ARMY RESEARCH LABORATORY

Alternative Asymmetric Hypothesis Tests for Hyperspectral Imagery

by Dalton Rosario

ARL-TR-3712 **February 2006**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Citation of manufacturers' or trade names does not constitute an official endorsement or approval of the use thereof.

Part of the material in ARL-TR-3712 is the core for a patent application.
Patent Application Number 11/171,366; Dated on 08-17-2005.

DESTRUCTION NOTICE—Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Adelphi, MD 20783-1145

ARL-TR-3712 **February 2006**

Alternative Asymmetric Hypothesis Tests for Hyperspectral Imagery

Dalton Rosario
Sensors and Electron Devices Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
February 2006	Final	FY05

4. TITLE AND SUBTITLE
Alternative Asymmetric Hypothesis Tests for Hyperspectral Imagery

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dalton Rosario

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Sensors and Electron Devices Directorate, ARL
(ATTN: AMSRD-ARL-SE-SE) rosario@arl.army.mil
Adelphi, MD 20783-1145

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-3712

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Adelphi, MD 20783-1145

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report focuses on the development of statistical anomaly detection techniques aimed at accentuating the presence of meaningful objects, e.g., a land vehicle, as a collection of localized anomalies in reference to a scene dominated by natural clutter backgrounds. The report presents a significant improvement in anomaly detection performance via a principle of indirect comparison, where samples are not compared to each other as individual entities, but as individual entities compared to the union of these entities. Let X and Y denote two random samples, and let Z = X U Y, where U denotes the union. X can be indirectly compared to Y by comparing instead some of the corresponding distribution attributes of Z and Y. This idea led to the development of four unconventional techniques for hyperspectral (HS) anomaly detection. The first technique is based on some of the advances made on semiparametric (large sample) inference, where a logistic model and its maximum likelihood method are presented, along with the analysis of its asymptotic behavior. The second technique is based on fundamental theorems from large sample theory and is developed to approximate performance of the former technique, albeit free from its implementation drawbacks. A third detector is developed, as an alternative, using the same principle and a known property of the F-distribution family. The introduction of this third detector, which has an asymptotic F-distribution behavior, was motivated by the classic one-way ANOVA, which under its null hypothesis and the normality assumption has a test statistic governed by an exact F distribution. Finally, a fourth and significantly more compact technique is developed based on an asymmetric variance test between the union of two samples and one of the individual samples. Like the previous three techniques, this compact form is free from distribution assumptions, although under its null hypothesis the test statistic converges to a known distribution. Theoretical analyses are shown for the power of the test applying these new detectors to two types of problems: (i) local anomaly detection through the perspective of a top view, where size uncertainties of known objects are not an issue, and (ii) scene anomaly detection through the perspective of a ground-level view, where uncertainties of objects' sizes and shapes are major issues. Experimental results using real HS data are presented to illustrate the effectiveness of the new detectors over conventional techniques.

15. SUBJECT TERMS hyperspectral, anomaly detection, asymmetric hypothesis test

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dalton Rosario
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	32	19b. TELEPHONE NUMBER (*Include area code*) (301) 394-4235

Standard Form 298 (Rev. 8/98)

Contents

Figures

1. Introduction

1.1 Background

In the context of machine vision, it would be desired to have a relatively simple automatic approach capable of focusing its attention in the same way a human analyst would observing the same set of images. For reasons well known and highlighted in standard computer-vision books (see, for instance, [1]), that expectation, however, is rarely met with experimental results, despite of the fact that sometimes the scenes in reference are characterized by image analysts as *easy* to focus their attentions to certain types of objects.

Humans, of course, use a combination of knowledge-based, local and global information to aid in the analysis of a scene, a capability maybe reproduced by applying, for instance, layers of unsupervised learning methods complementing each other to perform this single task. For example, a suite of algorithms that includes an edge detector, an edge elongation, a clustering method, and a morphological size test might reproduce the humans' performance in certain conditions, albeit with a huge cost: computational time. Needless to say, the topic of achieving meaningful automatic focus of attention (FOA) is an open and quite active area of research [1].

I seek to achieve humans' performance using a single unsupervised learning algorithm. Unsupervised learning algorithms, contrary to supervised learning methods, do not require a priori information of targets (objects of interest) and of non-targets for training purposes. Examples of unsupervised learning methods are anomaly detection algorithms, and examples of supervised learning methods are artificial neural networks.

To accomplish our goal, I opted to use hyperspectral (HS) rather than broadband imagery, and to focus our algorithmic development on adaptive anomaly detection rather than on a particular type of material detection, also known as target detection.

Hyperspectral sensors are passive sensors that simultaneously record images for many contiguous and narrowly spaced regions of the electromagnetic spectrum. In the context of FOA, especially at the ground-level view, this property would eliminate uncertainties of objects' sizes and shapes—a tremendous advantage over broadband sensors operating in the same region of the electromagnetic spectrum [2]–[4].

Our reason for choosing anomaly detection over target detection is that often the exact material of interest is not known a priori, or the number of spectra in a material of interest library is simply too exhaustive to search for all possible materials. The goal of an anomaly detector is to identify outliers, i.e., data points that are atypical compared to the rest of the data. An anomaly detector that properly detects all, or a significant portion, of the pixels representing meaningful objects (targets) while at the same time having hundreds of meaningless detections (false alarms) has little practical value.

1.2 Prior Art

I present in this section the more important results from our literature research on the topic of anomaly or target detection using hyperspectral imagery.

It is quite evident from the literature that due to the large amounts of data that are collected with hyperspectral sensors, much of the prior work has focused strictly on compression of the data sets for storage and transmission. More recently, work has been published about reducing the number of spectral bands used for processing detection and classification data. These algorithms fall into two basic categories: spectral-only and spatial-spectral algorithms. The spectral-only algorithms almost all rely on a known spectral signature for the target or targets of interest; they are generally geared to perform classification rather than detection tasks. Algorithms that fall into this category are the spectral matched filter by Crist, et al., [5], the spectral angle mapper by Haskett and Sood [6], and linear mixture models by Grossmann, et al., [7], Chang, et al., [8], and Slater, et al., [9]. The main limitation of these spectral-only algorithms is that they require a known target signature. Reliable target signatures are difficult to ascertain due to variations in the target signature that result from atmospheric and illumination effects.

Spatial-spectral algorithms can be further divided into local anomaly and global anomaly detectors. Local anomaly detectors, as mentioned earlier, process small windows of the HS data in order to compare the spatial and spectral properties of the centrally located pixels in the window with the properties of the surrounding pixels. Those pixels that are spatially-spectrally different from their surrounding backgrounds are considered detections. Yu, et al., [10] proposed an algorithm commonly referred to as RX algorithm, which has become a benchmark for multispectral data, based on this principle. The RX algorithm is a maximum likelihood (ML) anomaly detection procedure that simplifies the clutter to being spatially white. Researchers have also used classical approaches, such as, Fisher's Linear Discriminant, Principal Component Analysis (PCA), in the same spirit of the RX algorithm. PCA has been mostly used prior to another detection or classification algorithm for purposes of reducing the dimensionality of the hyperspectral data sets, thus making the applied detection and classification algorithms computationally efficient. The reduction of redundant information with PCA is based on reconstructing the data using a subset of the principal components. Often, the components used for reconstruction are those associated with largest eigenvalues. Crist, et al., [5] have shown, however, that components associated with lower order eigenvalues often contain important features for target discrimination. Thus, there is ambiguity as to what are the most appropriate principal components to use for data reduction.

I will present later a brief insightful discussion on some of the more prominent local anomaly detectors and their performances; this discussion will also include a recently published detector, the kernel RX [11].

In global anomaly detectors, the image scene is first segmented into its constituent classes. Detection then is achieved by determining the outliers of these classes. In general, the algorithms vary in the method of segmentation, but tend to use maximum

likelihood (ML) detection once the classes are determined. The number of classes is assumed to be known a priori for these algorithms, which is also a weakness. Stocker [12] discusses one of these hybrid algorithms—the stochastic expectation maximization clustering algorithm (see also Masson and Pieczynski [13]) coupled with ML-detection.

From the discussion above, it is evident that most conventional anomaly or target detectors use multivariate models to define the spectral variability of the data, and the majority of the data pixels are assumed to be spectrally homogeneous and are modeled using a multivariate probability density function with a single set of parameters. Until now, no significant work had been done to find non-normal statistical models, or unconventional alternatives for the development of anomaly detection techniques geared toward hyperspectral data. As I will show in this work, conventional anomaly detectors may detect the presence of targets in hyperspectral data, but in the process they yield a large number of false alarms. This sort of performance has little practical value.

1.3 A Notional Breakthrough

I present in this section a breakthrough in this research that led to the developments shown in this report.

1.3.1 Most Probable Study Cases in Anomaly Detection

To gain a better insight into the general behavior of local anomaly detectors, I decomposed their expected performances into what I considered to be the three most probable study cases (to be discussed shortly), and applied a few conventional anomaly detectors to actual HS imagery in order to compare their local responses with this decomposition model. From this comparison I made a simple but important discovery and a key recognition.

Discovery: The reason conventional techniques produce high numbers of meaningless detections in digitized scenes is not only because the assumed data models are unrealistic, but also because these techniques are not developed to address—explicitly—all three of the most probable spatial/spectral variability occurrences observed locally in the imagery. What this claim really means is that improving data models for various object classes will not necessarily improve performances of anomaly detectors based on those improved models.

To appreciate this claim, consider the decomposition of anomaly detection problems into three most probable study cases: Case 1, Case 2, and Case 3, where Case 1 represents a comparison between two sample sets from different distributions (e.g., land vehicle and grass); Case 2 represents a comparison between a two-material sample set and a sample set representing one of the two materials (e.g., a spatial transition between tree shadows and surrounding grass), and Case 3 represents a comparison between two sample sets from the same distribution (e.g., grass and grass). Using this simple decomposition model to judge the quality of the detectors' results revealed to us that the *application* of conventional techniques to local anomaly detection problems using digitized scenes is essentially *flawed*. Conventional

detectors are developed to account explicitly for Case 1 and Case 3, but not for one of the more abundant cases—Case 2. Case 2 occurs often on digitized scenes, representing major transitions of regions, or strong edges. With anomaly detection, the consequence of not accounting for Case 2 arguably leads to a significant increase of meaningless detections (e.g., edges between tree shadows and surrounding grass), often obscuring the locations of meaningful detections (e.g., a land vehicle parked in a road). This observation applies to conventional techniques based on parametric or nonparametric approaches, as it will be shown shortly. In summary, I claim that a significant performance improvement of anomaly detectors will not be achieved by proposing more accurate HS data models, but rather by proposing methods that can account, in some form, for at least all three most probable study cases discussed in this section. *Accounting* in this context means being able to accentuate a response categorized as Case 1 and suppress responses categorized as Cases 2 and 3.

Key recognition: Our discovery led to a key recognition after taking a closer look at Case 2. Case 2 (the case study that yields anomalous responses along with Case 1) is equivalent to comparing the union of two distinct sample sets with one of these sets. This recognition infers that two sample sets may be indirectly and effectively compared in the context of anomaly detection by comparing instead the union of both sets with one of the two sets. Our discovery and key recognition led us to a principle, to be discussed shortly, and served as a breakthrough in the developments presented in this report.

1.3.2 Principle of Indirect Comparison

I propose a plausible idea for the development of anomaly detection algorithms that accounts for all three study cases: Compare samples indirectly by combining them, i.e., compare samples not as individual entities, but as individual entities and the union among these entities. Let X and Y denote two random samples. Let X be reference sample and let $Z = X \cup Y$, where U denotes the union. Features of the distribution of X can be indirectly compared to features of the distribution of Y by comparing instead features of the distributions of Z and X. I will show that anomaly detection algorithms, based on this principle, enjoy the desirable outcome of preserving what is often characterized by image analysts as meaningful detections (e.g., a manmade object in an open terrain), while significantly reducing the number of meaningless detections (e.g., transition of different regions).

1.3.3 Proof of Principle Simulation

This subsection explains the advantage of applying the principle of indirect comparison to anomaly detection problems.

Figure 1 shows simulated realizations of random samples and their corresponding empirical distributions. A random sample, by definition [14], is a sequence of random variables, e.g., $X = (X_1, X_2, \ldots, X_n)$, where X_i is independent of X_j $(i \neq j)$. Our focus is on two study cases, labeled *Case 1* and *Case 2* in figure 1, where *Case 1* depicts the realization of two random samples from different distributions, and *Case 2* depicts the realization of a composite sample and a pure one.

The plot for *Case 1* under SIMULATION shows the simulated realizations of two random samples, X and Y; their random variables are Normally distributed having the same variance (σ^2), but significantly different means, *100* and *20*, or $X_i \sim N(100, \sigma^2)$ and $Y_i \sim N(20, \sigma^2)$, where X_i and Y_i are random variables of X and Y, respectively, and $i = 1, \ldots, 100$. The vertical axis represents realized values and the horizontal axis represents index i. The plot for *Case 2* under SIMULATION shows the simulated realizations of an additional random sample, S, which is composed of a sequence from two Normal distributions, $S \sim [N(100, \sigma^2)$ or $N(20, \sigma^2)]$, and the same realization of Y. Let X and S be reference samples and Y be a test sample. Then, compare X to Y using the conventional way (i.e., comparing samples as individual

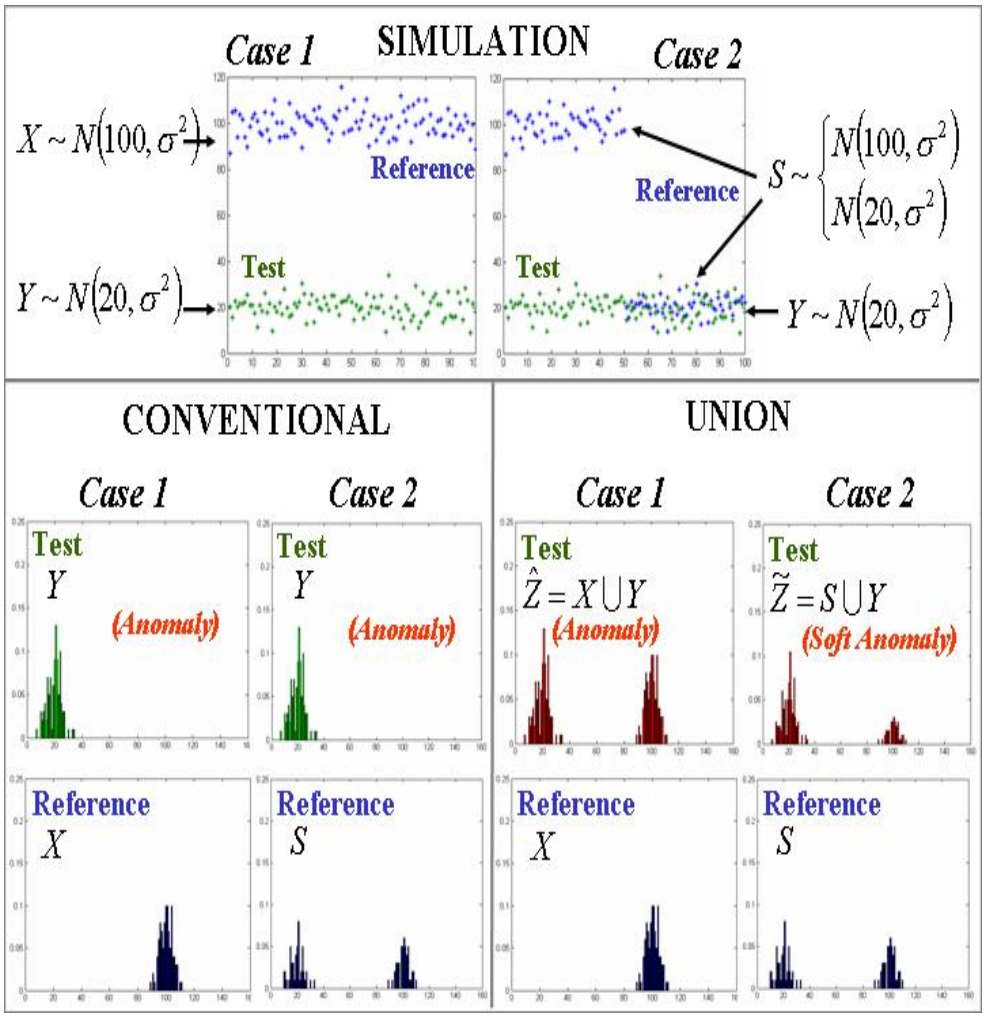

Figure 1. A principle of indirect comparison: the number of meaningless detections (Case 2) may be significantly reduced by comparing, instead, the union of candidate samples against a reference candidate sample. Another desirable outcome using this principle is that the number of meaningful detections (Case 1) is preserved.

entities) and using our proposed form (i.e., comparing individual entities to the union of entities), and repeat that comparison between *S* and *Y*. Comparison between

random samples often implies a comparison among the moments and/or central moments of their distributions. Thus, I computed the *empirical* distributions (normalized histograms) of these random samples and invite the reader to perform those comparisons by visual inspection.

In figure 1, under CONVENTIONAL, *Case 1,* the empirical distributions of the test sample *Y* and the reference sample *X* are shown. Both empirical distributions resemble a relatively tight Gaussian distribution, having the same variance but centered at different means, as expected. By visual inspection, one would expect that statistical methods using the conventional way would be able to distinguish the distribution of *Y* from the distribution of *X* in the basis of their mean difference, where *Y* would likely be declared as an anomaly with respect to *X*. In *Case 2,* under CONVENTIONAL, by visual inspection alone, one would also expect *Y* to be declared as an anomaly in respect to *S* for the obvious fact that the bimodal distribution of *S* is quite different from the unimodal distribution of *Y*. Correct as this declaration may be, it is also unfortunate, because these two study cases are often found together in real image processing problems. For instance, in a real scenario, *Case 1* could represent a comparison between a random sample *X* from a motor vehicle and a sample *Y* from the surrounding natural terrain. Similarly, *Case 2* could represent a comparison between a composite sample (*S*) from a transition of regions (terrain and tree shadow) and the pure sample (*Y*) from terrain. Based on this rationale, a conventional anomaly detector may not be able to distinguish between *Case 1* and *Case 2*. Furthermore, in many circumstances, it may even declare *Case 2* a stronger anomaly than *Case 1*, yielding instead results that are more comparable to those of edge detectors.

Under UNION in figure 1, visual inspection should convince the reader that the empirical distribution of the sample union $\hat{Z}$, which is bimodal in *Case 1*, is quite different from the corresponding unimodal distribution of *X.* This fact shall preserve the desirable declaration that *X* and *Y* are samples from different distributions. What shall not be preserved under UNION, however, is the unfortunate outcome under CONVENTIONAL, *Case 2*, since the empirical distributions of the sample union $\tilde{Z}$, in *Case 2,* and the composite sample *S* have the same general characteristics: they are bimodal. Therefore, under UNION, one would expect the differences between *X* and *Y* in *Case 1* to be accentuated and the differences between *S* and *Y* in *Case 2* to be suppressed, as desired.

Another study case, where both reference and test samples belong to the same distribution, was not included because the outcomes under both CONVENTIONAL and UNION are expected to be trivial and comparable.

1.4 Organization of Report

This report will focus on the development of statistical anomaly detection techniques aimed at accentuating the presence of meaningful objects, e.g., a land vehicle, as a collection of point anomalies in reference to a scene dominated by natural clutter

backgrounds. These techniques will exploit the principle of indirect comparison discussed in Section 1.

Section 2 discusses well known conventional techniques and their applications to an anomaly detection problem using actual HS data.

In Section 3, I formulate two types of remote sensing, anomaly detection problems and propose four unconventional anomaly detection techniques to address these problems. They will be referred in this report as: Semiparametric (SemiP), Approximation to SemiP (AsemiP), Asymptotic F-distribution Test (AFT), and Asymmetric Variance Test (AVT). The SemiP detector is based on a logistic model, having as inputs independent, identically distributed observations (iid), and on large sample theory. The utility of a logistic model indicates this approach is neither parametric nor nonparametric, but semiparametric. The implementation of the SemiP technique requires an unconstrained maximization subroutine, which depends on parameter initialization. I will show that to circumvent this dependence and still benefit from the effectiveness of the SemiP test statistic, the AsemiP technique will be developed. The AsemiP detector is also free from distribution assumption, although I will show that, under its test null hypothesis, its test statistic tends in law to a known distribution, as the number of samples increases. The AFT detector was developed as an alternative unconventional detector, tending in law to an F-distribution, under a null hypothesis, as the number of samples increases. The classic one-way ANOVA (analysis of variance), which has an exact F-distribution test statistic, will also be implemented in the context of anomaly detection for comparison purposes. Finally, I will show the development of a compact form that exploits the principle of indirect comparison, the AVT detector. In essence, the AVT detector performs an asymmetric hypothesis test using only the estimates of second central moments. The AVT test statistic also tends to a known distribution, under a null hypothesis, as the number of samples increases. I will then present theoretical analyses of the *power of the test* for two distinct types of problems: (i) local anomaly detection through the perspective of a top view and (ii) scene anomaly detection through the perspective of a ground-level view. The power of the test is an important theoretical analysis to determine the asymptotic behavior of statistical tests [15]. Performance comparison among conventional and unconventional anomaly detectors will be presented through computed receiver's operating characteristics (ROC) curves and output surfaces. As a proof of principle experiment, I will end Section 3 by showing that an effective unconventional anomaly detector may be extended to function as a classifier.

Section 4 presents a summary of the work presented in this report, with an emphasis on the contributions of this work to the field of hyperspectral image processing. Also discussed are limitations and a look towards future work.

2. Conventional Anomaly Detection and Some Results

In this section, I discuss some of the most prominent anomaly detection techniques for hyperspectral data. As I mentioned in Section 1, object detection within hyperspectral data is a highly desirable goal. The data lends itself to the ability to search large spatial areas, ideally in an automated and timely fashion. In order for the detector to have value, it should have a high rate of detection and a low rate of false alarms. Anomaly detectors are in particular desired because they fall under the category of unsupervised learning methods, i.e., they do require offline training using samples of targets and nontargets, as artificial neural networks do.

In the next few subsections, I discuss the following conventional unsupervised learning methods, which will be used for comparison in this research. They are: Fisher's linear discriminant (FLD) detector [16], dominant principal component (DPC)/Eigen separation transform (EST) detectors [17], the industry standard Reed-Xi (RX) detector [10], and the kernel-based RX (KRX) detector [11]. These techniques, or variants of them, arguably represent a list of the most distinct approaches for anomaly detection. But conspicuously missing from this list are techniques based on Markov Chain (MC) theory. Anomaly detectors based on MC theory were excluded from this effort because their performances have been shown to be comparable, not improved, to that of the industry standard technique—the RX detector. Some of these MC detectors, however, have been shown to be significantly more efficient computationally than the RX detector (see, for instance, [18]).

As a preliminary note before the discussion of these techniques, consider two sets of spectral samples that will be used for comparison. This set is organized as two matrixes $\mathbf{X}_{\text{in}}$ (B x n_{in}) and $\mathbf{X}_{\text{out}}$ (B x n_{out}), test and reference, respectively; where B is the number of spectral bands, and n_{in} and n_{out} are the number of spectral samples of $\mathbf{X}_{\text{in}}$ and $\mathbf{X}_{\text{out}}$, respectively.

2.1 Fisher's Linear Discriminant (FLD)

Fisher's linear discriminant analysis is a standard technique in pattern recognition. It projects the original high dimensional data onto a low dimensional space, where all the classes are well separated by maximizing the *Raleigh quotient*, i.e., the ratio of between-class scatter matrix determinant to within-class scatter matrix determinant. The application of the FLD detector to hyperspectral imagery has been investigated for anomaly detection [17] and for object classification [19], where a classification algorithm was derived based on FLD, having different classes forced to be along different directions in a low dimensional space. Multi-object classification is beyond the scope of this report. Hence, our focus will be limited to adapting FLD to a two class problem in HS imagery.

A version of FLD for the two-class (anomaly or not anomaly) problem is show below:

$$Z_{FLD} = \left| \mathbf{E}^{\mathbf{t}}_{\mathbf{S}_B/\mathbf{S}_w} \left(\overline{\mathbf{x}}_{\mathbf{in}} - \overline{\mathbf{x}}_{\mathbf{out}} \right) \right|, \tag{1}$$

where $\overline{\mathbf{x}}_{in}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{in}}$, $\overline{\mathbf{x}}_{out}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{out}}$, $|\bullet|$ denotes the absolute value operator, and $E^{t}_{S_B S_W^{-1}}$ is the eigenvector decomposition of the scatter matrices ratio $S_B S_W^{-1}$, where

$$\mathbf{S}_{\mathbf{W}} = \sum_{i=1}^{n_{in}} (\mathbf{x}^{(i)}_{\mathbf{in}} - \overline{\mathbf{x}}_{in})(\mathbf{x}^{(i)}_{\mathbf{in}} - \overline{\mathbf{x}}_{in})^t + \sum_{i=1}^{n_{out}} (\mathbf{x}^{(i)}_{\mathbf{out}} - \overline{\mathbf{x}}_{out})(\mathbf{x}^{(i)}_{\mathbf{out}} - \overline{\mathbf{x}}_{out})^t, \tag{2}$$

$$\mathbf{S}_{\mathbf{B}} = \sum_{i=1}^{n_{in}} (\mathbf{x}^{(i)}_{\mathbf{in}} - \overline{\mathbf{x}}_{\mathbf{total}})(\mathbf{x}^{(i)}_{\mathbf{in}} - \overline{\mathbf{x}}_{\mathbf{total}})^t + \sum_{i=1}^{n_{out}} (\mathbf{x}^{(i)}_{\mathbf{out}} - \overline{\mathbf{x}}_{\mathbf{total}})(\mathbf{x}^{(i)}_{\mathbf{out}} - \overline{\mathbf{x}}_{\mathbf{total}})^t, \tag{3}$$

$\overline{\mathbf{x}}_{total}$ is the total sample mean using the columns of both $\mathbf{X}_{\text{in}}$ and $\mathbf{X}_{\text{out}}$ as input, $\mathbf{x}^{(i)}_{in}$ and $\mathbf{x}^{(i)}_{out}$ are the *i*-th columns of $\mathbf{X}_{\text{in}}$ and $\mathbf{X}_{\text{out}}$, respectively, and n_{in} and n_{out} are the sample sizes of $\mathbf{X}_{\text{in}}$ and $\mathbf{X}_{\text{out}}$, respectively.

2.2 Dominant Principle Component (DPC) and Eigen Separation Transform (EST)

The DPC and EST techniques are both based on the same general principle, i.e., data are projected from their original high dimensional space onto a significantly lower dimensional space (in our case, only one dimension) using a criterion that promotes the highest sample variability within each domain in this lower dimensional space. Differences between DPC and EST are better appreciated through their mathematical representations:

$$Z_{DPC} = \left| \mathbf{E}^{\mathbf{t}}_{out} \left(\overline{\mathbf{x}}_{\mathbf{in}} - \overline{\mathbf{x}}_{\mathbf{out}} \right) \right|, \tag{4}$$

$$Z_{EST} = \left| \mathbf{E}^{\mathbf{t}}_{\Delta \mathbf{C}} \left(\overline{\mathbf{x}}_{\mathbf{in}} - \overline{\mathbf{x}}_{\mathbf{out}} \right) \right|, \tag{5}$$

where $\overline{\mathbf{x}}_{in}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{in}}$, $\overline{\mathbf{x}}_{out}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{out}}$, E^{t}_{out} is the transposed highest energy eigenvector from the principal component decomposition using as input the covariance matrix estimated from the rows of $\mathbf{X}_{\text{out}}$, and $E^{t}_{\Delta C}$ is the transposed highest positive energy eigenvector from the principal component decomposition using as input the difference between the estimated covariance matrix from the rows of $\mathbf{X}_{\text{in}}$

and the estimated covariance matrix from the columns of $\mathbf{X}_{\text{out}}$, and $|\bullet|$ denotes the absolute value operator.

2.3 Reed-Xi (RX) Algorithm

Reed and Yu in [10] derived in a fully adaptive multiband spectral detector. This detector was a generalized version of the adaptive spectral matched filter; the problem was formulated to detect objects of a known spatial pattern but unknown spectral distribution against a clutter background with *unknown* spectral distribution against a clutter background with unknown spectral covariance. This detection test has a constant false alarm rate (CFAR) property.

The basis of the fully adaptive spectral detector is to detect the spectral differences between a region to be tested and its surrounding neighboring pixels. This detector has been claimed to be one of the most robust detection techniques for the detection of a spectral anomaly in multispectral imagery [19], [20]. It was employed by the DARPA (then called ARPA) MUSIC program to detect military vehicles in an intense clutter background in [21]. This approach became known in the community as the RX anomaly detector, and eventually it became the industry standard for utility and comparison.

A popular version of the RX anomaly detector is shown below:

$$Z_{RX} = \left(\overline{\mathbf{x}}_{\mathbf{in}} - \overline{\mathbf{x}}_{\mathbf{out}}\right)^{\mathbf{t}} \mathbf{C}_{\mathbf{out}}^{-\mathbf{1}} \left(\overline{\mathbf{x}}_{\mathbf{in}} - \overline{\mathbf{x}}_{\mathbf{out}}\right), \tag{6}$$

where $\overline{\mathbf{x}}_{in}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{in}}$, $\overline{\mathbf{x}}_{out}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{out}}$, and $\mathbf{C}_{out}^{-1}$ is the sample covariance matrix using as input the rows of $\mathbf{X}_{\text{out}}$.

2.4 Kernel RX Algorithm

The conventional RX detector does not take into account the higher order relationships between the spectral bands at different wavelengths. The nonlinear relationships between different spectral bands within the target or clutter spectral signature were exploited recently by Kwon and Nasrabadi [11] using a kernel-based version of the RX model. The authors named this approach: the kernel RX (KRX) algorithm.

An interpretation is that the KRX algorithm extends the utility of the RX algorithm from a lower dimensional data space to a higher dimensional nonlinear feature space by applying a well known kernel trick (see, for instance, [11]) in order to *kernelize* the corresponding generalized likelihood ratio test (GLRT) expression of the conventional RX approach. The result of kernelization significantly improved the RX detector's performance. The GLRT expression of the kernel RX is similar to the conventional RX, but every term in the expression is in kernel form, which can be readily calculated in terms of the input data in its original data space.

The notion of applying nonlinear kernels as a means to extract features from data is not new. The most prominent algorithm using this application is the well known

support vector machine, as proposed by Vapnik [22]. Many other kernel-based versions of well known algorithms have been proposed in the literature, including PCA [23] and FLD [24]. The authors of the KRX detector, however, were the first to present to the hyperspectral community such a technique applied to the industry standard RX algorithm.

The KRX anomaly detector is compactly represented by the following test statistic:

$$Z_{KRX} = (\mathbf{K}_{\bar{\mathbf{x}}_{in}} - \mathbf{K}_{\bar{\mathbf{x}}_{out}})^t \mathbf{K}_{out}^{-1} (\mathbf{K}_{\bar{\mathbf{x}}_{in}} - \mathbf{K}_{\bar{\mathbf{x}}_{out}}) , \tag{7}$$

where $\mathbf{K}_{\bar{\mathbf{x}}_{in}} = \mathbf{K}(\bar{\mathbf{x}}_{in}, \mathbf{X}_{out})$ is a kernel-function based vector that uses as input $\bar{\mathbf{x}}_{in}$ and $\mathbf{X}_{out}$ representing the dot product between these two inputs nonlinearly mapped onto a higher dimensional space, $\bar{\mathbf{x}}_{in}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{in}}$, $\bar{\mathbf{x}}_{out}$ is the sample mean vector using the columns of $\mathbf{X}_{\text{out}}$, $\mathbf{K}_{\bar{\mathbf{x}}_{out}} = \mathbf{K}(\bar{\mathbf{x}}_{out}, \mathbf{X}_{out})$ is the same kernel function using instead the dot product between $\bar{\mathbf{x}}_{out}$ and $\mathbf{X}_{out}$, and $\mathbf{K}_{out}^{-1}$ is the inverse of $\mathbf{K}_{out} = \mathbf{K}(\mathbf{X}_{out}, \mathbf{X}_{out})$ using the dot product between $\mathbf{X}_{out}$ and itself. (Matrices $\mathbf{X}_{\text{in}}$ [B x n_{in}] and $\mathbf{X}_{\text{out}}$ [B x n_{out}] were defined earlier in this section as the test and reference samples, respectively.) The rationale for using $\mathbf{K}_{out}^{-1}$ as the normalizing matrix is based on the properties of the so-called kernel PCA. For a detailed discussion see, for instance, [11].

Finally, the kernel function used to implement the KRX detector in this research was the well known Gaussian (radial basis function) RBF kernel, or

$$k(\mathbf{x}_0, \mathbf{x}_1) = \exp\left(\frac{-\left\| \bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1 \right\|^2}{2\sigma^2} \right), \tag{8}$$

where $\| \bullet \|$ denotes the magnitude of a vector.

2.5 Performance with Actual Hyperspectral Data

The data from the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor were used to test the conventional local anomaly detectors described in this section. The HYDICE sensor records *210* spectral bands in the visible-to-near infrared (VNIR) and short-wave infrared (SWIR), *0.4-2.5 μm*, forming a cube of spatially registered pixels. Each pixel then in the scene represents a sequence of *210* components.

To challenge the local anomaly detectors, I extracted a sub-cube sufficiently large from the HYDICE dataset to include various levels of local complexity. The imagery used is from the so-called Forest Radiance I (FR-I) dataset and the spectral average (from *150* bands) of the sub-cube in reference is shown in figure 2 (far left), as a two dimensional (2D) image. (Water absorption and low signal-to-noise ration bands were discarded and only the remaining *150* bands were used. The discarded bands are:

23^{rd}-101^{st}, 109^{th}-136^{th}, and 152^{nd}-194^{th}.) From actual ground truth, it is known that the scene in FR-I contains *14* stationary motor vehicles on sparse grasses, near a forest in Aberdeen, Maryland. The vehicles in FR-I are considered in this report the objects of interest (targets). These targets and their shadows are quite noticeable in the scene shown in Fig. 2.8. Effective local anomaly detectors are expected to accentuate objects in the scene that are spectrally different from the local background and to suppress noise. Noise in this context also includes strong responses due to major transitions in local regions (e.g., grass and shadow).

To implement the five conventional algorithms (FLD, DPC, EST, RX, and KRX) as local anomaly detectors, I employed a standard sampling mechanism, where local background samples from the neighboring area of the pixel being tested are compared to the test samples. In order to perform this operation, at each test pixel location, a dual concentric rectangular window is used to separate a local area into two mutually exclusive regions: the inside window region (W_{IN}) and the outside window region (W_{OUT}). The size of W_{IN} is set to sample portions of potential targets and the distance between the W_{IN} and W_{OUT} is set to enclose the largest target size that is expected in the scene, given that the data is assumed to have been collected from a platform flying at a fixed altitude and that the sensor pixel resolution is known a priori. The size of W_{OUT} is set to include sufficient statistics from the neighboring background.

To implement the KRX detector, the Gaussian RBF kernel in (2.29) was used with the variance set to approximately *4.5*. The sizes of W_{IN} and W_{OUT} for the local kernel and covariance matrix estimations were 5 x 5 and 15 x 15 pixel areas, respectively.

The same concentric, dual-window sampling mechanism and window sizes were used to implement the RX, FLD, DPC, and EST detectors. The output surfaces of these detectors are shown in figure 2.

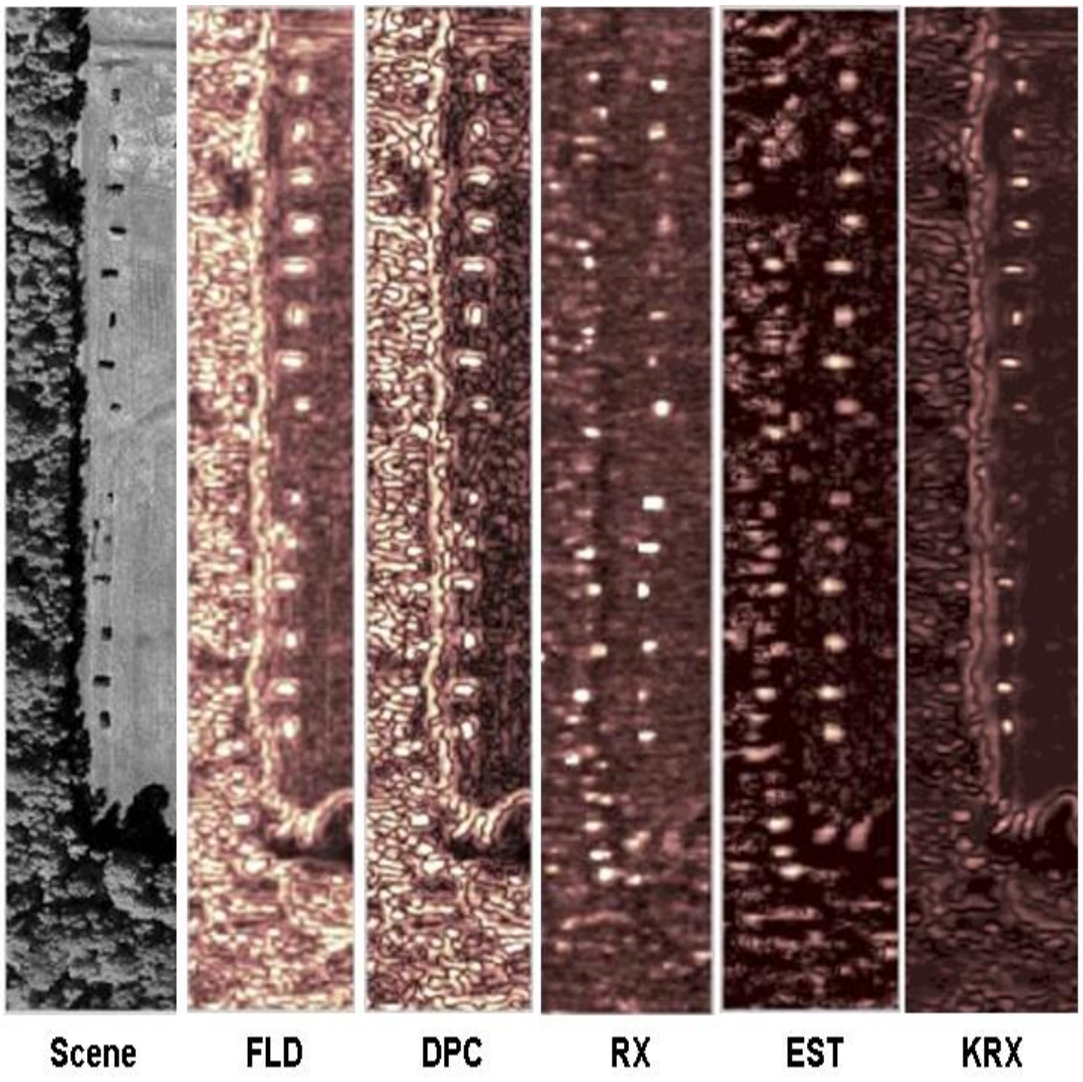

Figure 2. Decision surfaces testing the HYDICE FR-I hyperspectral data, forest radiance. The intensities of local peaks reflect the strength of anomaly evidences as *seen* by different detectors. Boundary issues were ignored in this test; surfaces were magnified to about the size of the original image for the purpose of visual comparison. The quality of performance shown in this capture is the state of the art using conventional approaches.

Notice in figure 2 that all five detectors perform as expected: they accentuate local anomalies in the scene, including of course the *14* targets near the treeline. The *colormap* used to display the four surfaces are exactly the same, and their values are relative only to maximum values in the individual surfaces.

Figure 2 suggests that the detectors based on conventional approaches are effective locating the presence of isolated objects in the scene (e.g., a motor vehicle parked in a terrain), albeit they clearly fail to suppress the responses from the scene that would not be regarded by an image analyst as important (e.g., a local patch consisting of a transition of two distinct regions—shadow and grass.) The quality of performance observed in figure 2 is arguably the state of the art using conventional approaches.

3. Asymmetric Hypothesis Tests

In this section, I formulate two types of anomaly detection problems and propose four anomaly detection techniques to address these problems.

Anomaly detection problems occurring in remote sensing applications are often characterized as the detection of local or global anomalies. Local anomaly detectors process small windows of the HS imagery in order to compare the spatial and spectral properties of the centrally located pixels in the window with the properties of the surrounding pixels. Those pixels that are spatially-spectrally different from their surrounding backgrounds are considered anomalies.

In this research effort, I address what are referred to as top view and ground view detection problems. The top view detection problems use imagery as input from a top view perspective between the sensor and an imaged scene. The application discussed in Section 2, using conventional anomaly detectors to test actual HS data, uses samples from top view imagery as input to detection algorithms. The ground view detection problems, on the other hand, use imagery from a ground-level view perspective between the sensor and the imaged scene. Applying detection algorithms to test ground view imagery is a significantly more challenging problem than applying the same algorithms to test top view imagery, as the sizes and shapes of potential targets are completely unknown a priori.

I also develop in this section four techniques for anomaly detection using both top view and ground view imagery. These techniques are all based on the indirect comparison approach discussed in Section 1. The names of these algorithms are: Semiparametric (SemiP) detector, Approximation to Semiparametric (AsemiP) detector, Asymptotic F-distribution Test (AFT) detector, and Asymmetric Variance Test (AVT) detector. (A fifth technique will be presented for comparison purposes only; this technique is based on the classic one-way ANOVA model.)

3.1 Formulation of Problems

I discuss in this section a data model that is suitable for our techniques and a detailed formulation of both types of problems discussed in this section, one from the perspective of a sensor's top view (top-view imagery) and another from the perspective of a sensor's ground-level view (ground-view imagery).

3.1.1 Simplified Data Modeling

This subsection describes briefly a data model for the hyperspectral reflectance phenomenology. For mathematical simplicity, a model, which is an idealization of a rather complicated optical sensor model, is used to represent object reflectance collected by, in this case, a visible-to-shortwave infrared (V-SWIR) hyperspectral sensor (the 0.4 to 2.4 μm bands) or by a visible-to-infrared (VNIR) hyperspectral sensor (the 0.38 to 0.97 μm bands). Although the sensor may produce many subspectral bands, only a portion of the bands are useful for detection since

atmospheric absorption causes some subbands not to provide any spectral information about the clutter or target.

Hyperspectral data are produced by a sensor that either scans or uses a focal plane array to collect the data in a rectangular grid about the region of interest. The sensor filters the data to provide a large number of narrow wavelength bands. Recall that each pixel then represents a resolution spot size on the ground. In order to appreciate how the atmospheric and illumination conditions affect the reflectance of an object in the ground, consider a relationship derived in [25] for the spectral radiance reaching an airborne or satellite sensor, which can be expressed in simplified form as

$$L_p = \int_R B_p \left[\left(GE_s(\lambda)\tau_1(\lambda)\cos\theta + FE_d(\lambda) \right) \tau_2(\lambda) \frac{r(\lambda)}{\pi} + L_u(\lambda) \right] d\lambda \qquad (9)$$

where R is the spectral region of interest centered at λ_p (the central wavelength in the p^{th} band in units of μm), L_p is the effective spectral radiance in the p^{th} band in units of $[Wcm^{-2}sr^{-1}\mu m^{-1}]$, $E_s(\lambda_p)$ is the exoatmospheric spectral irradiance from the sun in units of $[Wcm^{-2}\mu m^{-1}]$, $\tau_1(\lambda_p)$ is the transmission through the atmosphere along the Sun-object path, θ is the angle from the surface normal to the sun, F is the fraction of the spectral irradiance from the sky ($E_d(\lambda_p)$) incident on the object (i.e., not blocked by adjacent objects), G is the fraction of direct sunlight incident on the object, $\tau_2(\lambda_p)$ is the transmission along the object-sensor path, $r(\lambda_p)$ is the spectral reflectance factor for the object of interest (i.e., $r(\lambda_p)/\pi$ is the bidirectional reflectance in units of sr^{-1}), $L_u(\lambda_p)$ is the spectral path radiance $[Wcm^{-2}sr^{-1}\mu m^{-1}]$, and β_p is the normalized spectral response of the p^{th} spectral band of the sensor under study where

$$B_p = \frac{\rho_p(\lambda_p)}{\int_R \rho_p(\lambda) d\lambda} \qquad (10)$$

with $\rho_p(\lambda_p)$ being the peak normalized spectral response in R of the p^{th} band. Atmospheric and illumination conditions will affect all the radiometric terms in (9) (i.e., $E_s(\lambda_p)$, $\tau_1(\lambda_p)$, $\tau_2(\lambda_p)$, $E_d(\lambda_p)$, and $L_u(\lambda_p)$), which makes the task of predicting the responses of a particular object a formidable one. For a particular set of conditions during the data collection, the spectral radiance from a pixel-size location at the scene observed by a K-band sensor can be expressed as

$$\upsilon = \left(L_1, L_2, \cdots, L_p, \cdots, L_K \right), \qquad (11)$$

where, an additional subscript may be introduced to differentiate the spectral radiance of the j^{th} pixel, or

$$\upsilon_j = \left(L_{j1}, L_{j2}, \cdots, L_{jp}, \cdots, L_{jK} \right). \qquad (12)$$

3.1.2 Top View Anomaly Detection

In the top-view imagery, targets are expected to be stationary motor vehicles of unknown shape and material type, and the spatial size of the largest vehicle of interest is assumed known, using, of course, the sensor's pixel resolution and platform's

flying altitude. It is also assumed that the spectral radiance from targets are significantly different—hence, anomalous—with respect to a reference set of spectral radiance from natural clutter backgrounds. The sampling mechanism for the top-view problem will be discussed shortly, but first I will comment on circumventing local dependence of hyperspectral data, which is an assumption inherent in most statistical models.

The statistic tests proposed in this section rely on central limit theorem (CLT) to show that they converge in law to known distributions. The proof of CLT relies on the statistical independence of random variables, which is an assumption potentially at odds with the highly correlated local radiance often found in hyperspectral imagery. Researchers interested in multivariate solutions have used a local high-pass filter (HPF) spatially at each band to approximate this independence—the RX algorithm, for instance, expects the data to be preprocessed as such, as discussed in [10].

Although not emphasized in the literature, the following is the rationale for using a HPF to generate approximate independence in hyperspectral data: Let the random variables u_1, u_2 and u_3 be statistically dependent and let $h_1 = u_2 - u_1$ and $h_2 = u_3 - u_2$. It is not difficult to show (see, for instance, [26]) the plausibility that h_1 is statistically independent of h_2. This transformation is widely used by professional statisticians so that dependent random variables can be addressed using techniques based on statistical independence. Notice that a HPF may be implemented spatially by taking the systematic difference between the values of a pixel and its previous neighbor.

As our interest is to seek univariate solutions, for our sampling mechanism I aim at approximating independence taking in consideration both the spatial and spectral domains. Figure 3 illustrates the sampling mechanism and data preprocessing that I propose for transforming dependent random variables into approximately independent random variables for our statistical models.

Figure 3 introduces three window cells from which samples will be drawn from the top-view imagery. These windows are referred to as: test cell, reference cell, and variability cell. Information between the variability and reference cells will be used to form a control or reference feature values, and information between the variability and test cells will be used to form unknown test feature values. (Note: sequences of spectral radiances will be treated as vectors for the purpose of preprocessing, but their preprocessed versions will be treated as real-value sequences for the purposes of our statistical models. It should be noticeable from their notations.)

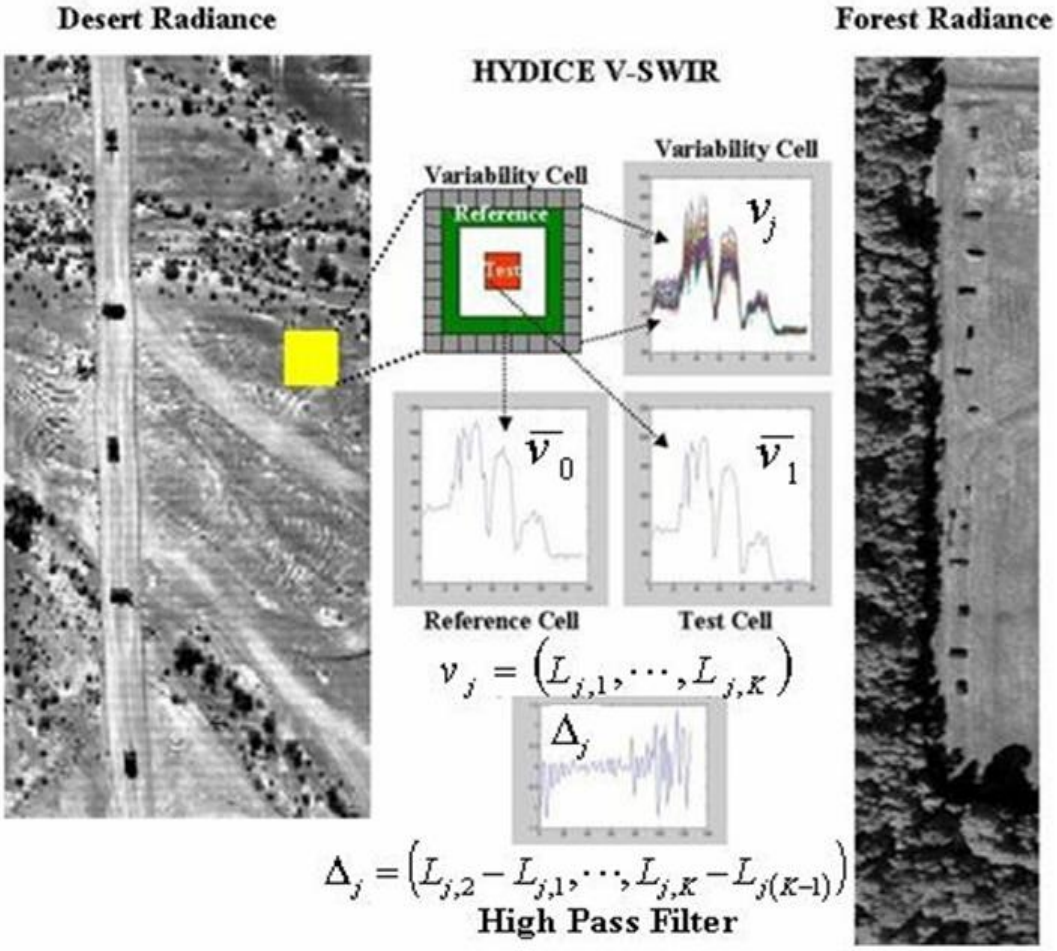

Figure 3. Sampling mechanism proposed to transform local HS radiance into plausible independent random samples.

The *test cell* will be used to provide a spectral average sequence $(\overline{v}_1)$ from a *w* x *w* window; the *reference cell* provides a spectral average sequence $(\overline{v}_0)$ from *M* spectral sequences surrounding a guard region, also known as the blind area, between test and reference cells to account for larger than (*w* x *w*) targets; and the *variability cell* provides *J* individual spectral vectors (v_j) each consisting of *k = 1,...,K* spectral responses (L_{jk}) at *K* distinct wavelengths, see figure 3.

Radiance values in the adjacent bands in (12) are highly correlated—hence, dependent—so to promote statistical independence, I apply a HPF filter in the spectral domain, transforming v_j into $\mathbf{\Delta_j}$ (see fig. 3), and then use $\mathbf{\Delta_j}$ to compute a feature that further promotes independence. The feature is the angle difference between two vectors, or

$$x_{ij} = \frac{180}{\pi} \arccos\left(\frac{\mathbf{\Delta}_\mathbf{j}^\mathbf{t} \overline{\mathbf{\Delta}}_\mathbf{i}}{\|\mathbf{\Delta}_\mathbf{j}\| \|\overline{\mathbf{\Delta}}_\mathbf{i}\|} \right), \tag{13}$$

where, $\mathbf{\Delta_j} = \left(L_{j,2} - L_{j,1}, \cdots, L_{j,K} - L_{j,(K-1)}\right)^\mathbf{t}$ is the high pass filtered version of v_j; $\overline{\mathbf{\Delta}}_\mathbf{i}$ is the high pass filtered version of $\overline{v}_i$; *i = 0* (reference cell)*, 1* (test cell); *j = 1,...,J* (*J* is the total number of pixels in the variability cell) $x_i = (x_{ij}) = (x_{i1}, \cdots, x_{iJ})$ is a random sequence of angle differences ranging from *0* to *90* degrees (*0* representing minimum class difference between reference and test samples and *90* representing the maximum class difference between these samples); the operator $\| \mathbf{z} \|$ denotes the squared root of $\mathbf{z}^\mathbf{t}\mathbf{z}$; and $[\cdot]^t$ denotes the vector transpose operator. Figure 4 depicts the transformed version of a highly spatially/spectrally correlated set of hyperspectral samples from a terrain in California using the SOC-700 sensor (additional details will be discussed later). The transformed result shown in figure 4 (right) is considered in this report a good approximation of a set of statistically independent feature values. Hence, they will be used as input to the detectors presented in this section.

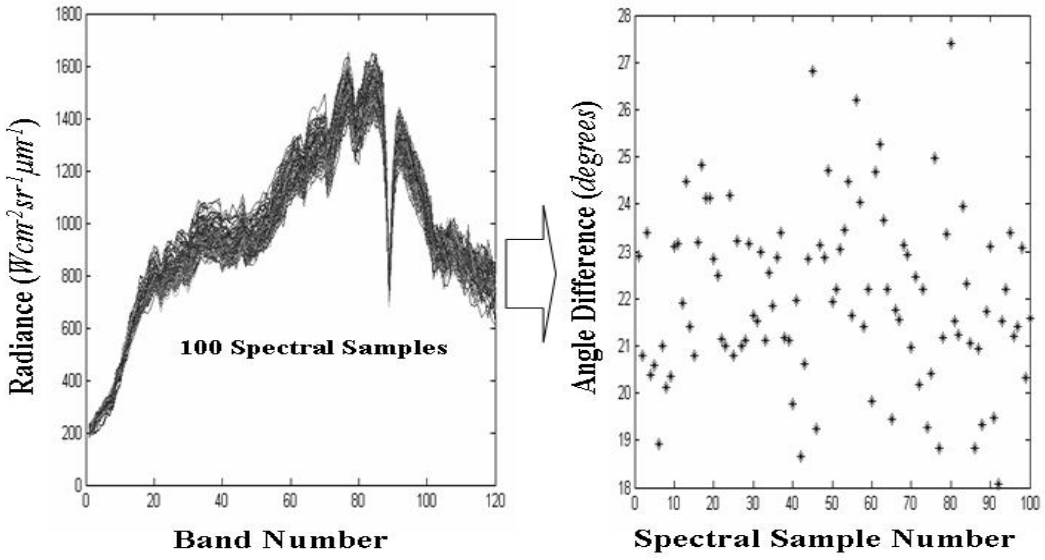

Figure 4. A set of *100* spectral samples from a highly correlated natural clutter background (left) is transformed using a high pass filter in the spectral domain, followed by an angle difference mapping—as described in text, to yield a set of approximately independent observations (right). Band 1 represents 0.4 μm and band 120 represents 0.97μm.

Suppose now that x_0 denotes the reference random sequence and that x_1 the test random sequence, let both sequences be distributed (~) under unknown joint distributions f_0 and f_1, respectively, or

$$\begin{aligned} x_1 &= \left(x_{11},\ldots, x_{1n_1}\right) \sim f_1(x) \\ x_0 &= \left(x_{01},\ldots, x_{0n_0}\right) \sim f_0(x), \end{aligned} \tag{14}$$

where, $n_0 = n_1 = J$ in this particular implementation.

The window cells are expected to draw samples and to move systematically across the entire imagery, and at each location a detector attempts to answer the following question: Do x_0 and x_1 belong to the same population, or class? If the answer is *no*, the test location would be labeled as an anomaly with respect to its surroundings at that location.

3.1.3 Ground View Anomaly Detection

The problem of anomaly detection using ground-view HS imagery is quite different from the one described for the top view anomaly detection, because the range between the sensor and objects in the scene are typically not known. In essence, the question I attempt to answer in this case is as follows: Will a *meaningful* object of unknown size, shape, or material type, which may be found completely immersed in some clutter background, be detectable when compared to a small set of spectral signatures representing the most abundant classes of objects in the given scene?

This problem is a significantly harder problem to address, since the distance between objects and the sensor are added to other unknown data collection factors. Using top-view imagery, one can exploit information on the expected platform flight altitude and the sensor's pixel resolution to fix a maximum expected size of targets.

To answer the question above, I first assume that a small spectra set of the most abundant object classes in a scene is available from the scene—in this case, tree and terrain, or at least from the general geo-location where the data were collected. This assumption is not as farfetched as it may sound; such a capability is currently under consideration by the Army Research Laboratory for a hyperspectral sensor application, where a miniaturized hyperspectral sensor similar in size and appearance to a gun scope would be available to the user, giving that user the ability to collect spectral samples from a scene using a trigger. Collected hyperspectral samples would be stored in a library—featured in the device—and be available to an electronic chip housing an anomaly detection algorithm.

Figure 5 depicts an illustrative ground level HS scene—the average of 120 bands between 0.40 μm and 0.97 μm (visible to near infrared, VNIR)—and a correspondent set of two spectral classes in that scene, i.e., sparse grass (terrain) and tree leaves. From actual ground truth, it is known that the center of the scene consists of three stationary vehicles and a standing person out in the open field. The two overlaid white boxes in the image show approximately the locations where those samples were taken from; the trees are visible in the upper part of the image. Within each box, approximately 1,000 spectral samples were drawn, which from each set a subset of 100 samples were randomly chosen to represent a corresponding class. Two subsets of 100 spectral samples each are available to represent two different classes. More about the sensor and the data collection will be discussed later.

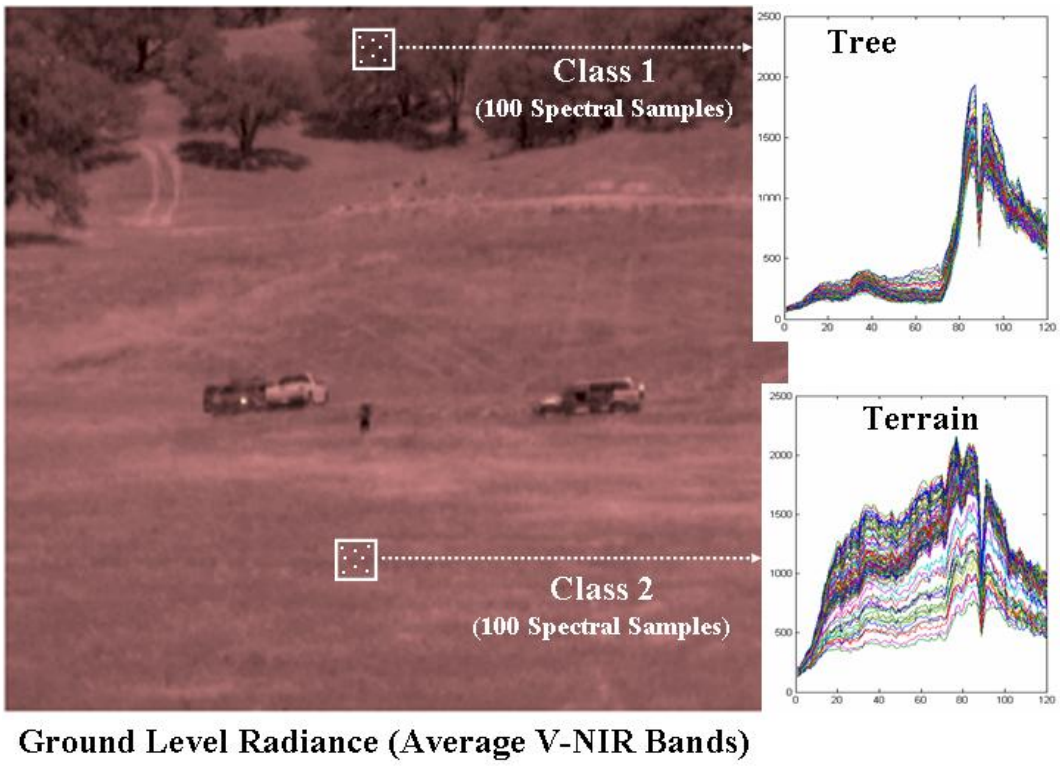

Figure 5. Ground level HS scene—the average of 120 bands between 0.40 μm and 0.97 μm—and a correspondent set of two spectral classes in that scene, i.e., sparse grass (terrain) and tree leave. The two overlaid white boxes in the image show approximately the locations where those samples were drawn from.

I propose to use the spectral transformations discussed for the top-view imagery, albeit each object class would be transformed and labeled as different classes. For instance, for the two hyperspectral sets shown in figure 5 (right), let the spectral average $\bar{v}_0^{(p)}$ from *n* hyperspectral samples $v_j^{(p)}$ (*j = 1, ... , n*), where in this example *p* = (*1* [*tree class*], *2* [*terrain class*]), be high pass filtered along with all $v_j^{(p)}$ yielding $\overline{\Delta}_0^{(p)}$ and $\Delta_j^{(p)}$, respectively. The angles between each $\Delta_j^{(p)}$ and $\overline{\Delta}_0^{(p)}$ form the p^{th} sequence $y_0^{(p)}$; figure 4 (right) shows an example of $y_0^{(p)}$. Suppose, now, that a *w* x *w* window collects spectral samples across the image and that, at a given location in the image, the high pass filtered version of the corresponding spectral average is denoted by $\overline{\Delta}_1$. The angles between each $\Delta_j^{(p)}$ and $\overline{\Delta}_1$ form the p^{th} sequence $y_1^{(p)}$ — notice that $\overline{\Delta}_1$ is fixed for all *p*'s in a given location. Let $y_0^{(p)}$ be the reference random sequence of the p^{th} object class and $y_1^{(p)}$ be the corresponding test random sequence, and let these sequences be distributed (~) under unknown joint distributions $f_0^{(p)}$ and $f_1^{(p)}$, respectively, or

$$y_1^{(p)} = \left(y_{11}^{(p)},..., y_{1n}^{(p)}\right) \sim f_1^{(p)}(x)$$
$$y_0^{(p)} = \left(y_{01}^{(p)},..., y_{0n}^{(p)}\right) \sim f_0^{(p)}(x). \quad (15)$$

I would like to know whether $f_0^{(p)}$ and $f_1^{(p)}$ are statistically different from each other for the p^{th} object class. If indeed it is the case for all object classes, then the spectral information at the given testing location in the image would be labeled as an anomaly with respect to all classes sampled from the scene.

The random sequences x_0 and x_1 in (14) and $y_0^{(p)}$ and $y_1^{(p)}$ in (15) will be used as inputs to the model discussed next.

3.2 Semiparametric Inference

In this section, I describe how to apply semiparametric inference to both types of anomaly detection problems. I start by describing a logistic model that is suitable for the notion of indirect comparison discussed in Section 1, and then I describe the adaptation of this model to the problem of anomaly detection using top view imagery and its adaptation to ground view imagery.

The logistic model in reference is based on case-control data, and its mathematical development depends on some of the advances made on the theory of semiparametric inference [27], [28], [29]. Semiparametric approaches are commonly used in analyzing binary data that arise in studying relationships between disease and environment of genetic characteristics [30], [31], [32]. The logistic model that will be discussed shortly has its roots in the standard logistic function having the general form

$$P(z) = \frac{\exp(\lambda + \beta z)}{1 + \exp(\lambda + \beta z)}, \quad (16)$$

where λ is a scale parameter and β interpreted as a constant rate both defining a proportion *P,* which is dependent on a variable *z*. The logistic function was invented in the 19th century for the description of the growth of populations and the course of autocatalytic chemical reactions. Pierre-Francois Verhulst (1804-1849) named (16) as the *logistic* function and published his suggestions between 1838 and 1847. For a complete historical account, see [33].

3.2.1 A Logistic Model

Suppose two random samples (real valued, not vector valued) x_0 and x_1 (of sizes n_0 and n_1, respectively) are independent and have their components independent and identically distributed (*iid*), as shown in the following model:

$$x_1 = (x_{11},..., x_{1n_1}) \quad iid \sim g_1(x)$$
$$x_0 = (x_{01},..., x_{0n_0}) \quad iid \sim g_0(x) \quad (17)$$

where

$$\frac{g_1(x)}{g_0(x)} = \exp(\alpha + \beta x). \tag{18}$$

In the context of HS imagery, the random samples x_0 and x_1 may represent the transformed versions of local HS radiances, where this transformation is geared to promote statistical independence both in spectral and spatial domains.

Notice in (18) that since g_1 is a density, $\beta = 0$ must imply that $\alpha = 0$, since α merely functions as a normalizing parameter. Notice also that if $\beta = 0$ then x_0 and x_1 must belong to the same population (i.e., $g_1 = g_0$). Using this fact, a local anomaly detector can be designed to test the following hypotheses:

$$\begin{aligned} H_0 &: \quad \beta = 0 \quad (g_1 = g_0) \qquad \text{anomaly absent} \\ H_1 &: \quad \beta \neq 0 \quad (g_1 \neq g_0) \qquad \text{anomaly present.} \end{aligned} \tag{19}$$

Testing (19) locally and repeating this test across the image yields a binary surface consisting of values *1* and *0*, representing H_1 and H_0, respectively. An isolated anomalous object is expected to produce a cluster of *1*'s in this surface.

The detector relies on the asymptotic behavior of the ML estimate of β, $\hat{\beta}$, which can be shown to be Normally distributed [28], as the sample size tends to infinity $(n \to \infty)$, or

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \xrightarrow[n\to\infty]{} N\left(0, \frac{\rho^{-1}(1+\rho)^2}{Var(t)}\right), \tag{20}$$

where the combined random sample, or the union of x_0 and x_1, is $t = (x_{11},...,x_{1n_1},x_{01},...,x_{0n_0}) \equiv (t_1,...,t_n)$, $Var(t)$ is the variance of t, β_o is the true value of β, $\rho = n_1/n_0$, $n = n_1+n_0$, and $\to$ means *converges to*.

Finding the ML estimate (MLE) of $Var(t)$, $\hat{V}(t)$, and normalizing the left side of (20) by this MLE, then setting $\beta_o = 0$, and squaring the final result, one can test H_0 in (19) with the following expression (see Appendix A for details):

$$Z_{SemiP} = n\rho(1+\rho)^{-2}\hat{\beta}^2\hat{V}(t) \xrightarrow[n\to\infty]{} \chi_1^2, \tag{21}$$

which has the chi-square distribution asymptotic behavior with one degree of freedom, χ_1^2. A decision is based on the value of Z_{SemiP} in (21), i.e., high values reject hypothesis H_o, declaring anomalies.

The SemiP anomaly detector, as shown in (21), relies on profiling and on a theorem applicable to extremum estimators (see Appendix A), but its implementation has some undesirable requirements. The most prominent one is the fact that one cannot find an explicit solution for the ML estimators of α and β, $\hat{\alpha}$ and $\hat{\beta}$, respectively. Therefore, the alternative is to maximize using some optimization algorithm the log likelihood function,

$$\log[\zeta(\alpha,\beta,g_0)] = \sum_{j=1}^{n_1}(\alpha+\beta x_{1i}) - \sum_{i=1}^{n}\log[1+\rho\exp(\alpha+\beta t_i)], \quad (22)$$

which is a direct result from the likelihood function

$$\begin{aligned}\zeta(\alpha,\beta,g_0) &= \prod_{i=1}^{n_0} g_0(x_{0i}) \prod_{j=1}^{n_1} \exp(\alpha+\beta\ x_{1j})\ g_0(x_{1j}) \\ &= \prod_{i=1}^{n=n_1+n_0} g_0(t_i) \prod_{j=1}^{n_1} \exp(\alpha+\beta\ x_{1j}),\end{aligned} \quad (23)$$

in order to find $\hat{\alpha}$ and $\hat{\beta}$. Both ML estimators are required to find $\hat{V}(t)$ in (21), where

$$\hat{V}(t) = \hat{E}(t^2) - \hat{E}^2(t), \quad (24)$$

$$\hat{E}(t^k) = \sum_i t_i^k \hat{g}_0(t_i) \quad (25)$$

and

$$\hat{g}_0(t_i) = \frac{1}{n_0}\frac{1}{1+\rho\exp(\hat{\alpha}+\hat{\beta}t_i)}. \quad (26)$$

Incidentally, notice in (23) that only a portion of the likelihood function uses information from the union of x_0 and x_1 (represented by t), the other portion uses only information from x_1.

A system of score equations that maximizes (22) over (α,β) is shown below [29],

$$\begin{aligned}\frac{\partial l(\alpha,\beta)}{\partial\alpha} &= -\sum_{i=1}^{n}\frac{\rho\exp[\alpha+\beta\ t_i]}{1+\rho\exp[\alpha+\beta\ t_i]} + n_1 = 0 \\ \frac{\partial l(\alpha,\beta)}{\partial\beta} &= -\sum_{i=1}^{n}\frac{t_i\rho\exp(\alpha+\beta\ t_i)}{1+\rho\exp(\alpha+\beta\ t_i)} + \sum_{j=1}^{n_1} x_j = 0.\end{aligned} \quad (27)$$

The system of equations shown in (27) is a key element for justifying the asymptotic behavior shown in (20), see Appendix A. With this behavior shown in (20), under the null hypothesis, one can set a level of confidence for the hypothesis test. However, for implementation purposes, (27) is not much of a help. Instead, one must be concerned about finding a way to maximize (22) by employing, for instance, an unconstrained maximization subroutine—an iterative algorithm—to estimate the likelihood values of α and β that maximizes the log function in (22). This requirement is potentially a serious drawback, since such an algorithm is often sensitive to arbitrary initial conditions. This drawback was not readily observed using imagery from the visible to short-wave infrared (V-SWIR) region of the spectrum, but later I observed the initialization problems (i.e., maximization subroutine could not converge given its initial condition), using imagery from the long-wave infrared (LWIR) region. (Results from the LWIR imagery could not be released in this report because of restrictions imposed by the U.S. Army, which owns the LWIR dataset.)

When the employed maximization subroutine can converge, given a chosen initial condition, the SemiP detector performs remarkably well accentuating meaningful detections in a scene and suppressing the meaningless ones, as it will be shown later. I will also show that other asymmetric hypothesis tests can enjoy the same level of performance, achieved by the SemiP detector, but free from such a drawback.

I show next how to adapt the SemiP algorithm to the anomaly detection problem using both top view and ground view imagery.

3.2.2 Top View Anomaly Detection

The expression shown in (28) constitutes the SemiP detector for the top view imagery,

$$Z_{SemiP} = n\rho(1+\rho)^{-2}\hat{\beta}^2\hat{V}(t)\,, \tag{28}$$

where $\hat{\beta}$ and $\hat{V}(t)$ are the estimators of the parameter β in model (18) and the variance of the union of samples, respectively; and ρ is a function of n_0 and n_1, see (20).

A decision is based on the value of Z_{SemiP} in (28), where high values reject hypothesis H_o in (19), declaring then x_0 and x_1 as anomalies.

I now present some helpful hints on the implementation of the SemiP algorithm.

Sampling Mechanism: I used the mechanism described in this Subsection 3.1 to sample a pair of random feature sequences x_{ij} ($i = 0$ [*reference*], *1* [*test*]; $j=1,\ldots,J$) from HS imagery. I used a *9*-pixel (*3x3*) *test window*, a *56*-pixel *reference window,* and a *60*-pixel *variability window*, as shown in figure 3. Note that the size of the *variability window* determines the size of the feature vectors; x_0 and x_1 have the same size, $J = 60$.

Statistical Independence: An attempt should be made to promote statistical independence in HS data. See the discussion in Subsection 3.7.1.

Function Maximization: In order to implement (28), I perform an unconstrained maximization of the log maximum likelihood function in (22), or alternatively one could minimize the negative version of (22), to obtain the extremum estimators $\hat{\alpha}$ and $\hat{\beta}$. For this research effort, I used one of the conventional, unconstrained, nonlinear optimization algorithms—the Simplex Method [34], which is available in MATLAB™ software (Release 13) under the function name *fminsearch*. The Simplex Method is a direct search method that does not use numerical or analytic gradients. If n is the length of **x**, a simplex in n-dimensional space is characterized by the $n+1$ distinct vectors that are its vertices. For instance, in two-space, a simplex is a triangle; in three-space, it is a pyramid. At each step of the search, a new point in or near the current simplex is generated. The function value at the new point is compared with the function's values at the vertices of the simplex and, usually, one of the vertices is replaced by the new point, giving a new simplex. This step is repeated until the diameter of the simplex is less than the specified tolerance. It is obvious

from this description that one limitation of such a method is that it only may give local solutions. So, the initial guess may prove to be critical in some cases. I set the initial values to ($\alpha=0, \beta=0$).

Variance under the Null Hypothesis: $\hat{V}(t)$ in (28) should be computed using (24), (25), and (26).

Decision Threshold: Using (28), high values of Z_{SemiP} reject hypothesis H_o in (19), declaring then x_0 and x_1 as anomalies. One can set a decision threshold based on the type I error, i.e., based on the probability of rejecting H_o given that H_o is true. Using a standard integral table for the chi square distribution, with *1* degree of freedom, find a threshold that yields an acceptable probability of error (e.g., 0.001), or alternatively find and use a suitable threshold that yields a value at the *knee* of the SemiP's corresponding ROC curve. Results using the latter recommendation will be shown later in this section.

3.2.3 Ground View Anomaly Detection

The discussion thus far in this section is readily applicable to the problem formulated for the top-view imagery. For the ground-view imagery, however, where an output is computed for each sampled class in the scene, it would be more appealing to fuse, in some form, these results into a single decision value. I propose the following fusing logic for the ground level problem: At a given testing location in the image, index the expression in (21) for each corresponding class and denote $Z_{SemiP}^{(p)}$ the detector's output for the p^{th} class and have Z denote the collection of outputs for N classes; A single decision value at each testing location is attained by

$$\widetilde{Z}_{SemiP} = \min\left\{ Z = \left(Z_{SemiP}^{(1)}, Z_{SemiP}^{(2)}, \cdots, Z_{SemiP}^{(N)} \right) \right\}. \tag{29}$$

Notice in (29) that if the local spectral radiance is significantly anomalous to all N classes, then $\widetilde{Z}_{SemiP}$ would be a relatively large value. Otherwise, it would be a relatively small value indicating that the local spectral radiance probably belongs to at least one of the N classes. The two likely hypotheses for the multiclass problem using (29) are

$$\begin{aligned} H_2 &: \text{ at least one } \beta^{(p)} = 0 \quad \left(p = 1, \cdots, N\right) \\ H_3 &: \text{ all } \beta^{(p)} \neq 0. \end{aligned} \tag{30}$$

where $\beta^{(p)}$ is the unknown parameter associated with the corresponding p^{th} object class. Note that the model in (18) is implicitly indexed by p.

Testing the null hypothesis H_2 in (20) with (21) indexed by p, and using the random sequences $y_0^{(p)}$ and $y_1^{(p)}$ in (15) as inputs, constitutes the adaptation of the SemiP anomaly detector to the ground-level view problem. In addition, it is expected that the recommendations discussed in Subsection 3.2.2 referring to the implementation of this detector will be applied.

Finally, the initialization drawback discussed earlier for the SemiP detector, using top view imagery, is also relevant to the ground view imagery.

3.3 Approximation to the Semiparametric Detector

In reference to the SemiP detector shown in (28), there are two major factors working in harmony and in complementary fashion to promote maximum separation between signal (anomalies) and noise (non-anomalies): (i) the squared value of $\hat{\beta}$ and (ii) the estimated variance from the union of samples, $\hat{V}(t)$, which is also quadratic.

These factors work in the following way: If two samples from the same homogenous class are compared [i.e., is H_0: $\beta = 0$ ($g_1 = g_0$) true?], the term $\hat{\beta}^2$ tends to approach zero very fast, especially for $\hat{\beta}$ less than unity. On the other hand, if two samples from distinct classes are compared, the term $\hat{V}(t)$ tends to a relatively high number also very fast, asserting the fact that a combined sample vector *t* consists of components belonging to distinct populations.

Motivated by these properties, I shall state and prove an approximation algorithm based on large sample theory to replace complicated SemiP equations with simpler ones describing the same phenomenon.

3.3.1 Derivation

I start off by proposing the AsemiP algorithm, as follows:

Proposition 1 (*AsemiP Algorithm*). *Let*

$$x_1 = \left(x_{11}, \cdots, x_{1n_1}\right) \text{ be iid}, \quad E(x_{1i}) = \mu_1, \quad Var(x_{1i}) = \sigma_1^2 < \infty;$$
$$x_0 = \left(x_{01}, \cdots, x_{0n_0}\right) \text{ be iid}, \quad E(x_{0i}) = \mu_0, \quad Var(x_{0i}) = \sigma_0^2 < \infty;$$

assume that x_0 and x_1 are independent and that, for some x_0 and x_1, the union of samples *t*

$$t = \left(x_{11}, \cdots, x_{1n_1}, x_{01}, \cdots, x_{0n_0}\right) = \left(t_1, ..., t_{n=n_1+n_0}\right) \text{ is iid}, E(t_i) = \mu, Var(t_i) = \sigma^2 < \infty;$$

and define

$$\tilde{\beta} \equiv \mu_1 - \mu_0; \quad \zeta_1 \equiv \frac{\sigma^2}{\sigma_0^2}, \zeta_2 \equiv \frac{\sigma^2}{\sigma_1^2}.$$

If the null hypothesis H0: $\left(\tilde{\beta} = 0; \zeta_1 = \zeta_2 = 1\right)$ is true and

$$\hat{\tilde{\beta}} = \bar{x}_1 - \bar{x}_0, \qquad \bar{x}_i = n_i^{-1} \sum_{k=1}^{n_i} x_{ik}, \quad i = 0,1;$$

$$\tilde{V}(t) = \sum_{i=1}^{n} (t_i - \bar{t})^2 \cdot g(x_1, x_0), \qquad n = n_1 + n_0, \ \ where \ \ \bar{t} = n^{-1} \sum_{i=1}^{n} t_i \ ,$$

$$g(x_1, x_0) = \frac{(n-2)^2}{\left[\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_0} (x_{0i} - \bar{x}_0)^2 \right]^2}; \ \ and \quad \tilde{\rho} = \left(\frac{1}{n_1} + \frac{1}{n_0} \right)^{-1};$$

where, $\hat{\tilde{\beta}}$ is an estimate of $\tilde{\beta}$, then the random variable

$$Z_{AsemiP} = \tilde{\rho}(n-1)^{-1} \hat{\tilde{\beta}}^2 \tilde{V}(t) \rightarrow \chi_1^2 \tag{31}$$

converges in distribution to a chi-squared distribution with 1 degree of freedom.

I first make a few comments on (31) prior to presenting its proof.

By inspection, one should readily recognize the behavior of our chosen function $\tilde{\beta}$, in Proposition 1, to replace the unknown parameter β in the SemiP algorithm. If two samples from the same population are compared using (31), the estimate of $\tilde{\beta}$, $\hat{\tilde{\beta}}$, would also tend to approach zero—as the sample size increases, and tend otherwise for samples belonging to distinct populations.

The real challenge, however, is to derive a relatively simple estimator to replace $\hat{V}(t)$, as defined in (24). The estimator $\hat{V}(t)$ is a sum of squared errors individually weighted by their probability of occurrence. In Proposition 1, $g(x_1,x_0)$ shall provide that probability feature, but as a normalizing fixed value for all the occurrences, instead. In this sense, comparing two samples from distinct populations would produce very high cumulative square errors using the union of samples t, but appropriately weighted by a fixed proportion.

In principle, the overall behavior of (31) seems to track that of (21), and both random variables are asymptotically identically distributed as χ_1^2. Note that the AsemiP's performance will not asymptotically approach that of the SemiP's performance, as the number of samples increases; the former approximates the general behavior of the latter, i.e., it promotes a high separation between signal (objects) from noise (homogeneous and non-homogenous local regions).

Proof: If hypothesis H_0: $\left(\tilde{\beta} = 0; \zeta_1 = \zeta_2 = 1 \right)$ is true in Proposition 1, then $\sigma^2 = \sigma_1^2 = \sigma_0^2$ and, using the independent assumptions of x_1 and x_0, and CLT, it follows that

$$\frac{\hat{\tilde{\beta}}}{\sqrt{\frac{1}{n0}+\frac{1}{n_1}}\sigma_0} = \frac{\hat{\tilde{\beta}}}{\sqrt{\frac{1}{n0}+\frac{1}{n_1}}\sigma_1} \xrightarrow[\substack{n_0\to\infty \\ n_1\to\infty}]{} N(0,1), \tag{32}$$

$\hat{\tilde{\beta}}$ as defined in Proposition 1; in addition, the following estimators of σ_1^2 and σ_0^2 are known to be consistent [see, for example, [14]]:

$$S_1^2 = \frac{\sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)^2}{n_1-1}, \text{ and } \quad S_0^2 = \frac{\sum_{i=1}^{n_0}(x_{0i}-\bar{x}_0)^2}{n_0-1}. \tag{33}$$

Using both samples x_1 and x_0, let the following be another estimator of σ_0^2 (or σ_1^2), given that under H_0, $\sigma_0^2 = \sigma_1^2$,

$$S^2 = \frac{(n_1-1)S_1^2 + (n_0-1)S_0^2}{(n_1-1)+(n_0-1)}. \tag{34}$$

The estimator S^2 is unbiased under H_0, as its expected value $E[S^2]$ is equal to σ_0^2 and σ_1^2:

$$\begin{aligned} E(S^2) &= \frac{(n_1-1)E(S_1^2)+(n_0-1)E(S_0^2)}{(n_1-1)+(n_0-1)} \\ &= \frac{(n_1-1)\sigma_1^2+(n_0-1)\sigma_0^2}{(n_1-1)+(n_0-1)} \\ &= \sigma_0^2. \end{aligned} \tag{35}$$

True because S_0^2 and S_1^2 are consistent estimators and, under H_0, $\sigma_0^2 = \sigma_1^2$. I want to prove now a *weak law of large numbers* (WLLN) [14] for S^2 to verify that S^2 is indeed a consistent estimator. Using *Chebychev's inequality* [35], I have under H_0:

$$P(|S^2-\sigma_0^2| \geq \varepsilon) \leq \frac{E(S^2-\sigma_0^2)^2}{\varepsilon^2} = \frac{Var(S^2)}{\varepsilon^2} \tag{36}$$

and, thus, a sufficient condition that S^2 converges in probability to σ_0^2, or σ_1^2, is that $Var(S^2) \xrightarrow[(n_0,n_1)\to\infty]{} 0$.

Note that $Var(S^2)$ can be expressed as

$$Var(S^2) = \left[\frac{(n_1-1)}{(n_1-1)+(n_0-1)}\right]^2 Var(S_1^2) + \left[\frac{(n_0-1)}{(n_1-1)+(n_0-1)}\right]^2 Var(S_0^2) \tag{37}$$

and, since the *sample variance* is known to be consistent, S_0^2 and S_1^2 are both consistent estimators, which implies their variances must converge to zero,

$$\begin{aligned} Var(S_1^2) &\xrightarrow[n_1\to\infty]{} 0 \\ Var(S_0^2) &\xrightarrow[n_0\to\infty]{} 0, \end{aligned} \tag{38}$$

also

$$\left[\frac{(n_k-1)}{(n_1-1)+(n_0-1)}\right]^2 \xrightarrow[(n_1,n_0)\to\infty]{} 0;\ k=0,1. \tag{39}$$

Then $Var(S^2) \xrightarrow[(n_1,n_0)\to\infty]{} 0$ and, therefore, under H_0:

$$\frac{\sigma_0^2}{S^2} = \frac{\sigma_1^2}{S^2} \to 1, \text{ as } (n_0, n_1) \to \infty. \tag{40}$$

Using the same argument to arrive at (40), one can also show that under H_0:

$$\frac{S_t^2}{\sigma_0^2} = \frac{S_t^2}{\sigma_1^2} \to \zeta_1 = \zeta_2 = 1, \quad \text{as} \quad n = (n_1 + n_0) \to \infty, \tag{41}$$

where S_t^2 (the sample variance using t) is a consistent estimator of σ^2, or

$$S_t^2 = (n-1)^{-1} \sum_{i=1}^{n} (t_i - \bar{t})^2 \xrightarrow[n\to\infty]{} \sigma^2, \ \bar{t} = n^{-1} \sum_{i=1}^{n} t_i. \tag{42}$$

Note that S_t^2 can be expressed as $S_t^2 = n(n-1)^{-1} \left\{ n^{-1} \sum_{i=1}^{n} (t_i - \mu)^2 - (\bar{t} - \mu)^2 \right\}$, where the summation term (which does not include $\bar{t}$) tends to σ^2 in probability by the WLLN, and the term that includes $\bar{t}$ tends to *zero* in probability also by the WLLN. Consequently, from the definition of convergence in probability, the result in (42) follows, which in turn under H_0, implies that

$$\frac{S_t}{\sigma_0} = \frac{S_t}{\sigma_1} \to 1, \quad \text{as} \quad n = (n_1 + n_0) \to \infty. \tag{43}$$

To finalize the proof, consider Theorem 3.1 below.

Theorem 3.1 *(Slutsky)* [14]. *Let X_n tend to X in distribution and Y_n tend to c in probability, where c is a finite constant. Then*

(i) $X_n + Y_n$ tend to $X+c$ in distribution;

(ii) $X_n\, Y_n$ tend to cX in distribution;

(iii) X_n/Y_n tend to X/c in distribution, if c is not zero.

See proof in [14].

Using (32), (40), (43) and the *Slutsky* Theorem, I conclude that

$$\left[\left(\frac{\hat{\tilde{\beta}}}{\sqrt{\frac{1}{n_0}+\frac{1}{n_1}}\sigma_0}\right)\left(\frac{\sigma_0^2}{S^2}\right)\right]\left(\frac{S_t}{\sigma_0}\right) \xrightarrow[\substack{n_0\to\infty \\ n_1\to\infty}]{} N(0,1) \tag{44}$$

and that by squaring (44) and using results from [35], I can also conclude that

$$\frac{\hat{\tilde{\beta}}^2}{\left(\frac{1}{n_0}+\frac{1}{n_1}\right)} \frac{S_t^2}{S^4} \xrightarrow[\substack{n_0\to\infty \\ n_1\to\infty}]{} \chi_1^2, \tag{45}$$

which can be readily reformatted into (31) using the definitions given in Proposition 1 and in this proof.

3.3.2 Top View Anomaly Detection

In contrast to the SemiP algorithm, the AsemiP algorithm is significantly simpler to implement, as the latter does not require specialized subroutines (unconstrained minimization) to perform its function. Using the sampling mechanism described in this section, the variables in Proposition 1 are straightforward to implement. Alternatively, one may use the expression in (46) below as the AsemiP anomaly detector,

$$Z_{AsemiP} = \frac{\hat{\tilde{\beta}}^2}{\left(\frac{1}{n_0} + \frac{1}{n_1}\right) S^4} \frac{S_t^2}{}, \tag{46}$$

where S^2 and S_t^2 are defined in (34) and (42), respectively, and $\hat{\tilde{\beta}}$ is defined in Proposition 1.

One is also expected to promote statistical independence and to take a sufficiently large number of samples (larger than *30* for our formulated problems) to justify the use of approximation theorems of mathematical statistics. I used the sampling mechanism proposed in this section to obtain a pair of random feature sequences x_0 (reference) and x_1 (test). I also used a *9*-pixel (*3*x*3*) *test window*, a *56*-pixel *reference window,* and a *60*-pixel *variability window*, as shown in figure 3, where $J = 60$. For a statistical decision, high values obtained by using (48), or equivalently (31), reject the hypothesis H_o in Proposition 1, thus, declaring sequences x_0 and x_1 as anomalies. One may set a decision threshold based on a choice of type I error using, as the base distribution, the chi-square distribution with *1* degree of freedom. Or, alternatively, one can find and use a suitable threshold that yields a value at the *knee* of the AsemiP's corresponding ROC curve. Practitioners usually rely on the latter approach to set decision threshold.

3.3.3 Ground View Anomaly Detection

For anomaly detection using ground-view imagery, I can fuse the individual results produced by each object class in the same manner that I described for the SemiP detector, i.e., for a given testing location in the image, I index the expression in (46) for each corresponding class and denote $Z_{AsemiP}^{(p)}$ the detector's output for the p^{th} class and collect the outputs for N classes; A single decision value at a given testing location is attained by

$$\tilde{Z}_{AsemiP} = \min\left\{ Z_{AsemiP}^{(1)}, Z_{AsemiP}^{(2)}, \cdots, Z_{AsemiP}^{(N)} \right\}. \tag{47}$$

In (47), if the local spectral radiance happens to be significantly anomalous to all N classes then $\tilde{Z}_{AsemiP}$ would be a relatively large value. Otherwise, it would be a relatively small value indicating that the local spectral radiance probably belongs to at

least one of the N classes. The null hypothesis for this multiclass problem using (37) is

$$H_4: \text{ For at least one class } p, \begin{cases} \widetilde{\beta}^{(p)} = 0, \\ \zeta_1^{(p)} = \zeta_2^{(p)} = 1 \end{cases}; \quad p = 1, \cdots, N, \quad (48)$$

where $\widetilde{\beta}^{(p)}$, $\zeta_1^{(p)}$ and $\zeta_2^{(p)}$ are the unknown parameters associated with the p^{th} class. Note that the model defined in Proposition 1 is implicitly indexed by p.

Testing the null hypothesis H_4 in (48) using (46), and the random sequences $y_0^{(p)}$ and $y_1^{(p)}$ in (15) as inputs, constitutes the adaptation of the AsemiP anomaly detector to the ground-level view problem. Also, it is expected that the recommendations discussed in Subsection 3.3.2 referring to the implementation of this detector will be applied.

Note that the multiclass version of the AsemiP detector can enjoy the same level of performance, achieved by the SemiP detector, but free from parameter initialization. Results will be shown later in this section.

3.4 F Distribution Algorithms

I mentioned in Section 1 that the principle of indirect comparison can be implemented in many forms. I have shown thus far that the solution of a logistic model (a semiparametric approach) and an approximation of its performance by the application of a few fundamental theorems of large sample theory, and exploiting the behavior of main components of the SemiP expression, are two different ways to implement such a notion. In this section, I present a third technique, a technique also based on the same fundamental theorems, albeit this time I aim at using a known property of the *F distribution* to design the new detector. Our interest to introduce a detector having an asymptotic behavior governed by the F distribution was motivated by the existence of a technique known as *analysis of variance*, which will be also discussed in this section. The analysis of variance (commonly referred to as the ANOVA) is one of the most widely used statistical techniques, and it called our attention for this paper because it also yields an F test statistics. In its simplest form, the ANOVA is a method of estimating the means of several populations often assumed to be normally distributed. The ANOVA, contrary to what its name infers, is not concerned with analyzing variances but rather with analyzing *variation in means*.

I first derive this third technique, and for convenience it will be called the asymptotic F test (AFT) detector, followed by its adaptation to both types of imagery. Finally, I discuss what will be referred in this report as the ANOVA detector.

3.4.1 The F distribution

Sir Ronald Aylmer Fisher (1890-1962) introduced the *F* probability density function (pdf) to statisticians early in the 20th century while working on ML estimation problems. The "F" in the *F* distribution was given in his honor.

Let X and Y be random variables such that

- X and Y are independent;
- X is distributed (~) as the chi-square distribution with p degrees of freedom $\left(X \sim \chi_p^2\right)$; and
- $Y \sim \chi_q^2$, the chi-square distribution with q degrees of freedom.

Define a new random variable Z by

$$Z = \frac{X / p}{Y / q}. \tag{49}$$

Then the distribution of Z is called the central F distribution, or simply the F distribution with p and q degree of freedom, denoted by $\left(Z \sim F_{p,q}\right)$. By transformation of X and Y, one can show (see, for instance, [35]) that the probability density function (pdf) of the F distribution of Z has the form:

$$f_Z(x) = \frac{p^{p/2} q^{q/2}}{B\left(\frac{p}{2}, \frac{q}{2}\right)} \frac{x^{(p/2)-1}}{(px + q)^{(p+q)/2}}, \tag{50}$$

where, $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and Γ is the gamma function.

If $X \sim \chi_p^2(\lambda)$, the non-central chi-square distribution with p degrees of freedom and non-centrality parameter λ, with Y and Z defined as above, then the distribution of Z is called the non-central F distribution with p and q degrees of freedom and non-centrality parameter λ.

Useful remarks:

- If $X \sim F_{p,q}$, then $1/X \sim F_{p,q}$.
- If $X \sim t_p$, the t-distribution with p degrees of freedom, then $X^2 \sim F_{1,p}$.
- If $X \sim F_{p,q}$, then its expected value $E(X)$ and its variance $Var(X)$ are

$$E(X) = \begin{cases} \dfrac{q}{q-2} \end{cases} \textbf{ for } \boldsymbol{q > 2} \textbf{ and } Var(X) = \begin{cases} \dfrac{2q^2(p+q-2)}{p(q-2)^2(q-4)} \end{cases} \textbf{ for } \boldsymbol{q > 4.}$$

3.4.2 Asymptotic F Test

Let random variables x_{ij} be observed according to the model

$$\begin{aligned} x_{1j} &= \theta_1 + \varepsilon_{1j}, \quad j = 1,..., n_1 \\ x_{2k} &= \theta_2 + \varepsilon_{2k}, \quad k = 1,..., n_2 \end{aligned} \tag{51}$$

where θ_1 and θ_2 are unknown parameters and

i. $E\varepsilon_{ij} = 0, Var\varepsilon_{ij} = \sigma_i^2 < \infty$, for $i = 1,2$ and all j. $Cov(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$ for all $i, i', j,$ and j' unless $i = i'$ and $j = j'$.

ii. The ε_{ij} are independent under an unknown distribution.

Let the union of samples be represented by

$$y = (y_1, ..., y_n) = (x_{11}, ..., x_{1n_1}, x_{21}, ..., x_{2n_2}), \tag{52}$$

where, $n = n_1 + n_2$, and let the expected value and variance of its components be $Ey_i = \theta$ and $Var\ y_i = \sigma^2 < 0$, respectively. Now define

$$\begin{aligned} \beta_1 &\equiv \theta_2 - \theta_1 \\ \beta_2 &\equiv \theta - \theta_2 \end{aligned} \tag{53}$$

and consider the hypothesis

$$H: \begin{cases} \beta_1 = \beta_2 = 0 \\ \sigma_1^2 = \sigma_2^2 = \sigma^2 \end{cases}. \tag{54}$$

Without the Normality assumption in assumption (ii), deriving a test for hypothesis H can be quite difficult. But as I anticipate a relatively large sample size in anomaly detection applications using HS data, I shall rely again on the CLT to design this new detector.

The application of WLLL ensures us that the set of parameters $(\theta_1, \theta_2, \theta, \sigma_1^2, \sigma_2^2, \sigma^2)$ can be estimated by the following consistent estimators: $(\bar{x}_1, \bar{x}_2, \bar{y}, s_1^2, s_2^2, s^2)$, respectively, where

$$\begin{aligned} &\bar{x}_i = n_i^{-1} \sum_{k=1}^{n_i} x_{ik}, \quad \bar{y} = n^{-1} \sum_{k=1}^{n} y_k, n = n_1 + n_2; \\ &s_i^2 = \tfrac{1}{n_i - 1} \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_i)^2, \ i = 0,1; \\ &s^2 = \tfrac{1}{n-1} \sum_{k=1}^{n} (y_k - \bar{y})^2. \end{aligned} \tag{55}$$

Following (55),

$$\hat{\beta}_1 = \bar{x}_2 - \bar{x}_1 \tag{56}$$

and

$$\hat{\beta}_2 = \bar{y} - \bar{x}_2 \tag{57}$$

also constitute a set of consistent estimators for β_1 and β_2, respectively. Recall that statistical consistency implies that the estimator's mean is asymptotically unbiased (i.e., it converges to an intended value) and that its variance converges to *zero*, as the number of samples increases.

Using the independence assumptions in (51) and the results in (55), the expected values and variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ are readily attained as

$$E(\hat{\beta}_1) = \theta_2 - \theta_1, Var(\hat{\beta}_1) = \frac{1}{n_2}\sigma_2^2 + \frac{1}{n_1}\sigma_1^2; \tag{58}$$

and

$$E(\hat{\beta}_2) = \theta - \theta_2, Var(\hat{\beta}_2) = \frac{1}{n}\sigma^2 + \frac{1}{n_2}\sigma_2^2. \tag{59}$$

Using the independence assumptions and (58), (59), if the hypothesis *H* in (54) is true (let $\tau^2 = \sigma_1^2 = \sigma_2^2 = \sigma^2$), a direct application of CLT ensures that the random variables z_1 and z_2, below, converges ($\rightarrow$) in law to the standard Normal distribution $N(0,1)$, or

$$z_1 = \frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{Var(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{1}{n_1}\sigma_1^2 + \frac{1}{n_2}\sigma_2^2}} = \frac{\hat{\beta}_1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\tau} \xrightarrow[\substack{n_1\to\infty \\ n_2\to\infty}]{} N(0,1) \tag{60}$$

and equivalently

$$z_2 = \frac{\hat{\beta}_2 - E(\hat{\beta}_2)}{\sqrt{Var(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{1}{n_1}\sigma^2 + \frac{1}{n_2}\sigma_2^2}} = \frac{\hat{\beta}_2}{\sqrt{\frac{1}{n} + \frac{1}{n_2}}\tau} \xrightarrow[\substack{n\to\infty \\ n_2\to\infty}]{} N(0,1). \tag{61}$$

Using known properties of the family of chi square distributions, the following are also true:

$$z_1^2 = \frac{\hat{\beta}_1^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\tau^2} \xrightarrow[\substack{n_1\to\infty \\ n_2\to\infty}]{} \chi_1^2 \tag{62}$$

and

$$z_2^2 = \frac{\hat{\beta}_2^2}{\left(\frac{1}{n} + \frac{1}{n_2}\right)\tau^2} \xrightarrow[\substack{n\to\infty \\ n_2\to\infty}]{} \chi_1^2, \tag{63}$$

where χ_1^2 is the chi-square probability density function (pdf) with *1* degree of freedom (dof).

Under *H*, I propose two estimators of τ^2. One to be used in (62) and the other in (63), they are

$$S_1^2 = \frac{(n_2 - 1)s_2^2 + (n_1 - 1)s_1^2}{(n_2 - 1) + (n_1 - 1)} \tag{64}$$

and

$$S_2^2 = \frac{(n-1)s^2 + (n_2 - 1)s_2^2}{(n - 1) + (n_2 - 1)}, \tag{65}$$

respectively, where, s_i^2 ($i = 1,2$) and s^2 are defined in (55), and $n = n_1+n_2$. Using the same argument presented for the proof of consistency of (34), one can readily show

that both estimators S_1^2 and S_2^2 are consistent, under the null hypothesis H. Consistency of S_1^2 and S_2^2 also implies that the ratios shown below converges in probability to a constant, or

$$\kappa_1 = \frac{\tau^2}{S_1^2} \xrightarrow[\substack{n_1 \to \infty \\ n_2 \to \infty}]{} 1 \tag{66}$$

and

$$\kappa_2 = \frac{\tau^2}{S_2^2} \xrightarrow[\substack{n \to \infty \\ n_2 \to \infty}]{} 1. \tag{67}$$

Using (62), (63), (66), (67), and the Slutsky theorem, one can show that

$$Z_1 = z_1^2 \kappa_1 = \frac{\hat{\beta}_1^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\tau^2} \frac{\tau^2}{S_1^2} \xrightarrow[\substack{n_1 \to \infty \\ n_2 \to \infty}]{} \chi_1^2, \tag{68}$$

$$Z_2 = z_2^2 \kappa_2 = \frac{\hat{\beta}_2^2}{\left(\frac{1}{n} + \frac{1}{n_2}\right)\tau^2} \frac{\tau^2}{S_2^2} \xrightarrow[\substack{n \to \infty \\ n_2 \to \infty}]{} \chi_1^2, \tag{69}$$

and that under H, using a property of F distributions (49) with $p = q = 1$,

$$Z_{AFT} = \frac{Z_1}{Z_2} \xrightarrow[\substack{n \to \infty \\ n_2 \to \infty}]{} F_{1,1}. \tag{70}$$

3.4.3 Top View Anomaly Detection

To apply (70) to anomaly detection problem, notice that (70) is readily reformatted into

$$Z_{AFT} = \rho \frac{\hat{\beta}_1^2}{\hat{\beta}_2^2} \frac{S_2^2}{S_1^2}, \tag{71}$$

where, $\rho = \left(n^{-1} + n_2^{-1}\right)\left(n_1^{-1} + n_2^{-1}\right)^{-1}$, $\hat{\beta}_1^2$, $\hat{\beta}_2^2$, S_1^2, and S_2^2 are defined in (56), (57), (64), and (65), respectively.

Testing hypothesis *H* in (3.41) for local anomalies using (71) constitutes the AFT detector. A decision threshold *T* can be determined via $\int_T^\infty F_{1,1}(w)\,dw = \alpha$, where α is the type I error (i.e., the probability of rejecting *H*, given that *H* is true). The user chooses α, and for values of $Z_{AFT} > T$, hypothesis *H* is rejected implying that $x_1 = \left(x_{11}, ..., x_{1n_1}\right)$ and $x_2 = \left(x_{21}, ..., x_{2n_2}\right)$ are most likely sampled from different distributions; hence, they are anomalous to each other. Otherwise, they are likely sampled from the same distribution. Note that the comparison between x_1 and x_2 via (71), is done indirectly using Z_1 in (68), which holds information of both samples individually, and Z_2 in (69), which holds information of the union of samples *y* in (52).

Finally, for Z_{AFT} to converge in law to a central F distribution, Z_1 and Z_2 must be independent, which ultimately means that $\hat{\beta}_1$ and $\hat{\beta}_2$ must be independent. Let the

random variables u_1, u_2 and u_3 be statistically dependent and let $h_1 = u_2 - u_1$ and $h_2 = u_3 - u_2$. It can be shown (see, for instance, [26]) that h_1 is plausibly independent of h_2. This transformation is widely used by practicing statisticians so that dependent random variables can be addressed using techniques based on statistical independence. So, in (56) and (57), even if $\bar{x}_1$, $\bar{x}_2$, and $\bar{y}$ are dependent random variables, which most likely are the case for local samples in HS data, their differences $\hat{\beta}_1$ and $\hat{\beta}_2$ are independent random variables. Since κ_1 and κ_2 converge in probability to a constant and not to a distribution, there is no concern about the independence of S_i^2 from Z_i, $i = 1, 2$. Thus, Z_1 and Z_2 are independent random variables.

3.4.4 Ground View Anomaly Detection

For anomaly detection using ground-view imagery, I can fuse the individual results produced by each object class in the same manner that I described for the SemiP detector, i.e., for a given testing location in the image, I index the expression in (71) for each corresponding class and denote $Z_{AFT}^{(p)}$ the detector's output for the p^{th} class and collect the outputs for N classes; A single decision value at each testing location in the image is attained by

$$\widetilde{Z}_{AFT} = \min\left\{Z_{AFT}^{(1)}, Z_{AFT}^{(2)}, \cdots, Z_{AFT}^{(N)}\right\}. \tag{72}$$

In (72), if the local spectral radiance happens to be significantly anomalous to all N classes then $\widetilde{Z}_{AFT}$ would be a relatively large value. Otherwise, it would be a relatively small value indicating that the local spectral radiance probably belongs to at least one of the N classes. The null hypothesis for the multiclass problem using (72) is

$$H_6: \text{ For at least one class } p \begin{cases} \beta_1^{(p)} = \beta_2^{(p)} = 0 \\ \left(\sigma_1^2\right)^{(p)} = \left(\sigma_2^2\right)^{(p)} = \left(\sigma^2\right)^{(p)} \end{cases} \tag{73}$$

where $p = 1,\ldots,N$. Note that the model defined in (51) is implicitly indexed by p.

Testing the null hypothesis H_6 in (73) using (72), and the random sequences $y_0^{(p)}$ and $y_1^{(p)}$ in (15) as inputs, constitutes the adaptation of the AFT anomaly detector to the ground-level view problem. Also, it is expected that the recommendations discussed in this Subsection 3.4.3 referring to the implementation of this detector will be applied.

As it was the case for the multiclass version of the AsemiP detector, the AFT detector can also enjoy the same level of performance achieved by the SemiP detector, but free from parameter initialization. Results will be shown later in this section.

3.4.5 ANOVA F-distribution Test

ANOVA (analysis of variance) is one of the most widely used statistical techniques, and it called our attention for this report because it also yields an F test statistics. In

its simplest form, the ANOVA is a method of estimating the means of several populations often assumed to be normally distributed. The ANOVA, contrary to what its name infers, is not concerned with analyzing variances but rather with analyzing variation in population means. I will briefly describe the most common type of ANOVA, the oneway ANOVA. For a thorough treatment of the different facets of ANOVA designs, there is the classic text by [36].

In the oneway ANOVA, data (x_{ij}) are assumed to be independent observations, according to the model

$$x_{ij} \sim N(\mu_i, \sigma^2), \quad i = 1,..., k, \quad j = 1,..., n_i. \tag{74}$$

In other words, data are normally distribution having unknown equal or unequal means, μ_i, but having the same variance σ^2; these parameters are unknown.

The classical ANOVA test is a test of the null hypothesis H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$. I want to make inferences about μ_i's without the knowledge of σ^2. Therefore, I want to replace σ^2 with an estimate. In each population, if I denote the sample variance by s_i^2 and the sample mean by $\bar{x}_i$,

$$s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \; \bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}, \quad i = 0,..., k, \tag{75}$$

then s_i^2 is an estimate of σ^2, and by property of Normal family of distributions (see, for instance, in [35])

$$(n_i - 1) \frac{s_i^2}{\sigma^2} \sim \chi^2_{n_i - 1}. \tag{76}$$

Furthermore, under the ANOVA assumptions, since each s_i^2 estimates the same σ^2, one can improve the estimators by combining them, or

$$s_p^2 = \frac{1}{N - k} \sum_{i=1}^{k} (n_i - 1) s_i^2, \tag{77}$$

where, $N - k = \sum (n_i - 1)$. Since s_i^2 are independent, using properties of chi square distributions,

$$Z_{2'} = (N - k) \frac{s_p^2}{\sigma^2} \sim \chi^2_{N-k}. \tag{78}$$

It can also be shown (see, for instance, [36]) that

$$Z_{1'} = \frac{1}{\sigma^2} \sum_{i=1}^{k} n_i \left[(\bar{x}_i - \bar{\bar{x}}) - (\mu_i - \bar{\mu}) \right]^2 \sim \chi^2_{k-1}, \tag{79}$$

where,

$$\overline{\overline{x}} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{x_{ij}}{n_i}, \quad \overline{\mu} = \sum_{i=1}^{k} \frac{\mu_i}{k}, \quad i = 0,..., k, \tag{80}$$

and that

$$Z_{ANOVA} = \frac{Z_{1'}}{Z_{2'}} \sim F_{k-1, N-k}. \tag{81}$$

Note in (81) that, contrary to the result shown in (70), Z_{ANOVA} is governed *exactly* by an F distribution of $k-1$ and $N-k$ degrees of freedom. Thus, it does not need to rely on large sample theory to arrive at (81). The independent observation Normal assumption in (74) plays a major role for arriving at (78), (79) and for concluding that $Z_{1'}$ and $Z_{2'}$ are independent.

If $k = 2$, which is in our case, and the null hypothesis H_0: $\mu_1 = \mu_2$ (given that the variances are the same) is true, then $\mu_1 = \mu_2 = \overline{\mu}$ and the $\mu_i - \overline{\mu}$ terms drop out of (79); I would reject H_0 if

$$Z_{ANOVA} = \frac{\sum_{i=1}^{2} n_i \left(\overline{x}_i - \overline{\overline{x}}\right)^2}{s_p^2} > F_{1, n_1+n_2-2, \alpha}, \tag{82}$$

where α is the chosen type I error and $F_{1, n_1+n_2-2, \alpha}$ is the threshold that yields α. Of course, the quality of the detector in (82) will be dependent on how close the data satisfy the assumptions of sample normality having equal variances.

3.5 Asymmetric Variance-Based Hypothesis Test

In this section, I present our fourth and last technique, a technique also based on the same fundamental theorems of large sample theory. This time, however, I aim at designing (arguably) the most compact form to implement the notion of indirect comparison discussed in Section 1. I will show how a simple asymmetric hypothesis test based only on a central moment can be designed to exploit the distinction between two samples. I first derive this fourth technique (for convenience referred to in this text as the asymmetric variance test (AVT) detector), followed by its adaptation to both types of imagery.

3.5.1 Derivation

Suppose that two random samples x_0 and x_1 are observed according to the model

$$\begin{aligned} x_1 &= (x_{11},..., x_{1n_1}) \quad iid \sim g_1(x) \\ x_0 &= (x_{01},..., x_{0n_0}) \quad iid \sim g_0(x), \end{aligned} \tag{83}$$

where, x_1 (the test random sample of size n_1) and x_0 (the reference random sample of size n_0) are independent, g_1 and g_0 are unknown, and that

$$Ex_{0j} = \mu_0, \quad Var\ x_{0j} = \sigma_0^2 < \infty,$$
$$Var\left(x_{0j} - \mu_0\right)^2 = \zeta_0^2 < \infty. \tag{84}$$

Consider the hypotheses

$$H_0 : \sigma_0^2 = \tau \quad (\tau > 0)$$
$$H_1 : \sigma_0^2 \neq \tau. \tag{85}$$

In (85), I would like to test the null hypothesis that the variance from a reference sample is equal to an arbitrary positive value. At a glance, the null hypothesis does not seem to be very effective, since τ can take any positive value, and the variance excludes an additional discriminant feature: the mean.

However, one can cleverly incorporate the indirect comparison approach discussed earlier to test (85), designing in the process a rather effective anomaly detector. A solution follows.

Let t represent the union of x_1 and x_0, or

$$t = (t_1,..., t_n) = (x_{01},..., x_{0n_0}, x_{11},..., x_{1n_1}), \tag{86}$$

where, $n = n_0+n_1$, and suppose that under certain conditions the components of t have the same finite variance, i.e., $Var\, t_k = \sigma_u^2 < \infty$. The last assumption may not be satisfied for all t_k, but would certainly be satisfied when x_1 and x_0 are sampled from the same population, in which case one could set $\tau = \hat{\sigma}_u^2$ in (85), where $\hat{\sigma}_u^2$ estimates σ_u^2, and test for validity of this equality.

Denoting the symbols $>>$ as *much greater then*, $\approx$ as *approximately equal to*, $\in$ as *belonging to*, and $P(\cdot)$ as the population of a random variable, the implications of setting $\tau = \hat{\sigma}_u^2$ in (85), using the symbols of the study cases in figure 1, are as follows:

- Case 1: $x_0 \in P(X)$ and $x_1 \in P(Y)$ would imply that $\hat{\sigma}_u^2 >> \sigma_0^2$, yielding a *strong anomaly*, since the difference between $P(\hat{Z})$ and $P(X)$ is so significant—especially for tight distributions with their first moments significantly different from each other.
- Case 2: $x_0 \in P(S)$ and $x_1 \in P(Y)$ would imply that $\hat{\sigma}_u^2 < \sigma_0^2$ or $\hat{\sigma}_u^2 \approx \sigma_0^2$, yielding a *softer anomaly*, since $P(\tilde{Z})$ and $P(S)$ have the same overall characteristics: they are bimodal.
- Case 3: $x_0 \in P(Y)$ and $x_1 \in P(Y)$ would imply that $\hat{\sigma}_u^2 \approx \sigma_0^2$, yielding a *non-anomaly*—a trivial case not included in figure 1.

Without the Normality assumption in (83), deriving a statistic of known distribution to test the null hypothesis in (85) can be quite difficult. Hence, I shall rely again on the CLT to design the new detector. (Our past experience using HS data has ensured us that a sample size greater than *30* satisfies the large sample requirement in methods based on large sample theory.)

Applying WLLN the set of parameters (μ_0, σ_0^2) can be estimated by the set of *consistent* estimators $(\bar{x}_0, s_0^2)$, respectively, where

$$\bar{x}_0 = \sum_{j=1}^{n_0} \frac{x_{0j}}{n_0}, s_0^2 = \sum_{j=1}^{n_0} \frac{(x_{0j} - \bar{x}_0)^2}{n_0 - 1}. \tag{87}$$

Following (87), a direct application of CLT, using the denotations in (84), ensures us that the random variable z_1, below, converges in law to the standard normal distribution $N(0,1)$, as the sample size n_0 increases (see, for instance, [37]),

$$z_1 = \sqrt{n_0} \frac{s_0^2 - \sigma_o^2}{\sqrt{\zeta_0^2}} \xrightarrow[n_0 \to \infty]{} N(0,1). \tag{88}$$

To estimate ζ_0^2 using a consistent estimator $(\hat{\zeta}_0^2)$, consider this rationale: Let $\vartheta_j = (x_{0j} - \mu_0)^2$ and note that, based on (84), $E(\vartheta_j) = \sigma_0^2$ and $Var(\vartheta_j) = \zeta_0^2$. A consistent estimator of $Var(\vartheta_j)$ then would qualify for application in (88). An obvious estimator of $Var(\vartheta_j)$ is $\hat{V}_\vartheta = \sum_{j=1}^{n_0} \frac{(\vartheta_j - \bar{\vartheta})^2}{n_0 - 1}$, where $\bar{\vartheta}$ is the sample average using all ϑ_j's. Notice that $\hat{V}_\vartheta$ can be also expressed by the following decomposition $\hat{V}_\vartheta = n_0 (n_0 - 1)^{-1} \left\{ n_0^{-1} \sum_{i=1}^{n_0} (\vartheta_i - \sigma_0^2)^2 - (\bar{\vartheta} - \sigma_0^2)^2 \right\}$, where the normalized summation term (which does not include $\bar{\vartheta}$) tends to ζ_0^2 in probability by the WLLN, and the term that includes $\bar{\vartheta}$ tends to *zero* in probability also by the WLLN. Therefore, $\hat{V}_\vartheta$ is a consistent estimator of ζ_0^2. In addition, using results from (87), notice that s_0^2 is also a consistent estimator of $E(\vartheta_j)$. I then propose the following consistent estimator of $\zeta_0^2 = E[\vartheta_j - E(\vartheta_j)]^2$:

$$\hat{\zeta}_0^2 = \sum_{j=1}^{n_0} \frac{\left[(x_{0j} - \bar{x}_0)^2 - s_0^2 \right]^2}{n_0 - 1}. \tag{89}$$

Consistency of (89) implies that the ratio κ, below, converges in probability to a constant, as the sample size increases, or

$$\kappa = \frac{\zeta_0^2}{\hat{\zeta}_0^2} \xrightarrow[n_0 \to \infty]{} 1. \tag{90}$$

which also implies that $\sqrt{\kappa} \xrightarrow[n_o \to \infty]{} 1$. Setting $\tau = \hat{\sigma}_u^2$ in (85), where

$$\hat{\sigma}_u^2 = \sum_{j=1}^{n} \frac{\left(t_j - \bar{t}\right)^2}{n-1}, \quad \bar{t} = \sum_{j=1}^{n} \frac{t_j}{n}, \quad n = n_0 + n_1, \tag{91}$$

if the null hypothesis in (85) is true with $\tau = \hat{\sigma}_u^2$, using (88), (90) and the application of the *Slutsky* theorem, the following must also be true:

$$z_2 = \sqrt{n_0} \frac{s_0^2 - \hat{\sigma}_u^2}{\sqrt{\zeta_0^2}} \sqrt{\kappa} \xrightarrow[n_0 \to \infty]{} N(0,1). \tag{92}$$

The next two subsections show how to adapt (92) to the two problem types discussed in this section, top view and ground level view

3.5.2 Top View Anomaly Detection

Squaring the standard-normal, random variable z_2 in (92), yields under the null hypothesis with $\tau = \hat{\sigma}_u^2$ the chi square distribution shown below,

$$Z_{AVT} = z_2^2 = n_0 \frac{\left(s_0^2 - \hat{\sigma}_u^2\right)^2}{\hat{\zeta}_0^2} \xrightarrow[n_0 \to \infty]{} \chi_1^2, \tag{93}$$

where χ_1^2 is the chi-square pdf with *1* degree of freedom (dof).

Testing hypothesis H_0 in (85) using (93) constitutes the AVT anomaly detector for the top-view problem. A decision threshold T can be determined via $\int_T^\infty \chi_1^2(w)dw = \alpha$, where α is the type I error (i.e., the probability of rejecting H_0, given that H_0 is true). The user chooses α, and for values of Z_{AVT} greater then T, hypothesis H_0 is rejected, implying that x_0 and x_1 are most likely sampled from different populations. Hence, they are anomalous to each other. Otherwise, they are not significantly anomalous to each other.

Note that the comparison between x_0 and x_1 via (93) is performed indirectly using $\hat{\sigma}_u^2$, which holds information about the union of the samples t, and the other estimators, which only hold information about the reference sample x_0.

3.5.3 Ground View Anomaly Detection

For anomaly detection using ground-view imagery, I fuse the individual results produced by each object class in the same manner that I described for the SemiP detector. In other words, for a given testing location in the image, I index the expression in (93) for each corresponding class and denote $Z_{AVT}^{(p)}$ the detector's output for the p^{th} class and collect the outputs for N classes; a single decision value of this detector per testing location is attained by

$$\widetilde{Z}_{AVT} = \min\left\{Z_{AVT}^{(1)}, Z_{AVT}^{(2)}, \cdots, Z_{AVT}^{(N)}\right\}. \tag{94}$$

In (94), if the local spectral radiance happens to be significantly anomalous to all N classes then $\widetilde{Z}_{AVT}$ would be a relatively large value; otherwise, it would be relatively small, indicating that the local spectral radiance probably belongs to at least one of the N classes. Two likely hypotheses for the multiclass problem using (94) are

$$\begin{aligned} H_7 &: \text{ at least one } \left(\sigma_0^2\right)^{(p)} = \tau^{(p)} \quad \left(\tau^{(p)} > 0; p = 1, \cdots, N\right) \\ H_8 &: \text{ all } \left(\sigma_0^2\right)^{(p)} \neq \tau^{(p)}. \end{aligned} \tag{95}$$

where $\left(\sigma_0^2\right)^{(p)}$ is the reference sample variance associated with the corresponding p^{th} class, $\tau^{(p)} = \left(\sigma_u^2\right)^{(p)}$, $\left(\sigma_u^2\right)^{(p)}$ is the sample union variance associated with the p^{th} class.

Testing the null hypothesis H_7 in (95) using (94) constitutes the adaptation of the AVT anomaly detector to the ground-level view problem.

3.6 Analysis: Power of the Test

In deciding to accept or reject the null hypothesis, a detector is expected to make mistakes. Usually, hypothesis tests are evaluated and compared through their probabilities of making mistakes, as I discussed in Section 2. In this section, I discuss how these error probabilities can be determined, or at least approximated, for both types of problems, top view and ground level view. For the top view anomaly detection problem, I will use as examples two of the algorithms covered in this section—the AVT and AsemiP detectors. For the ground view anomaly detection problem, I will use as an example the AVT detector, since the analyses of its error probabilities are readily applicable to any detector that is asymptotically governed by a known distribution family, under its null hypothesis.

3.6.1 Top View

Using as a reference the AVT detector shown in (93), figure 6 shows a decision threshold T separating two hypotheses H_0 and H_1. In the context of our discussion, values in the abscissa Z_{AVT} greater than T are automatically labeled as anomalies. As it was discussed earlier, decision errors are unavoidable. I would like to know whether the asymptotic behaviors of these errors can be determined, and whether they are favorable. The *power function* can provide those answers. The power function of the AVT detector for the top view problem is the following:

$$\psi\left(\sigma_0^2\right) = \begin{cases} P_{\sigma_0^2=\tau}\left(Z_{AVT} > T\right) \\ P_{\sigma_0^2\neq\tau}\left(Z_{AVT} > T\right) \end{cases}. \tag{96}$$

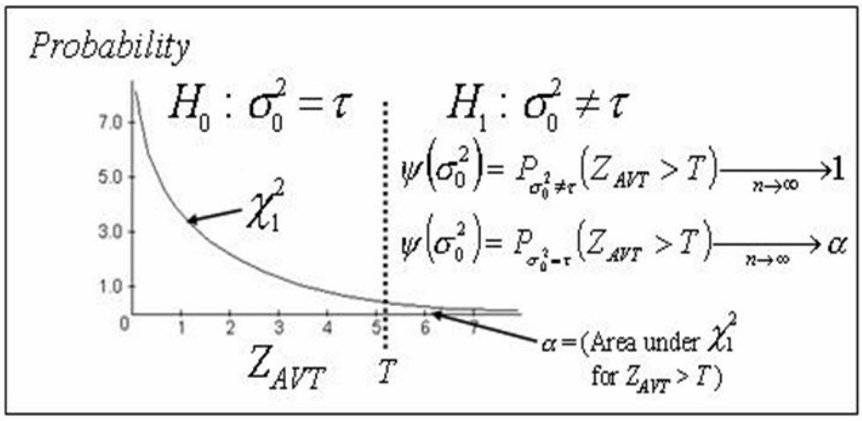

Figure 6. The asymptotic behavior of the AVT anomaly detector and the desirable asymptotic behavior of its *power function* ψ for the top view problem.

In essence, the power function Ψ yields the cumulative probability P of rejecting the null hypothesis H_0, when either H_0 ($\sigma_0^2 = \tau$) or H_1 ($\sigma_0^2 \neq \tau$) is true. This rejection region is $Z_{AVT} > T$, where Z_{AVT} is defined in (93) and T is a decision threshold. Notice in (96) that Ψ under H_0 corresponds to the well known type I error (i.e., the probability of rejecting H_0, given that H_0 is true) and that Ψ under H_1 corresponds to the complement of the type II error (i.e., one minus the probability of rejecting H_1, given that H_1 is true). The type I and type II errors constitute the only error types encountered in the context of our discussion. In the ideal case, ψ yields *0* when H_0 is true and *1* when H_1 is true. Except in trivial situations, this ideal cannot be attained. So, one of our goals is to show that Ψ tends in probability to α (a scalor controlled by the user), when the null hypothesis H_0 is true, and that Ψ tends in probability to *1*, when the alternative hypothesis H_1 is true. Figure 6 illustrates this desirable behavior.

In this subsection, the equality $\tau = \sigma_u^2$ (the sample variance from the union of two observed sequences) is always set to be true for every location in the image. If H_o in (85) is true, the AVT detector has the asymptotic behavior shown in (93), and the type-I error is readily obtained by

$$P_{\sigma_0^2=\tau}(Z_{AVT} > T) \xrightarrow[n\to\infty]{} P(\xi > T) = \alpha, \tag{97}$$

where ξ is a chi-square distributed random variable with *1* degree of freedom, Z_{AVT} as defined in (93), and T a nonnegative real value.

Setting $\psi(\sigma_0^2) = P_{\sigma_0^2=\tau}(Z_{AVT} > T)$, ψ is indeed an asymptotic size α test, which is controlled by the user.

Now consider an alternative parameter value, such that $\sigma_0^2 \neq \tau$, and let $\eta = \sigma_0^2 - \tau \neq 0$. From (93) I can write

$$Z_{AVT} = \left\{ \left[\underbrace{\left(\frac{\left(s_0^2 - \tau\right) - \eta}{\sqrt{\frac{\zeta_0^2}{n_0}}} \right)}_{A} + \underbrace{\left(\frac{\eta}{\sqrt{\frac{\zeta_0^2}{n_0}}} \right)}_{B} \right]^2 \underbrace{\left(\frac{\zeta_0^2}{\hat{\zeta}_0^2} \right)}_{C} \right\}. \tag{98}$$

Since, for a constant τ, $E\left(s_0^2 - \tau\right) = \eta$ and $Var\left(s_0^2 - \tau\right) = Var\ s_0^2 = \zeta_0^2 / n_0$, the application of CLT ensures us that term *A* in (98) will converge in distribution to the standard Normal, *N(0,1)*, as n_0 tends to $+\infty$, no matter what the value of η. Notice that term *B* will converge to $+\infty$ or $-\infty$ in probability, as *n* goes to $+\infty$, depending on whether η is positive or negative. And finally, notice also that no matter what the value of η, the estimator proposed in (89) is consistent. The term *C* then will converge to *1* in probability, as n_0 goes to $+\infty$. Thus, Z_{AVT} will converge in probability to $+\infty$, and the probability of rejecting H_o (given that $\sigma_0^2 \neq \tau$) tends to *1*, or

$$P_{\sigma_0^2 \neq \tau}\left(Z_{AVT} > T\right) \xrightarrow[n \to \infty]{} 1. \tag{99}$$

In this way, the AVT anomaly detector for the top-view problem has the properties of asymptotic size α and asymptotic power *1,* as it is desired.

Similar analysis can be made for other detectors presented in this section, for instance, using now as a reference the AsemiP detector, consider the following: If H_o is true in Proposition 1, the type I error probability is

$$P_{\tilde{\beta}=0}\left(Z_{AsemiP} > \gamma\right) \xrightarrow[(n_0, n_1) \to \infty]{} P(\xi > \gamma) = \alpha, \tag{100}$$

where ξ is a chi-square distributed random variable with *1* degree of freedom, Z_{AsemiP} as defined in (31) and expressed in a different form in (46), and γ an arbitrary scalor. $P_{\tilde{\beta}=0}\left(Z_{AsemiP} > \gamma\right)$ is indeed an asymptotically size α test, which is controlled by the user.

Now consider an alternative parameter value $\tilde{\beta} \neq 0$. In this case, $\sigma_0^2 \neq \sigma_1^2$ and from (46) I can write

$$Z_{AsemiP}=\left\{\left[\underbrace{\left(\frac{\hat{\tilde{\beta}}-\tilde{\beta}}{\sqrt{\frac{1}{n_0}\sigma_0^2+\frac{1}{n_1}\sigma_1^2}}\right)}_{A}+\underbrace{\left(\frac{\tilde{\beta}}{\sqrt{\frac{1}{n_0}\sigma_0^2+\frac{1}{n_1}\sigma_1^2}}\right)}_{B}\right]^2\underbrace{\left(\frac{S_t^2}{S^4}\right)}_{C}\right\}. \quad (101)$$

Note that the term *A* in (101) converges in distribution to the standard Normal, *N(0,1)*, as n_0 and n_1 go $+\infty$, no matter what the values of $\tilde{\beta}$, σ_0^2, or σ_1^2 are. Note also that the term *B* converges to $+\infty$ or $-\infty$ in probability, as n_0 and n_1 go $+\infty$, depending on whether $\tilde{\beta}$ is positive or negative. S^2 converges in probability to zero, as does S_t^2, but the term *C* converges to $+\infty$ because $\left(S^2\right)^2 = S^4$ is in the denominator. Thus, Z_{AsemiP} converges to $+\infty$ in probability and

$$P_{\tilde{\beta}\neq 0}\left(\text{reject}\,H_0\right)=P_{\tilde{\beta}\neq 0}\left(Z_{AsemiP}>\gamma\right)\xrightarrow[(n_0,n_1)\to\infty]{}1. \quad (102)$$

In this way, the test in Proposition 1 also has the properties of asymptotic size α and asymptotic power *1,* as desired.

3.6.2 Ground-Level View

From the discussion in this section, I learned that the output of the AVT detector for top-view anomaly detection problem has two asymptotic outcomes:

$Z_{AVT} \xrightarrow[n_0\to\infty]{} \chi_1^2$ (in distribution, if the null hypothesis is true) or

$Z_{AVT} \xrightarrow[n_0\to\infty]{} +\infty$ (in probability, if H_1 is true). For the ground view anomaly detection, refer to the null hypothesis H_7 and the alternative hypothesis H_8 shown in 957), and consider the following: For a given spatial location in a ground-view imagery, let $Z_{AVT}^{(p)}$ be the AVT detectors' output for the p^{th} object class, and assume, without loss of generality, that each one of the first *W* outputs in the independent sequence of results ($1 \le W \le N$; *N* is the total number of classes) has the asymptotic chi-square behavior shown in (93), and that each one of the remaining results have the asymptotic behavior shown in (99). Using (94) and results from this section (Top View), I have

$$\widetilde{Z}_{AVT} = \min\left\{\begin{array}{l} Z_{AVT}^{(1)} \xrightarrow{n\to\infty} \chi_1^2 \\ Z_{AVT}^{(2)} \xrightarrow{n\to\infty} \chi_1^2 \\ \vdots \\ Z_{AVT}^{(W)} \xrightarrow{n\to\infty} \chi_1^2 \\ Z_{AVT}^{(W+1)} \xrightarrow{n\to\infty} +\infty \\ \vdots \\ Z_{AVT}^{(N)} \xrightarrow{n\to\infty} +\infty \end{array}\right\}. \qquad (103)$$

Notice in (103) that $\widetilde{Z}_{AVT}$ is bounded and that, as $n \to \infty$, $\widetilde{Z}_{AVT}$ will converge in law to the distribution of the lowest *order statistics*. (The order statistics of a random sample $Z_1, \ldots, Z_n$ are the sample values placed in ascending order. They are often denoted by $Z_{(1)}, \ldots, Z_{(n)}$, where $Z_{(1)} = \min_{i\le i\le n} X_i$ and $Z_{(n)} = \max_{i\le i\le n} X_i$.) To attain an approximation of the type I error using (103), I first ignore all the components in (103) that converge in probability to $+\infty$, since they do not converge in distribution but in probability. Then I consider only the components that converge in distribution, i.e., $\left(Z_{AVT}^{(1)}, Z_{AVT}^{(2)}, \cdots, Z_{AVT}^{(W)}\right)$. The distribution of $Z_{AVT(1)} = \min_{i\le i\le W} Z_{AVT}^{(i)}$ from the culled sequence can be attained with the application of Theorem 3.2.

Theorem 3.2: Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample from a continuous population with cumulative distribution function (cdf) F(x) and pdf f(x). Then the pdf of $X_{(j)}$ is

$$f(x) = \frac{n!}{(j-1)!(n-j)!} f(x)\left[F(x)\right]^{j-1}\left[1-F(x)\right]^{n-j}, \qquad (104)$$

where $(\cdot)!$ denotes the factorial operator.

The proof of Theorem 3.2 can be found in [35].

Setting j = 1 and n = W in (104), the pdf of $Z_{AVT(1)}$ is

$$g(z) = Wf(z)\left[1 - F(z)\right]^{W-1}, \qquad (105)$$

where *f(z)* is the chi square pdf with *1* degree of freedom in the case of the AVT detector (also for SemiP and AsemiP), and *F(z)* is the equivalent cdf. Note that *f(z)* is the $F_{1,1}$ for the AFT detector.

Denote $\left(\sigma_0^2\right)^{(m)}$ the reference sample variance associated with the minimum order statistics $Z_{AVT(1)}$, and let $\left(\sigma_u^2\right)^{(m)}$ be the combined sample variance associated with $Z_{AVT(1)}$. As the number of the reference sample associated with $Z_{AVT(1)}$, $n = n^{(m)}$,

increases; setting $\tau^{(m)} = \left(\sigma_u^2\right)^{(m)}$, the probability of rejecting H_7 in (3.87), when $\left(\sigma_0^2\right)^{(m)} = \tau^{(m)}$, converges to

$$\hat{\psi}\left[\left(\sigma_0^2\right)^{(m)}\right] = P_{\left(\sigma_0^2\right)^{(m)}=\tau^{(m)}}\left(\widetilde{Z}_{AVT} > T_1\right) \approx P_{H_2}\left(Z_{AVT(1)} > T_1\right) \xrightarrow[n\to\infty]{} P(\xi > T_1) = \varepsilon, \quad (106)$$

where $\approx$ means *approximately equal to*, ξ is a random variable distributed by *g(z)*, as defined in (105); T_1 a nonnegative real value; and ε a positive real value, controlled by the user. The variable $\hat{\psi}$ in (106) is the type I error, and it is indeed an asymptotically size ε test.

Now consider the alternative hypothesis H_8 in (95), where all $\left(\sigma_0^2\right)^{(i)} \neq \tau^{(i)} = \left(\sigma_u^2\right)^{(i)}$, $i = 1, \cdots, N$. From (103) I write

$$\widetilde{Z}_{AVT} = \min\left\{\begin{array}{c} Z_{AVT}^{(1)} \xrightarrow{n\to\infty} +\infty \\ \vdots \\ Z_{AVT}^{(N)} \xrightarrow{n\to\infty} +\infty \end{array}\right\}. \quad (107)$$

From (107), $\widetilde{Z}_{AVT}$ will converge in probability to $+\infty$. Hence, the probability of rejecting the null hypothesis H_7 [given that all $\left(\sigma_0^2\right)^{(i)} \neq \tau^{(i)}$] tends to *1*, or

$$P_{H_8}\left(\widetilde{Z}_{AVT} > T_1\right) \xrightarrow[n\to\infty]{} 1. \quad (108)$$

In this way, the AVT expression in (94) for the ground-view problem has the desired properties of asymptotic size ε, which is controlled by the user, and asymptotic power *1*. (This discussion is readily applicable to any anomaly detector, whose test statistic tends in law to a known distribution, including of course the detectors SemiP, AsemiP, AFT, and RX.)

3.7 Results and Discussion

In this section, the performance of the conventional and non-conventional anomaly detectors that were previously discussed are evaluated using the imagery collected by two sensors: (i) the HYDICE sensor, which provided top view perspectives, and (ii) the commercially available hyperspectral sensor SOC-700, which provided ground view perspectives. I will start off by making a few comments on the data preprocessing used for the different types of algorithms, and proceed by showing various performance results for these detectors operating on the top-view and ground-view imagery. I also include a subsection describing a proof of principle experiment, where the discriminant power of an anomaly detector is adapted to function as an unsupervised learning classifier.

3.7.1 Data Preprocessing

In this subsection, I discuss the data preprocessing used for the four approaches based on the union of samples, followed by the data preprocessing used for the other types of detectors. The discussion in this subsection is relevant to both types of imagery, top view and ground level view.

The models based on the union of samples (i.e., SemiP, AsemiP, AFT, and AVT) are clearly based on idealized assumptions. In the context of using relatively high resolution imagery, at best one could hope that, in the presence of certain types of terrain, those assumptions would not be grossly violated. The assumptions dictate that not only the random samples x_0 must be statistically independent of x_1, their corresponding components x_{0j} and x_{1j} must be *iid*. For those assumptions not to be violated using HS data, the information in the spatial domain must be independent, as well as the information in the spectral domain.

In Subsection 3.1, I proposed two transformations aimed at promoting statistical independence in both domains: apply a HP filter in the spectral domain, followed by a spatial SAM. Both transformations use the same basic idea: They take the difference between dependent random variables. For instance, the difference among three highly dependent random variables will produce two independent random variables. The application of a HP filter, which is equivalent to a first order differentiation, in the spectral domain, followed by an angle difference mapping jointly will produce approximately *iid* random variables. An output result depicting transformed spectral samples, as described, was already shown in figure 4. The data preprocessed just discussed were used for all the detectors based on the union of samples, and for the ANOVA method.

For the other anomaly detectors discussed in this section (i.e., FLD, DPC, EST, RX, and KRX), I used the data preprocessing suggested by their corresponding authors. In other words, I applied a HP filter in the spatial domain of the actual reference and test hyperspectral samples (thus promoting spatial independence) and aimed at capitalizing on the spectral correlation of natural clutter background, which in essence constitutes the rationale for the development of detectors RX and KRX. Since neither the detectors DPC and EST are based on an assumed statistical model, but on principal component decomposition, data preprocessing was not applied to the actual hyperspectral samples.

3.7.2 HYDICE Top-View Hyperspectral Imagery

An experiment was carried out on data set from the hyperspectral digital imagery collection experiment (HYDICE) sensor. Recall that the HYDICE sensor records *210* spectral bands in the visible-to-near infrared (VNIR) and short-wave infrared (SWIR), *0.4-2.5* μm, forming a cube of spatially registered pixels. Each pixel then in the scene represents a sequence with *210* components.

To challenge this new family of local anomaly detectors, I used two sub-cubes from the HYDICE dataset. These sub-cubes depict the radiance from two different types of

terrains, forest and desert. The imagery used are the so-called Forest Radiance I (FR-I)—the same one used to test the conventional anomaly detectors, see figure 2—and Desert Radiance II (DR-II). Their spectral averages—from *150* bands—were shown in figure 3, as two dimensional (*2-dim*) images. (Water absorption and low signal-to-noise ration bands were discarded. Hence, only remaining *150* bands were used; the discarded bands are: 23^{rd}-101^{st}, 109^{th}-136^{th}, and 152^{nd}-194^{th}.) Recall that in DR-II, five stationary military vehicles can be observed aligned on a road in Yuma, AZ. In FR-I, *14* stationary military vehicles can be observed on sparse grasses, near a forest in Aberdeen, MD. The military vehicles in both scenes are considered as the targets in this report; they vary in sizes in both images. The HS images shown in figure 3 were magnified differently to fit in the same capture. FR-I consists of 600 x 140 pixels with a ground resolution of about *1.3*-m per pixel. The DR-II subcube consists of about 320 x 140 pixels with the same ground resolution per pixel.

The goal of local anomaly detectors on these types of scenes is to hopefully detect all objects that seems clearly anomalous to its immediate surroundings in some predetermined feature space. The local sampling mechanism is discussed next.

The SemiP, AsemiP, AFT, ANOVA, and AVT detectors were implemented using the sampling mechanism discussed in *Section: Formulation of Problems* and data preprocessing as discussed in *Subsection: Data Preprocessing*. The test window consisted of *9* pixels, while the reference window consisted of *56* pixels, and the variability window consisted of *60* pixels. The random sequence of angles obtained from the variability and the reference windows, as described in *Section: Formulation of Problems*, was labeled as the reference random sample x_0 for the new detectors. The sequence of angles obtained from the variability and test windows was labeled as the test random sample x_1. Note that the size of the *variability window* determines the size of the random samples, that is, x_0 and x_1 have the same size, *60.*

From empirical results using the top-view imagery, I learned that sample sizes above *40* comfortably satisfied the large sample requirement of the detectors based on large sample theory, i.e., estimated values did not change significantly using additional samples. The reason may be related to the pixel resolution of *1.3-m* of the HYDICE imagery being relatively low, which implies that the radiances from multiple objects (e.g., grass and dirt) were integrated in the sensor as being originated from a single object. The number of pixels in a single object, which is dependent on the sensor's pixel resolution and on the altitude of the data collection platform, will possibly influence the minimum required sample size for any method based on large sample theory. This dependence, however, is not very sensitive, as the reader will be able to verify later in this report.

Figure 7 shows again the output surfaces of the conventional detectors FLD, DPC, RX, and EST on the FR-I data, which were also shown in figure 2, in addition to the output performance of ANOVA. Figure 8 depicts output surfaces of the detectors based on the union of samples testing the FR-I data, in addition to KRX's output surface.

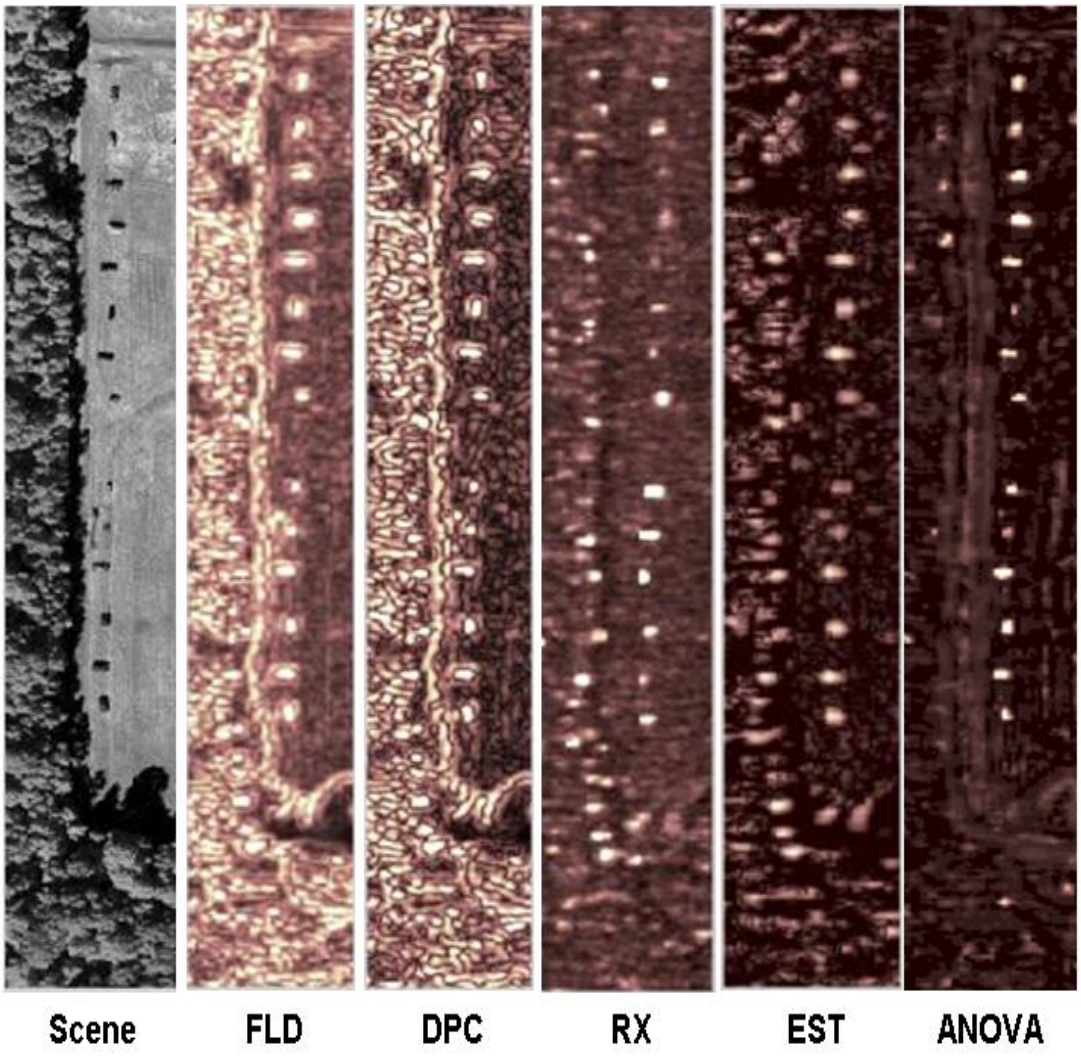

Figure 7. Decision surfaces using the HYDICE FR-I data, forest radiance. The intensities and heights of local peaks reflect the strength of anomaly evidences as *seen* by different detectors.

Note that the surfaces of FLD, DPC, RX, KRX, and EST did not require a suitable clipping threshold for the purpose of display. On the other hand, the detectors SemiP, AsemiP, AFT, AVT, and ANOVA required the application of suitable thresholds for the only purpose of display. All ten output surfaces shown in figure 7 and figure 8 were mapped using the same 2^8 pseudo-color map (colormap), as shown.

Notice in figure 8 that, for a particular initialization (i.e., $\left[\hat{\alpha}, \hat{\beta}\right] = [0,0]$), the SemiP detector suppresses very well what would be considered by an image analyst as meaningless detections from forest radiance, and that the other detectors based on the same principle of indirect comparison, but having no dependence on initial conditions, performs about the same.

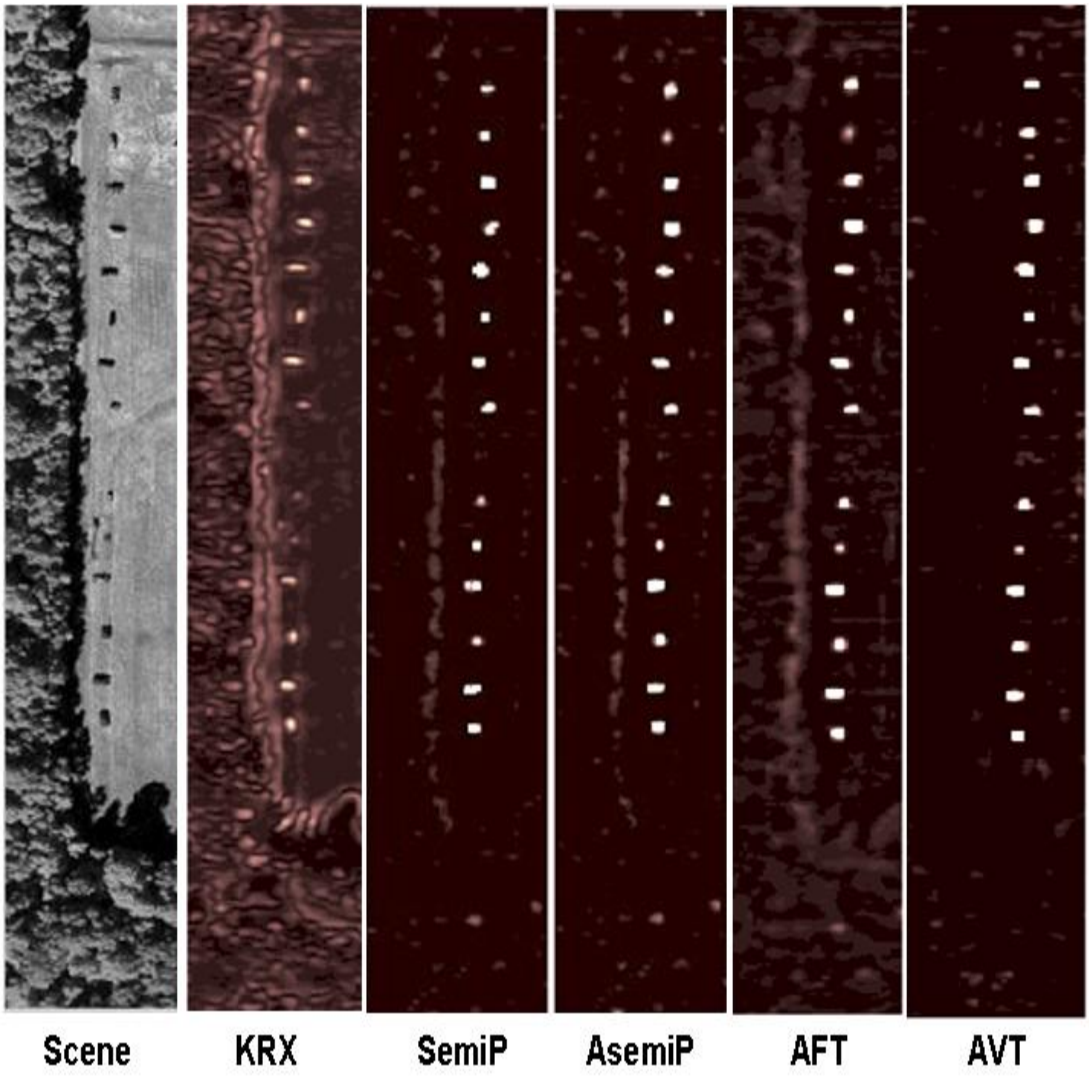

Figure 8. Decision surfaces for the HYDICE FR-I data.

By visual inspection alone of the output surfaces shown in figure 7, figure 8, one would be hard pressed not to ignore the advantage of applying our proposed principle of indirect comparison to the problem of local anomaly detection. These output surfaces suggest that our semiparametric and nonparametric detectors outperform conventional techniques by being able to significantly suppress *noise*, hence, accentuating in that scene the presence of meaningful objects.

Recall that the detectors SemiP, AsemiP, AFT, and AVT are based on the union of samples and that their assumptions do not depend on parametric models. Recall that the ANOVA detector does depend on the normality assumption, albeit it enjoys—partially—the advantage of using the union of samples by comparing the means of individual random samples to the mean of the random samples combined. Recall also that the theories of FLD, RX, and KRX detectors are based on the properties of normality and that DPC and EST detectors are merely based on the scores of random samples on the Eigen space. These differences explain the disparity in performance

between the two groups. For instance, when the spectral radiance of a grassy area is compared to a composite set of grass and shadow, the composite sample violates the normality assumption in those conventional models.

In order to provide a better appreciation for the indirect-comparison detectors, I present in figure 9 and figure 10 the 3D perspectives of a selected number of output surfaces, they are: AsemiP, AVT, AFT, ANOVA, and KRX. These 3D surfaces are the same surfaces shown in figure 7 and figure 8.

Notice in figure 9 that the clipping thresholds applied to the AsemiP, AVT, AFT, and ANOVA surfaces are *8000, 3000, 80, and 300*, respectively. These thresholds were

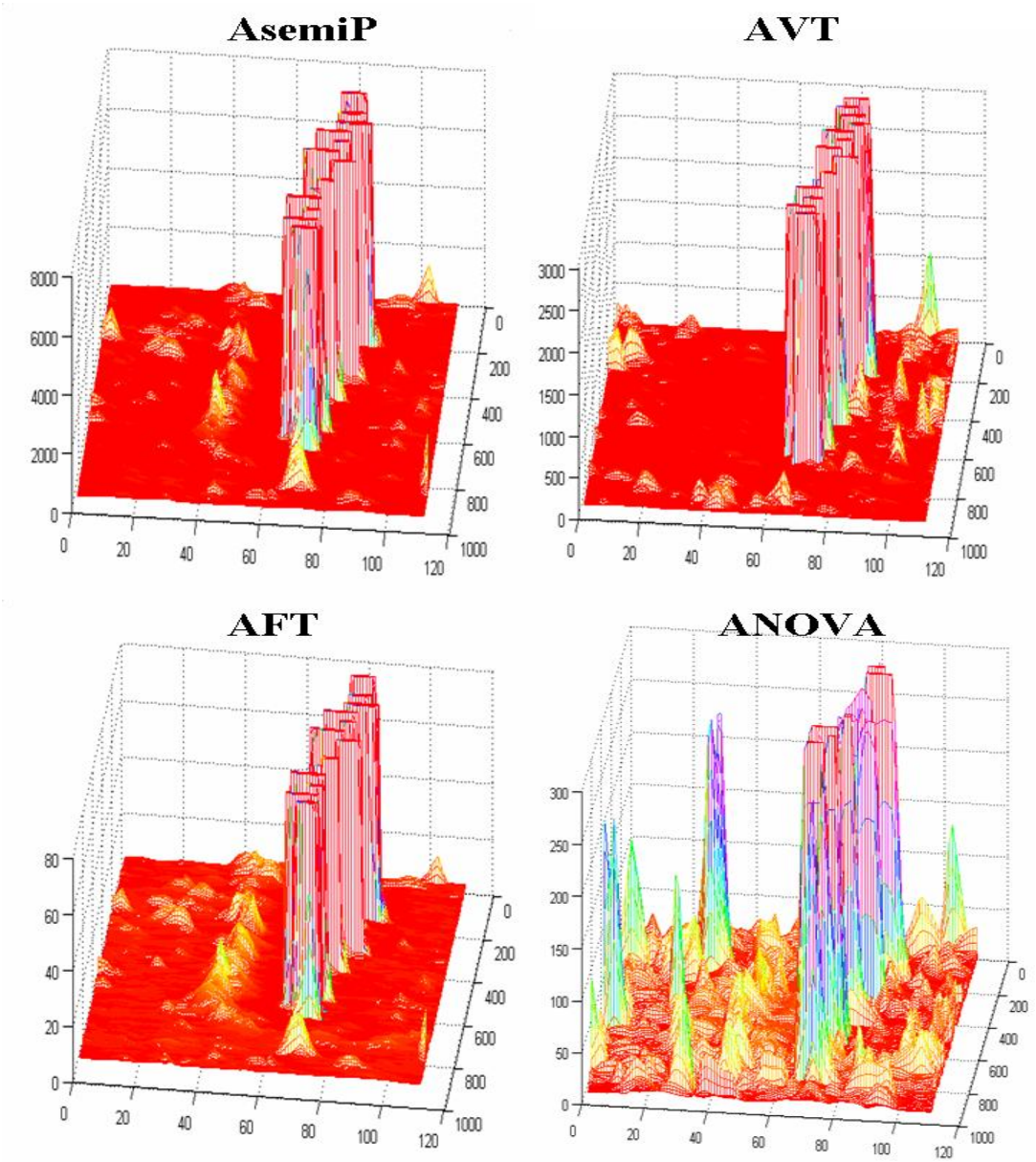

Figure 9. Decision surfaces (3D) produced by the detectors AsemiP, AVT, AFT, ANOVA testing on FR-I. Surface clipping applied.

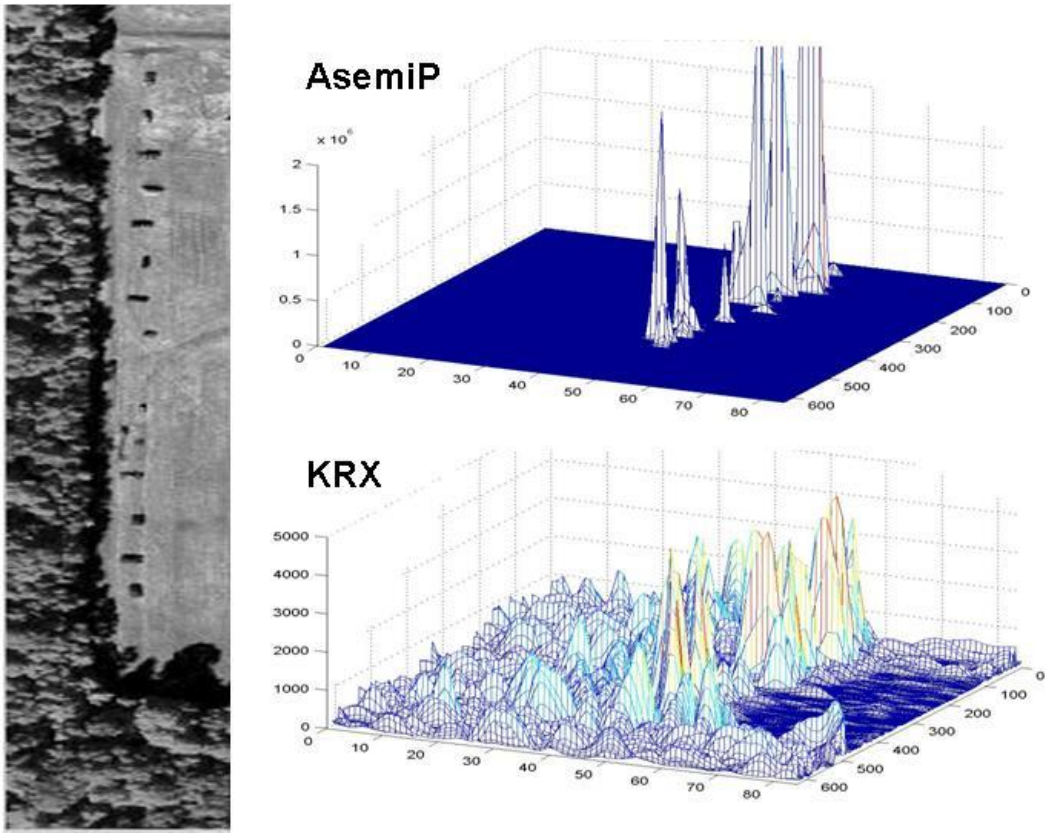

Figure 10. Decision surfaces (3D) produced by the detectors AsemiP and KRX testing FR-I data. Virtually no surface clipping applied.

applied and the results stretched so that the reader could better appreciate the intensity of targets' responses in contrast to the clutter background's in the entire image.

The high intensity peaks in all four surfaces correspond to the presence of the stationary land vehicles in the scene, although the ANOVA detector also accentuated some meaningless signs of local anomalies due to region discontinuities. The surfaces shown in figures 7–9 were clipped because some of their dominant peaks do continue to relatively higher numbers, completely obscuring the presence of less dominant target responses. The criterion for deciding on clipping values was based on the peak value of the weakest target response in each surface.

In figure 10, I present the same output surface of the AsemiP detector shown in figure 9, but in this case clipped at a significantly higher threshold (i.e., 2×10^4), and the KRX surface, which required no clipping. They are put side by side in figure 10 for visual comparison. As mentioned earlier, in both *3-dim* surfaces, the fourteen land vehicles by the treeline responded as the most dominant peaks in those surfaces, indicating that the spectral characteristics of the vehicles' paint and vehicles' shadows are significantly different from their immediate surroundings. The difference between the output surfaces produced by detectors AsemiP and KRX, shown in figure 10, is quite dramatic. That difference emphasizes the inherent ability of indirect-comparison

based detectors to suppress the clutter background and to accentuate what would be characterized by image analysts as meaningful detections in that scene, when compared to conventional detectors.

Similar results could also be observed testing these two detectors on the DR-II data, see figure 11. The five most dominant peaks shown in the 3D AsemiP surface correspond to the presence of the five stationary vehicles on a desert road (see fig. 11, left). Figure 12 shows 3D output surfaces produced by detectors SemiP, AFT, ANOVA testing the DR-II data.

To obtain quantitative results from performances of the different technique types, I use ROC curves. Figure 13 shows ROC curves produced by the output of the ten algorithms on FR-I. Detection performance was measured using the ground truth information for the HYDICE imagery.

I used the coordinates of all the rectangular target regions and their shadows to represent the ground truth target set; call it *TargetTruth*. If I denote the region outside the *TargetTruth* as *ClutterTruth,* then the intersection between *TargetTruth* and *ClutterTruth* is *zero* and the entire scene is the union of *TargetTruth* and *ClutterTruth*. In this text, for a given decision threshold, the proportion of target detection (PD) is measured as the proportion between the number of detected pixels belonging to *TargetTruth* over all pixels belonging to *TargetTruth*.

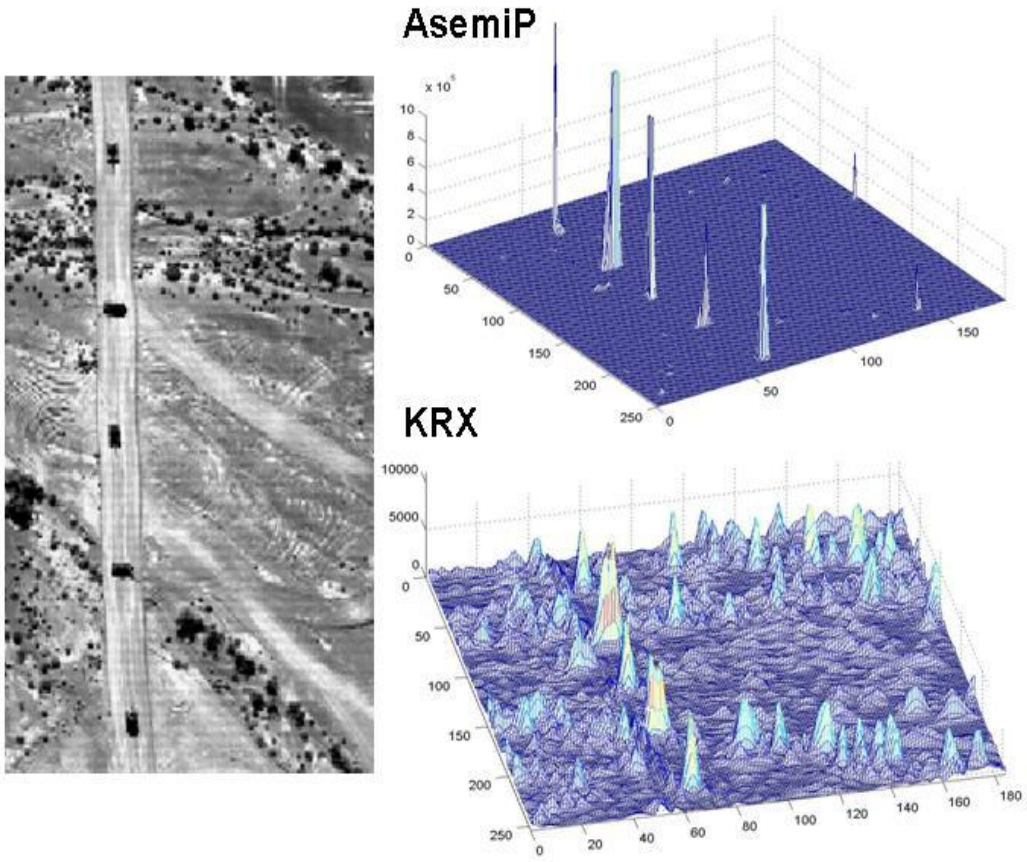

Figure 11. Decision surfaces (3D) produced by the detectors AsemiP and KRX testing on DR-II. No surface clipping applied.

On the other hand, the proportion of false alarms (PFA) is measured as the proportion between the detected pixels belonging to *ClutterTruth* over all pixels belonging to *ClutterTruth.*

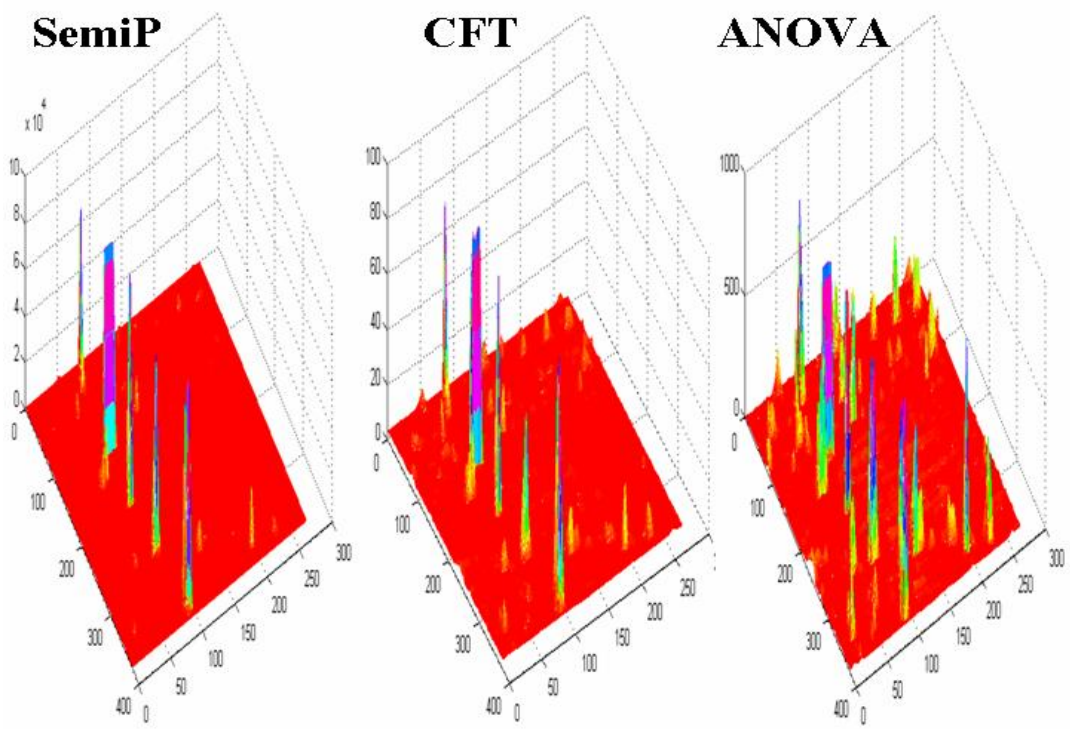

Figure 12. Decision surfaces (3D) produced by the detectors AsemiP, AFT ANOVA testing on DR-II.

Using the ROC curve as a metric, figure 13 further suggests the significant level of improvement produced by the indirect-comparison based techniques over alternative approaches. The differences in performance are better appreciated in figure 13 (right), where PFA is further restricted to a maximum value of *0.01*—an extremely low PFA. Although the results shown in figure 13 (right) help the reader appreciate the contrast in performance among the ten detectors, they do not do full justice to the quality of the indirect-comparison detectors. For instance, the threshold that yielded a PD of *0.55* using the SemiP, AsemiP, AFT, and AVT detectors comfortably found the 14 targets in FR-I, but not necessarily all the pixels on those targets. In other words, these detectors were able to detect sizeable portions of all 14 stationary land vehicles, yielding in the test a *zero* PFA.

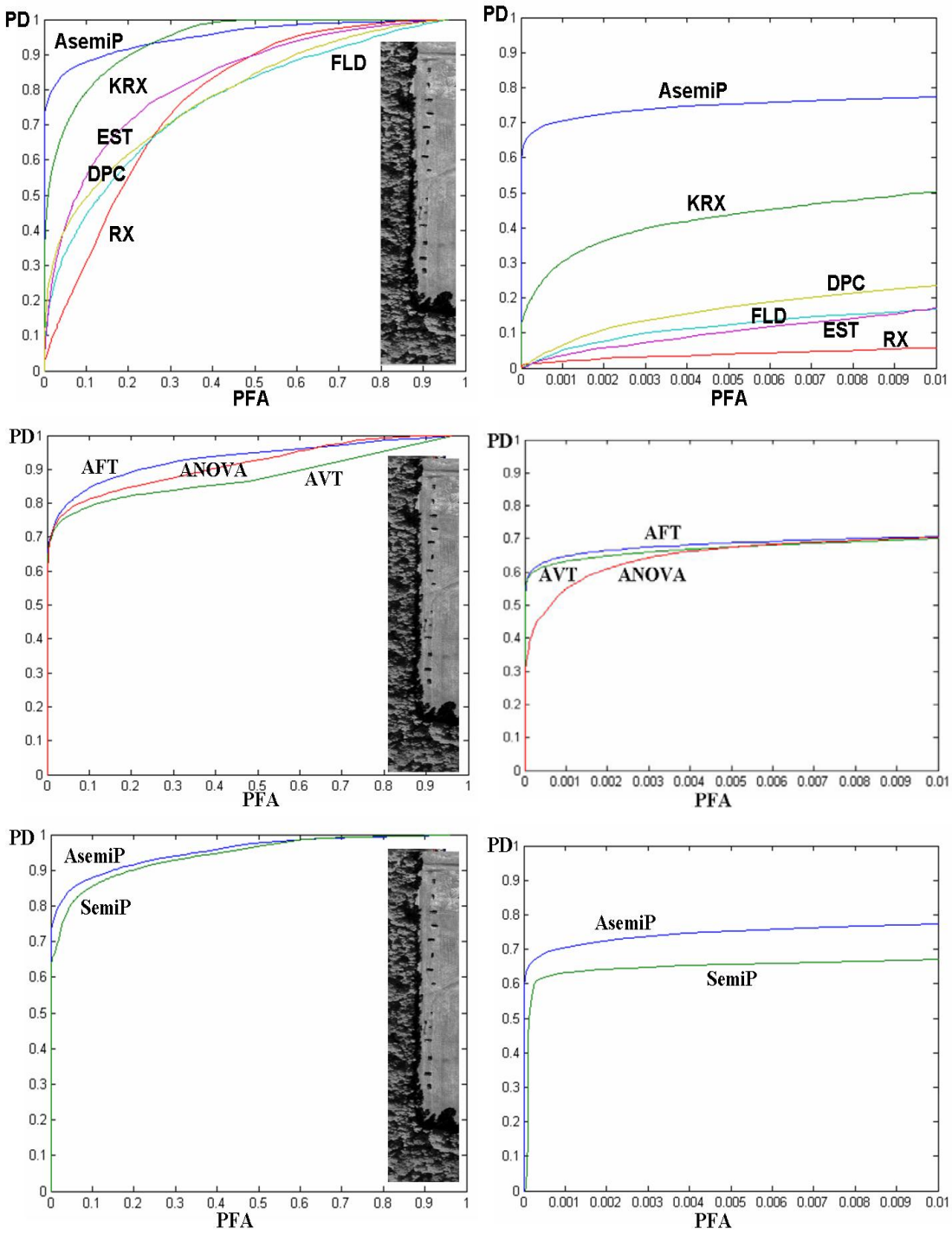

Figure 13. ROC curves using the HYDICE data scene FR-I (forest radiance). These figures suggest that indirect comparison based detectors are noticeably less sensitive to different decision thresholds compared to alternative conventional methods. An ideal ROC curve resembles a step function starting at point (PFA = 0.0, PD = 1.0).

Asymptotic Performance: In Appendix B, I present asymptotic performances of detectors SemiP and AsemiP under their null hypotheses. Their empirical distributions were computed from their output (FR-I) surfaces and qualitatively compared to empirical distributions generated from two epochs of 2,000 simulated realizations of a random variable following a chi square distribution with *1* dof. The SemiP and AsemiP detectors yielded a good fit between their empirical distributions and two empirical distributions computed from these simulated realizations.

Processing Time: I report the processing time in minutes (min) for a cube *600* x *140* (pixels) x *150* (bands) using a personal computer (CPU speed: *1.80* GHz; RAM memory: *1.0* Gbytes), MATLAB™ software (release *13*), and three detectors (RX, AsemiP, and SemiP). The recorded times were: *20.6* min (RX), *13.4* min (AsemiP), *42.9* min (SemiP). Computing the local variance-covariance matrix and its inverse dominated the RX processing time. Applying locally a HP filter in the spectral domain and applying SAM on the resulting vectors dominated the AsemiP processing time. Finally, applying locally a HP filter and a spatial SAM, and using an unconstrained minimization routine dominated the SemiP processing time.

The detection results presented in figure 13 using the FR-I data were consistent with results produced by these detectors using DR-II. The ROC curves corresponding to the performances of the anomaly detectors discussed in this section are shown in figure 14. Results presented in figure 13 and figure 14 suggest that performance disparities conventional (e.g., KRX), are significantly larger testing scenes dominated by major transitions of class regions. The overall scene background in FR-I has clearly more transitions of class regions (e.g., shadow and grass) than observed in DR-II.

The processing times of the detectors using the DR-II were proportional to the results shown using FR-I, proportional to the cube size of DR-II.

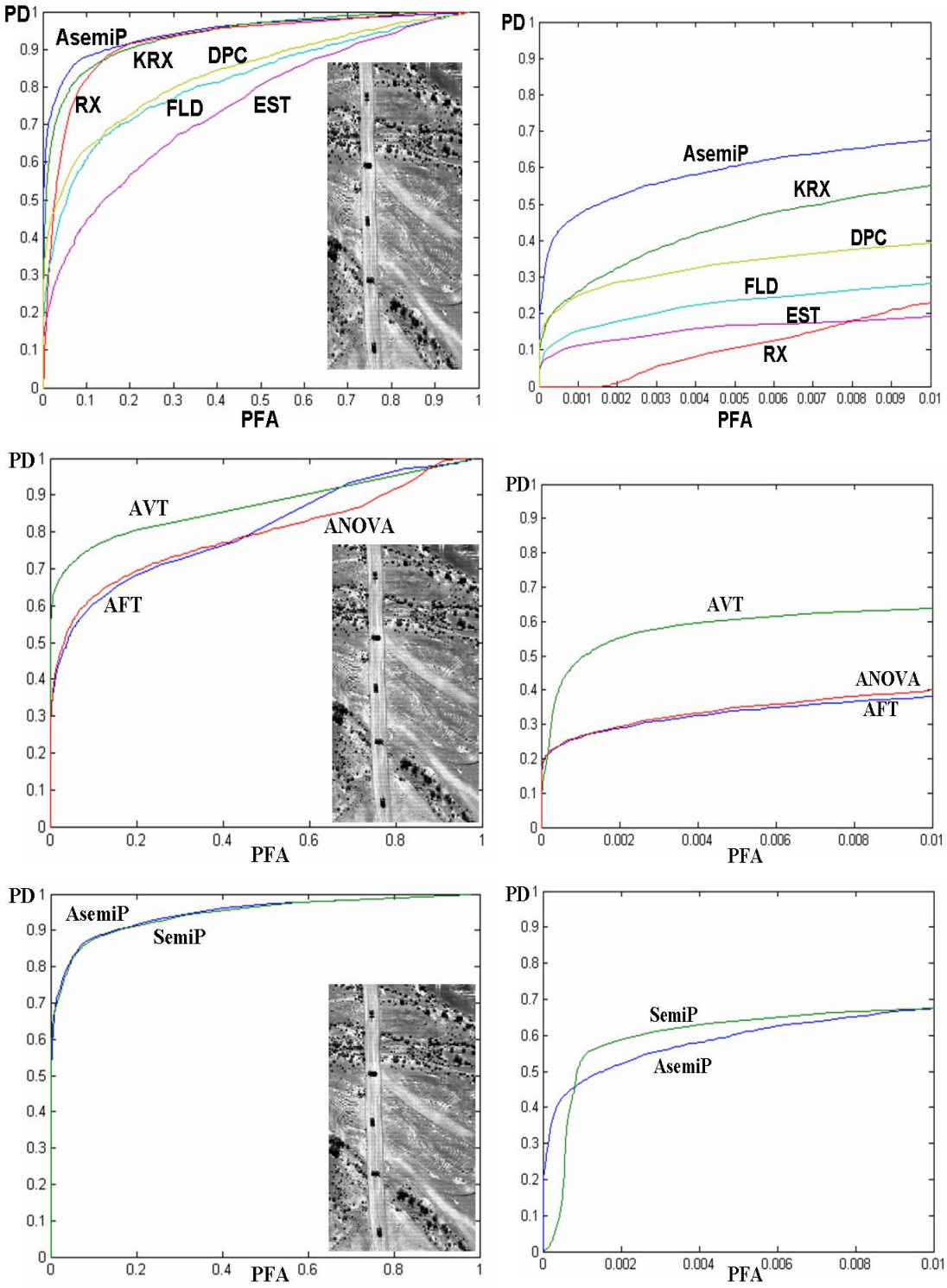

Figure 14. ROC curves using the HYDICE data scene DR-II (desert radiance).

3.7.3 SOC 700 Ground-Level View Hyperspectral Imagery

The ground-view imagery used for this work was collected with a novel visible to near-IR spectral imager (SOC-700) from Surface Optics Corporation, San Diego, CA. The system is a relatively small, portable hyperspectral imager, which collects a hyperspectral cube consisting of 640 x 640 pixels x 120 spectral bands and has a spectral range covering 0.38 to 0.97 μm. The sensor is commercially available off the shelf [38].

The data were collected during June, 2004 in Fort Hunter-Liggett, CA, to support a research effort by the U.S. Army. Three scenes from that data collection were used for this study. The first row in figure 15 shows the photos of those scenes, which were taken using a standard digital photo camera, and the second row depicts those scenes as the average of 120 bands, which were collected using the SOC-700 HS camera. Although not important to the impact of this work, notice that the photos and the HS imagery were not taken precisely at the same time, which explains some of the differences between the two types of images.

From actual ground truth, it is known that Scene 1 contains three motor vehicles and a standing person in the center of that scene (i.e., two pick-up trucks to the left in proximity to each other, a man slightly forward from the vehicles in the center, and a reflections from certain parts of the vehicles captured by the sensor in Scene 1 and 2 are not as dominant in Scene 3 because the vehicle there is in the shadow; hence, the terrain in Scene 3 appears to be a strong reflector.

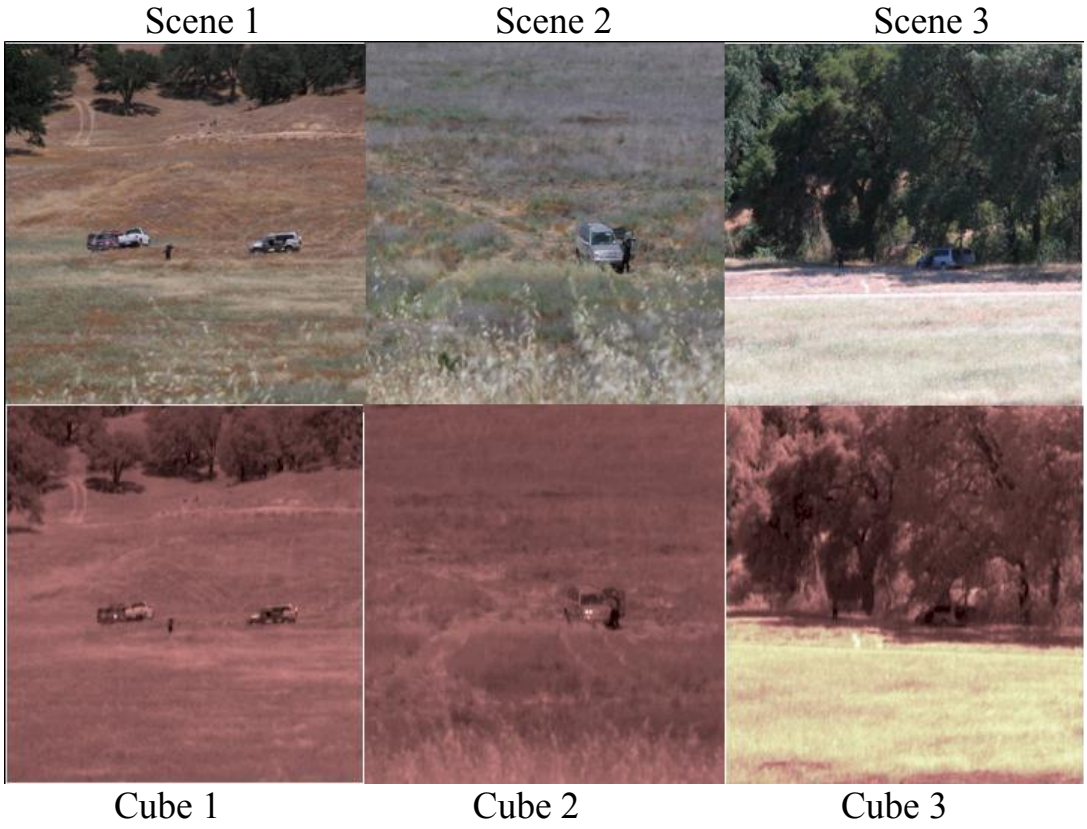

Figure 15. Scene photos and their corresponding SOC-700 hyperspectral cubes (band averages).

In essence, I would like to capture in this study the overall behavior of an anomaly detector in two ways: (i) seeking for global anomalies in a natural clutter background, given that only a few spectral samples from the most abundant object classes in the background (in this case, trees and terrain) are drawn from the same HS data and presented as references to the detectors and (ii) seeking for global anomalies, given that the reference samples are not drawn from the same HS data. I would be able to determine in (i) the effectiveness of these detectors within the same data of a particular area in the valley and in (ii) their effectiveness and robustness using data from different areas in the valley. Results are shown in figures 16–19.

Cubes RX AsemiP

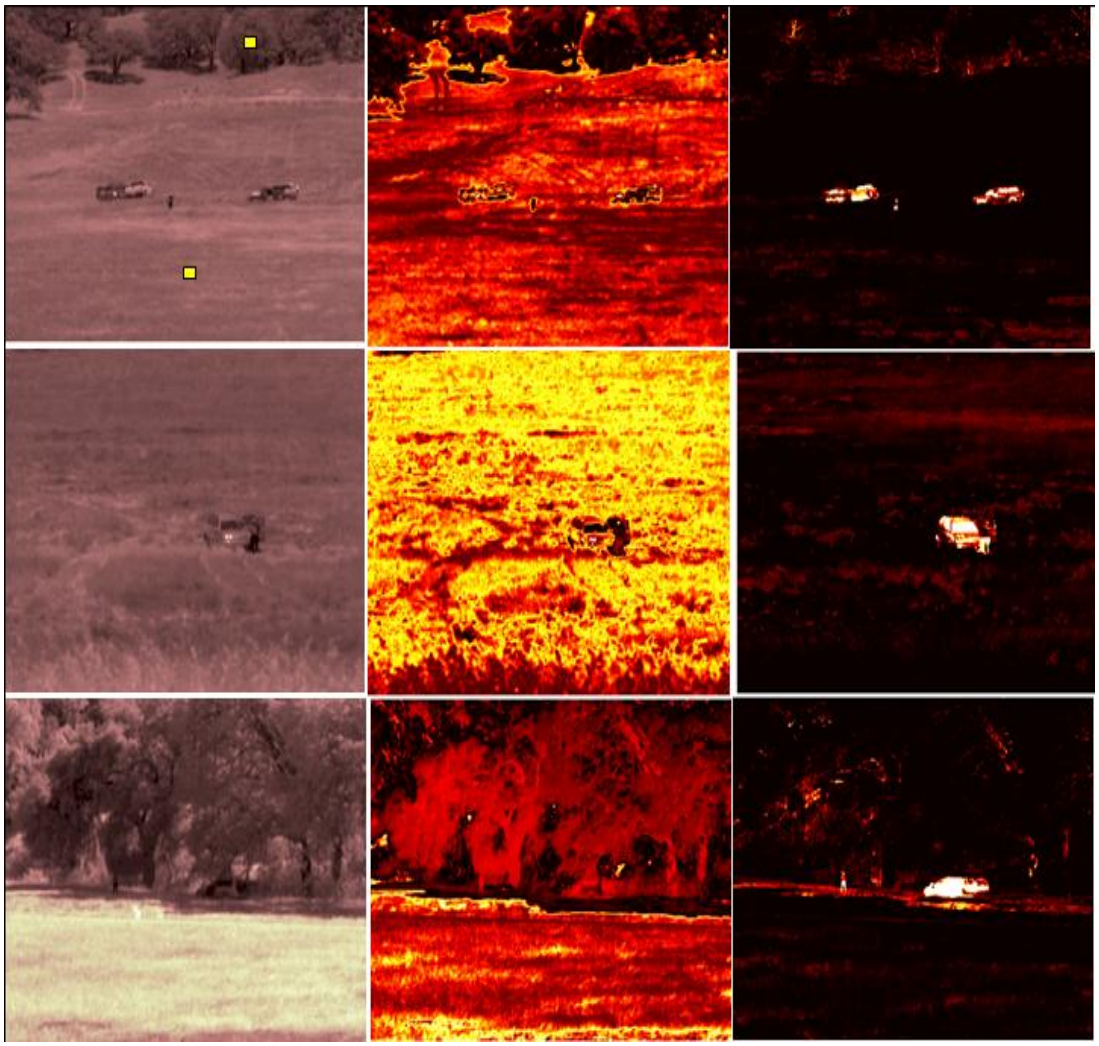

Figure 16. Scene anomaly detection using two reference sets of spectral samples (their locations are shown as yellow boxes in the top scene) from California tree leaves and valley terrain. The unconventional AsemiP anomaly detector was developed based on a principle of indirect comparison, and the conventional RX anomaly detector is the standard technique for anomaly detection. The RX and AseemiP output surfaces are displayed using the same pseudo color map, where *white* depicts the strongest sign of anomalies, *yellow* strong, *red* intermediate, and *black* lowest sign of anomalies.

I applied the RX and the AsemiP detectors to those scenes and present their output surfaces in figure 16, columns 2 and 3, respectively. (Using the initial condition $[\hat{\alpha}, \hat{\beta}] = [0, 0]$, the SemiP detector could not converge to a solution at every location in those images. Thus, I excluded its incomplete performances from this subsection.)

The sampling mechanism and data preprocessing used for the AsemiP detector were described in detail in Subsection 3.1, where, in this implementation, the test window

consisted of a 3 x 3 cell (nine 120-band spectral samples) and the two reference sample sets (one representing responses from tree leaves and another from a patch of terrain) consisted each of 100 spectral samples, for a total of 200 reference samples—a mere 0.05% {[200 / (640 x 640)] x 100} of the image area. The two yellow boxes shown in HS Cube 1 (row 1 and column 1 in figure 16) represent the general locations—chosen arbitrarily—where the two reference sets were drawn from. Using our proposed data preprocessing, both a reference angle sequence $y_0^{(p)}$ and a test angle sequence $y_1^{(p)}$ were obtained for each test location, as in (15), where, in this implementation, p = 1 denotes tree leaves and p = 2 denotes terrain. Note that the size of a reference set determines the size of its equivalent random sequences of angles, thus, $y_0^{(p)}$ and $y_1^{(p)}$ in this implementation have the same size, *100*, as only the average of 9 test samples is used to generate the test sequence of angles with each reference sample set.

The AsemiP detector is expected to systematically compare across the imagery the preprocessed test samples with the fixed preprocessed samples pertaining to both reference sets. If a local set of preprocessed test samples is significantly different from the fixed reference sets, the AsemiP detector should produce an accentuated value at that location indicating this fact; otherwise, it should produce a suppressed value. This expectation can be achieved by fusing results using (47) for $p = \{1, 2\}$, representing the two classes.

I adapted the RX detector to the ground-view problem using the recommended data preprocessing discussed in [10], i.e., a spatial high pass filter was applied to the untransformed hyperspectral samples belonging to the same reference sets used for the AsemiP detector, and also to the samples from the test samples across the imagery. This procedure removes the spatially nonstationary mean, which is not useful for the RX detector, and promotes spatial independence, allowing this detector to exploit an expected correlation in the spectral domain among samples belonging to the same class. Under the assumptions given in the RX model, this detector is expected to produce an accentuated value when the simplified *Mahalanobis* distance between a high-pass filtered reference set and a test set is significantly high; otherwise, it is expected to produce a suppressed value. Since I have, in this implementation, two fixed reference sets, the minimum between the two distances was also used as a means to produce a final result per location in the imagery. Recall that by using this decision logic, an anomalous test sample to both reference sets would still produce a high value, since both results would likely yield high values.

In figure 16, I present the output surfaces produced by both detectors and invite the reader to make a visual comparison between the corresponding surfaces. The output surfaces of the RX and AsemiP detectors are shown in columns 2 and 3, respectively, for the corresponding HS cubes in column 1. I used a suitable *colormap* to emphasize anomalies with respect to the reference samples by their false-color (intensity) levels, i.e., white is equivalent to the strongest anomalies, yellow to strong anomalies, red to

intermediate anomalies, brown to weak anomalies, black to weakest anomalies. The false colors change gradually and are relative only to those results within the same surface, for instance, a yellow pixel in one surface does not mean necessarily that its value is equivalent to another yellow pixel in another surface.

The local results shown in the first RX surface (row 1, column 2) are consistent with the study cases discussed in Section 1, see figure 1. A detector based on conventional methods performs well suppressing objects in the scene having low variability and belonging to the same class of a reference set (*Case 3*)—the trees were suppressed. Likewise, it performs well accentuating objects that are significantly different from the reference set (*Case 1*)—for instance, some parts of the vehicle at the right hand side (row 1, column 2) were highly accentuated. (One can actually observe white pixels within the boundaries of those vehicles by zooming close enough on both RX surfaces (rows 1 and 2, column 2), which indicates that those portions are significantly different from the reference sets. Unfortunately, as it was observed in the top-view problem, local areas characterized by class mixtures (transition of regions) may be also accentuated by these detectors, obscuring therefore the presence of meaningful objects in that scene. In fact, for the HS cubes presented in figure 16, the RX detector seems to perform more as an edge detector than as an anomalous object detector.

The AsemiP detector, on the other hand, was able to suppress virtually all the background of Cube 1, and to accentuate large portions of the vehicles and of the standing man. In a qualitative sense, test samples consisting of, say, a mixture of shadow and terrain were likely suppressed due to the indirect comparison between the mixture itself and the union between that mixture and a component of that mixture, in this case, terrain.

Next, compare the impact of shadowed objects to a reference consisting of the non-shadowed version of the same object.

The reason our indirect comparison based detectors work so well suppressing shadowed patches in the ground may be explained by the following: Regions characterized by tree shadows, for instance, may be interpreted as *partially obscured terrain* because tree leaves do partially obscure the incident solar light; however, since significant spectral radiances are still reflected from the partially shadowed terrain, such a region will be suppressed when compared to the union of itself and the reference set of open terrain.

Now, let us shift our attention to the results shown for Cube 2 in figure 16 (row 2, column 2 and 3). The RX surface shown in rows 2, column 2, suggests that the RX detector may be susceptible to subtle spectral differences of the same terrain when observed by the same HS sensor in a different area. Recall that Scenes 2 and 3 were tested using the same reference sets drawn from Scene 1. The surface shown in row 2, column 3, suggests that the AsemiP detector is significantly more robust to spectral differences of the same terrain. The concern of such robustness was addressed as one of our examination goals cited in (ii).

Shifting our attention now to the results corresponding to Cube 3 in figure 16 (row 3, column 2 and 3), the interpretation of a shadowed object as a partially obscured object is especially relevant to the interpretation of output results for Scene 3. The output surface shown in figure 16, row 3, column 2, emphasizes the fact that the RX anomaly detector performs as expected: it detects local anomalies in the scene. However, as I have been discussing throughout the report, these local anomalies are not guaranteed to be meaningful to an image analyst in the context of our problem. For instance, in reference to the RX output surface for Cube 3, notice that some of the tracks made by the shadowed vehicle, and the transition between the shadowed and the non-shadowed terrain were the most anomalous regions in the scene, as *seen* by the RX detector. Fortunately, with the indirect comparison approach that is inherent in the AsemiP detector, these same regions were virtually suppressed, while the more meaningful anomalous structures (vehicle and human pants) were accentuated, (see the corresponding AsemiP surface in figure 16 (row 3, column 3)).

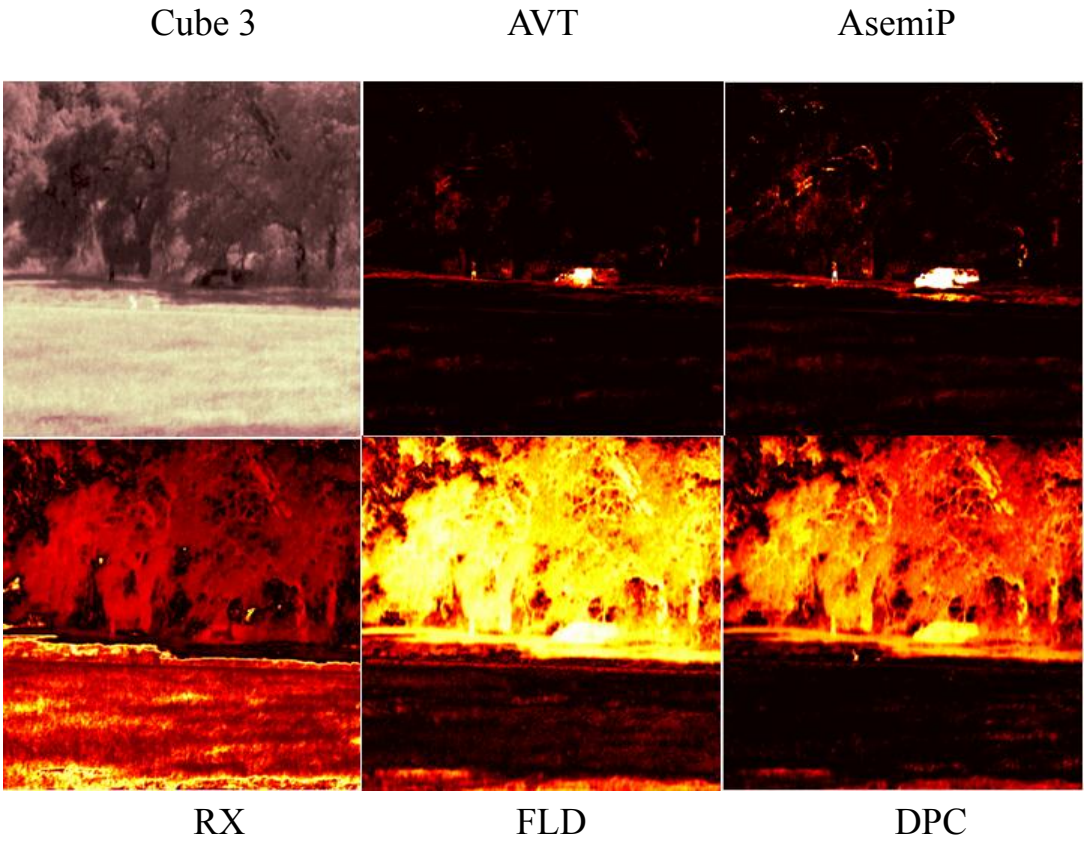

Figure 17. Scene anomaly detection using two reference sets of spectral samples from California tree leaves and valley terrain.

For additional performance results, refer to figure 17 and figure 18, where I present consistency in performance between the detectors AVT and AsemiP using HS Cube 3 and an additional cube (Cube 4), and also the results produced by the detectors FLD and DPC. The RX and AsemiP surfaces shown in figure 17 (row 2, column 1) and (row 1, column 3) are exactly the same ones corresponding to those detectors in figure 16. Notice that the FLD output surface shown in figure 17 (row 2, column 2) emphases the spectral differences between the shadowed tree region and the two reference sets, which incidentally are the same reference sets drawn from HS Cube 1. Notice that the FLD detector accentuates significantly a large portion of the shadowed land vehicle and of the person, among other shadowed objects in that region, e.g., shadowed tree trunks and leaves. The DPC detector, on the other hand, focused on a portion of the vehicle's tire tracks as being the most anomalous object class in the entire scene when compared to the reference sets. Taking a closer look at the DPC surface in figure 17 (row 2, column 3) did reveal that about three pixels within the boundaries of the tire tracks are actually white (highest intensity). Yellow pixels shown within the boundaries of the vehicle, within the boundaries of the person, and within the boundaries of other object classes in the shadowed region indicate that those shadowed object classes are the next lower level of anomalies in respect to the reference spectral sets, as *seen* by the DPC detector. (Incidentally, the RX, FLD, and

DPC surfaces in figure 17 are shown without clipping their values, which is also true for the RX surfaces shown in figure 16. The AVT and AsemiP surfaces, however, required some clipping for the reasons discussed earlier using top view imagery.)

The results in figure 16 and figure 17 further support our claim that conventional approaches are flawed in anomaly detection, as described in this report. In other words, they either account for a clear presence of anomalies when compared to a homogeneous background or no presence, but they do not account for transition of distinct regions, which unfortunately are quite abundant in digital imagery of natural clutter background. The indirect comparison detectors, on the other hand, account inherently for all three study cases, as described in Section 1.

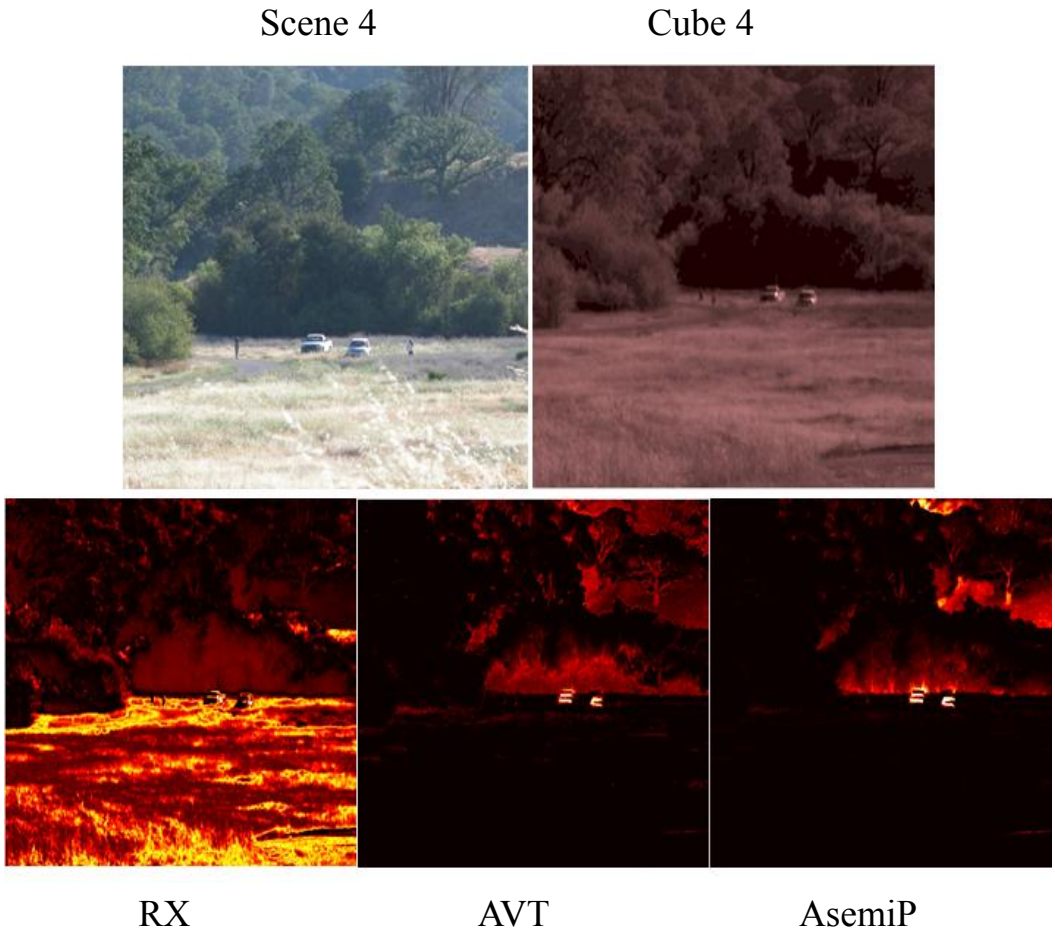

Figure 18. Scene anomaly detection using two reference sets of spectral samples from California tree leaves and valley terrain.

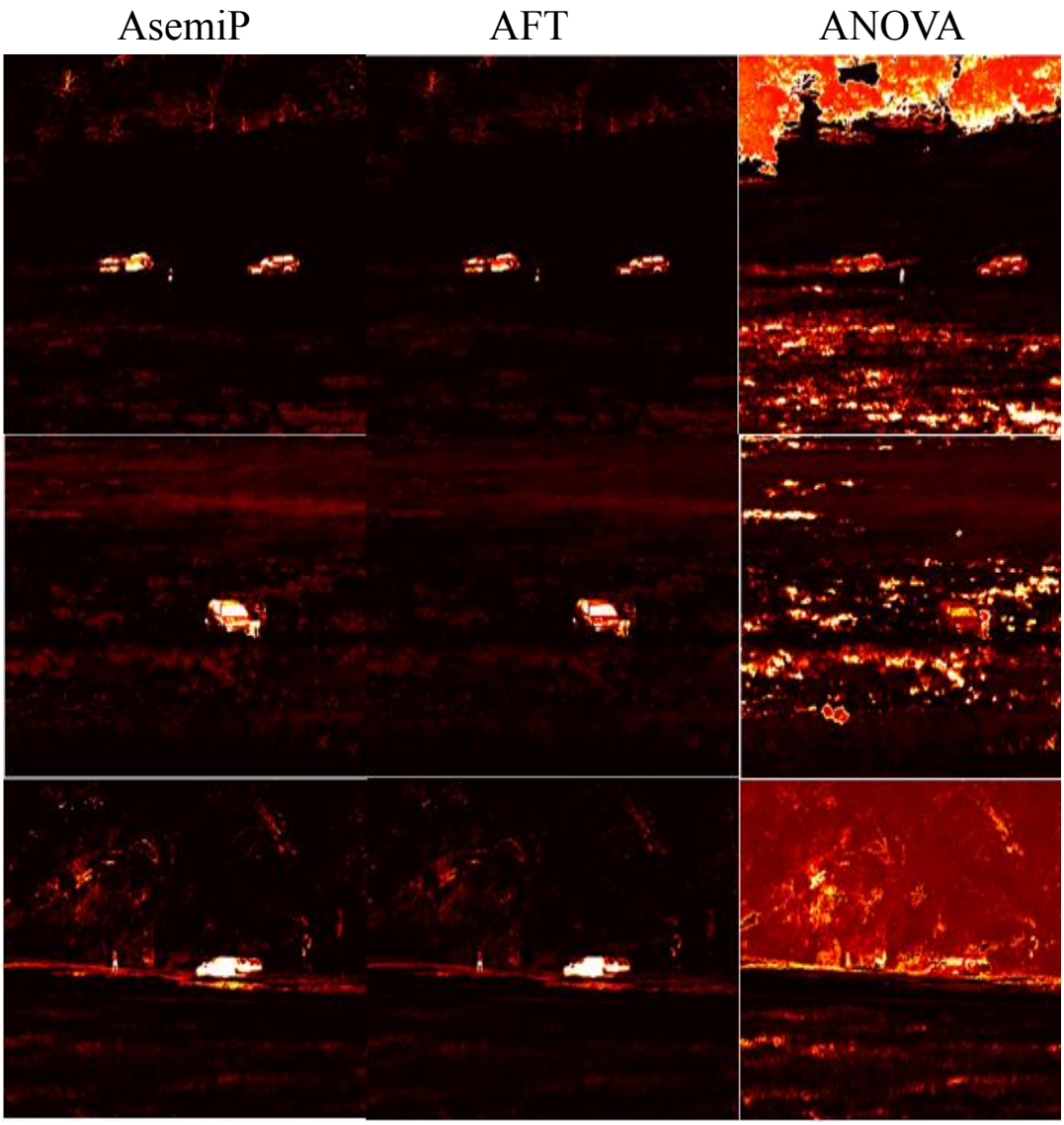

Figure 19. Performance results of detectors AsemiP, AFT and ANOVA testing ground level imagery (Cubes 1, 2, and 3, shown in fig. 3.13).

digital imagery of natural clutter background. The indirect comparison detectors, on the other hand, account inherently for all three study cases, as described in Section 1. In figure 18, a new scene in presented, where the photo was not taken exactly at the same time the SOC-700 camera collected its data. Notice that the performances of detectors AsemiP, AVT, and RX in figure 18 are consistent with their corresponding results shown in figures 16 and 17. I included HS Cube 4 primarily due to the relatively small scale of two of the targets in that scene, i.e., two persons consisting of very few pixels on them. In fact, as it is evident from the output surfaces shown in figure 18, these human targets were not even detected as anomalies by either one of the indirect-comparison detectors, AVT or AsemiP. Part of the problem is that the farther away a target is from the sensor, the more attenuated its total radiance will be

due to the atmospheric transmission properties. Moreover, the target radiance will be corrupted by the radiance of adjacent object classes. In addition, the most discriminatory feature of both persons (the material of their pants) was significantly immersed in high grass. These facts made those two targets not so discriminatory from the two sets of spectral samples used as references: tree leaves and terrain.

Figure 19 shows additional results for a direct comparison between the F-distribution detectors, AFT and ANOVA. Those results reinforce the fact the normality assumptions in the ANOVA model can degrade performance. The results between the AFT detector, which does not assume normality, and the ANOVA detector, which does, were reasonably comparable testing the top view imagery, but this comparability dissipated testing ground level view imagery. A conclusion that I can draw from the output surfaces presented in this section is that results of all indirect comparison based approaches developed in this research were all consistent, whether the problem used top view or ground level view imagery. To complete this subsection, I present some results related to the sensitivity of our approach to varying sample size, (see fig. 20). Using the same sampling mechanism described for Scene 1 (see fig. 5 and fig. 16), I varied the sample size per class and used the AsemiP detector as a benchmark to test Cube 1. Denoting N the sample size per class, I used $N = 30, 60, 100$ and 500 spectral samples to represent the two classes: tree leaves and general terrain. ($N = 100$ has been used by default in our discussion in this subsection.) The results shown in figure 20 suggest that the AsemiP detector is not significantly sensitive to sample sizes greater than 30, a desirable property. Even at $N = 30$, there are strong detection markers on the vehicles and on the man's pants, which would be sufficient to extract those objects as being anomalous to the two references, tree and terrain. The extraction of objects from their background using anomaly detection markers via post image processing will be discussed next.

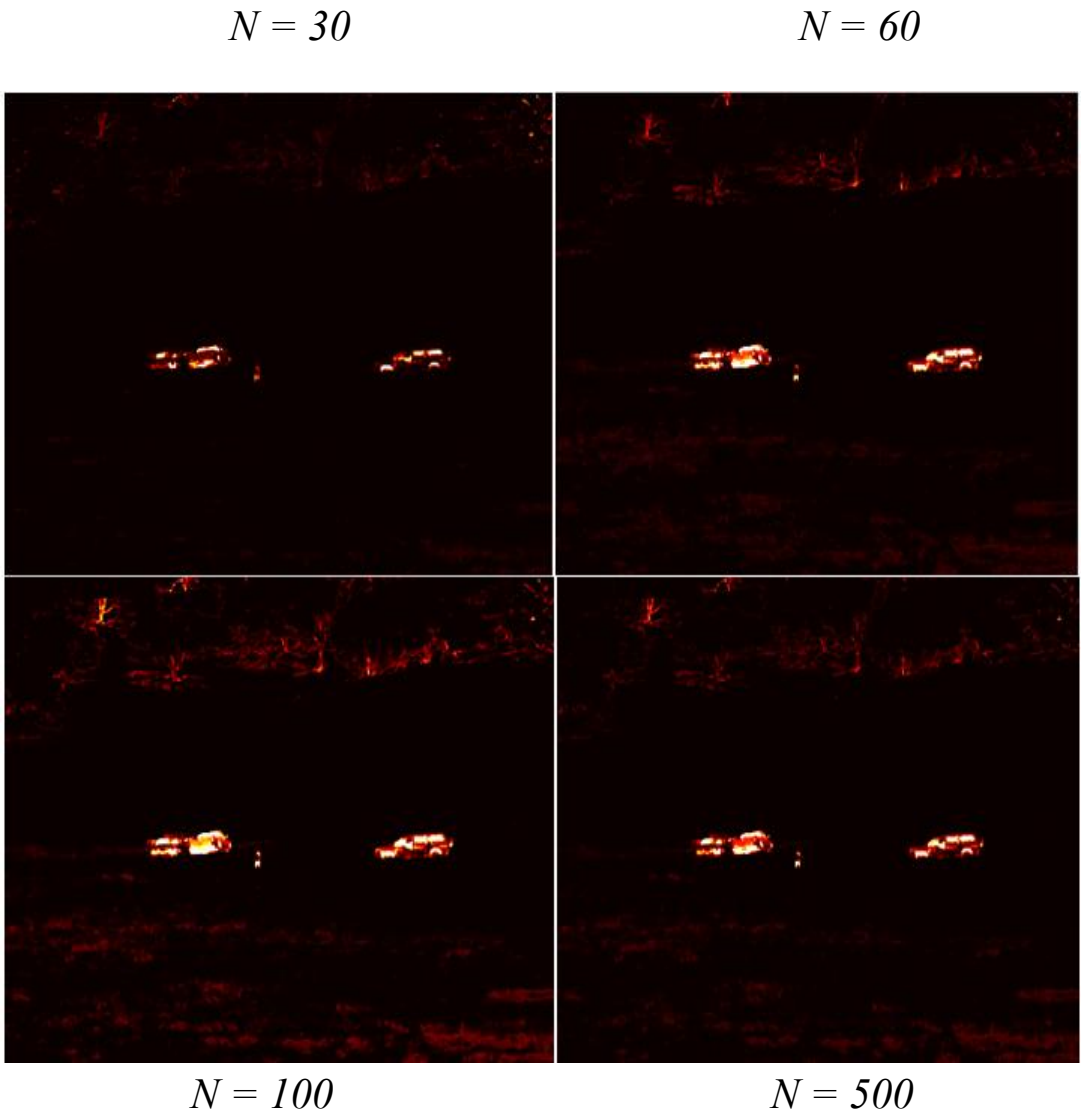

Figure 20. Sensitivity of the AsemiP detector to varying sample sizes. Denoting N the sample size per object class (two classes: tree leaves and terrain) the output surfaces are presented for N = 30, 60, 100, and 500.

3.7.4 Extension to Unsupervised-Learning Based Classification

It is well known that an effective anomaly detection technique may be adapted to function as an unsupervised learning classifier. In this subsection, I adapt our indirect comparison, anomaly detection approach to function as a *self classifier* and present a proof-of-principle experiment. Figure 21 depicts the concept. The notion of self classification, in the context of our discussion, simply means that a given algorithm suite consisting of two stages can be used to detect meaningful objects (stage 1) as a collection of point anomalies in respect to some reference set (available *a priori*). Upon applying a clustering algorithm to separate these detections as mutually exclusive clusters, the anomaly detection engine would be reused to function as a

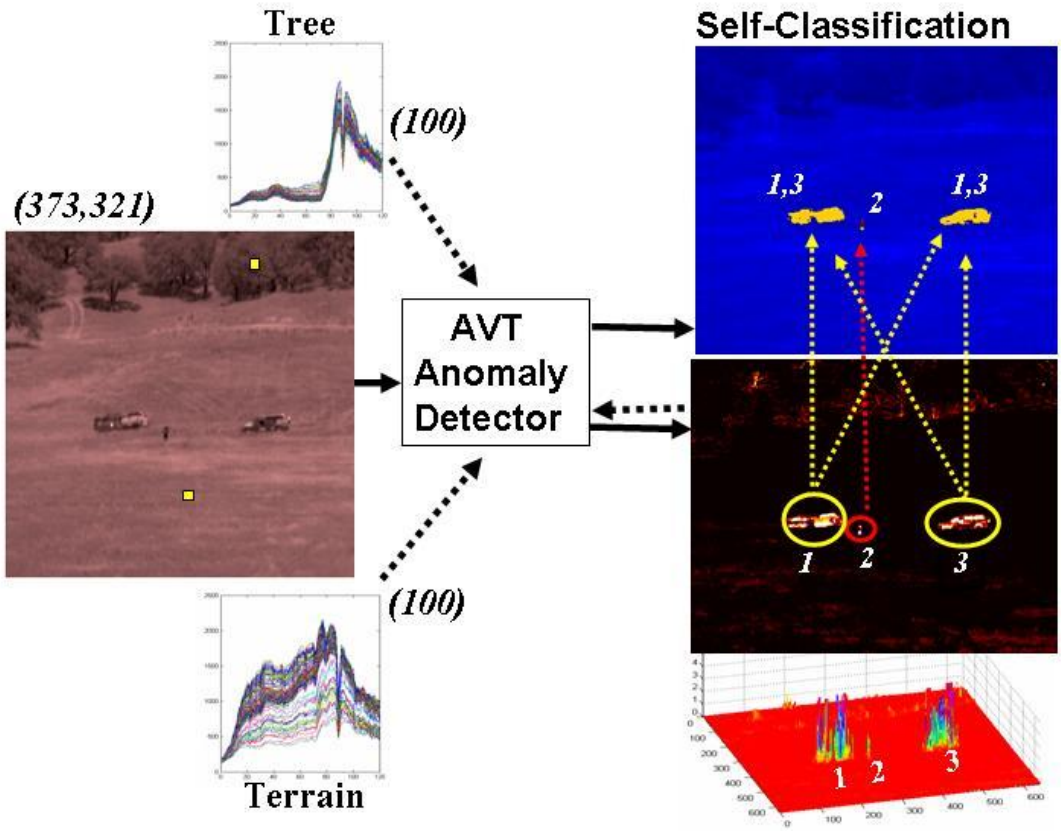

Figure 21. Proof of principle experiment illustrating a concept of self-classification using the AVT anomaly detector twice in the loop.

classifier by reintroducing, to the anomaly detector, samples from the clustered detections, as references, and—relying on the anomaly detector's ability to discriminate—to determine the classification among the detected objects. For instance, suppose that samples from two objects are detected and then clustered into two groups, Class 1 and Class 2. It would be of further interest to know rather Class 1 and Class 2 are the same or different classes. If these classes are the same, one could use the same color code to indicate this fact. Otherwise, the two classes would be displayed with different colors.

Using this procedure, it would be appealing to have, for instance, land vehicles that are detected in the same HS cube being able to retain the same color code, a color code that would be different from the one obtained by a different object class (e.g., human beings) also present in the scene. Note that self-classification, in this context, would not provide information on the actual classes of the objects, although it separates objects by class membership using the discriminatory power of the anomaly detector—given that these objects are separable.

The output results shown in figure 21 depict this notion of self-classification using AVT as the chosen anomaly detector. The red surface in figure 21 (lower right hand corner) represents the AVT detector's output surface for Cube 1. That surface clearly

shows strong anomaly peaks due to the presence of the vehicles and to the standing person in the open field. Based on the proximity of the two vehicles in the left of the person (reader's perspective), they seem to form a single anomalous object, which is labeled at this preliminary stage as class 1. The person was labeled at this stage as belonging to another class, 2. And the vehicle at the right of the person was labeled as belonging to a third class (labeled as 3). The surface immediately above the red surface is exactly the same surface, but displayed as a 2D surface using a different colormap. The circles around the anomalous structures were put artificially in this 2D surface to emphasize the fact that automatic post processing can be applied to exploit the detection markers and to spatially bound each unknown individual object. I used standard morphological filters (i.e., logical combinations of *dilation* and *erosion* to function as an opening operation to reduce noise, and as a closing operation to fill holes within the same object) to produce a silhouette for each class using the most dominant peaks as detection markers.

Spectral samples from within each silhouette (class 1, 2, or 3) were used as a reference set—one at a time—through the same anomaly detector to decide whether the other two classes belonged to the reference class. There could only have five outcomes: (i) the three classes are about the same, (ii) the three classes are significantly different, (iii) classes 1 and 2 are about the same but different from class 3, (iv) classes 1 and 3 are about the same but different from class 2, or (v) classes 2 and 3 are about the same but different from 1.

The blue surface shown in figure 21 shows the final result from the self classification procedure just described, where the blue region represents the suppressed clutter background after stage 1. The AVT detector produced the outcome (iv), i.e., the vehicles fell into the same class (depicted by yellow) and an overwhelming portion of the standing person fell into a different class (depicted by red). In summary, using initially two sample sets as references drawn from the scene shown in figure 21, the AVT anomaly detector was able to find three spatially independent objects as scene anomalies and could conclude that two of them (the three vehicles) belonged to the same class, and the remainder one (the standing person) most likely belonged to a class of its own.

A similar proof of principle experiment was carried out using the AsemiP anomaly detector. It produced the output results shown in figure 22. Figure 22 also depicts the output result produced after stage 1 and the post processing procedure that spatially clustered the mutually exclusive objects, as *seen* by the detector, see surface at the upper left hand side in figure 22. One may interpret the joint functions of anomaly detection and the follow-on post image processing as the extraction of meaningful objects from the scene, or as a meaningful focus of attention, this interpretation is emphasized by the 3D surface shown at the upper right hand side in figure 22. The spectral samples shown at the lower right and lower left hand sides are samples of the corresponding objects, person and right side vehicle, and were drawn from the scene using the detection masks (produced by post image processing) shown

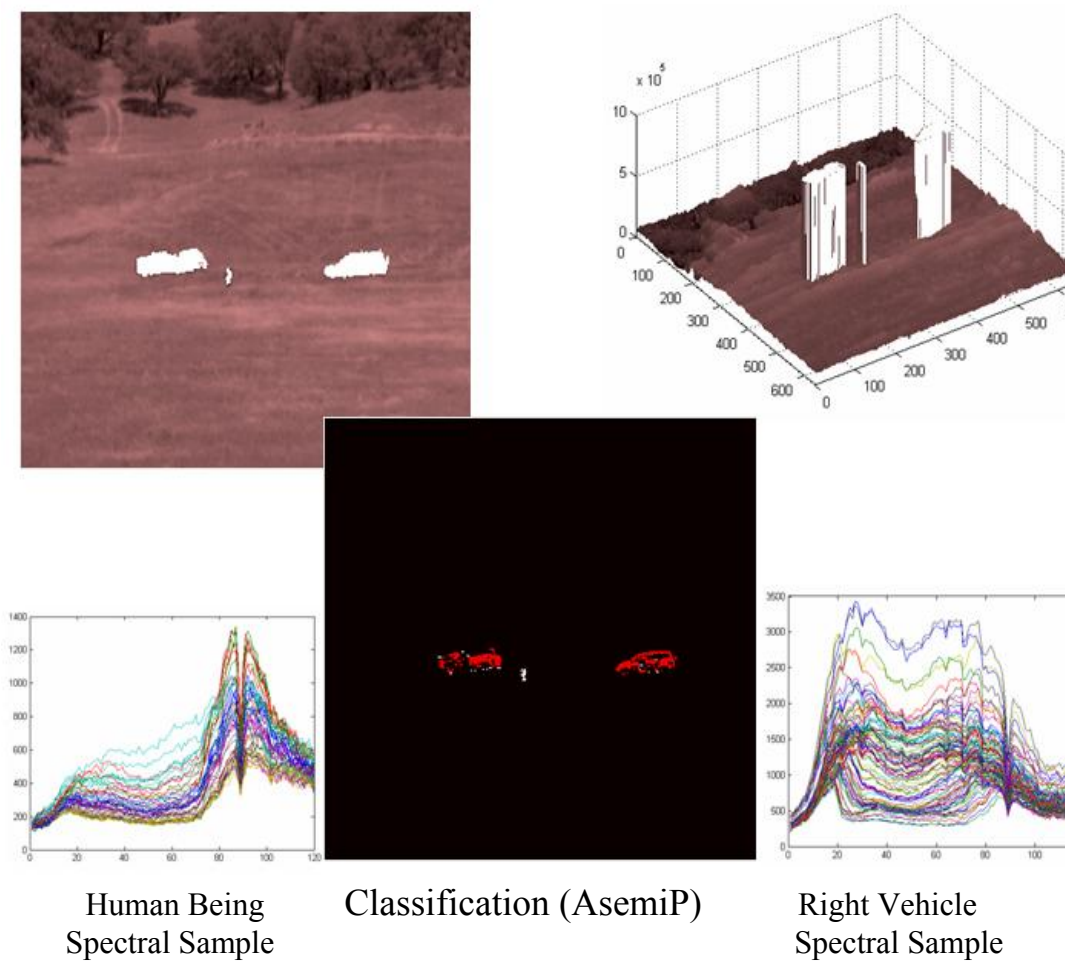

Figure 22. Proof of principle experiment illustrating a concept of self classification using the AsemiP anomaly detector twice in the loop.

in figure 22 as white solid shapes. The vehicles at the left side of the person have similar spectral responses. The final output surface produced by reintroducing the new reference sets of spectral samples from the vehicles at the left, from the vehicle at the right, and from the standing person to the AsemiP detector is shown at the lower center in figure 22. The final AsemiP result is consistent with the corresponding result produced by the AVT detector, as I expected.

4. Conclusions

The objective of this work was to develop statistical techniques with applications to a fundamental problem in machine vision: Empower a machine with the ability to focus its attention to meaningful objects in a scene, without human intervention. Meaningful objects are of course subjective, which may bring in prominence the individuality of an image analyst, or the objective of a particular surveillance task. In this work, meaningful objects are characterized by their material properties being significantly anomalous to an overwhelming presence of other types of materials forming a background, e.g., foliage clutter. Examples of meaningful objects are stationary equipments on natural terrain, a standing person in an open field, etc.

Although not evident from the problem statement, such a capability implies that the automatic procedure must be highly effective performing subtasks that are well known in the image processing community for being challenging problems by themselves. For instance, a challenging subtask is the ability to automatically suppress the entire clutter background in a digitized scene that may be dominated in abundance by local transitions of different types of material regions. Another challenging subtask is the ability to automatically accentuate the presence of certain types of objects, as a collection of localized anomalies, with respect to sets of predetermined material types.

To accomplish this work, I opted to use hyperspectral rather than broadband imagery and to focus our algorithmic development on adaptive anomaly detection rather than on a particular type of material detection. A key benefit for choosing hyperspectral data over broadband is that a particular type of material may be identified by testing a few pixels of the tested object, independently of the object's orientation, elevation angle, and distance from the sensor. A key benefit for choosing anomaly detection over a particular type of material detection is that often the exact material of interest is not known a priori, or the number of spectra in a material of interest library is simply too exhaustive to search for all possible materials. The prospect then of using HS imagery jointly with anomaly detection techniques holds the prospect of detecting both known and unknown targets of any shape, size (assumed to be greater than the sensor's pixel resolution), and material type as statistical outliers. This outcome has an important practical value, if, and only if, the final product yields a significantly low false alarm rate compared to the prior art.

Most conventional anomaly detectors use multivariate models to define the spectral variability of the data, and the majority of the data pixels are assumed to be spectrally homogeneous and are modeled using a multivariate probability density function with a single set of parameters. Until now, no significant work had been done to find non-normal statistical models, or unconventional alternatives, for the development of anomaly detection techniques using hyperspectral data. This work shows that conventional anomaly detectors can detect the presence of targets using hyperspectral data from the HYDICE and SOC-700 sensors, but also yield in the process a large number of false alarms. This type of performance has little practical value.

I aimed at improving overall performance by implementing a principle of indirect comparison, where samples are not compared to each other as individual entities, but as individual entities compared to the union of these entities. I implemented this principle in different forms and showed that they outperform significantly conventional techniques on two types of anomaly detection problems: one from the top view perspective and another from a ground level view perspective.

The more important findings and developments of this report are summarized next.

4.1 Summary

This subsection summarizes the more important findings and developments of this research in eight parts: Hyperspectral Data, Conventional Anomaly Detection, Principle of Indirect Comparison, Semiparametric Anomaly Detector, Approximation of the Semiparametric Anomaly Detector, F-distribution Anomaly Detectors, Asymmetric Variance Based Anomaly Detector, and Impact of Work.

- Hyperspectral Data: HS imagery played a major role in the quality of the results shown using the unconventional anomaly detectors developed in this work, especially in detection problems from the perspective of a ground level view, where the size, material type, object to sensor range, and pose of potential targets are unknown and virtually impossible to account for all their possible permutations. I further state that hyperspectral imagery may give some *hope* for object detection scenarios typically characterized by an image analyst as *hopeless* using an alternative sensor type, such as broadband. Examples of challenging situations where the use of HS data may help over broadband data are: partially obscured targets, targets parked in tree shadows, camouflaged targets, and (if operating in the long wave infrared region of the spectrum) stationary relatively cold targets. Although not discussed in this work, sensors that operate over a few number of bands (e.g., ten)—known as multispectral, may enjoy the same advantage of HS sensors, but this advantage may depend on the material type of potential targets and on the operational bands of multispectral sensors. This work shows that using HS sensors can help on challenging anomaly detection problems (e.g., be able to find targets in tree shadows), albeit the impact of this help was shown to be highly dependent on the effectiveness of the anomaly detector.

- Conventional Anomaly Detection: I discovered that conventional techniques do not adequately address all of the most common spatial/spectral variability occurrences that may be observed locally in HS imagery. Therefore, they often produce an intolerable high number of false alarms, as it would be characterized by an image analyst performing the same task in the image. In the strict statistical sense, these false alarms are actually justifiable detections, i.e., they actually represent local anomalies when compared to their immediate surroundings. To better understand the behavior of conventional anomaly detectors on actual HS data, I applied five known techniques on HYDICE top view imagery, including the industry standard, and decomposed spatial/spectral variability occurrences into three most probable study cases: Case 1, Case 2, and Case 3. Case 1 represents a comparison between two samples from distinct distributions, Case 2 represents a comparison between a two-material sample and a sample of one of the two materials, and Case 3 represents a comparison between two samples from the same distribution. I concluded through simulation and inspection of these detectors' performances on actual HS data that the *application* of conventional techniques to local anomaly detection problems is *flawed*. They are developed to account for Case 1 and Case 3, but not Case 2. Case 2 occurs quite often on digitized scenes, representing major transitions of regions (e.g., a spatial transition between tree shadows and surrounding terrain), or simply as strong edges owing, for instance, to the presence of manmade objects in a natural clutter background. This discovery applies to conventional techniques based on parametric or nonparametric approaches. (Although the application of a strict nonparametric technique was not included in this research, inspection of the empirical distributions shown in figure 1 should convince the reader.) I had a plausible idea for the development of algorithms aimed at explicitly accounting for all three study cases: compare samples indirectly by combining them.

- Principle of Indirect Comparison: I proposed to compare samples not as individual entities, but as individual entities and the union among these entities. This idea was motivated by a discovery and a key recognition. I first realized that improving data models would not improve performance of anomaly detectors based on these models, as Case 2 would still be a cause of anomalous responses using these detectors. In addition, I recognized that Case 2 may be interpreted as an indirect comparison between two samples from different populations, where the union of these samples are compared to one of the samples. In the context of anomaly detection, let X and Y denote two random samples, and let $Z = X \cup Y$, where U denotes the union. Features of the distribution of X can be indirectly compared to features of the distribution of Y by comparing instead features of the distributions of Z to Y. Distribution features correspond to lower and/or higher moments and central moments. I developed and showed that anomaly detection algorithms

based on this simple principle enjoyed the desirable outcome of preserving what is often characterized by image analysts as meaningful detections (e.g., a manmade object in natural clutter), while significantly reducing the number of meaningless detections (e.g., transition of distinct regions). These algorithms are discussed next.

- Semiparametric Anomaly Detector: I used a statistical approach that implements the principle of indirect comparison naturally in its mathematical development through a semiparametric model (a logistic model). This model assumes that the distributions of, for instance, two random samples X and Y are related by an exponential distortion. A statistical hypothesis test is then applied to decide whether the exponential distortion is insignificant. If this null hypothesis is rejected, then X and Y are declared anomalous in respect to each other. This model requires that all the components of X and Y are independent, identically distributed (iid) by their corresponding distributions. (I proposed a data preprocessing technique, which applies a first order differentiation in the spectral domain followed by angle mapping, to transform highly correlated spectral samples into approximately iid samples). Under this null hypothesis, the test statistic tends in law to a chi square distribution, as the number of samples increases to infinity. I tailored this technique (SemiP) to address the top view anomaly detection problem using HYDICE HS (V-SWIR) data and showed a significant improvement over conventional techniques suppressing natural clutter backgrounds and accentuating, as a collection of localized anomalies, the presence of stationary land vehicles in the scene. I tested the SemiP detector on HS radiance data from two different types of backgrounds, forest and desert. Its performance was consistent in both backgrounds. The implementation of this detector, however, revealed a major drawback. It requires an iterative optimization algorithm which may be sensitive to arbitrary initial conditions. A fixed initialization used for the HYDICE data worked very well on that dataset, but not so well on the SOC-700 HS (V-NIR) data, which was collected to address target detection as a collection of anomalies observed from a ground level perspective. The problem was that using a fixed initialization, the detector could not converge to a solution at each tested location in the SOC-700 imagery, so, in order to continue its function at those locations, I would have to account for varying initial guesses, which of course is computationally too expensive. I developed other algorithms based on the same principle of indirect comparison as alternatives to the semiparametric approach.

- Approximation of the Semiparametric Anomaly Detector: I developed an alternative detector based on the functional behaviors of the different components in the semiparametric test statistic. By defining new functions and applying fundamental theorems of large sample theory, I showed that under its null hypothesis the new test statistic (AsemiP) also converged in law to the same chi square distribution of the semiparametric test statistic, as the number of samples increased. I also showed that performance of the

AsemiP detector is comparable to performance of the SemiP detector on HYDICE data by comparing ROC curves and decision surfaces. Prominent advantages of the AsemiP detector over the SemiP detector are that the former is free from parameter initialization, computationally less expansive, and significantly simpler to implement. Its independence from initialization allowed us to test the AsemiP detector on SOC-700 imagery, using as references two sets of spectral samples from tree leaves and general terrain, as a priori information. The same reference sets were used to test additional HS cubes, including a cube having a land vehicle and a standing person being almost invisible in the tree shadows. Its results were compared to the RX detector (industry standard) and to other conventional detectors on this dataset by computing decision surfaces. This experiment showed the difficulty of attempting to find manmade objects in natural clutter as a collection of localized anomalies in respect to some fixed spectral sets used as reference. The results produced by the conventional techniques were virtually useless for the intended purpose. The AsemiP detector, on the other hand, using the principle of indirect comparison was able to suppress almost entirely the clutter background and to accentuate the manmade objects (vehicles and a man) in the SOC-700 HS data.

- F-distribution Anomaly Detectors: I developed a third detector using the principle of indirect comparison and large sample theory, as an alternative, albeit this time I aimed at using a known property of the F-distribution family as a model. Our interest to introduce a detector having an asymptotic behavior governed by the F-distribution family was motivated by the classic one-way ANOVA, which has a test statistic governed by an F distribution—exactly, under its null hypothesis and model's assumptions. The ANOVA model uses the normality assumption. I tested both techniques on the HYDICE data (forest and desert radiance), computed ROC curves, and concluded that their performances are highly comparable to each other on desert radiance (sparse vegetation), but the SemiP and AsemiP detectors significantly outperformed both F-distribution based detectors at a region of extremely low false alarm rate. In the forest radiance data (where region discontinuity is quite abundant in the scene), both F-distribution based detectors performed comparably to results produced by the SemiP and AsemiP detectors, i.e., they significantly outperformed conventional detectors. The ANOVA detector yielded significantly more false alarms in the forest radiance data at a region of extremely low false alarm rate compared to the results produced by the indirect comparison detectors. The reason may be arguably attributed to the normality assumption in the ANOVA model. This reason may also have contributed to the differences in performance between the AFT and ANOVA detectors testing the ground view data from the SOC-700 sensor.

- Asymmetric Variance Based Anomaly Detector: I developed a fourth detector, albeit this time I aimed at designing the most compact form to implement the notion of indirect comparison. I showed how effective a

simple asymmetric hypothesis test (based exclusively on variances) can be determining whether random samples are governed by different distributions. I tested this detector (AVT) on HYDICE and SOC-700 data and compared to other detectors through ROC curves and decision surfaces. The AVT detector is simple, elegant, and performs comparably with the other indirect comparison based detectors. It significantly outperformed the conventional detectors presented in this research. Using both AVT and AsemiP detectors, I also showed the result of a proof of principle experiment that extended the utility of an effective anomaly detector from merely performing a first level of object detection on a HS scene to further discriminating these detected objects by their own classes. I named this notion self-classification. The notion is that after an anomaly detector tests a HS data using pre-stored reference sets of spectral samples, spectral samples from the most accentuated anomaly clusters (which can be interpreted as taking samples from spatially independent multi-pixel objects) are reintroduced to the detector as a new reference set of spectral samples aimed at separating these clusters by class. In this context, I showed that in a scene consisting of three stationary land vehicles and a standing person, the vehicles were classified to belonging to the same class and the person classified to belonging to a different class.

- Impact of Research: The results of our research have introduced novelty in the concept and development of algorithms to a difficult problem of localized anomaly detection using a passive sensor device. The impact of this research is summarized as follows: (i) A principle of indirect comparison (i.e. given random samples X and Y, let $Z = X \cup Y$ and compare instead in some form features of the populations of Z and Y) has been proposed as novelty to address a computer vision problem; (ii) a semiparametric approach has been proposed as novelty to address object detection problems in the geoscience and remote sensing, image processing, and pattern recognition communities significantly outperforming conventional techniques; (iii) alternative techniques have been developed using the same principle and shown to perform comparably with the semiparametric approach, albeit free from its potential implementation drawback ; (iv) the role of anomaly detectors testing digitized scenes has been elevated from performing mere screening to performing *focus of attention* in a form that is meaningful to an image analyst; (v) the presence of stationary manmade objects under heavy tree shadows has been shown to be detectable as a collection of localized anomalies using visible to near infrared HS imagery. Philosophically, it has been remarkable to learn through this research that a relatively simple set of rules (rules to preprocess spectral data followed by rules to test the transformed data) applied locally to HS imagery produce spatially independent results, which are not very useful independently owing to their atomic nature, but that once they are cumulatively assembled in some logical form (e.g., as a 2D surface), they produce spatial structures that can virtually agree with an outcome produced by the analysis of an image analyst performing a surveillance task in the

same data. The most remarkable part is that by default these local rules are completely *unaware* of the global scene, or of the spatial object patterns, which is not the case for image analysts. Image analysts use all the information (local and global) that they can sense from digitized scenes to perform shape analysis and pattern recognition before focusing their attention to objects they characterize as meaningful.

4.2 Limitations

In this subsection, I focus on foreseeable limitations of the overall approach developed in this research. Some of the limitations have been already discussed in a previous subsection, such as our decision to use HS imagery rather than broadband imagery in order to realize robust object detection in natural clutter background, and the dependence of the semiparametric detector on an initial parameter guess. This problem was remedied by developing alternative detectors.

Perhaps the most important limitation in applying our approach to anomaly detection is that targets in some background can only be detected, if indeed they are *detectable*—this statement also applies to conventional or non-conventional detectors. In other words, the presence of manmade objects, for instance, in some scene can only be accentuated using our approach, if in fact samples from these targets have measurable differences from samples referred here as *references*. For example, if a soldier is hiding somewhere in a natural foliage background using a camouflage sniper suit, which is designed to have similar material spectral characteristics of foliage, the camouflaged sniper will certainly be concealed and be able to deceit casual or even critical observers of that scene. If manufacturers of camouflaged sniper suits are in fact successful achieving their goals, the approach developed in this work would not be able to assist an image analyst in this scenario.

The main strength of our approach using HS data is that it can significantly suppress areas in digitized scenes that are characterized by transitions of regions, and it is more tolerant to spectral variability of objects belonging to similar classes. I expect this strength to be diminished by some measure when our approach is applied to anomaly detection problems using broadband imagery, for reasons already discussed in this section. But independent of which sensor type is used, our approach should never be applied to the so-called subpixel target detection problems. Subpixel targets are objects of interest that are smaller than the pixel resolution of the data, given the range between target and sensor. A pixel consisting of a subpixel target displays the integrated radiances of both target and clutter, thus, our indirect comparison detectors may actually suppress the value of such a pixel.

Another foreseeable limitation is that a sensor, to have a practical value, must be able to produce a digitized scene in a rate comparable to that of a video rate (e.g., 30 to 60 frames per seconds), which is significantly above the rate of the state of the art portable hyperspectral sensors (1 to 10 cubes [640 pixels x 640 lines x 120 bands] per minute). This fact would impose a major practical constraint attempting to apply our approach to an actual surveillance task using a HS camera as the primary sensor to collect data. Advances in technology, however, have been occurring in remarkable

speeds since the 1990's, especially in the field of electronic material technology, which make us believe that such a limitation will no longer exist in the next few years. This concern on HS hardware speed can be also extended to the computational time required to execute our approach in hardware. Algorithms that are developed to perform detection tasks using HS data are notorious for being slow (taking hours, sometimes days to operate on a cube), not necessarily because of the algorithm itself, but because of the vast amount of data a HS cube actually represents, i.e., a cube represents the same digitized scene as a collection of B images having some size, B denoting the number of bands. A method that is often used to reduce the computational time of HS algorithms, making then feasible to apply our approach to a practical surveillance task, requires instead a compromise between HS and broadband sensors. The method is known as band selection, which is briefly discussed in the next section.

4.3 Future Work

In the future, work is needed to develop more insight into the following:

- Spectral Band Selection: A method that may be used to circumvent the speed limitation issue discussed in this section is to use instead a sensor that is a compromise between hyperspectral and broadband, i.e., a sensor that collects radiance using only a few spectral bands (e.g., 10), forming in the process a multispectral cube. Notice that, by default, a multispectral sensor should be able to collect data faster by an order of magnitude or two than a hyperspectral sensor can, given the same swath coverage. In addition, the computational cost of detection algorithms due to this reduced amount of data representing a scene may decrease by the same order of magnitude. A key decision, however, that must be made before developing multispectral sensors is to determine how many of these frequency bands and which ones should feature in these devices. A long list of contributions can be found in the literature (see, for instance, [39]) devoted exclusively to answer this question. The conclusions of these contributions, however, independently of the method applied share explicitly, or implicitly, a common message: *It depends*. It depends on the type of materials one is interested in detecting. It depends on the number of material classes one expects to find in the same scene, and it depends on the region of the EM spectrum the sensor is expected to operate, etc. To follow on with our research, I plan to use a test statistic (e.g., AsemiP) as a decision criterion, and also I would like to find all types of manmade objects in natural clutter backgrounds to determine the minimum number of combination of bands that will maximize performance in a HS dataset from a particular sensor (e.g., SOC-700).

- Randomly Sampling the Scene: It was evident from our discussion using imagery from a perspective of the ground level view that I used two reference sets of spectral samples obtained a priori for the online operation of our detectors. I plan to evaluate different types of random sampling techniques to study the effects of eliminating the need for a priori spectral information, while attempting to perform the same task. I suspect that some

parameter settings may be required, either by the user or already built into the detector, to determine the optimum number of samples. This determination will require a given set of known parameters (e.g., image size, pixel resolution) and a given set of unknown parameters (e.g., expected maximum range between targets and the sensor, expected maximum size of a target).

- Self Classification: A proof of principle experiment was presented to show the feasibility of using an effective anomaly detector, as a solo discriminant engine, to perform both detection and self classification among the detected objects. The preliminary results were quite promising, which motivate us to pursue this avenue further using additional data and different types of materials.

- Cultural Clutter Backgrounds: Another natural extension of this work is to evaluate the behavior of our approach as it attempts to detect the presence of certain types of targets (e.g., standing personnel, stationary motor vehicles) in an urban scenario, often referred to in the target detection community as cultural clutter backgrounds. Thus, I am interested in determining whether the introduction of spectral samples from cultural clutter (e.g., painted walls of local buildings, sidewalks and asphalt from the streets), as reference samples, to our approach would produce a performance level comparable to its equivalent performance level on natural clutter backgrounds. I are actively searching for such a HS dataset of cultural clutter to perform this evaluation.

Appendix A.–Asymptotic behavior of the SemiP algorithm

References are made to models (17)-(18) and to hypotheses (19). Lemma 1A is relevant to estimators based on function maximization with respect to unknown parameters.

Lemma 1A [40]. Assumptions:

(i) Let Θ be an open subset of the Euclidean K-space. (Thus the true value θ_o is an interior point of Θ.)

(ii) $Q_T(y,\theta)$ is a measurable function of vector y for all $\theta \in \Theta$ and $\partial Q_T / \partial \theta$ exists and is continuous in an open neighborhood $N_1(\theta_o)$ of θ_o. (Note that this implies $Q_T(y,\theta)$ is continuous for $\theta \in N_1$, where T is the sample size.)

(iii) There exists an open neighborhood $N_2(\theta_o)$ of θ_o such that $T^{-1}\, Q_T(\theta)$ converges to a nonstochastic function $Q(\theta)$ in probability uniformly in θ in $N_2(\theta_o)$, and $Q(\theta)$ attains a strict local maximum at θ_o.

Let Θ_T be set of roots of the equation

$$\frac{\partial Q_T}{\partial \theta} = 0 \tag{1A}$$

corresponding to the local maxima. If that set is empty, set Θ_T equals to {0}. Then, for any $\varepsilon > 0$,

$$\lim_{T\to\infty} P[\inf_{\theta\in\Theta_T} (\theta-\theta_0)'(\theta-\theta_0) > \varepsilon] = 0. \tag{2A}$$

In essence, Lemma 1A affirms that there is a consistent root of (1A). (For the proof, see [40]. Under certain conditions, a consistent root of (1A) is asymptotically Normal. The affirmation is shown in Theorem 1A, where asymptotic convergence is denoted by $A \to B$.

Theorem 1A [40]. *Assumptions:*

(i) All the assumptions of Lemma 1.

(ii) $\frac{\partial^2 Q_T}{\partial\theta\partial\theta'}$ exists and is continuous in an open, convex neighborhood of θ_o.

(iii) $\frac{\partial^2 Q_T}{\partial\theta\partial\theta'}$ converges to a finite nonsingular matrix $S(\theta_0) = \lim E[T^{-1}(\frac{\partial^2 Q_T}{\partial\theta\partial\theta'})_{\theta_0}]$

in probability for any sequence θ_T^* such that $\theta_T^* = \theta_0$.

(iv)
$$\sqrt{T}(\frac{\partial Q_T}{\partial\theta})_{\theta_0} \to N[0, V(\theta_0)], \text{ where} \tag{3A}$$

$$V(\theta_0) = \lim E[T^{-1}(\frac{\partial Q_T}{\partial \theta})_{\theta_0} \times (\frac{\partial Q_T}{\partial \theta'})_{\theta_0}]. \tag{4A}$$

Let $\{\hat{\theta}_T\}$ be a sequence obtained by choosing one element from Θ_T defined in Lemma 1 such that $\hat{\theta}_T \to \theta_0$.

Then
$$\sqrt{T}(\hat{\theta}_T - \theta_0) \to N(0,\Sigma), \text{ where} \tag{5A}$$
$$\Sigma = S(\theta_0)^{-1} V(\theta_0) S(\theta_0)^{-1}. \tag{6A}$$

For the proof, see also [40].

The semiparametric model's MLE solution satisfies the assumptions of Lemma 1A, including of course (1A) via (27). Therefore, by Lemma 1, ML estimators $\hat{\alpha}$ and $\hat{\beta}$ are consistent and, as I shall see by Theorem 1A, it converges asymptotically to a Normal distribution.

Under H_0: $\beta = 0$ ($g_1 = g_0$), I shall use the following notation for the moments of t (the union of the samples x_0 and x_1) with respect to the reference distribution g_0:

$$E(t^k) \equiv \int t^k g_0(t)dt,$$
$$Var(t) \equiv E(t^2) - E^2(t) \tag{7A}$$

Let (α_0,β_0) be the true value of (α,β) under model (17)-(18) and assume $\rho = n_1/n_0$ remains constant as both n_1 and n_0 go to *infinity*. Define $\nabla \equiv \left(\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta}\right)$ and notice from (27) that $E\left[\nabla l(\alpha_0,\beta_0)\right] = 0$. Under the null hypothesis (H_0: $\beta = 0$ [$g_1 = g_0$]), using (18), (26), (27), and (7A) one can verify that

$$-\frac{1}{n}\frac{\partial^2 l(\alpha_0,\beta_0)}{\partial\alpha\partial\beta} \to K_1 \int \frac{t\exp(\alpha_0+\beta_0 t)}{1+\rho\exp(\alpha_0+\beta_0 t)} g_0(t)dt$$

$$= K_2 \int t \cdot g_0(t)[\exp(\alpha_0+\beta_0 t) g_0(t)]dt$$
$$= \frac{\rho}{1+\rho} E(t), \tag{8A}$$

where K_1 and K_2 are constants involving (n_1,n_0) and $\rho/(1+\rho) = n_1/n$ (where $n = n_1 + \tilde{n_0}$. Using similar argument to arrive at (8A) and the application of WLLN, one can use assumption *(iii)* in Theorem 1A to recognize that

$$-\frac{1}{n}\nabla\nabla' l(\alpha_0,\beta_0) \to S = \frac{\rho}{1+\rho}\begin{pmatrix} 1 & E(t) \\ E(t) & E(t^2) \end{pmatrix} \tag{9A}$$

in probability as $n \to \infty$. It follows that S is nonsingular and its inverse is

$$S^{-1} = \frac{1}{E(t^2) - E^2(t)} \begin{pmatrix} E(t^2) & -E(t) \\ -E(t) & 1 \end{pmatrix} \frac{1+\rho}{\rho}. \tag{10A}$$

Our interest is only in the parameter β, so, let S_β denote the lower-right component of the expanded version of S^{-1} and use (7A) to obtain

$$S_\beta = \frac{1}{E(t^2) - E^2(t)} \frac{1+\rho}{\rho} = \frac{1}{Var(t)} \frac{1+\rho}{\rho}. \tag{11A}$$

Using also the application of CLT in Theorem 1A *(iv)* and the fact that

$$E\left[\nabla l(\alpha_0, \beta_0)\right] = 0\,, \tag{12A}$$

from (27), one can write

$$\sqrt{n}\left[\nabla l(\alpha_0, \beta_0)\right] \to N[0, V(\alpha_0, \beta_0)], \tag{13A}$$

where

$$V(\alpha_0, \beta_0) = \frac{\rho}{1+\rho} \begin{bmatrix} 1 & E(t) \\ E(t) & E(t^2) \end{bmatrix} - \rho \begin{bmatrix} 1 \\ E(t) \end{bmatrix} \begin{bmatrix} 1 & E(t) \end{bmatrix}. \tag{14A}$$

$V(\alpha_0,\beta_0)$ is a direct result from (4A), see, for instance, [28]. Using the conclusion of Theorem 1A, or (5A)-(6A), in terms of S_β in (11A) and the lower-right component of the expanded version of $V(\alpha_0,\beta_0)$ in (14A), I can conclude that

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \xrightarrow[n\to\infty]{} N\left(0, \frac{\rho^{-1}(1+\rho)^2}{Var(t)}\right), \tag{15A}$$

and having the left side of (15A) normalized by the asymptotic variance and then squared, one can conclude that the resulting random variable

$$Z_{SemiP} = n\rho(1+\rho)^{-2}\hat{\beta}^2\hat{V}(t) \xrightarrow[n\to\infty]{} \chi_1^2, \tag{16A}$$

converges to a chi square distribution with *1* degree of freedom, where $\hat{V}(t)$ estimates *Var(t)*. A multivariate solution is presented in [27]. □

Appendix B.–Asymptotic Performances of Detectors SemiP and AsemiP

To complement our analysis between detectors SemiP and AsemiP, I took a closer look at the performances of (28) and (46), under their corresponding null hypotheses, and display our results in figure 1B. Recall that under these null hypotheses (and having assumptions not grossly violated), the random outcome of both SemiP and AsemiP detectors should converge to a chi square distribution with *1* dof, as the number of samples increases to infinity. I checked for this behavior by empirically estimating the pdf using values from both output surfaces and comparing it to an empirical distribution obtained from independent realizations of an equivalent number of random samples from a chi square pdf with *1* dof. (MATLAB™ software was used to generate the chi-square samples.) To achieve our goal, I only used samples from the SemiP and AsemiP surfaces with values less than 5.0—about 2,200 samples from each surface—since the probability of obtaining the realization of chi-square (*1* dof) random variables above this value is less than 0.001. Results are shown in Fig. 1B in the form of bar plots (empirical pdf obtained from samples of the SemiP and AsemiP surfaces) and line plots (empirical pdf obtained from a set of 2,000 independent chi-square realizations, with *1* dof).

Figure 1B shows a remarkable agreement between the empirical distributions of SemiP and AsemiP for output results below 5.0; it also shows an even more remarkable fit of their asymptotic behaviors to the chi-square distribution with *1* dof, as it was predicted from both theories, under their null hypotheses and idealized assumptions. The quality of those fits gives also a vote of confidence to our choice of suitably transforming highly spatially/spectrally correlated HS data with the applications of a HPF (or first order differentiation in the spectral domain) followed by SAM (or angle difference in the spatial domain) to promote statistical independence in HS data for a test statistic that do not assume normality.

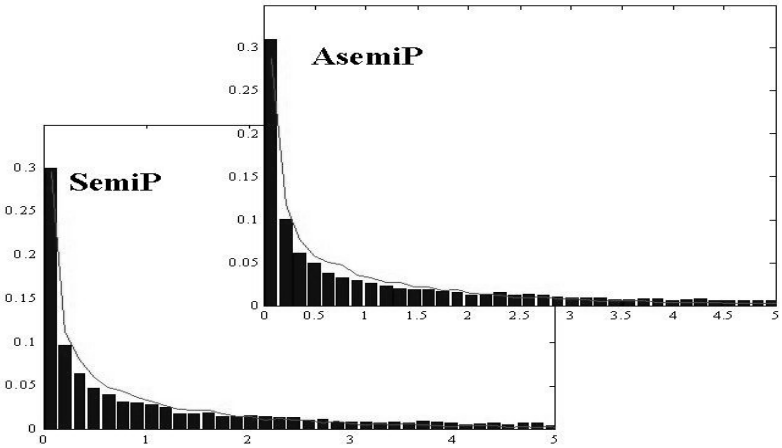

Figure 1B. Asymptotic behaviors of detectors SemiP and AsemiP. The bar plots represent the empirical distributions of samples from the SemiP and AsemiP output surfaces. The line plots represent the empirical distribution obtained from 2,000 independent realizations

References

1. S. Ullman. High-Level Vision: Object Recognition and Visual Cognition, Cambridge, MA: MIT Press, pp. 4-12, 1996.
2. R. A. Schowengerdt. Remote Sensing, Models and Methods for Image Processing, 2nd ed. San Diego, CA: Academic Press, 1997.
3. J. B. Campbell. Introduction to Remote Sensing, 2nd ed. New York: Guilford, 1996.
4. T. M. Lillesand and R. W. Kiefer. Remote Sensing and Image Interpretation. New York: Wiley, 1994.
5. E. Crist, C. Schwartz, and A. Stocker, "Pairwise adaptive linear matched-filter algorithm," in Proc DARPA Adaptive Spectral Reconnaissance Algorithm Workshop, Jan 1999.
6. H. T. Haskett, and A. K. Sood, "Adaptive real-time endmember selection algorithm for sub-pixel target detection using hyperspectral data," in Proc. 1997 IRIS Specialty Group Camouflage, Concealment, Deception, Oct. 1999.
7. J. Grossmann, J. Bowles, D. Hass, J. Antoniades, M. Grunes, P. Palmadesso, D. Gillis, K. Tsang, M. Baumback, M. Daniel, J. Fisher, and I. Triandaf "Hyperspectral analysis and target detection system for the adaptive-spectral reconnaissance program (ASRP)," Proc SPIE, vol. 3372, pp.2-13, Apr. 1998.
8. C. Chang, X. Zhao, M. Althouse, and J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," IEEE Trans. Geosci. Remote Sensing, vol. 36, pp. 898-912, May 1998.
9. D. Slater and G. Healey, "Exploiting an atmospheric model for automated invariant material identification in hyperspectral imagery," Proc. SPIE, vol. 3372, pp. 60-71, Apr. 1998.
10. X. Yu, L. Hoff, I. Reed, A. Chen, L. Stotts, "Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach," IEEE Tran. Image Processing, vol. 6, pp. 143-156, Jan. 1997.
11. H. Kwon, N. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral," IEEE Trans. on Geoscience and Remote Sensing, vol. 43, no. 2, February, 2005.

12. A. Stocker, “Stochastic expectation maximization (SEM) algorithm,” in Proc. DARPA Adaptive Spectral Reconnaissance Algorithm Workshop, Jan. 1999.

13. P. Masson and W. Pieczynski, “SEM algorithm and unsupersivsed statistical segmentation of satellite images,” IEEE Trans. Geosci. Remote Sensing, vol. 31, pp. 618-633, May 1993.

14. E. L. Lehmann. Theory of Point Estimation, Pacific Grove, CA: Wadsworth & Brooks, 1991, pp. 333-336, and Chapter 5.

15. E. L. Lehmann. Testing Statistical Hypotheses, New York: transferred to Chapman & Hall, 2nd Edition, pp 68, 74, Chapter 3, 1993.

16. R.O. Duda, P.E. Hart. Pattern Classification Scene Anal., New York: J. Wiley & Sons, 1973.

17. H. Kwon, S.Z. Der, and N.M. Nasrabadi, “Adaptive anomaly detection using subspace separation for hyperspectral imgery,” Opt. Eng., v. 42 (11), November 2003, pp. 3342-3351.

18. S. M. Schweizer and J. M. F. Moura, “Hyperspectral imagery: clutter adaptation in anomaly detection,” IEEE Trans. Information Theory August 2000, 46, no. 5, pp. 1855–1871.

19. Margalit and I. S. Reed, “Adaptive detection of stationary optical and IR targets using correlated scene,” Ph.D dissertation, Univ. of Southern Calif., Los Angeles, Nov. 1984.

20. Stocker, I. S. Reed, and X. Yu, “Multidimensional signal processing for electro-optical target detection,” Proc. SPIE, Orlando, FL, vol. 1035, Apr 1990.

21. L. B. Stotts, E. M. Winter, L. E. Hoff, and I. S. Reed, “Clutter rejection using multispectral processing,” Proc. SPIE, Orlando, FL, vol. 1305, Apr 1990.

22. V. N. Vapnik. The Nature of Statistical Learning Theory. Springer, 1995.

23. Schokopf, A. J. Smola, and K-R Muller, “Kernel principal component analysis,” Neural Computation, (10):1299-1319, 1999.

24. G. Baudat and F. Anouar, “Generalized discriminant analysis using a kernel approach,” Neural Computation, (12):2385-2404, 2000.

25. J. R. Schott. Remote Sensing: The Imaging Chaing Approach. Oxford University Press, New York, 1997.

26. Kedem. Time Series Analysis by Higher Order Crossings, New York: IEEE Press, 1994, pp. 260-266.

27. K. Fokianos, B. Kedem, J. Qin, "A semiparametric approach to the one-way layout," Technometrics, pp. 56--65, 2001.

28. J. Qin and B. Zhang, “A goodness of fit test for logistic regression models based on case-control data,” Biometrika, vol 84, 609-618, 1997.

29. J. Qin and J. F. Lawless, “Empirical likelihood and general estimation equation,” Ann. of Statistics, vol 22, 1994, pp. 300-25.

30. J. A. Anderson, “Separate sample logistic discrimination,” Biometrika, vol. 59, 19-35, 1972.

31. R. Prentice and R. Pyke, “Logistic disease incidence models and case-control studies,” Biometrika, vol. 66, pp. 403-411, 1979.

32. R. Cox, “Some procedures associated with the logistic qualitative response curve,” Research Papers in Statistics: Festschrift for J. Neyman, New York: John Wiley, 1966, pp. 55-71.

33. J. S. Cramer, “The Origin of Logistic Regression,” Tinbergen Institute, U. of Amsterdam: TI 2002-119/4. Website: http://ideas.repec.org/p/dgr/uvatin/20020119.html.

34. Lagarias, J.C., J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," SIAM Journal of Optimization, Vol. 9 Number 1, pp. 112-147, 1998.

35. G. Casella and R. L. Berger. Statistical Inference, Belmont, CA: Duxbury Press, 1990, pp. 112-120, 184, 222, 232.

36. H. Scheffe. The Analysis f Variance, New York: John Wiley & Sons, 1959, Chapter 3.

37. P. Sen and J. Singer. Large Sample Methods in Statistics: An Introduction with Applications , Washington, D.C.: Chapman & Hall/CRC, 1993, pp. 125-127.

38. Surface Optics Corp., San Diego, CA, official website: www.surfaceoptics.com.

39. N. Keshava and P. Boettcher, “On the relationships between physical phenomena, distance metrics, and best band selection in hyperspectral processing,” in Proc. SPIE Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII, vol. 4381, Apr 2001.

40. T. Amemiya. Advanced Econometrics, Harvard U. Press: Cambridge, MA, 1985, pp. 110–112.

Distribution list

Admnstr
Defns Techl Info Ctr
ATTN DTIC-OCP (Electronic copy)
8725 John J Kingman Rd Ste 0944
FT Belvoir VA 22060-6218

DARPA
ATTN C Schwartz
ATTN IXO S Welby
ATTN R Hummell
3701 N Fairfax Dr
Arlington VA 22203-1714

Ofc of the Secy of Defns
ATTN ODDRE (R&AT)
The Pentagon
Washington DC 20301-3080

US Army RDECOM AMRDEC
ATTN ATCD-B
FT Monroe VA 23651-5850

Army Rsrch Physics Div
ATTN AMSRD-ARL-RO-MM
R Launer
PO Box 12211
Research Triangle Park NC 27709-2211

US Army TRADOC
Battle Lab Integration & Techl Dirctrt
ATTN ATCH-B
10 Whistler Lane
FT Monroe VA 23651-5850

CECOM NVESD
ATTN AMSRD-CER-NV-OD
J Ratches
10221 Burbeck Rd Ste 430
FT Belvoir VA 22060-5806

US Army Aberdeen Test Center
ATTN CSTE-DT-AT-WC-A
F Carlen
400 Colleran Road
Aberdeen Proving Ground MD 21005-5009

US Army Aberdeen Test Center
ATTN CSTE-DTC-AT-TC-N
D L Jennings
400 Colleran Road
Aberdeen Proving Ground MD 21005-5059

US Army ARDEC
ATTN AMSTA-AR-TD
Bldg 1
Picatinny Arsenal NJ 07806-5000

US Army Aviation & Mis Lab
ATTN AMSRD-AMR-SG-IP
H F Anderson
Bldg 5400
Redstone Arsenal AL 35809

Commanding General
US Army Avn & Mis Cmnd
ATTN AMSAM-RD
W C McCorkle
Redstone Arsenal AL 35898-5000

US Army CERDEC, NVESD
ATTN AMSRD-CER-NV-xx J Hilger
ATTN AMSRD-CER-NV-xx P Perconti
ATTN AMSRD-CER-NV-xx R Driggers
10221 Burbeck Rd Ste 430
FT Belvoir VA 22060-5806

US Army Natick RDEC Acting Techl Dir
ATTN SBCN-TP
P Brandler
Kansas Street Bldg 78
Natick MA 01760-5056

Distribution list

US Army RDECOM AMRDEC
ATTN AMSRD-AMR-SG-IP B Pittman
ATTN AMSRD-AMR-SG-IP R Sims
Bldg 5400
Redstone Arsenal AL 35898

US Army RDECOM AMRDEC
ATTN AMSRD-AMR-WS-PL
W Davenport
Bldg 7804
Redstone Arsenal AL 35898

US Army RDECOM ARDEC
ATTN AMSRD-AAR-QES P Willson
Radiographic Laboratory, B.908
Picatinney Arsenal NJ 07806-5000

US Army RDECOM TARDEC
ATTN AMSRD-TAR-R G R Gerhart
MS 263
Warren MI 48397-5000

US Army Soldier & Biological Chem Ctr
ATTN AMSSB-RRT-DP W Loerop
ATTN AMSSB-RRT-xx F D'Amico
Edgewood Chem & Biological Ctr
Bldg E-5554
Aberdeen Proving Ground MD 21010-5424

US Army Topograpphic Engrg Ctr
ATTN CEERD-RR-S R Rand
7701 Telegraph Rd
Alexandria VA 22315

Commander
USAISEC
ATTN AMSEL-TD Blau
Building 61801
FT Huachuca AZ 85613-5300

AFRL/SNAA
ATTN M Jarratt
2241 avionics Circle Area B, Bldg 620
Wright Patterson AFB OH 45433-7321

CMTCO
ATTN MAJ A Suzuki
1030 S Highway A1A
Patrick AFB FL 23925-3002

SITAC
ATTN H Stiles
ATTN K White
ATTN R Downie
11981 Lee Jackson Memorial Hwy Suite 500
Fairfax VA 22033-3309

US Army Rsrch Lab
ATTN AMSRD-ARL-WM-BF G Haas
ATTN AMSRD-ARL-WM-BF W Oberle
Aberdeen Proving Ground MD 21005-5067

US Army Rsrch Lab
ATTN AMSRD-ARL-RO-EL W Sander
PO Box 12211
Research Triangle Park NC 27709-2211

US Army Rsrch Office
ATTN AMSRD-ARL-RO-PP
R Hammond
PO box 12211
Research Triangle Park NC 27709-2211

National Geospatial-Intelligence Agency
Wayne Hallada
Mail Stop N-06
1200 1st Street SE
Washington, D.C., 20303-0001

U.S. Dept. of the Interior
U.S. Geological Survey
Gregory L. Stensaas
EROS Data Center
Mundt Federal Building
Sioux Falls, SD 57198

US Army Rsrch Lab
ATTN AMSRD-ARL-CI-OK-T Techl Pub (2 copies)
ATTN AMSRD-ARL-CI-OK-TL Techl Lib (2 copies)
ATTN AMSRD-ARL-D A Grum
ATTN AMSRD-ARL-D J M Miller
ATTN AMSRD-ARL-SE J Pellegrino
ATTN AMSRD-ARL-SE-S J Eicke
ATTN AMSRD-ARL-SE-SE D Rosario (6 copies)
ATTN AMSRD-ARL-SE-SE N Nasrabadi
ATTN AMSRD-ARL-SE-SE P Gillespie
ATTN IMNE-ALC-IMS Mail & Records Mgmt