
Doctoral Dissertations

Student Theses and Dissertations

Fall 2007

Detection algorithms for spatial data

Spandan Tiwari

Follow this and additional works at: https://scholarsmine.mst.edu/doctoral_dissertations



Part of the [Electrical and Computer Engineering Commons](#)

Department: Electrical and Computer Engineering

Recommended Citation

Tiwari, Spandan, "Detection algorithms for spatial data" (2007). *Doctoral Dissertations*. 2153.
https://scholarsmine.mst.edu/doctoral_dissertations/2153

This thesis is brought to you by Scholars' Mine, a service of the Missouri S&T Library and Learning Resources. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

DETECTION ALGORITHMS FOR SPATIAL DATA

by

SPANDAN TIWARI

A DISSERTATION

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI-ROLLA

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2007

S. Agarwal, Advisor

D. C. Wunsch, II

R. P. Roe

R. H. Moss

S. Madria

© 2007

Spandan Tiwari

All Rights Reserved

ABSTRACT

This dissertation addresses the problem of anomaly detection in spatial data. The problem of landmine detection in airborne spatial data is chosen as the specific detection scenario. The first part of the dissertation deals with the development of a fast algorithm for kernel-based non-linear anomaly detection in the airborne spatial data. The original Kernel RX algorithm, proposed by Kwon et al. [2005a], suffers from the problem of high computational complexity, and has seen limited application. With the aim to reduce the computational complexity, a reformulated version of the Kernel RX, termed the Spatially Weighted Kernel RX (SW-KRX), is presented. It is shown that under this reformulation, the detector statistics can be obtained directly as a function of the centered kernel Gram matrix. Subsequently, a methodology for the fast computation of the centered kernel Gram matrix is proposed. The key idea behind the proposed methodology is to decompose the set of image pixels into clusters, and expediting the computations by approximating the effect of each cluster as a whole. The SW-KRX algorithm is implemented for a special case, and comparative results are compiled for the SW-KRX vis-à-vis the RX anomaly detector. In the second part of the dissertation, a detection methodology for buried mine detection is presented. The methodology is based on extraction of color texture information using cross-co-occurrence features. A feature selection methodology based on Bhattacharya coefficients and principal feature analysis is proposed and detection results with different feature-based detectors are presented, to demonstrate the effectiveness of the proposed methodology in the extraction of useful discriminatory information.

ACKNOWLEDGMENTS

My first word of thanks goes to my advisor Dr. Sanjeev Agarwal, who has been my mentor and advisor for over six years, ever since the beginning of my graduate studies. He has been a source of constant guidance and support, not only in my academic pursuits, but in other aspects of my life as a graduate student. This dissertation work would not have been possible without his help with problem solving ideas, mathematical derivations and programming, and I thank him for that.

My sincere thanks to Drs. R. H. Moss, D. C. Wunsch, R. P. Roe and S. Madria for agreeing to serve on my committee. They have been most generous in accommodating my tight deadlines in their busy schedules. I would also like to thank the Countermine Division of Night Vision and Electronic Sensors Directorate (NVESD), for providing the financial support, data and the opportunity to work on this project. In particular, my special thanks to Mr. Anh Trang from NVESD for his guidance, discussions and support during the course of this project. Also, I would like to thank Mr. Shareq Soofi from the ARIA Lab at UMR, for his help on numerous occasions during the preparation of this dissertation.

My heartfelt thanks go to my parents who have been a constant source of support and guidance throughout my life, and especially during the highs and lows of graduate school. Finally, I would like to thank my wife, Garima Pathak, for her selfless and unwavering love, support, and patience all these years.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGMENTS	iv
LIST OF ILLUSTRATIONS	viii
LIST OF TABLES	x
SECTION	
1. INTRODUCTION	1
1.1. SPATIAL DATA	2
1.2. LANDMINE DETECTION AND AIRBORNE SPATIAL DATA	4
1.3. OVERVIEW OF THE DISSERTATION	6
2. RELATED WORK	10
2.1. DETECTION IN AIRBORNE SPATIAL DATA: REVIEW	10
2.2. NON-UNIFORMLY SAMPLED SPATIAL DATA: COMPARISON AND REVIEW	14
3. SPATIALLY WEIGHTED KERNEL-RX (SW-KRX)	19
3.1. RX ANOMALY DETECTOR	20
3.2. KERNEL-RX ANOMALY DETECTOR	22
3.3. SPATIALLY WEIGHTED KERNEL-RX	24
3.3.1. Modified Detector Statistics	24
3.3.2. Selection of Weights	30
3.4. RESULTS: SIMULATED DATA	31
4. FAST APPROXIMATE COMPUTATION OF KERNEL GRAM MATRIX	36
4.1. COMPUTATIONAL BURDEN IN SW-KRX	36
4.2. SPECTRAL-SPATIAL CLUSTERING	36
4.2.1. Clustering Techniques: Review	37
4.2.2. Watershed-based Clustering	40
4.3. FAST COMPUTATION OF KERNEL GRAM MATRIX	42
4.4. FAST APPROXIMATE CENTERING OF KERNEL GRAM MATRIX: DIAGONAL ELEMENTS	47
4.5. COMPUTATIONAL ACCURACY AND APPROXIMATIONS	50

4.5.1. Determination of α	51
4.5.2. Determination of β	51
4.5.3. Spatial Far-field Approximation	52
4.5.4. Spectral Far-field Approximation	53
4.5.5. Spectral Zeroth Order Approximation	54
4.5.6. Spectral Near-Field Approximation.	55
4.6. MULTIPLE PIXEL TARGET.....	56
4.7. COMPUTATIONAL COMPLEXITY	56
5. SW-KRX: RESULTS AND DISCUSSION.....	61
5.1. DATA DESCRIPTION	61
5.2. DETECTION PERFORMANCE: SW-KRX (SPECIAL CASE) VS. RX.....	63
5.2.1. Results: Green Band.....	65
5.2.2. Results: Multispectral.....	67
5.3. COMPUTATIONAL SPEED: SW-KRX VS. KERNEL-RX	71
5.4. SW-KRX: COMPUTATIONAL SPEED-UP VIS-À-VIS FULL GRAM MATRIX COMPUTATION	74
5.5. CONCLUSION.....	76
6. BURIED MINE DETECTION USING CO-OCCURRENCE TEXTURE FEATURES	78
6.1. BURIED MINE DETECTION: REVIEW	78
6.2. DATA OVERVIEW	81
6.3. CO-OCCURRENCE TEXTURE FEATURES	85
6.3.1. Gray-Level Co-occurrence Texture Features	85
6.3.2. Cross Co-occurrence Texture Features	87
6.4. DISCRIMINATORY FEATURE SELECTION	89
6.4.1. Discriminatory Texture Feature Reduction.....	89
6.4.2. Feature Selection Using Principal Feature Analysis	93
6.5. FEATURE BASED DETECTORS	95
6.5.1. Vegetation Mask Detector.....	98
6.5.2. Matched Filter Detector.....	96
6.5.3. AND Fusion Detector.....	97
6.5.4. Feature-based SW-KRX Detector	98

6.6. RESULTS AND DISCUSSION	101
6.7. CONCLUSION	109
7. CONCLUSIONS AND FUTURE WORK	110
APPENDICES	
A. DETAILED DERIVATION OF THE KERNEL-RX DETECTOR STATISTICS	113
B. RELATIONSHIP BETWEEN COVARIANCE AND KERNEL GRAM MATRIX IN HIGH DIMENSIONAL SPACE	115
C. CENTERING OF KERNEL GRAM MATRIX	118
D. RELATIONSHIP BETWEEN THE KERNEL GRAM MATRIX AND THE WEIGHTED CENTERED KERNEL GRAM MATRIX	120
BIBLIOGRAPHY	122
VITA	135

LIST OF ILLUSTRATIONS

	Page
Figure 1.1. Airborne Mine and Minefield Detection Scenario	6
Figure 1.2. Typical Multispectral Segment.....	7
Figure 3.1. Relative Sizes and Shapes of Various Masks Used in RX Detector.	21
Figure 3.2. Simulated Patch for Gaussian Distributed Data.	32
Figure 3.3. Comparative Detection Performance for Various Detectors for Gaussian Distributed Data.....	33
Figure 3.4. Simulated Patch for Mixed Background in 80:20 Proportion.	34
Figure 3.5. Comparative Detection Performance for Various Detectors for Mixed Background with Different Proportions.	35
Figure 5.1. Clearly Visible Surface Mine Signatures.	62
Figure 5.2. Poorly Visible Surface Mine Signatures.	63
Figure 5.3. Typical Frame with False Alarms Along Edges.....	64
Figure 5.4. Typical Frame with Edge-Suppressed False Alarms.....	65
Figure 5.5. Detection Performance of SW-KRX vis-à-vis RX Detector for Green Band Data for Dataset 1.	66
Figure 5.6. Detection Performance of SW-KRX vis-à-vis RX Detector for Green Band Data for Dataset 2.	66
Figure 5.7. Detection Performance of SW-KRX vis-à-vis RX Detector for Multispectral (4-Band) Data for Dataset 1.	67
Figure 5.8. Detection Performance of SW-KRX vis-à-vis RX Detector for Multispectral (4-Band) Data for Dataset 2.	68
Figure 5.9. Example Frame with Targets in Non-Homogeneous Background.....	70
Figure 5.10. Typical Frames with False Alarms.....	70
Figure 6.1. Typical Data Segment (RGB Bands) Containing Buried Mine Signatures. ..	82
Figure 6.2. Examples of Clearly Distinguishable Buried Mine Signatures (RGB Band). 82	82
Figure 6.3. Examples of Poorly Distinguishable Buried Mine Signatures (RGB Band).. 82	82
Figure 6.4. Typical False Alarm Patches (RGB Bands).	83
Figure 6.5. Typical Data Segment (NIR Band) Containing Buried Mine Signatures.	83
Figure 6.6. Examples of Clearly Distinguishable Buried Mine Signatures (NIR Band).. 83	83
Figure 6.7. Examples of Poorly Distinguishable Buried Mine Signatures (NIR Band).. 84	84
Figure 6.8. Typical False Alarm Patches (NIR Band).	84
Figure 6.9. Kernel Density Estimates of the PDF for the Three Selected Features for Mine (red) and Non-mine (blue) Patches from Dataset 1.....	91

Figure 6.10. Bhattacharyya Coefficient Values for Different Features for Dataset 1..	92
Figure 6.11. Example 1 of a Feature Image as a Composite RGB Image.	95
Figure 6.12. Example 2 of a Feature Image as a Composite RGB Image.	95
Figure 6.13. Example 1 of a Feature Image as Composite RGB Image from Dataset 1. .	99
Figure 6.14. Example 2 of a Feature Image as Composite RGB Image from Dataset 2.	100
Figure 6.15. RX Detector on Energy-RG Feature Image from Dataset 1.....	102
Figure 6.16. Vegetation Based Masking of RX Detector Statistics.....	102
Figure 6.17. Comparison of Detection Performance of the Matched Filter Detector for Different Number of Representative Mine Signatures, vis-à-vis Multiband RX Anomaly Detector.....	104
Figure 6.18. Comparison of Detection Performance of the AND Fusion Detector, vis-à-vis Multiband RX Anomaly Detector.	105
Figure 6.19. Comparison of Detection Performance of the Feature-based SW-KRX Detector, vis-à-vis Multiband RX Anomaly Detector.	106
Figure 6.20. Detection Performance of the Vegetation Mask Detector vis-à-vis Multiband RX Anomaly Detector for Dataset 1.....	107
Figure 6.21. Detection Performance of the Vegetation Mask Detector vis-à-vis Multiband RX Anomaly Detector for Dataset 1.....	108

LIST OF TABLES

	Page
Table 5.1. Average Computation Time per Pixel for Full Kernel RX and Fast Implementation, for Single Band Data..	73
Table 5.2. Average Computation Time per Pixel for Full Kernel RX and Fast Implementation, for Multispectral Data.....	73
Table 5.3. Computational Speed-up via Taylor Series Approximation, for Single Band Data..	75
Table 5.4. Computational Speed-up via Taylor Series Approximation, for Multispectral Data..	75

1. INTRODUCTION

The rapid advances in the science and spur of technological innovations has enabled us to collect massive amounts of data, like never before in the history of humankind. Technology has not only enabled us to collect large amounts of data, relatively cheaply, in the traditional fields like geology, astronomy and remote sensing; but it has also fostered unconventional data driven fields like genomics, bioinformatics, biometrics and hyperspectral image processing. Especially, the advent of low cost sensor technology has revolutionized the way data is thought about and utilized. The ever burgeoning amounts of data have opened up opportunities for innovation and novel applications, which could not be fathomed fifty years ago. At the same time, it has also brought with it, new challenges related to its meaningful handling, storage and processing. David Donoho in his lecture during the AMS conference on Math Challenges in the 21st Century (held in celebration of the historic Hilbert's lecture in 1900) reverberated Tukey's views from sixty years ago, and stressed the need of utilizing mathematics for data analysis [Donoho, 2000]. He emphasized the need for new mathematical ideas and techniques for data analysis in light of the fact that most of the current techniques for data analysis were developed sixty years ago, and the incredible improvements in computation speeds have already pushed these to their limits.

Data can be classified based on the independent variable(s) in the data like time-series, spatial, spatio-temporal and others. In this work, spatial data, and processing techniques thereof, are considered. Spatial data or signals can be defined broadly as the data where the independent variables are the spatial dimensions. Spatial data has been gathered and used since centuries, with the earliest manifestations appearing in the form of data maps giving information regarding the geography, landforms and weather. In the modern world, spatial data can be seen in several classical fields like geology, soil science, mining, meteorology, and material science. For example, in mining, ore grade samples are collected over a mining block and analyzed so as to estimate the ore grade. Similarly, in meteorology, temperature, pressure and precipitation data is collected over a region to predict short term and long term weather trends. Soil scientists try to map the soil properties of a region based on a small number of soil samples at known locations

throughout the region. For instance, soil pH in water, mineral content, and soil-water infiltration are of particular interest to agriculturalists, who are interested in issues like fertility and water top-soil runoff. Note that in the fields mentioned above, the data is typically irregularly sampled.

Another important field where spatial data is used extensively is remote sensing. In remote sensing, earth orbiting satellites typically gather reflectance data for the electromagnetic waves at different frequencies like visible frequencies and infrared. It is an efficient means for collecting large amounts of spatial data in a relatively short time. This data is typically gathered in the form of regular sampled images, where each pixel essentially depicts the reflectance, or some other property, integrated over a small rectangle on the earth. Remote sensing data is used for numerous applications like predicting weather patterns, agriculture, land cover and mineral distribution. With long strides being taken in the area of sensor development, remote sensing is growing at a fast pace, with ever new spatial datasets coming up for analysis. Along with the conventional applications, remote sensing is increasingly being used for military applications, especially with the advent of high resolution multispectral and hyperspectral imaging sensors.

1.1. SPATIAL DATA

A generic data location in an n -dimensional Euclidean space can be defined as $\mathbf{x} \in \mathcal{R}^n$, and the corresponding datum at location \mathbf{x} as $\mathbf{F}(\mathbf{x})$. Then varying the location over an indexed set $X \subset \mathcal{R}^n$ and collecting the data in a set gives us a multivariate random field:

$$\{\mathbf{F}(\mathbf{x}) \mid \mathbf{x} \in X\} \quad (1.1)$$

A specific realization of the random field from Eqn. (1.1) gives us a spatial dataset, and is denoted as: $\mathbf{f}(\mathbf{x}), \forall \mathbf{x} \in X$. In this dissertation, two-dimensional spatial data, i.e., $n = 2$, is considered. Note that although this model is termed random field data model, the dataset itself does not have to be random and can be deterministic (or sampled from a deterministic function).

Spatial data can be classified broadly into three categories based on the underlying spatial location model used to define the random field [Cressie, 1991], as follows.

Geostatistical Data: In terms of spatial location, the spatial index \mathbf{x} can vary continuously over the subset $X \subset \mathcal{R}^n$, i.e., the locations at which the data is collected can be continuously varying and randomly scattered (in no specific order). This category of spatial data is most commonly seen in areas like mining, geology and soil science, where the soil samples are collected at random (but known) locations and then used for estimation and prediction of various properties like ore grade, soil pH, and composition.

Lattice Data: In case of lattice data, the spatial index \mathbf{x} can vary over a countable set of spatial sites at which data is observed. Although the term lattice gives rise to a picture of regularly spaced points in space, Cressie [1991] includes irregularly spaced lattice data into the same category. Thus, there are two types of lattice data: regular lattice data and irregular lattice data. However, the common factor between the two is that both models are supplemented with neighborhood information, i.e., each spatial site is connected to a countable set of sites called neighbors. The neighborhood of a site can be defined differently for different problems. Once the neighborhood is defined for the lattice, it becomes a graph with each site being a node, connected to its neighbors by edges. This graph theoretic formalism has given rise to several models and processing based thereon, for lattice data. One of the most common example of lattice data is that of an image, where every pixel is a spatial site with an intensity (or color) value, connected to say its four nearest neighbors. The image itself can come from different sources, like remote sensing (large scale) data, where each pixel is assumed to represent the center of the rectangle on the earth it covers; or medical imaging (small scale) where each pixel may represent the transmission property of a small area of the body.

Point Patterns: Point patterns are spatial data where the important variable is the spatial location of the “event” itself, and not the data recorded at the spatial location. In other words, the random field is completely defined by the set of spatial locations itself $\{\mathbf{x} \mid \mathbf{x} \in X \subset \mathcal{R}^n\}$ and the value of the function $\mathbf{f}(\mathbf{x})$ may be secondary (or can be assumed to be a constant over the entire set). The set X is a random set or more specifically a spatial point process and its realization is called a spatial point pattern.

Spatial point patterns are good models for “events” occurring at random locations, where the goal is to model or predict the location of the event itself. Some major areas for point pattern data are seismology (modeling and prediction of earthquakes) and plant population studies. It is noted that the third category of spatial data can sometimes be converted to one of the other two categories, by associating a feature or property, like number of neighbors, or mean distance from neighbors, as the value of the function $f(\mathbf{x})$.

The first two categories of spatial data are of particular interest to us. This dissertation presents algorithms for detection in spatial data. It considers the problem of detection of landmines in airborne multispectral image data. These spatial datasets fall broadly into the second category of lattice data. However, in general, problems in mine and minefield detection use concepts from the third category of point patterns, and entail the treatment of the data in a similar fashion.

1.2. LANDMINE DETECTION AND AIRBORNE SPATIAL DATA

Detection and remediation of landmines and minefields has been an area of intensive research owing to its significance from the humanitarian demining and tactical countermine perspectives. Effective detection of these landmines, which are typically spread over wide areas, is a challenging problem due to the inherent variability in the nature of the mines (shape, size, material) and terrain (vegetation, soil type, geography). Several data acquisition and detection systems have been proposed for this problem. These systems can be broadly classified into two categories: ground-based and airborne systems. Ground-based systems are based on technologies such as metal detectors [Brown, 2002], ground-penetrating radar [Amazeen, 1996; Witten, 1998; Kaskett, 1999; Sato, 2003; Sun, 2005], pulsed magnetic induction [Sower, 2001], forward looking radar (FLIR) [Zhao, 1998], nuclear quadropole resonance (NQR) [Garroway, 2001; Tan, 2002], and acoustic sensing [Xiang, 2001]. However, the ground-based systems are typically restricted in their search rate, and consequently have limited coverage area. Moreover, due to the close proximity to the ordnance, they may have a higher degree of hazard to the operator and equipment. Airborne systems, on the other hand, do not have the aforementioned limitations and offer low-risk standoff detection and a quick turnaround time. As a result of these advantages, airborne systems have gained popularity

in recent years. Some of the recent airborne minefield detection research programs are the Airborne Far IR Minefield Imaging System (AFIRMIS) [Simrad, 1998], Remote Minefield Detection System (REMIDS) [Bishop, 1998; Poulter, 2001], Cobra Reconnaissance and Analysis System (COBRA) [Witherspoon, 1995], and Lightweight Airborne Multi-spectral Minefield Detection System (LAMMD) [Haskett, 2001].

Figure 1.1 shows a typical airborne minefield detection scenario. Similar to remote sensing applications, airborne landmine detection systems typically use electro-optical sensors (usually operating in the visual and infrared wavelengths) mounted on an aerial platform. The sensors are typically arranged in a 2-D rectangular array. Each sensor in the array records the IR, MWIR or multispectral reflectance intensity integrated over a rectangular area of the ground below. The data from the sensor array is arranged on a two-dimensional regular lattice in R^2 to form two-dimensional images, with each pixel representing some finite area on the ground. The scale of the data is specified in terms of the Ground Sample Distance (GSD) of the data, which essentially conveys the resolution of the airborne imagery. The aerial platform transmits the captured data to the base station on the ground, where the data is processed with an aim to detect the presence of landmines. The idea is that the reflectance properties of the landmines in the various bands of the electromagnetic spectrum are different from that of the background like soil, vegetation and rocks. Then the problem of landmine detection essentially is to distinguish the spectral-spatial signature of the landmines from that of the background.

The data used in this dissertation is the airborne multispectral image dataset, collected as part of the airborne landmine detection program at Night Vision and Electronic Sensors Directorate (NVESD). The aim of the program is to develop and evaluate fast and accurate mine and minefield detection algorithms for the airborne spatial imagery. The data is captured in the form of a sequence of image frames from the sensor mounted on an aerial platform flying over simulated minefields, at various times of day and terrain conditions. The aerial platform is flown over the minefield area at a predefined altitude and speed, with a gimbal to collect frames of images in a specified pattern. A specified number of image frames create a segment/field of regard (FoR). The set of segments collected from one flight constitute a *run*. As depicted in Figure 1.1, the geo-locations of each frame along with other information is collected using onboard

Global Positioning System (GPS) and Inertial Measurement Unit (IMU), which constitutes the Meta data (data of data). The Meta data and any available image overlap are used to reconstruct the ground image for the FoR.

Figure 1.2 shows a typical segment of airborne multispectral data used in the dissertation. The 21 frames of the segment are co-registered to give a consolidated view of the FoR.

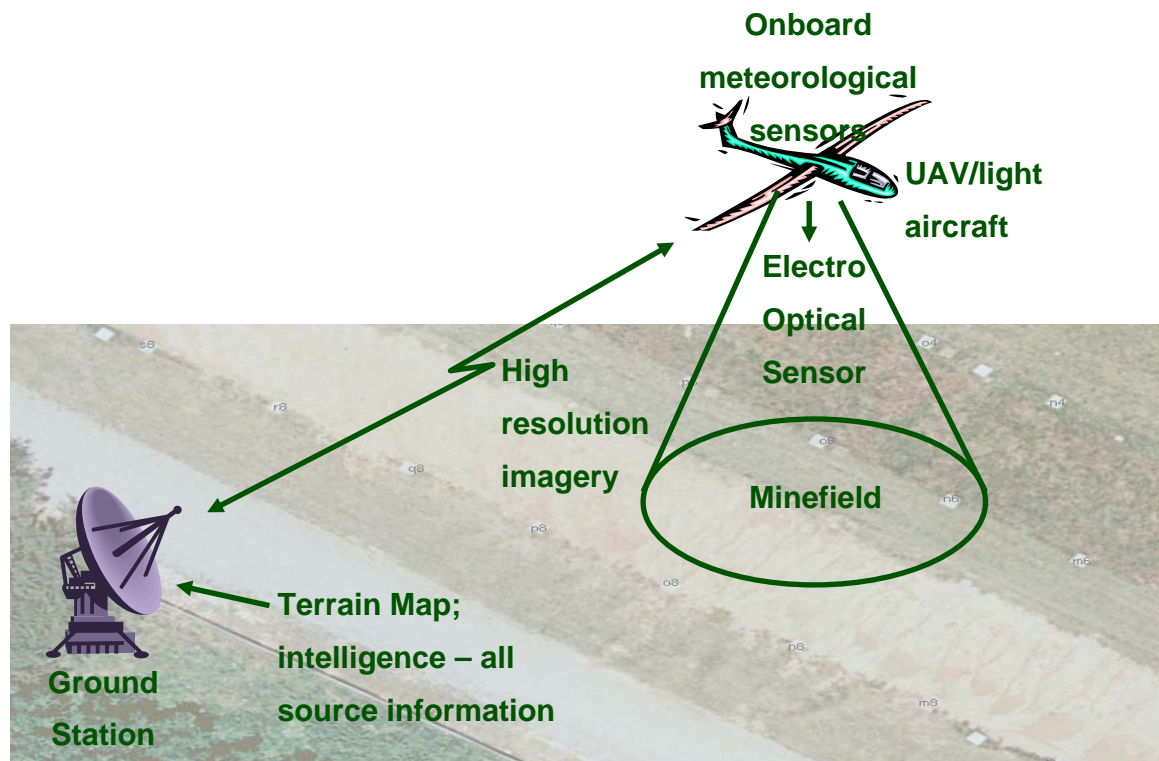


Figure 1.1. Airborne Mine and Minefield Detection Scenario

1.3. OVERVIEW OF THE DISSERTATION

In this dissertation, two different algorithms are presented for the problem of detection in spatial data. As mentioned earlier, the spatial dataset used in this work is the airborne multispectral image data collected as part of the airborne landmine detection effort. In the first part of the dissertation, a fast algorithm for kernel-based non-linear

anomaly detection in the airborne spatial data is presented. The kernel-based non-linear version of the RX anomaly detector, Kernel-RX, was introduced by Kwon et al. [Kwon, 2005a]. However, due to the high computational complexity of the kernelized version, its application is limited to problems with small sample sizes for target signature and background estimation. First, a reformulation of the Kernel-RX algorithm is presented which is termed Spatially Weighted Kernel-RX (SW-KRX). It is shown that under this reformulation, the detector statistics can be obtained as a function of the centered kernel Gram matrix calculated over the entire image. Finally, a methodology for the fast computation of the centered kernel Gram matrix is proposed in Section 4. The key idea behind the proposed methodology is to decompose the set of image pixels into clusters, and expediting the computations by approximating the effect of each cluster as a whole. The SW-KRX algorithm is implemented for a special case, and comparative results are compiled for the SW-KRX vis-à-vis the RX anomaly detector.

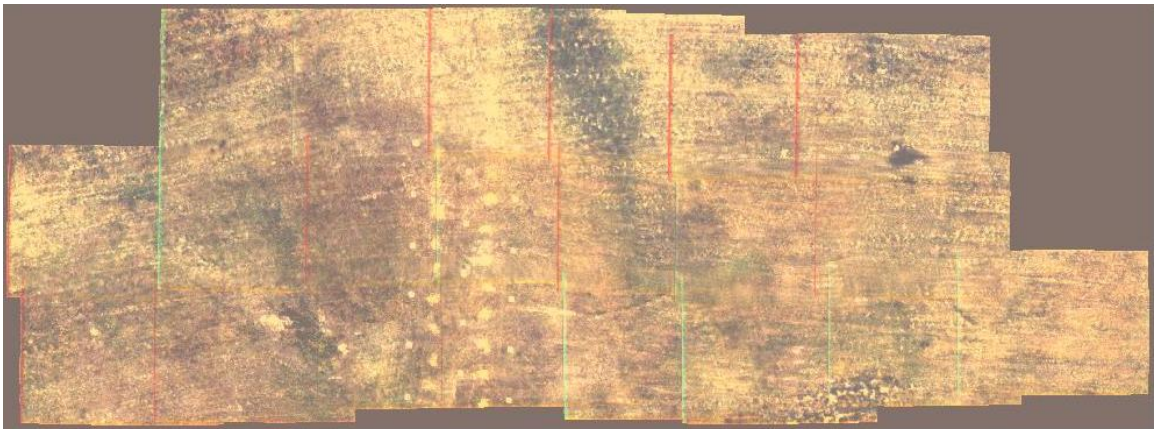


Figure 1.2. Typical Multispectral Segment

The second part of the dissertation deals with the problem of detection of buried mines in multi-spectral airborne imagery. Buried mine detection is a difficult problem because the spectral signature of the target pixels is very similar to spectral signature of the background constituents. The primary goal behind this work is to develop a methodology to extract the information in the spatial distribution of the spectral vectors

in the airborne imagery for the purposes of effective detection. This is achieved using cross-co-occurrence features from the imagery to capture the color texture information. As part of this work, a methodology for extraction, and subsequent selection of discriminatory features based on Bhattacharya coefficients and principal feature analysis of the raw features, is also presented.

It should be noted that although the two algorithms are tested on the airborne multispectral image data, the algorithms are not limited to any specific spatial dataset, and are applicable to a broad class of spatial data.

A brief overview of the organization of this dissertation is as follows. In Section 2, the problem of detection and classification in airborne landmine imagery is introduced, followed by a review of the previous work done in the area of detection in spatial data for both uniformly and non-uniformly sampled data. A detailed discussion of non-uniformly sampled data with its various advantages and disadvantages, and the methods is presented.

In Sections 3, 4 and 5, the development and details of the fast implementation of the non-linear anomaly detector, Kernel-RX, are presented. Section 3 begins by introducing the RX anomaly detector, followed by a detailed development of the Kernel-RX algorithm. The detailed development of the proposed Spatially Weighted Kernel-RX (SW-KRX) is presented, followed by the comparative results for a special case of SW-KRX on a simulated dataset.

Section 4 presents the theory behind the fast implementation of the Spatially Weighted Kernel-RX. Since the proposed methodology is based on cluster-based presentation of the data, first, a brief review of the spectral-spatial clustering methods is presented. Next, the details of the multivariate Taylor series-based methodology for fast computation of the centered kernel Gram matrix are presented. Finally, a comparative analysis of the complexity of the proposed algorithm as compared to the original Kernel-RX is presented.

Section 5 essentially compiles the various results on the proposed SW-KRX algorithm on various datasets. The results for SW-KRX are generated for a special case of the detector, details of which are explained in Section 4. The results are presented in two categories. First, the results on the detection performance of the proposed algorithm

vis-à-vis the RX anomaly detector are presented. Next, the results on the computational gains achieved by the proposed method over the original Kernel-RX are depicted.

Section 6 presents the algorithm for detection of buried mines in multi-spectral airborne imagery in detail. First, the problem of buried mine detection is introduced, followed by a brief description of the data that is used for presenting the performance results of the proposed algorithm. Then, the details of the methodology for feature extraction and discriminatory feature selection are presented. Next, detailed description of the various feature-based detectors is presented. Finally, comparative results on detection performance of the proposed algorithm vis-à-vis RX detector are presented, for various datasets.

Finally, this dissertation is concluded in Section 7 followed by a discussion on the possible directions of research in the future.

2. RELATED WORK

The problem of target detection in multispectral and IR spatial data is a long standing one. There are several sources of spatial data for which detection methods have been proposed. Some of the major sources are ground-penetrating radar [Amazeen, 1996; Witten, 1998; Kaskett, 1999; Sato, 2003], forward looking radar (FLIR) [Zhao, 1998], pulsed magnetic induction [Sower, 2001], acoustic sensing [Xiang, 2001], terahertz imaging [Hwu, 2004], and airborne IR and multispectral/hyperspectral imaging [Schweizer, 2001; Thai, 2002; Bowman, 1998] systems. A lot of research effort has been devoted to the detection problem, and there exists a huge body of literature on the subject. In Section 2.1, the detection problem specifically from the viewpoint of airborne spatial data is presented. Some of the detectors that have been proposed in the past, both for uniformly and non-uniformly sampled spatial data are also reviewed in the same section. Almost all of the detectors that have been proposed are for regular lattice spatial data (uniformly sampled images), and very little work has been done specifically on detection in non-uniformly sampled spatial data. A detailed discussion on the relative merits and demerits of non-uniformly sampled data is presented in section 2.2. Some of the basic tools and techniques that have been developed for non-uniformly sampled data processing are also reviewed.

2.1. DETECTION IN AIRBORNE SPATIAL DATA: REVIEW

In recent years, airborne spatial data has increasingly been explored for various detection applications due to its capability for low-risk standoff detection and quick turnaround time [Agarwal, 2001; Bishop, 1998; Engel, 1998; Grosch, 1995; Maksymenko, 1997; Simrad, 1998; Suzukawa, 1995; Tiwari, 2007; Witherspoon, 1995]. Accurate target detection in various forms of airborne data like hyperspectral (HSI) [Harsanyi, 1994a, 1994b; Manolakis, 2001, 2003; Kwon, 2005a], multispectral (MSI) [Clark, 2000; Tiwari, 2007] and infra-red (IR) image data [Agarwal, 2001; Bowman, 1998] is a challenging problem. The goal behind a detection problem is to distinguish whether a given spectral data vector (in case of MSI/HSI data) belongs to the target class or not. In mathematical terms, the problem of target detection is usually posed as a

statistical binary hypothesis test [VanTrees, 1968]. Although there are several methods of systematic design of detection algorithms, the likelihood ratio (LR) test is particularly popular. This is because LR test allows us to assign a specific cost to the risk associated with an incorrect decision and also produces detectors which are optimum for several performance criteria [Manolakis, 2003]. The LR test based target detectors can be classified broadly into two classes, based on the information available about the target: matched filter approaches and anomaly detection approaches.

In the matched filter approach, it is assumed that the statistics of the target and the background are known. The distributions of the spectra of the two classes are typically modeled as multivariate normal distributions. The detector statistic is then calculated as an appropriate distance of the observed vector from the means of the target and background classes. The matched filter is a constant false alarm rate (CFAR) detector, i.e., an exact threshold, which would give the desired false alarm probability, can be determined. In case the covariance matrices of the two classes are the same, the matched filter detector is just the Fisher's Linear Discriminant for the two class problem, [Fisher, 1936]. In the case when the different spectral vectors for both classes have uncorrelated components with equal variance, this detector is equivalent to the popular minimum variance distortionless-response beamformer [Haykin, 2001]. Several detection methodologies have been proposed which are based on this approach [Manolakis, 2000; Robey, 1992].

However, there are some problems in accurate target detection using the matched filter approach. One of the major problems is that of target signature variability, which arises due to variations in atmospheric conditions, sensor noise, material composition and terrain. Due to this, complete unique characterization of target signature in spatial data is practically impossible. Although it is possible to generate a library of possible target spectral signatures, the configuration space of the possible signatures is typically large and it limits the performance of the detector to the space of the signatures in the library. Also, in a typical data, there are relatively few targets as compared to the background, and consequently, sufficient information is not available to reliably estimate the statistics of the target class. This limitation brings out the major difference between detection and classification problems. Although theoretically detection and classification problems are

the same (detection is essentially a binary classification problem), in practice they pose different challenges. Typically, for classification there are enough data points to estimate the statistics of either class. But in detection problems the data points belonging to the target class are much fewer, which poses robust estimation problems. Also, in certain scenarios, the signature of the potential target is completely unknown and the system has to detect the target based on what it has “seen” previously. In such situations where there is not enough statistical information regarding the target, the anomaly detector approach is adopted.

In the anomaly detection approach, it is assumed that there is not sufficient *a priori* information regarding the target class. In such a case, the only option is to somehow measure how different (hence the word “anomaly”) a given spectral vector is from the background. The detector statistic in case of anomaly detection is just the Mahalanobis distance of the spectral vector from the mean of the background class [VanTrees, 1968]. The statistics of the background class can be reliably estimated from the data itself, in light of the fact that very few of the spectral vectors in the data actually belong to the target class. Such a LR test based anomaly detector is optimal for the Neyman-Pearson criterion. However, in almost all practical situations the conditional densities in the LR test depend on some unknown background parameters. In such a case, the maximum-likelihood estimates of the unknown parameters are used for the conditional densities. The LR test based on the estimated densities is then called a Generalized Likelihood Ratio Test (GLRT). Moreover, if the unknown parameters are estimated adaptively over the data (spatially varying), then such a detector is called an *adaptive* anomaly detector. Reed and Yu [Reed, 1990] proposed one of the first adaptive anomaly detectors based on GLRT, called the RX anomaly detector. Although, such an adaptive anomaly detector does not have the optimality property in the Neyman-Pearson sense, it has been found to be effective in practice [Yu, 1993].

Recently, a kernel-based nonlinear version of the RX anomaly detector was proposed by Kwon and Nasrabadi (2005a), and termed Kernel-RX. The essential idea in the Kernel RX detector is that the data vectors are first transformed to a feature space by non-linear feature transformation. The detector statistics are calculated in the feature space, instead of original data space. The computations can be devised in terms of the

“kernel function” associated with the feature transformation, which makes the non-linear computations tractable. Early studies comparing RX and Kernel-RX have shown that the latter can give better performance on some spatial datasets [Kwon, 2005a]. In the preliminary results shown in Section 6, it is also found that Kernel RX performs at par or better, for simulated spatial data. Kernel RX is particularly useful in cases where the background is non-homogeneous, in which case Kernel RX gives significantly better performance than the RX detector.

It is noted that in the above description of the two classes of detectors only the spectral information in the data was exploited. The detection was “non-literal” in the sense that no shape information was utilized. However, it is possible to have spatial-spectral versions of both the detectors (i.e., spatial-spectral matched filter and spatial spectral anomaly detector). It differs from the non-literal version only in that the test spectral vector is not just the current vector under consideration, but is obtained as a function of spectral vectors in the spatial region corresponding to the target shape, around the test pixel vector.

Another major problem in target detection in airborne imagery arises with the finite spatial resolution of the data. Consider the case when the target actually occupies only part of the area on the ground which is captured in a pixel. Since the rest of the area is covered by background, the pixel records a mixture of the spectral signature of the target and background. This problem is sometimes referred to as spatial mixing. As can be seen, the detectors mentioned above are not designed for the case of sub-pixel targets and spatially mixed signatures. In fact the aforementioned detectors can be classified as full-pixel target detectors, and detectors designed specifically for the spatial mixing problem, as sub-pixel target detectors [Manolakis, 2001; Keshava, 2003].

In case of spatial mixing, it is reasonable to assume that the spectral vector is a linear combination of the target and background spectra. Therefore, the basic difference in this case is that the variability of the target model is represented using a subspace model, i.e. the target vector $\mathbf{r} = \mathbf{S}\mathbf{a}$, where columns of \mathbf{S} are the “endmembers” or representative signatures and the vector \mathbf{a} provides their relative proportions. The variability of the background is modeled mathematically using either a statistical model or a structural model. Use of different models for the background leads to different

detection techniques. For instance, in the case of the statistical models, the background is modeled as multivariate normal distribution and this led to different algorithms like adaptive coherence/cosine estimator (ACE) [Kraut, 1999, 2001] and adaptive matched filter (AMF) [Chen, 1991; Robey 1992]. For the case of the subspace model for background, techniques like Orthogonal Subspace Projection (OSP) [Harsanyi, 1994b] and matched subspace detector [Scharf, 1994] have been proposed. A detailed comparison on the detection performance of the subspace background models can be found in [Manolakis, 2001]. Recently, kernel-based non-linear versions of these various sub-pixel target detectors have been proposed, which the authors argue improve the detection performance on hyper-spectral imagery (HSI) datasets [Kwon, 2005a, 2005b, 2006a, 2006b].

Few techniques exist that address the problem of detection in irregularly sampled spatial data. One of the important techniques has been proposed by Carlotto [2005]. They proposed a new unified approach for anomaly and change detection in hyperspectral imagery. Their technique, called the Cluster Based Anomaly Detection (CBAD) and Cluster Based Change Detection (CBCD), as a first step involves clustering of the spectral vectors in the image into disjoint clusters. Statistics on the distribution of the clusters are retained, which are then used for detecting anomalies and changes. This work is unique in the sense that it obtains a non-uniformly sampled representation from uniformly sampled data, so as to improve the robustness and performance of the algorithm. Other work on similar lines i.e., cluster-based detection, has also shown improvements in the computational efficiency of the algorithms using the cluster-based representation [Stein, 2002; Carlotto, 2002].

2.2. NON-UNIFORMLY SAMPLED SPATIAL DATA: COMPARISON AND REVIEW

Although, several algorithms have been proposed for regular grid data, very few have addressed the detection problem in non-uniformly sampled spatial data. There are several advantages of working with non-uniformly sampled data like prevention of oversampling, compression and anti-aliasing. Also, in certain scenarios, the data naturally occurs in non-uniformly sampled form. It is desirable that the detection and other algorithms can be applied directly on the non-uniformly sampled data, instead of

converting it into uniformly sampled data by interpolation of some kind, as is usually done in such situations. In this section, a detailed discussion on non-uniform sampling and a review of techniques and methods for working with non-uniformly sampled data is presented.

Uniformly sampled data has some definite advantages over non-uniformly sampled data. That is the reason why significant research and development effort has been devoted to the field of uniformly sampled data analysis and processing. However, with the challenges brought on with the advent of the age of data, there has been a resurgence of interest in the development of analysis and processing tools and techniques for the non-uniformly sampled data. For instance, there are situations where non-uniformly sampled data is all that is available. Conventional methods go around this problem by first converting the non-uniformly sampled data into uniformly sampled data, by some form of interpolation. However, research in the area of non-uniform sampling over the past few decades has suggested that non-uniformly sampled representation of data has certain advantages, and it is not always desirable to work with uniform samples. In this section, some advantages and disadvantages of working with non-uniformly sampled spatial data are presented. This is followed by a brief review of some of the tools and techniques that have been proposed for non-uniformly sampled data.

Shannon with his presentation of the sampling theorem in 1948 [Shannon, 1948] gave a tremendous boost to the area of digital signals. Although he did not claim this theory to be his own, he is largely credited with formalizing the sampling theorem and introducing it to the fields of communications and signal processing. With the sampling theorem, Shannon laid down the foundation for a coherent and sound mathematical theory for uniformly sampled signals¹. With the solid mathematical foundations coming from the sampling theorem and the linear systems theory, the use of uniformly sampled signals became tremendously popular. With the well-defined notion of frequency and a general mechanism for converting analog signals into a string of numbers, the world of science and engineering was quick to adopt uniformly sampled signals as the standard for

¹ In this section, the terms regular grid signals and uniformly sampled signals are used interchangeably. Typically, the former is used more in context of two or higher dimension signals, and the latter, more for one dimensional temporal signals, as seen in communications and signal processing.

representing real world signals. Although the sampling theorem has undoubtedly revolutionized the field of modern engineering and signal processing, there were practical problems with the basic premise of the theorem: the uniform sampling of signals. The one major problem with uniform sampling is the requirement that the signals should be bandlimited (to decide the sampling rate). Real-world signals are rarely bandlimited, and this leads to the problem of aliasing, since the frequency spectrum of the uniformly sampled signals repeats itself outside the interval $[-f_s/2, f_s/2]$, where f_s is the sampling frequency. Also, the ideal low-pass filters which are needed as anti-aliasing (or also during signal reconstruction) do not exist in practice, so it becomes difficult to remove aliasing in the reconstructed signal entirely.

Another related problem with the uniform sampling is that of oversampling. In most of the real-world signals, the high frequencies are irregularly distributed i.e., they are sporadic and occur in spurts. This phenomenon can be observed in several real-world signals of natural scene images, speech and music. In the slowly varying part of such signals (low-frequencies), there is a strong correlation amongst neighboring samples. Uniform sampling at a rate based on the highest frequency in the signal will typically result in oversampling. Although, oversampling does not produce distortion, it can lead to significant waste of computational and storage resources.

Besides aforementioned problems, there are scenarios in which it is not even possible to obtain uniformly sampled signals or data. For instance in very high rate signal processing, special instrumentation (analog-to-digital converter) is needed to ensure that the sampling is accurate, and even then there are formidable hardware issues. Also, in certain applications it is fairly frequent that one or more of the samples are “missing” from the recorded data [Goodsill, 1993], due to problems like data corruption or transmission loss. In certain fields like geology and geophysics, owing to the limited accessibility of the terrain and other reasons, it is not possible to take samples of geophysical data (electrical conductivity, soil composition, magnetic potential etc.) at regularly spaced locations. Thus, the data samples are inherently irregularly spaced and typically, highly clustered. Besides these, there are several other fields and applications where use of non-uniformly sampled data is widespread, like astronomy, computer tomography and seismology.

The primary challenge in working with non-uniformly sampled data comes from the lack of unified and coherent mathematical theory and tools for the same. Although, the field has seen strong resurgence since the late eighties [Strohmer, 1995, 2006; Gröchenig, 1992; Yellott, 1982; Tarczynski, 2004; Bilinskis, 1992], it is still in the nascent stages and there are several open fundamental questions [Summers, 1988]. For instance, the notion of frequency for non-uniformly sampled data is hard to define. Also, standard signal processing tools like the fast Fourier transform (FFT), fast convolution and filtering are no longer directly applicable.

However, there are certain significant fundamental advantages in working with non-uniformly sampled data. It is possible to adopt intelligent sampling schemes and avoid oversampling. As a result, there is a broader scope for designing computationally efficient algorithms for non-uniformly sampled data. For instance, Greitans [1997] showed that non-uniform sampling allows the processing of signals employing fewer signal samples as compared to what is required by the Nyquist sampling criterion.

Another major advantage of working with non-uniformly samples signals is the possibility to eliminate the aliasing effects, i.e., it is possible to exceed the Nyquist limit. In other words, in the case of regular sampling, by definition the highest possible frequency that can be distinguished unambiguously is $f_s/2$. If there exist frequencies higher than this limit they get folded back into the interval $[-f_s/2, f_s/2]$. With irregular sampling this restriction is removed and spectral information from outside this band can also be analyzed [Yellott, 1982; Tarczynski, 2004; Bilinskis, 1992]. Several methods for spectral analysis for non-uniformly sampled signals have been proposed [Tarczynski, 2002, 2004; Mednieks, 1999]. In fact, owing to this property of non-uniformly sampled signals, a whole new area of research called Digital Alias-free Signal Processing (DASP) has emerged, which already is seeing some interesting applications in instrumentation [Artyukh, 1997] and Digital Radio [Wojtiuk, 2000]. The interested reader is referred to some early fundamental papers in the DASP area [Shapiro, 1960; Masry, 1978], and some review papers on the various techniques proposed in the field [Bilinskis, 1992; Wojtiuk, 2000; Martin, 1998]. While on the topic of spectral analysis, it is important to note that methods exist for the Fourier transform of non-uniformly sampled signals (Non-uniform Discrete Fourier Transform, NDFT) [Bagchi, 1999], and its fast calculation

[Dutt, 1995; Potts, 2001]. However, it should be noted that although the forward Fourier transform for non-uniformly sampled data exists, it is not always invertible. Also, filtering techniques have also been proposed for non-uniformly sampled data, like the fast multipole method based filtering proposed by Gumerov and Duraiswami [2004].

Although several methods have been proposed for spectral analysis of non-uniformly sampled signals, the problem of signal reconstruction is still very much open. Most of the signal reconstruction methods for non-uniformly sampled signals that have been proposed are for bandlimited signals [Feichtinger, 1992a, 1992b; Gröchenig, 1992; SaySong, 1995; Seip, 1987; Strohmer 1995]. A few techniques have also been proposed for fast reconstruction [Feichtinger, 1995; Strohmer 2006]. Although a few spline based reconstruction techniques have been proposed which do not have a strict requirement of “bandlimitedness”, they too are not readily applicable to arbitrary signals due to the regularization condition they enforce [Vazquez, 2005; Arigovindan, 2005].

Another advantage that comes with non-uniform sampling is that of compression. As mentioned earlier, relating the sampling rate with the signal rate can reduce storage and processing requirements [Irvine, 2002; Arigovindan, 2005]. Recently, Arigovindan et al. [2005] proposed a variational method using multi-resolution splines, for reconstructing an image from few randomly selected pixels. Their work clearly demonstrates that most of the information in a uniformly sampled representation of an image is largely redundant, and it can almost completely be captured in fewer irregularly spaced samples.

Non-uniform sampling can also be seen in natural systems like the human retina. In the human retina the photoreceptors are distributed in a non-uniform fashion, which research has shown, allows the human visual resolution limit to beat the Nyquist rate, and discern fine patterns with spatial frequencies as high as 1.5 times higher than the nominal Nyquist rate [Williams, 1987]. Vision scientists have repeatedly conjectured that besides getting an alias-free representation [Yellott, 1982], compression is another major goal of the photoreceptor distribution and signal coding in the retina [Olhausen, 1997].

3. SPATIALLY WEIGHTED KERNEL-RX (SW-KRX)

Anomaly detectors have been used extensively in the past for various applications like image analysis and target detection. Recently with the rapid advances in infrared (IR) sensor technology, there has been an explosion in the availability of sensors in the form of visible band, multispectral (MSI) and hyperspectral imagery (HSI) for analysis. Accurate detection and classification of targets of interest in this realm of sensor image data has been a long standing problem and extensive research has been done. Since the target signatures are spectrally and spatially different from the local background, most well-known frameworks developed for target detection in spectral imagery data employ anomaly detectors in the initial stages of processing [Agarwal, 2001; Chang, 2002]. One of the most widely used anomaly detectors is the RX (Reed-Xiao) algorithm, which was first proposed by Reed and Yu [Reed, 1990]. Recently a kernel-based nonlinear version of the RX anomaly detector was proposed by Kwon and Nasrabadi [Kwon, 2005a], and termed Kernel-RX. Owing to its non-linear nature, the proposed Kernel-RX algorithm takes into account spatial correlation and higher-order interactions amongst various spectral bands of the data in the original image space. This is unlike the original RX, which had a set of restrictive assumptions like zero mean, uncorrelated and Gaussian distributed data. With the non-linear model, improved target detection performance of Kernel-RX vis-à-vis linear RX was reported by Kwon et al. (Kwon, 2005a) for HSI data.

However, like other kernel-based methods, Kernel-RX has a major drawback of high computational complexity. There exist fast implementations of the conventional RX which exploit its linear nature to do the computations in the Fourier domain, greatly reducing the computation time [Holmes, 1995]. However, no technique has been proposed for fast computation of the Kernel-RX algorithm. Thus, in spite of improved detection performance, the Kernel-RX has seen limited applications, especially for real-time detection scenarios like detection of military targets in an incoming HSI data stream. In this section, the problem of the high computational load of Kernel-RX is addressed. A reformulated version of the Kernel-RX algorithm is proposed, termed the Spatially Weighted Kernel RX (SW-KRX). As will be demonstrated in the next section, the proposed version is more malleable for faster implementation.

3.1. RX ANOMALY DETECTOR

Since the target signatures are typically statistically different from the local background, a statistical likelihood ratio test can be used to identify such “anomalies” in the local background. A detector based on such a test is called an *anomaly detector*. The RX anomaly detector described here was first proposed by Reed and Yu [Reed, 1990]. The RX detector assumes the input image to be uncorrelated and Gaussian distributed, with a zero mean. The assumption of Gaussian distribution of gray values is true of most of the images captured using electro-optical sensors, but most images are not zero mean, although they can be assumed to be slow-varying in most cases. Therefore, as a pre-processing step to RX anomaly detection, a locally zero mean image is obtained from the raw input image. This is done by subtracting a non-stationary local mean from the original image. For the work in this dissertation, a specific implementation of the RX anomaly detector, proposed by Holmes et al. [1995], is used. This particular implementation uses three masks namely, target mask, blanking mask and demeaning mask. The target mask is taken to be circular, with a specific radius called the target radius, and it defines the shape of the targets expected in the image. The demeaning mask is also taken to be circular with a radius called the demeaning radius. The demeaning circle is used to estimate local mean of the background and is also used for the demeaning the original image. Finally, the annular region (since the masks are circular) between the blanking and the demeaning masks is used for estimation of the background statistics. Blanking radius, which is always less than the demeaning radius, is used to prevent any pixels of the target from being used in background estimation. The relative mask shape and sizes are shown in Figure 3.1.

As the first step, the original image is demeaned by subtracting the estimated local mean from each pixel. The RX anomaly detector takes the zero-mean image as input and gives an “RX image” as output, which is of the same size as the input image. The input image is convolved with the set of masks, and this computation is done in the frequency domain (multiplication of FFTs) for efficiency.

The RX anomaly detector can be used for multi-band images in general. Consider a J band demeaned image I . Let the p^{th} pixel in the image be at location (i,j) .

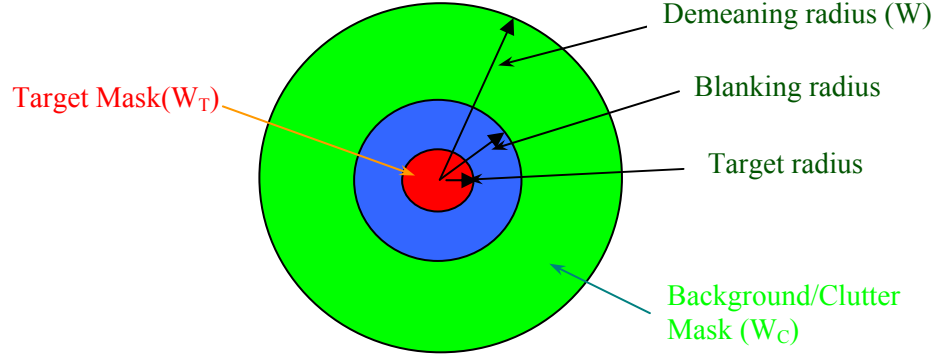


Figure 3.1. Relative Sizes and Shapes of Various Masks Used in RX Detector.

Then, for the p^{th} pixel, $I(i, j) = \mathbf{y}_p$, the RX statistic $RX(i, j)$ is given by:

$$RX(i, j) = \left(\boldsymbol{\mu}_S^T \mathbf{C}^{-1} \boldsymbol{\mu}_S \right) \quad (3.1)$$

where \mathbf{C} is the locally estimated $J \times J$ covariance matrix given by:

$$\mathbf{C} = \frac{1}{N_C} \sum_{\mathbf{y}_l \in W_C} \mathbf{y}_l^T \mathbf{y}_l \quad (3.2)$$

Also, $\boldsymbol{\mu}_S$ is the mean target signature given by:

$$\boldsymbol{\mu}_S = \frac{1}{N_T} \sum_{\mathbf{y}_l \in W_T} \mathbf{y}_l \quad (3.3)$$

where W_C and W_T are the set of clutter pixels and target pixels around location (i, j) used in estimating the background and target statistics, respectively, and N_C and N_T are the number of clutter and target pixels used in estimating \mathbf{C} and $\boldsymbol{\mu}_S$, i.e.,

$$N_C = |W_C|, N_T = |W_T| \quad (3.4)$$

where $|\cdot|$ denotes cardinality of a set. The RX detector statistics is the Mahalanobis distance of the target spectral vector from the sample distribution of the background as estimated from the spectral vectors in the clutter mask.

Once the RX output is obtained from the raw image data it is usually followed by some post-processing steps like thresholding at a pre-specified value called the RX

threshold, and a non-max suppression (on the thresholded RX image) to suppress multiple points for the same target from showing up.

3.2. KERNEL-RX ANOMALY DETECTOR

Kernel-RX is fundamentally very similar to the conventional RX algorithm in that both calculate a test statistic for a generalized likelihood ratio test (GLRT) which tests the presence of a target at a pixel. The same concept of the estimation of the background and target over the background and clutter masks is used for the kernel detector. However, unlike the conventional RX algorithm which is based on the statistics of the input image data, Kernel-RX's GLRT is based on the statistics on the non-linear transformation of the image pixels. Let us denote the data space of the original image by Γ i.e. $\Gamma \subseteq \mathfrak{R}^J$, where J is the dimensionality of the MSI/HSI data. This data is mapped into a feature space F through a nonlinear mapping denoted by $\Phi: \Gamma \rightarrow F$. The individual pixel vectors are then mapped into potentially higher (possibly infinite) dimensional feature space: $\mathbf{y} \rightarrow \Phi(\mathbf{y})$. The key idea behind kernel-based algorithms is that reformulating any linear algorithm in the feature space gives a non-linear algorithm on the original data space. Kernel-RX is also a direct re-formulation of the conventional RX in the feature space. The feature set mapped from the set of clutter pixels \mathbf{Y} is denoted as:

$$\mathbf{Y}^\Phi \equiv \{\Phi(\mathbf{y}(1)), \Phi(\mathbf{y}(2)), \dots, \Phi(\mathbf{y}(N_C))\} \quad (3.5)$$

The mapped target pixel \mathbf{y}_r is denoted as $\Phi(\mathbf{y}_r)$. For the Kernel RX algorithm, the RX test statistic given in Eqn. 1 in the feature space can be re-written as follows:

$$KRX(\mathbf{y}_r) = (\Phi(\mathbf{y}_r) - \boldsymbol{\mu}_C^\Phi)^T \hat{\mathbf{C}}_\Phi^{-1} (\Phi(\mathbf{y}_r) - \boldsymbol{\mu}_C^\Phi) \quad (3.6)$$

where,

$$\boldsymbol{\mu}_C^\Phi = \frac{1}{N_C} \sum_{\mathbf{y}_k \in Y_C} \Phi(\mathbf{y}_k) \quad (3.7)$$

Let us denote the *centered* feature vectors as: $\Phi_c(\mathbf{y}_i) = \Phi(\mathbf{y}_i) - \boldsymbol{\mu}_C^\Phi$, and the set of *centered* feature vectors as: $\mathbf{Y}_C^\Phi \equiv \{\Phi(\mathbf{y}(1) - \boldsymbol{\mu}_C^\Phi), \Phi(\mathbf{y}(2) - \boldsymbol{\mu}_C^\Phi), \dots, \Phi(\mathbf{y}(N_C) - \boldsymbol{\mu}_C^\Phi)\}$.

Then, the estimate of background sample covariance matrix in the feature space can be given as:

$$\hat{\mathbf{C}}_{\Phi} = \frac{1}{N_C} \sum_{\mathbf{y}_k \in \mathcal{Y}_C} (\Phi(\mathbf{y}_k) - \boldsymbol{\mu}_C^{\Phi})(\Phi(\mathbf{y}_k) - \boldsymbol{\mu}_C^{\Phi})^T = \frac{1}{N_C} \mathbf{Y}_C^{\Phi} \mathbf{Y}_C^{\Phi T} \quad (3.8)$$

Direct computations in the non-linear feature space are often not feasible due to the higher dimensionality of the data. For instance, the matrix $\hat{\mathbf{C}}_{\Phi}$ can be too large (potentially infinite) to compute. However, with the use of the “kernel trick”, the dot products of vectors in the feature space can be computed using a *kernel function* ‘ k ’ as:

$$k(\mathbf{a}, \mathbf{b}) = \Phi(\mathbf{a}) \cdot \Phi(\mathbf{b}) \quad (3.9)$$

Note that the kernel function k inherently determines the non-linear mapping Φ . Commonly used examples of kernel functions are the polynomial kernel and Gaussian kernel. In this work, the popular Gaussian kernel function is adopted. The Gaussian kernel of width $\sigma > 0$ is given by:

$$k(\mathbf{a}, \mathbf{b}) = \exp\left(-\|\mathbf{a} - \mathbf{b}\|^2 / \sigma^2\right) \quad (3.10)$$

where $\|\cdot\|$ denotes the L_2 Euclidean norm. The $N_C \times N_C$ kernel Gram matrix is defined for the feature vector set \mathbf{Y}^{Φ} as:

$$[\mathbf{K}]_{ij} = \Phi(\mathbf{y}_i) \cdot \Phi(\mathbf{y}_j) = k(\mathbf{y}_i, \mathbf{y}_j), \quad i, j = 1, 2, \dots, N_C \quad (3.11)$$

Then, the Kernel RX statistic is given as (detailed derivation in Appendix A):

$$KRX(\mathbf{y}_r) = \hat{\mathbf{k}}_r^T (\boldsymbol{\alpha} \boldsymbol{\Lambda}^{-2} \boldsymbol{\alpha}^T) \hat{\mathbf{k}}_r \quad (3.12)$$

where,

$$\hat{\mathbf{k}}_r = \mathbf{Y}_C^{\Phi T} \Phi_C(\mathbf{y}_r), \quad (3.13)$$

$\boldsymbol{\alpha}$ is the matrix containing the eigenvectors of matrix *centered* kernel Gram matrix $\hat{\mathbf{K}}$ along its columns, and $\boldsymbol{\Lambda}$ is the diagonal matrix containing the eigenvalues of matrix $\hat{\mathbf{K}}$. The *centered* kernel Gram matrix $\hat{\mathbf{K}}$ can be obtained from the original Gram matrix \mathbf{K} , defined in Eqn. (3.11) as (see Appendix C for the detailed derivation):

$$\hat{\mathbf{K}} = \mathbf{K} - \mathbf{1}_{N_C} \mathbf{K} - \mathbf{K} \mathbf{1}_{N_C} + \mathbf{1}_{N_C} \mathbf{K} \mathbf{1}_{N_C}$$

where, $[\mathbf{1}_{N_C}]_{ij} = 1/N_C$ is a $N_C \times N_C$ matrix. Note that the second and the third terms are the column and row means of the original Gram matrix. The fourth term is the mean of the entire matrix.

3.3. SPATIALLY WEIGHTED KERNEL-RX

As mentioned earlier, the Kernel RX algorithm has high computational complexity. The algorithm has primarily two computational bottlenecks. The first bottleneck is the expensive computation of the Gram matrix for each pixel of the image, which takes $O(N_C^2 J)$ computations, where N_C is the number of background samples and J is the dimensionality of the data. Secondly, the computation of the Kernel-RX value involves the inversion of the Gram matrix, which is a $O(N_C^3)$ process, followed by a matrix vector multiplication. The number of background samples N_C is governed by the image resolution and the target size. For large value mask sizes and high resolution, the computational complexity of the Kernel-RX algorithm becomes prohibitively large. In this section, the details of a modified version of the Kernel RX algorithm, called Spatially Weighted Kernel RX (SW-KRX), are presented. The proposed version is more amenable to faster implementation, the details of which are presented in the next section.

3.3.1. Modified Detector Statistics. Consider again, a J band image with dimensions $M \times M$, and total $N = M^2$ pixels. In the original Kernel RX algorithm, the background set \mathbf{Y} is defined by the clutter mask (the annular region between blanking and demeaning radii). In the proposed SW-KRX method, the concept of masks is eliminated, and the neighborhood of a given pixel is defined as the entire image domain, i.e., the set of background pixels of a given target pixel is the set of all the pixels in the image, i.e., $\mathbf{Y} = \{\mathbf{y}_i \mid i = 1, 2, \dots, N\}$. The mapped feature vector set is denoted as \mathbf{Y}^Φ in Eqn. (3.5) defined for all the N pixels in the image, and not just the N_c pixels falling within the clutter mask area.

The adaptivity of the anomaly detector is effectuated by the use of weights. The exact selection of the weights is described in detail in Section 3.3.2. Let the target pixel under consideration be denoted by \mathbf{y}_r . A weight w_i is associated with each feature vector

$\Phi(\mathbf{y}_i)$ of the image. These weights are with respect to the current target pixel under consideration. A weighted version of the Kernel RX is now developed using these weights. A similar reformulation for the RX anomaly detector was proposed by [Ren, 2005]. It should be noted that the set \mathbf{Y}^Φ comprises all the pixel feature vectors in the image including the target pixel under consideration, assuming that the detector statistic is being calculated for all the pixels in the image. This is unlike other anomaly detectors including the original version of the Kernel RX, where the set of target feature vectors and the background feature vectors are disjoint. Although the goal of the anomaly detector is to “distinguish” the target pixel from the background data, it can be safely assumed that one data sample in the background set does not significantly alter the detector statistic.

Given the target pixel under consideration, the weighted background mean and the weighted background covariance matrix can be written as:

$$\boldsymbol{\mu}_{CW}^\Phi = \sum_{i=1}^N w_i \Phi(\mathbf{y}_i) \quad (3.14)$$

$$\hat{\mathbf{C}}_{\Phi W} = \sum_{i=1}^N (\Phi(\mathbf{y}_i) - \boldsymbol{\mu}_{CW}^\Phi) w_i (\Phi(\mathbf{y}_i) - \boldsymbol{\mu}_{CW}^\Phi)^T \quad (3.15)$$

If the *centered* feature vectors is denoted as: $\Phi_c(\mathbf{y}_i) = \Phi(\mathbf{y}_i) - \boldsymbol{\mu}_{CW}^\Phi$, and the set of *centered* feature vectors as: $\mathbf{Y}_{CW}^\Phi \equiv \{\Phi(\mathbf{y}(1) - \boldsymbol{\mu}_{CW}^\Phi, \Phi(\mathbf{y}(2) - \boldsymbol{\mu}_{CW}^\Phi, \dots, \Phi(\mathbf{y}(N_c) - \boldsymbol{\mu}_{CW}^\Phi)\}$, then the weighted covariance matrix is given by:

$$\hat{\mathbf{C}}_{\Phi W} = \mathbf{Y}_{CW}^\Phi \mathbf{W}_0 \mathbf{Y}_{CW}^{\Phi T} \quad (3.16)$$

where, \mathbf{W}_0 is a diagonal matrix with the various weights on its diagonal, i.e. $[\mathbf{W}_0]_{ii} = w_i$ and $[\mathbf{W}_0]_{ij} = 0 \forall i \neq j$. Let us assume that the covariance matrix $\hat{\mathbf{C}}_\Phi$ has an eigen-decomposition:

$$\hat{\mathbf{C}}_{\Phi W} = \mathbf{U}_{\Phi W} \boldsymbol{\Lambda}_{\Phi W} \mathbf{U}_{\Phi W}^T \quad (3.17)$$

where $\mathbf{U}_{\Phi W} = [\mathbf{u}_{\Phi W}^1, \mathbf{u}_{\Phi W}^2, \dots, \mathbf{u}_{\Phi W}^N]$ is the matrix containing the eigenvectors along its columns, $\boldsymbol{\Lambda}_{\Phi W}$ is a diagonal matrix and $\mathbf{U}_{\Phi W}^T \mathbf{U}_{\Phi W} = \mathbf{U}_{\Phi W} \mathbf{U}_{\Phi W}^T = \mathbf{I}$.

The pseudo inverse of the covariance matrix is given as:

$$\hat{\mathbf{C}}_{\Phi}^{\#} = \mathbf{U}_{\Phi W} \mathbf{\Lambda}_{\Phi W}^{-1} \mathbf{U}_{\Phi W}^T \quad (3.18)$$

Then, multiplying Eqn. (3.17) by $\mathbf{U}_{\Phi W}$, and substituting Eqn. (3.16):

$$\mathbf{U}_{\Phi W} \mathbf{\Lambda}_{\Phi W} = \hat{\mathbf{C}}_{\Phi W} \mathbf{U}_{\Phi W} = \mathbf{Y}_{CW}^{\Phi} \mathbf{W}_0 \mathbf{Y}_{CW}^{\Phi T} \mathbf{U}_{\Phi W} \quad (3.19)$$

Multiplying both sides by $\mathbf{Y}_{CW}^{\Phi T}$:

$$\mathbf{Y}_{CW}^{\Phi T} \mathbf{U}_{\Phi W} \mathbf{\Lambda}_{\Phi W} = (\mathbf{Y}_{CW}^{\Phi T} \mathbf{Y}_{CW}^{\Phi}) \mathbf{W}_0 \mathbf{Y}_{CW}^{\Phi T} \mathbf{U}_{\Phi W}$$

Again pre-multiplying both sides by $\mathbf{\Lambda}_{\Phi W}^{-1/2}$ on the right and $\mathbf{W}_0^{1/2}$ on the left, and rearranging the terms:

$$\mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \mathbf{U}_{\Phi W} \mathbf{\Lambda}_{\Phi W}^{-1/2} \mathbf{\Lambda}_{\Phi W} = \mathbf{W}_0^{1/2} (\mathbf{Y}_{CW}^{\Phi T} \mathbf{Y}_{CW}^{\Phi}) \mathbf{W}_0^{1/2} \mathbf{W}_0^{1/2} (\mathbf{Y}_{CW}^{\Phi T} \mathbf{U}_{\Phi W} \mathbf{\Lambda}_{\Phi W}^{-1/2}) \quad (3.20)$$

Let

$$\mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \mathbf{U}_{\Phi W} \mathbf{\Lambda}_{\Phi W}^{-1/2} = \mathbf{a}_W \quad (3.21)$$

Note that columns of \mathbf{a}_W are a set of orthonormal vectors, i.e., $\mathbf{a}_W^T \mathbf{a}_W = \mathbf{a}_W \mathbf{a}_W^T = \mathbf{I}$.

Now, Eqn. (3.20) can be written as:

$$\mathbf{a}_W \mathbf{\Lambda}_{\Phi W} = (\mathbf{W}_0^{1/2} \mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^{\Phi} \mathbf{W}_0^{1/2}) \mathbf{a}_W \quad (3.22)$$

Note that the term $\mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \mathbf{Y}_{CW}^{\Phi} \mathbf{W}_0^{1/2}$ is the centered kernel Gram matrix with the weight terms included. Denoting the weighted centered kernel Gram matrix as:

$$\hat{\mathbf{K}}_W = \mathbf{W}_0^{1/2} \mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^{\Phi} \mathbf{W}_0^{1/2} \quad (3.23)$$

The $(i,j)^{\text{th}}$ element of this matrix will be:

$$[\hat{\mathbf{K}}_W]_{ij} = \sqrt{w_i} \sqrt{w_j} \Phi_C(\mathbf{y}_i)^T \Phi_C(\mathbf{y}_j) = \sqrt{w_i} \sqrt{w_j} k(\mathbf{y}_i, \mathbf{y}_j) \quad (3.24)$$

Then Eqn. (3.22) can be written in terms of the weighted centered kernel Gram matrix as:

$$\mathbf{a}_W \mathbf{\Lambda}_{\Phi W} = \hat{\mathbf{K}}_W \mathbf{a}_W \quad (3.25)$$

It can be seen from Eqn. (3.25) that the columns of the matrix \mathbf{a}_W are the eigenvectors of the weighted-centered kernel Gram matrix. Let us denote the j^{th} eigenvector of the

weighted kernel Gram matrix $\hat{\mathbf{K}}_W$ as \mathbf{a}_W^j , i.e., $\mathbf{a}_W = [\mathbf{a}_W^1, \mathbf{a}_W^2, \dots, \mathbf{a}_W^N]$. From Eqn. (3.19) and (3.21), the eigenvectors of the covariance matrix can be expressed in terms of the eigenvectors of the weighted centered kernel Gram matrix as:

$$\mathbf{U}_{\Phi W} = \mathbf{Y}_{CW}^{\Phi} \mathbf{W}_0^{1/2} \mathbf{a}_W \mathbf{\Lambda}_{\Phi W}^{-1/2} \quad (3.25)$$

Then from Eqn. (3.18) and (3.25), the pseudo inverse of the covariance matrix $\hat{\mathbf{C}}_{\Phi}$ can be written as:

$$\hat{\mathbf{C}}_{\Phi W}^{\#} = \mathbf{U}_{\Phi W} \mathbf{\Lambda}_{\Phi W}^{-1} \mathbf{U}_{\Phi W}^T = \mathbf{Y}_{CW}^{\Phi} \mathbf{W}_0^{1/2} \mathbf{a}_W \mathbf{\Lambda}_{\Phi W}^{-2} \mathbf{a}_W^T \mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \quad (3.26)$$

Writing Eqn. (3.6) in terms of the pseudo inverse of the covariance matrix from Eqn. (3.26):

$$SW \text{ KRX}(\mathbf{y}_r) = \Phi_C(\mathbf{y}_r)^T \mathbf{Y}_{CW}^{\Phi} \mathbf{W}_0^{1/2} \mathbf{a}_W \mathbf{\Lambda}_{\Phi W}^{-2} \mathbf{a}_W^T \mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \Phi_C(\mathbf{y}_r) \quad (3.27)$$

where, $\Phi_C(\mathbf{y}_r)$ is the *centered* target vector. Bringing in the weight associated with the target pixel w_r , Eqn. (3.27) can be written as:

$$SW \text{ KRX}(\mathbf{y}_r) = w_r^{-1/2} w_r^{1/2} \Phi_C(\mathbf{y}_r)^T \mathbf{Y}_{CW}^{\Phi} \mathbf{W}_0^{1/2} \mathbf{a}_W \mathbf{\Lambda}_{\Phi W}^{-2} \mathbf{a}_W^T \mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \Phi_C(\mathbf{y}_r) w_r^{1/2} w_r^{-1/2} \quad (3.28)$$

Thus,

$$SW \text{ KRX}(\mathbf{y}_r) = w_r^{-1/2} [\mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \Phi_C(\mathbf{y}_r) w_r^{1/2}]^T [\mathbf{a}_W \mathbf{\Lambda}_{\Phi W}^{-2} \mathbf{a}_W^T] [\mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \Phi_C(\mathbf{y}_r) w_r^{1/2}] w_r^{-1/2} \quad (3.29)$$

Defining the column vector:

$$\hat{\mathbf{k}}_{rW} = \mathbf{W}_0^{1/2} \mathbf{Y}_{CW}^{\Phi T} \Phi_C(\mathbf{y}_r) w_r^{1/2} \quad (3.30)$$

Note that this Eqn. (3.30) is the weighted dot product of the target vector with the entire centered feature vector set, so that Eqn. (3.29) can be written as:

$$SW \text{ KRX}(\mathbf{y}_r) w_r = \hat{\mathbf{k}}_{rW}^T (\mathbf{a}_W \mathbf{\Lambda}_{\Phi W}^{-2} \mathbf{a}_W^T) \hat{\mathbf{k}}_{rW} \quad (3.31)$$

Eqn. (3.31) gives the detector statistics for one target pixel \mathbf{y}_r . However, the modified expression for the detector statistics given in Eqn. (3.31) helps in circumventing a major computational hurdle. It should be noted that the weighted centered Gram matrix $\hat{\mathbf{K}}_W$ defined in Eqn. (3.23) remains the same for all the target pixels in the image,

irrespective of the position of the target pixel under consideration. This is unlike the original Kernel RX algorithm, in which the background pixel set \mathbf{Y}^Φ changes from pixel to pixel, due to which the Gram matrix has to be re-computed for each pixel. This fundamental drawback of re-computation of the Gram matrix is eliminated in the proposed reformulation of the Kernel RX detector.

Since the Kernel matrix does not change from pixel to pixel, the SW-KRX detector statistics of Eqn. (3.31) can be combined in vector form, for all the pixels vectors in the image. From Eqn. (3.23), it can be seen that the vector $\hat{\mathbf{k}}_{r_w}$ represents a column of the centered kernel Gram matrix $\hat{\mathbf{K}}_w$. The detector statistics \mathbf{q} for all the pixels in the image in combined vector form can be written as:

$$\mathbf{W}_0 \mathbf{q} = \text{diag}(\hat{\mathbf{K}}_w^T (\boldsymbol{\alpha}_w \boldsymbol{\Lambda}_{\Phi w}^{-2} \boldsymbol{\alpha}_w^T) \hat{\mathbf{K}}_w) \quad (3.39)$$

where \mathbf{q} is the $N \times 1$ vector containing the SW-KRX detector statistic for each pixel, i.e., $\mathbf{q} = [SWKRX(\mathbf{y}_1) \ SWKRX(\mathbf{y}_2) \ \dots \ SWKRX(\mathbf{y}_N)]^T$. Eqn. (3.39) gives the weighted detector statistic for all the pixels, and denoting the $\mathbf{q}_w = \mathbf{W}_0 \mathbf{q}$, it can be written as:

$$\mathbf{q}_w = \text{diag}(\hat{\mathbf{K}}_w^T (\boldsymbol{\alpha}_w \boldsymbol{\Lambda}_{\Phi w}^{-2} \boldsymbol{\alpha}_w^T) \hat{\mathbf{K}}_w) \quad (3.40)$$

Eqn. (3.40) gives the expression for the weighted SW-KRX detector statistics.

Recall that Eqn. (3.23) gives the definition of the weighted centered kernel Gram matrix $\hat{\mathbf{K}}_w$. The $(i,j)^{\text{th}}$ element of this matrix is from Eqn. (3.24) is given as:

$$[\hat{\mathbf{K}}_w]_{ij} = \sqrt{w_i} \sqrt{w_j} k(\mathbf{y}_i, \mathbf{y}_j)$$

Since the term of the centered kernel matrix is based on both the feature vectors, the total weight can be written as a function of the matrix entry position, i.e.,

$$\sqrt{w_i} \sqrt{w_j} = w_{ij} \quad (3.41)$$

Defining \mathbf{W} as the matrix which has the combined weights w_{ij} as entries, i.e.

$$[\mathbf{W}]_{ij} = w_{ij} \quad (3.42)$$

Note that the diagonal elements of the weight matrix \mathbf{W} are the same as that of the weight matrix \mathbf{W}_0 . Then, the weighted centered kernel Gram matrix can be written:

$$\hat{\mathbf{K}}_W = \hat{\mathbf{K}} * \mathbf{W} \quad (3.43)$$

where “*” denotes the element wise multiplication operation on two matrices, and $\hat{\mathbf{K}} = \mathbf{Y}_{CW}^\Phi{}^T \mathbf{Y}_{CW}^\Phi$ is the kernel Gram matrix defined on *centered* feature dataset. The matrix $\hat{\mathbf{K}}$ can be computed in terms of the weight matrix \mathbf{W} and the *unweighted uncentered* kernel Gram matrix \mathbf{K} , as follows:

$$\hat{\mathbf{K}} = \mathbf{K} - \mathbf{\Omega} \mathbf{W}^T \mathbf{K} - \mathbf{K} \mathbf{W} \mathbf{\Omega} + \mathbf{\Omega} \mathbf{W}^T \mathbf{K} \mathbf{W} \mathbf{\Omega} \quad (3.44)$$

where, \mathbf{W} is the weight matrix as defined in Eqn. (3.42) and $\mathbf{\Omega}$ is diagonal matrix given as:

$$[\mathbf{\Omega}]_{ii} = \frac{1}{\sum_{j=1}^N w_{ij}} \text{ and } [\mathbf{\Omega}]_{ij} = 0 \quad \forall \quad i \neq j \quad (3.45)$$

The detailed derivation of the relationship in Eqn. (3.44) is provided in Appendix D.

At this point a special case of the SW-KRX detector is considered. Eqn. (3.40) can be re-written in terms of the Eqn. (3.25) as:

$$\mathbf{q}_W = \text{diag}((\mathbf{a}_W \mathbf{\Lambda}_{\Phi W} \mathbf{a}_W^T)^T (\mathbf{a}_W \mathbf{\Lambda}_{\Phi W}^{-2} \mathbf{a}_W^T) (\mathbf{a}_W \mathbf{\Lambda}_{\Phi W} \mathbf{a}_W^T)) \quad (3.46)$$

In the case when the constituting dimensions of the mapped feature vector $\Phi(\mathbf{y}_i)$ are unit variance and uncorrelated, i.e., $\mathbf{\Lambda}_{\Phi W} = \mathbf{I}$, Eqn. (3.46) can be written as:

$$\mathbf{q}_W = \text{diag}(\mathbf{a}_W (\mathbf{a}_W^T \mathbf{a}_W) (\mathbf{a}_W^T \mathbf{a}_W) \mathbf{a}_W^T) = \text{diag}(\mathbf{a}_W \mathbf{a}_W^T) \quad (3.47)$$

Since in the special case, $\hat{\mathbf{K}}_W = \mathbf{a}_W \mathbf{a}_W^T$, the SW-KRX detector statistics case is:

$$\mathbf{q}_W = \text{diag}(\hat{\mathbf{K}}_W) \quad (3.48)$$

Since $\mathbf{q}_W = \mathbf{W}_0 \mathbf{q}$, and given Eqn. (3.44), Eqn. (3.48) can be written as:

$$\mathbf{W}_0 \mathbf{q} = \text{diag}(\hat{\mathbf{K}} * \mathbf{W}) \quad (3.49)$$

Since the diagonal elements of the weight matrix \mathbf{W} are the same as that of the weight matrix \mathbf{W}_0 , the SW-KRX detector statistics (unweighted) for the special case is given as:

$$\mathbf{q} = \text{diag}(\hat{\mathbf{K}}) \quad (3.50)$$

That is, the detector statistics is the diagonal of the weighted-centered kernel Gram matrix.

3.3.2. Selection of Weights. As described in Eqn. (3.42), $(i,j)^{\text{th}}$ element of the weight matrix W , i.e. w_{ij} , is some function of the i^{th} and j^{th} pixels from the background set. Although the concept of masks is eliminated in the aforementioned version of Kernel RX, as mentioned earlier, the local adaptivity of the detector is enforced using the set of weights. To this end, the weights in the weighted kernel matrix are chosen to be functions of their relative positions. Let $\mathbf{p}_i = (r_i, c_i)$ be the spatial position vector of the i^{th} pixel in the image. Let $d_w(\mathbf{p}_i, \mathbf{p}_j)$ be a distance function between the i^{th} and j^{th} pixels. For the proposed SW-KRX, a continuous 2-D distance function is adopted. The continuous function chosen as the distance function is the same as the kernel function k^2 . For the work in this dissertation, the kernel function, and hence the spatial distance function, is chosen to be the exponential function, i.e.,

$$d_w(\mathbf{p}_i, \mathbf{p}_j) = k(\mathbf{p}_i, \mathbf{p}_j) = \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{\gamma^2}\right) \quad (3.51)$$

Similar to the binary distance function, the exponential distance function gets smaller for the pair of pixels that are spatially separated, however in a continuous fashion. The width of exponential distance function γ essentially governs the falloff of the function with respect to the spatial distance. The smaller the width γ of the exponential distance function, the smaller the region of influence around any given pixel.

For the Gaussian kernel, the exponential spatial weight term $w_{ij} = d_w(\mathbf{p}_i, \mathbf{p}_j)$, and the exponential kernel term $k(\mathbf{y}_i, \mathbf{y}_j)$ can be combined into one term. The augmented spatial-spectral pixel vectors in the image are defined as:

$$\mathbf{x}_i = [\mathbf{y}_i \quad \frac{\sigma}{\gamma} r_i \quad \frac{\sigma}{\gamma} c_i]^T \quad \forall i = 1, 2, \dots, N \quad (3.52)$$

² This is assuming that the kernel function ‘ k ’ is a valid distance function, like the Gaussian kernel function used in this work.

where \mathbf{y}_i is the $J \times 1$ pixel vector at location $\mathbf{p}_i = (r_i, c_i)$. Thus, \mathbf{x}_i is a $d \times 1$ vector, where $d = J + 2$. The creation of the augmented spectral-spatial vectors allows us to combine the weights and the kernel terms together, and define it directly in terms of the kernel operation on the augmented vectors. This allows us to use the same machinery as used for fast computation of the kernel matrix, proposed in the next section, for the computation of the weight terms too.

3.4. RESULTS: SIMULATED DATA

In this section, some preliminary detection performance results of the proposed SW-KRX detector over simulated data are presented. For the purposes of comparison and accuracy of approximation, detection performance results for four other detectors, namely the RX detector, the Kernel RX detector, and two modified versions of Kernel RX detector are also presented.

The data is simulated for two different types of background distribution, namely single Gaussian and Gaussian mixture distributions. For each type of background distribution, 25×25 patches are generated at random. 1000 patches are generated with a target pixel at the center, which has a different mean from the background. Another 1000 patches are generated without any target pixel, and these are the patches with only the background. The number of bands in the spectral vectors for both the simulated datasets is $J = 3$.

Note that the proposed method is slightly different from the Kernel RX algorithm in that it computes the kernel Gram matrix over all the pixels in the image, including the target pixel(s). However, if the number of target pixels is small in comparison to the total number of background pixels, their effect on the final Kernel RX statistic is expected to be negligibly small. To demonstrate this, the results for the first modified version of the Kernel RX, termed T-KRX (Target-KRX), are included. This detector is the same as the original Kernel RX, except that the kernel Gram matrix is defined including the target pixel(s). In the second modified version of the Kernel RX algorithm, instead of the computing the Kernel-RX statistic as in Eqn. (3.12), it is obtained directly from the weighted-centered kernel Gram matrix as in Eqn. (3.50) (the special case). This modified version is termed Direct T-KRX (“Direct”, since no matrix inversion is needed). Note

that the Direct T-KRX detector is the same as the proposed method, except that there is no spatial weighting based on the spatial dimensions. The idea behind showing the detection performance results for the T-KRX and Direct T-KRX detectors is to systematically analyze the effect on the two key modifications in the proposed method over the original Kernel RX detector, namely the inclusion of the target pixel in kernel Gram matrix definition, and direct computation of the statistic from the centered weighted kernel Gram matrix.

For RX, Kernel RX, T-KRX and Direct T-KRX, a target mask diameter of 1, i.e., only one target pixel, is chosen. The width of the Gaussian kernel function for Kernel RX, T-KRX, Direct T-KRX, and the proposed method is $\sigma = 32.33$. The width of the Gaussian kernel used as the spatial distance function is set at $\gamma = 13$.

Figure 3.2 shows examples of simulated patches for the Gaussian distributed data. Figure 3.2a shows a non-mine patch, i.e., no target is present, and Figure 3.2b shows a patch with a single pixel target present at the center. Figure 3.3 shows the detection performance results using ROC curves for the Gaussian distributed background data. The standard deviation of the background data is 5, and the target mean is $[9 \ 8 \ 9]^T$ and $[11 \ 11 \ 12]^T$ for Figures 3.3a and 3.3b, respectively, corresponding to the medium SNR and high SNR. It can be seen that the detection performances of all the detectors are similar.

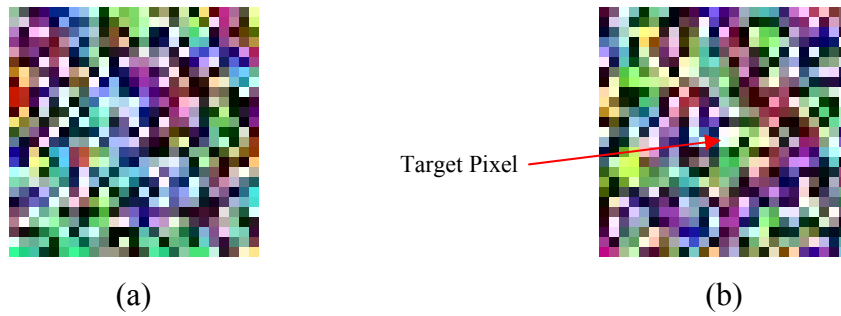


Figure 3.2. Simulated Patch for Gaussian Distributed Data (a) No Target Present (b) Target Present.

The ROC curves for the T-KRX and Direct T-KRX detectors are almost the same as the original Kernel RX algorithm, corroborating the accuracy of the detector statistic

computation as given in Eqn. (3.50), and the fact that the inclusion of the target pixel has negligible effect. It should be noted that in case of Gaussian distributed background, the RX algorithm is theoretically optimal, and should, therefore, give the best performance amongst all detectors.

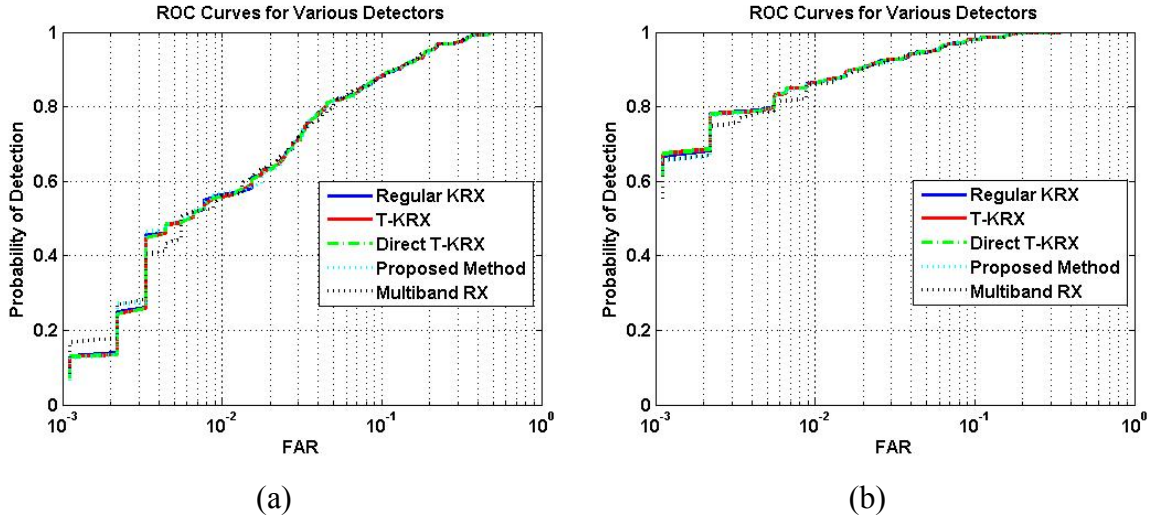


Figure 3.3. Comparative Detection Performance for Various Detectors for Gaussian Distributed Data (a) Medium SNR (b) High SNR.

However, it is noted that the performance of the Kernel RX algorithm is at par with the RX detector in this case as well. Also, as can be seen, the performance of the proposed detector, even with spatial weighting, is as good as the RX and the Kernel RX detectors.

Although, the RX detector is based on the assumption that the background is zero mean and Gaussian distributed, it is not always the case with real world spatial data and images, due to primarily two reasons. In case of remote sensing and aerial imaging, due to thermal noise in the sensors, salt and pepper type noise is observed in the data, which are outliers to the underlying Gaussian distribution. Secondly, it is fairly common to see spatial data containing a mixture of different types of distributions. For example, in case of the airborne multispectral image data, it is common to see images containing two or

more different types of backgrounds, like vegetation, road, sand, and rivers. The data at the interface of two or more of these regions cannot be modeled accurately as a single Gaussian distribution.

In the second set of results, the performance of the various detectors in the aforementioned scenarios is depicted. Figure 3.4 shows sample simulated patches for the case when the data is a mixture of two distinct background regions, similar to the condition mentioned above. The proportion of the two background regions in the patches is 80:20. Figure 3.4a shows a patch where target is not present, and the patch shown in Figure 3.4b contains a target at the center.

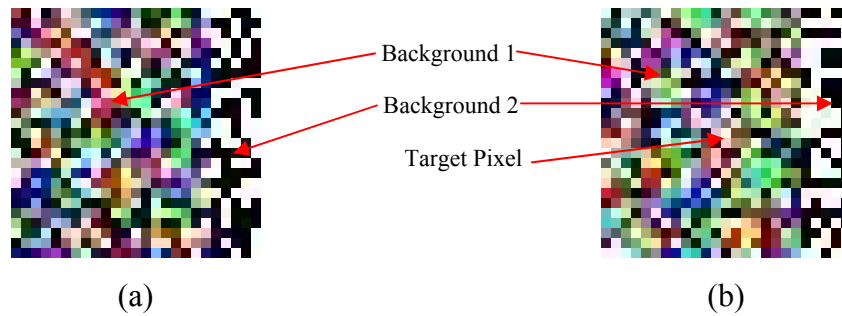


Figure 3.4. Simulated Patch for Mixed Background in 80:20 Proportion (a) No Target Present (b) Target Present.

Figure 3.5 depicts the detector performance results using Receiver Operating Characteristic (ROC) curves for data simulated for this case. The x-axis depicts the probability of detection and the y-axis depicts the False Alarm Rate (FAR). The proportion of the two regions in the background is 90:10, 80:20, 70:30, and 60:40 for Figures 3.5a, 3.5b, 3.5c and 3.5d, respectively. It can be seen from Figure 3.5 that, since the data is no longer Gaussian distributed, the performance of the RX detector deteriorates rapidly. However, the performance of the Kernel RX, T-KRX, Direct T-KRX, and the proposed detector, does not suffer as badly and is much superior to the RX detector. This robustness of the Kernel RX and the proposed detectors can be attributed to their ability to effectively reject outliers from the data due to non-linear weighting.

Therefore, these results are indicative of the fact that Kernel RX detector is likely to detect targets which lie on the interface on two or more regions, or data with outliers much more accurately than the RX detector.

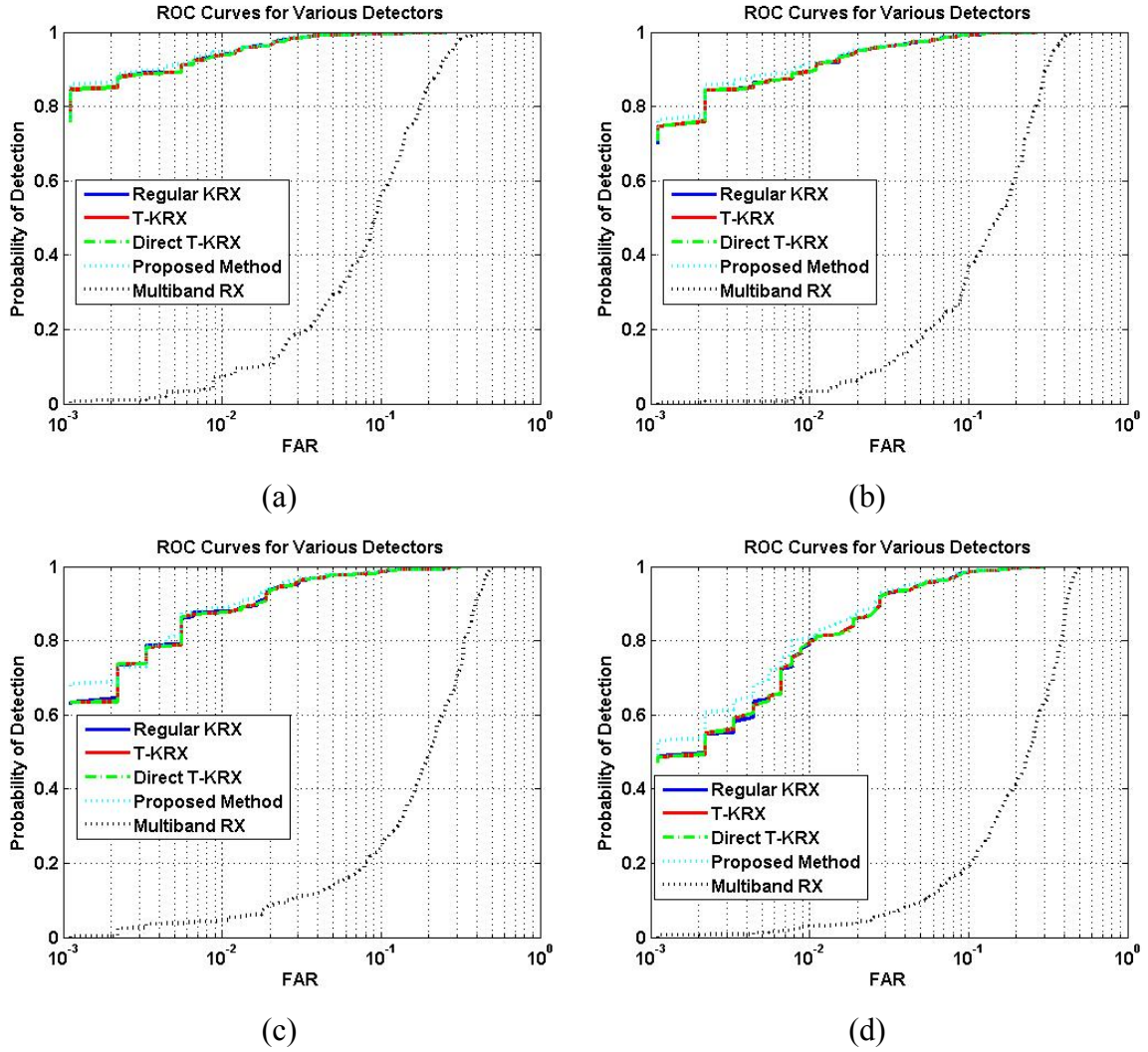


Figure 3.5. Comparative Detection Performance for Various Detectors for Mixed Background with Different Proportions (a) 90:10, (b) 80:20, (c) 70:30, and (d) 60:40 Mixture.

4. FAST APPROXIMATE COMPUTATION OF KERNEL GRAM MATRIX

4.1. COMPUTATIONAL BURDEN IN SW-KRX

In the SW-KRX algorithm presented in Section 3, the detector statistic for the image can be obtained directly from the *centered* Gram matrix computed over the entire image. The problem then becomes that of computing the centered weighted Gram matrix, $\hat{\mathbf{K}}_w$ and consequently the original Gram matrix \mathbf{K} efficiently. It should be noted that the matrix \mathbf{K} under this framework is an $N \times N$ matrix, which even for small size images can be extremely large in dimensions. For example, for a small 64x64 image, \mathbf{K} will be a 4096x4096 matrix. The brute force calculation of the Gram matrix is a $O(N^2J) = O(M^4J)$ computation, which is impracticable even for small images. In this section, the methodology behind fast calculation of the Gram matrix \mathbf{K} , and the weighted centering for the diagonal elements is developed for the widely used and well-known Gaussian kernel.

4.2. SPECTRAL-SPATIAL CLUSTERING

The key idea behind the development of a fast method for the computation of the kernel Gram matrix for the SW-KRX detector statistic is to move from individual pixel based computations to cluster-based computations. That is, instead of computing statistics for one pixel at a time, the pixels can be grouped into clusters and the computations can be done for the whole cluster at a time efficiently. Similar ideas for fast computation of the Gauss transforms has been proposed in techniques like Improved Fast Gauss Transform (IFGT) [Raykar, 2005a, 2005b] and KD-trees [Gray, 2001]. The Improved Fast Gauss Transform (IFGT) method is based on a wider class of algorithms called Fast Multipole Methods, which were developed for problems in computational physics [Greengard, 1987]. The basic idea behind the IFGT method is also to cluster the data points involved in the Gauss transform, and approximate the effect of clusters as a whole, instead of individual data points.

As a first step to the fast implementation of the proposed SW-KRX detector, the spectral vectors in the image need to be clustered into spectral-spatial clusters. Several algorithms exist in the literature, based on techniques like vector quantization, watershed

segmentation [Vincent, 1991; Gauch, 1999], quad-tree or oct-tree decomposition [Spann, 1985, 1989; Wilson, 1988; Lee, 1989], and Voronoi tessellation [Okabe, 1992; Horn, 1986], which can possibly be used for spatial-spectral clustering of the image. In Section 4.2.1, a brief review of some of the methods that can be used for the aforementioned spatial-spectral clustering is presented. This is followed by the description of the watershed-based clustering technique adopted for spectral-spatial clustering in this work.

4.2.1. Clustering Techniques: Review. For the fast implementation of the SW-KRX algorithm, the spatial data has to be “tessellated” with spectral-spatial clusters. Consider a discrete spatial dataset, with a set of points $P = \{\mathbf{p}_i\}_{i=1}^N \subseteq \mathbb{R}^d$. In case of the uniformly sampled image dataset, set P is the location of all the pixels in the image and the spatial dimensionality is $d = 2$. Then, a set $S = \{s_i\}_{i=1}^G$ is called a “tessellation” of P , if $s_i \cap s_j = \emptyset \forall i \neq j$ and $\cup_{i=1}^G s_i = P$. In this work, the goal is to tessellate the spatial data using a criterion based on spectral similarity and spatial vicinity. Several techniques in image segmentation and data clustering literature have been proposed for the generation of such tessellations on spatial data. A comprehensive review of such techniques is beyond the scope of this dissertation. The focus in this work being on fast computation, the important issue in selecting a clustering technique is that of computational speed. In this section, a brief review of the some of the more suitable techniques is presented.

One important category of useful techniques is that of algorithms based on Voronoi tessellations. The first formal studies of spatial tessellations were done by Dirichlet [1850] and Voronoi [1908]. Their main contribution was the formalization of the idea of partitioning of a space by considering a set of source points and assigning each point to the “closest” source point. There are several applications of the concept of Voronoi tessellations. For a detailed description of the various concepts in Voronoi tessellations and the computational algorithms and their variants, the reader is referred to [Okabe, 1992]. Earlier work based on Voronoi tessellations in image segmentation and vision can be found in [Horn, 1986]. A variant of Voronoi tessellation is the concept of Centroidal Voronoi Tessellations (CVT). The key idea in CVT is that the source points

are also the centroids of the clusters, which are determined algorithmically in an iterative fashion. Several applications of CVT have been proposed [Du, 1999, 2002; Hausner, 2001; Kanungo, 2002], and they have found several applications in image processing, specifically image segmentation [Du, 2006]. CVT is an intuitive clustering strategy and its simplest form is the same as the popular k -means clustering strategy [Kanungo, 2002; Inaba 1994]. Understanding the k -means algorithms in the mathematical framework of CVT allows significant extensions and improvements in the k -means algorithm, and also other clustering strategies. Arbelaez et al. [2003] recently proposed the concept of generalized Voronoi tessellations for the segmentation of vector-valued images. Their key contribution is the definition of a pseudo-metric on the vector-valued image data and a novel technique for the selection of the source points. Their pseudo-metrics are able to accurately describe the structure of the images with relatively homogeneous regions.

Another set of techniques is region-based split and merge segmentation techniques. There is an extensive body of literature on this category of techniques. The goal is to tessellate the image into regions, based on some homogeneity criteria. In merging, first a primitive tessellation of the image is derived. As a correction step, neighbors with similar characteristics, as defined by some criteria, are merged. The starting point in splitting techniques, is the entire rectangular image, which is iteratively split into four rectangular regions (in case of 3D data, eight regions) until each region satisfies a homogeneity criterion. The key component in these techniques is the homogeneity criteria, which decides the final segmentation outcome. For example, in the early work, Chen et al. [1981] proposed a quad-tree like linked pyramidal structure and a statistical criterion for combining global and local region statistics for image modeling. Improvements and variants of their technique were later developed in other efforts [Burt, 1981; Hong, 1984a, 1984b]. Spann and Wilson [1985] were amongst the first to propose a quad-tree structure based method which used clustering of a histogram at a low spatial resolution, followed by boundary refinement. Other algorithms on similar lines were also developed in [Wilson, 1988; Spann, 1989]. Schroeter et al. [1995] experimented with different clustering algorithms and also proposed an improved boundary refinement step using adaptive filtering. Another interesting method in this category was that proposed by Lee [1989]. Although the goal there was primarily compression, the basic idea was to

determine an optimal tessellation of the image. This method combines the regular splitting, and the quad-tree data structure of the split and merge techniques with the general threshold based region extraction method of the recursive splitting techniques.

Another important category of techniques for generating tessellation on images is watershed transformation based techniques. Watershed transformation, which is inspired from the notions in topography, has received considerable attention in the last two decades. Watershed is defined as the lines that divide a given region into the so called *catchment basins*, which are essentially the minima in the topographical surface. Any drop of water falling in the catchment basins will reach the minima associated with it. Watershed transformation in images essentially is the process of finding the watershed pixels, and consequently the catchment basins or segments, in the topographical surface created by the image function. This idea is used for image segmentation and tessellation in that anytime a secondary image can be formed such that it has high ridges at the region edges and low values on the region themselves, the watershed pixels will outline the various regions. Watershed transformation is particularly effective when combined with other morphological and region handling techniques.

The introduction of watershed transformation to the field of image processing is credited to Digabel et al. [1978]. However, Vincent and Soille [1991] proposed one of the first popular computational algorithms for computing the watershed transformation. Their algorithm is based on immersion simulation, and was the basis of several subsequent efforts. Another significant effort was that by Gauch et al. [1999], where the watershed segmentation was done on the gradient magnitude image, followed by non-linear filtering process. The resulting algorithm was simpler than the Vincent and Soille algorithm, but gave similar results. Although, the watershed algorithm results in a natural tessellation of the image, in most real images, it leads to over-segmentation. Therefore, several algorithms have been proposed which use region growing and region merging on the primitive regions defined by watersheds, for achieving accurate segmentation [Haris, 1998; Wang, 1997; Bleau, 2000; Makrogiannis, 2001]. For instance, Haris et al. [1998] proposed a gradient and region-based hybrid method, which used the primitive regions generated by watersheds and applies an efficient region merging process based on Region Adjacency Graphs (RAG) for final segmentation. Hernandez et al. [2000] also proposed a

hybrid region merging criterion based on edge integrity and region homogeneity for reducing oversegmentation. In an example of the application of watershed in real-world systems, Chen et al. [2007] have proposed a watershed-based adaptive skin lesion segmentation technique for dermoscopy images. They propose a object histogram based region merging method for reducing the oversegmentation, and conclude that watershed-based segmentation method outperforms several other automated techniques for this application. With the analysis of multiscale watersheds [Gauch, 1993; Pratikakis, 1999; Olsen, 1997; Jackway, 1996], several segmentation algorithms using morphological pyramids [Salembier, 1995; Meyer, 1997; Wright, 1997; Bosworth, 1998] and multiscale hierarchies [Pratikakis, 1999; Olsen, 1997] have also been proposed.

The clustering literature is vast, and there are several other techniques which may be applicable to the present problem. The interested reader is referred to the excellent surveys of data clustering techniques by Xu and Wunsch [2005] and Jain et al. [1999].

4.2.2. Watershed-based Clustering. For the SW-KRX algorithm presented in this section, the watershed-based clustering technique has been adopted. Watersheds produce a natural tessellation of the image and are can be computed speedily. The popular Vincent and Soille algorithm [Vincent, 1991], as implemented in Mathworks Inc.'s Matlab[®] package, was used for watershed transformation.

Although watersheds typically lead to the problem of oversegmentation, this was not a major issue in the implementation of the SW-KRX detector algorithm. The reason for this is that the proposed SW-KRX detector entails an upper bound on cluster size along each dimension, for more accurate approximation using multivariate Taylor series. This bound should hold for all clusters and for all the dimensions, including the spatial dimensions. A detailed explanation and development of this bound is presented in following Section 4.3. In light of the bound and the empirical results on watershed on the spatial image data, it was found that the natural clusters generated by watershed were of appropriate size and oversegmentation was not a major issue.

As a first step, each image frame is passed through a Gaussian smoothing filter, followed by the computation of the multiband gradient magnitude image. Watershed transformation was then applied to the gradient magnitude images. Each watershed region (which has a different label) is considered a spatial-spectral cluster, and is

subsequently checked for the spatial and spectral bounds given above. If a cluster fails to meet the bound along any dimension, it is split along the median point of the cluster along that dimension, and the resulting clusters are checked again for all dimensions. At the end of this process, each cluster meets the bound and consists of a group of spatially close and spectrally similar pixels.

It should be noted that the watershed pixels themselves do not belong to any cluster, and in that sense watershed transformation does not give a complete tessellation. However, instead of posing a problem, this creates a choice in the clustering process. Since the watershed transformation is done on the gradient magnitude image, the watershed pixels typically lie on region edges. The edge pixels typically are an overlap of pixel intensities from different regions and belong to a particular region only partially. In such a case, it might be better to leave the edge pixels out of any cluster instead of “corrupting” the cluster, which, as shown later, leads to more computation. However, since the SW-KRX detector statistic is calculated only for the pixels belonging to the clusters, it might not be a viable option to leave out edge pixels from the clusters when the statistics have to be computed for the entire image (or on any set including any subset of the set of the edge pixels). In such a case, there are potentially different ways to distribute the watershed pixels into clusters, so as to make them part of the detector statistics computation.

In this work, the following approach is adopted for this distribution of watershed pixels. Each edge pixel is compared with each one of eight neighbors in its immediate vicinity. The edge pixel is assigned to the cluster to which its spectrally closest neighbor belongs. There might be cases where the most similar neighbor of a watershed pixel is also a watershed pixel (in which case both of them will remain watershed pixels). In such cases, each watershed pixel is considered a single-pixel cluster in itself. In the results depicted in this dissertation, the watershed pixels are assigned to clusters using the technique outlined above. At the end of the watershed-based spectral-spatial clustering, the entire image is tessellated with compact clusters, which satisfy the bound.

4.3. FAST COMPUTATION OF KERNEL GRAM MATRIX

Let us assume that the augmented spatial-spectral vectors $\{\mathbf{x}_i\}$ of the image are clustered into G clusters. The clusters are denoted by $\{s_k \mid k = 1, 2, \dots, G\}$, and the number of pixels vectors in the k^{th} cluster by N_k . The spatial centroid of the k^{th} cluster is denoted by \mathbf{c}_k^P , and the spectral centroid is denoted by \mathbf{c}_k^S . The combined spatial-spectral centroid is denoted by $\mathbf{c}_k = [\mathbf{c}_k^S \ \mathbf{c}_k^P]^T$. The clustering is such that it should satisfy the following three conditions:

$$\begin{aligned}
 &1. \mathbf{x}_i \in s_k, \quad k \in \{1, 2, \dots, G\} \quad \forall i = 1, 2, \dots, N \\
 &2. \mathbf{x}_i \in s_k, \quad \Rightarrow \mathbf{x}_i \notin s_l \quad l \neq k \\
 &3. \bigcup_{k=1}^G s_k = \{\mathbf{x}_i \mid i = 1, 2, \dots, N\} \Leftrightarrow \sum_{i=1}^G N_i = N
 \end{aligned} \tag{4.1}$$

Since, the pixels in the image are clustered into spectral-spatial clusters, they can be indexed in the order of the clusters they belong to, i.e., all the pixels in the first cluster together and then the second cluster and so on.

$$\mathbf{K} = \begin{matrix} & \begin{matrix} \text{1st Cluster} & \text{2nd Cluster} & \dots & \text{Gth Cluster} \end{matrix} \\ \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & \dots & k(\mathbf{x}_1, \mathbf{x}_N) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \dots & \dots & k(\mathbf{x}_2, \mathbf{x}_N) \\ - & - & \dots & \dots & - \\ - & - & \dots & \dots & - \\ k(\mathbf{x}_N, \mathbf{x}_1) & - & \dots & \dots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \end{matrix} \tag{4.2}$$

As shown in Eqn. (4.2), since the indexing of pixels is such that the pixels under a cluster are placed together, the kernel Gram matrix \mathbf{K} can be thought of as a block matrix. Each block in the matrix denotes a sub-matrix formed by calculating the kernel distance between elements of one cluster with elements of another. Let us assume that the image is clustered into G spectral-spatial clusters. Let the G spectral-spatial clusters, be denoted by $\{s_k \mid k = 1, 2, \dots, G\}$.

Gram matrix \mathbf{K} can be written as:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} & \cdots & \mathbf{K}_{1G} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} & \cdots & \mathbf{K}_{2K} \\ - & - & - & \cdots & - \\ - & - & - & \cdots & - \\ \mathbf{K}_{G1} & \mathbf{K}_{G2} & \mathbf{K}_{G3} & \cdots & \mathbf{K}_{GG} \end{bmatrix} \quad (4.3)$$

where, the sub-matrix element \mathbf{K}_{kl} of the block matrix \mathbf{K} is given by:

$$[\mathbf{K}_{kl}]_{ij} = \exp\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right), \quad \forall \mathbf{x}_i \in s_k \text{ \& } \mathbf{x}_j \in s_l \quad (4.4)$$

Thus, \mathbf{K}_{kl} is a $N_k \times N_l$ matrix, the entires of which are calculated between elements of clusters k and l . It is noted that the blocks on the diagonal of the matrix are square and symmetric, and also $\mathbf{K}_{kl} = \mathbf{K}_{lk}^T$.

The goal is to calculate the matrix \mathbf{K} one block at a time, efficiently. A methodology is now proposed to get an approximation for each block, which can be computed efficiently. Consider the (i, j) element of the block \mathbf{K}_{kl} from Eqn. (4.3), which is the combination of the points from clusters s_k and s_l . It can be written as:

$$\begin{aligned} [\mathbf{K}_{lk}]_{ij} &= \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{c}_l + \mathbf{c}_l - \mathbf{x}_j + \mathbf{c}_k - \mathbf{c}_k\|^2}{\sigma^2}\right) \\ &= \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_l) + (\mathbf{c}_l - \mathbf{c}_k) - (\mathbf{x}_j - \mathbf{c}_k)\|^2}{\sigma^2}\right) \end{aligned} \quad (4.5)$$

where, \mathbf{c}_k and \mathbf{c}_l are the centroids of the clusters s_k and s_l , respectively. This can be expanded as

$$\begin{aligned} [\mathbf{K}_{kl}]_{ij} &= \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_k)\|^2 + \|(\mathbf{x}_j - \mathbf{c}_l)\|^2 + \|(\mathbf{c}_k - \mathbf{c}_l)\|^2}{\sigma^2}\right. \\ &\quad \left.- \frac{2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{x}_j - \mathbf{c}_l) + 2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{c}_l - \mathbf{c}_k) - 2(\mathbf{x}_j - \mathbf{c}_l) \cdot (\mathbf{c}_k - \mathbf{c}_l)}{\sigma^2}\right) \end{aligned} \quad (4.6)$$

Regrouping the terms, Eqn. (4.26) can be written as:

$$[\mathbf{K}_{kl}]_{ij} = \exp\left(-\frac{\|(\mathbf{c}_k - \mathbf{c}_l)\|^2}{\sigma^2}\right) q_{ikl} q_{jlk} \exp\left(\frac{2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{x}_j - \mathbf{c}_l)}{\sigma^2}\right), \mathbf{x}_i \in S_k, \mathbf{x}_j \in S_l \quad (4.7)$$

where,

$$q_{ikl} = \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_k)\|^2}{\sigma^2}\right) \exp\left(-\frac{2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{c}_k - \mathbf{c}_l)}{\sigma^2}\right) \quad (4.8)$$

The last term in Eqn. (4.8) can be approximated using the truncated multivariate Taylor series [Raykar, 2005a, 2005b]. To this end, at this point, the multi-index notation from [Raykar, 2005a] is introduced. Let $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_d]^T$ is a vector of the same length as the dimension of the spectral-spatial vectors \mathbf{x}_i , and following operations are defined over the vector $\boldsymbol{\alpha}$:

$$1. |\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + \dots + \alpha_d \quad (4.9a)$$

$$2. \mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_d^{\alpha_d} \quad (4.9b)$$

$$3. \boldsymbol{\alpha}! = \alpha_1! \alpha_2! \dots \alpha_d! \quad (4.9c)$$

$$4. (\mathbf{x} \cdot \mathbf{y})^n = \sum_{|\boldsymbol{\alpha}|=n} \mathbf{x}^{\boldsymbol{\alpha}} \mathbf{y}^{\boldsymbol{\alpha}} \quad (4.9d)$$

Using the multi-index notation, the multivariate Taylor series approximation of the last term in Eqn. (4.8) can be written as [Raykar, 2005a]:

$$\exp\left(\frac{2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{x}_j - \mathbf{c}_l)}{\sigma^2}\right) = \sum_{|\boldsymbol{\alpha}| \leq p-1} \frac{2^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \left(\frac{\mathbf{x}_i - \mathbf{c}_k}{\sigma}\right)^{\boldsymbol{\alpha}} \left(\frac{\mathbf{x}_j - \mathbf{c}_l}{\sigma}\right)^{\boldsymbol{\alpha}} \quad (4.10)$$

It is desirable that an accurate Taylor series approximation, up to a desired accuracy, be obtained in the fewest possible terms. For a compact Taylor series representation, it is essential that the argument of the exponential function be a small number. To this end, an upper bound of 1 is enforced on the argument of the exponential function in Eqn. (4.10). The exponential on the dot product can be written as the multiplication of individual exponentials on individual dimensions. Using this, the bound

on the argument of the exponential can be translated into the following bound on each individual dimension:

$$\left| \frac{2(x_{in} - c_{kn}) \cdot (x_{jn} - c_{ln})}{\sigma^2} \right| \leq 1 \quad \forall \quad \mathbf{x}_i \in s_k \text{ \& } \mathbf{x}_j \in s_l \text{ \& } n = 1, 2, \dots, d \quad (4.11a)$$

where, x_{in} and c_{kn} denote the n^{th} dimension of the spectral-spatial vectors \mathbf{x}_i and \mathbf{c}_k , respectively. In terms on each clusters this translates into the bound:

$$\max_{x_{in} \in s_k} \frac{(x_{in} - c_{kn})}{\sigma} \leq \frac{1}{\sqrt{2}} \quad \forall \quad n = 1, 2, \dots, d \quad (4.11b)$$

Eqn. (4.11b) should hold for all clusters and for all the dimensions. This is taken care of during the first step of spatial-spectral clustering.

Note that Eqn. (4.10) gives the truncated decomposition, with the truncation number p . This decomposition can be made arbitrarily accurate by choosing a high enough truncation number, although at the expense of more computation. The issue of approximation accuracy and computational burden is dealt with in Section 4.4.

However, such decomposition allows us to disentangle the two vectors $(\mathbf{x}_i - \mathbf{c}_k)$ and $(\mathbf{x}_j - \mathbf{c}_l)$ from the dot product and separate them out. Using Eqn. (4.10), Eqn. (4.7) and (4.8) can be modified as follows:

$$[\mathbf{K}_{kl}]_{ij} = \exp\left(-\frac{\|(\mathbf{c}_k - \mathbf{c}_l)\|^2}{\sigma^2}\right) \sum_{|\alpha| \leq p-1} \frac{2^\alpha}{\alpha!} \kappa_{ikl}^\alpha \kappa_{jlk}^\alpha \quad (4.12a)$$

where,

$$\kappa_{ikl}^\alpha = \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_k)\|^2}{\sigma^2}\right) \exp\left(-\frac{2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{c}_k - \mathbf{c}_l)}{\sigma^2}\right) \left(\frac{\mathbf{x}_i - \mathbf{c}_k}{\sigma}\right)^\alpha, \quad \mathbf{x}_i \in s_k \quad (4.12b)$$

From a computational perspective, note that in Eqn. 10, in the calculation of the terms κ_{ikl}^α and κ_{jlk}^α , the terms $\exp\left(-\frac{2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{c}_k - \mathbf{c}_l)}{\sigma^2}\right)$ have to be computed for each point in the cluster and for coupling with every other cluster, which is a $O(N_k G)$ operation. Instead, this term can be approximated by using the truncated multivariate Taylor series again, i.e.,

$$\exp\left(-\frac{2(\mathbf{x}_i - \mathbf{c}_k) \cdot (\mathbf{c}_k - \mathbf{c}_l)}{\sigma^2}\right) = \sum_{|\beta_k| \leq q-1} \frac{2^{\beta_k}}{\beta_k!} \left(\frac{\mathbf{x}_i - \mathbf{c}_k}{\sigma}\right)^{\beta_k} \left(\frac{\mathbf{c}_l - \mathbf{c}_k}{\sigma}\right)^{\beta_k} \quad (4.13a)$$

$$\exp\left(-\frac{2(\mathbf{x}_j - \mathbf{c}_l) \cdot (\mathbf{c}_l - \mathbf{c}_k)}{\sigma^2}\right) = \sum_{|\beta_l| \leq r-1} \frac{2^{\beta_l}}{\beta_l!} \left(\frac{\mathbf{x}_j - \mathbf{c}_l}{\sigma}\right)^{\beta_l} \left(\frac{\mathbf{c}_l - \mathbf{c}_k}{\sigma}\right)^{\beta_l} \quad (4.13b)$$

where q and r are the truncation number for the Taylor series approximation for clusters k and l , respectively. With this approximation, the terms $\left(\frac{\mathbf{x}_i - \mathbf{c}_l}{\sigma}\right)^\beta$ and $\left(\frac{\mathbf{c}_l - \mathbf{c}_k}{\sigma}\right)^\beta$ can be computed separately. Using the decomposition in Eqn. (4.13), Eqn. (4.12) is written as:

$$[\mathbf{K}_{kl}]_{ij} = \sum_{|\alpha| \leq p-1} \sum_{|\beta_k| \leq q-1} \sum_{|\beta_l| \leq r-1} b_{kl}^{\beta_k \beta_l} \kappa_{ik}^{\alpha \beta_k} \kappa_{jl}^{\alpha \beta_l} \quad (4.14)$$

where,

$$\kappa_{ik}^{\alpha \beta_k} = \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_k)\|^2}{\sigma^2}\right) \left(\frac{\mathbf{x}_i - \mathbf{c}_k}{\sigma}\right)^{\alpha + \beta_k} \quad (4.15)$$

$$b_{kl}^{\beta_k \beta_l} = \frac{2^\alpha}{\alpha!} \frac{2^{\beta_k}}{\beta_k!} \frac{2^{\beta_l}}{\beta_l!} \exp\left(-\frac{\|(\mathbf{c}_k - \mathbf{c}_l)\|^2}{\sigma^2}\right) \left(\frac{\mathbf{c}_k - \mathbf{c}_l}{\sigma}\right)^{\beta_k} \left(\frac{\mathbf{c}_l - \mathbf{c}_k}{\sigma}\right)^{\beta_l} \quad (4.16)$$

The terms in Eqns. (4.14), (4.15) and (4.16) are for the (i, j) element of the block matrix \mathbf{K}_{kl} . These equations can be written in vector form for the entire block matrix. Let

$$\mathbf{\kappa}_k^{\alpha \beta_k} = \begin{bmatrix} \kappa_{1k}^{\alpha \beta_k} \\ \kappa_{2k}^{\alpha \beta_k} \\ \vdots \\ \kappa_{N_k k}^{\alpha \beta_k} \end{bmatrix}, \quad \mathbf{x}_i \in s_k, \quad \mathbf{\kappa}_l^{\alpha \beta_l} = \begin{bmatrix} \kappa_{1l}^{\alpha \beta_l} \\ \kappa_{2l}^{\alpha \beta_l} \\ \vdots \\ \kappa_{N_l l}^{\alpha \beta_l} \end{bmatrix}, \quad \mathbf{x}_j \in s_l$$

where N_l and N_k are the number of vectors in the clusters s_l and s_k respectively. Thus, $\mathbf{\kappa}_k^{\alpha \beta_k}$ and $\mathbf{\kappa}_l^{\alpha \beta_l}$ are $N_k \times 1$ and $N_l \times 1$ vectors respectively. Using this notation, the block \mathbf{K}_{kl} as defined by Eqn. (4.14), can be written in vector form as outer product decomposition:

$$\mathbf{K}_{kl} = \sum_{|\alpha| \leq p-1} \sum_{|\beta_k| \leq q-1} \sum_{|\beta_l| \leq r-1} b_{kl}^{\alpha \beta_k \beta_l} \mathbf{\kappa}_k^{\alpha \beta_k} \mathbf{\kappa}_l^{\alpha \beta_l T} \quad (4.17)$$

Eqn. (4.17) can be interpreted as a combination of inter-cluster and intra-cluster interactions. Note that the vectors $\kappa_k^{\alpha\beta_k}$ and $\kappa_l^{\alpha\beta_l}$ depend only on the spectral-spatial vectors from cluster s_k and s_l , respectively. Therefore, the vectors $\kappa_l^{\alpha\beta}$ and $\kappa_k^{\alpha\beta}$ represent the *intra-cluster* interaction for the k^{th} and the l^{th} clusters. Also, the term $b_{kl}^{\beta_k\beta_l}$ in Eqn. (4.17) depends only on the centroids of the k^{th} and the l^{th} clusters, it represents the *inter-cluster* interaction between the two clusters. Thus, the block \mathbf{K}_{kl} can be interpreted as the result of interplay between the inter-cluster and intra-clusters vectors of the k^{th} and the l^{th} clusters.

Note that for the computation of the block lying on the diagonal of the kernel Gram matrix, i.e. \mathbf{K}_{kk} , the computations are considerably simplified. Going back to Eqns. (4.15), (4.16) and (4.17), it can be seen that the diagonal blocks represent the interaction of a cluster with itself. Since $\mathbf{c}_k = \mathbf{c}_l$, all the terms involving β_k and β_l vanish, and Eqn. (4.17) is simplified to:

$$\mathbf{K}_{kk} = \sum_{|\alpha| \leq p-1} \frac{2^\alpha}{\alpha!} \kappa_k^\alpha \kappa_k^{\alpha T} \quad (4.18)$$

where,

$$\kappa_{ik}^\alpha = \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_{kl})\|^2}{\sigma^2}\right) \left(\frac{\mathbf{x}_i - \mathbf{c}_k}{\sigma}\right)^\alpha \quad \& \quad \kappa_k^\alpha = [\kappa_{1k}^\alpha \quad \kappa_{2k}^\alpha \quad - \quad - \quad \kappa_{N_{kk}}^\alpha]^T \quad (4.19)$$

From Eqn. (4.17) it can be seen that the blocks of the kernel Gram matrix can be decomposed into weighted outer products of vectors, defined on the two interacting clusters. In Section 4.4, it is described how this decomposition can be used to efficiently compute the detector statistics for the special case given in Eqn. (3.36).

4.4. FAST APPROXIMATE CENTERING OF KERNEL GRAM MATRIX: DIAGONAL ELEMENTS

Once the outer product decomposition is obtained from either Eqn. (4.17) or (4.18), the centered version of the entire kernel Gram matrix needs to be computed for the computation of the SW-KRX statistics. However, recall that for the special case of the SW-KRX detector, i.e., unit variance of the various dimensions in the feature space, the

statistics is given by: $\mathbf{k} = \text{diag}(\hat{\mathbf{K}})$. In this dissertation, the implementation of the SW-KRX is limited to the special case, and this section presents the method for fast computation of these diagonal elements of the weighted centered Gram matrix.

From Appendix D, the centered kernel Gram matrix can be written in terms of the *uncentered unweighted* kernel Gram matrix \mathbf{K} (defined only on the spectral vectors) as:

$$\hat{\mathbf{K}}_w = \mathbf{K} - \mathbf{\Omega} \mathbf{W}^T \mathbf{K} - \mathbf{K} \mathbf{W} \mathbf{\Omega} + \mathbf{\Omega} \mathbf{W}^T \mathbf{K} \mathbf{W} \mathbf{\Omega} \quad (4.20)$$

where, $[\mathbf{W}]_{ij} = w_{ij} = k(\mathbf{p}_i, \mathbf{p}_j)$ and $\mathbf{\Omega}$ is a $N \times N$ diagonal matrix with $[\mathbf{\Omega}]_{ii} = \sum_{j=1}^N w_{ij}$.

Note that, in this section, the weights are incorporated explicitly by the use of the weight matrix \mathbf{W} , and are not assumed to be the part of the kernel Gram matrix. Similar to the kernel Gram matrix, the weight matrix \mathbf{W} can also be understood as a block matrix, with the block defined for the k^{th} and l^{th} clusters denoted as \mathbf{W}_{kl} . Moreover, since the spatial distance function is also a Gaussian function, the various elements of the block \mathbf{W}_{kl} can be computed efficiently using Taylor series decomposition. In terms of the $(i, i)^{\text{th}}$ pixel, the centering equation for the diagonal term can be written as:

$$\hat{\mathbf{K}}(i, i) = \mathbf{K}(i, i) - \frac{\sum_{k=1}^N w_{ki} \mathbf{K}(k, i)}{\sum_{k=1}^N w_{ki}} - \frac{\sum_{l=1}^N w_{il} \mathbf{K}(i, l)}{\sum_{l=1}^N w_{il}} + \frac{\sum_{l=1}^N \sum_{k=1}^N w_{ki} w_{il} \mathbf{K}(k, l)}{\sum_{k=1}^N w_{ki} \sum_{l=1}^N w_{il}} \quad (4.21)$$

The second term in Eqn. (4.21) is the weighted column mean and the third term is the weighted row mean of the Gram matrix. Since \mathbf{K} is a symmetric matrix the weighted row mean is the same as the weighted column mean.

$$\frac{\sum_{k=1}^N w_{ki} \mathbf{K}(k, i)}{\sum_{k=1}^N w_{ki}} = \frac{\sum_{l=1}^N w_{il} \mathbf{K}(i, l)}{\sum_{l=1}^N w_{il}} \quad (4.22)$$

Also, the diagonal terms in the Gram matrix are unity, i.e.,

$$\mathbf{K}(i, i) = 1 \quad (4.23)$$

The weighted row mean term are denoted by:

$$m_i \equiv \frac{\sum_{l=1}^N w_{il} \mathbf{K}(i, l)}{\sum_{l=1}^N w_{il}} \quad (4.24)$$

If the i^{th} pixel belongs to the k^{th} cluster, the weighted row mean for the i^{th} pixel can be written in terms of the block matrices as:

$$m_i = \frac{\sum_{n=1}^G \sum_{j=1}^{N_l} [\mathbf{W}_{kn}]_{ij} [\mathbf{K}_{kn}]_{ij}}{\sum_{n=1}^G \sum_{j=1}^{N_l} [\mathbf{W}_{kn}]_{ij}}, \mathbf{x}_i \in S_k \quad (4.25)$$

Since the spatial distance function is the same as the kernel function, the numerator term can be computed efficiently by working with augmented spectral-spatial vectors $\{\mathbf{x}_i\}$. For the denominator term the same computing framework can be used, except that the position vectors $\{\mathbf{p}_i\}$ alone are used in place of the augmented spectral-spatial vectors.

The fourth term in Eqn. (4.21) is the weighted Gram matrix mean. This term entails N^2 computations for the centering of each term of the $N \times N$ kernel Gram matrix. That is an order $O(N^4)$ complexity operation, and consequently has huge computational load. Therefore, instead of the actual weighted Gram matrix mean, an approximation is used. The weighted Gram matrix mean is approximated as the weighted average of the weighted row means, i.e.,

$$\frac{\sum_{l=1}^N \sum_{k=1}^N w_{ki} w_{il} \mathbf{K}(k, l)}{\sum_{k=1}^N w_{ki} \sum_{l=1}^N w_{il}} \approx \frac{\sum_{j=1}^N w_{ij} m_j}{\sum_{j=1}^N w_{ij}} \quad (4.26)$$

Eqn. (4.26) gives only an estimate of the actual fourth term, and is done only for computational purposes. The approximate term is denoted as:

$$\tilde{m}_i \equiv \frac{\sum_{j=1}^N w_{ij} m_j}{\sum_{j=1}^N w_{ij}} \quad (4.27)$$

Using Eqns. (4.23), (4.24), (4.26), and (4.27), the diagonal elements of the centered weighted Gram matrix can be written as:

$$\hat{\mathbf{K}}(i, i) = 1 - 2m_i + \tilde{m}_i \quad (4.28)$$

The centering equation for all the pixels in the k^{th} cluster can be written together in vector form. Defining the vector of weighted row means and the fourth term respectively as:

$$\begin{aligned} \mathbf{m}_k &= [m_1 \ m_2 \ \dots \ m_{N_k}]^T \\ \tilde{\mathbf{m}}_k &= [\tilde{m}_1 \ \tilde{m}_2 \ \dots \ \tilde{m}_{N_k}]^T \end{aligned} \quad (4.29)$$

Also, denoting the diagonal element of the weighted kernel Gram matrix of the k^{th} cluster as \mathbf{k}_k ($N_k \times 1$ vector), Eqn. (4.27) can be written as:

$$\mathbf{k}_k = 1 - 2\mathbf{m}_k + \tilde{\mathbf{m}}_k \quad (4.30)$$

Eqn. (4.30) is essentially the detector statistics for the k^{th} cluster in terms of the row sum of the block matrices.

4.5. COMPUTATIONAL ACCURACY AND APPROXIMATIONS

The outer product decomposition of a block \mathbf{K}_{kl} is given in terms of the truncated multivariate Taylor series of the various cross-terms between the clusters. This representation is approximate in nature although it can be made arbitrarily accurate by choosing a high enough truncation number. In this section, the details of choosing the truncation numbers and other parameters of the algorithm, based on the acceptable error in approximation, are presented. Also, based on the acceptable error some computational approximations are suggested, which can potentially improve the computational speed of the algorithm.

Although, the Taylor series approximation is done for function involving the augmented spectral-spatial vector, the acceptable error is specified for the spectral and spatial dimensions separately. Suppose the acceptable error per dimension for the spatial dimensions be denoted by ε_p and that in the spectral dimensions as ε_s .

4.5.1. Determination of α . In this section, the method for the selection of the set of vectors α employed in the multivariate Taylor series approximation in Eqn. (4.10), is outlined. Each element of the set α is a d -dimensional vector, where d is the dimensionality of the augmented spectral-spatial vectors. The truncation numbers for each dimension are chosen individually, and the truncated powers for the dimensions are combined later, so as to get the smallest set of α which give the desired accuracy. For a given block \mathbf{K}_{kl} , and each dimension $n = 1, 2, \dots, d$, the following is determined:

$$z_n = \max_i \left(\frac{x_{in} - c_{kn}}{\sigma} \right) \max_j \left(\frac{x_{jn} - c_{ln}}{\sigma} \right), \quad \mathbf{x}_i \in s_k \text{ \& } \mathbf{x}_j \in s_l \quad (4.31)$$

where, x_{in} denotes the n^{th} element (dimension) of the pixel vector \mathbf{x}_i . Then, the truncation number for the n^{th} dimension, q_n , is the smallest positive integer which satisfies the following:

$$\frac{(z_n)^{q_n+1}}{(q_n+1)!} < \varepsilon \quad (4.32)$$

where, $\varepsilon = \varepsilon_p$ for the spatial dimensions and $\varepsilon = \varepsilon_s$ for the spectral dimensions. Let Q_n denote the set: $Q_n = \{0, 1, 2, \dots, q_n\}$, then the set of α to compute the multivariate Taylor series in Eqn. (4.10) is given by the cross-products of the sets Q_n for all the dimensions, i.e.,

$$\alpha = Q_1 \times Q_2 \times \dots \times Q_d \quad (4.33)$$

4.5.2. Determination of β . The procedure for determination of the set β_k and β_l for the multivariate Taylor series in Eqn. (4.13a) and (4.13b) is similar to the determination of the set α . Again, the truncation numbers for each dimension are chosen individually, and the truncated powers for the dimensions are combined later. For a given block \mathbf{K}_{kl} , and each dimension $n = 1, 2, \dots, d$, the following is determined:

$$z_{kn} = \max_i \left(\frac{x_{in} - c_{kn}}{\sigma} \right) \left(\frac{c_{kn} - c_{ln}}{\sigma} \right), \quad \mathbf{x}_i \in s_k \quad (4.34)$$

Similarly, z_{in} is determined using cluster l . Then, the truncation number for the cluster k , and the n^{th} dimension, q_{kn} , is the smallest positive integer which satisfies the following:

$$\frac{(z_{kn})^{q_{kn}+1}}{(q_{kn}+1)!} < \varepsilon \quad (4.35)$$

where, $\varepsilon = \varepsilon_p$ for the spatial dimensions and $\varepsilon = \varepsilon_s$ for the spectral dimensions.

Similarly, let \mathcal{Q}_{kn} denote the set: $\mathcal{Q}_{kn} = \{0, 1, 2, \dots, q_{kn}\}$, then the sets of β_k and β_l to compute the multivariate Taylor series in Eqn. (4.13) are given by

$$\begin{aligned} \beta_k &= \mathcal{Q}_{k1} \times \mathcal{Q}_{k2} \times \dots \times \mathcal{Q}_{kd} \\ \beta_l &= \mathcal{Q}_{l1} \times \mathcal{Q}_{l2} \times \dots \times \mathcal{Q}_{ld} \end{aligned} \quad (4.36)$$

4.5.3. Spatial Far-field Approximation. Note that for the pair of clusters that are spatially far apart, the terms $\exp\left(-\frac{\|(\mathbf{c}_k - \mathbf{c}_l)\|^2}{\sigma^2}\right)$ in Eqn. (4.17) will be small, and approach zero as the inter-cluster distance approaches infinity. In terms of the Gram matrix \mathbf{K} , this means that the blocks \mathbf{K}_{kl} which correspond to a pair of clusters that are far apart (large spatial distance), will be populated by elements very close to zero. Therefore, since the contribution of these blocks is small, they can be approximated by block zero matrices, and the block need not be computed. This approximation is termed the spatial far-field approximation. This approximation can be enforced by putting a threshold on the distance between the centroids of the two clusters, i.e.,

$$\mathbf{K}_{kl} = \begin{cases} \mathbf{K}_{kl} & \text{if } d_w(\mathbf{c}_k^p, \mathbf{c}_l^p) \leq T_p \\ \mathbf{0}_{N_k N_l} & \text{if } d_w(\mathbf{c}_k^p, \mathbf{c}_l^p) > T_p \end{cases} \quad (4.37)$$

where $d(\mathbf{c}_k^p, \mathbf{c}_l^p)$ is a distance function. Therefore, the Gram matrix contains blocks of zeros corresponding to cluster pairs that are farther apart than the threshold T_p . Eqn. (4.38) shows the form of a typical sparse kernel Gram matrix.

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{0} & \text{-----} & \mathbf{0} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} & \text{-----} & \mathbf{0} \\ - & - & - & \text{-----} & - \\ - & - & - & \text{-----} & - \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{G3} & \text{-----} & \mathbf{K}_{GG} \end{bmatrix} \quad (4.38)$$

The threshold T_p is determined based on the acceptable error in the spatial dimensions as:

$$T_p = \sqrt{-\ln\left(\frac{\varepsilon_p}{2}\right)} \quad (4.39)$$

Now that the kernel matrix is modified using the spatial far-field approximation, the weighted row mean expression from Eqn. (4.25) needs to be modified. Note that the row means for the k^{th} cluster \mathbf{m}_k in Eqn. (4.25) are calculated for the entire row. However, due to the far field approximation, some of the blocks are negligibly small and not computed. To account for that, Eqn. (4.25) is modified to include the correct weight normalization. Let us denote the set of clusters which are in the “near-field” of the k^{th} cluster as:

$$\Delta_k = \{i \mid d(\mathbf{c}_k^p, \mathbf{c}_i^p) < T_p\} \quad (4.40)$$

Then, Eqn. (4.25) can be written as:

$$m_i = \frac{\sum_{n \in \Delta_k} \sum_{j=1}^{N_l} [\mathbf{W}_{kn}]_{ij} [\mathbf{K}_{kn}]_{ij}}{\sum_{n \in \Delta_k} \sum_{j=1}^{N_l} [\mathbf{W}_{kn}]_{ij}}, \mathbf{x}_i \in S_k \quad (4.41)$$

4.5.4. Spectral Far-field Approximation. Similar to the spatial far-field approximation, the spectral far-field approximation is also introduced. The idea is that from the clusters that lie in a given cluster’s spatial “near-field”, the block sub-matrices corresponding to the clusters which are far-off spectrally can be approximated as block zero matrices. The threshold for the spectral distance is given by:

$$T_s = \sqrt{-\ln\left(\frac{\varepsilon_s}{2}\right)} \quad (4.42)$$

Similar to Eqn. (4.40), the set of clusters which are in spatial-spectral near field of the k^{th} cluster is defined as:

$$\Gamma_k = \{i \mid d(\mathbf{c}_k^s, \mathbf{c}_i^s) < T_s\} \quad (4.43)$$

However, since these clusters lie in close spatial vicinity, the weights (which depend on spatial distance) will be non-negligible. Therefore, although the set of clusters used to estimate the numerators in the second and third terms of Eqn. (4.41) can be reduced further to those which have low spectral-spatial distance, the denominators have to be based on the set of all the clusters which are in spatial near-field. The detector statistics expression from Eqn. (4.41) can be written as:

$$m_i = \frac{\sum_{n \in \Gamma_k} \sum_{j=1}^{N_l} [\mathbf{W}_{kn}]_{ij} [\mathbf{K}_{kn}]_{ij}}{\sum_{n \in \Delta_k} \sum_{j=1}^{N_l} [\mathbf{W}_{kn}]_{ij}}, i \in S_k \quad (4.44)$$

4.5.5. Spectral Zeroth Order Approximation. Consider the case when the spectral expanse of say the k^{th} cluster is small, i.e., the pixels belonging to the cluster are spectrally similar. In such a case, the exponentials can be approximated with only a single term of the multivariate Taylor series, which is the “zeroth” order approximation. In zeroth order approximation the set of both multi-index variables \mathbf{a} and \mathbf{b} contain one vector each, which is the zero vector. Let $\mathbf{0}_J$ denote a $J \times 1$ vector full of zeros. A threshold is applied on the spectral expanse of the cluster to decide whether to use this approximation. That is, if

$$\min_i \left(\left| \exp \left(- \frac{\|(\mathbf{x}_i - \mathbf{c}_k)\|^2}{\sigma^2} \right) \right| \right) > 1 - \varepsilon_s, \mathbf{x}_i \in S_k$$

$$\Rightarrow \mathbf{a} = \mathbf{b}_k = \{\mathbf{0}_J\} \quad (4.45)$$

It is also noted that based on the multi-index notation:

$$\lim_{\mathbf{a} \rightarrow 0} \mathbf{a}^{0_J} = \mathbf{1}_J \quad (4.46)$$

where $\mathbf{1}_J$ denote a $J \times 1$ vector of ones.

From Eqn. (4.17), it can be seen that for the k^{th} cluster, whose spectral expanse satisfies the threshold in Eqn. (4.45), the second term on the right-hand side $\left(\frac{\mathbf{x}_i - \mathbf{c}_k}{\sigma}\right)^{\alpha + \beta_k}$ is just a vector of ones (similar to Eqn. (4.46)). Therefore, in the outer-product decomposition of the spectral block sub-matrix \mathbf{K}_{kl} (from Eqn. (4.17)), the left vectors $\boldsymbol{\kappa}_k^{\alpha\beta_k}$, which involve the k^{th} cluster, can be written directly using the spectral zeroth-order approximation as:

$$\boldsymbol{\kappa}_k^{\alpha\beta_k} = [\kappa_{1k}^{\alpha\beta_k} \ \kappa_{1k}^{\alpha\beta_k} \ \dots \ \kappa_{N_k k}^{\alpha\beta_k}]^T, \quad \kappa_{ik}^{\alpha\beta_k} = \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_k)\|^2}{\sigma^2}\right) \quad (4.47)$$

Since, only one term is used in the outer-product decomposition, it fastens the computation of the spectral block \mathbf{K}_{kl} . However, the block sub-matrix of weights \mathbf{W}_{kl} has to be computed using the full outer-product decomposition.

4.5.6. Spectral Near-Field Approximation. This approximation is similar to the spectral zeroth order approximation. Consider the case when the spectral centroids of the two interacting clusters representing the spectral block $[\mathbf{K}_{kl}]$, i.e. clusters k and l , lie close by. In such a case, the second term on the left-hand side in the expression for κ_{ikl} in Eqn. (4.12), can be approximated by 1. This leads similar outer-product decomposition as in the case of a diagonal blocks, given in Eqn. (4.18). A threshold is applied on the spectral inter-cluster distance to decide whether to use this approximation. That is, if

$$\|(\mathbf{c}_k - \mathbf{c}_l)\| < \varepsilon_s \quad (4.48)$$

then

$$\mathbf{K}_{kl} = \sum_{|\alpha| \leq p-1} \frac{2^\alpha}{\alpha!} \boldsymbol{\kappa}_k^\alpha \boldsymbol{\kappa}_l^{\alpha^T} \quad (4.49)$$

where,

$$\boldsymbol{\kappa}_{ip}^\alpha = \exp\left(-\frac{\|(\mathbf{x}_i - \mathbf{c}_p)\|^2}{\sigma^2}\right) \left(\frac{\mathbf{x}_i - \mathbf{c}_p}{\sigma}\right)^\alpha \quad \& \quad \boldsymbol{\kappa}_p^\alpha = [\kappa_{1p}^\alpha \ \kappa_{2p}^\alpha \ \dots \ \kappa_{N_{pp}}^\alpha]^T, \quad (4.50)$$

$$i \in s_p, \ p \in \{k, l\}$$

However, similar to the spectral zeroth order approximation, the block sub-matrix of weights \mathbf{W}_{kl} has to be computed using the full outer-product decomposition.

4.6. MULTIPLE PIXEL TARGET

In case of the RX anomaly detector, the case of targets with signatures that extend for more than one pixel can be handled conveniently. Since the RX detector is linear in nature, averaging the RX statistics over a target mask (which defines the expected target shape and size) is the same as finding the mean target pixel over the target window, and using it as the test vector. However, since KRX and SW-KRX are nonlinear in nature, the case of multiple pixel targets becomes slightly more complicated.

The case of multiple pixels is handled by computing the mean of the detector statistic over the target area. Given the form of the detector statistic for the special case of the SW-KRX detector in Eqn. (4.21), it can be seen that the detector statistic is essentially a weighted sum of the weighted row means of the kernel Gram matrix. Since the statistic is linear in the row means, the detector statistic for a multiple pixel target can be expressed as the linear combination of the detector statistics for individual pixels under the target region. Let the matrix \mathbf{S} be defined as the matrix with same dimensions as the image, containing the detector statistic for each pixel at its corresponding location, i.e,

$$[\tilde{\mathbf{S}}]_{ij} = SW\ KRX(\mathbf{y}(i, j))$$

Also, let \mathbf{W}_T denote the target mask, which has values 1 over the target region and 0 elsewhere. Then, the multiple pixel target detector statistics can be given as:

$$\tilde{\mathbf{S}} = \mathbf{S} \otimes \mathbf{W}_T \quad (4.51)$$

where, ' \otimes ' denotes the 2-D convolution operation.

4.7. COMPUTATIONAL COMPLEXITY

In this section, the analysis of the computational complexities of the Kernel RX and the SW-KRX algorithm is presented. Complexities are presented for the special case (unit variance and uncorrelated feature dimensions) for both the algorithms. Since Kernel RX operates pixel-wise and in a deterministic fashion, its exact complexity analysis is possible. However, the SW-KRX algorithm is data dependent in nature in that the

computation depends on several variables like cluster sizes and spectral and spatial distribution of data. Therefore, for SW-KRX complexity analysis is presented based on average values of the stochastic variables. Although, this analysis describes the algorithm complexity analysis for the “average” case only, it is effective in highlighting the salient computational differences in the two algorithms. It is noted that the operations counts are based primarily on the multiplication operation, although it is noted that operations like the exponentials take several more flops than the multiplication operation, on almost all processors.

Consider the computation for a single pixel using the Kernel RX algorithm. Let us assume that a total of N_C pixels are used for the estimation of the background statistics. Let the dimensionality of the data be J . Then, the number of operations needed for the computation of the kernel Gram matrix at each pixel is $N_C^2 J$. It is noted that the computational cost of the calculation of the exponential is taken to be the same as the multiplication operation, although, as mentioned earlier, exponentials take several more flops than the multiplication operation. The second step is that of the Gram matrix centering. The centering operation takes $2N_C^2$ additions. This is followed by centered Gram matrix inversion, which typically is a $O(N_C^3)$ operation. It is noted that there exist techniques for fast iterative inversion methods which are $O(N_C^2 m)$ complexity, where m is the number of iterations. However, for our purposes and in general complexity comparison literature, matrix inversion operation is considered $O(N_C^3)$. Similarly, the computation of the target vector \mathbf{k}_r takes approximately $N_C J$ operations. Again, it is noted that this step involves exponential computation, which is treated as a simple multiplication. And finally, the matrix-vector product to get the final detector statistic requires $N_C^2 + N_C$ computations. In total, the number of computations needed for the detector statistics for a single pixel is approximately:

$$C_{KRX} = N_C^3 + (J + 3)N_C^2 + (J + 1)N_C \quad (4.52)$$

As is evident, the main computational bottleneck is the matrix inversion operation, which for large masks sizes (higher values of N_C) can make the computation prohibitively expensive. Besides the matrix inversion, there are other computational steps

which are expensive too. At best, with the most efficient implementation, the complexity of the Kernel RX algorithm still remains at $O(N_C^2)$.

Now the complexity for the proposed SW-KRX algorithm is presented. Again, the goal is obtain the computational cost of computing the detector statistics for a single pixel. Let us consider a J band multi-dimensional data, with an average cluster size n . Let the spectral distribution of the members in the cluster be such that the average number of terms required for the α -multivariate Taylor series be A_s and the average number of terms required for β -multivariate Taylor series be B_s . Similarly, let the spatial distribution of the cluster be such that the α and β -multivariate Taylor series approximations require A_p and B_p number of terms on an average. Since the detector statistics is computed for a cluster as a whole, first the cost for the entire cluster is obtained and then divided by the cluster size. The computational complexity of the outer product decomposition for an average sized block will take the following number of computations:

$$\text{Single Outer Product Decomposition: } 2n(A_s B_s^2 A_p B_p^2 + A_p B_p^2)$$

Let us assume that there are, on an average, M clusters which lie within the vicinity of the given block and are not neglected based on spatial far-field approximation. This translates into the computation of the outer product decomposition for a total of M blocks. However, note that since the kernel Gram matrix is symmetric, the following relationship between the blocks holds: $\mathbf{K}_{kl} = \mathbf{K}_{lk}^T$. In terms of the computations, the outer product decomposition for the \mathbf{K}_{kl} block can be used for the \mathbf{K}_{lk} block, just by interchanging the left and right vectors. On an average this reduces the number of blocks for which the outer product decomposition is needed by a factor of 2. Therefore, on an average, the total number computations needed for outer product decompositions for computing the statistics for a given cluster are:

$$\text{Total Outer Product Decomposition: } Mn(A_s B_s^2 A_p B_p^2 + A_p B_p^2)$$

Once the outer-product decomposition is obtained, the next step is to compute the row means for each block. Given a single outer product decomposition, the row mean can be calculated efficiently in $2n$ computations instead of full n^2 operations. However, it

should be noted that although the outer product decomposition for the \mathbf{K}_{kl} block can be used for the \mathbf{K}_{lk} block too, the row means need to be calculated for both the blocks separately. Therefore, the total computations needed for row mean calculation is

$$\text{Row Mean Computation: } 2Mn(A_S B_S^2 A_P B_P^2 + A_P B_P^2)$$

Thus the total computations required for the computation of the detector statistics for the entire cluster is $3Mn(A_S B_S^2 A_P B_P^2 + A_P B_P^2)$. Therefore the average computations required for a single pixel is:

$$C_{SW-KRX} = 3M(A_S B_S^2 A_P B_P^2 + A_P B_P^2) \quad (4.53)$$

Comparing complexities of the two algorithms, it can be seen that Kernel RX complexity depends mainly on N_C . In case of SW-KRX, it depends mainly on the number of clusters in the vicinity M . It is noted that M typically is much smaller than N_C . In case of Kernel RX N_C grows as the square of the clutter mask radius, and is a very high number for large radii. However, in the case of SW-KRX, the value of M does not always go higher with the increase in the width of the spatial weighting function γ . At relatively small values of γ , M , which is the number of clusters in the vicinity, typically goes down. This is because with the increase in the value of γ , the allowed cluster sizes (based on the bounds in Eqns. (4.10) and (4.11)) also goes higher, leading to larger clusters. Thus, for small increases in the value of γ the region of influence remains almost the same, but the cluster sizes grow, leading to a falloff in the value of M . This translates into fewer computations with increasing value of γ . This phenomenon is observed in the empirical results on execution times of the SW-KRX algorithm for both multispectral and single band data in Tables 5.1 and 5.2 in Section 5. However, this is not true at large value of γ . This is because at higher values, the cluster sizes cease to grow as the cluster start to cut across different regions of the image. Also, at higher values the allowed cluster sizes get restricted due to the bounds on the spectral expanse of the clusters. Thus, as γ grows, the region of influence grows but the cluster sizes remain fixed, which leads to higher values of M , thus increasing computational loads.

Another factor that governs the load of the SW-KRX algorithm is the number of terms needed for multivariate Taylor series approximations. Fewer terms (low values of A_p, B_p, A_s and B_s) lead to better computational efficiency of the proposed algorithm. It was observed that these values were typically less than 10 for error tolerances of $\varepsilon_s = 0.02$ and $\varepsilon_p = 0.1$. This analysis highlights the need for the application on the bounds on the cluster sizes as discussed in Eqns. (4.10) and (4.11). Larger cluster expands in either spectral or spatial dimensions require a larger number of terms in the approximation, which lowers the computational efficiency. Limiting the cluster expands helps overcome this problem to a certain extent. On the other hand limiting the cluster to very small sizes also hurts the performance due to higher values of M . Also, with the increase in the allowed errors, the number of terms needed for approximation goes down, thereby reducing computational load. Finally, both algorithms have the same linear complexity in terms of the dimensionality of the data.

For comparison purposes, the computational requirements of the two algorithms are determined for a typical data scenario. For a typical frame from Dataset 1 for $\gamma = 7$, $\varepsilon_s = 0.02$ and $\varepsilon_p = 0.1$, the average values of the different variables are found to be:

$$M = 8.55, n = 28.82, A_s B_s^2 = 2.717, A_p B_p^2 = 3.665$$

Based on the average values, a total of approximately 350 operations are needed for the fast implementation of the SW-KRX detector (special case). Now consider the full implementation of the original Kernel RX detector. For the same size of the neighborhood used for background estimation used for SW-KRX detector, the following parameters are used:

$$\text{demeaning radius} = 10, \text{blanking radius} = 0 \Rightarrow N_C = 317$$

For these mask sizes, a total of approximately 3×10^6 operations are needed for the Kernel RX detector. From the number of operations needed, it can be seen that for an average scenario the fast implementation of the SW-KRX detector is several orders of magnitude faster than the Kernel RX detector.

5. SW-KRX: RESULTS AND DISCUSSION

This section presents the various results pertaining to the SW-KRX algorithm. The fast version of the SW-KRX detector has been implemented. However, it is only for the special case of the SW-KRX detector given in Eqn. (3.50), which is based on the assumption that the various dimensions in the feature space are unit variance and uncorrelated. Results are presented on broadly two lines. Section 5.2 presents the results on the detection performance of the SW-KRX detector vis-à-vis the popular multiband RX anomaly detector. These results are compiled for surface mines, on two different datasets. Sections 5.3 and 5.4 present results on the computational speed of the two detectors. Section 5.3 depicts the comparison of the execution times of the proposed detector and the original Kernel RX algorithm [Kwon, 2005a]. Section 5.4 presents the comparative results on the computational speed-up achieved by the proposed multivariate Taylor series based approximation of the Gaussian kernel. A brief summary of the results is presented in Section 5.5.

5.1. DATA DESCRIPTION

Airborne data has been collected as part of the airborne landmine detection program at Night Vision and Electronic Sensors Directorate (NVESD). In this work, results are reported on two datasets, collected at different times, and for different terrain conditions. From this point on, the two datasets are referred to as Dataset 1 and Dataset 2. Dataset 1 data was collected over a temperate site, whereas Dataset 2 was collected at an arid site.

Data collected in one sweep over a minefield is collectively called a “segment”, and it consists of 21 images (called frames). The multispectral data consists of four bands, of which three are in the visual range: red (R), green (G), and blue (B). The fourth band in the multispectral data is in the near-infrared (N) range. The data contains both buried and surface mine signatures. However, the current analysis presented in this section is limited to the surface mines. Datasets 1 and 2 evaluated here contain a total of 39 and 54 segments, respectively, containing surface mines. As a pre-processing step, the

dynamic range of various bands of the multispectral image is restricted to the following intervals:

Dataset 1: $50 < R < 1000$, $200 < G < 2600$, $100 < B < 600$, $400 < N < 2600$

Dataset 2: $400 < R < 4000$, $200 < G < 1800$, $200 < B < 1600$, $500 < N < 3200$

This is done to eliminate any bias in the further processing due to extremely high sensor values from noise or some ground features such as fiducial markers. It was found that some frames in the datasets had a “washed out” appearance due to saturation of the sensors. Such frames were eliminated from consideration and results have been compiled on only the good frames. The exact number of frames used for a particular result is provided in the discussion following each result. Figures 5.1 and 5.2 show typical surface mine signatures. Figure 5.1 shows surface mine signatures that are clearly visible and well-distinguished from the background. Figure 5.2 shows relatively obscure mine signatures, which are hard to distinguish from the background.

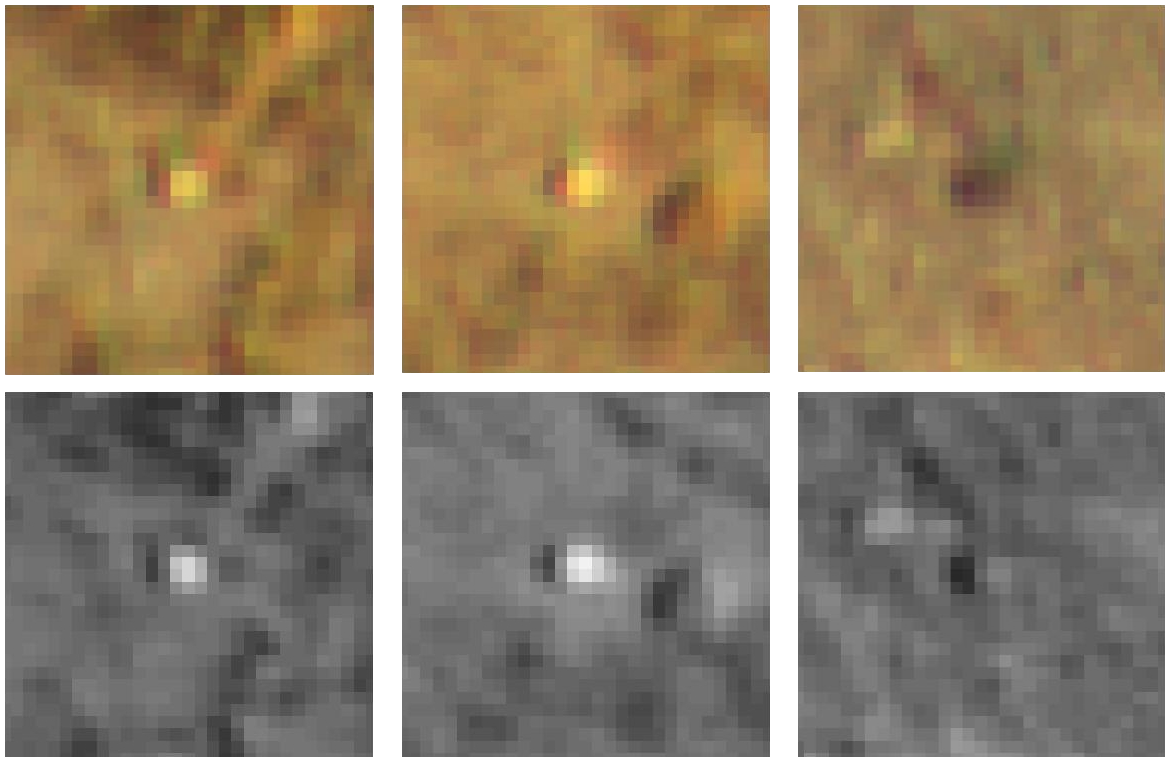


Figure 5.1. Clearly Visible Surface Mine Signatures (Top Row) RGB Composite (Bottom Row) NIR Band.

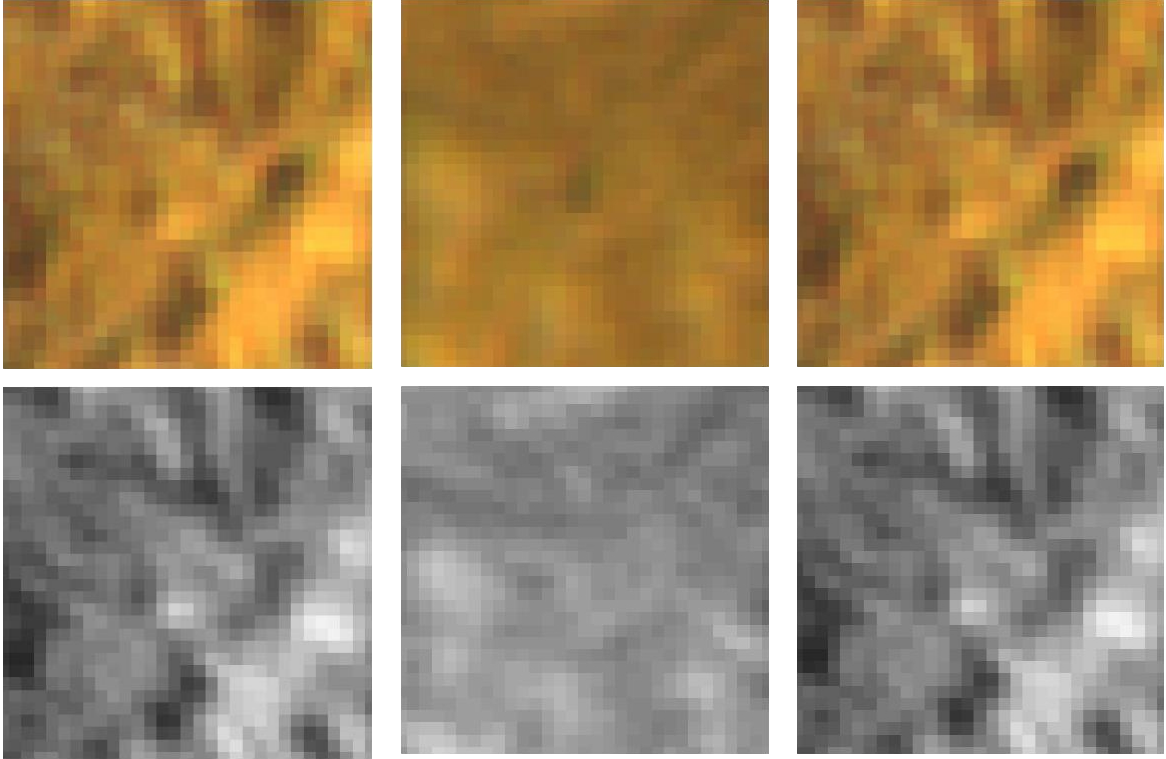


Figure 5.2. Poorly Visible Surface Mine Signatures (Top Row) RGB Composite (Bottom Row) NIR Band.

5.2. DETECTION PERFORMANCE: SW-KRX (SPECIAL CASE) VS. RX

This section presents the detection performance results for the fast implementation of the special case of SW-KRX detector (Eqn. (3.50)). The detection performance results for the popular RX anomaly detector are also presented for the purposes of comparison. Results on both the datasets are illustrated for two different cases. The first case is that of single band data, where only the green-band (G) is used for detection, both for the proposed SW-KRX and RX anomaly detector. The second case is that of multispectral data, where all four bands are used for detection. Comparative results using the RX anomaly detector are shown. For the RX detector the following mask sizes were used: *demeaning radius* = 10, *blanking radius* = 5 and *target radius* = 1, for all the results. The SW-KRX detector entails the selection of two parameters, namely spectral kernel width σ , and the spatial weighting function width γ . The spatial weight function width was chosen to be $\gamma = 7$, as that gives a region of influence of

approximately the same as that used for demeaning in the RX detector. The value of the spectral kernel width σ , is chosen adaptively for each segment, based on the variance of the frames of the segment. The width σ was determined as $\sigma = 5s$, where 's' is the mean standard deviation of the demeaned frames, averaged over the segment. Thus, the width σ is the same for all the frames in a segment, but for different segments. Typical ranges of values for σ are 350-600 and 150-300 for multispectral and single band data, respectively.

Also, it was found that the SW-KRX detector gives high detector statistics along the major edges in the images. This is expected because typically the edge pixels are spectrally "different" from the object regions on either side. Since SW-KRX is capable of detecting anomalies from non-homogeneous background, the SW-KRX detector gives a high response at the edge pixels and generates false alarms. Figure 5.3a shows a typical frame and Figure 5.3b shows the SW-KRX detector output. The RX and SW-KRX detections (at constant FAR of 10^{-2} FA/m²) are depicted in blue circles and red diamond markers, respectively. As can be seen, the SW-KRX detector response is high at the edges, and consequently there are several false alarms at the edges. To correct this problem, the SW-KRX detector statistics are modulated using an edge map of the image. This reduces the false alarms along the edges as can be seen in Figure 5.4.

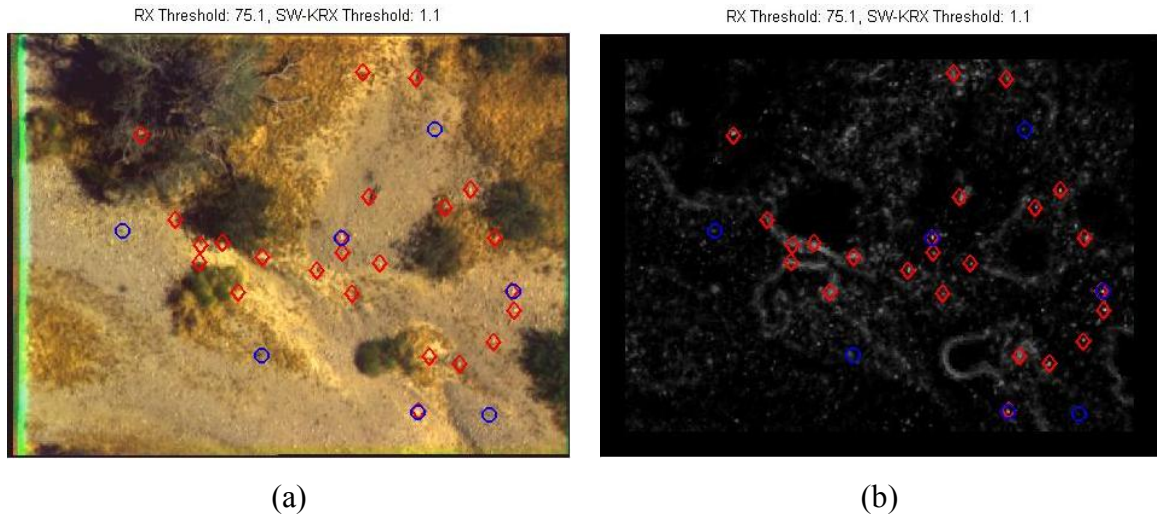


Figure 5.3. Typical Frame with False Alarms Along Edges (a) Original Frame, (b) SW-KRX Detector Output.

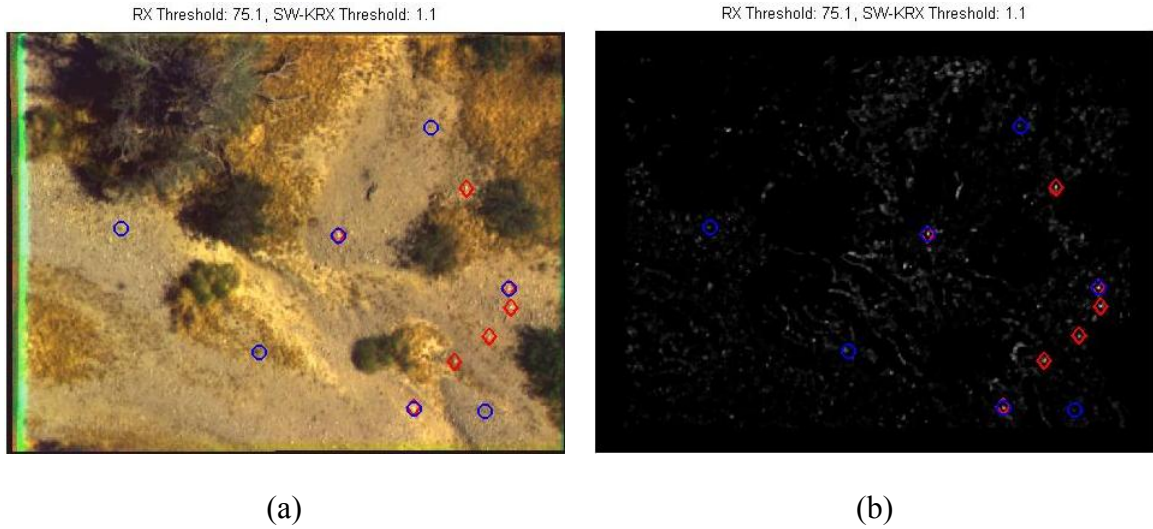


Figure 5.4. Typical Frame with Edge-Suppressed False Alarms (a) Original Frame, (b) Edge-Suppressed SW-KRX Detector Output.

5.2.1. Results: Green Band. As mentioned earlier, the results for the SW-KRX detector are presented only for the special case, assuming that the various dimensions in the feature space are uncorrelated and unit variance. Given this assumption, a mixed performance for the SW-KRX detector was achieved vis-à-vis the RX anomaly detector. It was found that on some segments the proposed detector gave superior results as compared to the RX detector and on others its performance was at par or slightly inferior to the RX algorithm. Figure 5.5a shows the Receiver Operating Characteristic (ROC) curves for both the detectors for the set of segments where SW-KRX performance was better as compared to the RX detector. The results are compiled for a set of 179 frames from 16 segments. As can be seen, the performance is distinctly superior on the selected segments, especially at the lower false alarm rates (FARs). Figure 5.5b depicts the ROC curves over the set of rest of the data, namely 226 frames from 23 segments. Again, it can be seen that the performance of the two detectors is largely at par with each other.

Figure 5.6 shows the detection performance for the two detectors for Dataset 2. Dataset 2 in general has lower contrast signatures of targets as compared to Dataset 1. Again, the special case SW-KRX detector shows superior performance for a certain set of

segments as compared to the RX algorithm, as can be seen in Figure 5.6a. The results in Figure 5.6a were compiled on 345 frames from 18 segments. Showing a similar trend as Dataset 1, the detection performance over the rest of the data (738 frames from 36 segments) was almost at par with the RX detector, with the detection rates at lower FARs marginally lower.

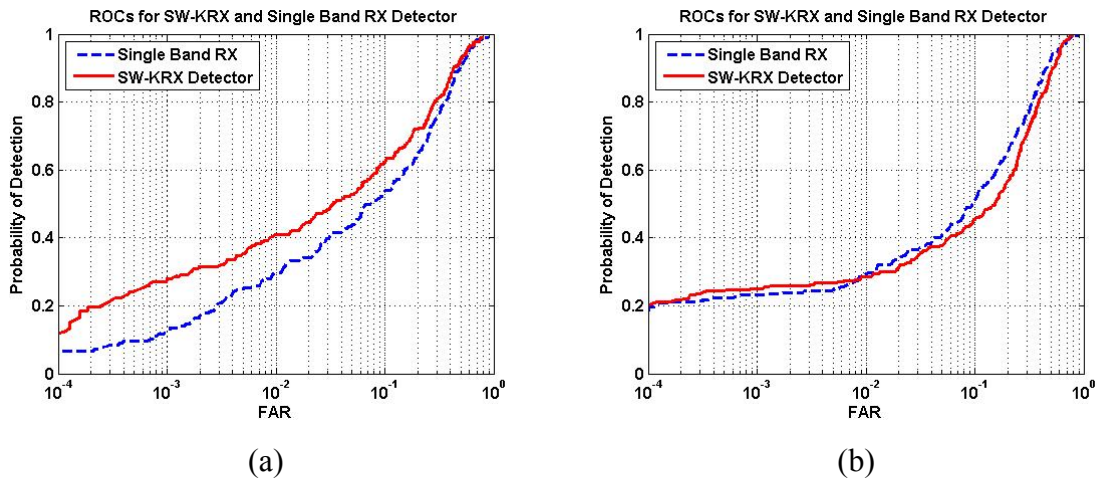


Figure 5.5. Detection Performance of SW-KRX vis-à-vis RX Detector for Green Band Data for Dataset 1 (a) Selected Segments (b) Rest of the Segments.

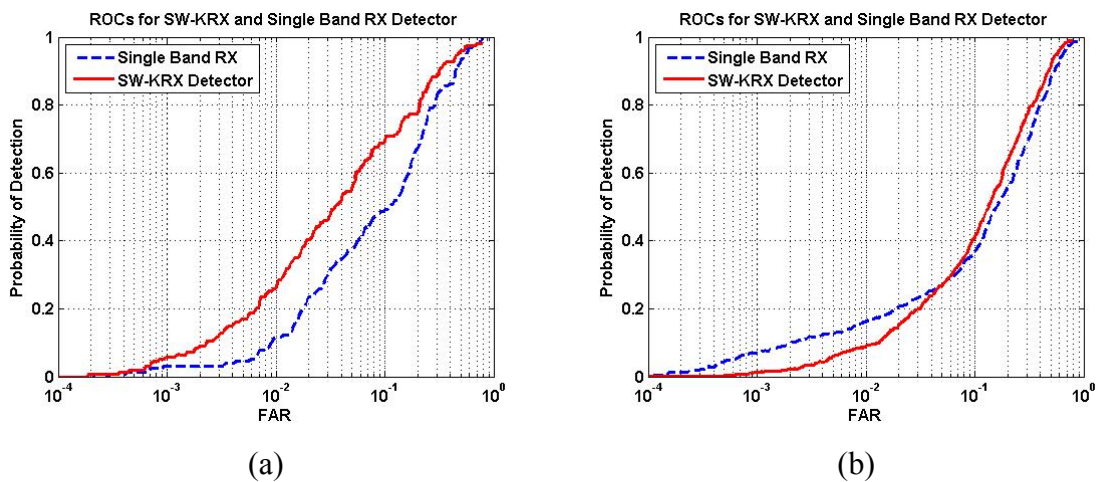


Figure 5.6. Detection Performance of SW-KRX vis-à-vis RX detector for Green Band Data for Dataset 2 (a) Selected Segments (b) Rest of the Segments.

The results in Figures 5.5 and 5.6 indicate that the non-linear SW-KRX detector, even with the restrictive assumption of uncorrelated feature dimensions, is largely at par with the RX detector, but gives improved detection under certain scenarios.

5.2.2. Results: Multispectral. The results on multispectral data for the two datasets are presented in this section. Figure 5.7 shows the ROC curves for Dataset 1. The performance of the SW-KRX detector is only marginally better in some segments, as is evident from Figure 5.7a, which is based on 46 frames from 4 segments. The performance on the rest of the data (359 frames from 35 segments) for the SW-KRX detector is worse than the multiband RX detector.

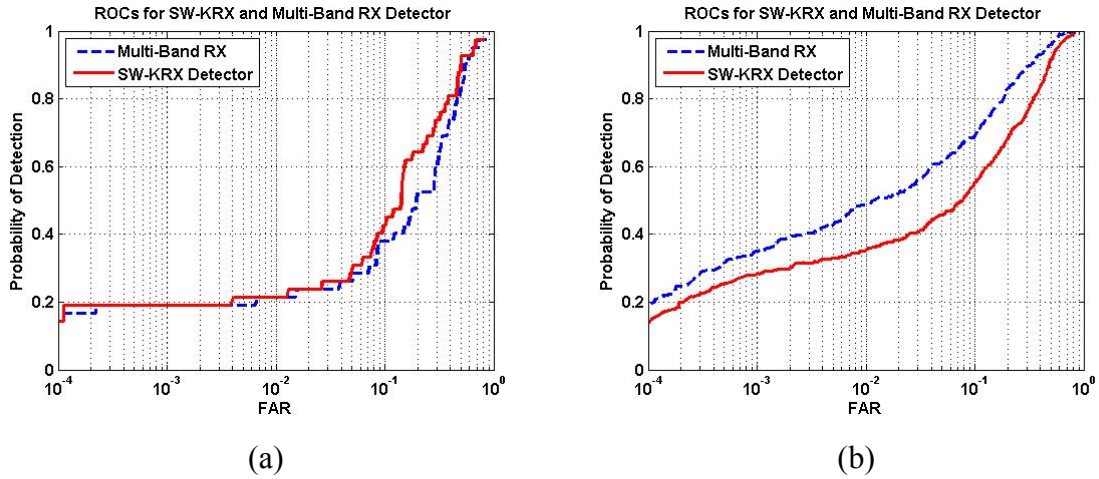


Figure 5.7. Detection Performance of SW-KRX vis-à-vis RX Detector for Multispectral (4-Band) Data for Dataset 1.

Visual inspection of the detection results reveals that the primary reason for the drop in the performance of the proposed detector is the inter-band misalignment in the multispectral data. As a result of this misalignment the mine signatures in case of multispectral data are more diffused than that in any single band. This misalignment may also result in poor initial clustering of the image data, which reduces performance. Based on empirical observations, this misalignment affects the proposed detector (in its current implementation) more than the RX detector.

In the case of Dataset 2, note that in Figure 5.8a the relative performance in some segments is slightly improved than that on Dataset 1 (Figure 5.7a). The results in Figure 5.8a are based on 93 frames from 5 segments. Figure 5.4b shows the results on the rest of the data consisting of 990 frames from 49 segments. As is evident, although SW-KRX performs marginally better at low FARs, its performance is largely worse at higher FARs. Comparing Figure 5.7a and 5.8a, it can be seen that the performance of SW-KRX relative to the RX detector is better on Dataset 2 as compared to Dataset 1. As mentioned earlier, Dataset 2 contains greater number of lower contrast target signatures as compared to Dataset 1. Although both RX and SW-KRX are sensitive to target to background contrast, SW-KRX shows better capability in the detection of low contrast targets in non-homogenous background, as compared to the RX detector, as shown later in Figure 5.9.

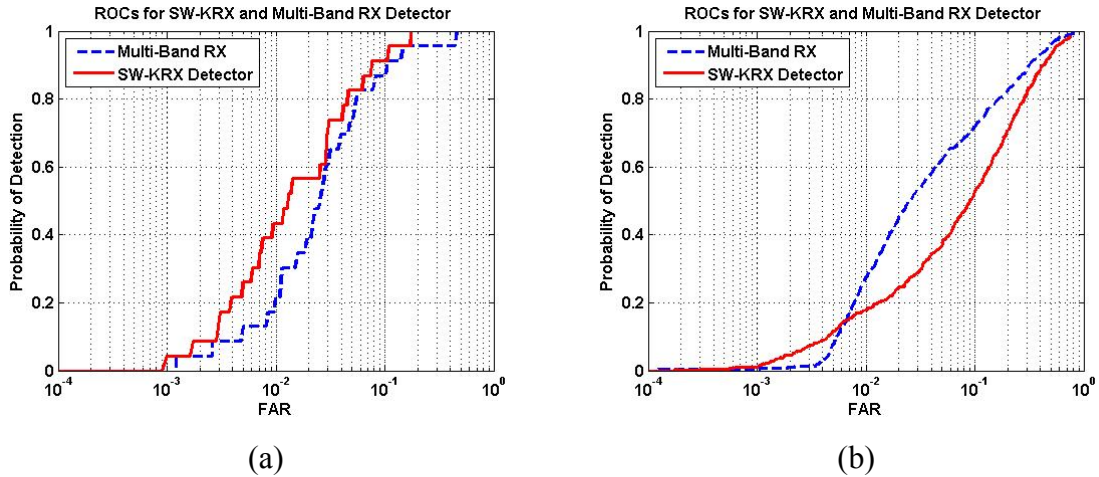


Figure 5.8. Detection Performance of SW-KRX vis-à-vis RX Detector for Multispectral (4-Band) Data for Dataset 2.

The proposed SW-KRX detector demonstrates a better capacity at detecting targets in a non-homogeneous background, as compared to the RX detector. This phenomenon can be explained more clearly by looking at the detector statistics of the special case of the SW-KRX detector, given in Eqn. (3.50), and the RX detector statistics. Since RX assumes a single Gaussian distribution for the background, it computes the

Mahalanobis distance of the target vector from the sample distribution obtained from the samples falling in the clutter mask. Therefore, in case of non-homogeneous background, where the data is not from a single Gaussian distribution, the RX statistic is essentially the distance of the target vector from the “mixture” of the two or more background distributions. On the other hand, looking at the SW-KRX detector (special case), it is observed that the detector statistic is essentially the sum of the kernel distances (exponential, in case of the Gaussian kernel) of the target vector from all the pixels in the “neighborhood”. It, unlike the RX detector, is not the distance of the target vector from the “mixture” of the background. Consider, for example, the case of a two class background, which are well separated. Now consider the case where the target vector lies somewhere in between the ranges of two background classes. The mean statistics of the “mixture” of the two distributions will lie close to the target vector, and consequently, the RX detector will give a poor response for the target pixel. However, since the target vector is well-separated from the two background classes, the SW-KRX detector will give a better response.

As an example, Figure 5.9 shows a frame which has several surface mine targets which are in non-homogeneous backgrounds. Thresholds for the detector statistics are chosen at a constant FAR of 10^{-2} FA/m². The thresholds for SW-KRX and RX detectors are 1.22 and 37.41, respectively. All targets which have detector statistics greater than the threshold are classified as detections. The green (square) markers show the actual mine locations, and the red (diamond) and blue (circle) markers show the detections given by the SW-KRX and RX detectors, respectively. As is evident, the SW-KRX detector is able to detect all the mine targets at the given FAR, whereas RX misses some of them, since the targets lie in a non-homogeneous background. The only target it detects lies in the region where majority of the pixels in the background are from a single class.

Figure 5.10 shows a typical frame with multiple false alarms generated by both the detectors. The detections are shown at a constant false alarm rate of 10^{-2} FA/m². It can be observed that the SW-KRX detector mostly responds to features which are similar to the surface mine signatures, although they are really false alarms and do not count as correct detections. It is noted again, that although RX also detects mine-like features, it does not detect such features when they lie in non-homogeneous background.

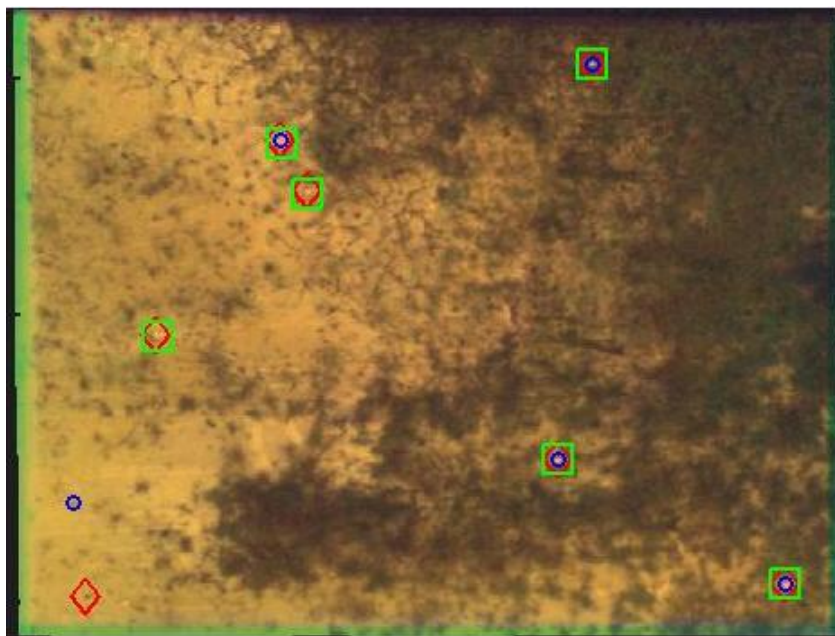


Figure 5.9. Example Frame with Targets in Non-Homogeneous Background.



Figure 5.10. Typical Frame with False Alarms.

5.3. COMPUTATIONAL SPEED: SW-KRX (SPECIAL CASE) VS. KERNEL-RX

In this section, the computational speeds of the fast implementation of the special case of the SW-KRX and the original Kernel RX algorithm are presented. The total computational time for the proposed implementation is separated into two categories: that used for initial clustering, and that for the detector statistics computation, based on the clustered data. Tables 5.1 and 5.2 show the average execution times per pixel for the proposed implementation and the original Kernel RX algorithm, for single band and multispectral data, respectively. Both the detectors were implemented in Matlab[®], and the execution times were recorded for a 3.2 GHz, Pentium[®] DuoCore[®] processor, Windows machine with 2 GB RAM.

Since the original implementation of the Kernel RX algorithm has a large computation time, the average execution time was computed over a set of 100 pixels. The execution times for the proposed implementation were recorded for a typical frame from Dataset 1, and the average time per pixel was computed over an entire frame. Since the computational speed of the proposed implementation depends on the width of the spatial weighting function γ , and the spectral and spatial error tolerance parameters ε_s and ε_p , the execution times were recorded for different value of the of these parameters. The clustering part of the algorithm does not change with the change in error tolerance parameters, and depends only the width of the spatial weighting function γ and the kernel width σ . Therefore, for a given value of γ , the clustering time remains the same for the various values of the spectral and spatial error tolerances. For faithful comparison, the radius of the clutter mask for the full Kernel RX algorithm D , and the width of the spatial weighting function γ , was chosen such that both the detectors take approximately the same background region under consideration. The parameters values are depicted in the first column of the tables.

As can be seen from Table 5.2, the proposed implementation provides computational speed-ups of several orders of magnitudes over the original kernel RX algorithm. It can be observed that in the original Kernel RX detector, as the clutter mask radius increases, the computations rise exponentially, due to the increase in dimensionality of the kernel Gram matrix. However, in case of the fast implementation the computation time actually reduces with the larger values of γ . The reason for this is

two-fold. Firstly, the clustering time reduces since larger clusters are now permitted to exist based on the upper bound on the cluster spatial dimensions and fewer clusters need to be corrected. In a related manner, since the data is clustered into larger clusters in general, the algorithm for the computation of the detector statistics has higher computational gains, as explained in Section 4.8 on computational complexity. Due to this, there is a falloff in the overall computation time per pixel, as observed. However, with the increase in the value of γ , the number of blocks that can be neglected based on the spatial far-field approximation reduce, and more computation is required. Therefore, the falloff can be expected only up to a limited value of γ .

Also, it is observed that the computation times reduce with higher value of error tolerances. This is expected because higher allowed error translates into lower truncation numbers for the multivariate Taylor series approximation, which reduces computational load. However, based on the execution times for $\varepsilon_s = 0.02$, $\varepsilon_p = 0.5$ and $\varepsilon_s = 0.1$, $\varepsilon_p = 0.1$, it is observed that the speed-up with the increase in spatial error tolerance is greater than that with spectral error tolerance. This is due to the fact that typically the spatial expanse of the clusters is larger than the spectral expanse and hence the reduction in the number of terms needed for the Taylor series approximation is greater for spatial dimensions. A similar trend is observed in the case of single band data as shown in Table 5.1, although the computation times are lesser than the multispectral case.

It is noted that the fast implementation (special case) of the SW-KRX detector gives a speed-up of up to 3-4 orders of magnitude over the original Kernel RX detector. However, in the theoretical complexity analysis discussed in Section 4.7, the speed-ups were up to 5-6 orders of magnitude, for the average scenario. The primary reason for this difference is that in the theoretical analysis of the computations required for Gram matrix inverse was taken as N_C^3 . In the Kernel RX implementation used for the results in Tables 5.1 and 5.2, the pseudo-inverse of the Gram matrix is used, since the Gram matrix is not full rank and hence non-invertible. For the pseudo-inverse computation, the eigenvector decomposition of the Gram matrix is computed which is lower complexity than the full matrix inverse. Therefore, the implementation of the Kernel RX detector is slightly faster than what the theoretical complexity was presented for.

Table 5.1. Average Computation Time per Pixel for Full Kernel RX and Fast Implementation, for Single Band Data. Times are in μ Secs.

Time (μ Sec)	Full Kernel RX	Clustering Time	$\varepsilon_s = 0.02, \varepsilon_p = 0.1$		$\varepsilon_s = 0.02, \varepsilon_p = 0.5$		$\varepsilon_s = 0.1, \varepsilon_p = 0.1$	
γ, D			SW-KRX Time	Total	SW-KRX Time	Total	SW-KRX Time	Total
$\gamma = 7,$ $D = 10$	1.16×10^5	40.56	207.97	248.53	101.16	141.72	189.48	230.04
$\gamma = 10,$ $D = 15$	9.13×10^5	26.71	132.09	158.80	60.64	873.44	156.81	183.51
$\gamma = 13,$ $D = 20$	4.32×10^6	21.79	131.96	153.74	56.03	77.812	154.63	175.81

Table 5.2. Average Computation Time per Pixel for Full Kernel RX and Fast Implementation, for Multispectral Data. Times are in μ Secs.

Time (μ Sec)	Full Kernel RX	Clustering Time	$\varepsilon_s = 0.02, \varepsilon_p = 0.1$		$\varepsilon_s = 0.02, \varepsilon_p = 0.5$		$\varepsilon_s = 0.1, \varepsilon_p = 0.1$	
γ, D			SW-KRX Time	Total	SW-KRX Time	Total	SW-KRX Time	Total
$\gamma = 7,$ $D = 10$	9.93×10^4	34.79	108.23	143.01	57.31	92.11	106.09	140.88
$\gamma = 10,$ $D = 15$	8.51×10^5	21.97	73.71	95.68	31.80	53.78	71.29	93.26
$\gamma = 13,$ $D = 20$	4.13×10^6	17.07	66.81	83.88	26.79	43.84	62.85	79.92

Also, since the clustering time is also included in the total execution time of the fast implementation of the SW-KRX detector in Tables 5.1 and 5.2, and were not included in the theoretical analysis, the computation time for SW-KRX is more than the one based on theoretical calculations.

5.4. SW-KRX: COMPUTATIONAL SPEED-UP VIS-À-VIS FULL GRAM MATRIX COMPUTATION

In the previous section, the broad comparison between the overall execution times between the two detectors was presented. This section presents the results on the computational speed-ups achieved due to the computation of the blocks of the kernel Gram matrix using multivariate Taylor series approximation, as compared to the full computation using direct exponentials.

Tables 5.3 and 5.4 depict the average computation time per block (in milliseconds) for a typical frame from Dataset 1, for single band and multispectral data, respectively. The values of the spectral and spatial kernel width are: $\gamma = 7$ and $\sigma = 140$, respectively. The first row of the tables shows the average time per block, when averaged over all the blocks of all sizes in the Gram matrix. The second row shows the average computation time per block for the blocks which have at least one dimension greater than 15, i.e. at least one of the two representing clusters contains 15 or more pixels. The third row is the same as the second except that only blocks with at least one dimension greater than 30 are considered. The idea is to demonstrate the effect on computational efficiency of the proposed method, with the increase in cluster sizes.

The execution times are recorded for all the non-zero blocks of the kernel Gram matrix. The columns in the tables denoted by “ t Approx” depict the average time (in milliseconds) per block for the computation of the weighted row means using approximate computations based on the multivariate Taylor series. The columns in the tables denoted by “ t Expo” show the average time (in milliseconds) per block for the computation of the weighted row means using full exponential based computation. The column denoted by “Factor” shows the ratio of the former to the latter, which essentially is the fraction of the full exponential computation time that is taken by the approximate computation.

Table 5.3. Computational Speed-up via Taylor Series Approximation, for Single Band Data. Times are in milliSecs.

Time(mSec)	$\varepsilon_s = 0.02, \varepsilon_p = 0.1$			$\varepsilon_s = 0.02, \varepsilon_p = 0.3$			$\varepsilon_s = 0.02, \varepsilon_p = 0.5$		
Clusters	t Approx.	t Expo.	Factor	t Approx.	t Expo.	Factor	t Approx.	t Expo.	Factor
All	0.7	0.8	0.91	0.4	0.8	0.50	0.4	0.8	0.49
> 15	0.7	0.9	0.87	0.4	0.8	0.47	0.4	0.8	0.46
> 30	0.8	1.0	0.76	0.4	1.0	0.40	0.4	1.0	0.40

Table 5.4. Computational Speed-up via Taylor Series Approximation, for Multispectral Data. Times are in milliSecs.

Time(mSec)	$\varepsilon_s = 0.02, \varepsilon_p = 0.1$			$\varepsilon_s = 0.02, \varepsilon_p = 0.3$			$\varepsilon_s = 0.02, \varepsilon_p = 0.5$		
Clusters	t Approx.	t Expo.	Factor	t Approx.	t Expo.	Factor	t Approx.	t Expo.	Factor
All	0.6	0.8	0.76	0.3	0.8	0.38	0.3	0.8	0.38
> 15	0.7	0.9	0.73	0.3	0.9	0.36	0.3	0.9	0.34
> 30	0.7	1.2	0.64	0.3	1.1	0.29	0.3	1.1	0.27

The various values are recorded for three different values of the spatial error tolerance ε_p . The goal is to observe the effect of approximation accuracy on the computational gain, and since the spatial dimensions have larger expanses, the effect is expected to be relatively clearer.

As can be seen, in case of both multispectral and single band data, there are larger computational gains with higher error tolerance, and the approximate computation takes smaller fraction of the time taken by full computation. Also, it is evident there are higher computational gains with increasing clusters sizes, as is expected based on the complexity analysis.

Comparing the speed-ups achieved for the single band and multispectral data, it is observed that the gains are higher for higher dimensionality of the data. Again, this is expected, since there are higher computational savings for the computation of the term $\|(\mathbf{x}_i - \mathbf{x}_j)\|^2$ that is used in the exponential computation, because the proposed method does not explicitly compute the term.

5.5. CONCLUSION

Section 3 presented the SW-KRX detector which is a reformulated version of the original Kernel RX algorithm. The reformulated version entails the computation of the centered kernel Gram matrix which was defined over the augmented spectral-spatial vectors from the entire image. Section 4 presented the detailed theoretical development of the methodology for the fast computation of the kernel Gram matrix. The proposed methodology shifts the paradigm from pixel-based to cluster-based computation. Based on the methodology, a fast implementation of the SW-KRX algorithm has been developed for the special case of uncorrelated feature dimensions. This section presented detailed results for the SW-KRX algorithm. It was found that the SW-KRX detector gave better detection performance as compared to the RX detector under certain scenarios even with the restrictive assumption of uncorrelated feature dimensions. However, the performance was not superior to RX in all the cases.

As for the computational gains of the proposed algorithm, it was demonstrated that the proposed multivariate Taylor series based block approximation provided computational gains over the direct block computation. These gains were greater for the

multispectral data as compared to single band data. It was also demonstrated that overall the proposed methodology achieved speed-ups of up to 3-4 orders of magnitude, over the Kernel RX algorithm. This makes the proposed methodology more lucrative for implementation, especially over large datasets, as compared to the original algorithm.

6. BURIED MINE DETECTION USING CO-OCCURRENCE TEXTURE FEATURES

Detection of buried landmines in the airborne multispectral imagery is a challenging problem. For this problem, the baseline processing in several airborne detection systems is the popular RX anomaly detector [Reed, 1990]. Several techniques for improvement over RX detector output have been proposed in the past; however, most of these methods have focused on detection of surface mines [Agarwal, 2001; Filippidis, 2000; Beaven 2004], and few algorithms have been proposed specifically for buried mine detection [Ling, 2006; Bowman, 1998; McFee, 1997]. As a result, notable success has been reported for surface mine detection, but results for buried mine detection are far from satisfactory. In this section, a methodology for the detection of buried mines in airborne multispectral imagery using co-occurrence texture features is presented.

6.1. BURIED MINE DETECTION: REVIEW

This section presents a review of some of the techniques for buried mine detection proposed in the past. Lundberg [2001] developed a parametric model for the thermal signature of buried landmines in order to capture, in a few parameters, the variability in the signature due to weather, soil type, moisture content etc. Lundberg [2001] models the signature as a convolution of the ideal shape (top-view) of the buried mine and a smoothing kernel specified by two parameters, one for scaling and one for smoothing the mine shape depending on the depth of burial. A likelihood ratio test (LRT) based detector, which assumes the noise in the thermal signature to be a quarter-plane autoregressive process, is also developed. McFee and Ripley [1997] conducted extensive experiments using surrogate mines and blocks of explosives, buried under soil and vegetative cover, and scanning them using a *casi* hyperspectral imager to estimate the receiver operating characteristics for buried mines. They used the Linear Correlation Coefficient (LCC) to determine the similarity between the average spectral reflectance vector (over a 20-30 cm² area) and a reference vector, to detect any surface disturbances that might reveal presence of a buried mine. They also used linear unmixing of the spectral reflectance vector using Orthogonal Subspace Projection (OSP) to isolate the

mine signature. They reported slightly better performance for LCC, but suggested the need for conducting more studies.

Recently, Ling et al. [Ling, 2006] proposed a methodology for detection of buried mines in airborne mid-wave infrared imagery. They proposed a supervised image “chip” or patch based classification technique, in which the pixel intensities from each image chip from the MWIR image are clustered using a 3-D Adaptive Self-Organizing Map, and an intensity difference vector amongst the various clusters is computed. To perform detection, these difference vectors are then compared to the reference vectors from the buried mine library using the Kolmogorov-Smirnov test. Their approach highlights the need for investigating the intensity variations in the buried mine signature, instead of basing the detection decision on intensity values themselves, as is the case with the RX detector [Reed, 1990]. Amongst other notable work on using IR imagery for buried landmine detection, Bowman et al. [Bowman, 1998] present a consolidated overview of the challenges in buried mine detection using both spectral and thermal signatures. They describe the physical mechanisms behind the observed signature for various categories of background types (vegetation, tilled soil, untilled soil) and mines (both buried and surface) in different sensor modalities.

The primary challenge with using electro-optical multispectral data for buried mine detection comes from the fact that the spectral signature of the mine pixels is very similar to the spectral signature of the background constituents. Any technique utilizing anomaly detection or linear unmixing, both of which work directly with the intensities of the pixels (i.e. *appearance-based* methods), is rendered ineffective due to the substantial overlap of the spectral subspaces defined by the mine pixels and the background. In this work, an alternate approach of extracting features from the patches on the images, instead of utilizing the pixel intensities directly, is explored. This is done so as to exploit the fact that the tilling of the soil also creates intensity variations in the thermal signature of the mine area due to small shadows. Moreover, in case of spectral data, the signature of the disturbed soil typically seen at and around the area where a buried mine is placed is substantially different from that of the undisturbed soil. This difference typically arises due to the disparity in the particle sizes at different depths, which arises due to the fact that smaller particles (order of tens of micrometers) at the surface are eroded away due to

wind and rain, leaving a relatively lower concentration at the surface. Therefore, long-exposed and recently-disturbed surfaces have distinct spectral signatures. The reader is referred to the work by Bowman et al. [1998] for a detailed analysis of the spectral signatures under various conditions. In light of the aforementioned phenomenology, it is reasonable to expect that the information in the spatial variations of intensity may be more useful for detection, as opposed to the intensities directly.

To capture the information in intensity variations, first the cross-co-occurrence matrices (CCM) based texture features from the various patches in the multispectral images are extracted. Cross-co-occurrence matrices are extension of the popular Gray-Level Co-occurrence Matrices (GLCM) for texture feature extraction to color images. Based on extensive analysis of the raw texture features from both the mine and background patches, a subset of features with relatively high discriminatory information is selected. Next, a few features are short listed from this subset for detection, using a technique for critical variable selection called Principal Feature Analysis [Cohen, 2002], which takes the interdependencies of the features into account and eliminates redundant features. Finally, a detection strategy based on feature-based anomaly detection is developed to generate the final detector statistics, to demonstrate the ability of the features to effectively capture the information in intensity variations.

The proposed approach compactly includes the intensity variation information in the detection process. Also, since it is based on co-occurrence features, it is inherently invariant to illumination changes in the images, which, to a certain extent, addresses the problem of variability of mine signature. Finally, the proposed technique is amenable to fast implementation, since methods exist for the fast calculation of the co-occurrence features over an image [Argenti, 1990].

In Section 6.2, a brief overview of the multispectral data used in this part of the dissertation is presented. Section 6.3 contains a detailed description of the cross co-occurrence texture features that are extracted from the imagery to capture the color texture information. Section 6.4 presents the details of a systematic approach for the selection of co-occurrence texture features. First the Bhattacharya coefficients are used for the initial selection of discriminatory texture features, followed by principal feature analysis of the selected features, to identify the minimal set of features with mutually

uncorrelated information. To show the utility of the selected features, four different detectors, namely the Matched Filter, AND Fusion, Feature-based SW-KRX, and the Vegetation Mask detectors, are used for detection. Section 6.5 contains the description of these detectors. Section 6.6 provides comparative results in the form of Receiver Operating Characteristic (ROC) curves for the proposed methodology and the multiband RX detector. Finally, this section on buried mine detection is concluded in Section 6.7.

6.2. DATA OVERVIEW

The data used in this part of the dissertation is essentially the same four-band multispectral data as that described in Section 5. However, unlike Section 5, the focus here is on buried mine targets in the imagery. The results are reported on both Datasets 1 and 2. A total of 168 frames from 22 segments from Dataset 1, and 201 frames from 50 segments from Dataset 2 are used for analysis.

Figure 6.1 shows a typical segment from the Dataset 1 in the combined RGB bands (color images) where the frames are co-registered for convenient visualization. Figures 6.2 and 6.3 show examples of clearly visible and poorly visible buried mine signatures (51x51 pixels) in the segment, respectively. Figure 6.4 depicts patches in the segment which typically show up as false alarms in the detection process. Figures 6.5, 6.6, 6.7 and 6.8 are the same as Figures 6.1, 6.2, 6.3 and 6.4, respectively, except the near-IR band signatures are shown. Figures 6.3 and 6.7 show that not all the buried mines have a distinct signature, and in some cases the signature is largely indistinguishable from the background to a human observer. Conversely, looking at Figures 6.4 and 6.8, it is apparent that several of the background patches are visually very similar to the buried mine signatures. It was also observed that some mine targets are in close vicinity of background features, like vegetation and other terrain features.

The goal of this work is to demonstrate the effective extraction of the texture information in a minimal set of features from the imagery for improved detection. Therefore, feature reduction and selection is an important step of the proposed methodology. For efficient selection of discriminatory features, a set of actual buried mine signatures and background patches are extracted from the imagery, using available groundtruth information.

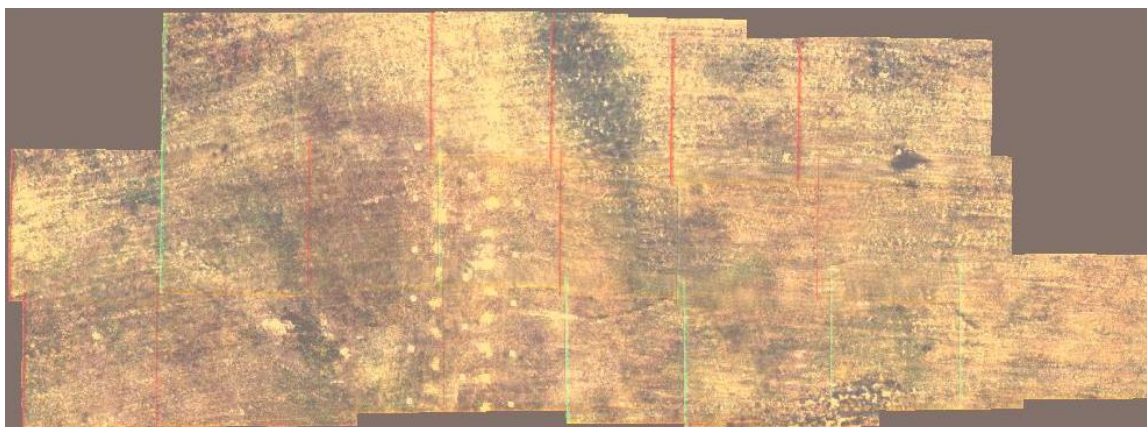


Figure 6.1. Typical Data Segment (RGB Bands) Containing Buried Mine Signatures.

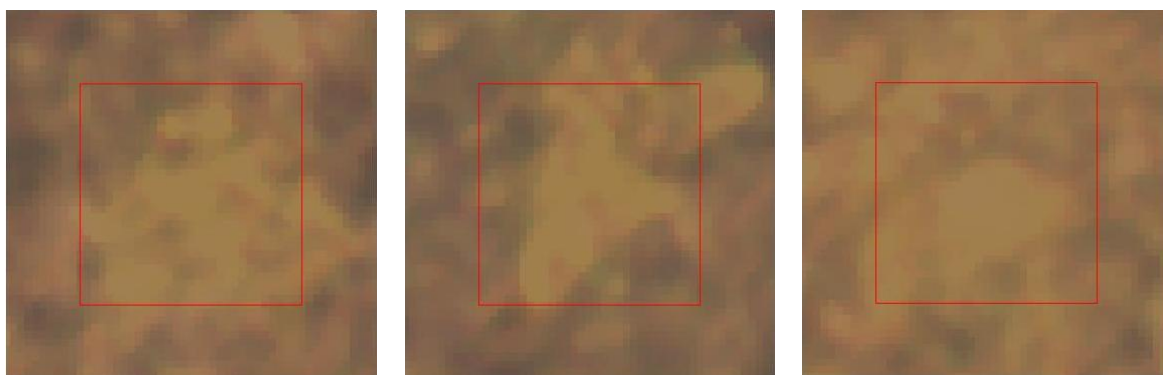


Figure 6.2. Examples of Clearly Distinguishable Buried Mine Signatures (RGB Bands).

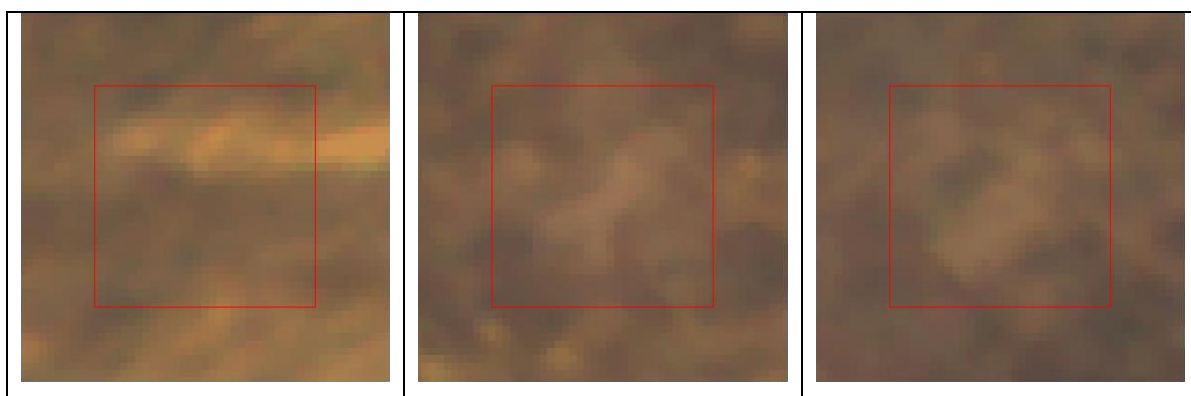


Figure 6.3. Examples of Poorly Distinguishable Buried Mine Signatures (RGB Bands).

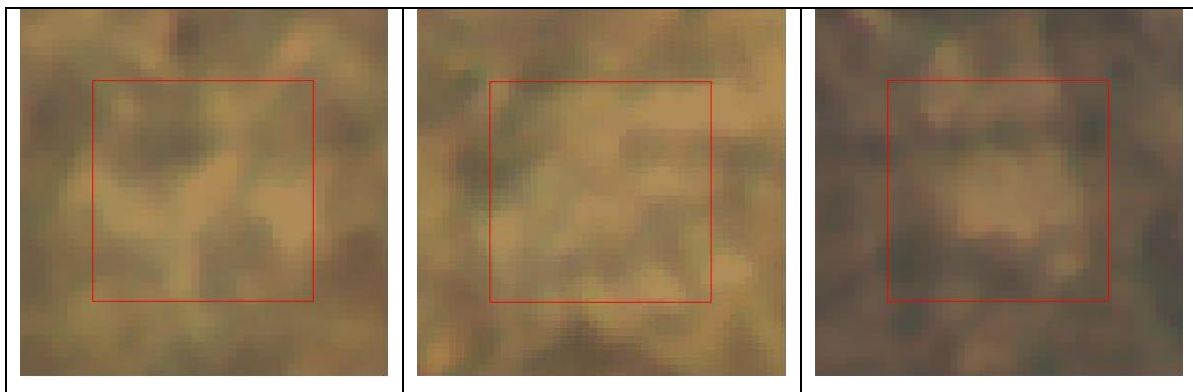


Figure 6.4. Typical False Alarm Patches (RGB Bands).

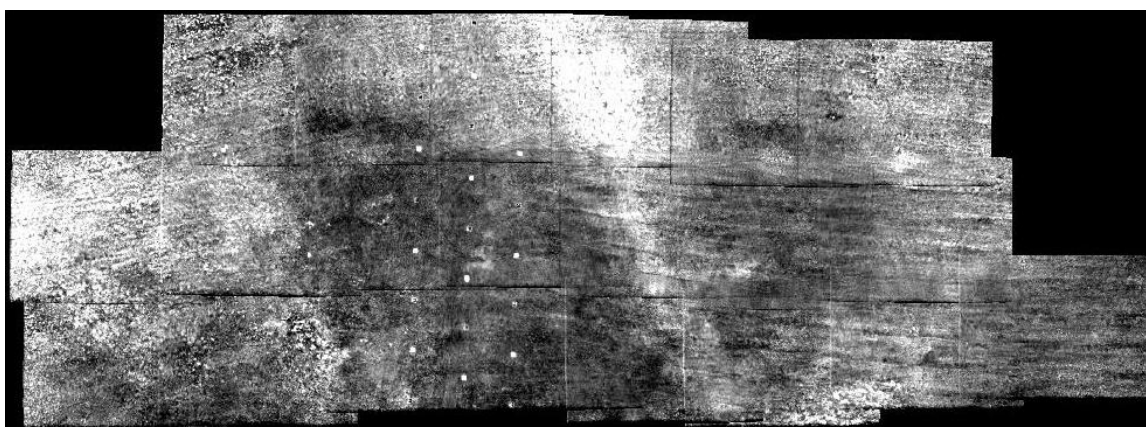


Figure 6.5. Typical Data Segment (NIR Band) Containing Buried Mine Signatures.

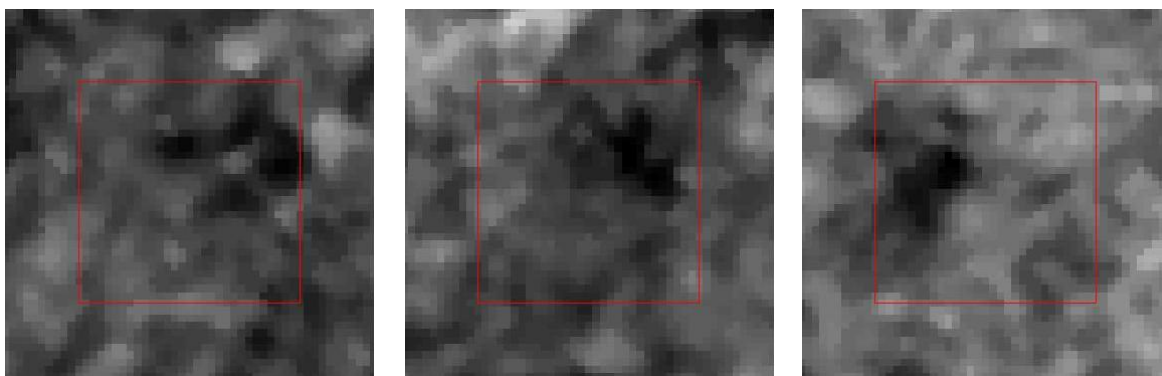


Figure 6.6. Examples of Clearly Distinguishable Buried Mine Signatures (NIR Band).

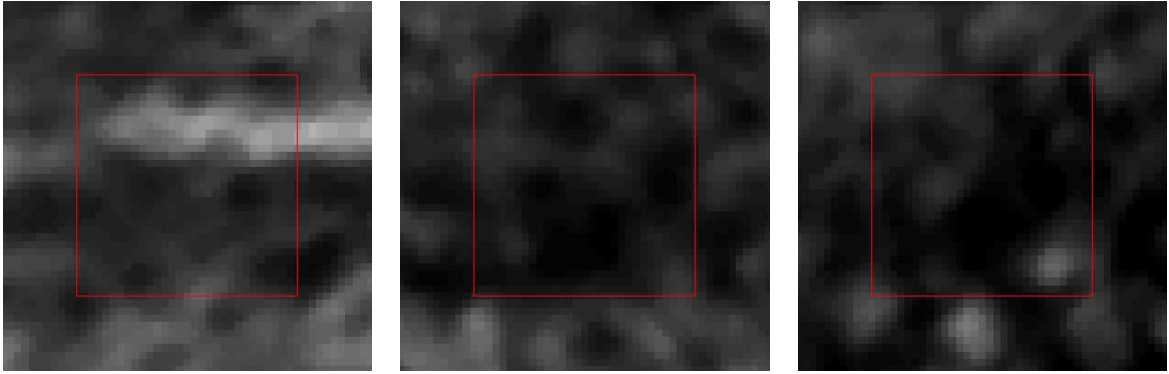


Figure 6.7. Examples of Poorly Distinguishable Buried Mine Signatures (NIR Band).

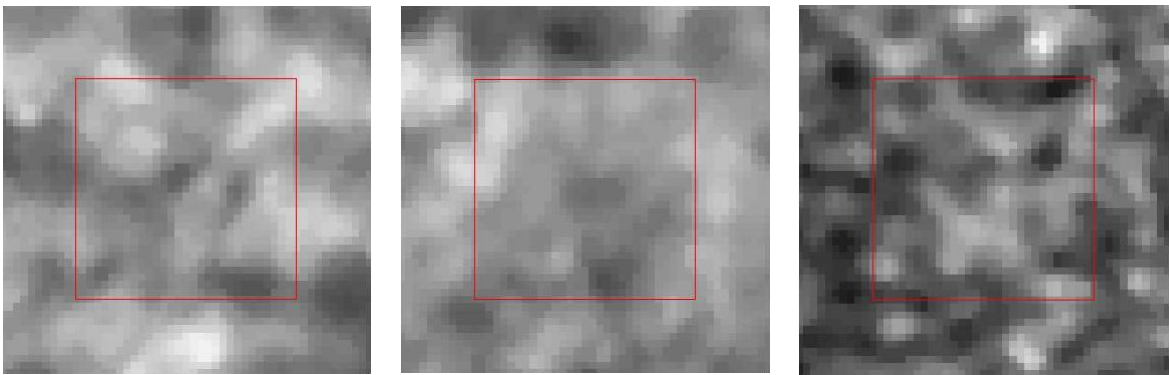


Figure 6.8. Typical False Alarm Patches (NIR Band).

The selection methodology for these patches is as follows. The set of actual buried mine patches are selected based on the ground truth for the dataset, which is available to us. However, the background patches are selected based on the values of the RX anomaly detector. From each frame of data, 40 locations with the highest RX values are selected and a patch (15x15 pixels) at each location is extracted as examples of background patches for further processing. Essentially, these patches are representative of the patches with high potential for generating false alarms. Any potential false alarm patches that spatially overlap with the actual mine patches or other known targets are removed to avoid repetition of patches. For Dataset 1, a total of 545 mine patches and 14,792 non-mine patches (potential false alarms) are obtained from 168 frames from 20

segments. For Dataset 2, a total of 1,251 mine patches and 6,709 non-mine patches (potential false alarms) are obtained from 201 frames from 50 segments.

6.3. CO-OCCURRENCE TEXTURE FEATURES

The current method is based cross-co-occurrence (CCM) texture features, extracted from a given patch (15x15 pixels) around each location. Since CCM texture features are a direct extension of the Gray-level Co-occurrence Matrix (GLCM) texture features to multispectral images, first the description of GLCM features is presented in Section 6.3.1. This description is followed by the description of the CCM features for multispectral images in Section 6.3.2.

6.3.1. Gray-Level Co-occurrence Texture Features. Gray-level Co-occurrence Matrix [Haralick, 1973, 1979] is a well known method for texture analysis. GLCM estimates second-order statistics from an image. A co-occurrence matrix is an estimate of the joint probability density function of gray-level pairs along a given direction and distance in any region of the image (typically over the target window). The idea is to capture the average information regarding the coarseness and direction of the texture in the region. For pattern recognition, detection or image segmentation, a set of these textural parameters are calculated over a window centered at a given pixel to define a texture feature vector.

To describe the gray-level co-occurrence matrix, an $M \times M$ window in the image denoted by W is defined, centered at the pixel under consideration. The pixel at location (p, q) in the window is denoted as w_{pq} . Let there be N gray levels in the image and let $\delta = (\delta_x, \delta_y)$ be the displacement vector. Then the gray-level co-occurrence matrix \mathbf{G} is an $N \times N$ square matrix whose entries are defined as follows: $\mathbf{G}^\delta = [G_{ij}^\delta]$, where $G_{ij}^\delta = |K_{ij}^\delta|$, where $|\cdot|$ denotes the cardinality of a set and the set K_{ij}^δ is defined as follows:

$$K_{ij}^\delta = \{(w_{pq}, w_{rs}) \mid r = p + \delta_x, s = q + \delta_y, w_{pq} = i, w_{rs} = j\} \quad (6.1)$$

Thus, each entry G_{ij}^δ of the gray-level co-occurrence matrix counts the number of times a given pair of gray values (i, j) occurs at a displacement of δ in the window W .

The co-occurrence matrix is typically normalized by the number of pixel pairs in the window (N_δ) to give the joint probability density estimate $P_{ij}^\delta = G_{ij}^\delta / N_\delta$. The displacement vector δ plays an important role in texture analysis. For a coarse texture, whose texture elements are larger than δ , most of the energy will be centered along the diagonal of the GLCM matrix. In case of finely-grained texture (relative to δ), the entries will be mostly off diagonal. Furthermore, in case of patterned textures the co-occurrence entries G_{ij}^δ are found at only a few locations in GLCM matrix.

Haralick [1979] proposed 14 different features defined over the co-occurrence matrices, for distinguishing between different co-occurrence matrices. Several studies have been conducted for testing the effectiveness of these features for texture description. GLCM texture features have been found to perform better texture classification than fractal, Markov Random Field, and Gabor filter features [Ohanian, 1992]. For our work in buried mine detection, seven of the more frequently quoted features in literature are evaluated, to identify which of them are useful for the problem at hand. The seven GLCM features are defined as:

$$\text{Maximum Probability:} \quad \text{Max. Prob.} = \max(P_{ij}^\delta) \quad (6.2a)$$

$$\text{Energy:} \quad \text{Energy} = \sum_{i,j} (P_{ij}^\delta)^2 \quad (6.2b)$$

$$\text{Contrast:} \quad \text{Contrast} = \sum_{i,j} (i-j)^2 P_{ij}^\delta \quad (6.2c)$$

$$\text{Inverse Difference Contrast:} \quad \text{IDC} = \sum_{i,j} \frac{1}{1+(i-j)^2} P_{ij}^\delta \quad (6.2d)$$

$$\text{Correlation:} \quad \text{Corr.} = - \sum_{i,j} \frac{(i-\mu_x)(j-\mu_y)}{\sqrt{\sigma_x \sigma_y}} P_{ij}^\delta \quad (6.2e)$$

$$\text{Variance:} \quad \text{Var.} = \sum_{i,j} (i-\mu_x)^2 P_{ij}^\delta \quad (6.2f)$$

$$\text{Entropy:} \quad \text{Ent.} = - \sum_{i,j} P_{ij}^\delta \log P_{ij}^\delta \quad (6.2g)$$

where the mean and variance along the row and the column of the co-occurrence matrix are defined as:

$$\mu_x = \sum_i i \sum_j P_{ij}^\delta$$

$$\mu_y = \sum_j j \sum_i P_{ij}^\delta$$

$$\sigma_x = \sum_i (i - \mu_x)^2 \sum_j P_{ij}^\delta$$

$$\sigma_y = \sum_j (j - \mu_y)^2 \sum_i P_{ij}^\delta$$

The texture features in Eqn. (6.2) are computed for a set of displacement vectors. The displacement vectors are defined using distances and angles, with distances from the set $dist = \{2, 4\}$ and angles from $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. The displacement vector at an angle of θ and distance $dist$ is denoted as d_θ^{dist} . Composite GLCM matrices are formed for a given distance by combining the GLCM for all the angles at that distance. This is done as follows. Let P_θ^d denote the co-occurrence matrix at a distance d and angle θ . Then the composite GLCM matrix for distance d , P^d , is given as:

$$P^d = \frac{1}{4}(P_0^d + P_{45}^d + P_{90}^d + P_{135}^d) \quad (6.3)$$

The composite GLCM matrix is computed for 2 different distances, $d = 2$ and $d = 4$. In addition to this, GLCM matrix is also computed for the zero displacement vector $d_0^0 = (0, 0)$ as the 3rd displacement vector. Thus, each location is defined by 3x7 GLCM features.

6.3.2. Cross Co-occurrence Texture Features. The Gray-level Co-occurrence Matrix described in the last section is defined for scalar images. Arvis et al. [2004] extended the concept of co-occurrence matrices to multispectral images and defined cross-co-occurrence matrices. The key idea in their work is to calculate the co-occurrences not just within, but also between the color bands, so as to take into account the correlations between the bands and get a complete *color* texture description. Several approaches to color texture description have been proposed in the past. Arvis et al. [2004] reported that cross-co-occurrence matrix texture features perform better classification, as compared to methods based on joint color-texture features (gray-scale texture and color

features like moments or histograms computed separately [Maenpaa, 2002; Drimbarean, 2001]) and co-occurrences based on color quantization (converting the color image to a gray-scale image using color binning [Chen, 2002; Hauta-Kasari, 1996]).

The cross-co-occurrence matrices for multispectral images are briefly described here. Consider a multispectral image with J bands. Let W be an $M \times M$ window in the multispectral image centered at the given pixel. The value of the g^{th} band at pixel location (p, q) in the multispectral image is denoted as w_{pq}^g . The cross-co-occurrence matrices are defined for a pair of bands. Thus, for every pixel in the J band image, there are $J(J+1)/2$ cross-co-occurrence matrices. The definition of cross-co-occurrence matrix between bands f and g (\mathbf{G}_{fg}^δ) is obtained by modifying the definition of set K_{ij}^δ for gray-level co-occurrence matrix from Eqn. (6.1) as follows:

$$K_{fg,ij}^\delta = \{(w_{pq}^f, w_{rs}^g) \mid r = p + \delta_x, s = q + \delta_y, w_{pq}^f = i, w_{rs}^g = j\} \quad (6.4)$$

Thus, $K_{fg,ij}^\delta$ is the set of pixel pairs between f^{th} and g^{th} bands of the image, which have intensities i and j , respectively, and $\mathbf{G}_{fg}^\delta = [G_{fg,ij}^\delta]$, where $G_{fg,ij}^\delta = |K_{fg,ij}^\delta|$.

Once the cross-co-occurrence matrices are computed, the aforementioned GLCM texture features can be computed and used to represent the matrices. It is noted that it is entirely possible to have different dynamic ranges and, hence, different numbers of quantization levels for different bands. In the case where the quantization levels of the two bands under consideration are not the same, the co-occurrence matrix is no longer square. However, that does not change the definitions of any of the seven texture features given in Eqn. (6.2).

In addition to the seven features mentioned above, another feature called Normalized Color Index (NCI) is included, which is defined as follows:

$$NCI = \sum_{i,j} \frac{(i-j)}{(i+j)} P_{ij}^\delta \quad (6.5)$$

Note that the NCI feature for the cross-co-occurrence matrix between the red and near-infrared bands and a displacement of $\delta = (0,0)$, is a modified version of the Normalized Difference Vegetation Index (NDVI), a well-known feature in field of remote sensing. At a pixel (i, j) , NDVI is defined as:

$$NDVI(i, j) = \frac{N(i, j) - R(i, j)}{N(i, j) + R(i, j)}$$

where N and R are the near-infrared and red bands of the multispectral image, respectively. NDVI, which was first formally proposed by Rouse et al. [Rouse, 1973], is a good indicator of the presence of vegetation in multispectral imagery. Note that the NCI feature defined over Red-Near-infrared CCM at $\delta = (0,0)$ is the mean NDVI that averages over the target window. However, it is noted that for the within-band co-occurrence matrices (RR, GG, BB, NN), the NCI feature at zero displacement $\delta = (0,0)$ will always be zero.

Thus, given these eight features for each cross-co-occurrence matrix, every detected location in the multispectral image is described by $3 \times 8 \times J(J+1)/2$ texture features.

6.4. DISCRIMINATORY FEATURE SELECTION

The amount of texture information captured in a particular feature depends on the texture type and the displacement vector used for computing co-occurrence matrices. The goal is to capture maximum information in the fewest of the texture features. A two stage approach for the selection of the minimal set of discriminatory features is adopted. The first step is to identify texture features with relatively high discriminatory information (in terms of their ability to separate false alarms from the mine signatures). This is done using the Bhattacharya coefficient based analysis of the features. In the second step, Principal Feature Analysis (PFA) is used to reduce the features selected from the first step to a set of uncorrelated features. Details of initial selection of the discriminatory texture features based on Bhattacharya coefficient is expounded upon in Section 6.4.1, and the uncorrelated feature subset selection using PFA is presented in Section 6.4.2.

6.4.1. Discriminatory Texture Feature Reduction. For this step, first the cross-co-occurrence features, as described in Section 6.3.2 are extracted from the entire set of mine and non-mine patches. All the eight cross-co-occurrence features are calculated for all ten possible band combinations, and for all the 3 displacements. Thus, a total of $3 \times 10 \times 8$ (240) features are extracted for each patch. Once the features are extracted for the entire set of mine and non-mine patches, the distribution for each feature over the mine

and non-mine patches is estimated separately. The idea is to calculate the disparity in the distribution of each feature between the mine and non-mine patches. The kernel density estimation method is used to obtain a non-parametric estimate of the distributions. This is done primarily to overcome the effect of different numbers of mine and non-mine sample patches, and compute a more robust and representative estimate of the distributions. The values for each texture feature are binned into $N = 50$ equidistant bins, between the minimum and maximum feature values, for mine and non-mine patches separately. Let us denote the vector of values in each bin for a feature f for the mine and non-mine patches as \mathbf{x}_m^f and \mathbf{x}_n^f , i.e.,

$$\mathbf{x}_m^f = [x_{m1}^f, x_{m2}^f, \dots, x_{mN}^f]$$

$$\mathbf{x}_n^f = [x_{n1}^f, x_{n2}^f, \dots, x_{nN}^f]$$

Also, the vector containing the bin centers is denoted as $\mathbf{b}^f = [b_1^f, b_2^f, \dots, b_N^f]$, which is the same for both the mine and non-mine patches. Then the kernel density estimate for the feature distribution over mine patches (q_m^f) and the non-mine patches (q_n^f) can be calculated as

$$q_m^f(y) = \frac{1}{N\sigma} \sum_{i=1}^N K\left(\frac{y - x_{mi}^f}{\sigma}\right) \quad \text{and} \quad q_n^f(y) = \frac{1}{N\sigma} \sum_{i=1}^N K\left(\frac{y - x_{ni}^f}{\sigma}\right) \quad (6.6)$$

where K is the kernel function. In this work, the Gaussian function with unit width and zero mean is used. Note that q_m^f and q_n^f are continuous density estimates. Figure 6.9 shows these distributions for mine and non-mine patches for three features, namely the Energy, IDC and the NCI features. From the feature definitions in Eqn. (6.2b), it is noted that the energy feature is high when the probability is high for only a few elements of the co-occurrence matrix. Similarly, the IDC feature values (Eqn. (6.2d)) tend to be higher when most of the high probability elements in the co-occurrence matrix lie along the diagonal. Both of these are true when target signature is relatively smooth. It was observed that the variation in intensity tends to be smaller for most of the mine signatures. This is reflected in the feature value distributions for the IDC and energy

features shown in Figure 6.9a and 6.9b, respectively, where the IDC and energy values for mine patches are higher than those of non-mine patches.

Next, the Bhattacharya coefficient between the feature distribution over the mine and non-mine patches is computed. Bhattacharya coefficient [Kailath, 1967], which is a divergence-type measure, is defined for the two densities for the feature f as:

$$B_f = \int \sqrt{q_m^f(y)q_n^f(y)}dy \quad (6.6)$$

The Bhattacharya coefficient can be interpreted geometrically as the cosine of the angle enclosed by the square root of the two density functions.

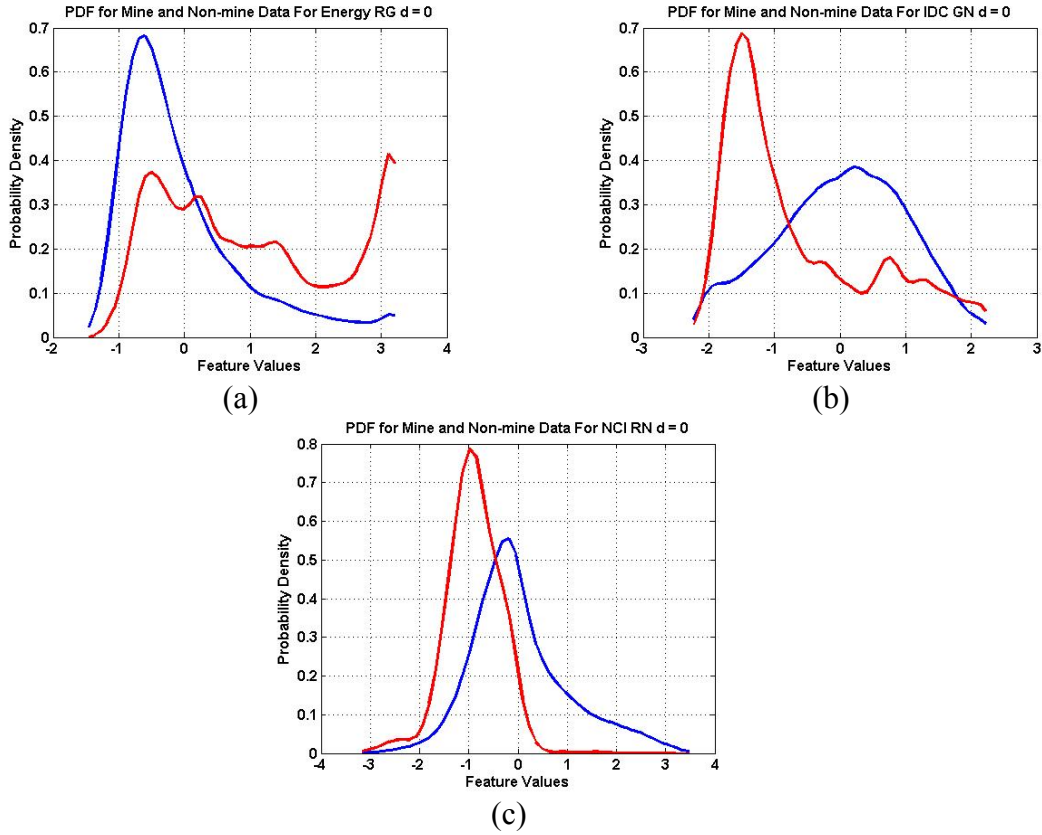


Figure 6.9. Kernel Density Estimates of the PDF for the Three Selected Features for Mine (red) and Non-mine (blue) Patches from Dataset 1. (a) Energy $d = 0$, RG, (b) IDC $d = 0$, GN, (c) NCI $d = 0$, RN.

Features with lower values of Bhattacharya coefficient have higher disparity between their mine and non-mine distributions, i.e., higher discriminatory information. Bhattacharya coefficient is computed in the aforementioned fashion for all the 240 features. Figure 6.10 graphically illustrates the relative Bhattacharya coefficient values for various features. The various displacement vectors are plotted along the rows, with the first, second and the third row corresponding to distances of $d = 2, 4$ and 0 , respectively. The columns are arranged separately for different band combinations. The first eight columns show the eight features over the RR cross-co-occurrence matrix, and the second set of eight columns shows values for the RG cross-co-occurrence matrix and so on, in the order shown in Figure 6.10.

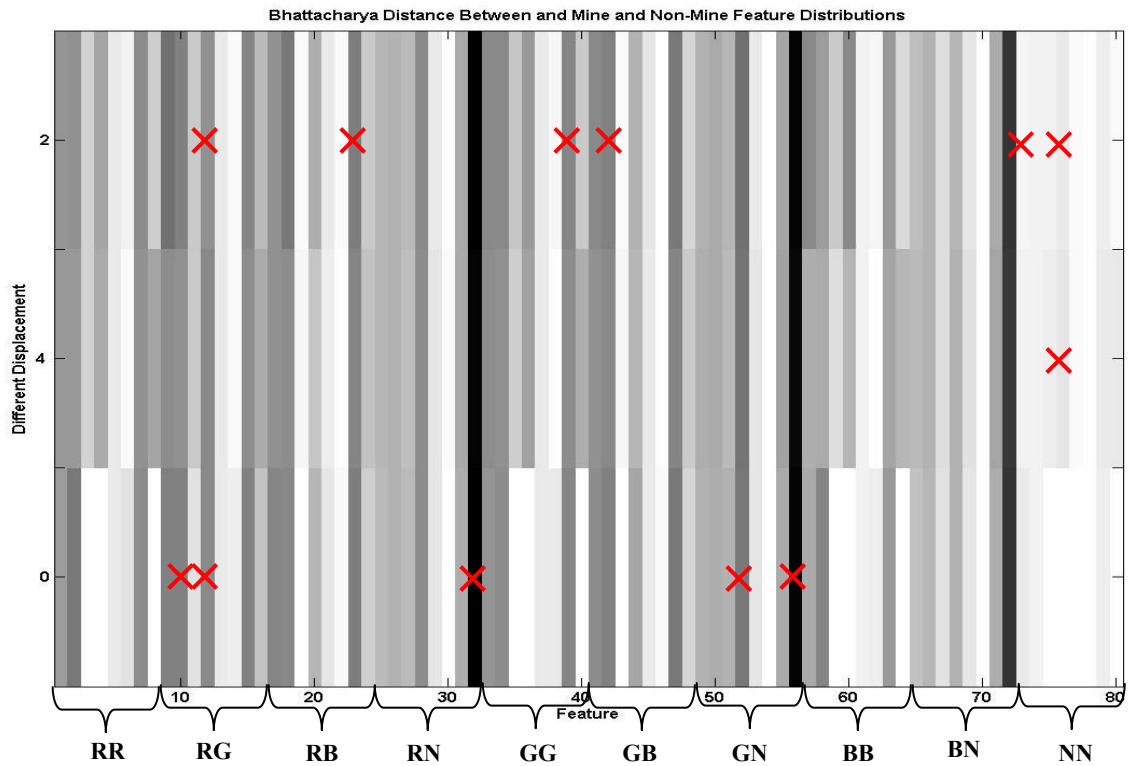


Figure 6.10. Bhattacharyya Coefficient Values for Different Features for Dataset 1. The Eight Features (1: Max. Prob., 2: Energy, 3: Contrast, 4: IDC, 5: Corr., 6: Var., 7: Entropy, 8: NCI) are Plotted for all Possible Cross-Co-Occurrence Matrices Along the Columns as Shown, for the Different Displacement Vectors that Vary Along the Rows.

As can be seen, some features have lower values (darker shade) of Bhattacharya coefficient and, thus, higher discriminatory information than others. These features with lower values of Bhattacharya coefficient are more suitable for mine and non-mine discrimination. The lower values of Bhattacharya coefficients are shown in darker shades in Figure 6.10. The coefficient values in Figure 6.10 range from 0.8056 to 1. Note that the NCI features for the RN, GN and BN band combinations (columns 32, 56 and 72, respectively) have low values of Bhattacharya coefficients. In order to reduce computational complexity, a set of 12 features with relatively low values of Bhattacharya coefficients were selected for further analysis. The selected 12 features are shown with cross-marks (red) in Figure 6.10.

6.4.2. Feature Selection Using Principal Feature Analysis. Feature selection based on Bhattacharya coefficient only helps to identify individual texture features that might be more useful to discriminate between mines and non-mines. However, it does not take into account any inter-dependencies and correlations amongst these features. Thus, features selected based on Bhattacharya coefficient may be correlated and thus will be redundant. The goal is to obtain a small set of features that will allow us to effectively discriminate between mine and non-mine detections. As a second step to the feature selection process, Principal Feature Analysis (PFA) is used for the selection of a subset of uncorrelated features from the 12 short listed features. The PFA method used here is proposed by Cohen et al. [Cohen, 2002], and is described here briefly in a stepwise fashion.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]^T$ be the matrix containing all the feature vectors for both the mine and non-mine patches, where M is the total number of mine and non-mine patches combined.

Step 1: Normalize each feature to zero mean and unit variance.

Step 2: Compute the sample covariance matrix \mathbf{C} from the data and calculate the eigenvectors and corresponding eigenvalues of the covariance matrix as $\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, where \mathbf{U} contains the eigenvectors along the columns and \mathbf{D} is a diagonal matrix containing the corresponding eigenvalues.

Step 3: Choose first q eigenvectors with highest eigenvalues; q can either be a fixed number or determined based on the energy to be retained in the selected

eigenvectors. The reduced set of q eigenvectors is denoted as \mathbf{U}_q . \mathbf{U}_q is an $M \times q$ matrix. Let the square of the rows (each element squared) of the matrix \mathbf{U}_q be denoted as $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$.

Step 4: Cluster the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$ into k clusters using k-means clustering algorithm. The distance used here for the k-means algorithm is Euclidean distance.

Step 5: For each cluster, select vector \mathbf{v}_i belonging to that cluster that has the lowest value of Bhattacharya coefficient. The feature corresponding to vector \mathbf{v}_i is called the principal feature for that cluster. Thus, k features are selected as principal features, from k clusters.

For PFA on the co-occurrence features from Dataset 1, the first three eigenvectors are used for clustering, i.e. $q = 3$. Also, the number of clusters k was chosen to be $k = 3$, which was determined empirically. The 12 selected features are clustered into 3 clusters, as follows:

Cluster 1: (1) $d = 0$, NCI-RN, (2) $d = 0$, NCI-GN, (3) $d = 0$, NCI-BN

Cluster 2: (1) $d = 0$, Energy-RG, (2) $d = 2$, Energy-GB, (3) $d = 2$, Entropy-RB, (4) $d = 2$, Entropy-GG, (5) $d = 2$, IDC-RG

Cluster 3: (1) $d = 0$, IDC-GN, (2) $d = 2$, Max. Prob.-BN, (3) $d = 2$, IDC-NN, (4) $d = 4$ IDC-NN

The 3 principal features obtained are:

(1) $d = 0$, NCI-RN, (2) $d = 0$, Energy-RG, (3) $d = 0$, IDC-GN

These 3 principal features are used to represent each patch in the data. Note that all the selected features are cross-band features between red, green and NIR bands. This indicates the importance of color as a feature. Also, note that the first feature is the average NDVI feature, which is indicative of presence of vegetation at a location. This feature was also amongst the set of selected features in an earlier reported work [Tiwari, 2007]. Moreover, similar to the selected feature set here, it was the IDC and Energy features that were short listed in the final set of features in [Tiwari, 2007], although for different spectral band combinations.

Figures 6.11a shows an example frame from Dataset 1. Figure 6.11b shows the composite color image where the values from the three features are mapped into an RGB

color image for the frame in Figure 6.11a. Figure 6.12 shows another example image and the corresponding composite feature image.

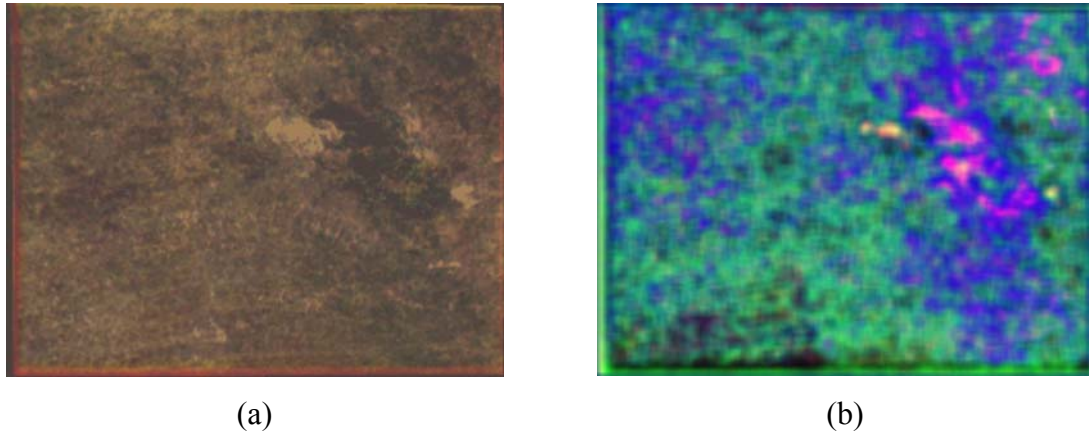


Figure 6.11. Example 1 of a Feature Image as a Composite RGB Image (a) Original Frame 1, (b) Corresponding Composite Feature Image.

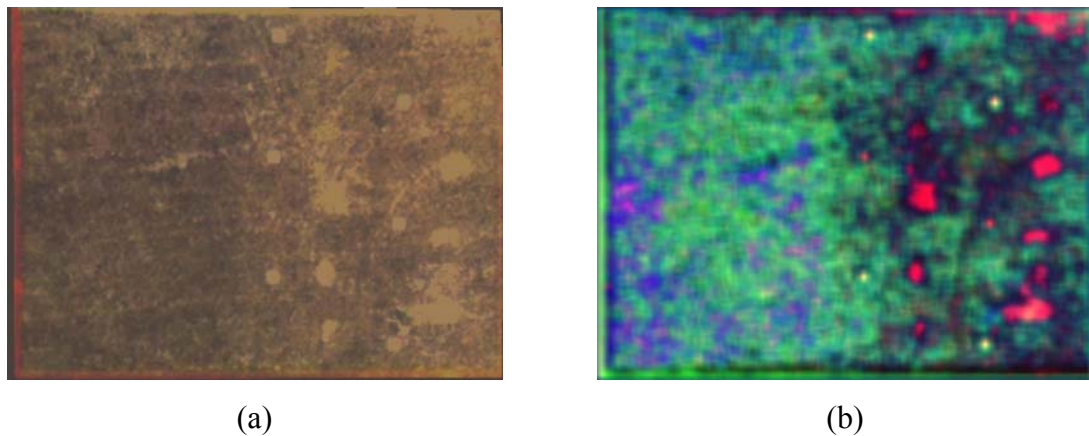


Figure 6.12. Example 2 of a Feature Image as a Composite RGB Image (a) Original Frame 2, (b) Corresponding Composite Feature Image.

6.5. FEATURE-BASED DETECTORS

To demonstrate the efficacy of the selected features in extracting useful discriminatory information, these are used in four different feature-based detectors for buried mine detection. The four detectors are named as the Matched Filter detector, the

AND Fusion detector, the Feature-based SW-KRX detector, and the Vegetation Mask detector. The Matched Filter detector and AND Fusion detector are supervised and semi-supervised in nature, respectively. The Feature-based SW-KRX and Vegetation Mask detectors are completely unsupervised in nature in that no information from the training set of mine and non-mine signatures are needed for computing the detector statistics. The Matched Filter, AND Fusion and Feature-based SW-KRX detectors are extension of similar standard detectors to the feature images. However, the Vegetation Mask detector is more heuristically based, motivated by the empirical observations on the selected features. The performance results for these four detectors are shown to demonstrate the effectiveness of the buried mine detection strategy using the selected features in general, irrespective of the detector used. The results highlight the performance that can typically be expected by using the selected features. However, it is possible to devise other detection and classification strategies which might improve the detection performance further.

It is pointed out that the detection is done for all the pixels in the images, i.e., the detectors are applied on the complete feature images. This is different from the detection methodology reported earlier in [Tiwari, 2007], where the detector statistics were computed for only those patches which gave a high multi-band RX detector statistics. In that sense, the earlier strategy was more akin to a false alarm mitigation strategy, whereas the current methodology is a direct anomaly detection strategy.

A brief description of the Matched Filter detector, the AND Fusion detector, the Feature-based SW-KRX detector, and the Vegetation Mask detector is presented in Sections 6.5.1, 6.5.2, 6.5.3 and 6.5.4, respectively. Corresponding detection performance results for the detectors are shown in Section 6.6.

6.5.1. Matched Filter Detector. Each pixel in the image data is described by a vector of three selected texture features. The matched filter detector essentially “matches” a given feature vector with typical signature(s) of buried mines. In this sense, it is a supervised detector as it looks for a specific type of feature vector. In this work, the typical buried mine signatures, which are termed *representative mine signatures* (RMSs), are obtained from the training data used for Bhattacharya coefficient based analysis.

Given a set of normalized³ mine and non-mine signatures, the mine feature vector which has the highest ratio of the mines to non-mines in a selected neighborhood around it in the feature space, is selected as the RMS. Since there can potentially be more than one type of mine signature, it is quite natural to select more than one RMS. In that case, first all the mines and non-mines within a certain neighborhood of the previously selected RMS are removed from consideration. Then, the same process is followed for selecting the second RMS on the reduced set. In this way, multiple RMSs can be selected for the detection process. The number of RMSs to use for the matched filter is determined empirically.

Suppose there are L representative mine signatures denoted by $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_L\}$. Then the matched filter detector test statistic for a given pixel, whose feature vector is denoted by \mathbf{t} , is defined as follows:

$$MFD(\mathbf{t}) = \exp\left(-\frac{\mathbf{d}^2}{\sigma_0}\right), \quad \mathbf{d} = \min\{\|\mathbf{t} - \mathbf{r}_1\|, \|\mathbf{t} - \mathbf{r}_2\|, \dots, \|\mathbf{t} - \mathbf{r}_L\|\} \quad (6.7)$$

Here, $\|\cdot\|$ denotes the Euclidean distance, and σ_0 is chosen to be $\sigma_0 = 1$. As can be seen from Eqn. (6.7), patches with feature vectors ‘similar’ to one of the RMSs will have higher value of the MCD detector output. Results for one and two RMSs are shown in Section 6.6

6.5.2. AND Fusion Detector. The AND Fusion detector does not require an explicit set of mine and non-mine signatures, like the matched filter detector. The idea here is to generate a composite test statistic based on the three features. To this end, first a mapping function for each feature is determined which maps the feature values between 0 and 1. This mapping function in this work is chosen to be a step-wise constant function. The feature range for a particular feature is divided into a certain number of bins (quantization levels) say ‘ m ’, whose edges are determined such that each bin covers the same area on the probability density function (PDF), estimated from the training data. Each bin is mapped to a different level between 0-1. The mapping function is determined for each feature separately, based on the corresponding non-mine PDF. Let the mapped

³ Here “normalization” refers to the transformation of making the feature set zero mean and unit variance.

features (say L of them) for a pixel with feature vector \mathbf{t} , be denoted by $\{\mathbf{r}_1^t, \mathbf{r}_2^t, \dots, \mathbf{r}_L^t\}$. Then the AND Fusion detector statistic is given as:

$$AFD(\mathbf{t}) = \max_{i=1}^L [\mathbf{r}_1^t, \mathbf{r}_2^t, \dots, \mathbf{r}_L^t] \quad (6.8)$$

In the results shown in Section 6.6, the number of features used is $L = 3$ and the number of quantization bins used for each of the three features is $m = 10$.

6.5.3. Feature-based SW-KRX Detector. The third detector is essentially the SW-KRX anomaly detector (special case implementation), as described in Sections 3, 4 and 5, applied to the feature images. The results for the SW-KRX anomaly detector are compiled for two different modes of operation. For the first case, the three feature images, obtained from computation of the three features at each pixel location, are arranged into a multiband image. The SW-KRX detector statistics are computed on the feature multiband images. In the second mode, the SW-KRX detector statistics are computed on combined multispectral and feature images. The feature images are added on to the multispectral images as additional bands (total of $4+3 = 7$ bands), and a scaling is applied to each band so as to bring each band to approximately the same mean value. The SW-KRX detector on these composite images gives the statistics in second mode. The feature-based SW-KRX detector is completely unsupervised as no information from the training set of mine and non-mine signatures is utilized.

6.5.4. Vegetation Mask Detector. This detection strategy is based on the heuristics derived from the empirical observations on the feature images. This detector uses two of the three selected features, namely Energy-RG and NCI-RN features. The first step in this detector is anomaly detection on the selected energy feature image, which is followed by the modulation of the anomaly detector statistics based on the NCI feature. Figures 6.13a shows an example frame from Dataset 1 with the buried mine signatures enclosed in green boxes.

Figure 6.13b shows the feature image for the Energy-RG feature for the same frame. It should be noted that the energy feature shows high values where there are less variations in the image i.e., relatively flat areas. Based on empirical observations it is found that the buried mine signatures are relatively smoother as compared to the

background, and give high values for the energy feature. Figure 6.13c show the feature image for the NCI-RN features. Since it is the average NDVI feature, it has high values where vegetation is present. Figure 6.14 depicts another example image from Dataset 2 and the corresponding feature images.

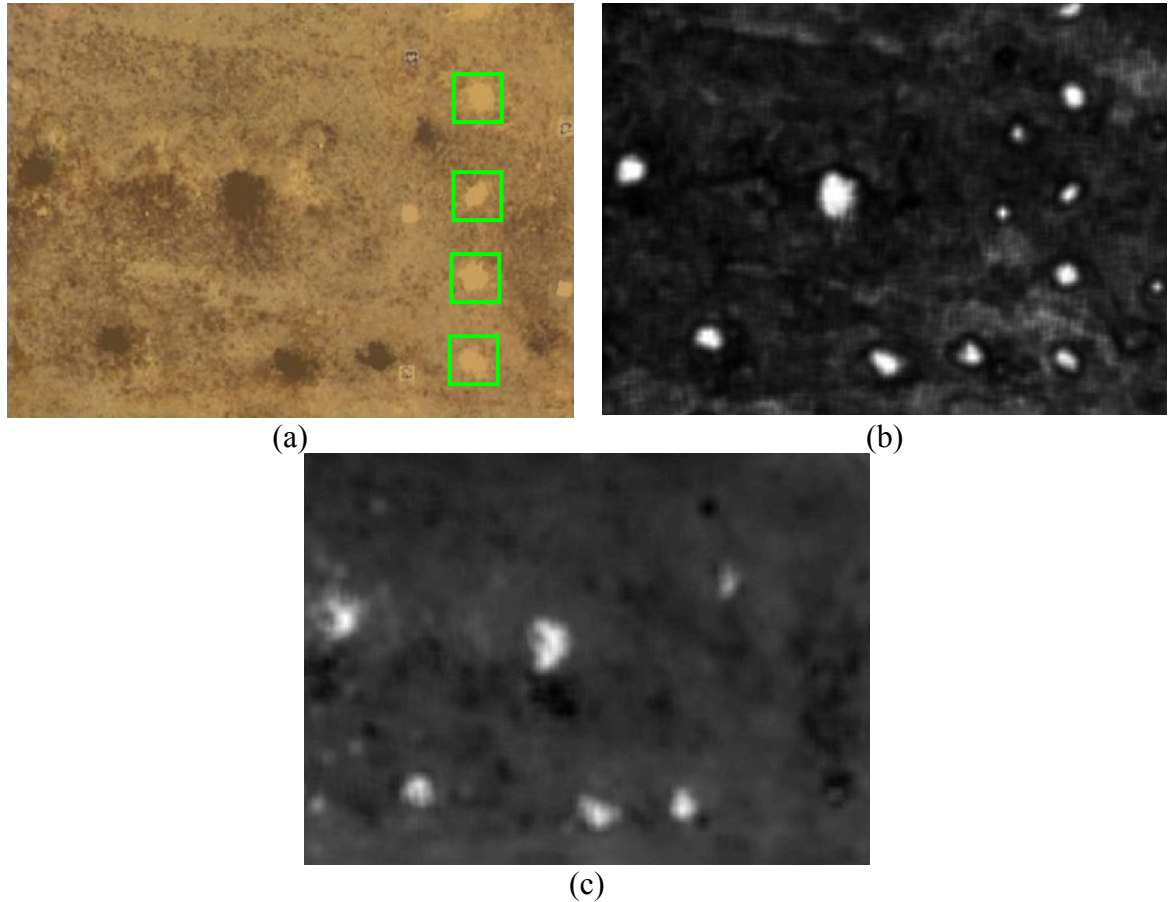


Figure 6.13. Example 1 of a Feature Image as a Composite RGB Image from Dataset 1 (a) Original Frame 1, (b) Corresponding Energy-RG Feature Image, (c) Corresponding NCI-RN Feature Image.

It can be seen in Figures 6.13a and 6.14a that the typical buried mine signatures (marked with green boxes) show relatively fewer variations in the color values as compared to the background, i.e. is relatively smoother. Therefore, the energy feature gives high values in and around the buried mine signature patches, as can be seen in Figures 6.13b and 6.14b. Moreover, the feature values are typically significantly different

from the background, and form a distinct region around the mine signature, like a local anomaly. Therefore, it is expected that anomaly detection on the energy feature image will successfully indicate the presence of buried mines. However, it is found that the major vegetation patches, containing bush or tree tops, also tend to have smaller spectral variations. Consequently, such vegetation patches also tend to give high values for the energy feature. This phenomenon can be observed clearly in Figures 6.13. Therefore, basing the detection only on energy feature gives false alarms on or around the vegetation patches.

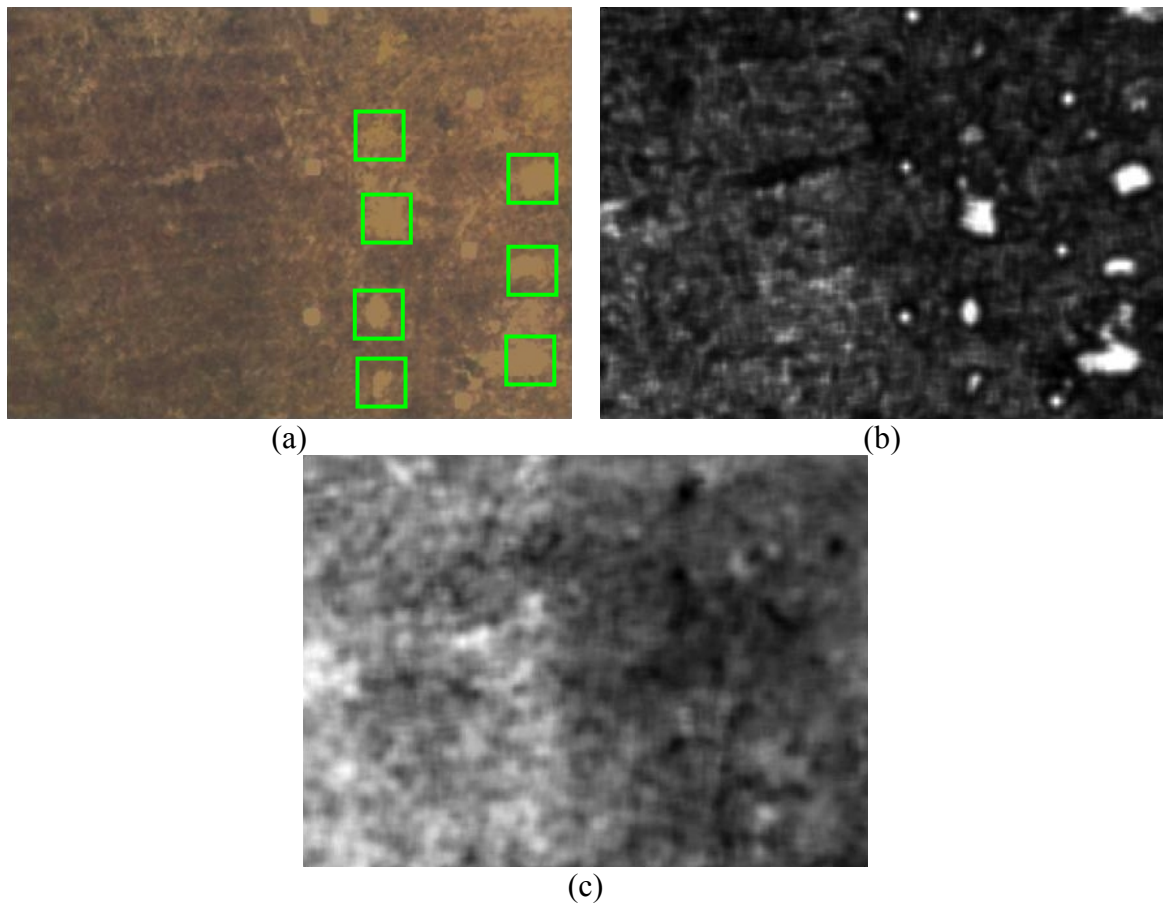


Figure 6.14. Example 2 of a Feature Image as a Composite RGB Image from Dataset 2
(a) Original Frame 2, (b) Corresponding Energy-RG Feature Image, (c) Corresponding
NCI-RN Feature Image.

However, note that the second selected feature, i.e. the NCI feature, is a robust indicator of the presence of the vegetation in the image. The idea is to modulate the detector statistics obtained by using the anomaly detector on the energy feature image, based on the NCI features.

The first step of the feature-based detection strategy is anomaly detection on the Energy-RG feature images. The RX detector is used for this anomaly detection step. Figure 6.15 shows the Energy-RG feature image from the example image shown in Figure 6.13, and the corresponding RX anomaly detector output. It can be seen that the RX detector gives high output in the regions with high values of the Energy-RG feature, including the vegetation regions. The second step is to develop a “mask” image based on the NCI-RN feature. First, a binary image is obtained by thresholding the NCI-RN feature image. The threshold is chosen empirically such that the feature values of the actual vegetation patches are greater than the threshold. This is followed by a morphological opening and closing operation on the binary images, to retain only the sizeable vegetation regions and removing the scattered smaller regions. Morphological image dilation is applied to the resulting image, to get the final “mask” image. The idea is that the NCI-RX feature-based binary “mask” image delineates the vegetation regions in the image. Figure 6.16a shows the NCI-RN mask image for the same image from Figures 6.15 and 6.13, obtained from the aforementioned morphological operations on the NCI-RN feature image.

As the final step, the RX detector statistics for each pixel is multiplied by the corresponding pixel value of the mask image, to give the final feature-based detector statistics. Figure 6.16b shows the result of the multiplication of the mask image with the RX detector output. It can be seen that the high values of the RX detector in and around the vegetation regions are suppressed. Note that the proposed Vegetation Mask detector is completely unsupervised as no information from the training set of mine and non-mine signatures is utilized for computation of the detector statistics.

6.6. RESULTS AND DISCUSSION

In this section, comparative results on the performance of the proposed buried mine detection methodology, vis-à-vis the multiband RX anomaly detector is presented,

to highlight the improvement in detection performance. Detection performance of the Matched Filter, AND Fusion and Feature-based SW-KRX detectors are shown for Dataset 1. The detection performance for the Vegetation Mask detector is shown for both Datasets 1 and 2.

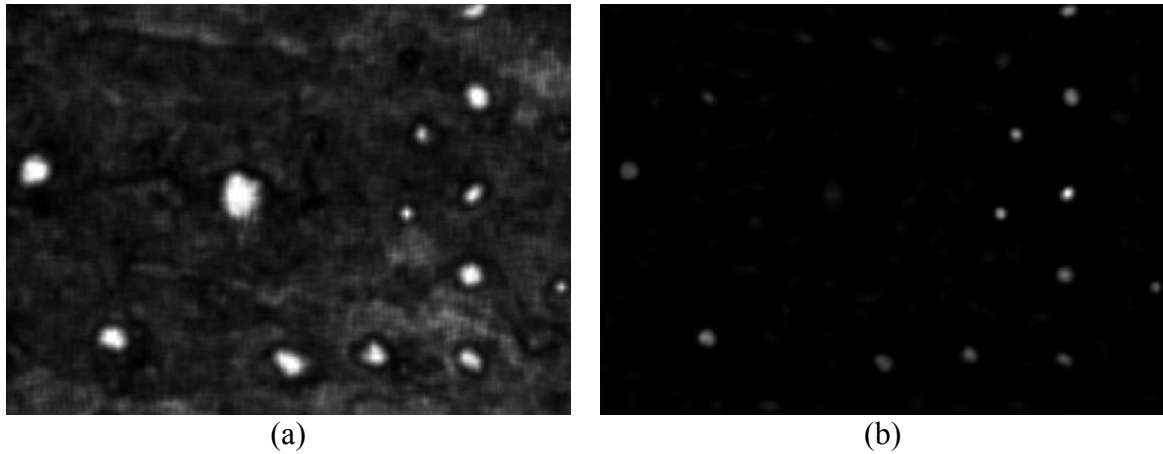


Figure 6.15. RX Detector on Energy-RG Feature Image from Dataset 1 (a) Energy-RG Feature Image, (b) Corresponding RX Detector Output.

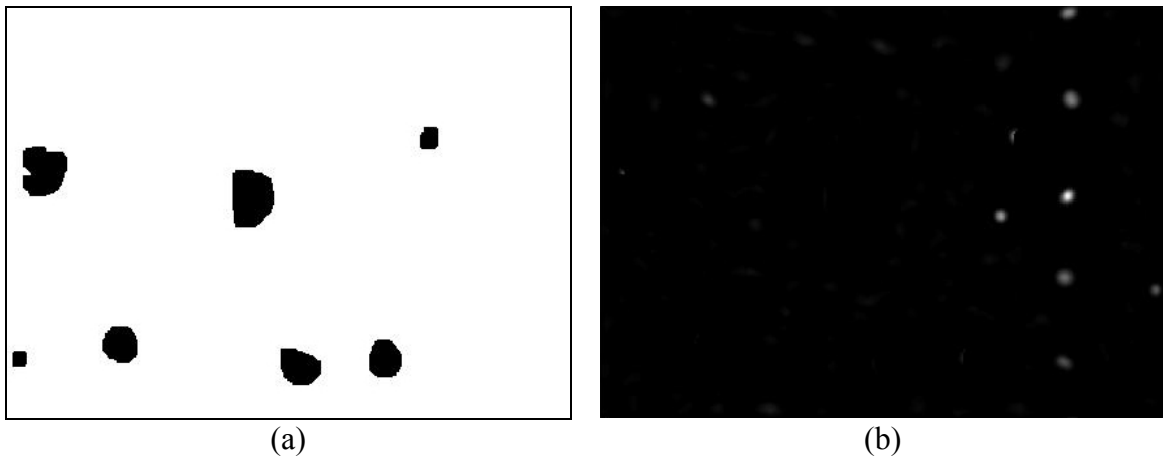


Figure 6.16. Vegetation Based Masking of RX Detector Statistics (a) Mask Image, (b) RX Detector Output with Vegetation Based Masking.

For the results presented here, the cross-co-occurrence matrices are calculated over a window size of 15x15 pixels. As a pre-processing step, the dynamic range of various bands of the multispectral image is restricted to the following intervals: $50 \leq R \leq 1000$, $200 \leq G \leq 2600$, $100 \leq B \leq 600$, and $400 \leq N \leq 2600$ (the letter N is used to denote the near-infrared band). This is done to eliminate any bias in the further processing due to extremely high sensor values for some ground features such as fiducial markers. Next, the mean and standard deviation of each band for a given segment are estimated, and the dynamic range of the bands is further restricted to two standard deviations around the mean. Next, the values in the various bands are uniformly quantized to eight levels before cross-co-occurrence matrices are calculated. The dynamic range reduction, based on mean and standard deviation, helps to avoid the loss of texture information in the mine signatures in case of low contrast signatures during the quantization step. The selected three CCM features as listed in Section 6.4 are extracted for all the pixels, which are subsequently used for detection. Note that the Vegetation Mask detector is uses only two of the three selected features. However, the rest of the three detectors use all the three features for detection. In addition to the proposed methodology, multiband RX detector's test statistics are calculated using the following mask sizes: *demeaning radius* = 25, *blanking radius* = 15 and *target radius* = 6.

First, the results for the Matched Filter detector are presented. Figure 6.17 shows the receiver operating characteristic (ROC) curves for the feature-based matched filter and the multiband RX anomaly detector. ROC curves are shown for different number of representative mine signatures. The selection mechanism for the RMSs is described in Section 6.5.1.

As can be seen, the detection performance is characteristically different for the case of single and two RMSs. The ROC curves are representative of the case when there are two different 'type' of mine signatures, which are fairly distinctly clustered into two clusters in the feature space. As can be seen, in case of a single RMS the performance is comparatively better for lower false alarm rates (FAR), but does not improve beyond a certain detection rate even at high FARs. This is indicative of the fact that there are mine signatures that do not give high detector statistics, i.e., they are significantly dissimilar from the RMS.

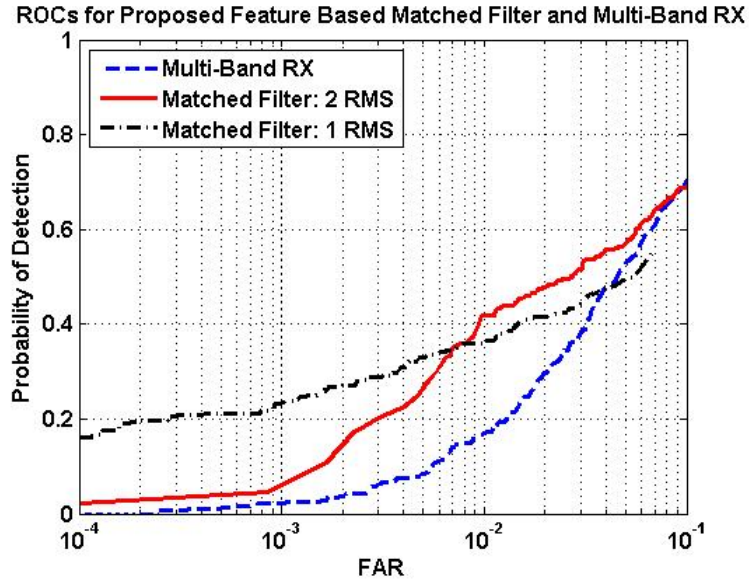


Figure 6.17. Comparison of Detection Performance of the Matched Filter Detector for Different Number of Representative Mine Signatures, vis-à-vis Multiband RX Anomaly Detector.

The ROC curve for two RMSs corroborates the conclusion that there are two different types of mine signatures, since the inclusion of the second RMS improves the detection performance at higher FARs. However, this improvement comes at the cost of higher false alarm rate, since the probability of a false alarm patch falling “close enough” to one of the cluster centroids also increases. In case of high degree of similarity between the mine patches, the detection performance of the matched filter detector is not expected to improve for higher number of RMSs. This is because the selected mine patches do not form distinct clusters in the feature space, the average distance of a patch feature vector from the nearest cluster does not improve significantly with increasing number of clusters. As seen in Figure 6.17, the proposed methodology has a significantly improved detection performance as compared to multiband RX.

Figure 6.18 presents the detection performance results for the AND fusion detector. The AND Fusion detector essentially maps the individual feature values into meaningful individual feature detector statistics, using a mapping function based on the feature PDF. It then takes the maximum value amongst all the mapped feature values as the final AND Fusion detector statistics. The detection performance for the detector is shown in Figure 6.18, which shows improvement in performance over multiband RX.

The performance is better than the matched filter with 2 RMSs at lower FAR (up to about 0.005 FA/m^2), but is similar at higher FARs.

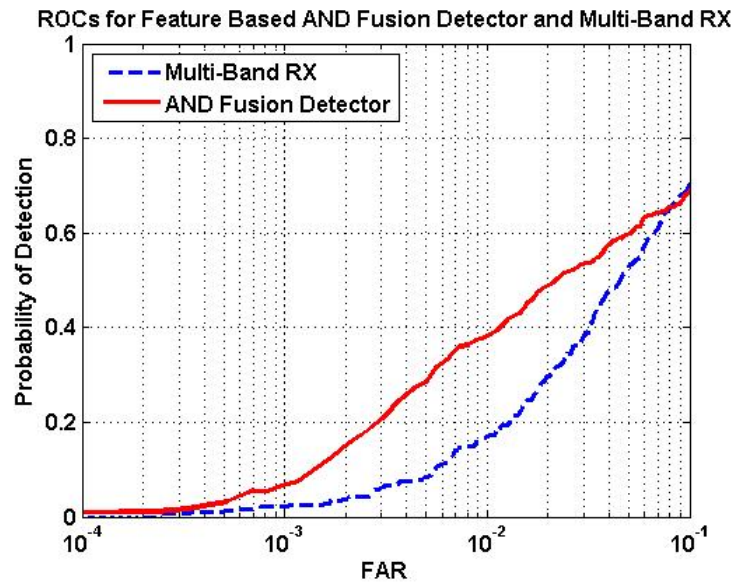


Figure 6.18. Comparison of Detection Performance of the AND Fusion Detector, vis-à-vis Multiband RX Anomaly Detector.

Next, the detection performance for feature-based SW-KRX detector is presented in Figure 6.19. As mentioned earlier, this detector is applied in two different modes. In the first mode, the detector is applied to the multiband image with each band representing each feature image. Thus, each pixel contains the 3-element vector representation of the texture. In the second mode, the 4 multispectral bands of the images are combined with the 3-band feature image to get 7-band composite image. The SW-KRX detector is applied on the composite image. The ROC curve in black shows the detection performance for the first mode. As can be seen in Figure 6.11b and 6.12b, the values for the 3-selected features are typically significantly different from the background, and form a distinct region around the mine signature, like a local anomaly. The SW-KRX is able to pick up on these anomalies and this is reflected in the detection performance of the detector, which shows improvement over not just the multiband RX algorithm, but also the other detectors. It should be noted that the detection rates show improvement at all

FARs upto approximately 0.05 FA/m^2 . The falloff around 0.05 FA/m^2 is due to the presence of mine signatures which are not significantly distinct from the background and do not show up as anomalies in the feature images. This problem is corrected slightly with the operation in the second mode, where the multispectral bands are also included. As can be seen, in the second mode (the red ROC curve) the detection performance at higher FARs (over 0.01 FA/m^2) improves over the first mode. Overall, the performance is better than the multiband RX upto 0.1 FA/m^2 , beyond which it is similar.

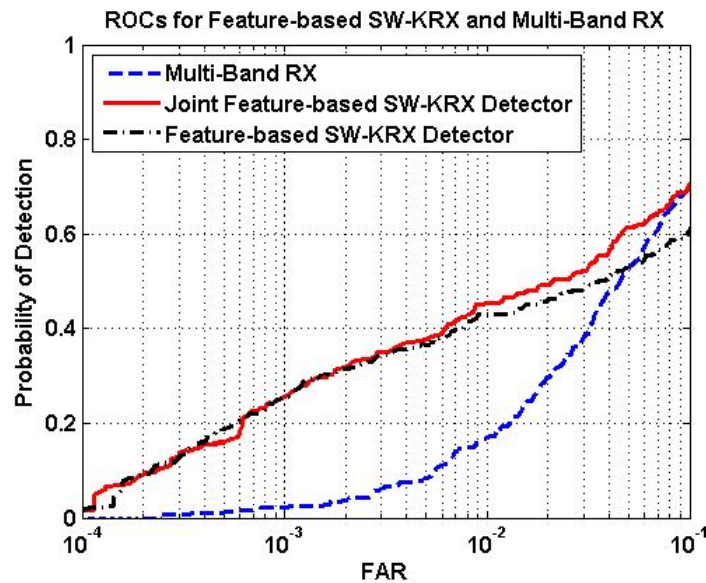


Figure 6.19. Comparison of Detection Performance of the Feature-based SW-KRX Detector (Black: Feature based SW-KRX, Red: Joint Multispectral and Feature-based SW-KRX), vis-à-vis Multiband RX Anomaly Detector.

The detection performance results for the Matched Filter, AND Fusion and Feature-based SW-KRX detector shown here are for Dataset 1 only. For Dataset 2, the performance of these feature-based detectors was found to be similar to that of the multiband RX detector. However, the results for the Vegetation Mask detector are shown for both Datasets 1 and 2.

For the Vegetation Mask detector, the RX detector statistics on the Energy-RG feature images are calculated for the following mask size values: *demeaning radius* = 30, *blanking radius* = 20 and *target radius* = 8. The threshold value for the thresholding of the NCI-RN feature image is chosen to be 0.6. For the morphological opening and

closing operations, a 5x5 square structuring element is used. For the morphological dilation step, a 7x7 square structuring element is used. In addition to the proposed methodology, multiband RX detector's test statistics are also calculated using the following mask size values: *demeaning radius* = 25, *blanking radius* = 15 and *target radius* = 6.

Figure 6.20 shows the receiver operating characteristic (ROC) curves for the Vegetation Mask detector and the multiband RX anomaly detector for Dataset 1. The feature-based detector is able to pick up on the anomalies in the Energy-RG feature image, and this is reflected in the detection performance of the detector, which shows significant improvement over the multiband RX algorithm. It should be noted that the detection rates show improvement at all FARs upto approximately 0.05 FA/m². The falloff around 0.05 FA/m² is due to the presence of mine signatures which are not significantly distinct from the background and do not show up as anomalies in the feature images.

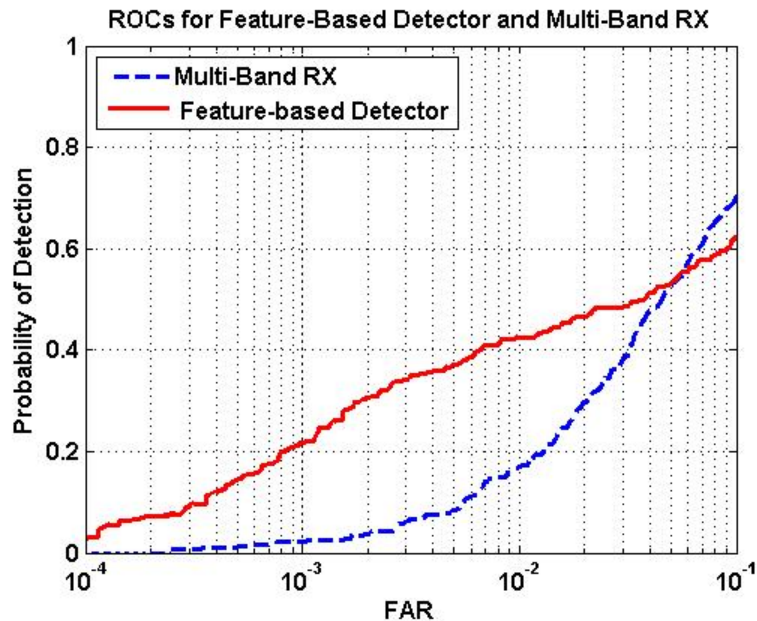


Figure 6.20. Detection Performance of the Vegetation Mask Detector vis-à-vis Multiband RX Anomaly Detector for Dataset 1.

Similarly, Figure 6.21 shows the receiver operating characteristic (ROC) curves for the feature-based detection strategy and the multiband RX anomaly detector for Dataset 2. Again, similar to Dataset 1, the detection performance of the feature-based detection strategy shows improvement over the multiband RX anomaly detector. However, the improvement in performance is not as high as that shown for Dataset 1. The primary reason for that, in general, the contrast of the images in Dataset 2 is not as high as that in images in Dataset 1. Due to this, the buried mine signatures show relatively greater distinction from the background in Dataset 1, than in Dataset 2. This is also reflected in the feature images, and consequently, in the final feature-based detector statistics.

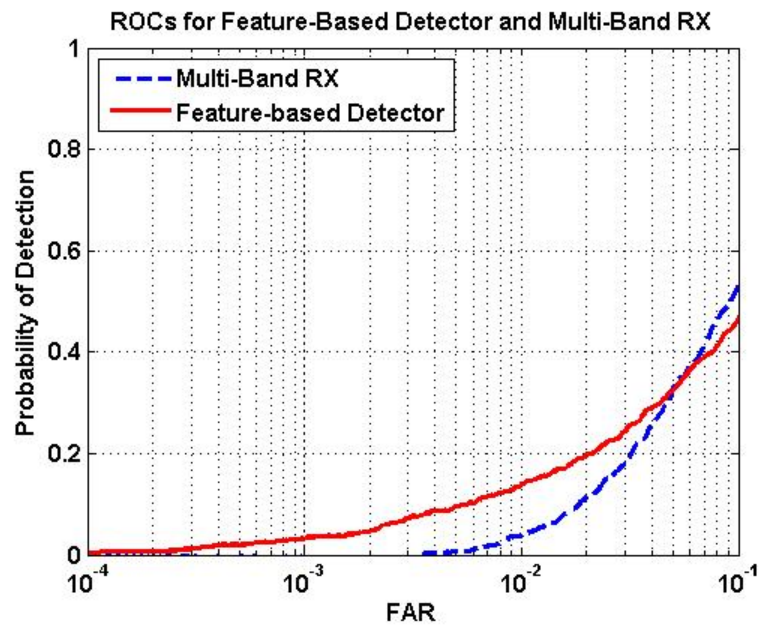


Figure 6.21. Detection Performance of the Vegetation Mask Detector vis-à-vis Multiband RX Anomaly Detector for Dataset 1.

These results provide the proof-of-concept that the selected texture features have substantial discriminatory information, and can substantially reduce false alarms and improve detection performance. It is noted that the results on the various detectors were

presented to demonstrate performance that can typically be expected by using the selected features. Further improvement in the detection performance is possible for different detection and classification strategies.

6.7. CONCLUSION

This section presents a methodology for buried mine detection in multispectral images based on cross co-occurrence texture features. First the raw CCM features are analyzed individually, based on the disparity of their distribution over the mine and non-mine patches. Principal feature analysis is used for selecting the final set of three features, which are then used for generating a test statistic for detection. Different feature-based detectors are presented, which can be classified as supervised, semi-supervised and unsupervised in nature. Comparative detection results for the detectors are depicted, which shows improvement over the traditional multiband RX approach for buried mine detection. The results are also indicative of the presence of the significant discriminatory information in the selected features, which can be harnessed for effective buried mine detection.

7. CONCLUSIONS AND FUTURE WORK

This work considers the problem of detection in airborne spatial data. Two different detection algorithms for spatial data have been proposed. Performance of the detection algorithms is demonstrated for the airborne landmine detection data. However, the proposed algorithms are not restricted to any specific spatial dataset.

The first part of the dissertation presents the development of a fast approximate implementation of the kernel-based nonlinear anomaly detector called Kernel RX. First a reformulated version, termed the Spatially Weighted Kernel RX (SW-KRX), of the original detector is proposed. A novel framework, which is based on computing the detector statistic using all the pixels in the image while maintaining local adaptivity, is presented. It is shown that under the proposed framework, the detector statistics can be computed as a function of the centered kernel Gram matrix defined over the entire image. Next, a detailed development of the fast computation of the kernel Gram matrix is presented. The proposed method uses a cluster-based representation of the data, to obtain a sparse block representation of the kernel Gram matrix. Also, a method for an outer product decomposition of each block of the kernel Gram matrix is presented, which allows for fast computation of the detector statistics. The details of the fast centering of the diagonal elements of the kernel Gram matrix are also presented. Based on the fast computation of the kernel Gram matrix and the fast centering of its diagonal elements, an implementation for a special case of the SW-KRX detector has been developed. The underlying assumption in the implementation is that of uncorrelatedness and unit variance of the various feature dimensions in the non-linear feature space. It is demonstrated that, even with this assumption, the SW-KRX detector shows better detection performance under certain scenarios, as compared to the RX anomaly detector. This has been demonstrated for both the multispectral and single band data, although better results have been obtained for single band data due to the absence of problems like band misalignment. In terms of the computational gains, the proposed methodology has been shown to be 3-4 orders of magnitude faster than the original kernel RX algorithm. Results have also demonstrated the efficacy of the multivariate Taylor series based block approximation in reducing the computational burden.

Although the current implementation of the SW-KRX (with the specific assumption) gives superior performance vis-à-vis RX detector under some scenarios, it is expected that the SW-KRX detector in its complete form will improve the detection performance universally. Therefore, as part of the future direction of work on this algorithm, a technique for the fast centering of the non-diagonal elements of the kernel Gram matrix needs to be explored. Moreover, for fast implementation of the general case detector, methods for fast eigenvalue decomposition of the centered kernel Gram matrix, based on the outer product decomposition has to be developed. This work provides the details of the underlying framework necessary for the aforementioned developments. The performance of the proposed SW-KRX detector on large hyperspectral datasets can also be explored.

The second part of the dissertation presented a methodology for the detection of buried mines in spatial data, through efficient extraction of the information in the spatial distribution of the spectral vectors. The proposed methodology extracts the spectral texture information using cross-co-occurrence features. Although, cross-co-occurrence texture features as color texture features have been proposed in the past, they have not been used extensively and are relatively less explored. A new color texture feature, termed Normalized Color Index (NCI), defined on the cross-co-occurrence matrices is proposed. This feature is similar to the Normalized Difference Vegetative Index, popular in the remote sensing community, but is more general in definition and scope. The algorithm is one of the first to exploit color texture information in the airborne multispectral images, using cross-co-occurrence texture features, for buried mine detection. A unique two stage scheme, using Bhattacharya coefficients and Principal feature analysis, is proposed for discriminatory feature selection. This feature selection process gives a minimal set of uncorrelated features, containing discriminatory information. Finally, details of the different feature-based detectors are presented. Comparative detection performance results are presented for the different feature-based detectors vis-à-vis the multiband RX detector. These results demonstrate the efficient extraction of texture information via the CCM features and the efficacy of the feature selection process. Currently, the feature selection process is semi-automatic in that the initial stage of the feature reduction process using Bhattacharya coefficients is based on

manual selection of features. In the future, techniques for automatic feature reduction can be explored. Also, methods for fast implementation of some of the co-occurrence features such as NCI exist, and have been implemented here as such. Techniques for fast computation of other features, like Energy and Entropy, need to be explored.

APPENDIX A.
DETAILED DERIVATION OF THE KERNEL-RX DETECTOR STATISTICS

The detector statistics for the Kernel RX detector can be given as:

$$KRX(\mathbf{y}_r) = \Phi_C(\mathbf{y}_r)^T \hat{\mathbf{C}}_\Phi^\# \Phi_C(\mathbf{y}_r) \quad (\text{A.1})$$

where $\hat{\mathbf{C}}_\Phi^\#$ is the pseudo-inverse of the covariance matrix in the feature space $\hat{\mathbf{C}}_\Phi$. Let the eigen-decomposition of $\hat{\mathbf{C}}_\Phi$ is given by:

$$\hat{\mathbf{C}}_\Phi = \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi \mathbf{U}_\Phi^T \quad (\text{A.2})$$

where, $\mathbf{U}_\Phi = [\mathbf{u}_\Phi^1, \mathbf{u}_\Phi^2, \dots, \mathbf{u}_\Phi^{N_c}]$, is the matrix containing the eigenvectors along its columns. Then, the pseudo inverse of the covariance matrix $\hat{\mathbf{C}}_\Phi$ can be written as:

$$\hat{\mathbf{C}}_\Phi^\# = \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi^{-1} \mathbf{U}_\Phi^T \quad (\text{A.3})$$

As shown in Appendix B, the relationship between the eigenvectors of the covariance matrix $\hat{\mathbf{C}}_\Phi$ and the centered kernel Gram matrix $\hat{\mathbf{K}}$ is given by:

$$\mathbf{U}_\Phi = \mathbf{Y}_C^\Phi \mathbf{\alpha} \mathbf{\Lambda}_\Phi^{-1/2} \quad (\text{A.4})$$

where $\mathbf{\alpha}$ is the matrix containing the eigenvectors for the centered kernel Gram matrix $\hat{\mathbf{K}}$ (see Appendix B).

Writing the KRX statistics from Eqn. (A.1) in terms of the pseudo inverse of the covariance matrix from Eqn. (A.3):

$$KRX(\mathbf{y}_r) = \Phi_C(\mathbf{y}_r)^T \mathbf{Y}_C^\Phi \mathbf{\alpha} \mathbf{\Lambda}_\Phi^{-2} \mathbf{\alpha}^T \mathbf{Y}_C^{\Phi T} \Phi_C(\mathbf{y}_r) \quad (\text{A.5})$$

where, $\Phi_C(\mathbf{y}_r)$ is the *centered* target vector.

Define:

$$\hat{\mathbf{k}}_r = \mathbf{Y}_C^{\Phi T} \Phi_C(\mathbf{y}_r) \quad (\text{A.6})$$

Then, Eqn. (A.5) can be re-written as:

$$KRX(\mathbf{y}_r) = \hat{\mathbf{k}}_r^T \mathbf{\alpha} \mathbf{\Lambda}_\Phi^{-2} \mathbf{\alpha}^T \hat{\mathbf{k}}_r \quad (\text{A.7})$$

APPENDIX B.
RELATIONSHIP BETWEEN COVARIANCE AND KERNEL GRAM MATRIX IN
HIGH DIMENSIONAL SPACE

In this Appendix, the relationship between the eigenvectors of the centered kernel Gram matrix and the covariance matrix is obtained. Consider the set of centered background feature vectors:

$$\mathbf{Y}_C^\Phi \equiv \{\Phi(\mathbf{y}(1) - \mu_C^\Phi), \Phi(\mathbf{y}(2) - \mu_C^\Phi), \dots, \Phi(\mathbf{y}(N) - \mu_C^\Phi)\}$$

The covariance matrix is given as:

$$\hat{\mathbf{C}}_\Phi = \sum_{i=1}^N (\Phi(\mathbf{y}_i) - \mu_C^\Phi)(\Phi(\mathbf{y}_i) - \mu_C^\Phi)^T = \mathbf{Y}_C^\Phi \mathbf{Y}_C^{\Phi T} \quad (\text{B.1})$$

Let us assume that the covariance matrix $\hat{\mathbf{C}}_\Phi$ has an eigen-decomposition:

$$\hat{\mathbf{C}}_\Phi = \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi \mathbf{U}_\Phi^T \quad (\text{B.2})$$

where, $\mathbf{U}_\Phi = [\mathbf{u}_\Phi^1, \mathbf{u}_\Phi^2, \dots, \mathbf{u}_\Phi^{N_c}]$, is the matrix containing the eigenvectors along its columns, $\mathbf{\Lambda}_\Phi$ is a diagonal matrix and $\mathbf{U}_\Phi^T \mathbf{U}_\Phi = \mathbf{U}_\Phi \mathbf{U}_\Phi^T = \mathbf{I}$. From Eqn. (B.2),

$$\mathbf{U}_\Phi \mathbf{\Lambda}_\Phi = \hat{\mathbf{C}}_\Phi \mathbf{U}_\Phi = \mathbf{Y}_C^\Phi \mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \quad (\text{B.3})$$

Multiplying both sides by $\mathbf{Y}_C^{\Phi T}$:

$$\mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi = (\mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi) \mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \quad (\text{B.4})$$

Again multiplying both sides from the left by $\mathbf{\Lambda}_\Phi^{-1/2}$:

$$\begin{aligned} \mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi \mathbf{\Lambda}_\Phi^{-1/2} &= (\mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi) \mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi^{-1/2} \\ \Rightarrow (\mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi^{-1/2}) \mathbf{\Lambda}_\Phi &= (\mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi) (\mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi^{-1/2}) \end{aligned} \quad (\text{B.5})$$

Define

$$\mathbf{Y}_C^{\Phi T} \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi^{-1/2} = \mathbf{a} \quad (\text{B.6})$$

So that Eqn. (B.5) can be written as:

$$\mathbf{a} \mathbf{\Lambda}_\Phi = (\mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi) \mathbf{a} \quad (\text{B.7})$$

Note that the term $\mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi$ is the centered kernel Gram matrix, i.e. $\hat{\mathbf{K}} = \mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi$. Then, Eqn. (B.7) becomes:

$$\mathbf{a} \mathbf{\Lambda}_\Phi = \hat{\mathbf{K}} \mathbf{a} \quad (\text{B.8})$$

The columns of \mathbf{a} are a set of orthonormal vectors, i.e., $\mathbf{a}^T \mathbf{a} = \mathbf{a} \mathbf{a}^T = \mathbf{I}$. Eqn. (B.8) implies that \mathbf{a} are the eigenvectors of the centered kernel Gram matrix $\hat{\mathbf{K}}$. Multiplying both sides of Eqn. (B.6) by \mathbf{a}^T ,

$$\hat{\mathbf{K}} = \mathbf{a} \mathbf{\Lambda}_\phi \mathbf{a}^T \quad (\text{B.9})$$

Multiplying both sides of Eqn. (B.3) by $\mathbf{\Lambda}_\phi^{-1}$ and substituting Eqn. (B.6):

$$\mathbf{U}_\phi = \mathbf{Y}_C^\Phi \mathbf{a} \mathbf{\Lambda}_\phi^{-1/2} \quad (\text{B.10})$$

APPENDIX C.
CENTERING OF KERNEL GRAM MATRIX

In this Appendix, the relationship between the kernel Gram matrix \mathbf{K} and the centered kernel Gram matrix $\hat{\mathbf{K}}$ is derived. The centered weighted kernel Gram matrix $\hat{\mathbf{K}}$ is defined on centered featured vectors. Let the feature vector set be denoted as \mathbf{Y}^Φ .

Then the centered feature vector set \mathbf{Y}_C^Φ can be written as:

$$\mathbf{Y}_C^\Phi = \mathbf{Y}^\Phi - \boldsymbol{\mu}_C^\Phi \quad (\text{C.1})$$

where, $\boldsymbol{\mu}_C^\Phi$ is the mean feature vector, which can be written in terms of the feature vector set as:

$$\boldsymbol{\mu}_C^\Phi = \mathbf{Y}^\Phi \mathbf{1}_{N_C} \quad (\text{C.2})$$

and, $\mathbf{1}_{N_C}$ is a $N_C \times N_C$ matrix with each element equal to $1/N_C$.

Then the centered kernel Gram matrix can be written as:

$$\begin{aligned} \hat{\mathbf{K}} &= \mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi \\ &= (\mathbf{Y}^\Phi - \boldsymbol{\mu}_C^\Phi)^T (\mathbf{Y}^\Phi - \boldsymbol{\mu}_C^\Phi) \\ &= (\mathbf{Y}^\Phi - \mathbf{Y}^\Phi \mathbf{1}_{N_C})^T (\mathbf{Y}^\Phi - \mathbf{Y}^\Phi \mathbf{1}_{N_C}) \\ &= \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi - \mathbf{1}_{N_C}^T \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi - \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi \mathbf{1}_{N_C} + \mathbf{1}_{N_C}^T \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi \mathbf{1}_{N_C} \end{aligned} \quad (\text{C.3})$$

Since, $\mathbf{K} = \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi$ and $\mathbf{1}_{N_C}^T = \mathbf{1}_{N_C}$, it implies:

$$\hat{\mathbf{K}} = \mathbf{K} - \mathbf{1}_{N_C} \mathbf{K} - \mathbf{K} \mathbf{1}_{N_C} + \mathbf{1}_{N_C} \mathbf{K} \mathbf{1}_{N_C} \quad (\text{C.4})$$

APPENDIX D.
RELATIONSHIP BETWEEN THE KERNEL GRAM MATRIX AND THE
WEIGHTED CENTERED KERNEL GRAM MATRIX

In this Appendix, the relationship between the kernel Gram matrix \mathbf{K} and the weighted *centered* kernel Gram matrix $\hat{\mathbf{K}}_w$ is derived. The centered weighted kernel Gram matrix $\hat{\mathbf{K}}_w$ is defined on centered featured vectors. Let the feature vector set be denoted as \mathbf{Y}^Φ . Then the centered feature vector set \mathbf{Y}_C^Φ can be written as:

$$\mathbf{Y}_C^\Phi = \mathbf{Y}^\Phi - \boldsymbol{\mu}_C^\Phi \quad (\text{D.1})$$

where, $\boldsymbol{\mu}_C^\Phi$ is the matrix containing the mean feature vectors. The i^{th} column of the matrix $\boldsymbol{\mu}_C^\Phi$ is denoted by $\mu_C^\Phi(i)$ and is given as:

$$\mu_C^\Phi(i) = \frac{1}{\sum_{j=1}^N w_{ij}} \mathbf{Y}^\Phi \mathbf{w}_i \quad (\text{D.2})$$

where, $\mathbf{w}_i = [w_{i1} \ w_{i2} \ \dots w_{iN}]^T$. Thus, the matrix $\boldsymbol{\mu}_C^\Phi$

$$\boldsymbol{\mu}_C^\Phi = \mathbf{Y}^\Phi \mathbf{W} \boldsymbol{\Omega} \quad (\text{D.3})$$

Where \mathbf{W} is the weight matrix given as $[\mathbf{W}]_{ij} = w_{ij}$, and $\boldsymbol{\Omega}$ is diagonal matrix given as:

$$[\boldsymbol{\Omega}]_{ii} = \frac{1}{\sum_{j=1}^N w_{ij}} \text{ and } [\boldsymbol{\Omega}]_{ij} = 0 \quad \forall \quad i \neq j \quad (\text{D.4})$$

Then the weighted *centered* kernel Gram matrix can be written as:

$$\begin{aligned} \hat{\mathbf{K}} &= \mathbf{Y}_C^{\Phi T} \mathbf{Y}_C^\Phi \\ &= (\mathbf{Y}^\Phi - \boldsymbol{\mu}_C^\Phi)^T (\mathbf{Y}^\Phi - \boldsymbol{\mu}_C^\Phi) \\ &= (\mathbf{Y}^\Phi - \mathbf{Y}^\Phi \mathbf{W} \boldsymbol{\Omega})^T (\mathbf{Y}^\Phi - \mathbf{Y}^\Phi \mathbf{W} \boldsymbol{\Omega}) \\ &= \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi - \boldsymbol{\Omega}^T \mathbf{W}^T \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi - \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi \mathbf{W} \boldsymbol{\Omega} + \boldsymbol{\Omega}^T \mathbf{W}^T \mathbf{Y}^{\Phi T} \mathbf{W} \boldsymbol{\Omega} \end{aligned} \quad (\text{D.5})$$

Since, $\mathbf{K} = \mathbf{Y}^{\Phi T} \mathbf{Y}^\Phi$ and $\boldsymbol{\Omega}^T = \boldsymbol{\Omega}$, it implies:

$$\hat{\mathbf{K}} = \mathbf{K} - \boldsymbol{\Omega} \mathbf{W}^T \mathbf{K} - \mathbf{K} \mathbf{W} \boldsymbol{\Omega} + \boldsymbol{\Omega} \mathbf{W}^T \mathbf{K} \mathbf{W} \boldsymbol{\Omega} \quad (\text{D.6})$$

BIBLIOGRAPHY

- Agarwal S., Sriram P., Palit P. P. and Mitchell R. O. (2001), "Algorithms for IR-Imagery-Based Airborne Landmine and Minefield Detection," *Proceedings, SPIE- Detection and Remediation of Mine and Minelike Targets VI*, Vol. 4394, pp. 284-295.
- Amazeen C. A. and Locke M. C. (1996), "US Army's New Handheld Standoff Mine Detection System (HSTAMIDS)," *IEE International Conference on the Detection of Abandoned Landmines*, Vol. 431, pp. 172-176.
- Arbelaez P. and Cohen L. (2003), "Generalized Voronoi Tessellations for Vector-Valued Image Segmentation," *Proceedings 2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision (VLSM'03)*, Nice, France.
- Argenti F., Alparone L. and Benelli G. (1990), "Fast Algorithms for Texture Analysis Using Co-occurrence Matrices," *Proceedings of the IEEE*, Vol. 137, Part F, No. 6, pp. 443-448.
- Arigovindan M., Suhling M., Hunziker P. and Unser M. (2005), "Variational Image Reconstruction From Arbitrarily Spaced Samples: A Fast Multiresolution Spline Solution," *IEEE Transactions on Image Processing*, Vol.14, No. 4, pp. 450-460.
- Artyukh Y., Bilinskis I., Greitans A. and Vedin A. (1997), "Signal Digitizing and Recording in DASP-Lab System," *Proceedings of the International Workshop on Sampling Applications*, pp. 357-360.
- Arvis V., Debain C., Berducat M. and Benassi A. (2004), "Generalization of the Cooccurrence matrix for Colour Images: Application to Colour Texture Classification," *Image Analysis and Stereology*, Vol. 23, pp. 63-72.
- Bagchi S. and Mitra S. K. (1999), *"The Non-uniform Discrete Fourier Transform and Its Applications in Signal Processing,"* Kluwer Academic Publishers, Boston, Dordrecht, London.
- Beaven S., Stocker A. D. and Winter E. W. (2004), "Joint Multisensor Exploitation for Mine Detection," *Proceedings of SPIE*, Vol. 5415, April 2004.
- Bilinskis I. and Mikelsons M. (1992), *"Randomized Signal Processing,"* Prentice Hall, London, U.K.
- Bishop P. K., Perry K. M. and Poulter M. A. (1998), "Airborne Minefield Detection," *2nd International Conference on Detection of Abandoned Land Mines*, pp. 213-217.

- Bleau A. and Leon L. J. (2000), "Watershed-based segmentation and region merging," *Computer Vision and Image Understanding*, Vol. 77 , No. 3, pp. 317 – 370.
- Bosworth J., Koshimizu T., and Acton S. T., "Automated Segmentation of Surface Soil Moisture from Landsat TM Data," in *Proc. IEEE Southwest Symp. Image Analysis and Interpretation*, pp. 70–74.
- Bowman A. P., Winter E. M., Stocker A. D. and Lucey P. G. (1998), "Hyperspectral Infrared Techniques for Buried Landmine Detection," *Second International Conference on Detection of Abandoned Landmines*, pp. 129-133.
- Brown C. L., Zoubir A. M., Chant I., J. and Abeynayake C. (2002), "Landmine Detection Using Single Sensor Metal Detectors," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'02)*.
- Burt P. J., Hong T. H., and Rosenfeld A. (1981), "Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchical Computation," *IEEE Trans. Sys, Man Cyber*, 11(12):802–809.
- Caralotto M. J., Nebrich M. A. and Crary D. (2002), "Order of Battle Change Detection, Fusion, and Tracking," *Proceedings of the SPIE*, Vol. 4729.
- Caralotto M. J. (2005), "A Cluster-Based Approach for Detecting Man-Made Objects and Changes in Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, No. 2, pp. 374-387.
- Chang C. I. and Shao-Shan C. (2002), "Anomaly Detection and Classification for Hyperspectral Imagery," *IEEE Transactions on Geosciences Remote Sensing*, Vol. 43, No. 6, pp. 1314-1325.
- Chen P. C. and Pavlidis T. (1981), "Image Segmentation as an Estimation Problem," In A. Rosenfeld, editor, *Image Modelling*, Academic Press, pp. 9–28.
- Chen W. S. and Reed I. S. (1991), "A New CFAR Detection Test for Radar," *Digital Signal Processing*, Vol. 1, No. 4, pp. 198-214.
- Chen K. M. and Chen S. Y. (2002), "Color Texture Segmentation using Feature Distributions," *Pattern Recognition Letters*, Vol. 23, No. 7, pp. 755-771.
- Chen X., Moss R. H., Stoecker W. V., Umbaugh S. E., Stanley J. R., Celebi E. M., Shreshta B., Malters J. M., Grichnik J. M., Marghoob A. A., Rabinowitz H. S. and Menzies S. W. (2007), "Skin Lesion Segmentation Using by an Adaptive Watershed Flooding Approach," submitted to *IEEE Transactions on Medical Imaging*.

- Clark G. A., Sengupta S. K., Aimonetti W. D., Roeske F. and Donetti J. G. (2000), "Multispectral Image Feature Selection for Land Mine Detection," *IEEE Transactions on Geosciences Remote Sensing*, Vol. 38, No. 1, pp. 308-311.
- Cohen I., Tian Q., Zhou, X. S. and Huang T. S. (2002), "Feature Selection Using Principal Feature Analysis," *IEEE International Conference in Image Processing (ICIP'02)*, June 2002.
- Cressie N. A. C. (1991), "*Statistics for Spatial Data*," Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons Inc.
- Digabel H. and Lantuejoul C. (1978), "Iterative Algorithms," *Proceedings of the 2nd European Symposium on Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, Caen, France.
- Dirichlet P. G. L. (1850), "Über Die Reduction Der Positiven Quadratischen Formen Mit Drei Unbestimmten Ganzen Zalhen," *Journal Für Die Reine Und Angewandte Mathematik*, Vol 40, pp. 209-227.
- Drimbarean A. and Whelan P. F. (2001), "Experiments in Color Texture Analysis," *Pattern Recognition Letters*, Vol. 22, pp. 1161-1167.
- Donoho D. L. (2000), "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality," *Aide-Memoire of the Lecture in AMS Conference Math Challenges of 21st Century*. URL: <http://www-stat.stanford.edu/~donoho/Lectures>.
- Du Q., Faber V., and Gunzburger M. (1999), "Centroidal Voronoi Tessellations: Applications and algorithms," *SIAM Review* 41, 1999, pp. 637-676.
- Du Q., Gunzburger M., and Ju L. (2002), "Meshfree, Probabilistic Determination of Point Sets and Support Regions for Meshless Computing," *Comput. Meths. Appl. Mech. Engrg.* 191, 2002, pp. 1349-1366.
- Du Q., Gunzburger M., Ju L., and Wang X. (2006), "Centroidal Voronoi Tessellation Algorithms for Image Compression, Segmentation, and Multichannel Restoration," *J. Math. Imaging and Vision* 24, 2006, pp. 177-194.
- Dutt A. and Rokhlin V. (1995), "Fast Fourier Transforms for Non-equispaced Data II," *Journal of Computational Harmonic Analysis*, Vol. 2, pp. 85-100.
- Engel T. G. and Nunnaly W. C. (1998), "Airborne Ground Penetrating Radars for Mine Detection," *SPIE- Detection and Remediation of Mine and Minelike Targets III*, Vol. 3392, pp. 586-597.

- Feichtinger H. G. and Gröchenig K. (1992), "Irregular Sampling Theorems and Series Expansions of Band-Limited Functions," *Journal of Mathematical Analysis and Applications*, Vol. 167, No. 2, pp. 530–556.
- Feichtinger H. G. and Gröchenig K. (1992), "Iterative Reconstruction of Multivariate Band-Limited Functions from Irregular Sampling Values," *SIAM Journal of Mathematical Analysis*, Vol. 23, No. 1, pp. 244–261.
- Feichtinger H. G., Gröchenig K. and Strohmer, T. (1995), "Efficient Numerical Methods in Non-uniform Sampling Theory," *Numerische Mathematik*, Vol. 69, pp. 423–440.
- Fisher R. A. (1936), "Multiple Measures in Taxonomic Problems," *Annals of Eugenics*, Vol. 7, pp. 179–188.
- Filippidis A., Jain L. C. and Martin N. (2000), "Multisensor Data Fusion for Surface Landmine Detection," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, Vol. 30, No. 1, pp. 176–186.
- Garroway A. N., Buess M. L., Miller J. B., Suits B. H., Hibbs A. D., Barrall G. A., Matthews R. and Burnett L. J. (2001), "Remote Sensing by Nuclear Quadrupole Resonance," *IEEE Transactions on Geosciences Remote Sensing*, Vol. 39, No. 6, pp. 1108–1118.
- Gauch J. M. and Pizer S. M. (1993), "Multiresolution analysis of ridges and valleys in grey-scale images," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 15, No. 6, pp. 635–646.
- Gauch J. M. (1999), "Image Segmentation and Analysis Via Multiscale Gradient Watersheds," *IEEE Transactions on Image Processing*, Vol. 8, No. 1.
- Goodsill S. J. (1993), "*The Restoration of Degraded Audio Signals*," Ph.D Dissertation, Dept. of Engineering, University of Cambridge.
- Gray A. and Moore A. W. (2001), "N-body Problems in Statistical Learning," *Advances in Neural Information Processing Systems (NIPS 2001)*, Vancouver, BC.
- Greengard L. and Rokhlin V. (1987), "A Fast Algorithm for Particle Simulation," *Journal of Computational Physics*, Vol. 73, pp. 325–348.
- Greitans M. (1997), "Iterative Reconstruction of Lost Samples Using Updating of Autocorrelation Matrix," *Proceedings of the Workshop on Sampling Theory and Applications (SampTA'97)*, pp. 155–160.
- Gröchenig K. (1992), "Reconstruction Algorithms in Irregular Sampling," *Mathematical Computation*, Vol. 59, No. 199, pp. 181–194.

- Grosch T. O., Lee C. F. and Adams E. M. (1995), "Detection of Surface and Buried Mines with a UHF Airborne SAR," *SPIE- Detection Technologies for Mines and Minelike Targets*, Vol. 2496, pp. 110-120.
- Gumerov N. A. and Duraiswami R. (2004), "Fast Multipole Method Based Filtering of Non-uniformly Sampled Data," *Unpublished*, URL: http://www.umiacs.umd.edu/~ramani/pubs/fftinterpolation_icassp_2004.pdf.
- Haralick R. M., Shanmugam K. and Dinstein I. (1973), "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 6, pp. 610-621.
- Haralick R. M. (1979), "Statistical and Structural Approaches to Texture," *Proceedings of the IEEE*, Vol. 67, No. 5, pp. 786-804.
- Haris K., Efstratiadis S. N., Maglaveras N. and Katsaggelos A. K. (1998), "Hybrid image Segmentation Using Watersheds and fast Region Merging," *IEEE Transactions on Image Processing*, Vol. 7, No. 12, pp. 1684-1699.
- Harsanyi J. C., Farrand W. and Chang C. I. (1994), "Detection of Subpixel Spectral Signatures in Hyperspectral Image Sequences," *Proc. Amer. Soc. Photogram. Remote Sens.*, pp. 236-247, Reno, NV.
- Harsanyi J. C., and Chang C. I. (1994), "Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, No. 4, pp. 779-785.
- Haskett H. T. and Reago D. A. (2001), "Identification of Optimal Bands in the 3-5 μ m Region for Lightweight Airborne Mine Detection System (LAMDS)," *SPIE- Detection and Remediation of Mine and Minelike Targets VI*, Vol. 4394, pp. 246 - 257.
- Hausner A. (2001), "Simulating Decorative Mosaics," *Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 573-580.
- Hauta-Kasuri M., Parkkinen J., Jaaskelainen T. and Lenz R. (1996), "Generalized Cooccurrence Matrix for Multispectral Texture Analysis," *13th International Conference on Pattern Recognition*.
- Haykin S. (2001), "Adaptive Filter Theory," 4th Edition, Prentice Hall.
- Hernandez S. E. and Barner K. E. (2000), "Joint Region Merging Criteria for Watershed-Based Image Segmentation," *IEEE International Conference on Image Processing (ICIP 2000)*, pp. 108-111.

- Holmes Q.A. (1995), "Adaptive Multispectral CFAR Detection of Landmines," *Proceedings of the SPIE, Detection Technologies for Mines and Minelike Targets*, Vol. 2496, pp. 421 – 432.
- Hong T. H. and Rosenfeld A. (1984), "Compact Region Extraction Using Weighted Pixel Linking in a Pyramid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 2, pp. 222–229.
- Hong T. H. and Shneier. M. (1984), "Extracting Compact Objects Using Linked Pyramids," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 2, pp. 229–237.
- Horn B. K. P. (1986), "*Robot Vision*," MIT Press, Cambridge.
- Hwu R. J. and Woolard D. L. (2004), "Terahertz for Military and Security Applications II," *Proceedings of the SPIE*, Vol. 5411.
- Inaba M., Katoh N. and Imai H. (1994), "Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering: (extended abstract)," *Proceedings of the Tenth Annual Symposium on Computational Geometry*, 1994, pp. 332–339.
- Irvine G. B. and Cumming D. R. S. (2002), "Algorithms for the Nonuniform Acquisition and Subsequent Reconstruction of Data," In *Proceedings of the 14th International Conference on Digital Signal Processing*, pp. 791–794.
- Jackway P. T. (1996), "Gradient Watershed In Morphological Scale-Space," *IEEE Transactions on Image Processing*, Vol. 5, pp. 913–921, June 1996.
- Jain A. K., Murty M. N. and Flynn P. J. (1999), "Data Clustering: A Review," *ACM Computing Surveys*, Vol. 31, No. 3.
- Kailath T. (1967), "The Divergence and Bhattacharya Distance Measures in Signal Selection," *IEEE Trans. on Communication Technology*, Vol. 15, pp. 52-60.
- Kanungo T., Mount D., Netanyahu N., Piatko C., Silverman R., and Wu A. (2002), "An Efficient K-Means Clustering Algorithm: Analysis and Implementation," *IEEE Trans. Pattern Anal. Machl Intel.* 24, 2002, pp. 881–892.
- Kaskett H. T. and Broach J. T. (1999), "Automatic Mine Detection Algorithm Using Ground Penetrating Radar Signatures," *Proceedings of the SPIE- Detection and Remediation of Mine and Minelike Targets IV*, Vol. 3710, pp. 942-952.
- Keshava N. (2003), "A Survey of Spectral Unmixing Algorithm," *Lincoln Laboratory Journal*, Vol. 14, No. 1, pp. 55-78.

- Kwon H. and Nasrabadi N. M. (2005a), "Kernel RX-Algorithm: A Nonlinear Anomaly Detector for Hyperspectral Imagery," *IEEE Transactions on Geosciences Remote Sensing*, Vol. 43, pp. 388-397.
- Kwon H. and Nasrabadi N. M. (2005b), "Kernel Orthogonal Subspace Projection for Hyperspectral Signal Classification," *IEEE Transactions on Geosciences Remote Sensing*, Vol. 43, No. 12, pp. 2952 - 2962.
- Kwon H. and Nasrabadi N. M. (2006a), "Kernel Matched Subspace Detectors for Hyperspectral Target Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 178-194.
- Kwon H. and Nasrabadi N. M. (2006b), "Kernel Adaptive Subspace Detector for Hyperspectral Imagery," *IEEE Geosciences Remote Sensing Letters*, Vol. 3, No. 2, pp. 272 - 275.
- Kraut S. and Scharf L. L. (1999), "The CFAR Adaptive Subspace Detector is Scale Invariant GLRT," *IEEE Transactions on Signal Processing*, Vol. 47, No. 9, pp. 2538-2541.
- Kraut S., Scharf L. L. and M Whorter L. T. (2001), "Adaptive Subspace Detectors," *IEEE Transactions on Signal Processing*, Vol. 49, No. 1, pp. 1-16.
- Lee C. H. (1989), "Image Surface Approximation with Irregular Samples," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 2, pp. 206-212.
- Ling B., Dabbiru S., Trang A. H. and Phan C. (2006), "Surface and Buried Mine Detection Using MWIR Images," *Proceedings of the SPIE- Detection and Remediation of Mine and Minelike Targets XI*, Vol. 6217, pp. 62170F1-62170F3, April 2006.
- Lundberg M. (2001), "Infrared Landmine Detection by Parametric Modeling," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, pp. 3157-3160, May 2001.
- Maenpaa T., Pietikainen M. and Viertola J. (2002), "Separating Color and Pattern Information for Color Texture Discrimination," *16th International Conference on Pattern Recognition*, Vol. 1, pp. 668-671, August 2002.
- Makrogiannis S., Vanhamel I., Sahli H., Fotopoulos S. (2001), "Scale Space Segmentation of Color Images Using Watersheds and Fuzzy Region Merging," *International Conf. Image Processing*, Vol. 1, pp.7-10.

- Maksymonko G. and Breiter K. (1997), "ASTAMIDS Minefield Detection Performance at Aberdeen Proving Ground Test Site," *Proceedings of SPIE*, Vol. 3079, pp. 726-737.
- Manolakis D., Shaw G. and Keshava N. (2000), "Comparative Analysis of Hyperspectral Matched Filter Detectors," *Proceedings of SPIE*, Vol. 4049, pp. 2-17.
- Manolakis D., Siracusa C. and Shaw G. (2001), "Hyperspectral Subpixel Target Detection Using the Linear Mixing Model," *IEEE Transactions on Geosciences Remote Sensing*, Vol. 39 No. 7, pp. 1392-1409.
- Manolakis D., Marden D. and Shaw G. A. (2003), "Hyperspectral Image Processing for Automatic Target Recognition," *Lincoln Laboratory Journal*, Vol. 14, No. 1, pp. 79-115.
- Martin R. J. (1998), *"Irregularly Sampled Signals: Theories and Techniques for Analysis,"* Ph.D. Dissertation, University College, London, U.K.
- Masry E. (1978), "Alias-Free Sampling: An Alternative Conceptualization and its Applications," *IEEE Transactions on Information Theory*, Vol. IT-24, pp. 317-324.
- McFee J. E. and Ripley H. T. (1997), "Detection of Buried Landmines Using a *casi* Hyperspectral Imager," *Proceedings of the SPIE- Detection and Remediation of Mine and Minelike Targets II*, Vol. 3079, pp. 738-749, April 1997.
- Mednieks I. (1999), "Methods for Spectral Analysis of Non-uniformly Sampled Signals," *Proceedings of the Workshop on Sampling Theory and Applications (SampTA'99)*, pp. 190-193.
- Meyer F., Oliveras A., Salembier P., and Vachier C. (1997), "Morphological Tools for Segmentations: Connected Filters and Watershed," *Annals of Telecom.*, pp. 367-380.
- Ohanian P. P. and Dubes R. C. (1992), "Performance Evaluation for Four Classes of Texture Features," *Pattern Recognition*, Vol. 25, No. 8, pp. 819-833.
- Okabe A., Boots B., Sugihara K., and Chiu S. N. (1992), *"Spatial Tessellations: Concepts and Applications of Voronoi Diagrams."* John Wiley & Sons, 2nd Edition.
- Olsen O. F. and Nielsen M. (1997), "Multi-scale Gradient Magnitude Watershed Segmentation," in *Int. Conf. on Image Analysis and Processing (ICIAP'97)*, pp. 6-13.
- Olshausen B. A. and Field D. J. (1997), "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed By V1?," *Vision Research*, Vol. 37, pp. 3311-3325.

- Potts D., Steidl G. and Tasche M. (2001), "Fast Fourier Transforms for Non-equispaced Data: A Tutorial," In *Modern Sampling Theory: Math and Applications*, Eds. J. Benedetto and P. Ferreira, pp. 253 - 274, Birkhäuser.
- Poulter M. A., and Tee H. S. (2001), "Remote Minefield Detection System (REMIDS): A UK Programme for Airborne Minefield Detection," *IEEE International Geosciences and Remote Sensing Symposium (IGARSS)*, Vol. 1, pp. 243-245.
- Pratikakis I., Sahli H. and Cornelis J. (1991), "Hierarchical Segmentation Using Dynamics of Multiscale Gradient Watersheds," *11th Scandinavian Conf. Image Analysis (SCIA 99)*, pp. 577-584.
- Raykar V. C., Yang C., Duraiswami R. and Gumerov N. A. (2005), "*Fast Computation of Sums of Gaussians in High Dimensions*," *Technical Report*, CS-TR-4767, Dept. of Computer Science, University of Maryland-College Park.
- Raykar V. C. and Duraiswami R. (2005) "The Improved Fast Gauss Transform with Applications to Machine Learning," *Workshop on Large Scale Kernel Machines, Advances in Neural Information Processing Systems (NIPS 2005)*, Whister, BC.
- Reed I. S. and Yu X. (1990), "Adaptive Multi-band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 38, No. 10, pp. 1760 – 1770.
- Ren H., Chen C. W. and Chen H. T. (2005), "Weighted Anomaly Detection for Hyperspectral Remotely Sensed Images," *Proceedings of the SPIE, Chemical and Biological Standoff Detection III*, Volume 5995, pp. 63-68.
- Robey F. C., Fuhrmann D. R., Kelly E. J. and Nitzberg R. (1992), "A CFAR Adaptive Matched Filter Detector," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 1, pp. 208-216.
- Rouse J. W., Hass R. H., Schell J.A. and Deering D. W. (1973), "Monitoring Vegetation Systems in the Great Plains with ERTS," *Third ERTS Symposium*, NASA SP-351 I, pp. 309-317.
- SaySong G. and Ong I. G. H. (1995), "Reconstruction of Bandlimited Signals from Irregular Samples," *Signal Processing*, Vol. 46, No. 3, pp. 315–29.
- Salembier P. and Serra J. (1995), "Flat Zone Filtering, Connected Operators, and Filters by Reconstruction," *IEEE Trans. Image Processing*, Vol. 4, pp. 1153–1160.
- Sato M., Guangyou F. and Zhaofa Z. (2003), "Landmine Detection by a Broadband GPR System," *IEEE International Symposium on Geosciences Remote Sensing*, Vol. 2, pp. 758-760.

- Scharf L. L. and Friedlander B. (1994), "Matched Subspace Detectors," *IEEE Transactions on Signal Processing*, Vol. 42, No. 8, pp. 2146-2157.
- Scholkopf B., Smola A. J. and Muller K. R. (1999), "Kernel Principal Component Analysis," *Neural Computation*, Vol. 10, pp. 1299-1319.
- Schroeter, P. and Bigun J. (1995), "Hierarchical Image Segmentation by Multi-Dimensional Clustering and Orientation-Adaptive Boundary Refinement," *Pattern Recognition*, Vol. 28, No. 5, pp. 695-709.
- Schweizer S. M. and Moura J. M. F. (2001), "Efficient Detection in Hyperspectral Imagery," *IEEE Transactions on Image Processing*, Vol. 10, No. 4, pp. 584-597.
- Seip K. (1987), "An Irregular Sampling Theorem for Functions Bandlimited in a Generalized Sense," *SIAM Journal of Applied Mathematics*, Vol. 47, No. 5, pp. 1112-1116.
- Shannon C. E. (1948), "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, pp. 379-423 and 623-656.
- Shapiro H. S. and Silverman R. A. (1968), "Alias-free Sampling of Random Noise," *SIAM Journal of Applied Mathematics*, Vol. 8, pp. 225-236.
- Simrad J. R. and Mathieu P. (1998), "Airborne Far IR Minefield Imaging System(AFIRMIS): Description and Preliminary Results," *SPIE- Detection and Remediation of Mine and Minelike Targets III*, Vol. 3392, pp. 84-95.
- Sower G. D., Kilgore R. and Eberly J. (2001), "GSTAMIDS Ulseid Magnetic Induction Metal Detector: Hardware Description and Data Processing Algorithms," *SPIE- Detection and Remediation of Mine and Minelike Targets VI*, Vol. 4394, pp. 1-7.
- Spann M. and Wilson R. G. (1985), "A Quad-Tree Approach to Image Segmentation Which Combines Statistical and Spatial Information," *Pattern Recognition*, 18(3/4):257- 269.
- Spann M. and Horne. C. (1989), "Image Segmentation Using a Dynamic Thresholding Pyramid," *Pattern Recognition*, 22(6):729-732.
- Stein D., Beaven S., Hoff L., Winter E., Schaum A. and Stocker A. (2002), "Anomaly Detection from Hyperspectral Imagery," *IEEE Signal Processing Magazine*, No. 1, pp. 58-69.
- Strohmer T. (1995), "On Discrete Band-Limited Signal Extrapolation," *Contemporary Mathematics*, Vol. 190, pp. 323-337.

- Strohmer T. and Tanner J. (2006), "Fast Reconstruction Methods for Bandlimited Functions from Periodic Nonuniform Sampling," *SIAM Journal on Numerical Analysis*, Vol.44, No. 3, pp. 1073-1094.
- Summers B. (1998), "*Extensions to the Sampling Theorem and Processing Techniques for Nonequispaced Sampled-Data Systems*," Ph.D Dissertation, University of Westminster.
- Sun Y., Li X. and Li J. (2005), "A Practical Landmine Detector Using Forward- Looking Ground Penetrating Radar," *IEEE Electronics Letters*, Vol. 41, No. 2, pp. 97-98.
- Suzukawa Jr. H. H. and Farber M. S. (1995), "Long-Range Airborne Detection of Small Floating Objects," *SPIE- Detection Technologies for Mines and Minelike Targets*, Vol. 2496, pp. 193-205.
- Tan Y., Tantum S. L. and Collins L. M. (2002), "Landmine Detection with Nuclear Quadrupole Resonance," *IEEE International Symposium on Geosciences Remote Sensing*, Vol. 3, pp. 1575 – 1578.
- Tarczynski A. (2002), "Spectrum Estimation of Non-uniformly Sampled Signals," *Proceedings of the 14th International Conference on Digital Signal Processing*, pp. 795-798.
- Tarczynski A. and Allay N. (2004), "Spectral Analysis of Randomly Sampled Signals: Suppression of Aliasing and Sampling Jitter," *IEEE Transactions on Signal Processing*, Vol. 52, No. 12, pp. 3324-3334.
- Thai B. and Healey G. (2002), "Invariant Subpixel Material Detection in Hyperspectral Imagery," *IEEE Transactions on Geosciences Remote Sensing*, Vol. 40, No. 3, pp. 599-608.
- Tiwari S., Agarwal S. and Trang A. (2007), "Buried Mine Detection in Airborne Imagery Using Co-occurrence Texture Features," *Proceedings of the SPIE- Detection and Remediation of Mine and Minelike Targets XII*.
- VanTrees H. L. (1968), "*Detection, Estimation, and Modulation Theory. Part I: Detection, Estimation, and Linear Modulation Theory*," John Wiley and Sons Inc.
- Vazquez C., Dubois E. and Konrad J. (2005), "Reconstruction of Nonuniformly Sampled Images in Spline Spaces," *IEEE Transactions on Image Processing*, Vol. 14, No. 6, pp. 713-725.
- Vincent L. and Soille P. (1991), "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 6, pp. 583-398.

- Voronoi G. M. (1908), "Nouvelles Applications Des Parametres Continus a La Theorie Des Formes Quadratiques. Deuxieme Memoire: Recherches Sur Les Paralleloedres Primitifs," *Journal Für Die Reine Und Angewandte Mathematik*, Vol 134, pp. 198-287.
- Wang D. (1997), "A Multiscale Gradient Algorithm for Image Segmentation Using Watersheds," *Pattern Recognition*, Vol. 30, No. 12, pp. 2043-2052.
- Williams D. R. and Coletta N. J. (1987), "Cone Spacing and the Visual Resolution Limit," *Journal of the Optical Society of America*, Vol. 4, pp. 1514-1523.
- Wilson R. and Spann. M. (1988), "Finite Prolate Spheroidal Sequence and Their Applications I, II: Image Feature Description and Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, pp.193-203.
- Witherspoon N.H., Holloway Jr. J. H., Davis K. S., Miller R. W. and Dubey A. C. (1995), "The Coastal Battlefield Reconnaissance and Analysis (COBRA) Program for Minefield Detection," *Proceedings, SPIE- Detection Technologies for Mines and Minelike Targets*, Vol. 2496, pp. 500-508, 1995.
- Witten T. T. (1998), "Present State of The Art in Ground-Penetrating Radars for Mine Detection," *Proceedings of the SPIE- Detection and Remediation of Mine and Minelike Targets III*, Vol. 3392, pp. 576-586.
- Wojtiuk J. J. (2000), "*Randomized Sampling for Radio Design*," Ph.D Dissertation, University of South Australia, Mawson Lakes, Australia.
- Wright A. S. and Acton S. T. (1997), "Watershed pyramids for edge detection," in *Proc. Int. Conf. on Image Processing*, Vol. II, Washington, DC.
- Xiang N. and Sabatier J. M. (2001), "Anti-Personnel Mine Detection Using Acoustic-to-Seismic Coupling," *Proceedings of the SPIE- Detection and Remediation of Mine and Minelike Targets VI*, Vol. 4394, pp. 535-541.
- Xu R. and Wunsch D. (2005), "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, Vol. 16, No. 3.
- Yellott Jr. J. I. (1982), "Spectral Analysis of Spatial Sampling by Photoreceptors: Topological Disorder Prevents Aliasing," *Vision Research*, Vol. 22, pp. 1205-1210.
- Yu X., Reed I. and Stocker A. D. (1993), "Comparative Performance Analysis of Multispectral Anomaly Detectors," *IEEE Transactions on Image Processing*, Vol. 41, No. 8, pp. 2639-2658.

Zhao H., Shah S., Choi J. H., Nair D. and Aggarwal J. K. (1998), "Robust Automatic Target Detection/Recognition System for Second Generation FLIR Imagery," *Fourth IEEE Workshop on the Applications of Computer Vision (WACV'98)*, pp. 262-263.

VITA

Spandan Tiwari was born on 11th May 1980 Kanpur, India. He received his primary, secondary, and high school education from various “Kendriya Vidyalaya” (Central School) schools across India. In 1997, he joined the National Institute of Technology, Bhopal (formerly known M.A.C.T, Regional Engineering College, Bhopal) in Department of Electronics and Communication Engineering. He graduated with the degree of Bachelor of Engineering in 2001, standing third in class of 70 students.

In Fall 2001, he joined the Department of Electrical and Computer Engineering at University of Missouri-Rolla and worked in the capacity of Graduate Research Assistant at Intelligent Systems Center, UMR. Spandan completed his research and obtained his Masters degree from UMR in summer 2003.

Spandan joined the PhD program at the Department of Electrical and Computer Engineering at University of Missouri-Rolla in January 2004. He was employed in the capacity of Graduate Research Assistant at the Airborne Reconnaissance and Image Analysis (ARIA) Lab at UMR. He also worked in the capacity of instructor and teaching assistant for several courses in the department from June 2004 to May 2007.

Spandan received his PhD from UMR in December, 2007.