IMPROVED HYPERSPECTRAL IMAGE TESTING USING SYNTHETIC IMAGERY AND FACTORIAL DESIGNED EXPERIMENTS

THESIS

Joseph P. Bellucci, Captain, USAF

AFIT/GOR/ENS/07-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GOR/ENS/07-01

IMPROVED HYPERSPECTRAL IMAGE TESTING USING SYNTHETIC IMAGERY AND FACTORIAL DESIGNED EXPERIMENTS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Joseph P. Bellucci, BS

Captain, USAF

March 2007

AFIT/GOR/ENS/07-01

IMPROVED HYPERSPECTRAL IMAGE TESTING USING SYNTHETIC IMAGERY AND FACTORIAL DESIGNED EXPERIMENTS

Joseph P. Bellucci, BS
Captain, USAF

Approved:

_____	_____
Dr. Kenneth W. Bauer (Chairman)	date
_____	_____
Dr. Marcus B. Perry (Member)	date

Abstract

The goal of any remote sensing system is to gather data about the geography it is imaging. In order to gain knowledge of the earth's landscape, post-processing algorithms are developed to extract information from the collected data. The algorithms can be intended to classify the various ground covers in a scene, identify specific targets of interest, or detect anomalies in an image. After the design of an algorithm comes the difficult task of testing and evaluating its performance. Traditionally, algorithms are tested using sets of extensively ground truthed test images. However, the lack of well characterized test data sets and the significant cost and time issues associated with assembling the data sets contribute to the limitations to this approach.

This thesis uses a synthetic image generation model in cooperation with a factorial designed experiment to create a family of images with which to rigorously test the performance of hyperspectral algorithms. The factorial designed experimental approach allowed the joint effects of the sensor's view angle, time of day, atmospheric visibility, and the size of the targets to be studied with respect to algorithm performance. A head-to-head performance comparison of the two tested spectral processing algorithms was also made.

To My Wife and Daughter

Acknowledgments

It is only through the combined efforts of several people that this thesis work materialized. First, I would like to extend my sincere thanks and appreciation to Dr. Kenneth Bauer for his attention, guidance, insight, and support during this research. His enthusiasm for this study was a constant motivation. In addition, special thanks are due to Dr. Marcus Perry for his constructive comments and suggestions during the development of the experimental designs in this analysis. I am also very grateful to Maj Timothy Smetek for his endless assistance with my onslaught of questions concerning the use of DIRSIG.

I would like to offer my genuine gratitude to the members of the Super Group. Thank you for sharing your knowledge and friendship. The continuous supply of comedy also helped to make the act of writing a thesis much more bearable.

Finally, but certainly not least, I would like to express my most heartfelt thanks to my family. To my wife, thank you for tolerating the many long hours I spent shackled to a computer at AFIT. To my daughter, thank you for always flashing me a beautiful, toothless smile whenever I came home…even if it was just gas.

Joseph P. Bellucci

Table of Contents

Page

Abstract iv

Dedication v

Acknowledgments vi

List of Figures ix

List of Tables xi

I. Introduction 1

Background 1
Methodology 2
Preview 4

II. Literature Review 6

Hyperspectral Imaging 6
Synthetic Imagery 11
Anomaly Detection Algorithms 14
RX Method 15
BACON Method 16
OC Curves 17
Experimental Design 18

III. Methodology 25

Overview of the Approach 25
Experimental Factors and Factor Levels for Creating Synthetic Images 28
Sensor View Angle 28
Presence of Shadows in the Scene 29
Atmospheric Haze Levels 30
Size of Targets in a Scene 31
2^4 Factorial Designed Experiment for Creating Synthetic Images 32
Experimental Factors and Factor Levels for Algorithm Testing 34
Factorial Design, Factors, and Factor Levels for the RX Method 34
Factorial Design, Factors, and Factor Levels for the BACON Method 36
Analysis Plan for the Collected Data 38
The Overall Nested Design 40

IV. Results and Analysis 46

Overview and Assumptions 46
Creating the Synthetic Hyperspectral Images 47
Results from the RX Method 51
Results from the BACON Method 57
An RX versus BACON Comparison 63
Nested Experimental Design 64
RX – BACON Comparison #1 67
RX – BACON Comparison #2 68

V. Discussion 73

Conclusions 73
Contributions 74
Recommendations 75

Appendix A. Responses for All Images Using the RX Method 77

Appendix B. Responses for All Images Using the BACON Method 78

Appendix C. Model Diagnostic Plots for the RX Method 79

Appendix D. Model Diagnostic Plots for the BACON Method 80

Appendix E. OC Curves and Binary Maps for All Images 81

Bibliography 97

Vita 100

List of Figures

Page

Figure 1. Hyperspectral Data Cube (Shaw and Manolakis, 2003: 13) 8

Figure 2. Example of Spectral Variability (Shaw and Burke, 2003: 10) 10

Figure 3. Contributors to Image Degradation (Shaw and Manolakis, 2002: 14) 11

Figure 4. DIRSIG Western Rainbow Scene 14

Figure 5. OC Curves for RX Method on Synthetic Image 3 17

Figure 6. Geometric View of the 2^3 Factorial Design 20

Figure 7. Flowchart of Thesis Study Approach 26

Figure 8. Graphical Depiction of Factor Levels for Sensor View Angle 29

Figure 9. Comparison of Full Scale and Half Scale Targets 31

Figure 10. Graphical Depiction of the 2^4 Factorial Experiment 33

Figure 11. Graphical Representation of the 2^2 Factorial Design for the RX Method 35

Figure 12. An Example of Band Aggregation 37

Figure 13. Representation of the 2^2 Factorial Design for the BACON Method 38

Figure 14. The 2^6 Factorial Design to Analyze the RX Method 40

Figure 15. Comparison of a Few Synthetic Images 49

Figure 16. Truth Image versus Truth Image with Buffer Pixels 50

Figure 17. "Untestable" Pixels for RX Processing Window Sizes 52

Figure 18. Half Normal Plot for RX Method 54

Figure 19. Main Effects Plots for RX Method 55

Figure 20. Control-by-noise Interaction Plots for the RX Method 56

Figure 21. Initial (a) and Improved (b) BACON Results for Image 1 59
Figure 22. Half Normal Plot for BACON Method 60
Figure 23. Main Effects Plots for BACON Method 61
Figure 24. Interaction Plots for the BACON Method 62
Figure 25. Schematic of Nested Factorial Design 64
Figure 26. Main Effects Plots for the Nested Experimental Design 66
Figure 27. AX Interaction Plot for the Nested Design 67
Figure 28. Comparison of RX and BACON TPFs for All Images 68
Figure 29. OC Curves for Image 17 70
Figure 30. RX and BACON Results for Image 17 71

List of Tables

Page

Table 1. The 2^3 Factorial Design Matrix 20

Table 2. Contrasts for Main Effects and Interaction Effects 22

Table 3. The ANOVA Table for the 2^3 Factorial Design in *n* Replications 23

Table 4. Factor Levels for Sensor View Angle..... 29

Table 5. Factor Levels for Time of Day 30

Table 6. Factor Levels for Atmospheric Haze 31

Table 7. Factor Levels for Size of Targets..... 32

Table 8. 2^4 Factorial Experiment for Creating Synthetic Images 33

Table 9. The 2^2 Factorial Design for the RX Method..... 35

Table 10. Factors and Factor Levels for the Nested Factorial Design..... 41

Table 11. Partial ANOVA Table for the Nested Factorial Design 42

Table 12. Expected Mean Squares and F-statistics for the Nested Factorial Design 44

Table 13. Summary of Factors and Factor Levels for Synthetic Image Creation..... 48

Table 14. ANOVA Table for RX Method 53

Table 15. ANOVA Table for BACON Method..... 60

Table 16. ANOVA Table for the Nested Experimental Design 66

Table 17. RX Predictions for Image 17 69

Table 18. BACON Predictions for Image 17..... 69

IMPROVED HYPERSPECTRAL IMAGE TESTING USING SYNTHETIC IMAGERY AND FACTORIAL DESIGNED EXPERIMENTS

I. Introduction

Background

From the first manned hot air balloon flight in 1783, earth observation technology has advanced in a manner that allows us to gain more information about the world we live in. Photographs taken from a balloon by Nadar in 1858 were the beginnings of the evolution of remote sensing (Schott, 1997). Scientists in the 1950s and 1960s were motivated by the concurrence of digital computers, advancements in pattern recognition technology, and the launching of Sputnik to visualize how to observe the earth from space to acquire information to better manage its renewable and nonrenewable resources (Landgrebe, 2002: 17). As a result, remote sensing has grown into what presently includes high-tech aerial and satellite based electro-optical sensors that provide us immense environmental, economical, commercial, and military value.

Basically, information in remote sensing data is expressed by the spatial and spectral distribution of energy that is either reflected or emitted from earth and is then collected by an imaging system (Shaw and Manolakis, 2002: 13). Spatial characteristics of an image relate directly to the size, shape, pattern, and geometry of objects within the image. Spectral information refers to the electromagnetic distribution of light or thermal

energy. It is the differences in the spectral signatures of materials that allow for the identification and discrimination of materials in a scene. The emergence and utility of hyperspectral remote sensing systems take advantage of the fact that all materials reflect, absorb, and emit electromagnetic energy in distinct patterns related to their molecular composition (Manolakis and Shaw, 2002: 29).

Just as our brain processes the information our eyes gather, post-processing algorithms extract information from the data collected by the electro-optical sensor. Post-processing algorithms include anomaly detectors, target detectors, classifiers, and unmixing algorithms. After the design and implementation of an algorithm comes the often difficult task of testing and evaluating its performance. Traditionally, algorithms have been tested using sets of extensively ground truthed test images. There are several limitations to this approach. The first is the availability of well characterized test data sets. For example, what if a data set is not available that contains a specific target of interest? Another restriction is the diversity in available data sets (DIRS, 2006: 11). Is the algorithm robust if the performance can be verified on only a single scene type (forested, urban, etc.)? Or under a single set of atmospheric conditions? Or for a single time of day? When a synthetic image generation model is being utilized, the user can create a large set of test images that feature variations such as scene type, time of day, and atmospheric conditions so that the algorithm can be assessed under more conditions.

Methodology

It has been established from a review of the literature, that algorithm developers only test their algorithms on a minimal number of images. The reader may wonder how

an analyst can consider their algorithm to be robust if it is only tested on one or two images. This leads to the main thrust behind this thesis study – an examination of how differing several in-scene parameters affects the performance of hyperspectral algorithms and the usefulness of the information derived from them. This will allow us to better predict and understand the effectiveness of the post-processing algorithms under different in-scene parameters.

In brief, the basic approach that will be followed in this research is to begin with a synthetic hyperspectral image created with the Digital Imaging and Remote Sensing Image Generation model, or DIRSIG. The advantage to using synthetically created images is that the user is supplied with per-pixel ground truth which allows the algorithm performance to be evaluated at every pixel rather than at a few selected ground truthed sites. Then, using a factorial designed experimental approach, the original image will be changed according to the specified experimental factors. For this research, the factors of interest are the sensor view angle, the time of day, the amount of atmospheric visibility, and the size of the targets in the image. Varying these factors from their low level to their high level will result in $2^4 = 16$ hyperspectral images. These images will then be subjected to two hyperspectral anomaly detection algorithms, the RX Method and the BACON Method.

A second experimental design will be implemented when utilizing each anomaly detector. For example, varying the number of principal components retained and the size of the processing window in the RX Method will give a better understanding of how the algorithm's settings influence its performance on each synthetic image. Separate

analyses of each algorithm will then be accomplished in order to determine which main effects and interactions have the greatest affect on algorithm performance.

One common measure of performance for anomaly detectors is the utilization of true positive fractions and false positive fractions in the form of an operator characteristic (OC) curve. However, large false positive fractions are not appealing since they imply that many background pixels are being declared as outliers. Therefore, the OC curves in this thesis will only show the region of false positive fractions from 0 to 0.05. For each image-detector combination, the responses for the designed experiments will be the true positive fraction when the false positive fraction is fixed at 0.01. This will provide a measure of how the performance of each algorithm is affected by concurrently varying in-scene image parameters and anomaly detector parameters. This approach will also provide a performance comparison of the two algorithms. Since a factorial designed experimental approach will be adopted for altering the levels of the in-scene and algorithm parameters, we will also be able to observe how the joint effects of the parameters affect algorithm performance.

Preview

Chapter 2 contains a wealth of information providing background to this research. In this chapter, the hyperspectral anomaly detection algorithms that will be tested and the metrics employed will be discussed. Additionally, a more detailed look at how a factorial designed experiment is essential to appropriately study and analyze each of the in-scene parameters and their interaction on algorithm performance is illustrated. Chapter 3 outlines the approach and experimental design of this thesis. In Chapter 4, the results of

the experiments are presented with an in-depth discussion of their analysis. Chapter 5 provides a synopsis of the work completed in this thesis study and makes recommendations towards future work that is needed in the field of algorithm performance comparison and testing.

II. Literature Review

Before jumping right into the methodology and subsequent results of this thesis, it is first important to highlight some of the main concepts used in the research. First, a thorough look into the field of hyperspectral imaging will be carried out. Following that, the advantages and disadvantages of using synthetic imagery will be discussed. Also, this literature review will discuss the two hyperspectral anomaly detector algorithms that will be tested – the RX and BACON methods. After that, operating characteristic (OC) curves, as well as the useful information that can be derived from them will be discussed. Finally, we will investigate the advantages experimental designs provide in determining the variables that are the most influential on a given experimental response.

Hyperspectral Imaging

Electro-optical remote sensing utilizes the fact that the numerous objects in a scene are made up of materials that reflect, absorb, and emit electromagnetic radiation in ways attributable to their molecular composition and shape. The field of spectroscopy measures, analyzes, and interprets this radiation that arrives back at a sensor (Shaw and Manolakis, 2002: 12). If this radiation is measured over a band of contiguous wavelengths, the resulting spectral signature can be used to uniquely characterize and identify any given material.

Shaw and Burke briefly describe the four sampling operations involved in the collection of hyperspectral image data. They include, spatial sampling, spectral sampling, radiometric sampling, and temporal sampling. The spatial sampling resolution

of the sensor is equal to the ground sample distance (GSD). The GSD is equivalent to the area on the ground represented in one pixel of the spectral image. For example, if the GSD is 1 meter, then each pixel in the image represents a 1 meter by 1 meter square on the ground. In general, the GSD varies from a fraction of a meter to tens of meters. The GSD is primarily established by the sensor aperture and the altitude of the airborne platform. Spectral sampling is accomplished by decomposing the radiance received by the sensor in each spatial pixel into a number of wavebands. A prism and interferometer are two means of spectral sampling. An analog-to-digital (A/D) converter samples the radiance measured in each spectral channel and produces digital data at a recommended radiometric resolution (Shaw and Burke, 2003: 6).

Finally, temporal sampling refers to the process of collecting multiple spectral images on the same scene separated in time. In other words, temporal sampling is using an airborne sensor to collect data over the same geographic area in the span of hours, days, weeks, or even months (Shaw and Burke, 2003: 8). This is especially important for studying natural changes in a scene. We see that seasonal variations present great changes in the spectral character of a scene. A forest, for example, changes dramatically from one season to the next.

Hyperspectral sensors have been developed to sample the reflective portion of the electromagnetic spectrum that extends from the visible region (0.4-0.7 μm) through the near-infrared (about 2.4 μm) in hundreds of adjacent bands. The high spectral resolution feature of hyperspectral sensors preserves significant aspects of the spectrum. It is this fact that makes the differentiation of materials on the ground possible. Figure 1 illustrates how the spatial and spectral information is formed to make a data cube (Shaw

and Manolakis, 2002: 13). In general, it is the distinguishing features from the pixel spectra that provide the principal device for detecting and classifying materials in a scene.

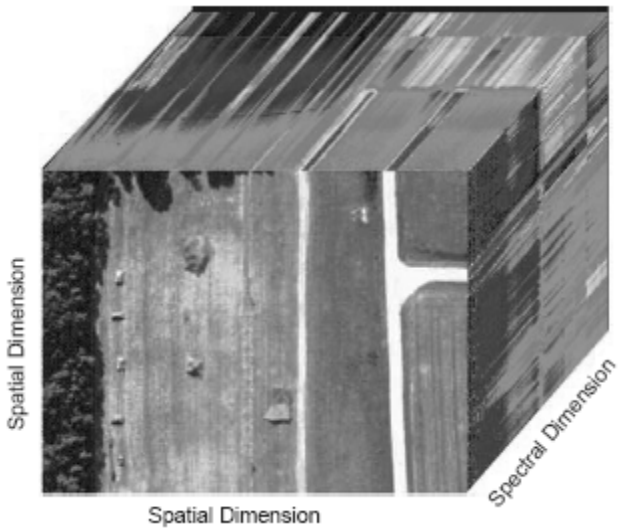


Figure 1. Hyperspectral Data Cube (Shaw and Manolakis, 2003: 13)

There are many environmental and sensor-related phenomena that can complicate the recovery of the reflectance spectra. Sensor resolution, atmospheric effects, spectral variability of a scene's surface materials, and other environmental and sensor effects are some of these issues. As a result, there are many issues that must be addressed when designing, implementing, and analyzing a spectral imaging sensor (Shaw and Burke, 2003: 8). The spatial resolution of the sensor and atmospheric absorption and scattering are the two most significant contributors to diminished image quality.

A sensor's spatial and spectral resolutions are matters that must be dealt with when designing a sensor. The cost of a sensor is highly dependent on the size of aperture being used. A smaller aperture reduces the cost of the sensor, but results in a degraded

spatial resolution, or a larger ground sample distance (GSD). In terms of the "goodness" of a spectral imager, the best detection performance is expected when the angular resolution of the sensor, which is specified in GSD, is proportionate with the size of the targets in the image (Shaw and Burke, 2003: 8). However, since targets vary in size, some targets may be fully resolved spatially while others may only fill a fraction of the GSD footprint. This leads to designing detection algorithms to perform well for both full pixel and sub-pixel targets.

Atmospheric absorption and scattering are also significant contributors to a diminished image quality. First, the atmosphere alters the spectrum of the solar illumination before it even reaches the ground. This must be known in order to separate the impinging solar radiance from the reflectance spectrum that characterizes the material of interest. Atmospheric gases, aerosols, and water vapor contribute greatly to the overall atmospheric transmission. Second, the atmosphere scatters some of the solar radiation into the field of view of the sensor without ever having reached the ground. This scattered light effects the reflected light traveling from the ground to the sensor and is called path radiance. Third, the solar radiation scattered by the atmosphere acts as another source of diffused colored illumination. Finally, the solar illumination that does reach the scene and is reflected back by the target of interest is further absorbed and scattered by the atmosphere on its way to the sensor (Shaw and Burke, 2003: 8-9).

A third concern when dealing with designing, implementing, and analyzing a spectral imaging sensor is spectral variability. The term "spectral signature" implies that there exists a unique association between a material and its reflectance spectrum. However, it has been observed in both laboratory and field data that there is some

variability in the spectrum of most materials. This may be due to unaccounted errors in the sensor, unaccounted atmospheric or environmental effects, material variation due to aging, and adjacency effects in which reflections from nearby objects alter the illumination of the material of interest (Shaw and Burke, 2003: 10). As stated earlier, we also see that seasonal variations present great changes in the spectral character of a scene. Figure 2 is an example of variability in reflectance spectra measured over multiple instances of vehicle paint in a scene. One can notice that the shapes of the spectra are fairly consistent, but the amplitudes vary greatly within the scene.

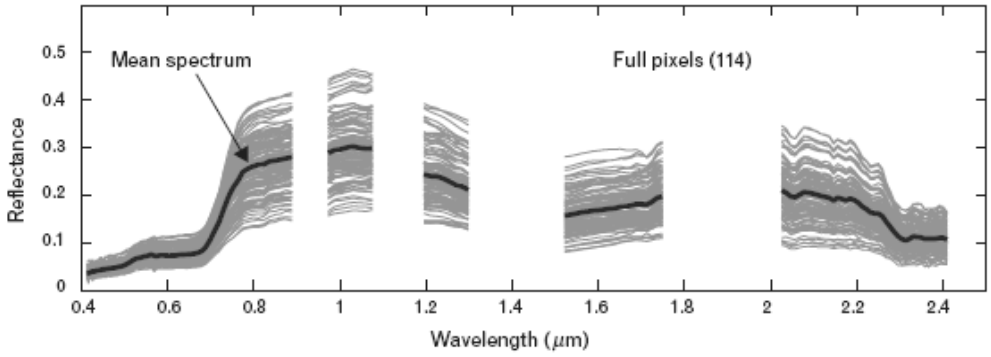


Figure 2. Example of Spectral Variability (Shaw and Burke, 2003: 10)

There also exist many other environmental and sensor effects that can make the recovery of the reflectance spectra difficult. In addition to absorption and scattering, several other environmental phenomena influence spectral imaging. The amount of light reflected into the sensor field of view is affected by the sun angle relative to zenith, the sensor viewing angle, and the target's surface orientation. Clouds and ground cover, such as trees, may change the illumination of a material by casting shadows on targets. Also,

nearby objects may reflect or scatter sunlight onto the target and thus change the overall illumination of the target. Figure 3 illustrates some of the many atmospheric and scene-related factors that can contribute to degradations in the imaging process.

As mentioned, spectral imaging sensors take advantage of the forward linear motion of the sensor platform to scan a scene. Any nonlinear motion of the sensor though, can also degrade the spectral image by mixing spectral returns from different parts of the image. In addition, the motion of targets in the scene can create artifacts in images. Also, the actual operation of the sensor may create artifacts and noise in the data (Shaw and Burke, 2003: 8-10).

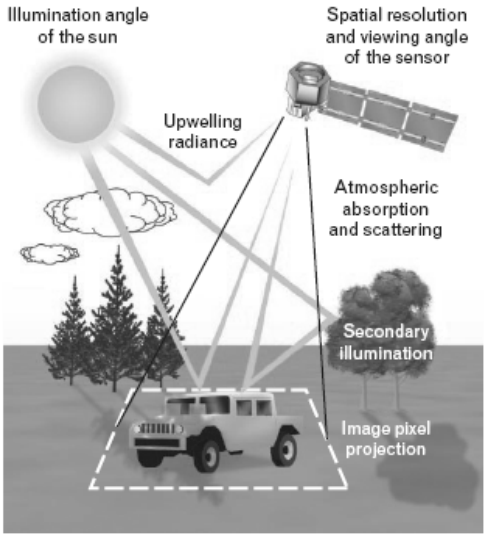


Figure 3. Contributors to Image Degradation (Shaw and Manolakis, 2002: 14)

Synthetic Imagery

Synthetic image generation (SIG) models are powerful tools used for the study of the image chain (Schott, 1997: 363). Rochester Institute of Technology's Digital Image

and Remote Sensing laboratory has developed the Digital Imaging and Remote Sensing Image Generation model, or DIRSIG, to assist the remote sensing community in several different application areas.

Along with instrument prototyping and algorithm training, DIRSIG can be utilized to help in the testing of algorithms. After the design and implementation of an algorithm comes the often difficult task of testing and evaluating the performance of the algorithm. Traditionally, algorithms have been tested using sets of extensively ground truthed test images. There are several limitations to this approach. The first is the availability of well characterized test data sets. For example, what if a data set is not available that contains a specific target of interest? Another restriction is the diversity in available data sets (DIRS, 2006: 11). Is the algorithm robust if the performance can be verified on only a single scene type (forested, urban, etc.)? Or under a single set of atmospheric conditions? Or for a single time of day? When a synthetic image generation model is being utilized, the user can create a large set of test images that feature variations such as scene type, time of day, and atmospheric conditions so that the algorithm can be assessed under more conditions.

SIG models look to effectively model nature and generate synthetic images that mimic real imagery. One of the many groups interested in SIG models are algorithm developers (Schott, 1997: 364). The interest is in developing and testing algorithms on scenes that contain a target of interest in a variety of forms and over a range of acquisition conditions. The lack of real well ground truthed imagery motivates this interest.

To date, efforts have focused on using synthetic images and the corresponding truth to support the tuning and testing of algorithms. Many complex hyperspectral algorithms have several modifiable parameters, such as weights and thresholds, which can be adjusted to improve performance. However, when working with real data, it is difficult to know whether an adjustment is having the desired impact (Schott, 2000: 23). An example of this may be if, after making a change to a parameter, two possible targets become more detectable while another target becomes less detectable. Has the algorithm improved? When using SIG models, the user knows exactly what is in an image and can then tell quantitatively what is in each pixel and where the errors are being made.

The use of synthetic data sets provides the user with many advantages. The primary reason for the use of synthetic imagery in image chain analysis is that all of the details of the constructed image are known. These details include the geometry of the scene and the spatial relationships of objects in the scene. SIGs supply the user with per-pixel ground truth which allows the algorithm performance to be evaluated at every pixel rather than at a few selected ground truthed sites. Also, the cost and time savings of producing synthetic data over extensive field collection campaigns is significant (DIRS, 2006: 11).

The DIRSIG image used as the base model for this thesis is the Western Rainbow scene. It is relatively homogeneous in the sense that the majority of the image is either normal vegetation or a mixture of grass and dirt. There are, however, several deciduous trees and bushes, as well as roughly 30 military targets (tanks, missile carriers, and trucks), scattered throughout the image. The image is 250 by 250 pixels in size with a ground sample distance of approximately 2 meters. A framing array sensor with an

AVIRIS spectral range of 0.4 to 2.5 μm was utilized at an altitude of 15,000 feet. The DIRSIG Western Rainbow image shown in Figure 4 was used in this thesis.

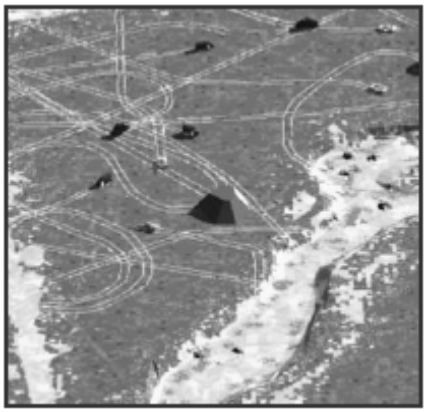


Figure 4. DIRSIG Western Rainbow Scene

The state-of-the-art image generation models are a very powerful tool in helping to visualize the image chain. However, with all of its benefits, it should not be perceived as a substitution to real imagery. Mirroring all aspects of the real world is an impossible task. The SIG models should be treated as tools that approximate the process, but not fully represent it. Synthetic image generation models can aid in designing, analyzing, and sometimes reducing the extent of field studies, but it cannot replace them.

Anomaly Detection Algorithms

Shaw and Manolakis describe detection as "the process of identifying the existence or occurrence of a condition." In the sense of anomaly detection, the problem becomes the existence or non-existence of a target in a given image pixel. Usually, the output of an anomaly detection algorithm is an OC curve signifying the true positive and

false positive fractions for the detector. A true positive refers to an outlier pixel being correctly classified as an outlier. A false positive refers to a pixel being incorrectly identified as an outlier pixel. The following paragraphs outline the two hyperspectral anomaly detection algorithms that will be used in this thesis. The detectors are the RX Method and the BACON Method.

RX Method

The RX Method employed in this thesis is a local anomaly detector developed by Reed and Yu. It is a local detector in the sense that it passes a user-defined processing window over every pixel in an image to find anomalies. The pixels enclosed inside the processing window are used to characterize the local background. The pixel at the center of the processing window, *x*, is then tested relative to the background to establish if it is anomalous. A test statistic is computed for each pixel using the following equation:

$$RX(x) = (x - \hat{\mu})^T \left(\frac{N}{N+1} C + \frac{1}{N+1} (x - \hat{\mu})(x - \hat{\mu})^T \right)^{-1} (x - \hat{\mu})$$

where,

$\hat{\mu}$ = the mean vector of the processing window

C = the covariance matrix of the processing window

N = the number of pixels in the processing window.

Thus, to determine if *x* is an anomalous pixel in the image, *RX(x)* can be compared to an appropriate quantile of the χ^2 distribution with *p* degrees of freedom, where *p* is the dimensionality of the data (Smetek and Bauer, 2006: 2).

BACON Method

The Blocked Adaptive Computationally Efficient Outlier Nominators (BACON) Method employed in this thesis was developed by Billor, Hadi, and Velleman. The algorithm for multivariate data begins by identifying an initial basic subset of $m > p$ observations that can safely be assumed free of outliers, where p is the dimension of the data and m is an integer chosen by the user. Discrepancies are then computed for each observation using the equation:

$$d_i(\bar{x}_b, C_b) = \sqrt{(x_i - \bar{x}_b)^T C_b^{-1} (x_i - \bar{x}_b)}, \qquad \text{i = 1, 2,…, } n,$$

where,

$\bar{x}_b$ = the mean vector of the observations in the basic subset b

C_b = the covariance matrix of the observations in the basic subset b.

A new basic subset to all points is then set with discrepancies less than $c_{npr}\chi_{p,\alpha/n}$, where, $\chi^2_{p,\alpha}$ is the 1 - α percentile of the chi square distribution with p degrees of freedom, $c_{npr} = c_{np} + c_{hr}$ is a correction factor, $c_{hr} = \max\{0, (h-r)/(h+r)\}$, $h = [(n+p+1)/2]$, r is the size of the current basic subset, and $c_{np} = 1 + \frac{p+1}{n-p} + \frac{1}{n-h-p} = 1 + \frac{p+1}{n-p} + \frac{2}{n-1-3h}$.

The algorithm iterates until the size of the basic subset no longer changes. The observations excluded from the final basic subset are nominated as outliers (Billor, Hadi, and Velleman, 2000: 286).

OC Curves

Results for anomaly detection algorithms can be summarized using operator characteristic (OC) curves. The OC curve describes the relationship between the probability of a true detection and the probability of a false alarm. A true detection, also known as a true positive, is a correct detection of a target in a pixel. A false alarm, or false positive, is when the algorithm claims that a target is present in a pixel, when in fact, there is not one. Spectral complexity of an image, the similarity of the target to the background, and the size of the target are some of the parameters that affect the performance of anomaly detection algorithms. The ideal algorithm would approach a 100% detection rate and a 0% false alarm rate. Figure 5 contrasts OC curves for the performance of the RX Method on Synthetic Image 3 as the size of the processing window changes along with the number of principal components retained for data reduction.

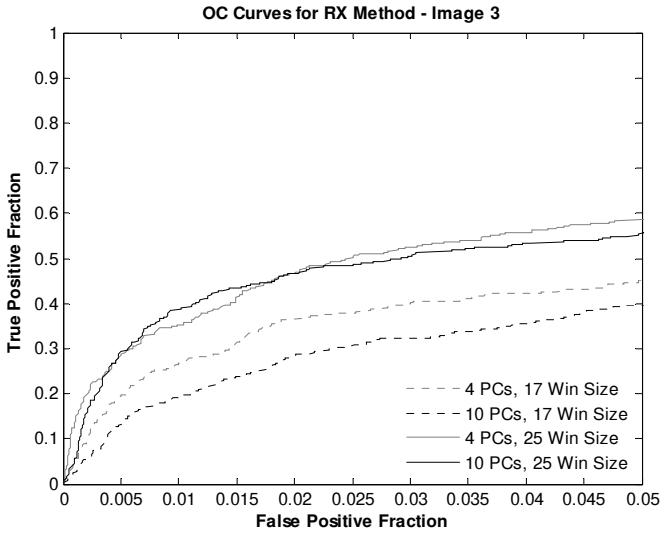


Figure 5. OC Curves for RX Method on Synthetic Image 3

To create an OC curve, the user must have access to the truth data that defines where the targets are in the imagery. For real imagery, the truth maps are constructed using extensive field collected ground truth measurements which are then mapped to the pixels in the scene. Due to the natural uncertainty of the target location, the detection algorithm can only be performed on a per-target basis. As was stated earlier, the use of synthetic image generators provides a per-pixel ground truth. Therefore, the location of every target in the image is known with 100% certainty and a pixel-based OC curve can be generated (Ientillucci and Brown, 2003:119-120).

Experimental Design

It has been recognized that in-scene parameters, such as the sensor view angle, the time of day, the atmospheric conditions, and the size of the targets hinder the ability of hyperspectral anomaly detection algorithms to find anomalies in a hyperspectral image. It has also been established from a review of the literature, that algorithm developers only test their algorithms on a minimal number of images. For the robustness of an algorithm to be determined, the algorithm should be tested against several images with varying factors. A factorial designed experiment is essential to appropriately study and analyze each of these in-scene parameters and their interaction on algorithm performance. Montgomery, as well as Wackerly, Mendenhall, and Scheaffer, fully detail the designing of experiments in their respective texts. For purposes of this thesis, a simplified explanation of factorial designed experiments in presented in the following paragraphs.

In experiments concerning several factors where it is essential to study the joint effect of the factors on a response, experimentalists turn to factorial designs because of

their efficiency. A factorial design is one in which a complete replication of the experiment contains all possible combinations of the levels of the investigated factors (Montgomery, 2005: 160). In a two factor experiment for example, if factor A has *a* levels and factor B has *b* levels, the experiment would consist of *ab* treatment combinations, or experimental runs. In general, an L^k factorial design is one in which we have an experiment with L levels of k factors. The advantage of using factorial designs can be easily demonstrated.

For the purposes of illustrating a factorial design, we will use a simple experiment involving three factors A, B, and C, each with two levels. We call these levels "low" and "high" and denote them "-" and "+," respectively. These levels may be quantitative, qualitative, or even the presence or absence of a factor. An experimental design such as this is known as the 2^3 factorial design. Thus, we will have a total of $2^3 = 8$ experimental runs. The eight runs are shown geometrically in Figure 6.

The eight treatment combinations can also be represented by lowercase letters, also shown in Figure 6. We can see from the figure that in a treatment combination, a lowercase letter represents the high level of that particular factor. The absence of a certain letter in that treatment corresponds to that factor at its low level. For example, the treatment combination *ac* denotes factors A and C at their high levels and factor B at its low level. The notation (1) depicts all factors at their low levels (Montgomery, 2005: 211-212). Table 1 lists the eight experimental runs of the 2^3 design in a design matrix.

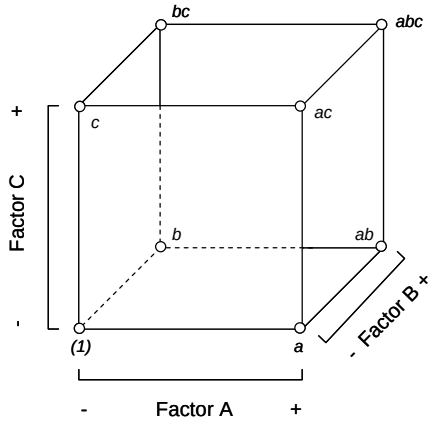


Figure 6. Geometric View of the 2^3 Factorial Design

Table 1. The 2^3 Factorial Design Matrix

Trial	Treatment Combination	Factor A	Factor B	Factor C
1	(1)	-	-	-
2	a	+	-	-
3	b	-	+	-
4	ab	+	+	-
5	c	-	-	+
6	ac	+	-	+
7	bc	-	+	+
8	abc	+	+	+

At this point, the experimental runs are ready to be conducted in a random order. Once the runs have been completed and the data have been collected, a statistical analysis of the data is necessary. Most of the time, this involves statistical procedures included in the analysis of variance, or ANOVA. The ANOVA procedure attempts to analyze the variation in a set of responses and assigns portions of this variation to the main effects, interaction terms, and experimental error of the design. The objective of ANOVA is to

locate the important factors and interactions and determine how they affect the response (Wackerly 2002).

In general, the variability of a set of n measurements is proportional to the sum of squares of deviations, $\sum_{i=1}^{n}(y_i - \bar{y})^2$, where y_i is an experimental observation or response and $\bar{y}$ is the overall mean of the observations. This quantity is then divided by $n - 1$ in order to calculate the sample variance. As before, ANOVA partitions this sum of squares of deviations, or the Total Sum of Squares (SS_T), into parts attributed by each of the main effects and interaction effects, plus a remainder associated with random experimental error (Wackerly, 2002: 629-630).

In order to determine the Sum of Squares for the ANOVA, we must estimate the main and interaction effects using contrasts. A contrast is also called the total effect of a factor or interaction of factors. The contrast for each main factor and interaction effect can be found in Table 2. For the 2^3 design, the Sum of Squares for each factor and interaction is then found by dividing each contrast by $2^3n = 8n$ (Montgomery, 2005: 215). The Sum of Squares for Error (SS_E) can be determined by subtraction, since the Total Sum of Squares is partitioned as $SS_T = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_E$.

The 2^3 design in n replications contains $2^3n - 1$ degrees of freedom between the eight treatment combinations. The main effects A, B, and C and the interaction effects AB, AC, BC, and ABC are each associated with one degree of freedom. The remaining $2^3(n - 1)$ degrees of freedom are associated with the experimental error.

Table 2. Contrasts for Main Effects and Interaction Effects

Factor or Interaction	Contrast
A	$a + ab + ac + abc - [(1) + b + c + bc]$
B	$b + ab + bc + abc - [(1) + a + c + ac]$
C	$c + ac + bc + abc - [(1) + a + b + ab]$
AB	$ab + abc + c + (1) - [a + b + ac + bc]$
AC	$ac + abc + b + (1) - [a + c + ab + bc]$
BC	$bc + abc + a + (1) - [b + c + ab + ac]$
ABC	$a + b + c + abc - [(1) + ab + ac + bc]$

Given that the Sum of Squares and the degrees of freedom for each main effect and interaction effect have been computed, the ANOVA table can now be completed. The ANOVA table for n replicates of the 2^3 design is shown in Table 3. The calculated ANOVA values allow the experimenter to test the observed results of each factor for statistical significance by means of an "F-test". That is, the experimenter can see which factors and interactions make a significant impact on the response variable of the experiment based on the magnitude and statistical hypothesis testing of F_0. The null hypothesis being tested is that a given factor or interaction between factors is not significant in the design. The alternative hypothesis is that the given factor or interaction between the factors is significant in the design. To test these hypotheses, F_0 is compared to an F test statistic with a specified degree of significance, α. Therefore, $F_0 < F_{\alpha,1,8(n-1)}$ implies that the factor of interest is not significant in the design. On the other hand, if $F_0 > F_{\alpha,1,8(n-1)}$, the experimenter concludes that the factor of interest is significant.

Table 3. The ANOVA Table for the 2^3 Factorial Design in *n* Replications

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_0
Factor A	SS_A	1	$MS_A = SS_A$	$F_0 = \frac{MS_A}{MS_E}$
Factor B	SS_B	1	$MS_B = SS_B$	$F_0 = \frac{MS_B}{MS_E}$
Factor C	SS_C	1	$MS_C = SS_C$	$F_0 = \frac{MS_B}{MS_E}$
AB Interaction	SS_{AB}	1	$MS_{AB} = SS_{AB}$	$F_0 = \frac{MS_{AB}}{MS_E}$
AC Interaction	SS_{AC}	1	$MS_{AC} = SS_{AC}$	$F_0 = \frac{MS_{AB}}{MS_E}$
BC Interaction	SS_{BC}	1	$MS_{BC} = SS_{BC}$	$F_0 = \frac{MS_{AB}}{MS_E}$
ABC Interaction	SS_{ABC}	1	$MS_{ABC} = SS_{ABC}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	8(n - 1)	$MS_E = \frac{SS_E}{8(n-1)}$	
Total	SS_T	8n - 1		

On occasion, available resources will only allow a single replicate of the design to be run. The risk in performing an experiment with only one run at each test combination is that, if the response is highly variable, then misleading conclusions may result. This results in fitting the model to noise. The reader can also see from Table 3 that only one replication of the design ($n = 1$) leads to zero degrees of freedom for error. Thus, there is no internal estimate for error. However, Montgomery appeals to the sparsity of effects principle. That is to say, most systems are dominated by some of the main effects and low-order interactions. The higher order interactions can be deemed negligible. Therefore, the negligible effects can be pooled as an estimate of error (Montgomery, 2005: 211-224).

The primary advantage of the experimental design approach is that it allows an examination of both the main factors and their interactions with respect to a response variable. Clearly, implementing a factorial designed experiment will help achieve the main objective of this thesis. The experimental approach will permit a thorough examination of the accuracy and effectiveness of hyperspectral target detection algorithms under differing in-scene factors. Also, it will allow us to observe the joint effects of these in-scene parameters with respect to algorithm performance.

III. Methodology

Overview of the Approach

Before exploring the approach and algorithms used in this thesis study, it is important to first review our main objective. The primary goal of this thesis is to examine the accuracy and effectiveness of hyperspectral anomaly detection algorithms under different values of sensor view angle, presence of shadows in the image, the amount of haze in the atmosphere, and the size of the targets in the scene. This will involve the use of synthetic hyperspectral images created by DIRSIG and two hyperspectral anomaly detector algorithms. The algorithms that will be used and tested are the RX method and the BACON method. Chapter 2 discusses these algorithms significantly.

Chapter 2 also details the algorithm evaluation metric that will be used in this thesis. The true positive fraction and the false positive fractions of each detector will be utilized to create an OC curve for each image-detector combination. These rates can be determined since synthetic imagery provides per-pixel ground truth and evaluation at every pixel in the scene is made possible. The true positive fraction when the false positive fraction is fixed at 0.01 will then be the response in the factorial experiments. Factorial designed experiments give us the ability to simultaneously analyze the main and joint effects of the in-scene parameters with respect to hyperspectral algorithm performance. The basic approach of this thesis can be best represented schematically as the flowchart in Figure 7.

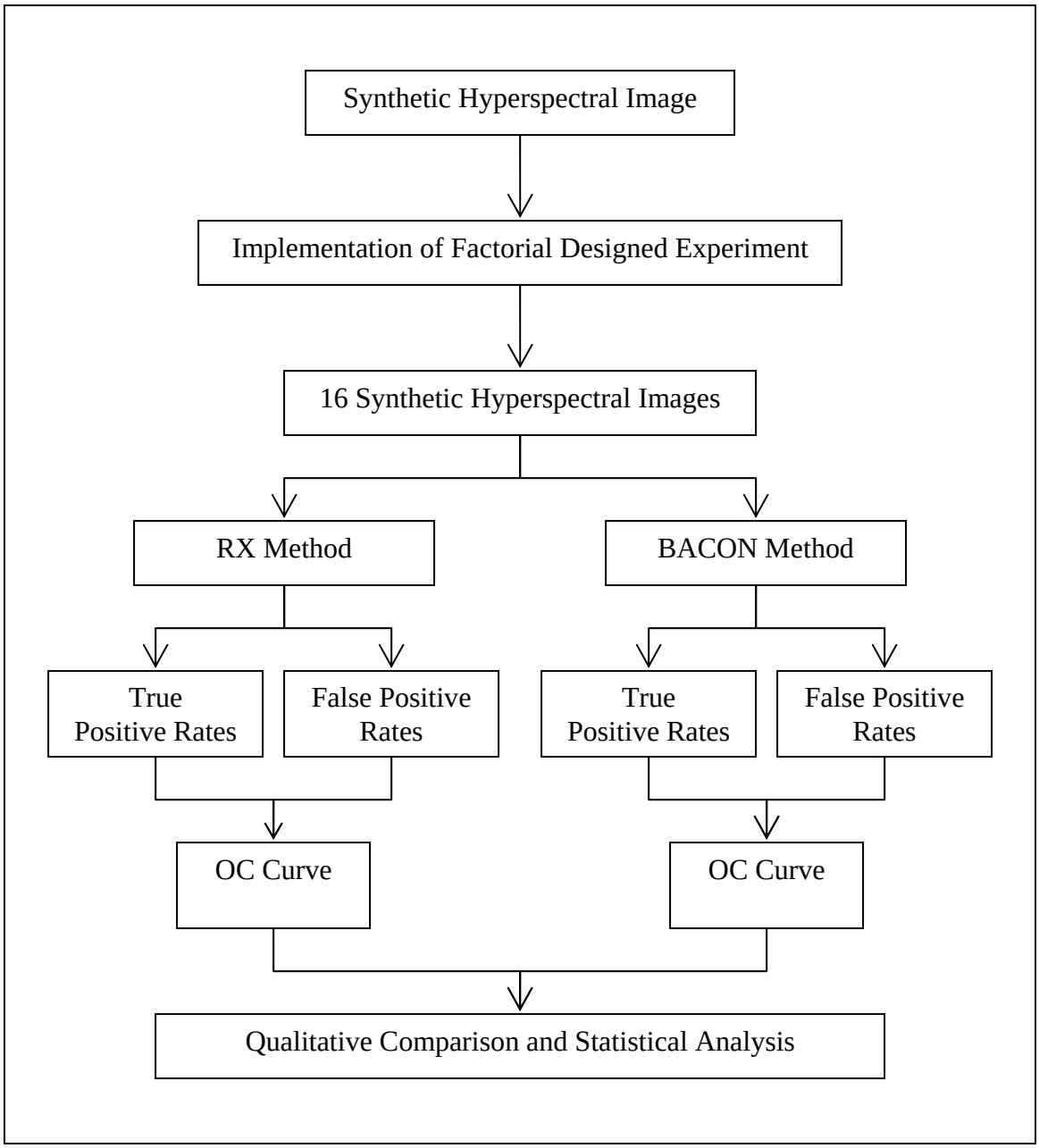


Figure 7. Flowchart of Thesis Study Approach

As seen in Figure 7, we begin with the DIRSIG Western Rainbow scene described in Chapter 2. This synthetic hyperspectral image is characterized by a framing array hyperspectral sensor positioned directly over the center of the scene at noon in the middle

of June. The targets are full scale and there is 23 km of atmospheric visibility. Ground truth is available for each pixel in the synthetic image. A 2^4 full factorial designed experiment will then be run to create the remaining 15 experimental synthetic images at the specified levels of all factors. These images, along with the initial synthetic image we began with, now become the 16 design points of our experiment.

From here, each design point will be run through the two hyperspectral anomaly detectors to acquire the experimental responses. To get these responses though, we will employ another factorial designed experiment for each detector. This time, a factorial experiment will be conducted on each anomaly detector with two algorithm-specific parameters being varied at two levels. This experiment will allow us to determine optimal settings to maximize each algorithm's performance. The result of this experiment will be an OC curve for each image-detector-parameter permutation. The false positive fraction will then be fixed at 0.01 and the corresponding true positive fractions will be calculated and become the responses of each experiment.

Analysis of these results can be achieved by employing statistical techniques such as ANOVA and plotting the results to establish trends visually. The analysis will reveal the degree of significance that the main effects and interactions of the sensor view angle, presence of shadows in the image, the atmospheric visibility, and the size of the targets in the scene have on anomaly detector performance. It is from these results and the corresponding analysis that we will also be able to determine the optimal parameter settings for each algorithm. Even more, using a nested factorial design, we will be able to determine if the algorithm used is a significant factor and which, if any, algorithm performed the best under the given levels of our experimental factors.

Experimental Factors and Factor Levels for Creating Synthetic Images

As stated previously, the intention of this thesis is to examine the accuracy and effectiveness of hyperspectral anomaly detection algorithms under different values of sensor view angle, shadows, atmospheric visibility, and the size of the targets in the scene. These are four factors that are believed to affect the performance of anomaly detectors.

Sensor View Angle

By varying the viewing angle of the hyperspectral sensor, more of the geometry and terrain of the actual image comes into play. Hills and valleys in the terrain that, when viewed from directly overhead are not as prominent in the image, now stand out and influence the path of the reflected or emitted electromagnetic energy back to the sensor.

Table 4 shows the low and high factorial levels of sensor view angle for the purpose of this thesis. A 0° degree view angle corresponds to the framing array sensor being positioned directly over the center of the image looking straight down. Figure 8 shows the graphical depiction of 0°, 20°, and 40° sensor view angles. When varying the view angle of the sensor, the distance from the sensor to the center of the image was held at a constant 15,000 feet. This allows the sensor to "look through" the same distance of atmosphere for each view angle setting. As a result, the altitude of the sensor varies with the view angle.

Table 4. Factor Levels for Sensor View Angle

Factor Level	*Sensor View Angle*
Low (-1)	0°
High (1)	40°

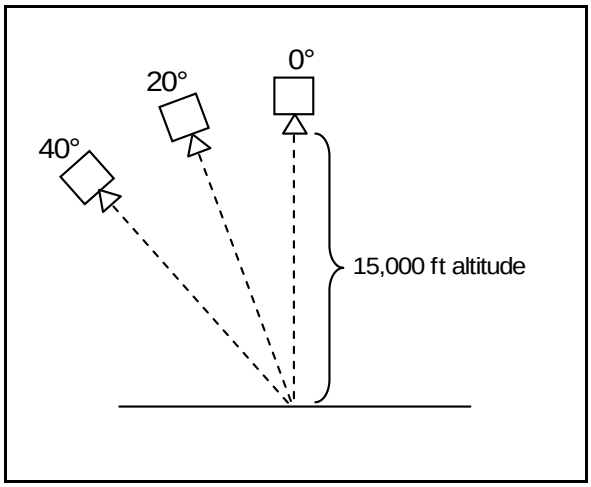


Figure 8. Graphical Depiction of Factor Levels for Sensor View Angle

Presence of Shadows in the Scene

The presence of shadows in a scene is one of the many environmental effects that can make the recovery of the reflectance spectra difficult. Clouds and ground cover, such as trees, may change the illumination of a material by casting shadows on targets. Also, nearby objects may reflect or scatter sunlight onto the target and thus change the overall illumination of the target. Shadows also influence the spectral variability of a material in a scene. Figure 3 illustrates how shadows and other unknown atmospheric and scene-related factors can contribute to degradations in the imaging process by having an effect on the spectral variability of a material.

Since the location of the sun determines where and how much shading appears in a scene, the controllable factor that will be used to vary the amount of shadows in a scene is the time of day. Obviously, there are more shadows cast during the morning and evening hours than in the middle of the afternoon. The simulation time in DIRSIG will be set to 15 June 2005 in order to replicate a day in which the sun is approximately directly overhead at noon time. Thus, to simulate a minimum level of shadows in the scene, the low level for time of day will be 1200. Alternatively, the time of day will be set to 1800 to simulate a high level of shadows in the scene. Table 5 shows the low and high factorial levels for the experimental factor Time of Day.

Table 5. Factor Levels for Time of Day

Factor Level	*Time of Day*
Low (-1)	1200
High (1)	1800

Atmospheric Haze Levels

Haze in the atmosphere is another environmental effect that can make detecting and identifying targets in a hyperspectral image difficult. Atmospheric gases, aerosols, and water vapor contribute greatly to the overall atmospheric transmission. Gases and particles in the atmosphere alter the spectrum of the solar illumination before it even reaches the ground. Also, the solar illumination that does reach the scene and is reflected back by the target of interest is further absorbed and scattered by the atmosphere on its way to the sensor. In DIRSIG, the user can regulate the atmospheric visibility in a scene. The low and high factor levels for atmospheric visibility are listed in Table 6.

Table 6. Factor Levels for Atmospheric Haze

Factor Level	*Atmospheric Haze*
Low (-1)	5 km visibility
High (1)	23 km visibility

Size of Targets in a Scene

The size of the targets in the scene will be varied from full scale targets down to half scale targets. A typical full scale target in the DIRSIG Western Rainbow scene is contained in roughly 15-16 pixels. When scaled by half, that same target is contained in about 4-5 pixels. Figure 9 compares several full scale targets (A and C) from Synthetic Image 1 to those same targets at half scale (B and D) from Synthetic Image 2. The figures are subsets of the DIRSIG generated truth images.

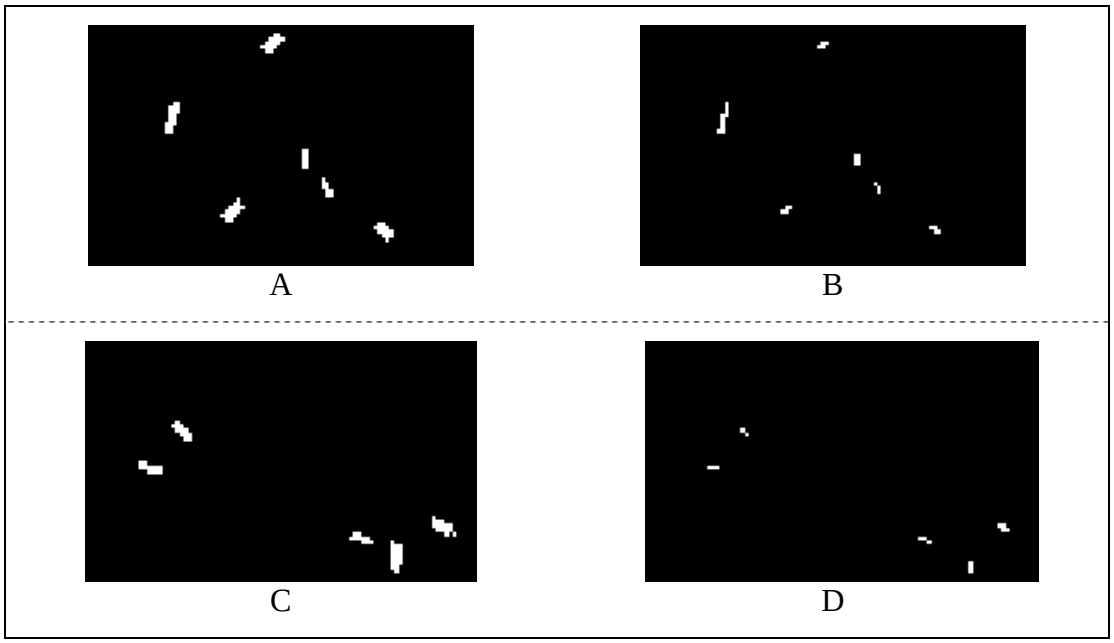


Figure 9. Comparison of Full Scale and Half Scale Targets

The size of the targets was chosen as a controllable factor because it is known that certain anomaly or target detection algorithms perform poorly with large in-scene targets. The RX Method for instance, does not perform well if a target is large relative to the size of the image processing window. The low and high factor levels for target size are shown in Table 7.

Table 7. Factor Levels for Size of Targets

Factor Level	*Scale of Targets*
Low (-1)	1/2
High (1)	1

2^4 Factorial Designed Experiment for Creating Synthetic Images

In order to determine if the main effects and interactions between the experimental factors contribute significantly to the performance of an anomaly detector, a factorial designed experiment will be used to create a family of synthetic images for which to test algorithms. Instead of testing an algorithm on one or two images, the experimental hypothesis of this thesis is that testing an algorithm against many images will aid in determining the robustness of the algorithm being investigated. A factorial designed experiment is essential to appropriately study and analyze each of these in-scene parameters and their interactions on algorithm performance. Thus, a 2^4 full factorial experiment will be run to create the 16 synthetic hyperspectral images needed for this study. The design points for the experiment are listed in Table 8 and a graphical depiction is shown in Figure 10.

Table 8. 2^4 Factorial Experiment for Creating Synthetic Images

Image	*A: View Angle*	*B: Time of Day*	*C: Atmospheric Visibility*	*D: Scale of Targets*
2	-	-	-	-
9	+	-	-	-
13	-	+	-	-
12	+	+	-	-
15	-	-	+	-
11	+	-	+	-
16	-	+	+	-
8	+	+	+	-
1	-	-	-	+
5	+	-	-	+
4	-	+	-	+
6	+	+	-	+
3	-	-	+	+
10	+	-	+	+
14	-	+	+	+
7	+	+	+	+

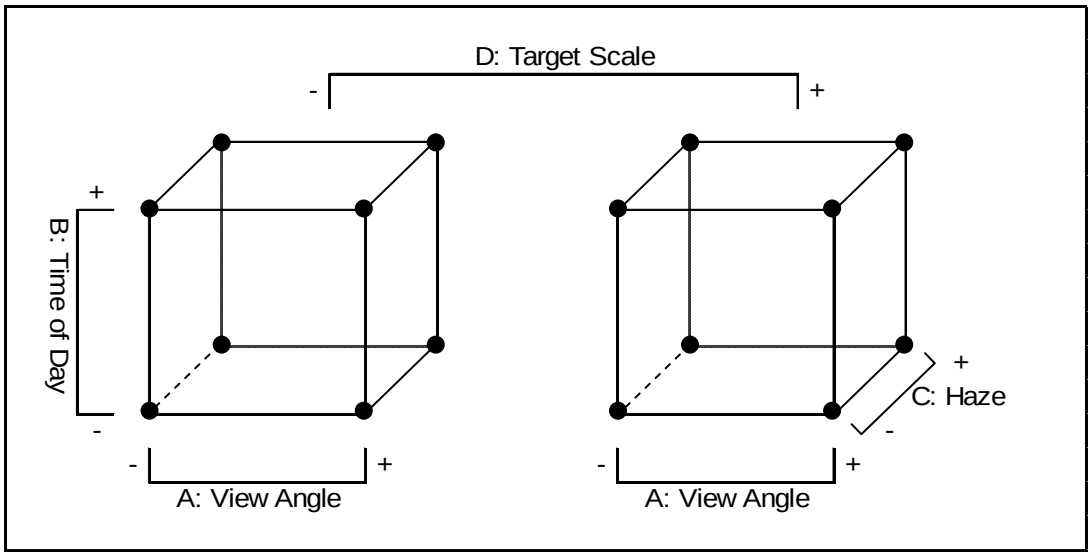


Figure 10. Graphical Depiction of the 2^4 Factorial Experiment

Experimental Factors and Factor Levels for Algorithm Testing

In the hyperspectral community, there are several types of image processing algorithms. Anomaly detectors, target detectors, and unmixing algorithms are just a few. For the purpose of this thesis, however, the focus will be on comparing the two different anomaly detectors detailed in Chapter 2, the RX Method and the BACON Method. Like many complex hyperspectral algorithms, the RX Method and the BACON Method are utilized by allowing the user to determine certain thresholds and weights for modifiable parameters in order to improve performance. It is unclear when working with real data, though, whether an adjustment to a parameter is having the desired impact. For example, after making a change to a parameter, two possible targets become more detectable while another target becomes less detectable. The uncertainty lies in determining whether the algorithm has improved or not. This leads to the main reason for implementing the use of synthetic imagery in this thesis study. With the material in every pixel being known with 100% certainty, all errors in detection will be able to be determined. In order to test the RX and BACON methods, additional factorial experiments will be developed to investigate how different levels of parameters for each algorithm affect their performances on the synthetic images.

Factorial Design, Factors, and Factor Levels for the RX Method

The RX Method contains several parameters that can be set by the analyst. This study will focus on how the size of the processing window and the number of retained principal components influence the performance of the RX anomaly detector. To examine this, a 2^2 factorial experiment is designed in which each of the two factors is

tested at two levels. The two factor levels for each of the experimental factors is shown in Table 9. Figure 11 depicts a graphical representation of the 2^2 factorial design.

Table 9. The 2^2 Factorial Design for the RX Method

Factor Level	*PCs Retained*	*Window Size*
Low (-1)	4	17
High (1)	10	25

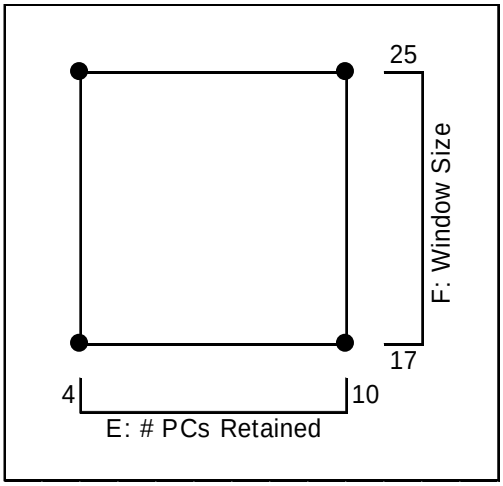


Figure 11. Graphical Representation of the 2^2 Factorial Design for the RX Method

The size of the processing window is an important factor in the RX algorithm. There is no rule stating that one window size works best because every image is different. A large processing window can contain too many anomalous pixels and may not correctly identify the "pixel of interest" as an outlier if, in fact, it is one. A small processing window, as stated in Chapter 2, does not perform well for multiple-pixel targets. Thus, analyzing an image using several window sizes will aid in the

investigation of how different levels of the processing window size affect the RX algorithm performance.

The RX Method also uses Principal Component Analysis to compress the image data into a few major principal components specified by the eigenvectors of the processing window covariance matrix. It is up to the analyst to decide how many principal components to use. Again, there is no rule stating how many principal components should be retained. As a result, like with the processing window size, we will experiment by retaining the top 4 and 10 principal components.

Factorial Design, Factors, and Factor Levels for the BACON Method

This study will focus on how different pre-processing data techniques, namely Band Aggregation and Clustering, can influence the performance of the BACON anomaly detector. A 2^2 factorial experiment will aid in this study. Principal Component Analysis, as stated earlier, will be used to compress the image data into a few major principal components. For this thesis, the top 10 principal components will be used to reduce the dimensionality of the image data prior to running BACON.

Band aggregation is a data reduction technique that averages every *n* bands to trim down the number of hyperspectral channels that defines an image. The DIRSIG synthetic images used in this thesis contain 224 hyperspectral channels. When atmospheric absorption bands have been removed, the image data still contains 163 channels. By aggregating every 10 bands, the dimension of the image can be greatly reduced. Figure 12 illustrates band aggregation on a sample pixel spectrum. The two

levels for this experimental factor are “Do Not Aggregate” and “Aggregate.” When band aggregation is used, every 10 band increment will be averaged.

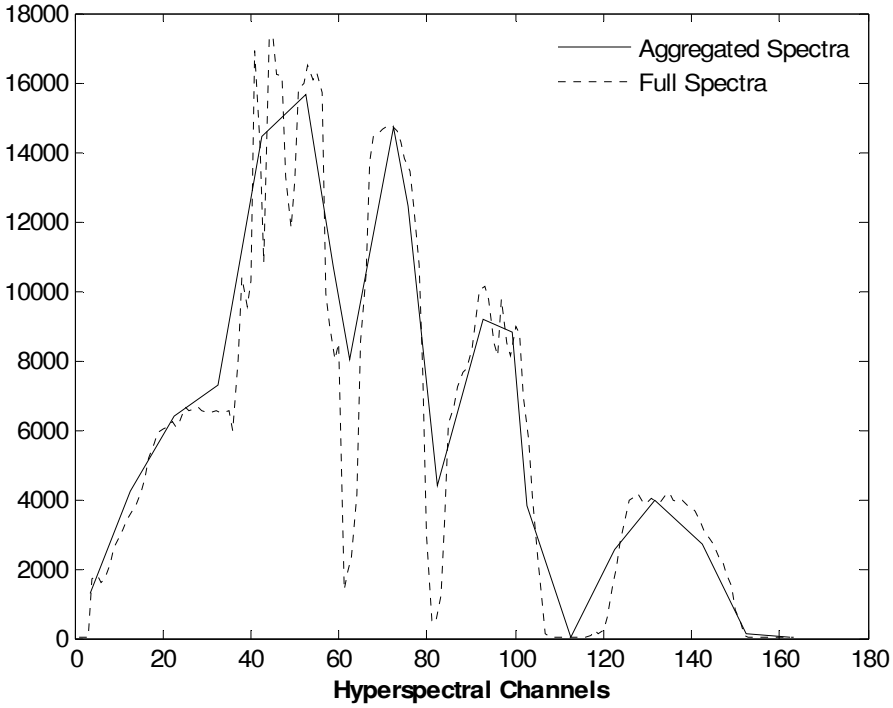


Figure 12. An Example of Band Aggregation

Prior to implementing BACON, MATLAB’s *k-means* function is utilized to partition the data set into *k* clusters. The value of *k*, which is determined by the user, will be the experimental factor for clustering in our study of BACON. The number of clusters detected in the scene can affect the performance of BACON. In the DIRSIG Western Rainbow scene, there can be 3 to 4 main clusters of pixels depending on whether trees and bushes can be differentiated from a grass field. The other two clusters are a road and a dirt-grass area. Thus, the number of clusters *kmeans* detects will be varied from 3 to 4

clusters. Figure 13 shows the graphical representation of the 2^2 factorial design used to study the BACON Method.

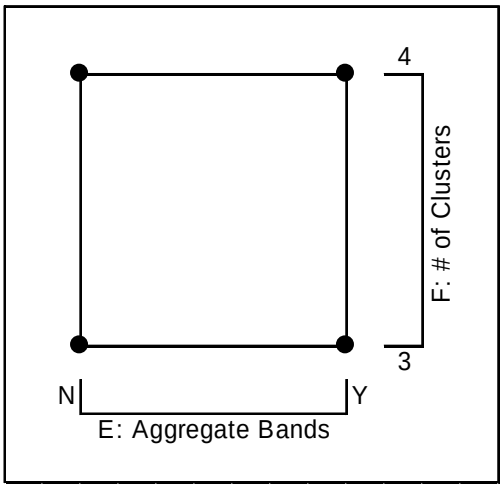


Figure 13. Representation of the 2^2 Factorial Design for the BACON Method

Analysis Plan for the Collected Data

At this point we will have created a family of synthetic hyperspectral images as proposed by the initial factorial designed experiment. Also, we will have processed these images through both the RX and BACON anomaly detectors based on two more factorial designed experiment using two algorithm-specific parameters as factors. The metric we are interested in is the true positive fraction of the algorithm when the false positive fraction is fixed at 0.01. This metric is a measure of the performance of each algorithm based on the levels of the in-scene parameters and algorithm-specific parameters. A statistical analysis is now required in order to examine the accuracy and effectiveness of the hyperspectral anomaly detection algorithms under different values of sensor view angle, time of day, atmospheric visibility, and the size of the targets in the scene.

One of the principal assumptions made in this thesis concerns the statistical analysis of the data using the ANOVA technique. As seen in Chapter 2, the application of ANOVA to the results from a factorial designed experiment allows the user to statistically test the significance of all main factors and their interactions. However, closer inspection of Table 3 reveals that it is necessary for at least two replications ($n \geq 2$) of the experiment be performed in order to properly calculate an error sum of squares, SS_E. In this thesis, though, only one run of the experiment is conducted since there will not be any variability in the response. It does not make sense to perform replicates of any portion of this experiment since the results will not change from run to run given that the factor levels remain the same. This implies that there are zero degrees of freedom dedicated to experimental error, SS_E. However, in his text, Montgomery appeals to the sparsity of effects principle. That is to say, most systems are dominated by some of the main effects and low-order interactions. The higher order interactions can be deemed negligible. Thus, a principal assumption made in this thesis is that only main effects and two-factor interactions are significant. This allows all three-factor interactions and higher to be pooled together as the experimental error.

Since the core focus of this thesis is to determine if the main effects and interactions among the image and algorithm experimental factors contribute significantly to the performance of an anomaly detector, we will first study the algorithms separately. Therefore, the 2^4 factorial design used to create the synthetic images will be combined with the 2^2 factorial design applied to test the RX Method on each synthetic image to yield an overall 2^6 factorial design. Figure 14 shows the graphical representation of the 2^6 factorial design that will be used to analyze the RX Method. For each image, the RX

Method will be performed with the 4 Principal Component/Window Size permutations to obtain the true positive fractions of the algorithm. Thus, $2^6 = 64$ responses will be collected. Similarly, the 2^6 factorial design will be used to test the effects on the BACON Method. The analysis of variance (ANOVA) procedure will be used to analyze the significance of the image and algorithm factors as it relates to the true positive fraction of each detector.

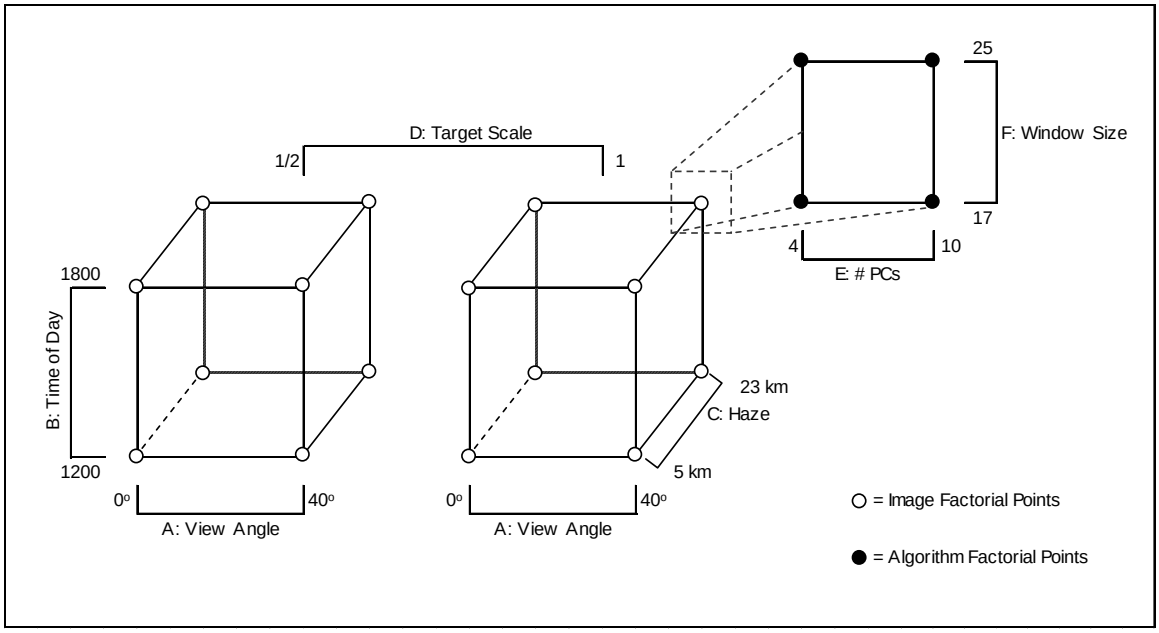


Figure 14. The 2^6 Factorial Design to Analyze the RX Method

The Overall Nested Design

At this time we will have determined which main effects and interactions among the image and algorithm experimental factors contribute significantly to the anomaly detection performance of each algorithm. However, the secondary focus of this thesis is

to determine whether or not the type of algorithm used is significant in finding anomalies in an image. A head-to-head comparison between the two algorithms is not possible by putting the two experimental designs from the previous sections side by side. This is a result of the fact that the algorithm experimental factors and levels are not the same from one algorithm to the next. Therefore, we are not comparing apples to apples. However, by utilizing a nested factorial design, we will be able to determine if the algorithm used is a significant factor. An "apples-to-apples" evaluation will then be made between the two algorithms.

A nested factorial design is used when the levels of one factor are similar, but not identical, for different levels of another factor. In this thesis, the levels of Algorithm Parameter 1 are not the same for the RX Method or the BACON Method. The same can be said for Parameter 2. Therefore, Parameter 1 and Parameter 2 are nested in Algorithm Type. Table 10 outlines the 7 factors and factor levels of the nested design. Table 11 displays the equations used to derive the sums of squares and degrees of freedom for the nested factorial design with 7 factors – two of which are nested within another factor.

Table 10. Factors and Factor Levels for the Nested Factorial Design

Factor Label	*Factor Name*	*Factor Levels*	
A	Sensor View Angle	0^o	40^o
B	Time of Day	1200	1800
C	Atmospheric Visibility	5 km	23 km
D	Target Scale	½	1
X	Algorithm Type	RX	BACON
E(X)	Parameter 1 nested in Algorithm Type	Low	High
F(X)	Parameter 2 nested in Algorithm Type	Low	High

Table 11. Partial ANOVA Table for the Nested Factorial Design

Source of Variation	*Sums of Squares*	*Degrees of Freedom*
A	$bcdxefg\sum_{i=1}^{a}\left(\bar{y}_{i\bullet\bullet\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$a-1$
B	$acdxefg\sum_{j=1}^{b}\left(\bar{y}_{\bullet j\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$b-1$
C	$abdxefg\sum_{k=1}^{c}\left(\bar{y}_{\bullet\bullet k\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$c-1$
D	$abcxefg\sum_{l=1}^{d}\left(\bar{y}_{\bullet\bullet\bullet l\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$d-1$
X	$abcdefg\sum_{m=1}^{x}\left(\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$x-1$
AB	$cdxefg\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{i\bullet\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet j\bullet\bullet\bullet\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(a-1)(b-1)$
AC	$bdxefg\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\bar{y}_{i\bullet k\bullet\bullet\bullet\bullet}-\bar{y}_{i\bullet\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet k\bullet\bullet\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(a-1)(c-1)$
AD	$bcxefg\sum_{i=1}^{a}\sum_{l=1}^{d}\left(\bar{y}_{i\bullet\bullet l\bullet\bullet\bullet}-\bar{y}_{i\bullet\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet l\bullet\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(a-1)(d-1)$
AX	$bcdefg\sum_{i=1}^{a}\sum_{m=1}^{x}\left(\bar{y}_{i\bullet\bullet\bullet m\bullet\bullet}-\bar{y}_{i\bullet\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(a-1)(x-1)$
BC	$adxefg\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\bar{y}_{\bullet jk\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet j\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet k\bullet\bullet\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(b-1)(c-1)$
BD	$acxefg\sum_{j=1}^{b}\sum_{l=1}^{d}\left(\bar{y}_{\bullet j\bullet l\bullet\bullet\bullet}-\bar{y}_{\bullet j\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet l\bullet\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(b-1)(d-1)$
BX	$acdefg\sum_{j=1}^{b}\sum_{m=1}^{x}\left(\bar{y}_{\bullet j\bullet\bullet m\bullet\bullet}-\bar{y}_{\bullet j\bullet\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(b-1)(x-1)$
CD	$abxefg\sum_{k=1}^{c}\sum_{l=1}^{d}\left(\bar{y}_{\bullet\bullet kl\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet k\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet l\bullet\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(c-1)(d-1)$
CX	$abdefg\sum_{k=1}^{c}\sum_{m=1}^{x}\left(\bar{y}_{\bullet\bullet k\bullet m\bullet\bullet}-\bar{y}_{\bullet\bullet k\bullet\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(c-1)(x-1)$
DX	$abcefg\sum_{l=1}^{d}\sum_{m=1}^{x}\left(\bar{y}_{\bullet\bullet\bullet lm\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet l\bullet\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet}+\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$(d-1)(x-1)$
E(X)	$abcdfg\sum_{m=1}^{x}\sum_{n=1}^{e}\left(\bar{y}_{\bullet\bullet\bullet\bullet mn\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet\bullet}\right)^2$	$x(e-1)$
F(X)	$abcd\,eg\sum_{m=1}^{x}\sum_{p=1}^{f}\left(\bar{y}_{\bullet\bullet\bullet\bullet m\bullet p}-\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet\bullet}\right)^2$	$x(f-1)$
EF(X)	$abcdg\sum_{m=1}^{x}\sum_{n=1}^{e}\sum_{p=1}^{f}\left(\bar{y}_{\bullet\bullet\bullet\bullet mnp}-\bar{y}_{\bullet\bullet\bullet\bullet mn\bullet}-\bar{y}_{\bullet\bullet\bullet\bullet m\bullet p}+\bar{y}_{\bullet\bullet\bullet\bullet m\bullet\bullet\bullet}\right)^2$	$x(e-1)(f-1)$
Error	By Subtraction	By Subtraction

Total	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{x}\sum_{n=1}^{e}\sum_{p=1}^{f}\sum_{q=1}^{g}\left(y_{ijklmnpq}-\bar{y}_{\bullet\bullet\bullet\bullet\bullet\bullet\bullet}\right)^2$	$abcdxefg-1$

In the partial ANOVA table, a lower case letter corresponds to the number of levels for the factor of the associated upper case letter. Therefore, since Factors A, B, C, D, X, E, and F all have 2 levels each, the values of *a*, *b*, *c*, *d*, *x*, *e*, and *f* are 2. Also, since only one replication can be run of the design, the value of g is 1. This will result in 127 total degrees of freedom. One degree of freedom is dedicated for each of the main effects and two-factor interactions among A, B, C, D, and X. It also produces 2 degrees of freedom for each of the nested factors E(X), F(X), and EF(X). Finally, since all three-factor interactions and higher will be considered insignificant, there are 106 degrees of freedom devoted to model error.

In order to complete the ANOVA table, the expected mean squares must be calculated to determine the appropriate F statistic for testing the effects of all factors. Since factors A, B, C, D, and X are fixed effects, it is assumed that $\sum_{i=1}^{a}\alpha_i=0$, $\sum_{j=1}^{b}\beta_j=0$, $\sum_{k=1}^{c}\gamma_k=0$, $\sum_{l=1}^{d}\delta_l=0$, $\sum_{m=1}^{x}\chi_m=0$, $\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}=0$, $\sum_{i=1}^{a}\sum_{k=1}^{c}(\alpha\gamma)_{ik}=0$, $\sum_{i=1}^{a}\sum_{l=1}^{d}(\alpha\delta)_{il}=0$, $\sum_{i=1}^{a}\sum_{m=1}^{x}(\alpha\chi)_{m}=0$, $\sum_{j=1}^{b}\sum_{k=1}^{c}(\beta\gamma)_{jk}=0$, $\sum_{j=1}^{b}\sum_{l=1}^{d}(\beta\delta)_{jl}=0$, $\sum_{j=1}^{b}\sum_{m=1}^{x}(\beta\chi)_{jm}=0$, $\sum_{k=1}^{c}\sum_{l=1}^{d}(\gamma\delta)_{kl}=0$, $\sum_{k=1}^{c}\sum_{m=1}^{x}(\gamma\chi)_{km}=0$, and $\sum_{l=1}^{d}\sum_{m=1}^{x}(\delta\chi)_{lm}=0$. That is, the treatment effects for factor A, for example, sum to zero. Also, since the nested factors E and F are fixed effects, $\sum_{n=1}^{e}\eta_{n(m)}=0$, $\sum_{p=1}^{f}\tau_{p(m)}=0$, and $\sum_{n=1}^{e}\sum_{p=1}^{f}(\eta\tau)_{np(m)}=0$ for $m = 1, 2,\ldots, x$. Thus, since all factors are fixed effects, F_0 can be determined by dividing each of the expected mean

squares by the expected mean square for experimental error (MS_E). If $F_0 > F_{\alpha, ndf, ddf}$, where α is the specified degree of significance, *ndf* is the number of numerator degrees of freedom and *ddf* is the number of denominator degrees of freedom, then the corresponding experimental factor is considered to be significant in the design. Table 12 displays the expected mean squares and F-statistics for each main effect and two-factor interaction.

Table 12. Expected Mean Squares and F-statistics for the Nested Factorial Design

Source	Expected Mean Squares	F_0	Source	Expected Mean Squares	F_0
A	$\sigma^2 + \frac{bcdxefg\sum_{i=1}^{a}\alpha_i^2}{a-1}$	$\frac{E(MS_A)}{E(MS_{Error})}$	BD	$\sigma^2 + \frac{acxefg\sum_{j=1}^{b}\sum_{l=1}^{d}(\beta\delta)_{jl}^2}{(b-1)(d-1)}$	$\frac{E(MS_{BD})}{E(MS_{Error})}$
B	$\sigma^2 + \frac{acdxefg\sum_{i=1}^{a}\beta_i^2}{b-1}$	$\frac{E(MS_B)}{E(MS_{Error})}$	BX	$\sigma^2 + \frac{acdefg\sum_{j=1}^{b}\sum_{m=1}^{x}(\beta\chi)_{jm}^2}{(b-1)(x-1)}$	$\frac{E(MS_{BX})}{E(MS_{Error})}$
C	$\sigma^2 + \frac{abdxefg\sum_{i=1}^{a}\gamma_i^2}{c-1}$	$\frac{E(MS_C)}{E(MS_{Error})}$	CD	$\sigma^2 + \frac{abxefg\sum_{k=1}^{c}\sum_{l=1}^{d}(\gamma\delta)_{kl}^2}{(c-1)(d-1)}$	$\frac{E(MS_{CD})}{E(MS_{Error})}$
D	$\sigma^2 + \frac{abcxefg\sum_{i=1}^{a}\delta_i^2}{d-1}$	$\frac{E(MS_D)}{E(MS_{Error})}$	CX	$\sigma^2 + \frac{abdefg\sum_{k=1}^{c}\sum_{m=1}^{x}(\gamma\chi)_{km}^2}{(c-1)(x-1)}$	$\frac{E(MS_{CX})}{E(MS_{Error})}$
X	$\sigma^2 + \frac{abcdefg\sum_{i=1}^{a}\chi_i^2}{x-1}$	$\frac{E(MS_X)}{E(MS_{Error})}$	DX	$\sigma^2 + \frac{abcefg\sum_{l=1}^{d}\sum_{m=1}^{x}(\delta\chi)_{lm}^2}{(d-1)(x-1)}$	$\frac{E(MS_{DX})}{E(MS_{Error})}$
AB	$\sigma^2 + \frac{cdxefg\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{E(MS_{AB})}{E(MS_{Error})}$	E(X)	$\sigma^2 + \frac{abcdfg\sum_{m=1}^{x}\sum_{n=1}^{e}\eta_{n(m)}^2}{x(e-1)}$	$\frac{E(MS_{E(X)})}{E(MS_{Error})}$
AC	$\sigma^2 + \frac{bdxefg\sum_{i=1}^{a}\sum_{k=1}^{c}(\alpha\gamma)_{ik}^2}{(a-1)(c-1)}$	$\frac{E(MS_{AC})}{E(MS_{Error})}$	F(X)	$\sigma^2 + \frac{abcdeg\sum_{m=1}^{x}\sum_{p=1}^{f}\tau_{p(m)}^2}{x(f-1)}$	$\frac{E(MS_{F(X)})}{E(MS_{Error})}$
AD	$\sigma^2 + \frac{bcxefg\sum_{i=1}^{a}\sum_{l=1}^{d}(\alpha\delta)_{il}^2}{(a-1)(d-1)}$	$\frac{E(MS_{AD})}{E(MS_{Error})}$	EF(X)	$\sigma^2 + \frac{abcdg\sum_{m=1}^{x}\sum_{n=1}^{e}\sum_{p=1}^{f}(\eta\tau)_{np(m)}^2}{x(e-1)(f-1)}$	$\frac{E(MS_{EF(X)})}{E(MS_{Error})}$
AX	$\sigma^2 + \frac{bcdefg\sum_{i=1}^{a}\sum_{m=1}^{x}(\alpha\chi)_{im}^2}{(a-1)(x-1)}$	$\frac{E(MS_{AX})}{E(MS_{Error})}$	Error	σ^2	

BC	$\sigma^2 + \frac{adxefg\sum_{j=1}^{b}\sum_{k=1}^{c}(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$\frac{E(MS_{BC})}{E(MS_{Error})}$	

IV. Results and Analysis

Overview and Assumptions

Prior to a full examination and discussion of the results, it is first important to outline how the results will be presented and discuss some of the assumptions used in the statistical analysis. Chapter 4 is divided into four main sections: a) the creation of the synthetic hyperspectral images, b) the results of the RX Method, c) the results of the BACON Method, and d) a head-to-head comparison of the RX and BACON Methods. As stated in Chapter 2, the use of the analysis of variance (ANOVA) procedure allows us to analyze the significance of each main effect and joint interaction on the true positive fraction of each algorithm. To assist in the calculations of the ANOVA and to provide model diagnostic plots, the Minitab statistical software package was used. The level of significance for each test in this chapter is $\alpha = 0.05$.

As stated earlier in regards to the statistical analysis of the experimental designs, only a single replication of the design is run since the results will not change from run-to-run given that the factor levels remain the same. Thus, in order to estimate experimental error, a principal assumption in this thesis is that only main effects and two-factor interactions are significant and higher order interaction terms can be pooled into experimental error.

OC curves were also used to present the overall detection performance of each algorithm. To build the OC curves, the truth maps were used to locate every man-made anomalous pixel in the image. These are the target pixels. Depending on the algorithm

used, the appropriate distance measure was then found for each corresponding target pixel. The distances were then sorted in descending order and were, one by one, defined as the threshold. Given each threshold, the true positive fractions and false positive fractions were recorded to construct the OC curves.

Binary image maps were also used in the comparison of algorithms and algorithm parameter settings among synthetic images. The binary maps display white pixels where the detectors declared an anomaly. The binary maps in this thesis display the detector results for a fixed false positive rate of 0.01. Appendix E reports a comparison of the RX and BACON methods using OC curves and binary maps for every image.

Creating the Synthetic Hyperspectral Images

The DIRSIG image used as the base model for this thesis is the Western Rainbow scene. It is relatively homogeneous in the sense that the majority of the image is either normal vegetation or a mixture of grass and dirt. There are, however, several trees and bushes, as well as roughly 30 military targets (tanks, missile carriers, and trucks), scattered throughout the image. A dirt road also runs along the right-hand side of the scene. The image, when viewed from directly overhead, is 250 by 250 pixels in size with a ground sample distance of approximately 2 meters. A framing array sensor with an AVIRIS spectral range of 0.4 to 2.5 μm was utilized at an altitude of 15,000 feet. The 16 synthetic hyperspectral images used in this thesis study were created using the 2^4 factorial design shown in Table 8. A summary of the experimental factors and factor levels for the image creation is listed in entirety for reference in Table 13.

Table 13. Summary of Factors and Factor Levels for Synthetic Image Creation

	Sensor View Angle	*Time of Day*	*Atmospheric Visibility*	*Scale of Targets*
Low (-1)	0^o	1200	5 km	1/2
High (1)	40^o	1800	23 km	1

Figure 15 displays a sample of the synthetic images generated by DIRSIG subject to the factor levels given in Table 8. The figure is shown only as a visual comparison between Image 3 and the other images since there is only a one factor difference between Image 3 and each of the others. Image 3 is viewed from directly overhead (0^o) at 1200 with 23 km of atmospheric visibility. Image 10, by comparison, is viewed at a 40^o angle. The lone variation between Image 1 and Image 3 is that there is 5 km of atmospheric visibility in Image 1. Image 14 varies from Image 3 only in the time of day. Image 14 was collected at 1800. In Image 14, as well as all images created with the time of day equal to 1800, the sun is setting to the left hand side of the image resulting in longer shadows to the right hand side of trees, hills, and military targets. The targets in all four images are at full scale.

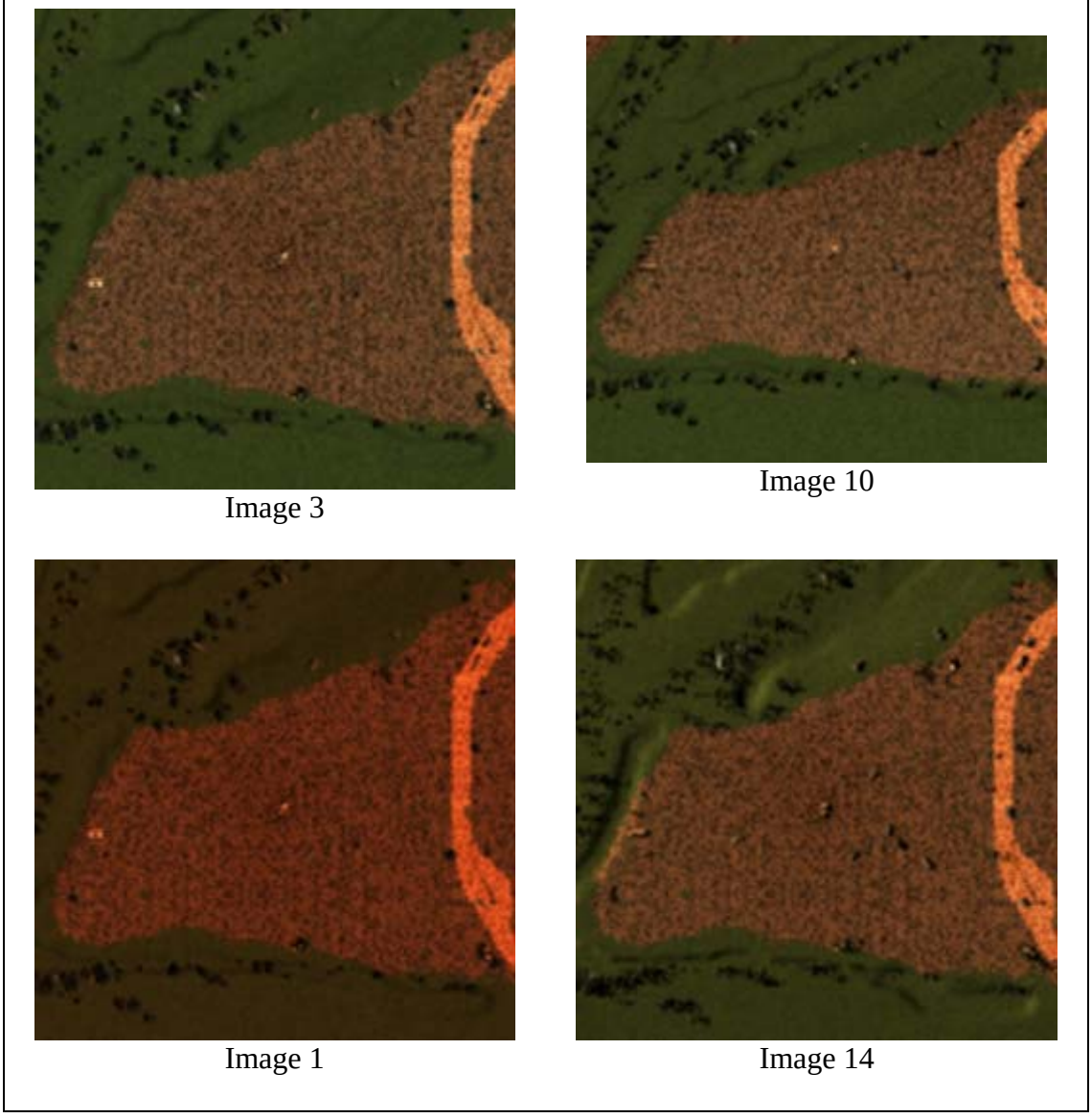
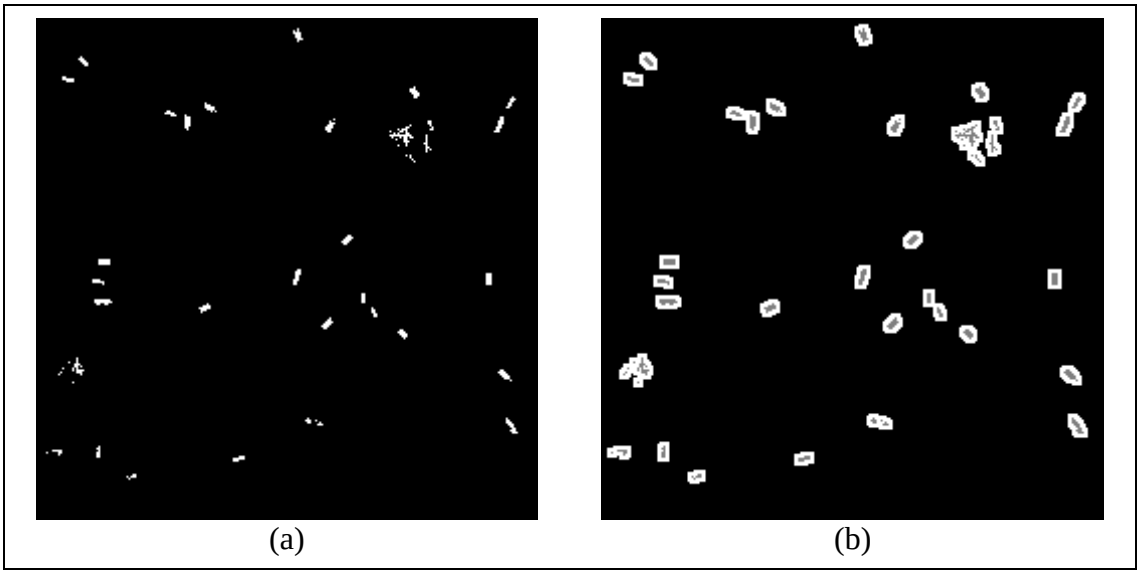


Figure 15. Comparison of a Few Synthetic Images

Truth images were also generated for each synthetic image. The truth images show exactly where the targets of interest are in the scene. Even though synthetic imagery provides a 100% per-pixel ground truth, 2 layers of "buffer" pixels were built around the true target points to account for spectral mixing between the edges of the

targets and the background. These buffer pixels were not taken into account for the determination of the OC curve. Figure 16(a) displays the actual truth image from Image 3 where white pixels represent target locations while Figure 16(b) shows the truth image with buffer pixels where gray pixels represent targets and the white pixels are the buffer pixels.



(a) (b)

Figure 16. Truth Image versus Truth Image with Buffer Pixels

Prior to processing the images through the anomaly detectors, the maximum values of the data in the synthetic image cube were found to be on the order of 10^{-5}. Since working with miniscule numbers like this may result in poor algorithm results, the images were each multiplied by 10^{9} in order to scale the data to useful values. Normal hyperspectral data falls in the range of 0 and 12000.

Additionally, before processing any of the synthetic images, it was important to delete the atmospheric absorption bands from the data. For these images, the following

hyperspectral channels were considered the "good bands:" 4-58, 63-79, 83-106, 120-151, and 184-218. From here the red, green, and blue color bands that were used to show the colored images in this thesis were computed. The red color band was determined by averaging bands 22 through 35. The green color band was determined by averaging bands 9 through 16. The red color band was determined by averaging bands 4 through 8.

Once all synthetic images and truth images were created, they were processed through the RX and BACON hyperspectral anomaly detectors using the factorial designed experiments described in Chapter 3. Significant main effects and interactions between the experimental factors were studied with the help of ANOVA and optimal algorithm parameter settings were determined. The following sections detail the results of the two designed experiments and a head-to-head comparison is made between the RX and BACON Methods.

Results from the RX Method

The RX Method utilized in this thesis is outlined in Chapter 2. The coded algorithm was borrowed from Capt Yuri Taitano's master's thesis entitled "Hyperspectral Imagery Target Detection Using the Iterative RX Detector." However, instead of implementing an iterative approach to finding anomalies in hyperspectral data as Capt Taitano studied, this thesis applied a single iteration of the RX algorithm. The RX Method applies a user-defined processing window to each of the pixels in the image. This results in a border around the image where several pixels cannot be processed. Figure 17 displays the "untestable" border pixels for each window size of the design in black. These images can be compared to the "borderless" Image 3 in Figure 15.

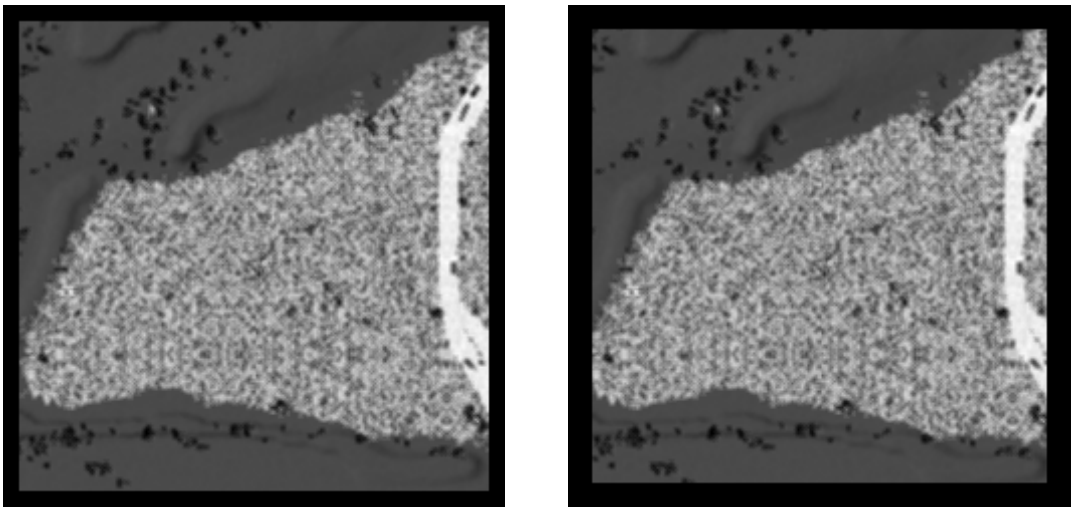


Figure 17. "Untestable" Pixels for RX Processing Window Sizes

All 16 synthetic images were processed by the four settings of the RX algorithm stated in Table 9. The four settings are a) 4 principal components retained with a processing window size of 17 pixels, b) 4 PCs retained with a window size of 25, c) 10 PCs retained with a window size of 17, and d) 10 PCs retained with a window size of 25. Given that there are 4 experimental image factors and 2 algorithm factors, each with 2 levels, a 2^6 factorial design was used to analyze the response of the RX Method. The response of interest was the true positive fraction (TPF) of the detector given that the false positive fraction (FPF) is fixed at 0.01. A summary of the 64 responses for the RX Method is listed in Appendix A.

Once all images had been processed by the four experimental settings of the RX Method and all responses were collected, the Minitab statistical software was used to calculate the ANOVA table and produce model diagnostic plots. Prior to relying on the

results from the ANOVA table, it is important to first check for model adequacy. That is, the assumption of normally and independently distributed error terms with constant variance must be investigated. These assumptions were easily checked and confirmed by inspecting the normal probability plot of the residuals and the plot of the residuals versus the fitted values. These plots are shown in Appendix C.

It can be seen from the normal probability plot that the errors are normally distributed since the residuals fall on the straight line, for the most part. There are no gross deviations from normality. Also, the plot of the residuals versus the fitted values displays a structureless "shotgun" pattern which implies that there is constant variance among the observations. Given that the original assumptions hold, we can examine the ANOVA table in Table 14 for significance among the main effects and two-factor interactions of the 6 experimental factors: 4 image-specific factors and 2 algorithm-specific factors. Figure 18 displays the Half Normal Plot for the RX Method.

Table 14. ANOVA Table for RX Method

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_0	*p-value*
Model	0.3818	15	0.0255	70.1905	< 0.0001
A (Sensor View Angle)	0.0309	1	0.0309	85.3473	< 0.0001
B (Time of Day)	0.0041	1	0.0041	11.2071	0.0016
C (Atmospheric Visibility)	0.0033	1	0.0033	9.0620	0.0042
D (Target Scale)	0.0087	1	0.0087	23.8891	< 0.0001
E (# of PCs Retained)	0.0146	1	0.0146	40.3410	< 0.0001
F (Window Size)	0.2543	1	0.2543	701.2443	< 0.0001
AB	0.0060	1	0.0060	16.5202	0.0002
AE	0.0068	1	0.0068	18.8146	< 0.0001
BD	0.0075	1	0.0075	20.6094	< 0.0001
BE	0.0092	1	0.0092	25.4275	< 0.0001
CE	0.0024	1	0.0024	6.6075	0.0133
DE	0.0078	1	0.0078	21.5252	< 0.0001
DF	0.0208	1	0.0208	57.4802	< 0.0001
EF	0.0039	1	0.0039	10.7375	0.0020
Residual	0.0174	48	0.0004		
Total	0.3992	63			

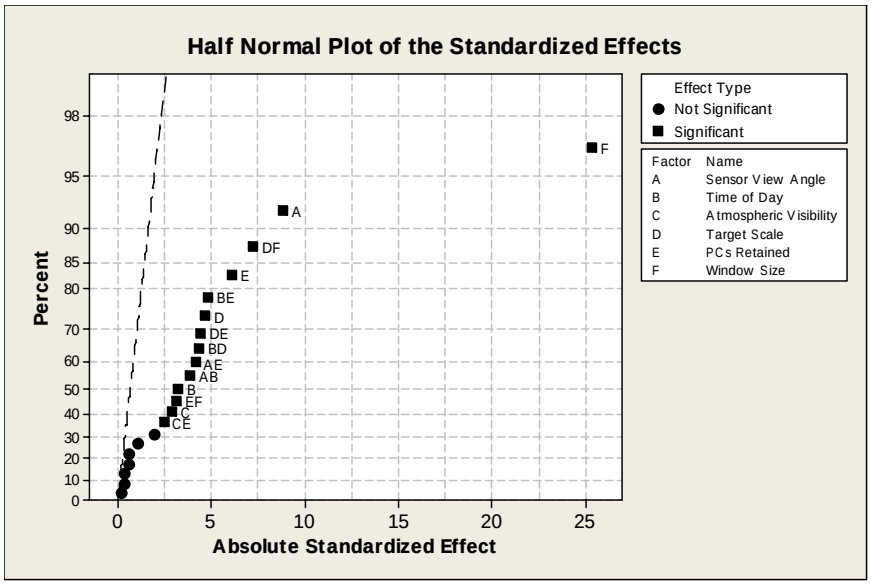


Figure 18. Half Normal Plot for RX Method

By inspecting the ANOVA table for the RX Method in Table 14, all 6 main effects can be seen to be significant since their p-values are less than 0.05. The ANOVA table also shows that factor F (Window Size) is the single most significant factor for the RX Method. Factor F contributes 66.6% to the model. In contrast, the next five highest contributors combined (A at 8.1%, DF at 5.5%, E at 3.8%, BE at 2.4%, and D at 2.3%) only account for 22.1% of the model. The R^2 value of this model is 0.9527 (R^2_{adj} = 0.9392, R^2_{pred} = 0.9192). This implies that the model explains 95% of the variability in the data. The plots of the main effects are shown in Figure 19. The processing window size definitely stands out as the most important factor. As the window size increases from it low level to its high level, the detection performance of the RX Method increases.

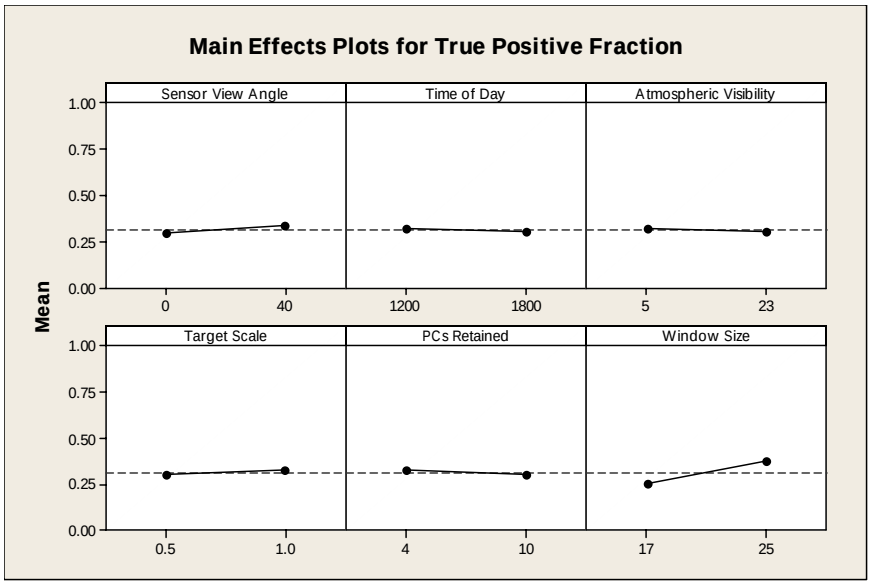


Figure 19. Main Effects Plots for RX Method

Even though the factors of view angle, time of day, atmospheric visibility, and target size can be controlled using DIRSIG, in the real world they are considered to be noise factors since they are difficult to control and keep at a specified target. The parameters of the RX method, on the other hand, can be controlled by setting the levels of the appropriate factors. The control-by-noise interaction plots can be studied to determine robust settings of the algorithm parameters. In other words, we can choose the algorithm settings that minimize the variability in the response transmitted from the in-scene image factors. The significant control-by-noise interaction plots are shown in Figure 20. In all interaction plots involving factor E, it can be seen that, regardless of the number of principal components retained to reduce the data, the true positive rate of the RX detector remains fairly stable. That is to say, the variability in the response is minimal for either setting of factor E. It can also be observed from the DF (Target Size × Window Size) interaction plot that the variability in the response is minimized for a

smaller processing window size even though better anomaly detection occurs with a larger processing window and larger targets. Thus, in order to minimize the variance in the response when studying the RX method, it is recommended to use a processing window of 17 pixels while retaining either the top 4 or 10 principal components.

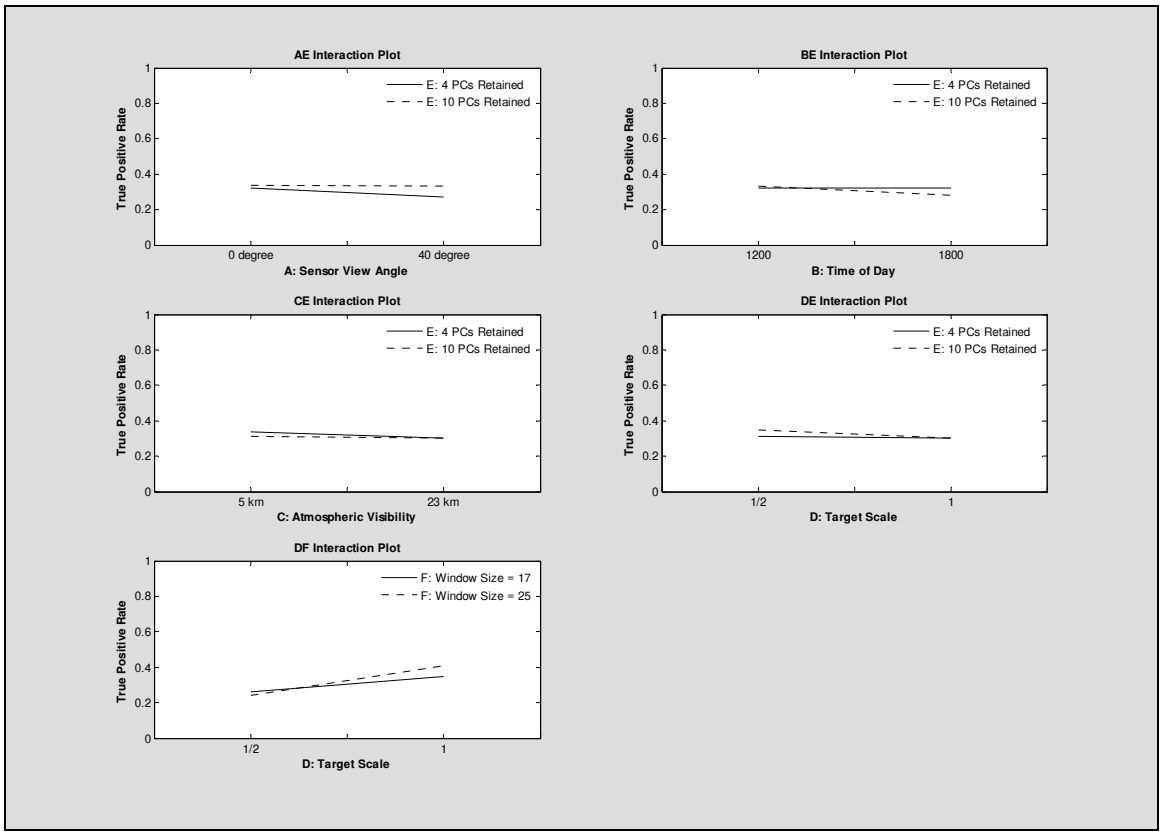


Figure 20. Control-by-noise Interaction Plots for the RX Method

Through the use of a factorial experimental design, all image and algorithm main effects were found to be significant in determining the performance of the RX Method. The data showed that the RX Method performed the best on 10 of the 16 synthetic images with the use of a 25 pixel processing window and by retaining the top 4 PCs. Of the

remaining 6 images, a 25 pixel window and retaining 10 PCs resulted in the best RX detection performance. Thus, for this family of synthetic images, a processing window of 25 pixels creates the best anomaly detection performance for the RX Method.

Results from the BACON Method

The BACON algorithm used in this thesis is outlined in Chapter 2. The algorithm is part of the auto_det function written by Maj Timothy Smetek as part of his dissertation. Auto_det is a multivariate outlier detector designed to detect anomalies in hyperspectral imagery. It combines a *k*-means clustering algorithm with the BACON algorithm proposed by Billor, Hadi, and Velleman (2000).

The first step prior to running the BACON algorithm was to cluster the image data using a *k*-means clustering algorithm. The *k*-means algorithm uses a two-phase iterative algorithm to minimize the sum of point-to-centroid distances summed over all *k* clusters. The user determines the value of *k*. The algorithm also contains a distance parameter in which *k*-means minimizes with respect to. This parameter can be set to "Cosine," "Correlation," "Squared Euclidean," "City Block," and "Hamming." For reasons not completely examined in this thesis, the cosine and correlation distances caused a breakdown in the algorithm when it was run on some images. However, there is belief that, since the breakdown occurred on images in which the overall illumination in the scene was diminished due to low atmospheric visibility or the time of day being set to 1800, the magnitudes of the some of the data were so small that they were effectively zero. Since the cosine distance measures the cosine of the angle between points and if all points were considered to be zero, then there would not be an angle to measure and the

data could not be divided into the required *k* clusters. Therefore, the distance metric chosen for *k*-means was the squared Euclidean distance.

Once the data was clustered, the BACON algorithm was applied to determine outliers in each of the clusters. Initial results showed that the alpha level for the χ^2 test was set too high. The original alpha was equal to 0.05 which implies that the cutoff value for the χ^2 test was 1- 0.05/*n* (where *n* was the size of the cluster). This, as shown in Figure 21(a), resulted in too many pixels being labeled as outliers. Indeed, the algorithm found the man-made outliers in the image (i.e. tanks, missile carriers, and trucks), but it also found the natural anomalies in the scene. Trees, bushes, vehicle tracks, the road, and boundaries between background materials, such as the dirt and grass, were considered outliers. This is mostly due to the fact that DIRSIG's spectral-spatial variability of backgrounds is not nearly as complex as the real world. This fact must be remembered when drawing any conclusions throughout this thesis when comparing the performance of the algorithms. Through some testing, it was found that the alpha level for the χ^2 test should be set a little bit lower in an attempt to suppress the natural anomalies, but still locate the man-made outliers in the image. An alpha value equal to 10^{-12} was determined to provide a reasonable χ^2 cutoff point of $1\text{-}10^{-12}/n$ for this family of synthetic images. Improved RX results can be viewed in Figure 21(b).

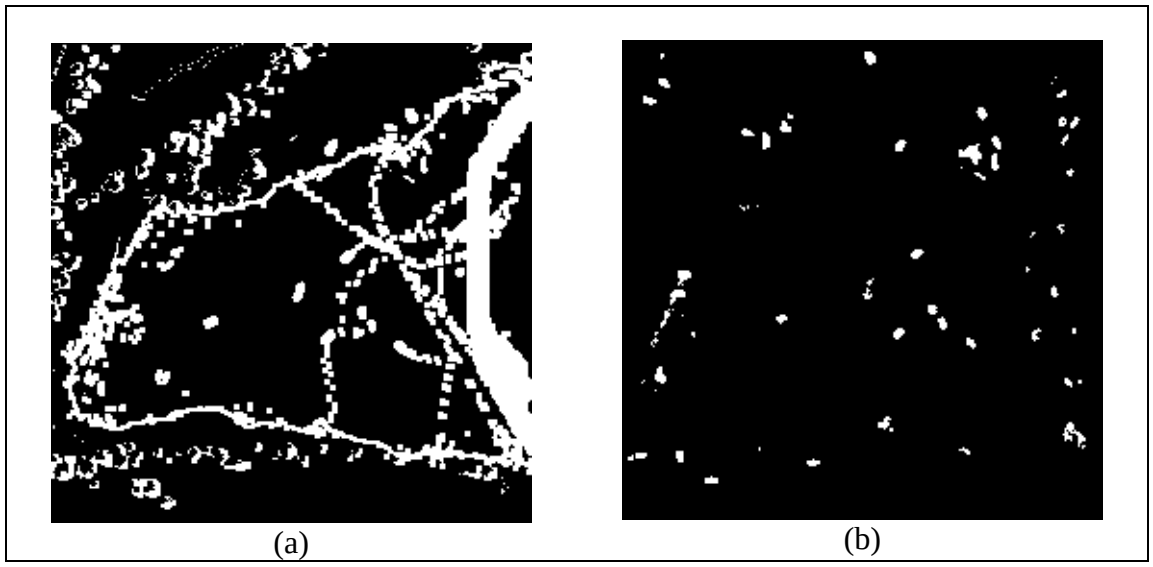


Figure 21. Initial (a) and Improved (b) BACON Results for Image 1

Again, the 16 synthetic images were processed by the settings of the BACON algorithm shown in Figure 13. Since there are 4 experimental image factors and 2 algorithm factors, each with 2 levels, a 2^6 factorial design was used to analyze the response of the BACON Method. The response of interest was still the true positive fraction (TPF) of the detector given that the false positive fraction (FPF) is fixed at 0.01. A summary of the 64 responses is listed in Appendix B. Once all images had been processed by the BACON Method, the Minitab statistical software was used to calculate the ANOVA table and produce model diagnostic plots. As with the RX Method, a check for model adequacy was performed. The assumptions of normally and independently distributed error terms with constant variance were confirmed by inspecting the normal probability plot of the residuals and the plot of the residuals versus the fitted values. These plots are shown in Appendix D. Given that these assumptions hold, we can examine the ANOVA table in Table 15 for significance among the main effects and

interactions of the 6 experimental factors: 4 image-specific factors and 2 algorithm-specific factors. Figure 22 displays the Half Normal Plot for the BACON Method.

Table 15. ANOVA Table for BACON Method

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_0	*p-value*
Model	1.5926	8	0.1991	8.3945	< 0.0001
A (Sensor View Angle)	0.3249	1	0.3249	13.6979	0.0005
B (Time of Day)	0.0256	1	0.0256	1.0791	0.3035
C (Atmospheric Visibility)	0.0136	1	0.0136	0.5739	0.4520
F (# of Clusters)	0.5242	1	0.5242	22.1018	< 0.0001
AC	0.1289	1	0.1289	5.4349	0.0234
AF	0.2441	1	0.2441	10.2925	0.0022
BF	0.1764	1	0.1764	7.4366	0.0086
CF	0.1551	1	0.1551	6.5392	0.0133
Error	1.3043	55	0.0237		
Total	2.8970	63			

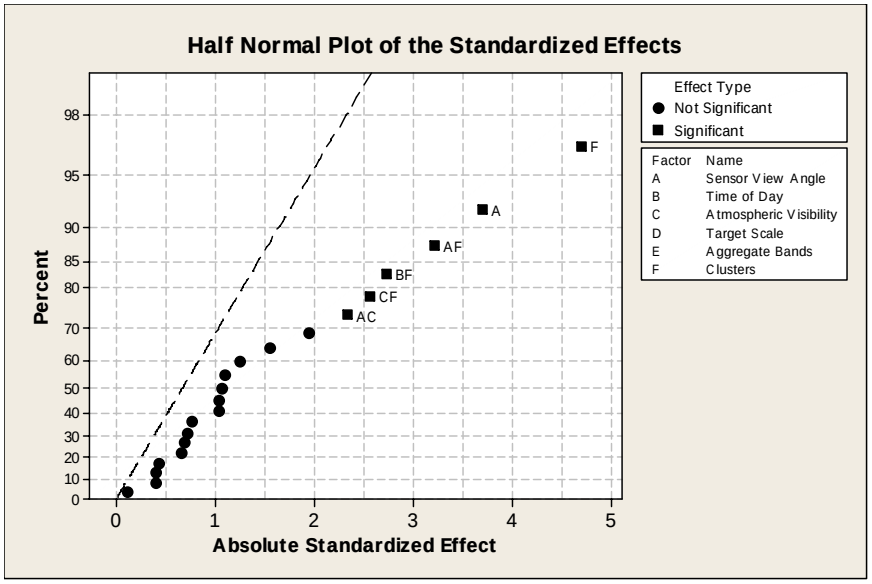


Figure 22. Half Normal Plot for BACON Method

The ANOVA table reveals the significant main effects and two-factor interactions for the BACON Method. Upon inspection, only 2 of the 6 main effects can be seen to be significant since their p-values are less than 0.05. The viewing angle (Factor A) of the hyperspectral sensor is the only image related factor that contributes significantly to BACON's detection performance. Likewise, the number of clusters (Factor F) in the scene is the only significant algorithm parameter. In fact, factors A, F, and AF contribute 68.6% to the model. The plots of the main effects are shown in Figure 23. It is shown that as both the sensor's view angle and the number of image clusters increase from their low levels to their high levels, BACON's detection performance decreases. The plots also show that as Factor E (Band Aggregation) varies from its low level to its high level, the average response does not change. Therefore, it can be concluded that algorithm performance is not diminished by averaging every 10 bands to reduce the dimensionality of the data. The R^2 value of this model is 0.5498 (R^2_{adj} = 0.4843, R^2_{pred} = 0.3904). This implies that the model explains 55% of the variability in the data.

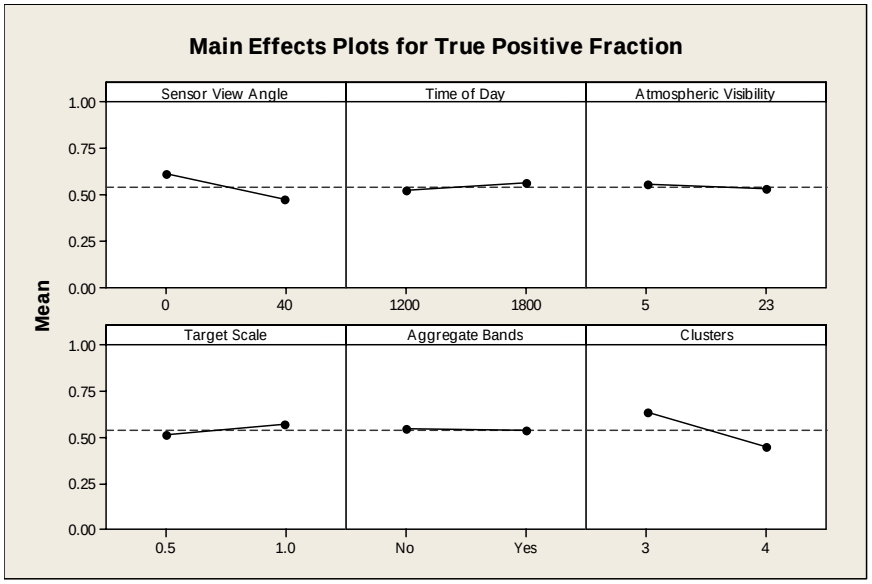


Figure 23. Main Effects Plots for BACON Method

Again, the control-by-noise interaction plots can be studied to determine the settings of BACON's parameters that minimize the variability in the response transmitted from the in-scene image factors. The significant interaction plots for the BACON method are shown in Figure 24. For the AF (View Angle × # of Clusters) and BF (Time of Day × # of Clusters) control-by-noise interaction plots, it is illustrated that using 3 clusters minimizes the variance in the response transmitted from the sensor's view angle and time of day. However, the CF (Atmospheric Visibility × # of Clusters) interaction plot shows that 4 data clusters minimize the variability in the true positive rate of the BACON detector from the atmospheric visibility.

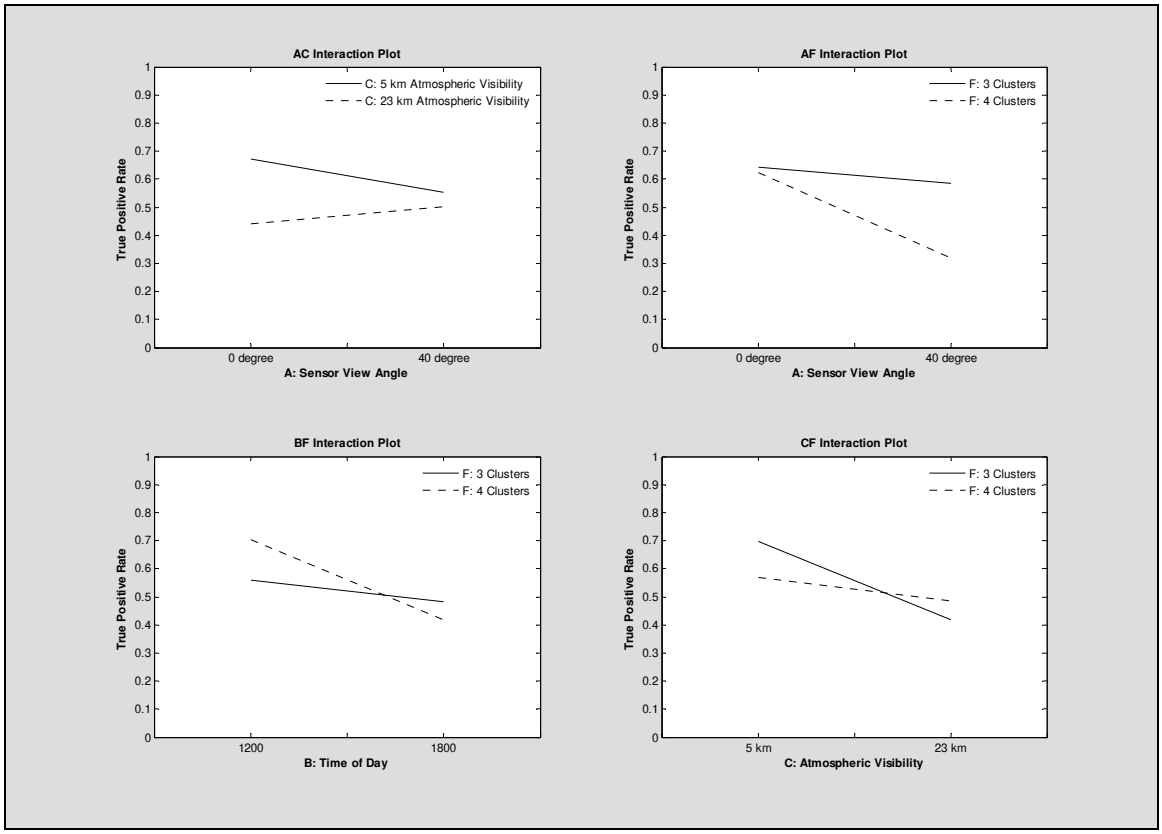


Figure 24. Interaction Plots for the BACON Method

The application of a factorial experimental design has shown that only the viewing angle of the sensor and the number of data clusters in the scene were found to be significant in determining the performance of the BACON Method. The data showed that BACON performed the best on 13 of the 16 synthetic images when 3 clusters are determined. The three images that performed the best with 4 clusters happened to have combination of a 0^o sensor view angle and time of day of 1200. As far as algorithm parameter settings go, it was shown that further data reduction by the means of Band Aggregation prior to running the image through BACON did not significantly increase the algorithm's performance.

An RX versus BACON Comparison

The RX and BACON anomaly detectors were chosen to be studied in this thesis for several reasons. The interest lies in the fact that classical Mahalanobis distance-based outlier detectors, like the RX Method, are limited. These methods rely on non-robust covariance matrix estimates that are highly sensitive to outlying observations. The BACON Method, on the other hand, is a multivariate outlier detector that uses robust estimates of the mean and covariance. This section of Chapter 4 will attempt to show that specific robust multivariate outlier detectors are superior to the classical methods. Two tests were designed to compare the RX and BACON anomaly detectors. Before comparing algorithms, though, a nested experimental design was used to determine if the anomaly detection algorithm used is statistically significant.

Nested Experimental Design

It is important to realize that a head-to-head comparison between the two algorithms is not possible by putting the two experimental designs from the previous sections side by side. This is a result of the fact that the algorithm experimental factors and levels were not the same from one algorithm to the next. Therefore, we are not comparing apples to apples. However, by utilizing a nested factorial design, we will be able to determine if the algorithm used is a significant factor. An "apples-to-apples" comparison can then be made between algorithms.

A nested factorial design is applied when the levels of one factor are similar, but not identical, for different levels of another factor. In this thesis, the levels of Algorithm Parameter 1 are not the same for the RX Method or the BACON Method. The same can be said for Parameter 2. Therefore, Parameter 1 and Parameter 2 are nested in Algorithm Type. Table 10 outlines the 7 factors and factor levels of the nested design. Figure 25 shows a diagram of the nested design.

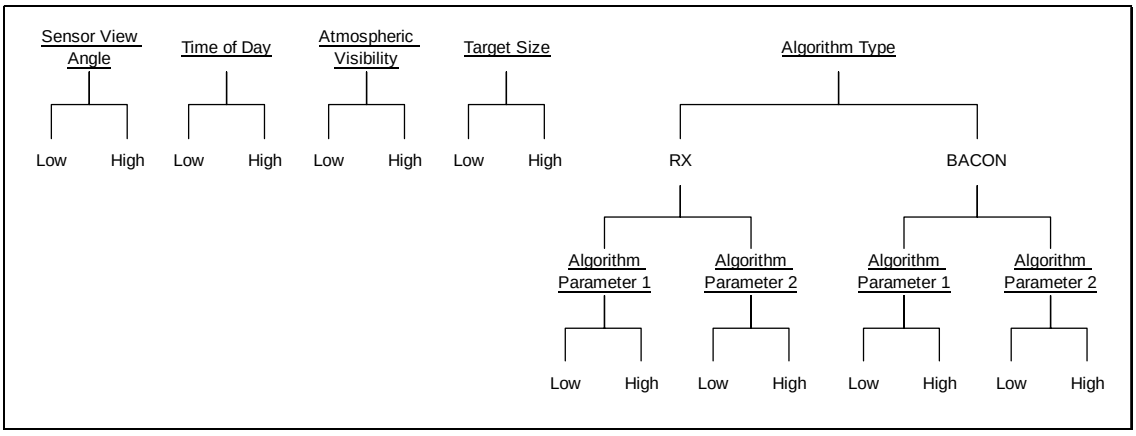


Figure 25. Schematic of Nested Factorial Design

To analyze the nested design, Minitab was used to construct the ANOVA table to determine the significant main effects and two-factor interactions. Again, three-factor interactions and higher were deemed irrelevant and were pooled into the estimate for error. The resulting ANOVA table is shown in Table 16. For an alpha level of 0.05, factors A, X, AX, and F(X) were found to be the significant effects. Indeed, it has been discovered that the algorithm used is important in detecting anomalies. It also makes sense that the sensor's view angle was established to be significant since it was the only image related factor to be significant for both the RX and BACON methods.

The main effects plots and the AX interaction plot are shown in Figure 26 and Figure 27, respectively. The key point to take from Figure 26 is that the average TPF is much larger for the BACON Method than it is for the RX Method. Also, since the sensor's view angle was found to be a significant factor, the main effects plot shows that slightly better anomaly detection occurs at a lower view angle. The AX interaction plot illustrates that BACON performs better for a 0^o view angle while RX performs somewhat better for a 40^o view angle.

Table 16. ANOVA Table for the Nested Experimental Design

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_0	*p-value*
Model	3.1473	21	0.1499	8.6880	< 0.0001
A	0.0776	1	0.0776	4.5003	0.0362
B	0.0046	1	0.0046	0.2684	0.6055
C	0.0151	1	0.0151	0.8774	0.3511
D	0.0550	1	0.0550	3.1867	0.0771
X	1.6797	1	1.6797	97.3702	< 0.0001
AB	0.0363	1	0.0363	2.1030	0.1500
AC	0.0670	1	0.0670	3.8856	0.0513
AD	0.0153	1	0.0153	0.8886	0.3480
AX	0.2782	1	0.2782	16.1255	0.0001
BC	0.0043	1	0.0043	0.2521	0.6166
BD	0.0302	1	0.0302	1.7499	0.1887
BX	0.0250	1	0.0250	1.4508	0.2311
CD	0.0454	1	0.0454	2.6328	0.1076
CX	0.0018	1	0.0018	0.1020	0.7500
DX	0.0106	1	0.0106	0.6130	0.4354
E(X)	0.0185	2	0.0093	0.5365	0.5864
F(X)	0.7784	2	0.3892	22.5633	< 0.0001
EF(X)	0.0042	2	0.0021	0.1214	0.8858
Error	1.8285	106	0.0173		
Total	4.9758	127			

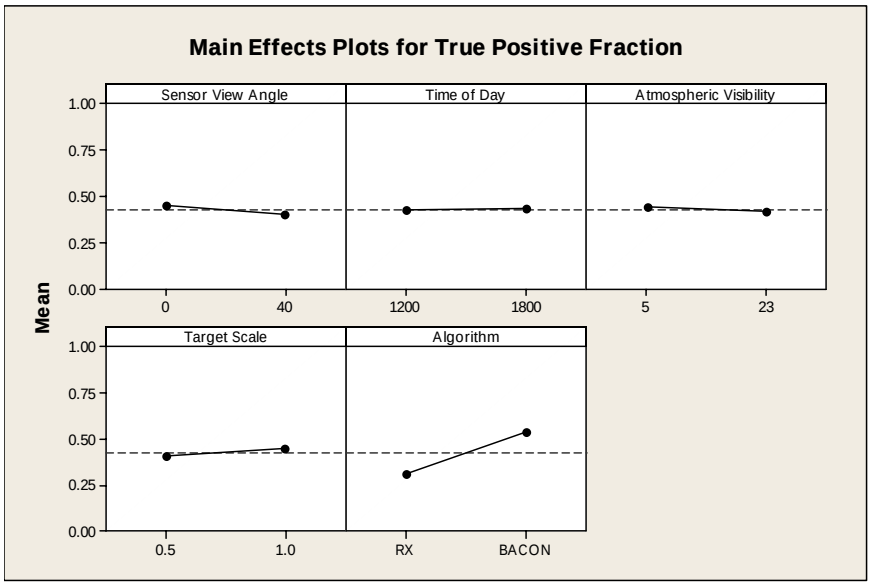


Figure 26. Main Effects Plots for the Nested Experimental Design

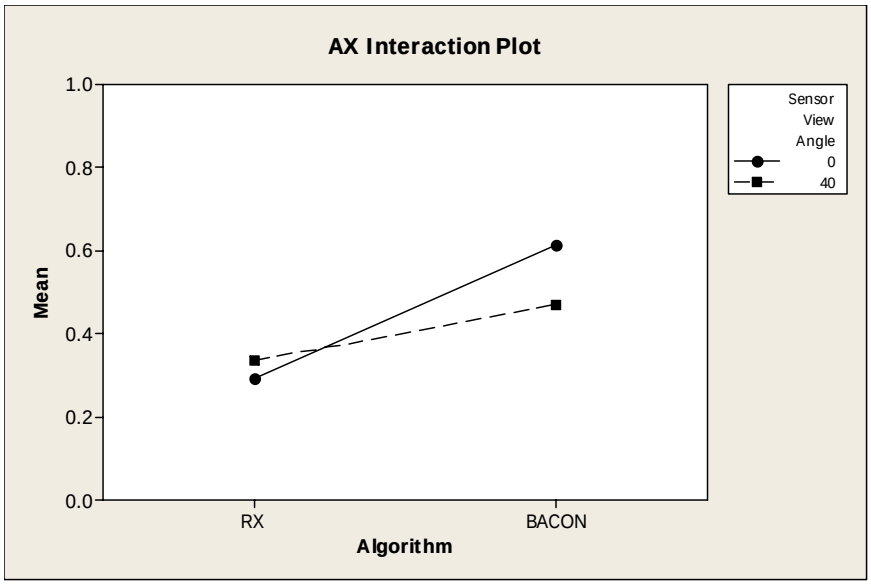


Figure 27. AX Interaction Plot for the Nested Design

RX – BACON Comparison #1

The first test involved plotting the maximum true positive fractions for each image-detector combination in Figure 28. The x-axis displays the 16 synthetic images from the experimental design. For each image, a lower case letter implies that the associated factor is set to its high level. The absence of a letter implies that the factor is set at its low level. It is clear to see from the plot that the BACON Method outperformed the RX Method for every image. In fact, it outperformed RX by 29.6% on average. A 95% confidence interval for the mean difference between the true positive fractions of the BACON and RX Methods is (0.247, 0.344).

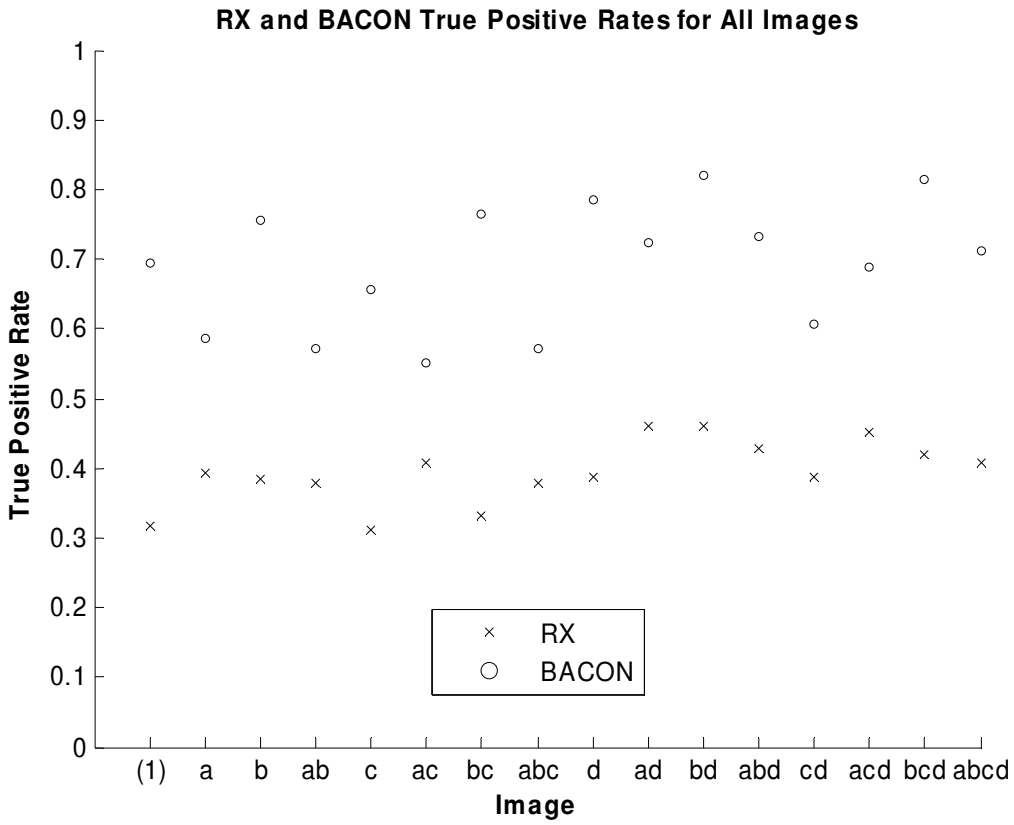


Figure 28. Comparison of RX and BACON TPFs for All Images

RX – BACON Comparison #2

The second test involved creating a new synthetic image. The models generated from the ANOVA and statistical testing were used to predict the best algorithm settings for an image created at the center of the image design space. However, since there is not an actual center level in DIRSIG for the Atmospheric Visibility factor, the low level (5 km) was arbitrarily used. Therefore, DIRSIG was used to created a 17^{th} synthetic image with a sensor view angle of 20^{o}, the time of day set to 1500, 5 km of atmospheric visibility, and targets in the scene scaled by three-fourths their natural size.

The model for the RX Method predicted a TPF of 0.3965 by retaining 4 principal components and using a window size of 25 pixels. A 95% prediction interval for RX's TPF is (0.36, 0.44). Meanwhile, the model for BACON predicted a TPF of 0.6962 for 3 clusters no matter if band aggregation is performed or not. A 95% prediction interval for BACON's TPF is (0.38, 1.01). Based on the predictions, BACON was expected to perform better than RX. Actual results from running the RX and BACON detectors on Image 17 with the "best" algorithm settings given the image factors generated TPFs of 0.4545 and 0.75, respectively. The predictions for each algorithm are summarized in Table 17 and Table 18. The OC Curves of the actual algorithm performance in Figure 29 illustrate that BACON dominates RX in the FPF region of 0 to 0.05.

Table 17. RX Predictions for Image 17

# of PCs Retained	*Window Size*	*Prediction*	*95% Prediction Interval*	
			Lower Bound	*Upper Bound*
4	17	0.2860	0.24	0.33
4	25	0.3965	0.36	0.44
10	17	0.2279	0.19	0.27
10	25	0.3696	0.33	0.41

Table 18. BACON Predictions for Image 17

Aggregate Bands	*# of Clusters*	*Prediction*	*95% Prediction Interval*	
			Lower Bound	*Upper Bound*
No	3	0.6962	0.38	1.01
No	4	0.4168	0.099	0.73
Yes	3	0.6962	0.38	1.01
Yes	4	0.4168	0.099	0.73

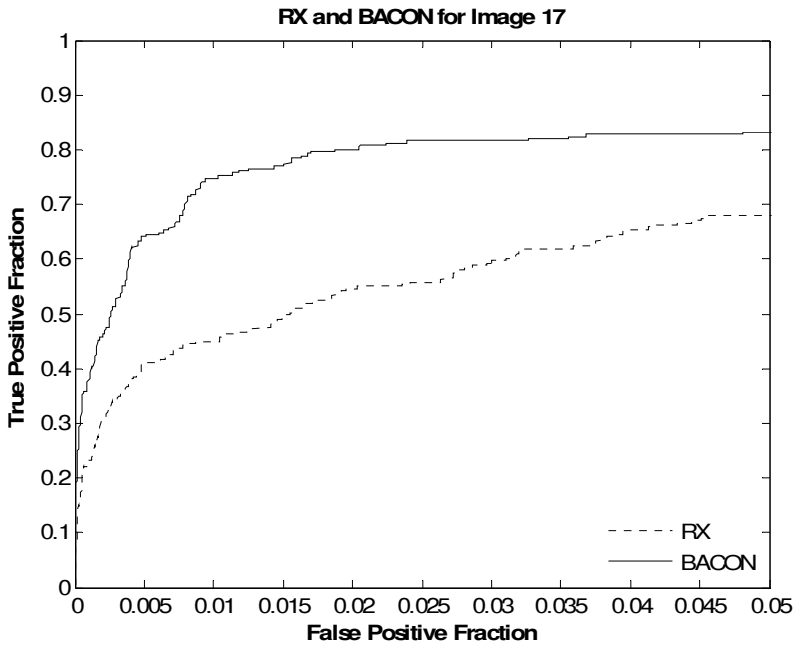


Figure 29. OC Curves for Image 17

The binary maps in Figure 30 display white pixels where the detectors declared an anomaly at a false positive rate of 0.01. The RX and BACON anomaly detectors perform well finding individual tanks, trucks, and missile carriers. However, BACON exceeds RX's performance in finding a large cluster of targets. This can be examined by observing the areas in the truth map labeled A and B. BACON does a very good job at detecting these anomalies while RX does not. This leads back to the hypothesis that the RX method performs poorly for large targets relative to the processing window size. The RX method also detects more of the natural scene anomalies, such as the trees, bushes, and vehicle tracks, as outliers in the data.

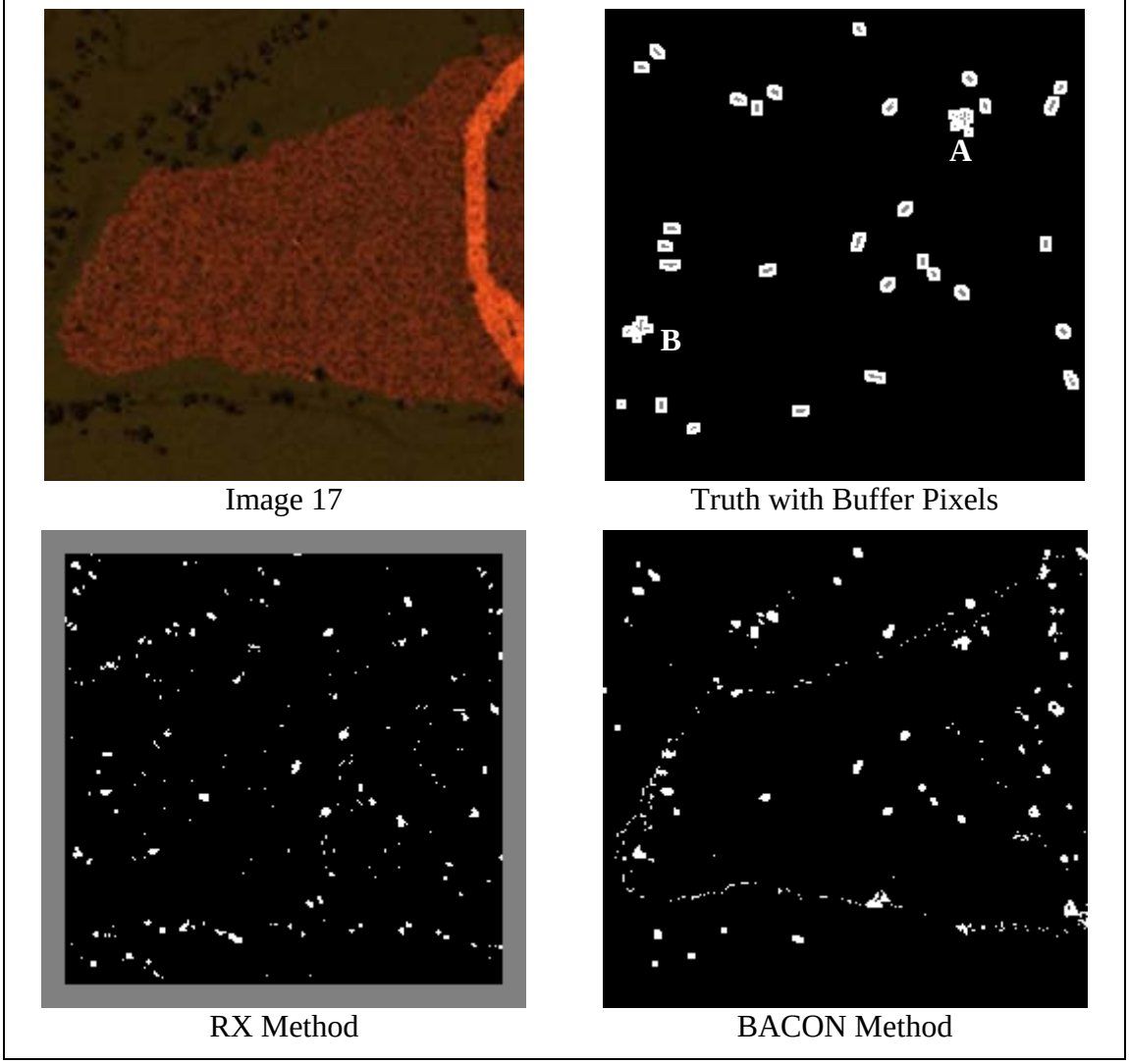


Figure 30. RX and BACON Results for Image 17

The previous two tests of this section were designed to compare the RX and BACON anomaly detectors head-to-head. The performance of the BACON algorithm, although not perfect, was far better than the performance of the RX algorithm. BACON functions very well in areas where RX performs poorly, such as in detecting large targets, or a cluster of targets. The conclusion that can be made from these tests is that, in the

case of the RX and BACON methods, the performance of the robust multivariate outlier detector surpassed that of the classical non-robust method. The main motivation for using robust multivariate outlier detectors lies in the fact that they are less sensitive to outlying observations in the data.

V. Discussion

Conclusions

The primary goal of this thesis was to conduct an examination into how differing values of sensor view angle, time of day, atmospheric visibility, and the size of the target in the scene affect the performance of hyperspectral anomaly detection algorithms. This was accomplished by utilizing the true positive fraction of the algorithms and employing a factorial experimental design. The main advantage of using an experimental design approach was that it allowed a thorough examination of both the main factors and their joint interactions with respect to the algorithm's performance. Also, the use and analysis of a nested factorial design allowed us to make a head-to-head comparison between the BACON and RX anomaly detectors.

In summary, DIRSIG was used to create a family of synthetic hyperspectral images in which to test the robustness of two anomaly detectors – the RX Method and the BACON Method. The interest of comparing these two algorithms originates from the fact that classical Mahalanobis distance-based outlier detectors, like the RX Method, are limited. These methods rely on non-robust covariance matrix estimates that are highly sensitive to outlying observations. The BACON Method, on the other hand, is a multivariate outlier detector that uses robust estimates of the mean and covariance.

Throughout the course of this thesis, it was discovered that BACON is a more robust hyperspectral anomaly detector than RX. The sensor's view angle, the time of day, the visibility in the atmosphere, and the size of the targets in the image were all

determined to be significant image factors in determining the true positive fraction of the RX Method. On the other hand, the sensor's view angle was the only image parameter that affected the performance of BACON. That implies that BACON performed equally well over differing values of time of day, atmospheric visibility, and target size. It was also illustrated that BACON's true positive fraction was approximately 30% greater than RX's.

It has been shown that the use of a factorial designed experiment is a very effective tool for testing algorithm performance. Furthermore, this experimental approach allows the investigator to study several different factors simultaneously. The output from this type of experimental design permits an easy identification of trends and allows statistical hypothesis testing to establish significance of the parameters on the algorithm's outcome.

The results of this thesis are useful and important in two main areas. First, as a proof of concept, we have demonstrated that the use of a factorial designed experiment is an excellent approach for simultaneously studying several in-scene factors and their impact on algorithm performance. Second, this thesis is a step towards determining robustness and superiority among hyperspectral image detectors, whether they are anomaly detectors, target detectors, or classification algorithms.

Contributions

This thesis has made several contributions to the field of hyperspectral remote sensing. For one, it has provided an innovative, new technique for determining the robustness of hyperspectral algorithms. The fusion of synthetic imagery and factorial

designed experiments allows the analyst to vary several parameters over many levels to resolve the main influences on algorithm performance. This will permit the analyst to make modifications to further improve their target/anomaly detectors. Second, it provides a means to compare the head-to-head performance of various algorithms. A nested factorial designed experiment is the only channel that allows an "apples to apples" comparison of algorithms.

The synthetic images created for this thesis were also used as testing and validation images for four other master's theses. Capt Yuri Taitano, whose thesis was mentioned earlier with regards to the RX method, used several DIRSIG images to validate the Iterative RX method. Capt Ryan Caulk and Capt Kevin Reyes each used the entire family of synthetic images to validate their new anomaly detection algorithms. Finally, Capt Jason Williams used a number of images to validate the robustness of several hyperspectral clustering algorithms.

Recommendations

The factorial designed experiments in this thesis have certainly proved to be effective. However, many of the algorithm performance trends witnessed in this study cannot be considered concrete conclusions without further testing. More scene parameters, as well as more levels of all parameters, need to be tested. More scenes and testing levels will result in the recognition of global trends with respect to algorithm performance. A random effects factorial designed experiment will allow conclusions to be developed with respect to a population of possible parameter levels. The testing that

was done in this thesis only allows us to draw conclusions about the levels that were tested.

Also, given that more images are tested at more parameter levels, a regression model could be developed for each hyperspectral algorithm tested. This would allow a user to input the image parameters of interest into a regression equation and subsequently be delivered the expected metric result.

Using DIRSIG to create synthetic hyperspectral images helps to provide a baseline for spectral algorithm performance. However, at the time, even though DIRSIG does a spectacular job at modeling the hyperspectral image chain, it is not quite as complex as the real world. The lack of “real-world-like” spectral and spatial variability of DIRSIG images provides a significant obstacle when conducting rigorous algorithm testing.

Appendix A. Responses for All Images Using the RX Method

				F: Window Size	-	-	+	+
				E: PCs Retained	-	+	-	+
Image	***A: View Angle***	***B: Time of Day***	***C: Atmospheric Visibility***	***D: Scale of Targets***				
2	-	-	-	-	0.2727	0.2323	0.3158	0.3053
9	+	-	-	-	0.2517	0.2867	0.3929	0.3857
13	-	+	-	-	0.2828	0.2172	0.3842	0.3053
12	+	+	-	-	0.2937	0.2587	0.3786	0.3643
15	-	-	+	-	0.2374	0.2273	0.2737	0.3105
11	+	-	+	-	0.2378	0.2797	0.3571	0.4071
16	-	+	+	-	0.2727	0.2172	0.3316	0.3053
8	+	+	+	-	0.2587	0.2727	0.3429	0.3786
1	-	-	-	+	0.2766	0.2008	0.3777	0.3863
5	+	-	-	+	0.3414	0.2876	0.4613	0.4586
4	-	+	-	+	0.2746	0.1352	0.4614	0.3541
6	+	+	-	+	0.2661	0.2097	0.4282	0.3923
3	-	-	+	+	0.2705	0.1988	0.3541	0.3863
10	+	-	+	+	0.3199	0.2796	0.442	0.4503
14	-	+	+	+	0.2541	0.1168	0.4206	0.3476
7	+	+	+	+	0.2527	0.2097	0.4088	0.3591

Response: True Positive Fraction (TPF) when False Positive Fraction (FPF) is Fixed at 0.01.

	A: View Angle	***B: Time of Day***	***C: Atmospheric Visibility***	***D: Scale of Targets***
Low	0 degrees	1200	5 km visibility	Half Scale
High	40 degrees	1800	23 km visibility	Full Scale

	E: PCs Retained	***F: Window Size***
Low	4	17
High	10	25

Appendix B. Responses for All Images Using the BACON Method

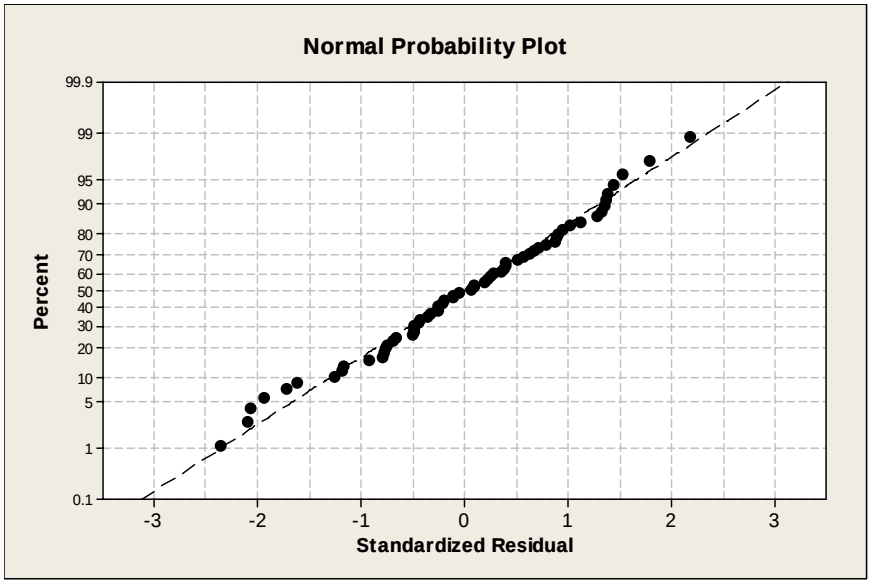
				F: # of Clusters	-	-	+	+
				E: Aggregate?	-	+	-	+
Image	***A: View Angle***	***B: Time of Day***	***C: Atmospheric Visibility***	***D: Scale of Targets***				
2	-	-	-	-	0.6600	0.6550	0.6950	0.6750
9	+	-	-	-	0.5850	0.5646	0.0408	0.0272
13	-	+	-	-	0.7550	0.7550	0.4050	0.3800
12	+	+	-	-	0.5714	0.5374	0.3197	0.2041
15	-	-	+	-	0.2200	0.3750	0.4300	0.6550
11	+	-	+	-	0.5510	0.5034	0.4626	0.4558
16	-	+	+	-	0.7650	0.7450	0.7550	0.7350
8	+	+	+	-	0.5714	0.4898	0.4014	0.4422
1	-	-	-	+	0.7863	0.7661	0.7762	0.7621
5	+	-	-	+	0.7236	0.6809	0.1482	0.2387
4	-	+	-	+	0.8206	0.8165	0.5242	0.5302
6	+	+	-	+	0.7337	0.7286	0.6080	0.3342
3	-	-	+	+	0.2520	0.2843	0.6069	0.5081
10	+	-	+	+	0.6884	0.6633	0.6307	0.6307
14	-	+	+	+	0.8145	0.8004	0.4577	0.4556
7	+	+	+	+	0.7111	0.6633	0.1206	0.0302

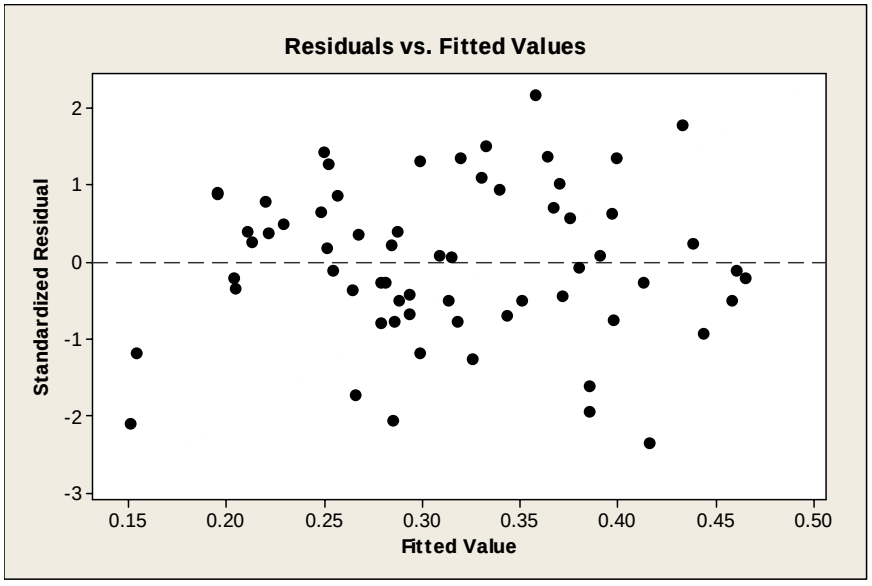
Response: True Positive Fraction (TPF) when False Positive Fraction (FPF) is Fixed at 0.01.

	A: View Angle	***B: Time of Day***	***C: Atmospheric Visibility***	***D: Scale of Targets***
Low	0 degrees	1200	5 km visibility	Half Scale
High	40 degrees	1800	23 km visibility	Full Scale

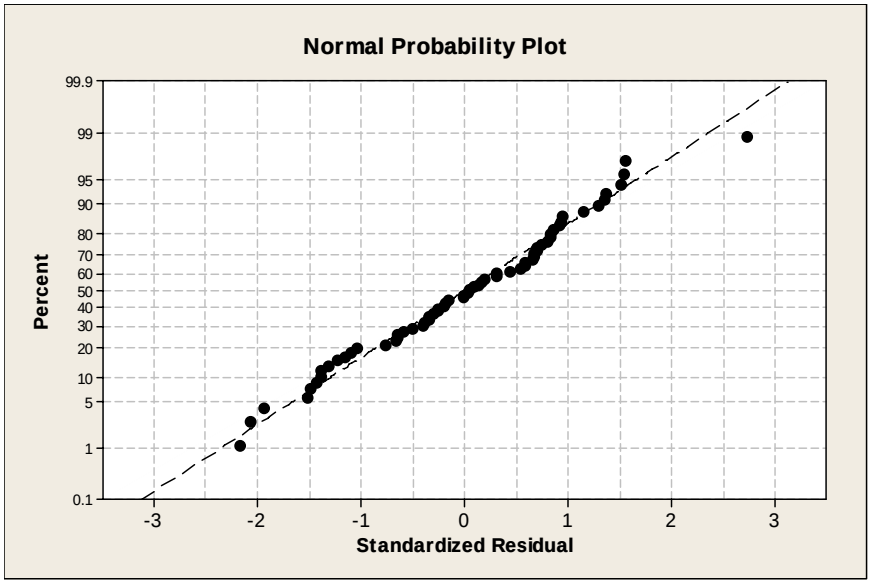
	E: Band Aggregation	***F: # of Clusters***
Low	N	3
High	Y	4

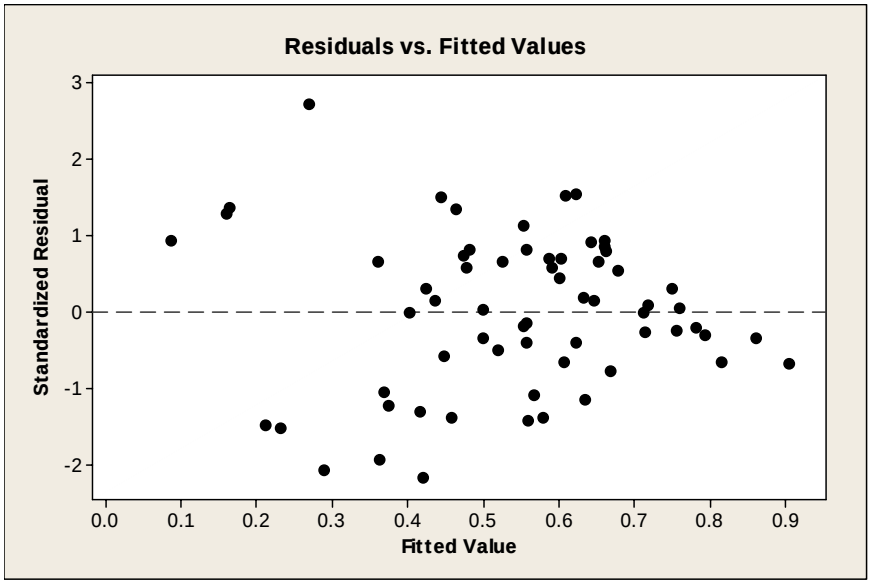
Appendix C. Model Diagnostic Plots for the RX Method





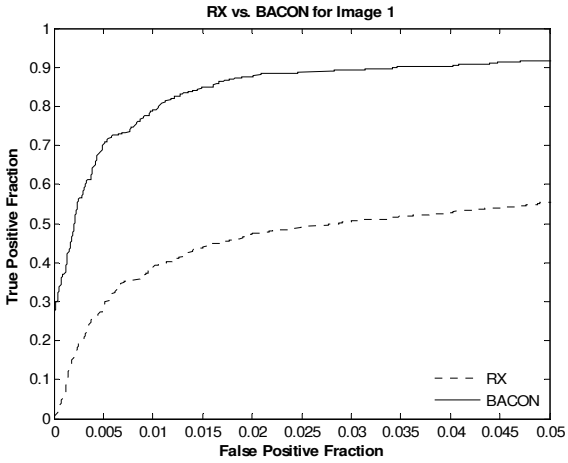
Appendix D. Model Diagnostic Plots for the BACON Method



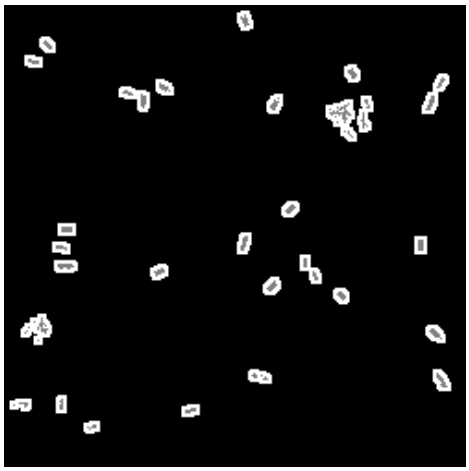


Appendix E. OC Curves and Binary Maps for All Images

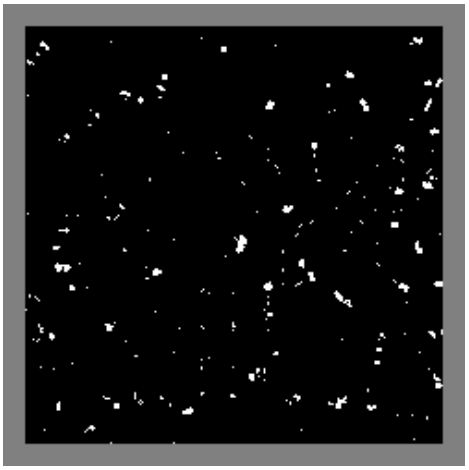
Image 1



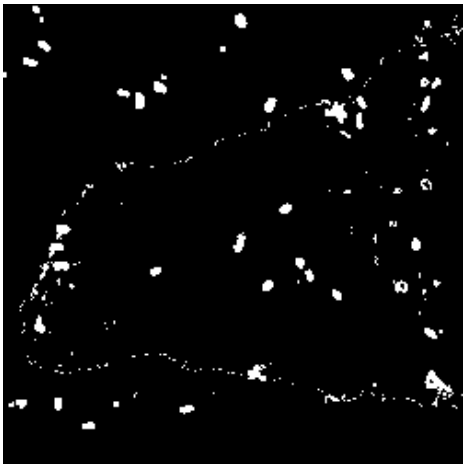
Truth



Truth with Buffer Pixels

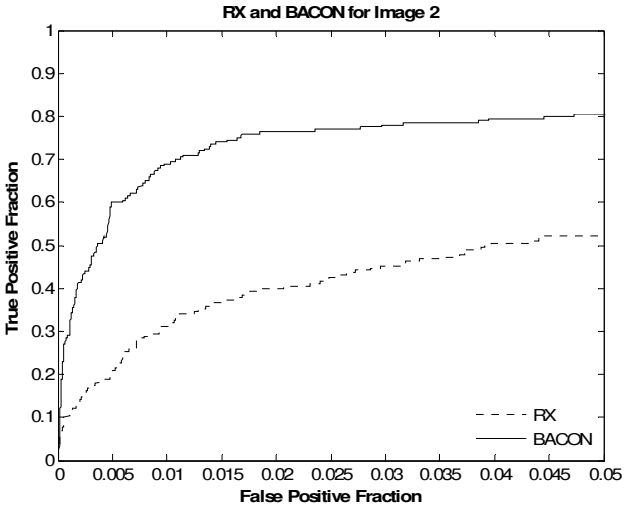


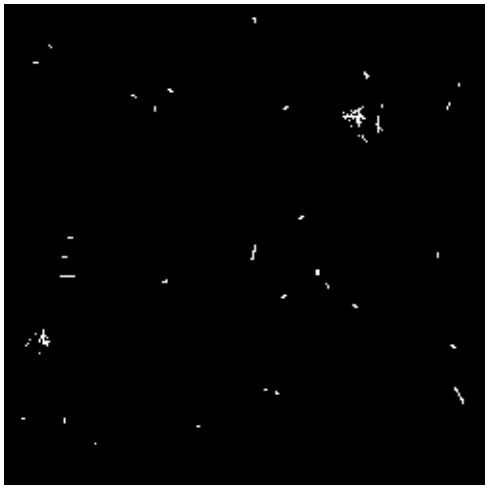
RX



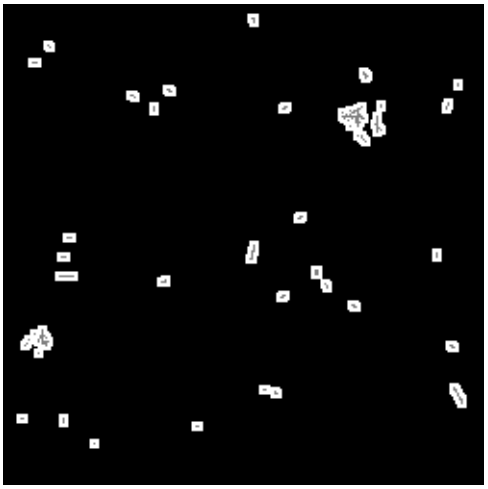
BACON

Image 2

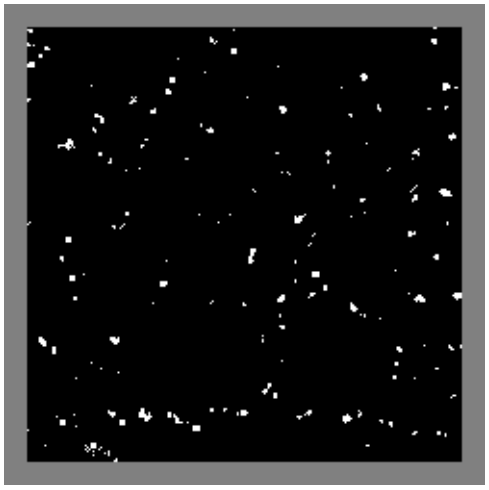




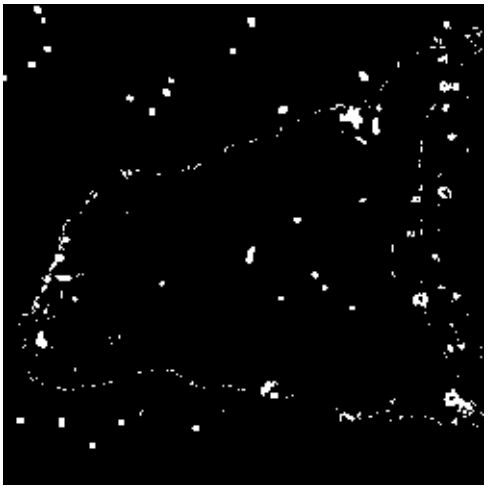
Truth



Truth with Buffer Pixels

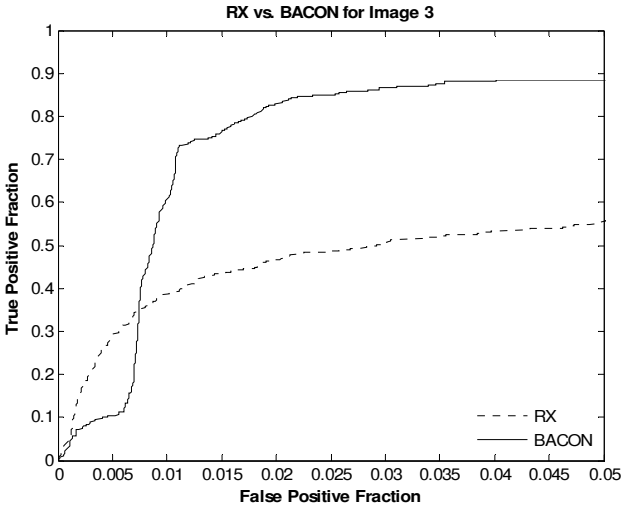


RX

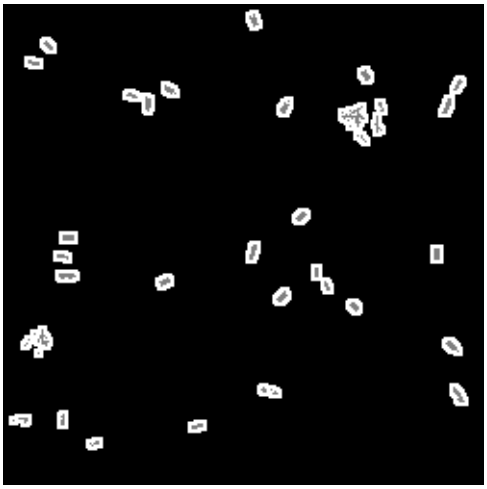


BACON

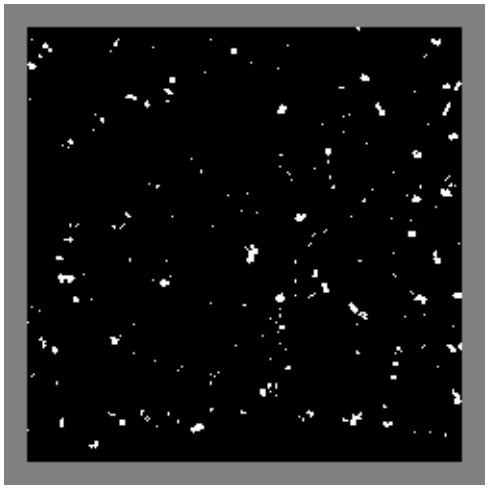
Image 3



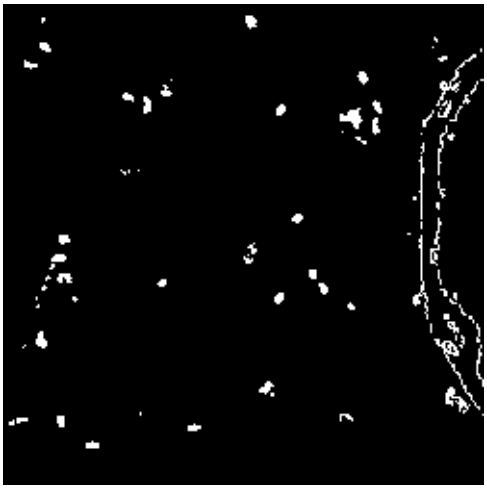
Truth



Truth with Buffer Pixels

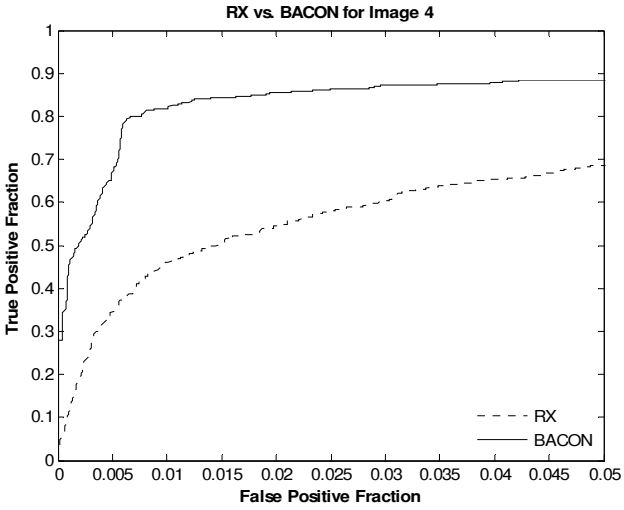


RX

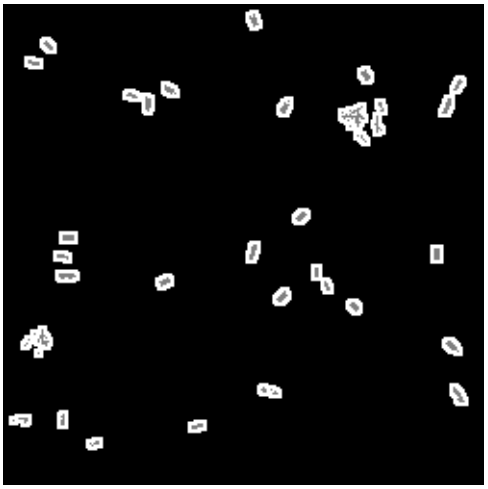


BACON

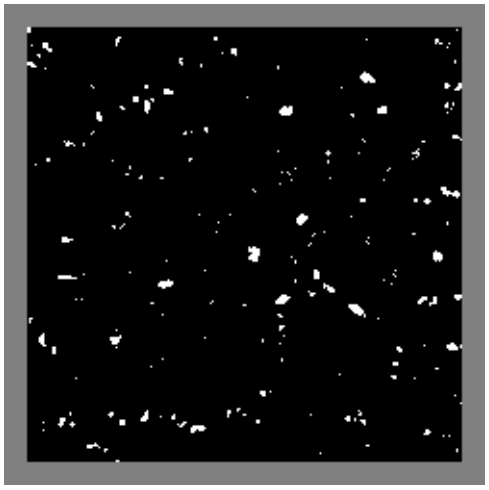
Image 4



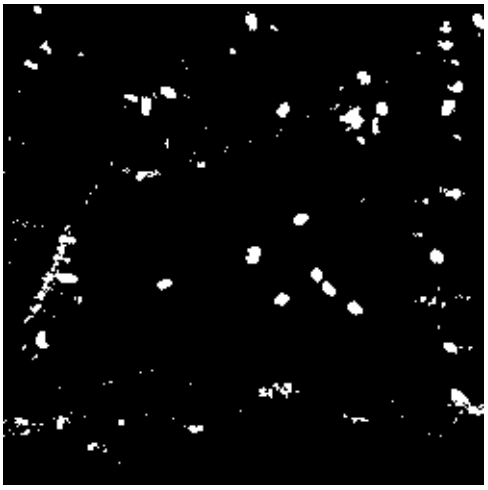
Truth



Truth with Buffer Pixels

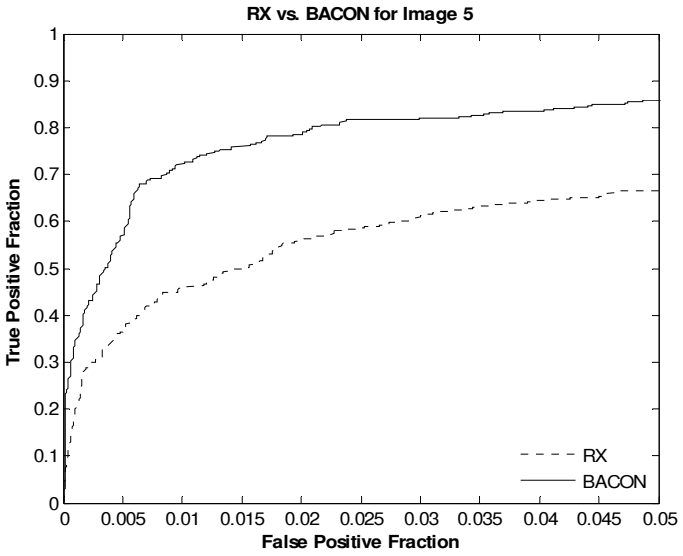


RX

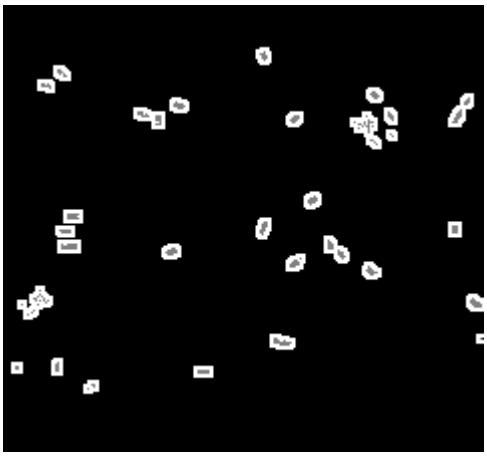


BACON

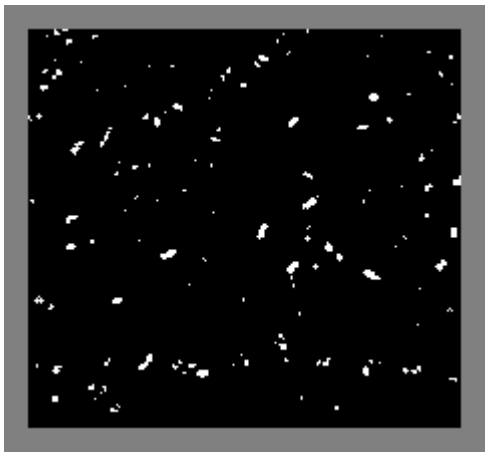
Image 5



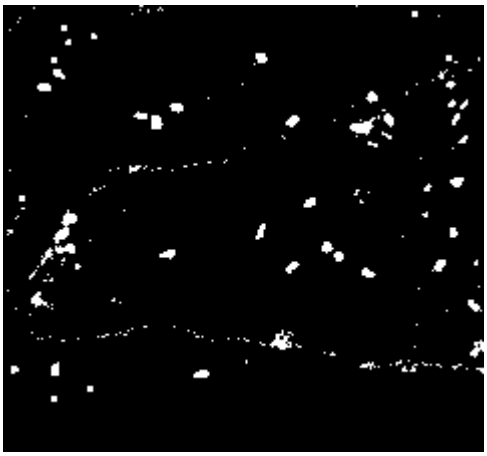
Truth



Truth with Buffer Pixels

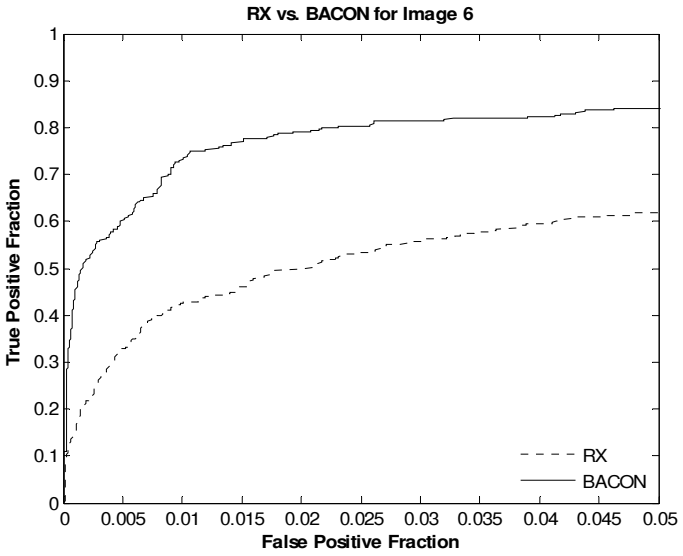


RX

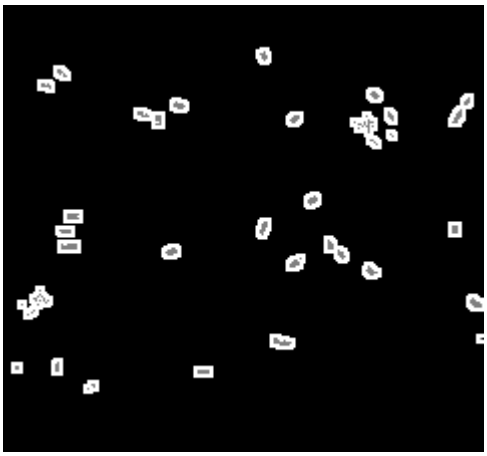


BACON

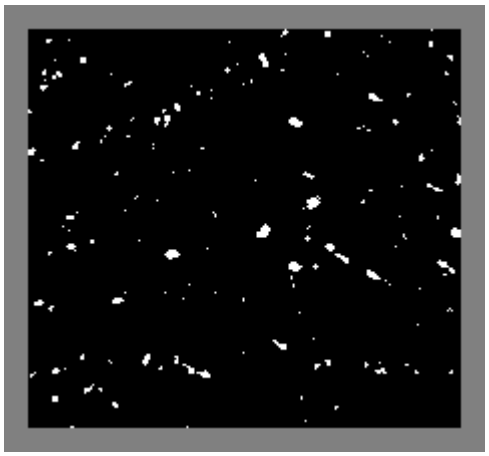
Image 6



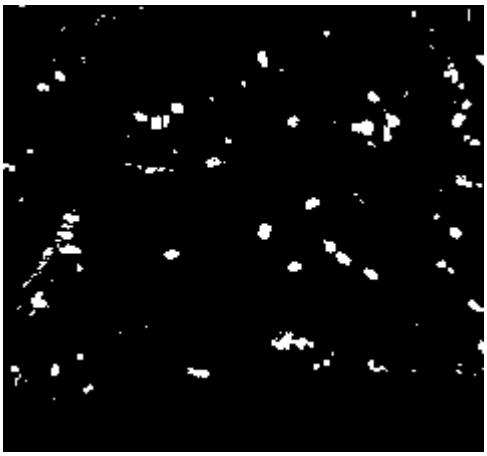
Truth



Truth with Buffer Pixels

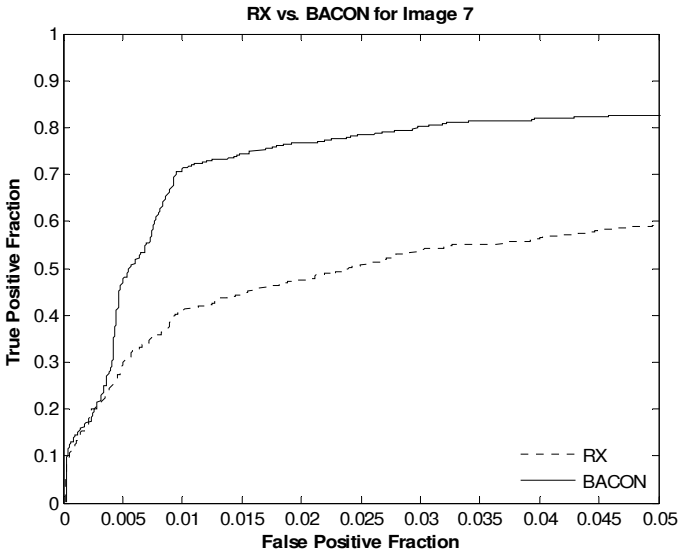


RX

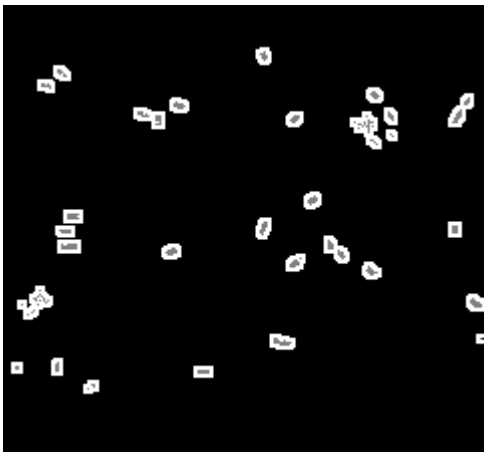


BACON

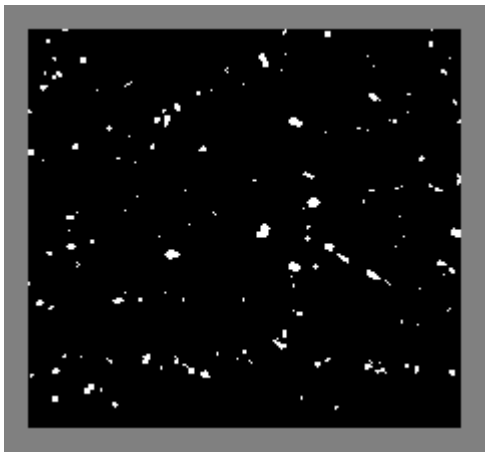
Image 7



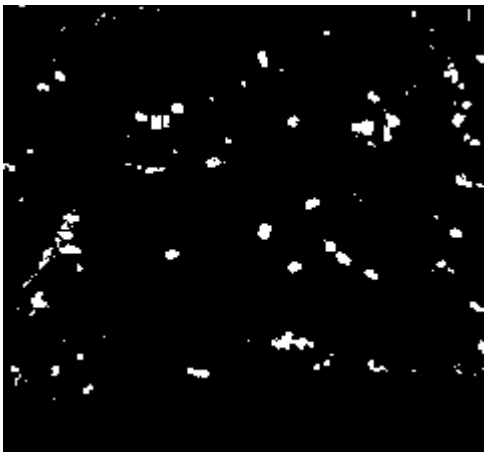
Truth



Truth with Buffer Pixels

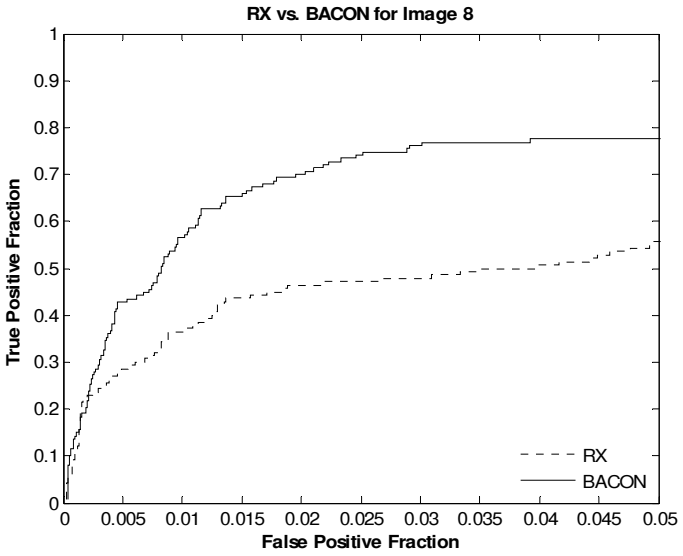


RX

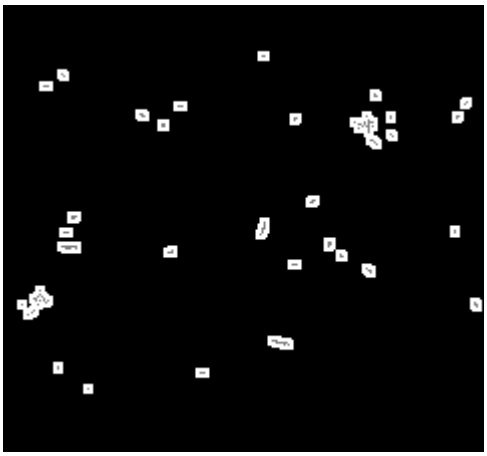


BACON

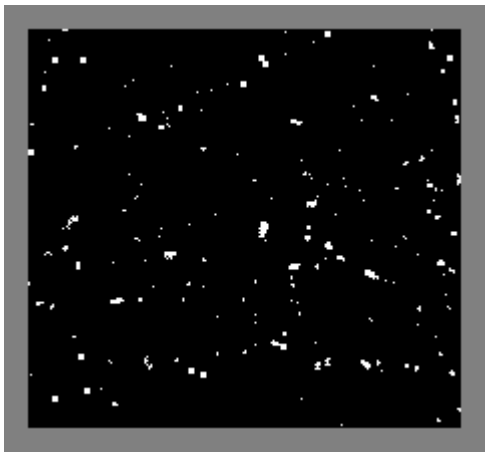
Image 8



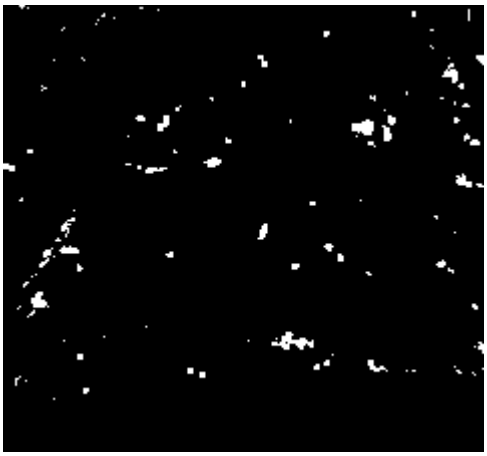
Truth



Truth with Buffer Pixels

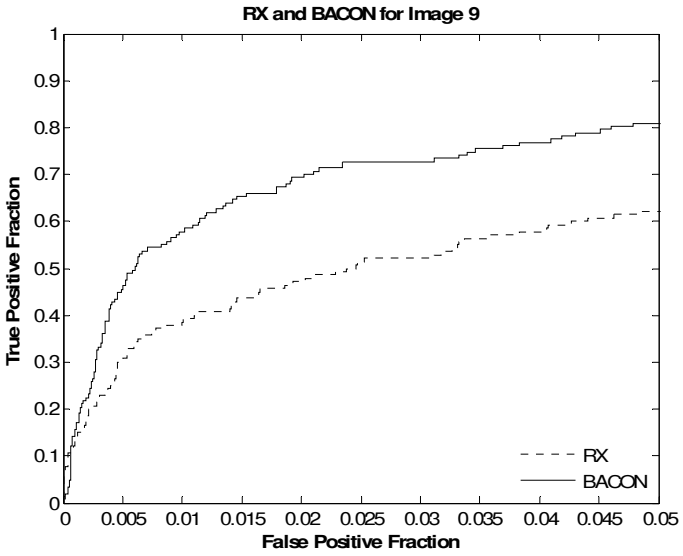


RX

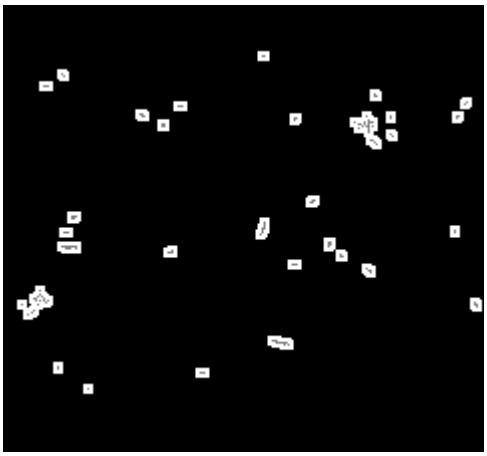


BACON

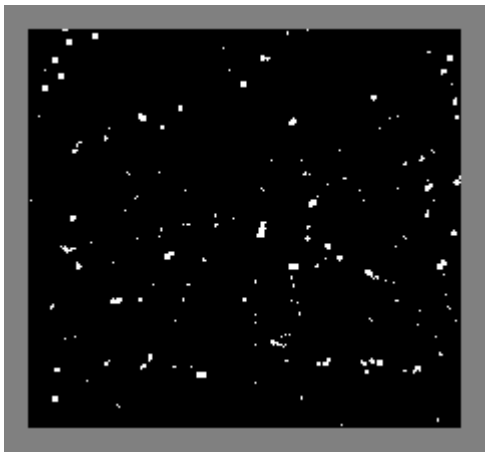
Image 9



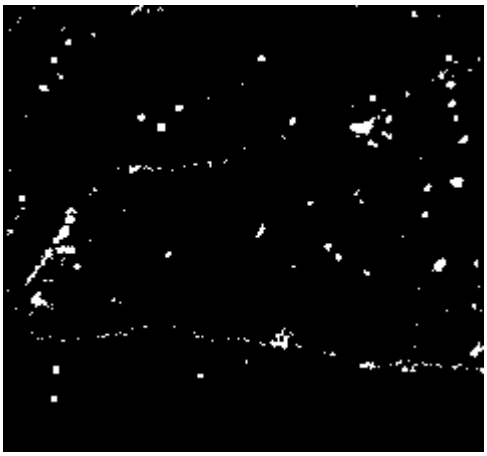
Truth



Truth with Buffer Pixels

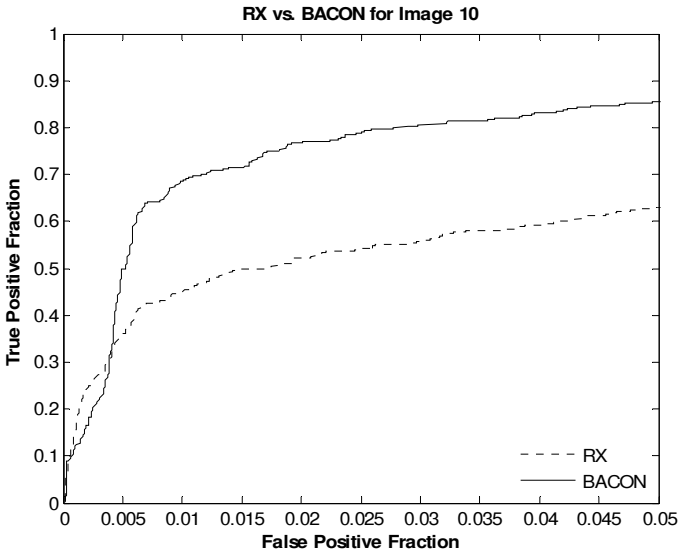


RX

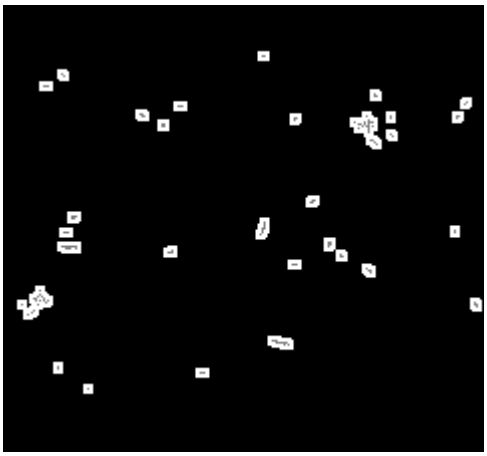


BACON

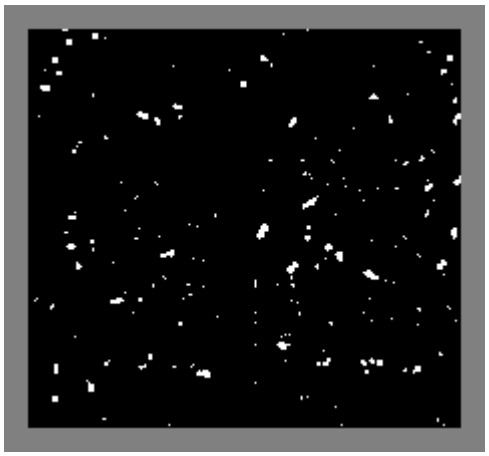
Image 10



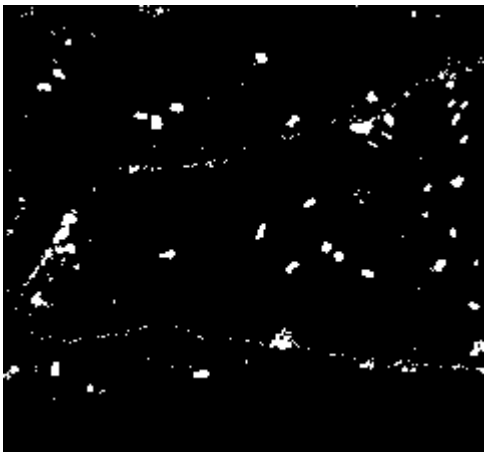
Truth



Truth with Buffer Pixels

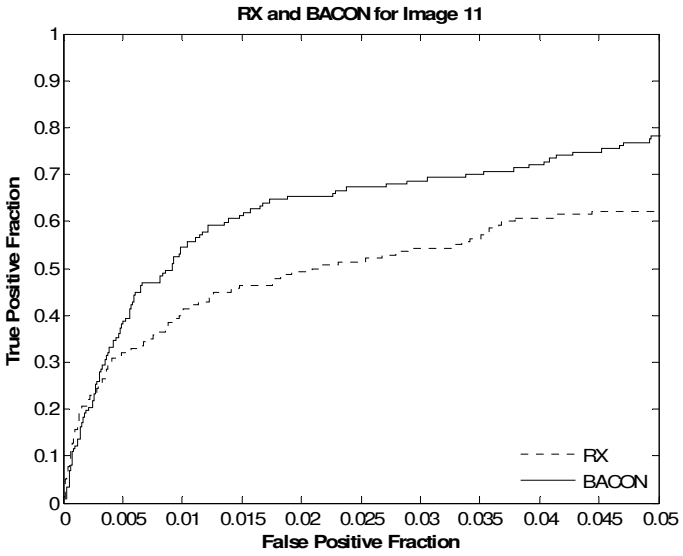


RX

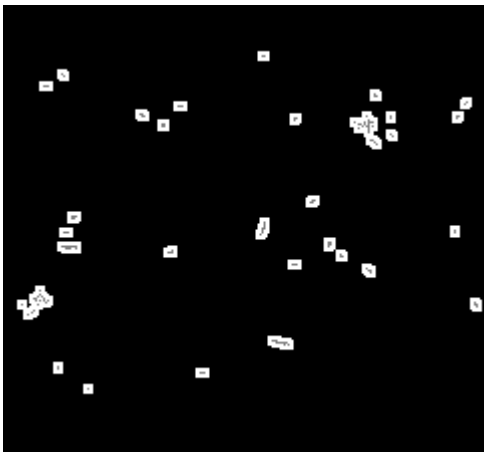


BACON

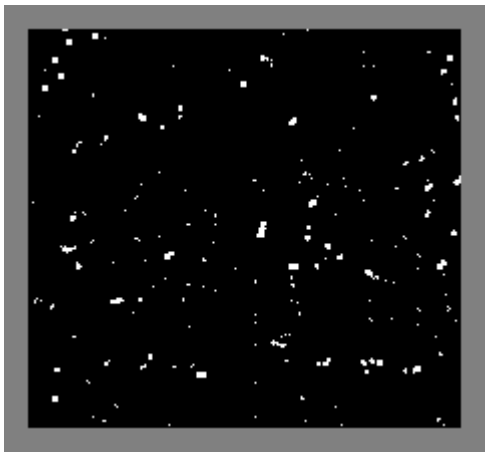
Image 11



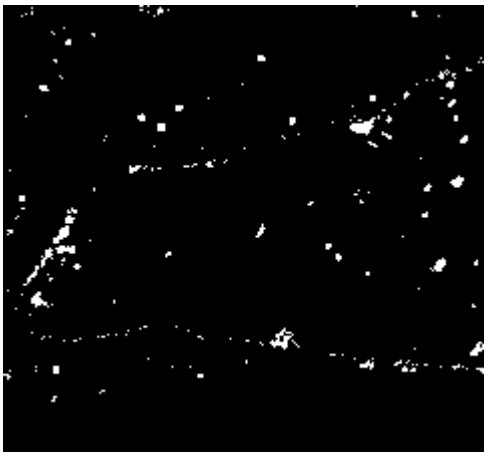
Truth



Truth with Buffer Pixels

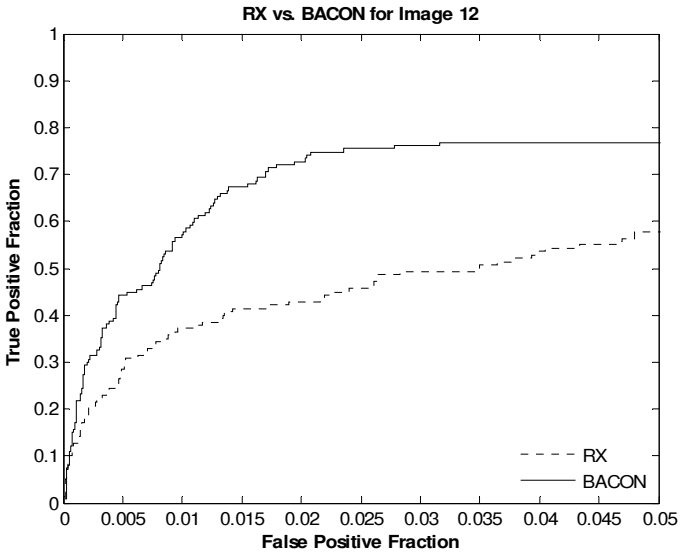


RX

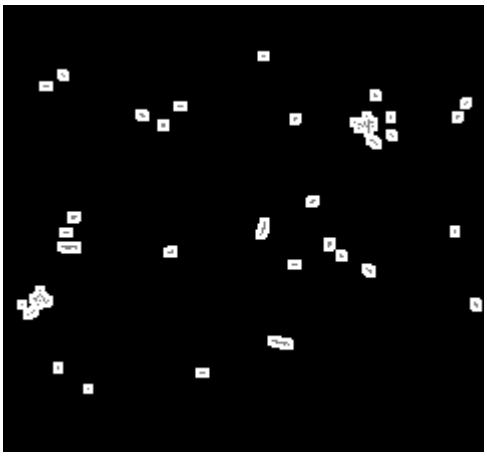


BACON

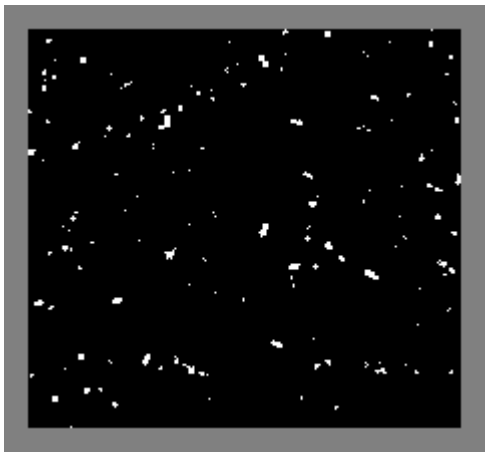
Image 12



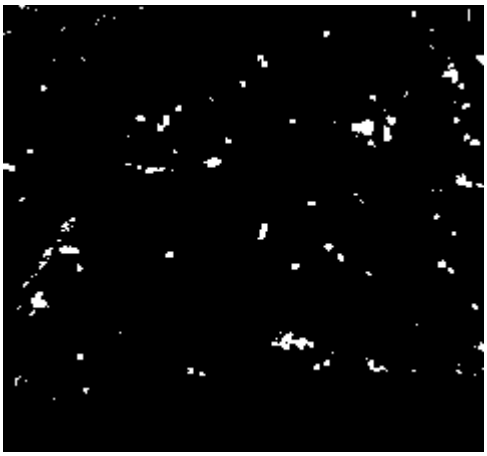
Truth



Truth with Buffer Pixels

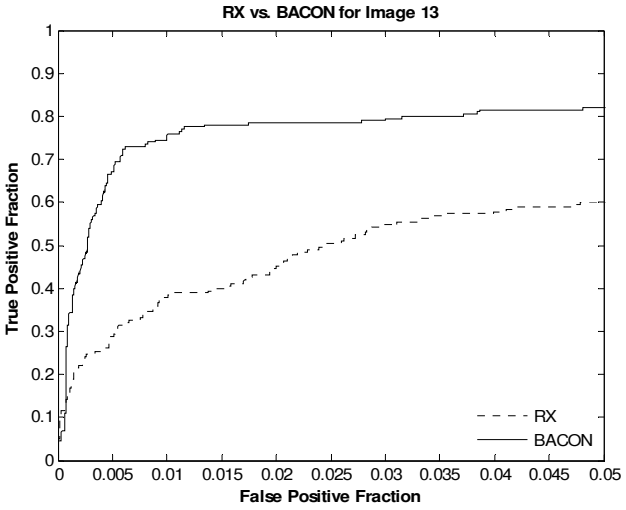


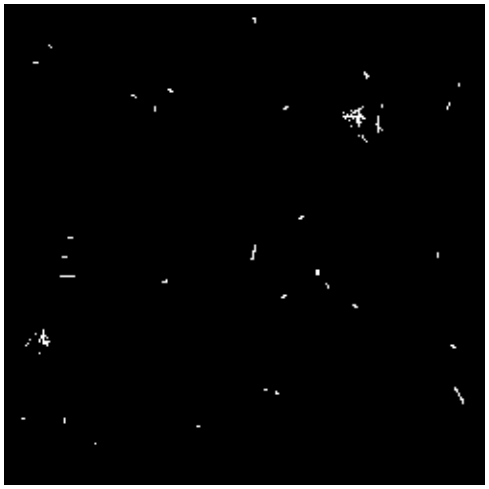
RX



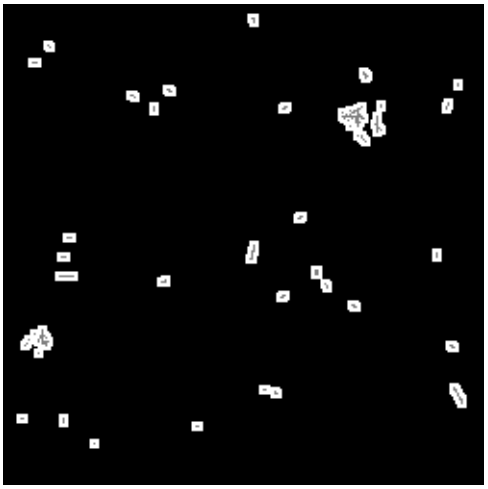
BACON

Image 13

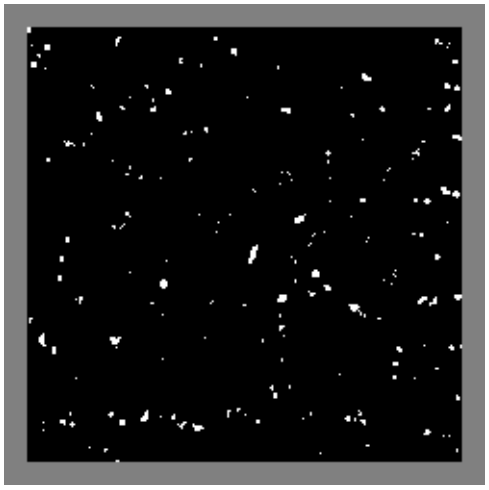




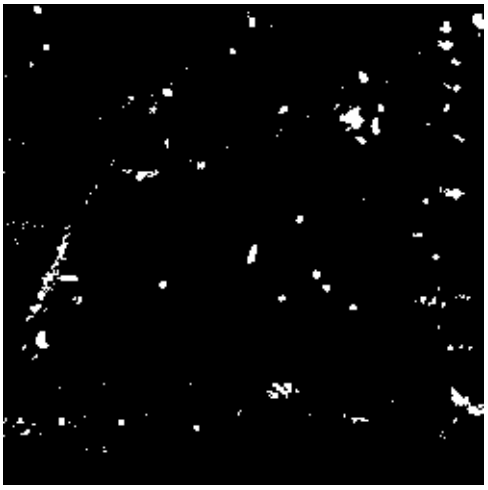
Truth



Truth with Buffer Pixels

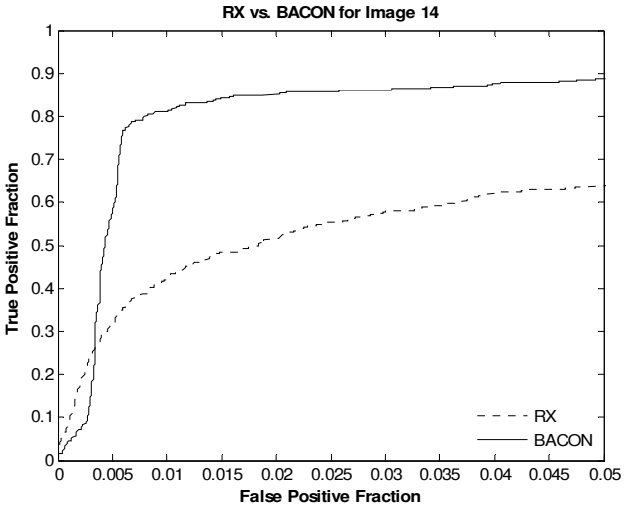


RX

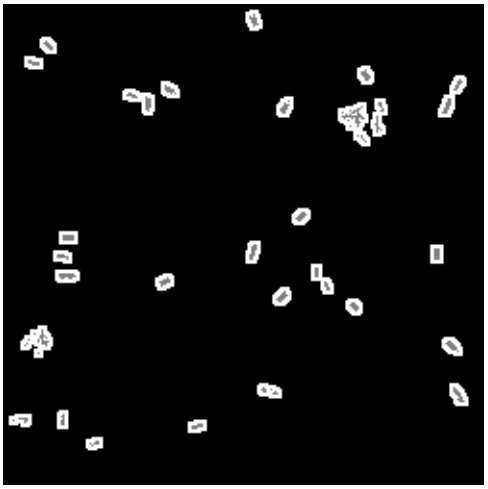


BACON

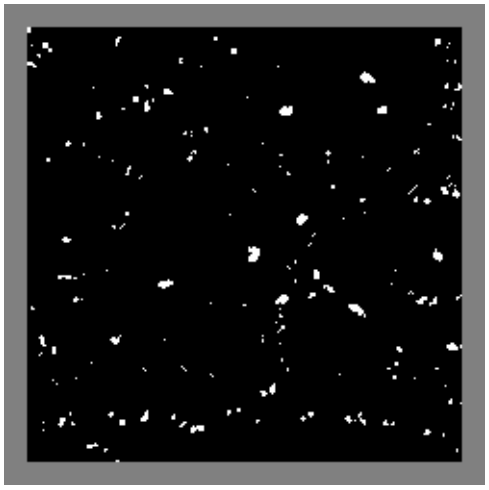
Image 14



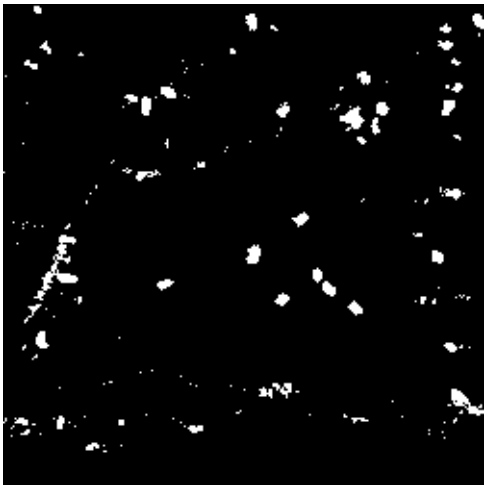
Truth



Truth with Buffer Pixels

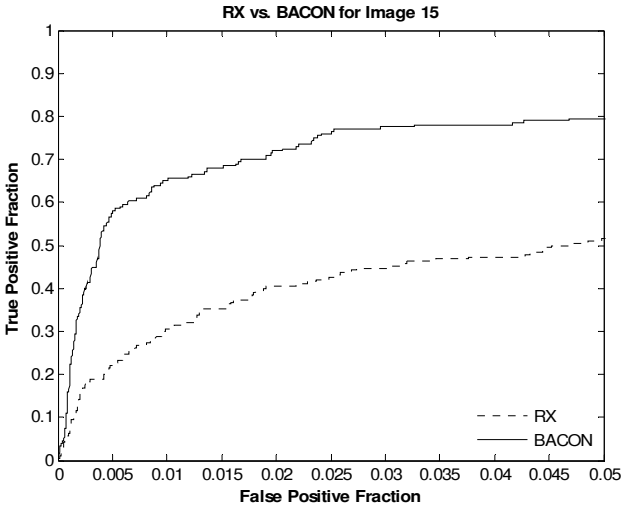


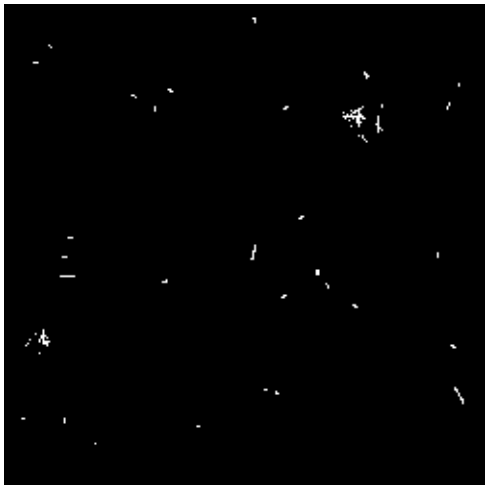
RX



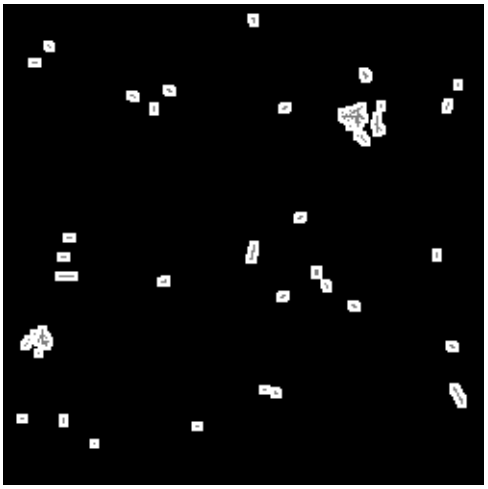
BACON

Image 15

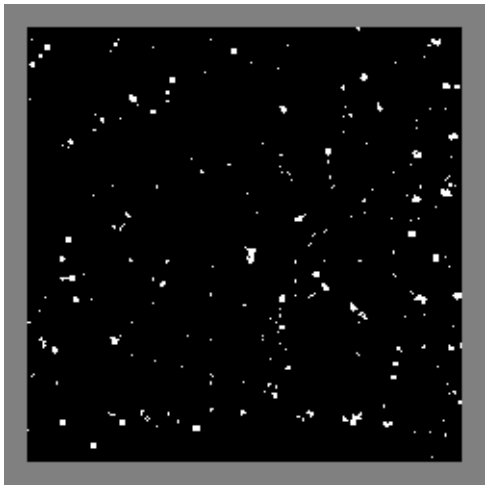




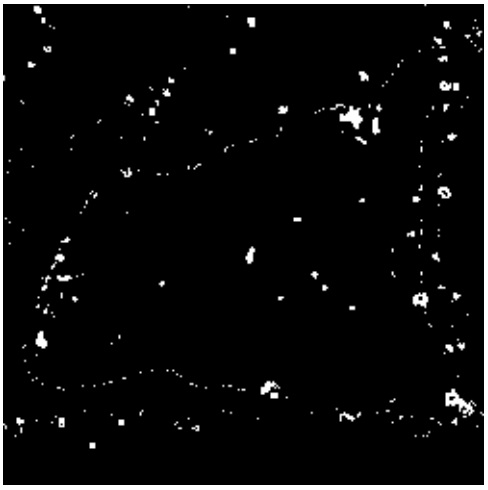
Truth



Truth with Buffer Pixels

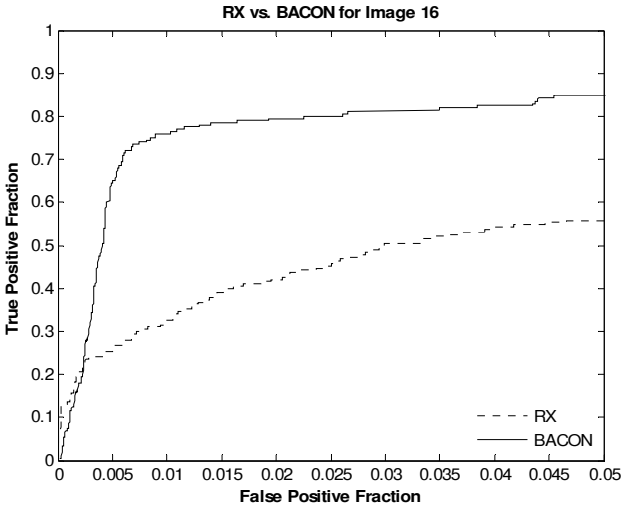


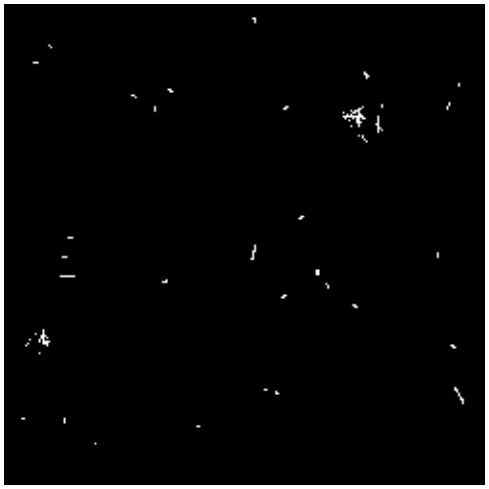
RX



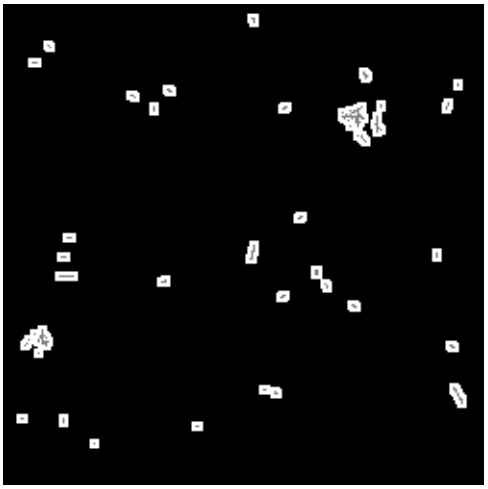
BACON

Image 16

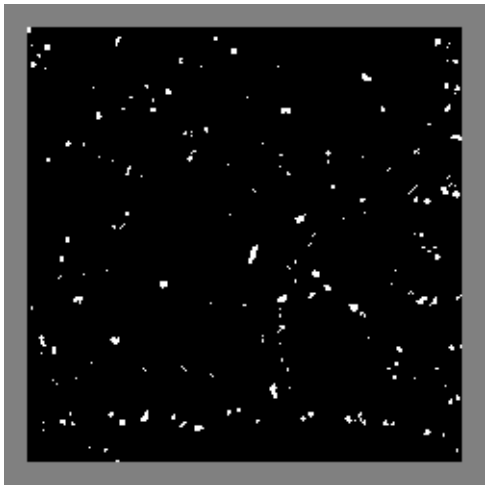




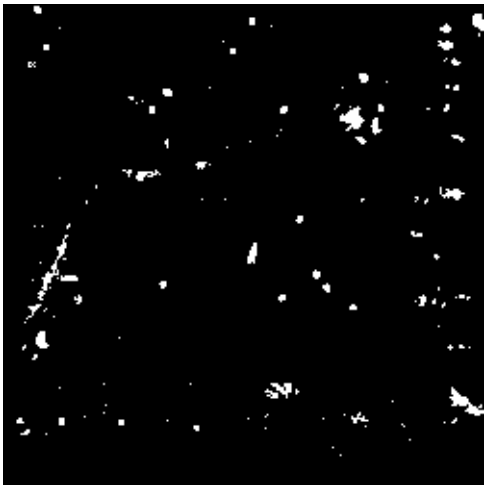
Truth



Truth with Buffer Pixels



RX



BACON

Bibliography

Billor, Nedret, Ali S. Hadi, and Paul F. Velleman. "BACON: Blocked Adaptive Computationally Efficient Outlier Nominators." *Computational Statistics & Data Analysis*, 34: 279-298 (2000).

Caulk, Ryan F. *Outlier Detection in Hyperspectral Imagery Using Closest Distance to Center with Ellipsoidal Multivariate Trimming.* MS thesis, AFIT/GOR/ENS/07-02. School of Operation Sciences, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2007.

Chiang, Shao-Shan and Chein-I Chang. "Discrimination Measures for Target Classification." IEEE 2001 International Geoscience and Remote Sensing Symposium (IGARSS). 4:1871-1873 (2001).

Chang, Chein-I and Shao_Shan Chiang. "Anomaly Detection and Classification fro Hyperspectral Imagery." *IEEE Transactions on Geoscience and Remote Sensing*. 40: 1314-1325 (2002).

Chang, Chein-I, Weimin Liu, and Chein-Chi Chang. "Discrimination and Identification for Subpixel Targets in Hyperspectral Imagery." *2004 International Conference on Image Processing (ICIP)*, 5:3339-3342 (October 2004).

Dillon, William R. and Matthew Goldstein. *Multivariate Analysis: Methods and Applications*. New York: John Wiley and Sons, 1984.

DIRS – Digital Imaging and Remote Sensing Laboratory. *The DIRSIG User's Manual*. Rochester Institute of Technology. 2006.

Ientillucci, Emmett J. and Scott D. Brown. "Advances in Wide Area Hyperspectral Image Simulation." *Proc SPIE*, 5075: 110-121 (2003).

Landgrebe, David. "Hyperspectral Image Analysis." *IEEE Signal Processing Magazine*, 19: 17-28 (January 2002).

Manolakis, Dimitris and Gary Shaw. "Detection Algorithms for Hyperspectral Imaging Applications." *IEEE Signal Processing Magazine*, 19:29-43 (January 2002).

Montgomery, Douglas C. *Design and Analysis of Experiments*. New York: John Wiley and Sons, 2005.

Myers, Raymond H. and Douglas C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: John Wiley and Sons, 2002.

Reed, Irving S. and Xiaoli Yu. "Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution." *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 38: 1760-1770 (1990).

Reyes, Kevin B. *Outlier Detection in Hyperspectral Imagery Using Principal Component Analysis.* MS thesis, AFIT/GOR/ENS/07-22. School of Operation Sciences, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2007.

Schott, John R. "Combining Image Derived Spectra and Physics Based Models for Hyperspectral Image Exploitation." *Proceedings of the 29th Applied Imagery Pattern Recognition Workshop (AIPR 2000)*, 15-24 (October 2000).

-----. *Remote Sensing: The Image Chain Approach*. New York: The Oxford University Press, 1997.

Shaw, Gary A. and Hsiao-hua K. Burke. "Spectral Imaging for Remote Sensing," *Lincoln Laboratory Journal*, 14:3-28 (2003).

Shaw, Gary and Dimitris Manolakis. "Signal Processing for Hyperspectral Image Exploitation." *IEEE Signal Processing Magazine*, 19:12-16 (January 2002).

Smetek, Timothy E. and Kenneth W. Bauer. "Finding Hyperspectral Anomalies Using Multivariate Outlier Detection." Wright-Patterson AFB OH: Air Force Institute of Technology, October 2006.

Smetek, Timothy E. "Hyperspectral Imagery Target Detection Using Improved Anomaly and Invariant Detection Methods." Dissertation Prospectus. Wright-Patterson AFB OH: Air Force Institute of Technology, November 2005.

Taitano, Yuri P. *Hyperspectral Imagery Target Detection Using the Iterative RX Detector*. MS thesis, AFIT/GOR/ENS/07-25. School of Operation Sciences, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2007.

Wackerly, Dennis D., William Mendenhall, and Richard L. Scheaffer. *Mathematical Statistics with Applications*. Pacific Grove CA: Duxbury, 2002.

Williams, Jason P. *Robustness of Multiple Clustering Algorithms on Hyperspectral Images.* MS thesis, AFIT/GOR/ENS/07-27. School of Operation Sciences, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2007.

Yuhas, Roberta H., Alexander F.H. Goetz, and Joe W. Boardman. "Discrimination Among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm." *Summaries of the Third Annual JPL Airborne Geoscience Workshop*. 147-149. 1992.

Vita

Captain Joseph P. Bellucci graduated from Spring Hill High School in Columbia, Tennessee. He entered undergraduate studies at Cumberland University in Lebanon, Tennessee where he graduated with a Bachelor of Science degree in Mathematics and Physics in May 2001. He was commissioned through the Air Force Officer Training School at Maxwell AFB in June 2002.

His first assignment was at Hanscom AFB where he served as a hyperspectral data analysis manager and acquisitions officer in June 2002. In August 2005, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology. Upon graduation, he will report to Squadron Officer School at Maxwell AFB enroute to his assignment to ACC/A9 at Langley AFB.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
22-03-2007	**Master's Thesis**	Jun 2006 - Mar 2007

4. TITLE AND SUBTITLE
Improved Hyperspectral Image Testing Using Synthetic Imagery and Factorial Designed Experiments

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Bellucci, Joseph, P., Captain, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GOR/ENS/07-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Intentionally left blank.

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

The goal of any remote sensing system is to gather data about the geography it is imaging. In order to gain knowledge of the earth's landscape, post-processing algorithms are developed to extract information from the collected data. The algorithms can be intended to classify the various ground covers in a scene, identify specific targets of interest, or detect anomalies in an image. After the design of an algorithm comes the difficult task of testing and evaluating its performance. Traditionally, algorithms are tested using sets of extensively ground truthed test images. However, the lack of well characterized test data sets and the significant cost and time issues associated with assembling the data sets contribute to the limitations to this approach. This thesis uses a synthetic image generation model in cooperation with a factorial designed experiment to create a family of images with which to rigorously test the performance of hyperspectral algorithms. The factorial designed experimental approach allowed the joint effects of the sensor's view angle, time of day, atmospheric visibility, and the size of the targets to be studied with respect to algorithm performance. A head-to-head performance comparison of the two tested spectral processing algorithms was also made.

15. SUBJECT TERMS
Algorithm Testing, Anomaly Detection, Design of Experiments, Factorial Design, Hyperspectral Imagery, Nested Experimental Design, Synthetic Imagery, Target Detection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kenneth W. Bauer (ENS)
U	U	U	UU	112	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636, ext 4328