)\right] \tag{27}$$

where $\kappa(\cdot)$ is the kernel function that is centered at each of the sample data $\{\boldsymbol{x}_n\}_{n=1}^{N}$ and $\boldsymbol{H}(\cdot)$ is the bandwidth matrix, a $d\times d$ matrix containing the kernel function widths, also referred to as *bandwidths*. The kernel function is a smooth function that decreases in intensity with the distance from the data sample in which it is centered. According to (27), the background PDF estimate is constructed by performing a weighted average of the values assumed in the test pixel $\boldsymbol{x}$ by the kernel function centered at each of the sample data. Regardless of the specific kernel function employed (which is generally taken to be a multivariate Gaussian function), the degree of smoothing of the PDF estimate is determined by the bandwidths.

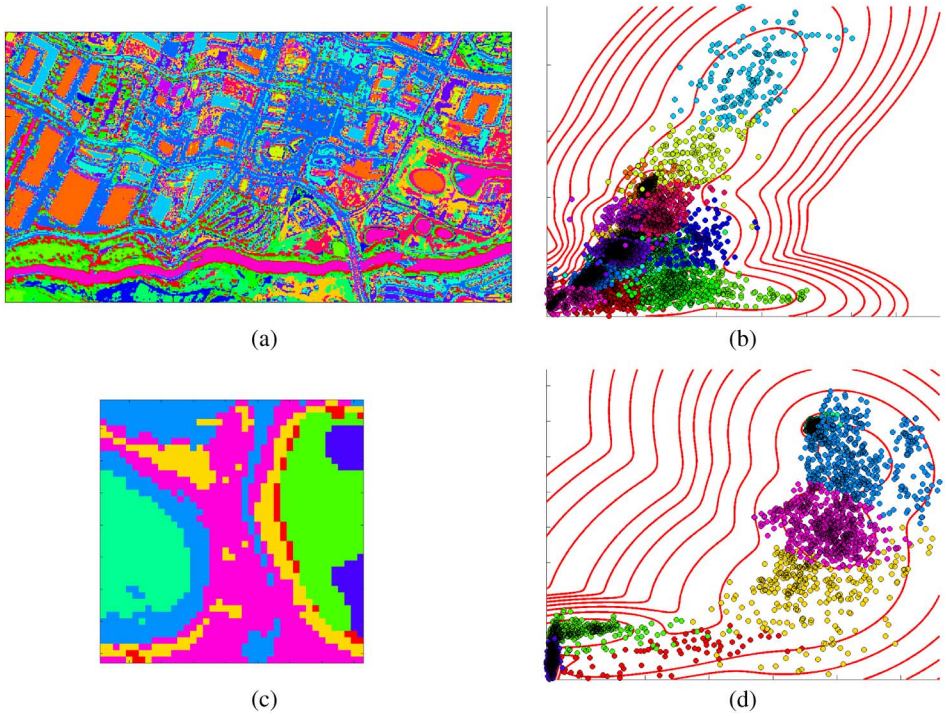


Fig. 7. Outcome of a GMM semi-parametric modeling as concerns the graphical example of Fig. 2. (a) Spatial domain with the image of Reno segmented into clusters by the EM algorithm. (b) Spectral domain with respect to two spectral channels. The image pixels belonging to a given cluster are represented as circles colored with the same color employed in the cluster map. GMM contours are plotted in red. (c) Spatial domain with the 20×20 image subset segmented into clusters by the EM algorithm. (d) Spectral domain with respect to two spectral channels. The pixels belonging to a given cluster are represented as circles colored with the same color employed in the cluster map. GMM contours are plotted in red.

The bandwidth matrix may be parameterized to be a diagonal matrix $\boldsymbol{H} = diag\{[h_1, \ldots, h_d]\}$, thus featuring a bandwidth value for each spectral dimension. It is more often the case that the matrix $\boldsymbol{H}$ is taken to be a scaled $d \times d$ identity matrix $\boldsymbol{H} = h\boldsymbol{I}_d$, which constrains the contours of the kernel function to be spherically symmetric. This latter simplification has been shown to be as effective as more complicated forms of $\boldsymbol{H}$ if the data are pre-scaled to prevent severe differences of spread in the various spectral directions [97]. Indeed, Fukunaga [33] suggested that whitening the data (i.e., linearly transforming the data to have an identity covariance matrix) and adopting $\boldsymbol{H} = h\boldsymbol{I}_d$ to estimate the PDF of the transformed data is equivalent to perform estimation in the original domain by using a bandwidth matrix with same structure of the data covariance matrix. Thus, in the following we simplify our analysis to the case $\boldsymbol{H} = h\boldsymbol{I}_d$:

$$p_b(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{h^d(\boldsymbol{x}, \boldsymbol{x}_n)}\kappa\left[\frac{\boldsymbol{x} - \boldsymbol{x}_n}{h(\boldsymbol{x}, \boldsymbol{x}_n)}\right]. \tag{28}$$

The detector corresponding to employment of (28) within the background log-likelihood function results in [69] and [74]

$$D(\boldsymbol{x}) = -log\left[\frac{1}{N}\sum_{n=1}^{N} \frac{1}{h^d(\boldsymbol{x}, \boldsymbol{x}_n)}\kappa\left[\frac{\boldsymbol{x} - \boldsymbol{x}_n}{h(\boldsymbol{x}, \boldsymbol{x}_n)}\right]\right]. \tag{29}$$

The most widely known form of the KDE is the so-called fixed KDE (FKDE) [97], also known as Parzen windowing [80]. In FKDE, the bandwidth is constant: $h(\boldsymbol{x}, \boldsymbol{x}_n) = h$. This makes the modeling ability of FKDE strongly depend on the specific value of h, which should be appropriately chosen. Although many techniques have been proposed and tested to properly choose the bandwidth [74], [97], [119], a unique bandwidth value might not exist that avoids over-smoothing the PDF body and, at the same time, under-smoothing the PDF on the tails and in low-density regions. This problem has been found to be more and more significant as the data dimension increases [97]. This aspect can be exemplified in the toy example represented in Fig. 8, featuring data generated following a mixture of two Gaussian distributions in a simplified bivariate domain. The true bivariate PDF is represented in Fig. 8(a), whereas Fig. 8(b) and (c) represents the outcomes of FKDE estimation obtained with different values of h. Specifically, the marginal PDFs, obtained by numerical integration of the joint PDF estimate, are reported for convenience. As it is evident, variation of h has a major impact on FKDE outcome, which, for most h values, does not properly respond to the variations of the true PDF. Too small h values lead to spurious structures in the estimate at some data sample locations, whereas too large h values tend to obscure the bimodal nature of the PDF due to over-smoothing.

In order to reduce KDE performance sensitivity to a fixed bandwidth value and, at the same time, increase the KDE capability of following the local data peculiarities across the data domain, bandwidths that vary on a per-pixel basis may be employed so as to adapt the amount of smoothing to the local density of samples in the data space. This approach has led to the variable-bandwidth KDE (VKDE) [97], [102], which has been shown to bring an increased capability of following complex backgrounds in hyperspectral images [73], [75], [117]. Within

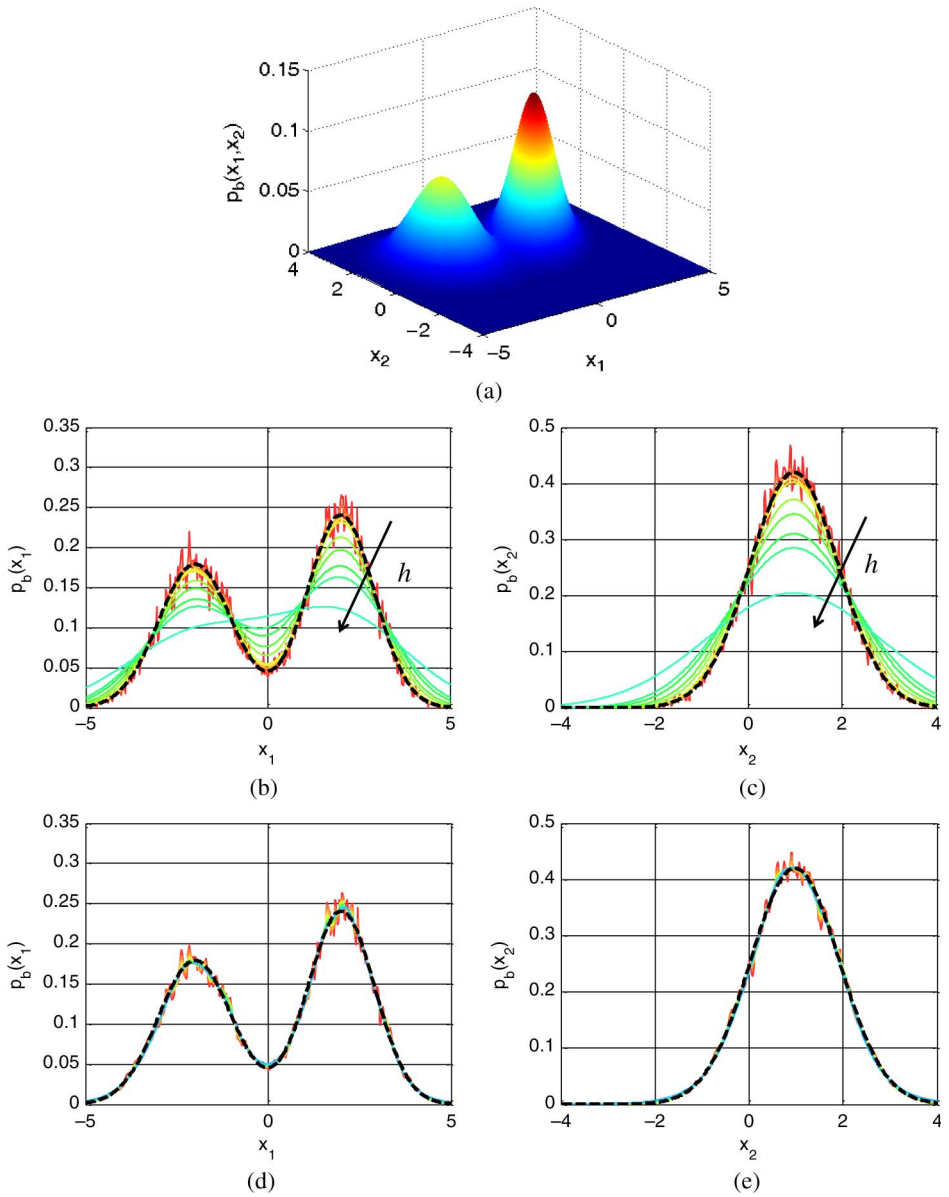


Fig. 8. (a) Bivariate PDF of a mixture of two Gaussians. Marginal PDFs estimates along the two axes, obtained by numerical integration of the bivariate PDF estimate. (b) and (c) FKDE estimates for different h values. (d) and (e) VKDE estimates for different k values. The True marginal PDFs are plotted in dashed black.

the VKDE framework, two distinct approaches of varying the bandwidth may be found. The Balloon Estimator (BE) varies the bandwidth at each test pixel $\boldsymbol{x}$, i.e., $h(\boldsymbol{x}, \boldsymbol{x}_n) = h(\boldsymbol{x})$, while keeping it constant for all sample data. The sample point estimator (SPE) varies the bandwidths at each of the sample data regardless of where the PDF is estimated, i.e., $h(\boldsymbol{x}, \boldsymbol{x}_n) = h(\boldsymbol{x}_n)$ [86]. Both approaches may be coupled with the k-nearest neighbor method (k-NN) as a strategy to let the bandwidth vary according to the local density of samples in the data space [97]. The k-NN-based BE results in taking $h(\boldsymbol{x}) = r_k(\boldsymbol{x})$, where $r_k(\boldsymbol{x})$ is the Euclidean distance of $\boldsymbol{x}$ to its kth nearest neighbor within the sample data $\{\boldsymbol{x}_n\}_{n=1}^N$. The SPE k-NN implementation is equivalent to taking $h(\boldsymbol{x}_n) = r_k(\boldsymbol{x}_n)$, with $r_k(\boldsymbol{x}_n)$ being the distance of $\boldsymbol{x}_n$ to its kth nearest neighbor within $\{\boldsymbol{x}_i\}_{i=1, i\neq n}^N$.

The VKDE—implemented with the k-NN strategy—has shown to provide better background characterization capabilities [73], [75] than the FKDE, avoiding over-smoothing and under-smoothing effects. Also, the dependence on the k parameter has been shown to be very weak [73], [75]. Fig. 8(d) and (e) provides a graphical illustration by displaying the marginal PDFs obtained by integrating the joint PDF estimate of the bivariate PDF in Fig. 8 (a), performed with VKDE applied in conjunction with the k-NN strategy for different values of k. Choosing the bandwidth according to the local data density makes VKDE much more data-responsive than FKDE. For most k values, VKDE estimates closely follow the true PDF both in the body and in the tails. Furthermore, the k parameter weakly affects the final outcome because it simply rules the cardinality of the sample data neighborhood considered for evaluating the bandwidth, which is computed as the distance to the kth neighbor and, thus, expressly tailored to the density of data samples in that neighborhood.

By recalling the graphical example depicted in Fig. 2, Fig. 9 represents the outcome of KDE (and, specifically, VKDE) as

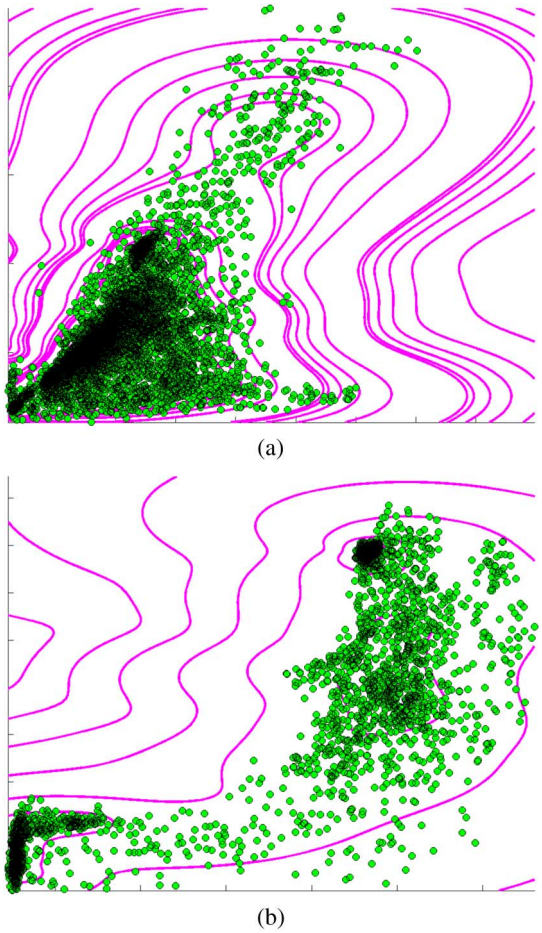


Fig. 9. Outcome of a VKDE parametric model as concerns the graphical example of Fig. 2. (a) Spectral domain with respect to two spectral channels. The image pixels are represented as green circles. VKDE contours are plotted in magenta. (b) Spectral domain with respect to two spectral channels. Only the pixels belonging to a 20×20 subset of the image are shown, again represented as green circles. VKDE contours are plotted in magenta.

concerns the two images considered in Fig. 2(a) and (c). VKDE contours in the simplified two-dimensional spectral domains show that this approach is capable of adapting to peculiar data distribution without assuming any specific statistical model for the data themselves.

Finally, it is interesting to observe that the FKDE may be seen as a simple Euclidean Distance detector applied in a higher-dimensional kernel-induced space obtained though a nonlinear mapping $\Phi(\cdot)$, as shown in [27] and [69]

$$\begin{aligned} ED[\Phi(\boldsymbol{x})] &= \left[\Phi(\boldsymbol{x}) - \boldsymbol{\mu}_\phi\right]^T \left[\Phi(\boldsymbol{x}) - \boldsymbol{\mu}_\phi\right] \\ &= \Phi^T(\boldsymbol{x})\Phi(\boldsymbol{x}) - \frac{2}{N}\sum_{n=1}^{N} \Phi^T(\boldsymbol{x})\Phi(\boldsymbol{x}_n) \\ &\quad + \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \Phi^T(\boldsymbol{x}_n)\Phi(\boldsymbol{x}_m) \end{aligned} \tag{30}$$

where $\boldsymbol{\mu}_\phi = \frac{1}{N}\sum_{n=1}^{N} \Phi(\boldsymbol{x}_n)$ is the estimate of the mean vector computed with the mapped data. By expressing the dot products in the kernel-induced space in terms of kernel functions $K(\boldsymbol{x}, \boldsymbol{y}) = \Phi^T(\boldsymbol{x})\Phi(\boldsymbol{y})$ a relationship with FKDE is obtained

$$\begin{aligned} ED[\Phi(\boldsymbol{x})] &= \mathrm{K}(\boldsymbol{x}, \boldsymbol{x}) - \frac{2}{N}\sum_{n=1}^{N} \mathrm{K}(\boldsymbol{x}, \boldsymbol{x}_n) \\ &\quad + \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \mathrm{K}(\boldsymbol{x}_n, \boldsymbol{x}_m) \\ &= const - \frac{2}{N}\sum_{n=1}^{N} \mathrm{K}(\boldsymbol{x}, \boldsymbol{x}_n) \\ &= const - 2p_b(\boldsymbol{x}) \end{aligned} \tag{31}$$

where $p_b(\boldsymbol{x})$ is the PDF estimate given by (28) considering $\mathrm{K}(\boldsymbol{x}, \boldsymbol{x}_n) = \frac{1}{h^d}\kappa\left(\frac{\boldsymbol{x}-\boldsymbol{x}_n}{h}\right)$. This holds for any stationary (translational invariant) kernel $\mathrm{K}(\boldsymbol{x}, \boldsymbol{x}_n) = \mathrm{K}(\boldsymbol{x} - \boldsymbol{x}_n)$.

b) Kernel-RX: A MVN Model in the Kernel-Induced Space: The kernel-RX (KRX) algorithm was derived by assuming an MVN model in the higher-dimensional space induced by the nonlinear mapping $\Phi(\cdot)$, and applying a Mahalanobis-distance-based detector [50]

$$MD[\Phi(\boldsymbol{x})] = \left[\Phi(\boldsymbol{x}) - \boldsymbol{\mu}_\phi\right]^T \boldsymbol{C}_\phi^{-1} \left[\Phi(\boldsymbol{x}) - \boldsymbol{\mu}_\phi\right] \tag{32}$$

where $\boldsymbol{\mu}_\phi = \frac{1}{N}\sum_{n=1}^{N} \Phi(\boldsymbol{x}_n)$ and $\boldsymbol{C}_\phi = \frac{1}{N}\sum_{n=1}^{N}[\Phi(\boldsymbol{x}_n) - \boldsymbol{\mu}_\phi]$ $[\Phi(\boldsymbol{x}_n) - \boldsymbol{\mu}_\phi]^T$ are the estimates of mean vector and covariance matrix computed with the mapped data. Note that $\boldsymbol{C}_\phi^{-1}$ is the pseudo-inverse of $\boldsymbol{C}_\phi$ since the covariance matrix is of lower rank than the dimension of the kernel-induced space. As with (30), the expression is not evaluated in the high-dimensional kernel-induced space, but employs kernel functions for dot products, leading to the main expression for Kernel-RX algorithm, introduced in [50], and corrected in [78]

$$D(\boldsymbol{x}) = KRX(\boldsymbol{x}) = \left(\boldsymbol{K}_x^T - \boldsymbol{K}_{\boldsymbol{\mu}}^T\right)^T \boldsymbol{K}_b^{-2} \left(\boldsymbol{K}_x^T - \boldsymbol{K}_{\boldsymbol{\mu}}^T\right). \tag{33}$$

In (33), $\boldsymbol{K}_x^T$ and $\boldsymbol{K}_{\boldsymbol{\mu}}^T$ are vectors defined in [50] and $\boldsymbol{K}_b$ is the $N \times N$ centered Gram matrix [93], [94].

Following [27], we can write $\boldsymbol{C}_\phi^{-1} = \boldsymbol{V}_\phi \boldsymbol{\Lambda}_\phi^{-1} \mathbf{V}_\phi^T$ where $\boldsymbol{V}_\phi$ is the matrix of eigenvectors and $\boldsymbol{\Lambda}_\phi$ is the diagonal matrix with r positive eigenvalues $\{\lambda_k\}_{k=1}^r$, of the covariance matrix $\boldsymbol{C}_\phi$. From (32)

$$\begin{aligned} D(\boldsymbol{x}) = MD[\Phi(\boldsymbol{x})] &= \Phi_0^T(\boldsymbol{x}) \boldsymbol{V}_\phi \boldsymbol{\Lambda}_\phi^{-1} \mathbf{V}_\phi^T \Phi_0(\boldsymbol{x}) \\ &= \sum_{k=1}^{r} \frac{\left[\mathbf{V}_\phi^{(k)} \Phi_0(\boldsymbol{x})\right]^2}{\lambda_k} \\ &= \sum_{k=1}^{r} \left[\sum_{i=1}^{N} \alpha_i^{(k)} K_0(\boldsymbol{x}_i, \boldsymbol{x})\right]^2 \end{aligned} \tag{34}$$

where the eigenvectors $\{\mathbf{V}_\phi^{(k)} = \sum_{i=1}^{N} \alpha_i^{(k)} \Phi_0(\boldsymbol{x}_i)\}_{k=1}^r$ lie in the span of the mapped centered data $\Phi_0(\boldsymbol{x}) = \Phi(\boldsymbol{x}) - \boldsymbol{\mu}_\phi$. The constants $\alpha_i^{(k)}$ can be derived from the eigenvectors and eigenvalues of $\boldsymbol{K}_b$, and $K_0(\cdot)$ is the centered kernel function,

expressed as follows [94]:

$$K_0(\boldsymbol{x}_i,\boldsymbol{x}) = K(\boldsymbol{x}_i,\boldsymbol{x}) - \frac{1}{N}\sum_{j=1}^{N}[K(\boldsymbol{x},\boldsymbol{x}_j)+K(\boldsymbol{x}_i,\boldsymbol{x}_j)] + \frac{1}{N^2}\sum_{j=1}^{N}\sum_{l=1}^{N}K(\boldsymbol{x}_l,\boldsymbol{x}_j). \quad (35)$$

Equations (34) and (35) show that, although entailing a more complex algebraic expression than a simple sum of kernel functions (as in FKDE), kernel-RX still involves computation of kernel functions over pairs of sample data and may, thus, be seen as a generalization of FKDE.

c) Estimating the Support of the Background Distribution: In [10], hyperspectral background characterization was performed by exploiting the SVDD [101], which has its foundations in the *support vector* approach [93]. The idea behind this approach is to estimate the data support region where most of the data lie, which should be easier than estimating the data PDF [94]. Within this framework, the SVDD aims at identifying the data support region by means of the minimum hypersphere enclosing (most of) the sample data in a higher-dimensional kernel-induced space obtained though a nonlinear mapping $\Phi(\cdot)$

$$\begin{aligned} SVDD[\Phi(\boldsymbol{x})] &= \left[\Phi(\boldsymbol{x}) - \sum_{n=1}^{N}\alpha_n\Phi(\boldsymbol{x}_n)\right]^T \times \left[\Phi(\boldsymbol{x}) - \sum_{n=1}^{N}\alpha_n\Phi(\boldsymbol{x}_n)\right] \\ &= \Phi^T(\boldsymbol{x})\Phi(\boldsymbol{x}) - 2\sum_{n=1}^{N}\alpha_n\Phi^T(\boldsymbol{x})\Phi(\boldsymbol{x}_n) + \sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m\Phi^T(\boldsymbol{x}_n)\Phi(\boldsymbol{x}_m). \end{aligned} \quad (36)$$

In (36), $\sum_{n=1}^{N}\alpha_n\Phi(\boldsymbol{x}_n)$ is the hypersphere center, obtained by solving a constrained optimization problem, and the $\{\alpha_n\}_{n=1}^{N}$ are scalar values (Lagrange multipliers) subject to non-negativity and sum-to-one constraints that identify the sample data $\{\boldsymbol{x}_n|\alpha_n>0\}_{n=1}^{N}$ lying on the boundary of the support region (i.e., on the hypersphere surface) [101]. These sample data are called *support vectors* because they are the only sample data needed to describe the hypersphere. By kernelizing (36)

$$\begin{aligned} D(\boldsymbol{x}) = SVDD(\boldsymbol{x}) &= \mathrm{K}(\boldsymbol{x},\boldsymbol{x}) - 2\sum_{n=1}^{N}\alpha_n\mathrm{K}(\boldsymbol{x},\boldsymbol{x}_n) + \sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m\mathrm{K}(\boldsymbol{x}_n,\boldsymbol{x}_m) \\ &= const - 2\sum_{n=1}^{N}\alpha_n\mathrm{K}(\boldsymbol{x},\boldsymbol{x}_n). \end{aligned} \quad (37)$$

Equation (37) is very similar to (31) with the difference that the kernel functions that contribute to the sum are centered only on the support vectors.

d) Summary of Kernel-Based Methods: Each of the three kernel-based algorithms described above (FKDE, KRX, and SVDD) can be derived by mapping the data into a high-dimensional kernel-induced feature space $\boldsymbol{x}\rightarrow\Phi(\boldsymbol{x})$, expressing the algorithm in terms of dot products, and employing the kernel trick. Although derived following different approaches, the three algorithms turn out to be deeply interrelated. Both FKDE and KRX compute a centroid in the kernel-induced space $\boldsymbol{\mu}_\phi = \frac{1}{N}\sum_{n=1}^{N}\Phi(\boldsymbol{x}_n)$, whereas SVDD employs an adaptive center $\sum_{n=1}^{N}\alpha_n\Phi(\boldsymbol{x}_n)$. Both FKDE and SVDD compute anomalousness in terms of Euclidean distance to their respective centers, whereas KRX employs the Mahalanobis distance.

As noted in [78], "SVDD is based on a *discriminative* model, which does not assume any distribution for the input data. On the other hand, KRX is a *generative* model that represents the data as a Gaussian distribution in the kernel induced high-dimensional feature space." In this terminology, FKDE is also generative; it represents data as Gaussian in the kernel-induced space with a covariance matrix that is the identity matrix. A further distinction between KRX and FKDE is that the FKDE Gaussian is of full rank in the feature space, whereas KRX projects the data in the feature space to a lower dimensional subspace spanned by $[\Phi(\boldsymbol{x}_1),\Phi(\boldsymbol{x}_2),\ldots,\Phi(\boldsymbol{x}_N)]$ before computing the Mahalanobis distance. The variant of KRX originally introduced by [27] introduces a regularization term in the covariance matrix before taking its inverse, and effectively computes a distance in the full rank of the kernel-induced space.

It is worth mentioning that other detectors have been developed following the idea behind KRX and assuming a simple model in the kernel-induced space so as to obtain models that more tightly fit the background data in the original data space [24], [31], [49], [51], [120]. More recently, other nonparametric background models, such as manifold-based models, widely used in classification applications [54], have been gaining popularity also in detection applications [55], [128]. As with the kernel density estimators described in Section III-B3a, a perennial challenge for all kernel methods is choosing a good value for the kernel bandwidth [51], [97], [73].

4) Subspace Separation: Although it is widely observed that hyperspectral data are inadequately represented by MVN models, a number of authors have observed that the data can be more Gaussian in some directions than in others [4], [9], [110]. That is, the projection of pixel vectors $\boldsymbol{x}$ onto scalar values by a dot product with some vector $\boldsymbol{a}$, yields a distribution of scalars $\boldsymbol{a}^T\boldsymbol{x}$, which may be more or less Gaussian depending on the direction $\boldsymbol{a}$. This suggests the use of models based on a separation of hyperspectral data into two subspaces: 1) mostly Gaussian; and 2) mostly non-Gaussian. With a Gaussian/non-Gaussian (G/NG)-separation algorithm, one can explicitly produce a probability density that is a product of the MVN density in the Gaussian directions and a more tailored distribution in the non-Gaussian directions. A simplified graphical illustration of G/NG subspaces for the image shown in Fig. 2(a) is reported in Fig. 10(a) and (b), where the data projected onto two 2-dimensional subspaces exhibit strongly non-Gaussian and mostly Gaussian behaviors.

In [104], the non-Gaussian directions were modeled with a simplex, producing an ellipsoid-simplex hybrid. In [114], a nonparametric histogram-based model of the non-Gaussian directions is employed. The approach in [4] treated each dimension

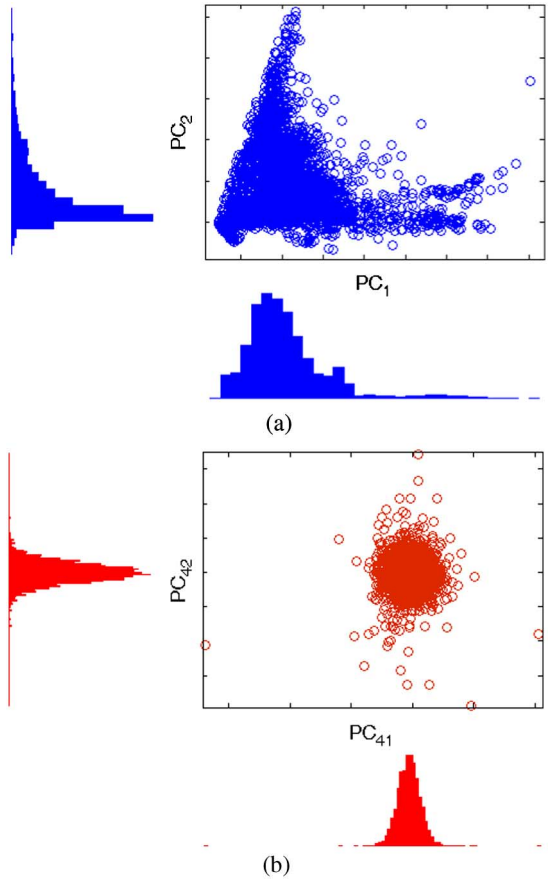


Fig. 10. Graphical example of G/NG subspace separation for the Reno urban image shown in Fig. 2(a). (a) Spectral domain with respect to two principal components where the data are strongly non-Gaussian. Univariate histrograms of the two principal components are plotted for convenience. (b) Spectral domain with respect to two principal components where the data are mostly Gaussian. Univariate histrograms of the two principal components are plotted for convenience.

of a principal components analysis as a separate subspace, and modeled those dimensions as individual EC distributions, with each direction getting its own measure of non-Gaussian-ness.

Another kind of subspace separation was described in [31], and referred to as a "folded subspace" in [91]. Here, the high-dimensional hyperspectral space is compacted into two dimensions corresponding to: 1) an MF projection of the data; and 2) the magnitude of the residual after MF subtraction. Although of reduced dimension, this space is rich enough to support a number of standard detectors (RX, MF, ACE, EC-GLRT, etc.) and furthermore permits new detectors, based on parametric and nonparametric modeling of the background in that lower-dimensional space.

C. Exploitation of Background Spatial Structure

As noted previously, there are both local and global versions for many of the background characterization algorithms described so far, with the local versions computing statistics from an annulus around the pixel of interest and global versions computing those statistics over the whole image. The advantage of the local versions is that they adapt to the changing statistics exhibited by spatially inhomogeneous backgrounds (and there is a multiple window variant that also adapts to different target sizes [52]), and as such implicitly incorporate the spatial structure of the background into the model. In this section, we will describe

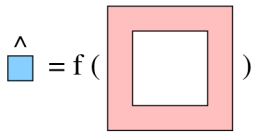


Fig. 11. Graphical representation of the exemplar-based background estimation concept. Background for the pixel under test is estimated as a general function of the pixels enclosed in the annulus. For many standard algorithms, the function $f(\cdot)$ is just the mean of the values in the annulus.

several more direct approaches for explicitly incorporating the background structure into the model.

1) Exemplar-Based Background Estimation: In general, local algorithms attempt to estimate the target-free value of a pixel in terms of pixels that surround it. For an algorithm such as RX, that estimate is based on the mean of pixels in an annulus. As also shown in the graphical example in Fig. 4, such local estimates can be effective, but they are challenged by edges and textures for the same reason that denoising by low-pass filtering (i.e., smoothing) is challenged by edges and textures.

Recent developments of image restoration techniques in traditional image processing [76] suggest an alternative approach that is more adaptive to the local image structure, and has the potential for substantially improving the estimate of the target-free (i.e., background) value at a pixel, and therefore enhancing the ability to detect whether a target is present. These *exemplar-based* methods identify patches in the image that are similar to each other (even though they may be well-separated on the image) and use them to collectively make inferences about each other. Problems such as denoising [23], [28], inpainting [123], [125], and super-resolution [126] are then treated as independent regression problems on these patches.

While the background estimation problem has not been directly addressed in mainstream image processing, this problem is closely related to denoising or inpainting, and is also amenable to attack via most of the standard exemplar-based techniques. Inspired by exemplar-based image processing, a regression framework was suggested in [103], in which the background for a pixel under test is estimated as a more general function of the annular pixels, as exemplified in Fig. 11. Specifically, this generic function is learned from the entire image and nonlocal algorithms are employed that exploit pixels from the k most similar annular patches in the image. This was shown to improve estimates of the central pixel for single-band, multispectral, and hyperspectral imagery. Adapting these exemplar-based approaches to the high-dimensional spaces encountered in hyperspectral imagery is an important challenge, but also a potential opportunity to more fully exploit both the spatial and spectral structure that is exhibited by hyperspectral backgrounds.

2) Standard Deviation MF: Pixels belonging to background structures such as edges between background classes have generally been found to be particularly susceptible to become false alarms [17], [89]. Simple detectors such as the MF may be slightly modified to mitigate false alarms induced by edges and background structures, using the Standard Deviation Matched Filter (SDMF) [17]

$$D(\boldsymbol{x}) = \left[(\boldsymbol{\Sigma} + \xi\bar{\boldsymbol{\Sigma}})^{-1}\boldsymbol{t}\right]^T \boldsymbol{C}_G{}^{-1}\left[(\boldsymbol{\Sigma} + \xi\bar{\boldsymbol{\Sigma}})^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right]. \quad (38)$$

Here, $\boldsymbol{C}_G$ is the background covariance matrix estimate computed globally over the whole image (other ways of locally and quasi-locally estimating the covariance matrix have been already discussed in Section III-A1), $\boldsymbol{\mu}$ is the local mean vector estimated over a spatial neighborhood of $\boldsymbol{x}$, $\boldsymbol{\Sigma}$ is a $d \times d$ diagonal matrix containing, for each spectral dimension, the standard deviation computed over a spatial neighborhood of $\boldsymbol{x}$. $\bar{\boldsymbol{\Sigma}}$ is a diagonal matrix that contains, for each spectral dimension, the average standard deviation computed over the whole image.

The multiplication by the matrix $\boldsymbol{\Sigma}^{-1}$ has been expressly designed to lower the detector output value in pixels characterized by high local standard deviations, i.e., in close proximity of background transitions. The term $\xi\bar{\boldsymbol{\Sigma}}$ has been introduced to avoid inflation of detector output value in areas of very low background variability (i.e., very low local standard deviations). In [17], it is specified that ξ should be chosen as a tradeoff between a small value, which allows to considerably reduce edge induced false alarms, and a large value, which makes targets on edges and background structures easier to detect. As highlighted in [17], the SDMF may be coupled with a more robust estimation of the local mean vector performed in a "directional" fashion. In fact, conventional local mean vector estimation would lead to misleading estimates for those pixels lying where statistically different background classes meet. By averaging a pair of pixels oriented symmetrically with respect to $\boldsymbol{x}$ and aligned along the direction characterized by the smallest variance, a more robust estimate should be obtained, at least in the case of transitions between two background classes [17].

3) Post-detection Spatial Analysis: False alarm pixels caused by background structures generally tend to be found in physical proximity to each other (i.e., tend to be grouped in clusters). There are a number of techniques that exploit this tendency, and by suitably filtering the gray-level image of detection results, can reduce false alarms. One approach [84] is to use morphological operators (such as combination of erosion and dilation operators) to reduce false alarms by eliminating, from the detector outcome, those groups of pixels characterized by detector high scores that cannot be associated with target-sized clusters. Another example is the *nonmaximum suppression* filter used in [26] that decreases the detector output based on the presence of higher scores in the local spatial neighborhood on a given pixel. Regardless the specific method used, post-detection spatial filters generally make use of *a priori* information about the expected target size.

IV. Conclusion

Given the huge variety of *actual* hyperspectral imagery, it is perhaps fitting that there is such a huge variety of *models* that have been proposed for hyperspectral image data. To characterize hyperspectral clutter—which is a high-dimensional and spatially inhomogeneous assortment of natural textures, man-made structures, and the various transitions between them—researchers have invoked kernels, manifolds, graphs, histograms, and exemplars; but they have also successfully employed low-dimensional subspaces and simple distributions.

We encourage the development of complex models for hyperspectral backgrounds, because those backgrounds are indeed complex. But we remind those researchers that complicated models may all-too-easily overfit the data, and what worked like a charm on Indian Pines might not fare so well on Jasper Ridge. We did not say much in this review about validation of hyperspectral models, but as models grow in complexity, reliable validation will grow in difficulty. As data acquisition rates continue to rise, so too will the raw size of hyperspectral datasets. Larger datasets are always welcomed by researchers, because they support better and more accurate models. But there has to be a practical tradeoff between sophisticated modeling and operational applicability. To be useful in real-world situations, the models should strive toward computational efficiency, ease of use by nonexperts, and robustness to the selection of free parameters.

The emphasis on background characterization over the more intuitively natural focus on the detection algorithms themselves represents a perspective that is becoming more prominent in the hyperspectral data analysis community. Although our ultimate interest is in effective detection of targets and anomalies, many detection algorithms can be understood in terms of how they model the background. In an attempt to keep the scope of this review bounded, we did not discuss change detection. But it is our conviction that the background modeling approaches described here can be productively applied in this problem domain as well.

Acknowledgment

The authors are enormously grateful to Prof. S. R. Rotman for his role in providing an impetus for this work and for intellectually stimulating conversations about background modeling in hyperspectral target detection applications.

References

[1] N. Acito, G. Corsini, and M. Diani, "Statistical analysis of hyper-spectral data: A non-Gaussian approach," *EURASIP J. Adv. Signal Process.*, vol. 2007, article no. 027673, 2007.

[2] N. Acito, G. Corsini, and M. Diani, "Adaptive detection algorithm for full pixel targets in hyperspectral images," *Proc. IEE Vis. Image Signal Process.*, vol. 152, no. 6, pp. 731–746, Dec. 2005.

[3] N. Acito, M. Diani, and G. Corsini, "On the CFAR property of the RX algorithm in the presence of signal-dependent noise in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, part 2, pp. 3475–3491, Jun. 2013.

[4] S. M. Adler-Golden, "Improved hyperspectral anomaly detection in heavy-tailed backgrounds," in *Proc. 1st IEEE Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS)*, 2009.

[5] J. A. Albano and D. W. Messinger, "Euclidean commute time distance embedding and its application to spectral anomaly detection," in *Proc. SPIE*, 2012, vol. 8390, p. 83902G.

[6] D. Avraham, N. Belogus, and S. R. Rotman, "Point target detection in segmented images," in *Proc. SPIE*, 2008, vol. 7086, p. 70860N.

[7] L. R. Bachega, J. Theiler, and C. A. Bouman, "Evaluating and improving local hyperspectral anomaly detectors," in *Proc. 40th IEEE Appl. Imagery Pattern Recognit. Workshop (AIPR)*, 2011.

[8] W. F. Baesner, "Clutter and anomaly removal for enhanced target detection," in *Proc. SPIE*, 2010, vol. 7695, p. 769525.

[9] P. Bajorski, "Maximum Gaussianity models for hyperspectral images," in *Proc. SPIE*, 2008, vol. 6966, p. 69661M.

[10] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.

[11] B. Basener, E. J. Ientilucci, and D. W. Messinger, "Anomaly detection using topology," in *Proc. SPIE*, 2007, vol. 6565, p. 65650J.

[12] O. Besson, S. Bidon, and J.-Y. Tourneret, "Bounds for estimation of covariance matrices from heterogeneous samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3357–3362, Jul. 2008.

[13] O. Besson, S. Bidon, and J.-Y. Tourneret, "Covariance matrix estimation with heterogeneous samples," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 909–920, Mar. 2008.

[14] N. E. Billor, A. S. Hadib, and P. F. Velleman, "BACON: Blocked adaptive computationally efficient outlier nominators," *Comput. Statist. Data Anal.*, vol. 34, pp. 279–298, 2000.

[15] C. M. Bishop, *Pattern Recognition*. New York, NY, USA: Springer-Verlag, 2006.

[16] C. E. Caefer, J. Silverman, O. Orthal, D. Antonelli, Y. Sharoni, and S. R. Rotman, "Improved covariance matrices for point target detection in hyperspectral data," *Opt. Eng.*, vol. 47, no. 7, p. 076402, Jul. 2008.

[17] C. E. Caefer, M. S. Stefanou, E. D. Nielsen, A. P. Rizzuto, O. Raviv, and S. R. Rotman, "Analysis of false alarm distributions in the development and evaluation of hyperspectral point target detection algorithms," *Opt. Eng.*, vol. 46, no. 7, p. 076402, Jul. 2007.

[18] N. A. Campbell, "Robust procedures in multivariate analysis I: Robust covariance estimation," *Appl. Statist.*, vol. 29, pp. 231–237, 1980.

[19] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal components?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, May 2011.

[20] G. Cao and C. A. Bouman, "Covariance estimation for high dimensional data vectors using the sparse matrix transform," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2008, vol. 21, pp. 225–232.

[21] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.

[22] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.

[23] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1635–1649, Apr. 2012.

[24] C. Chen, N. M. Nasrabadi, and T. D. Tran, "Kernel sparse representation for hyperspectral target detection," in *Proc. SPIE*, 2012, vol. 8390, p. 839005.

[25] J. Y. Chen and I. S. Reed, "A detection algorithm for optical targets in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, no. 1, pp. 394–405, Jan. 1987.

[26] Y. Cohen, D. G. Blumberg, and S. R. Rotman, "Subpixel hyperspectral target detection using local spectral and spatial information," *J. Appl. Remote Sens.*, vol. 6, no. 1, p. 063508, Jan. 2012.

[27] D. Cremers, T. Kohlberger, and B. Schölkopf, "Shape statistics in kernel space for variational image segmentation," *Pattern Recognit.*, vol. 36, no. 9, pp. 1929–1943, Sep. 2003.

[28] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[29] C. E. Davidson and A. Ben-David, "Performance loss of multivariate detection algorithms due to covariance estimation," in *Proc. SPIE*, 2009, vol. 7477, p. 74770J.

[30] M. T. Eismann, A. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.

[31] B. R. Foy, J. Theiler, and A. M. Fraser, "Decision boundaries in two dimensions for target detection in hyperspectral imagery," *Opt. Express*, vol. 17, no. 20, pp. 17391–17411, 2009.

[32] J. H. Friedman, "Regularized discriminant analysis," *J. Am. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, Mar. 1989.

[33] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego, CA, USA: Academic, 1990.

[34] C. C. Funk, J. Theiler, D. A. Roberts, and C. C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1410–1420, Jul. 2001.

[35] K. Gerlach, "The effects of signal contamination on two adaptive detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 297–309, Jan. 1995.

[36] N. Gorelik, D. Blumberg, S. R. Rotman, and D. Borghys, "Non singular approximations for a singular covariance matrix," in *Proc. SPIE*, 2012, vol. 8390, p. 839021.

[37] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[38] A. Hayden, E. Niple, and B. Boyce, "Determination of trace-gas amounts in plumes by use of orthogonal digital filtering of thermal-emission spectra," *Appl. Opt.*, vol. 35, no. 16, pp. 2802–2809, Jun. 1996.

[39] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 763–767, Jul. 1996.

[40] SpecTIR Advanced Hyperspectral and Geospatial Solutions, *Free data samples* [Online]. Available: http://www.spectir.com/free-data-samples/

[41] B. R. Hunt and T. M. Cannon, "Nonstationary assumptions of Gaussian models of images," *IEEE Trans. Syst. Man Cybern.*, vol. 6, no. 12, pp. 876–88, Dec. 1976.

[42] P. C. Hytla, R. C. Hardie, M. T. Eismann, and J. Meola, "Anomaly detection in hyperspectral imagery: Comparison of methods using diurnal and seasonal data," *J. Appl. Remote Sens.*, vol. 3, no. 1, pp. 033546-1–033546-30, Jan. 2009.

[43] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice Hall, 1998.

[44] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 2, pp. 115–127, Mar. 1986.

[45] J. P. Kerekes and D. Manolakis, "Improved modeling of background distributions in an end-to-end spectral imaging system model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2004, pp. 972–975.

[46] S. Khazai, A. Safari, B. Mojaradi, and S. Homayouni, "An approach for subpixel anomaly detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, part 3, pp. 769–778, Apr. 2013.

[47] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Process.*, vol. 53, no. 2, part 1, pp. 427–438, Feb. 2005.

[48] B. C. Kuo and D. A. Landgrebe, "A covariance estimator for small sample size classification problems and its application to feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 814–819, Apr. 2002.

[49] H. Kwon and N. M. Nasrabadi, "A comparative analysis of kernel subspace target detectors for hyperspectral imagery," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, article no. 029250, Jan. 2007.

[50] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.

[51] H. Kwon and P. Gurram, "Optimal kernel bandwidth estimation for hyperspectral kernel-based anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2010, pp. 2812–2815.

[52] W.-M. Liu and C.-I. Chang, "Multiple-window anomaly detection for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, part 2, pp. 644–658, Apr. 2013.

[53] E. Lo and J. Ingram, "Hyperspectral anomaly detection based on minimum generalized variance method," in *Proc. SPIE*, 2008, vol. 6966, p. 696603.

[54] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.

[55] L. Ma, M. M. Crawford, and J. Tian, "Anomaly detection for hyperspectral images based on robust locally linear embedding," *J. Infrared Millimeter Terahertz Waves*, vol. 31, no. 6, pp. 753–762, Jun. 2010.

[56] E. Madar, O. Kuybeda, D. Malah, and M. Barzohar, "Local-global background modeling for anomaly detection in hyperspectral images," in *Proc. IEEE 1st Workshop Hyperspecral Image Signal Process.: Evol. Remote Sens. (WHISPERS)*, 2009.

[57] D. G. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.

[58] D. Manolakis, "Taxonomy of detection algorithms for hyperspectral imaging applications," *Opt. Eng.*, vol. 44, no. 6, p. 066403, Jun. 2005.

[59] D. Manolakis, "Realistic matched filter performance prediction for hyperspectral target detection," *Opt. Eng.*, vol. 44, no. 11, p. 116401, 2005.

[60] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. Mag.*, vol. 14, no. 1, pp. 79–116, 2003.

[61] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 24–33, Jan. 2014.

[62] D. Manolakis, M. Rossacci, J. Cipar, R. Lockwood, T. Cooley, and J. Jacobson, "Statistical characterization of hyperspectral background clutter in the reflective spectral region," *Appl. Opt.*, vol. 47, no. 28, pp. F96–F106, 2008.

[63] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Hyperspectral algorithms: Use covariances or subspaces?" in *Proc. SPIE*, 2009, vol. 7457, p. 74570Q.

[64] D. Marden and D. Manolakis, "Modeling hyperspectral imaging data," in *Proc. SPIE*, 2003, vol. 5093, pp. 253–262.

[65] D. B. Marden and D. Manolakis, "Using elliptically contoured distributions to model hyperspectral imaging data," in *Proc SPIE*, 2004, vol. 5425, pp. 558–572.

[66] A. Margalit, I. S. Reed, and R. M. Gagliardi, "Adaptive optical target detection using correlated images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 21, no. 3, pp. 46–59, May 1985.

[67] S. Matteoli, M. Diani, and G. Corsini, "Effects of signal contamination in RX detection of local hyperspectral anomalies," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2012, vol. 1, pp. 4845–4848.

[68] S. Matteoli, M. Diani, and G. Corsini, "A Kurtosis-based test to efficiently detect targets placed in close proximity by means of local covariance-based hyperspectral anomaly detectors," in *Proc. IEEE 3rd Workshop Hyperspecral Image Signal Process.: Evol. Remote Sens. (WHISPERS)*, 2011.

[69] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[70] S. Matteoli, M. Diani, and G. Corsini, "Hyperspectral anomaly detection with kurtosis-driven local covariance matrix corruption mitigation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 532–536, May 2011.

[71] S. Matteoli, M. Diani, and G. Corsini, "Impact of signal contamination on the adaptive detection performance of local hyperspectral anomalies," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 1948–1968, Apr. 2014.

[72] S. Matteoli, M. Diani, and G. Corsini, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Opt. Eng.*, vol. 49, no. 4, p. 046201, 2010.

[73] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "Background density nonparametric estimation with data-adaptive bandwidths for the detection of anomalies in multi-hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 163–167, Jan. 2014.

[74] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "Models and methods for automated background density estimation in hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2837–2852, May 2013.

[75] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "A locally adaptive background density estimator: An evolution for RX-based anomaly detectors," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 323–327, Jan. 2014.

[76] P. Milanfar, "A tour of modern image filtering," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, Jan. 2013.

[77] J. W. V. Miller, J. B. Farison, and Y. Shin, "Spatially invariant Image sequences," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 148–161, Apr. 1992.

[78] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.

[79] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," in *Proc. SPIE*, 2008, vol. 6966, p. 696604.

[80] E. Parzen, "On the estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, 1962.

[81] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.

[82] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, no. 6, pp. 853–863, Nov. 1974.

[83] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–218, Jan. 1992.

[84] J. M. Romano, D. Rosario, and E. Niver, "Morphological operators for polarimetric anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 664–677, Feb. 2014.

[85] P. J. Rousseeuw and K. Van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[86] S. R. Sain, "Multivariate locally adaptive density estimation," *Comput. Statist. Data Anal.*, vol. 39, no. 2, pp. 165–186, 2002.

[87] L. L. Scharf and B. Frielander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[88] L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Proc. 30th Asilomar Conf. Signals Syst. Comput.*, 1996, pp. 1114–1117.

[89] A. Schaum, "A remedy for non-stationarity in background transition regions for real time hyperspectral detection," in *Proc. IEEE Aerosp. Conf.*, 2006.

[90] A. Schaum, "Advanced hyperspectral detection based on elliptically contoured distribution models and operator feedback," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop (AIPR)*, 2009.

[91] A. Schaum, "Design methods for continuum fusion detectors," in *Proc. SPIE*, 2011, vol. 8048, p. 804803.

[92] A. Schaum, "Spectral subspace matched filtering," in *Proc. SPIE*, 2001, vol. 4381, pp. 1–17.

[93] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[94] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[95] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.

[96] H. B. Segev, S. R. Rotman, and D. G. Blumberg, "Detecting anomalous objects in hyperspectral data using segmentation," in *Proc. SPIE*, 2008, vol. 7113, p. 711314.

[97] B. W. Silverman, *Density Estimation for Statist. and Data Analysis*. London, U.K.: Chapman and Hall, 1986.

[98] K. W. Smetek and T. E. Bauer, "Finding hyperspectral anomalies using multivariate outlier detection," in *Proc. IEEE Aerosp. Conf.*, 2007, vol. 1, pp. 1–24.

[99] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal. Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[100] B. Stevenson, R. O'Connor, W. Kendall, A. Stocker, W. Schaff, D. Alexa, J. Salvador, M. Eismann, K. Barnard, and J. Kershenstein, "Design and performance of the civil air patrol ARCHER hyperspectral processing system," in *Proc. SPIE*, Jul. 2005, vol. 5806, pp. 731–742.

[101] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, no. 11–13, pp. 1191–1199, Nov. 1999.

[102] G. R. Terrell and D. W. Scott, "Variable kernel density estimation," *Ann. Statist.*, vol. 20, pp. 1236–1265, 1992.

[103] J. Theiler and B. Wohlberg, "Regression framework for background estimation in remote sensing imagery," in *Proc. 5th IEEE Workshop Hyperspecral Image Signal Process.: Evol. Remote Sens. (WHISPERS)*, 2013.

[104] J. Theiler, "Ellipsoid-simplex hybrid for hyperspectral anomaly detection," in *Proc. 3rd IEEE Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS)*, 2011.

[105] J. Theiler, "The incredible shrinking covariance estimator," in *Proc. SPIE*, 2012, vol. 8391, pp. 83910P.

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

[107] J. Theiler and B. R. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 98–102, Jan. 2006.

[108] J. Theiler and B. Wohlberg, "Detection of spectrally sparse anomalies in hyperspectral imagery," in *Proc. IEEE Southwest Symp. Image Anal. Interpret. (SSIAI)*, 2012, pp. 117–120.

[109] J. Theiler and D. Hush, "Statistics for characterizing data on the periphery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2010, pp. 4764–4767.

[110] J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," in *Proc. SPIE*, 2005, vol. 5806, pp. 182–193.

[111] J. Theiler, B. R. Foy, and A. M. Fraser, "Nonlinear signal contamination effects for gaseous plume detection in hyperspectral imagery," in *Proc. SPIE*, 2006, vol. 6233, p. 62331U.

[112] J. Theiler, C. Scovel, B. Wohlberg, and B. R. Foy, "Elliptically contoured distributions for anomalous change detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 271–275, Apr. 2010.

[113] J. Theiler, G. Cao, L. R. Bachega, and C. A. Bouman, "Sparse matrix transform for hyperspectral image processing," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 424–437, Jun. 2011.

[114] G. A. Tidhar and S. R. Rotman, "Target detection in inhomogeneous non-Gaussian hyperspectral data based on nonparametric density estimation," in *Proc SPIE*, 2013, vol. 8743, p. 87431A.

[115] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[116] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "Fully unsupervised learning of Gaussian mixtures for anomaly detection in hyperspectral imagery," in *Proc. IEEE Int. Conf. Intell. Syst. Des. Appl. (ISDA)*, 2009, pp. 596–601.
[117] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "An anomaly detection architecture based on a data-adaptive density estimation," in *Proc. 3rd IEEE Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS)*, 2011.
[118] T. Veracini, S. Matteoli, M. Diani, G. Corsini, and S. U. De Ceglie, "A nonparametric approach to anomaly detection in hyperspectral images," in *Proc. SPIE*, 2010, vol. 7830, p. 78300B.
[119] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "Nonparametric framework for detecting spectral anomalies in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 666–670, Jul. 2011.
[120] T. Wang, B. Du, and L. Zhang, "A kernel-based target-constrained interference-minimized filter for hyperspectral sub-pixel target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, part 2, pp. 626–637, Apr. 2013.
[121] J. E. West, D. W. Messinger, and J. R. Schott, "Comparative evaluation of background characterization techniques for hyperspectral unstructured matched filter target detection," *J. Appl. Remote Sens.*, vol. 1, no. 1, pp. 1–15, Jan. 2007.
[122] C. J. Willis, "Hyperspectral target detection using heavy-tailed distributions," in *Proc. SPIE*, 2009, vol. 7482, p. 74820P.
[123] B. Wohlberg, "Inpainting by joint optimization of linear combinations of exemplars," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 75–78, Jan. 2011.
[124] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2009, vol. 22, pp. 2080–2088.
[125] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1153–1165, May 2010.
[126] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
[127] J. Yin, C. Gao, and X. Jia, "A new target detector for hyperspectral data using cointegration theory," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, part 2, pp. 638–643, Apr. 2013.
[128] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

Stefania Matteoli (S'07–M'11) received the B.S. and M.S. *(cum laude)* degrees in telecommunications engineering and the Ph.D. degree in remote sensing, all from the University of Pisa, Pisa, Italy, in 2003, 2006, and 2010, respectively.

She is currently a Post-Doctoral Researcher with the Department of Information Engineering, University of Pisa, Pisa, Italy, where she has been teaching laboratory classes of "Design and Simulation of Remote Sensing Systems." From May to October 2008, she was a Visiting Student with the Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA. Her research interests include signal and image processing, anomaly and target detection methodologies for multi-hyperspectral remotely sensed images, and physics-based forward modeling methodologies for spectral signature based target detection in hyperspectral images.

Dr. Matteoli has served as a reviewer for journals, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Journal of Applied Remote Sensing*, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS.

Marco Diani (M'93) received the Laurea degree *(cum laude)* in electronic engineering from the University of Pisa, Pisa, Italy, in 1988.

He is currently an Associate Professor with the Department of Information Engineering, University of Pisa, where he has been teaching "Statistical Signal Theory" and "Design and Simulation of Remote Sensing Systems." His works covered different topics such as data fusion, signal processing in imaging radars, image classification and segmentation, object detection and tracking in infrared image sequences, and target detection and recognition in multi-hyperspectral images. His research interests include image and signal processing with application to remote sensing.

Prof. Diani has served as a reviewer for many journals in the fields of remote sensing and image/signal processing, such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *Applied Optics*. He is a Member of SPIE.

James Theiler received the Ph.D. degree in physics from California Institute of Technology, Pasadena, CA, USA, in 1987, with a thesis on statistical and computational aspects of identifying chaos in time series.

He followed a nonlinear trajectory to the University of California San Diego, USA, MIT Lincoln Laboratory, Lexington, MA, USA, Los Alamos National Laboratory, Los Alamos, NM, USA, and the Santa Fe Institute, Santa Fe, NM, USA. His enthusiasm for algorithmic data analysis and his eagerness to have a real job were combined in 1994, when he joined the Space and Remote Sensing Sciences Group, Los Alamos, NM, USA. In 2005, he was named a Los Alamos Laboratory Fellow. His research interests include statistical modeling, machine learning, image processing, and remote sensing.