Night Vision and Electronic Sensors Directorate

AMSRD-CER-NV-TR-C258

Expectation Maximization and Its Application In Modeling, Segmentation and Anomaly Detection

Approved for Public Release; Distribution Unlimited

Fort Belvoir, Virginia 22060-5806

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2008	Thesis - 2006

4. TITLE AND SUBTITLE
Expectation Maximization and its Application In Modeling, Segmentation and Anomaly Detection

5. FUNDING NUMBERS
DAAB07-01-D-G601; 0073

6. AUTHOR(S)
Ritesh Ganju

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF MISSOURI - ROLLA
Rolla, Missouri

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate
10221 Burbeck Road
Fort Belvoir, Virginia 22060

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
AMSRD-CER-NV-TR-C258

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*

Expectation Maximization (EM) is a general purpose algorithm for solving maximum likelihood estimation problems in a wide variety of situations best described as incomplete data problems. The incompleteness of the data may arise due to missing data, truncated distributions, etc. One such case is a mixture model, where the class association of the data is unknown. In these models, the EM algorithm is used to estimate the parameters of parametric mixture distributions along with their probabilities of occurrence. In this thesis, the EM algorithm is employed to estimate different mixture models for raw single and multi-band electro-optical Infra Red (IR) data. The EM update equations for single and multi-band Gaussian and single-band Gamma and Beta mixture models are discussed. Gaussian mixture models are used for the raw image segmentation of single and multi-band imagery. Three different anomaly detection techniques based on EM-based image segmentation are discussed and evaluated. The Gamma and Beta mixture models are used to model the detection statistic of two different anomaly detectors. An adaptive CFAR (Constant False Alarm Rate) threshold selection based on the mixture model of the detection statistic has been implemented to determine potential target locations. These mixture models of detection statistics can also be used for multi-sensor or multi-algorithm fusion. The algorithms have been evaluated using single-band mid-wave IR airborne imagery for mine and mine field detection problems.

14. SUBJECT TERMS
Gamma and Beta mixture models, EM algorithm, Constant False Alarm Rate (CFAR), sensor fusion, echolocation systems, minefield detection, anomaly detector, STOCHASTIC EM, Image segmentation, SEM-based anomaly detectors, Locally Optimum Bayes Detection

15. NUMBER OF PAGES
118

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	None

Night Vision and Electronic Sensors Directorate

AMSRD-CER-NV-TR-C258

Expectation Maximization and Its Application In Modeling, Segmentation and Anomaly Detection

A Thesis Presented By

Ritesh Ganju

University of Missouri – Rolla

May 2008

Approved for Public Release; Distribution Unlimited

Countermine Division
FORT BELVOIR, VIRGINIA 22060-5806

EXPECTATION MAXIMIZATION AND ITS APPLICATION IN MODELING, SEGMENTATION AND ANOMALY DETECTION

by

RITESH GANJU

A THESIS

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI – ROLLA

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

2006

Approved by

_____	_____
Dr. S. Agarwal, Advisor	Dr. R. J. Stanley

_____

Dr. V. Samaranayake

ABSTRACT

Expectation Maximization (EM) is a general purpose algorithm for solving maximum likelihood estimation problems in a wide variety of situations best described as incomplete data problems. The incompleteness of the data may arise due to missing data, truncated distributions, etc. One such case is a mixture model, where the class association of the data is unknown. In these models, the EM algorithm is used to estimate the parameters of parametric mixture distributions along with their probabilities of occurrence. In this thesis, the EM algorithm is employed to estimate different mixture models for raw single and multi-band electro-optical Infra Red (IR) data. The EM update equations for single and multi-band Gaussian and single-band Gamma and Beta mixture models are discussed. Gaussian mixture models are used for the raw image segmentation of single and multi-band imagery. Three different anomaly detection techniques based on EM-based image segmentation are discussed and evaluated. The Gamma and Beta mixture models are used to model the detection statistic of two different anomaly detectors. An adaptive CFAR (Constant False Alarm Rate) threshold selection based on the mixture model of the detection statistic has been implemented to determine potential target locations. These mixture models of detection statistics can also be used for multi-sensor or multi-algorithm fusion. The algorithms have been evaluated using single-band mid-wave IR airborne imagery for mine and mine field detection problems.

ACKNOWLEDGEMENTS

I am extremely grateful to my advisor, Dr. Sanjeev Agarwal, for his constant encouragement and support throughout my master's program. Without his consistent support, this work would not have been possible. I would also like to thank Dr. Stanley and Dr. Samaranayake for being a part of my committee.

In addition, I would like to thank all the members of my research group at UMR's ARIA Labs, especially Deepak Menon for his help. I would also like to thank the Countermine Division of Night Vision and Electronic Sensors Directorate (NVESD) for providing the data and the opportunity to work on this project. I would like to thank Kathryn Mudd, at UMR's writing center, for her suggestions and editing the thesis.

Last, but definitely not the least, I would like to express eternal gratitude towards my family for being there for me whenever it mattered the most.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

Table Page

1. INTRODUCTION

1.1. THE EM ALGORITHM

The EM algorithm is a technique for maximum likelihood estimation in situations best described as incomplete data problems [1]. It is so called because of its two important steps—Expectation (E step) and Maximization (M step). The EM algorithm seeks to iteratively compute the maximum likelihood estimates and it is very useful in situations where algorithms such as Newton-Raphson, Prediction-Error, Sliding Window and Least-Squares turn out to be tedious and time consuming. EM has specifically gained importance because in certain incomplete data situations, the maximum likelihood estimation can be difficult due to the absence of the data. If the same problem is converted to a complete data problem with additional unknown parameters, then the problem can be solved more easily using EM iterations.

Although these incomplete data problems can arise in different situations, this thesis will study the incomplete data problems as applied to mixture models. In background modeling, the background data can be characterized as coming from a set of different probability distributions. This problem is an incomplete data problem in the sense that the class wise association of the data is unknown. Also the proportions of different classes are not known. Thus in these situations the EM algorithm can be applied to distribute the data into classes and to find the class proportions and parameters of distribution in a parametric mixture model. The EM algorithm and its concepts are discussed in detail in Section 2 of this thesis.

1.2. BACKGROUND MODELING

Background modeling is an efficient way to characterize the background data with certain probability distributions. These distributions are in the form of mixture models. Because the data sometimes is of high dimensionality, non-parametric methods of density estimation, such as kernel-based methods, would require large amounts of training data. This makes it important for us to study modeling using parameter estimation. Also, in

certain situations parametric modeling makes the analysis more robust. This is because if the system has been modeled using parametric distribution, then any unpredictability can be accounted for by adjusting the parameters of the model based on knowledge from past modeling experiences [6].

In many problems such as minefield detection, target recognition and echolocation systems, background characterization is required for a proper interpretation and analysis of the data. For example the EM-based anomaly detectors use the mixture model framework, where the concept of pixel membership is used to detect anomalies in the data. Modeling of detection statistic can also be performed. The detection statistic represents the non-homogeneities and spatial correlation in the data, and therefore its modeling into parametric distributions is very important. Modeling of detection statistic also helps in performing adaptive Constant False Alarm Rate (CFAR) threshold selection, which is important for practical detection systems and also for sensor fusion [7]. Modeling of the detection statistic is discussed in Section 5 of this thesis.

Background modeling is useful in image segmentation where the image is segmented into various regions. Background modeling is also used in anomaly detection. Here, the concept of data membership is exploited to separate the anomalies that are statistically different from the background. These concepts are discussed in Sections 4, 5 and 6 of this thesis.

1.3. APPLICATIONS OF BACKGROUND MODELING

1.3.1. Echolocation Systems. In the echolocation systems, radio and sound waves are transmitted, and from the echo of the reflected waves, it is possible to get information such as the location, direction and size of the target. In these systems, it is sometimes desirable and needed to estimate the statistical behavior of the target and the environment in which it operates. There are three main motivations for this [6]:

(i) An appropriate setting of the detection threshold of an echolocation system is required to control the false alarm rate. The reason for this is that in many

active SONAR and RADAR systems, the decision to declare the presence of a target depends on the amplitude of the matched filter output exceeding a certain threshold. The estimate of the background probability density function of the matched filter amplitude is required if the threshold is to be determined by parametric signal processing techniques [7].

(ii) It is advantageous if a system detection performance can be predicted by means of measurement of the historical data. For example, if the background is characterized using parametric distributions, then in the case of unfavorable circumstances, the parameters (such as frequency and wavelength) can be adjusted to combat those unfavorable circumstances.

(iii) Knowledge of the background statistics may be used to improve the design of the detector in an optimal sense.

1.3.2. Automatic Target Recognition (ATR) Systems. ATR systems generally have a multistage architecture for target detection or recognition. The first stage is generally a pre-screener. It selects the targets at a given CFAR that are passed on to the next stage, which is the discrimination stage. The discrimination stage is basically a false alarm mitigation scheme that rejects the false alarms based on certain features [34], [35]. Finally, the classification stage is used to detect the targets. The need for modeling comes in the pre-screener stage where an efficient modeling can be used for adaptive CFAR threshold selection. This helps in the selection of the optimum number of targets that are then passed to the next stage.

Also in these systems, understanding of the battlefield would be greatly enhanced if the surveillance systems used could also provide automated, reliable classification of objects in the areas surveyed. The volume of the data produced by surveillance makes it infeasible for the human interpretation of the data. The RADAR Range Profiles (RRPs) data are used in the ATR systems. Since the data are of a high dimensionality, non-parametric methods of density estimation, such as kernel-based methods, would require a

large amount of training data. Approximating the probability density by a Gaussian distribution alone may make the analysis rigid.

In these cases, mixture models provide a parametric method that is flexible for the modeling and analysis, as this would help in modeling a wide range of possible densities, including multi-modal densities. In case of the RADAR data, there is a strong motivation for the mixture model to unscramble returns from multiple targets [6]. In order to form the mixture models using parametric distributions, an efficient algorithm is needed. The EM algorithm is one such algorithm that efficiently models the background data with the probability models.

1.4. MINEFIELD DETECTION

In minefield detection, an efficient classification and characterization of the background is to be done. The aim is to separate the mines (anomalies) statistically from the background. Following are the motivations for background analysis:

(i) The detection performance of any system depends on the background characteristics and terrain. A basic knowledge of an anomaly detector such as an RX algorithm shows that the performance over different types of backgrounds depends on the correlation and non-homogeneities in the background. Thus in order to quantify the detector performance, it is necessary to model the detection statistics of the anomaly detector into probabilistic models that would help improve the performance of the anomaly detector.

(ii) Modeling helps to statistically differentiate the background and therefore helps in forming guidelines that would help in studying the detectability and likelihood of mines in different types of backgrounds and terrains.

(iii) Modeling also helps in extensive analysis of threshold determination for a given CFAR. This threshold is used to separate the background from the targets.

Automatic/Aided Target Recognition (ATR/AiTR) systems are often airborne in nature. Collection of data from a large area mandates the development of these systems. These systems have a pre-screening stage that is required to detect potential target locations. Background modeling of the detection statistic is very useful in the pre-screening stage, where an adaptive CFAR threshold selection is performed to select a threshold based on detection statistics. The detection statistic represents non-homogeneities and spatial correlation of the data. Therefore, the adaptive threshold selection based on detection statistic helps in selecting a threshold that is invariant to the background the detector is operating in.

Thus, it is evident that minefield detection forms an important application of background modeling. This application is explored in greater details in this thesis.

1.5. OVERVIEW OF THE THESIS

In Section 2, the EM algorithm is introduced. The concept and mechanism of the EM algorithm is studied in detail as the EM algorithm is the backbone of the mixture model architecture. The development and the mathematical formulation of the algorithm is presented.

In Section 3, the concepts of the Image Segmentation using the EM algorithm are presented, where the image is segmented into regions that are statistically different. The region belonging to a particular class appears as a separate segment in the image.

Section 4 covers the various anomaly detectors that have been implemented based on mixture modeling. The mechanism and concept of the different type of EM-based anomaly detectors is presented.

Section 5 introduces various mixture models that have been implemented to model the detection statistic. In order to test the modeling performance of these mixture models, the Chi-Square test is described that evaluates the modeling performance of a given distribution. The two parameter Beta and Gamma mixture models are tested on two different statistics for extensive airborne data. The performance of the mixture models is then compared.

Section 6 discusses the automatic CFAR threshold selection from the modeling results. These thresholds are calculated for a given CFAR value. These thresholds that are determined from the modeling results are used to determine the targets that are candidates for further processing.

Appendix—A lists various mixture models implemented along with their distributions and the update equations. It also shows the use of the update equations to estimate the parameters of the mixture models.

Appendix—B discusses different tests that evaluate the goodness of fit of the mixture models implemented.

2. THE EM ALGORITHM

2.1. EM—AN INTRODUCTION

The EM algorithm is a general purpose algorithm for maximum likelihood estimation in a wide variety of situations best described as incomplete data problems. The incomplete data problems arise, for example, where there are missing data, truncated distributions, censored or grouped distributions and also in situations where the missing data are not evident. One such case is a mixture model, where the class association of the data is unknown. The data is assumed to belong to a parametric mixture model but the proportion of each class is unknown.

It is to be noted that some problems at first sight may not seem to be incomplete data problems but they actually are. Direct solution of these problems can be tedious and unstable. There can be a great reduction in computation if the problem could be converted to a complete data problem because in these situations, the complexity of the Maximum Likelihood (ML) estimation reduces if the complete data is provided because the log-likelihood or the cost function for complete data problem is often of a nice and tractable form. However, in some cases the ML equations do not have explicit solutions, and therefore one has to resort to some iterative methods to arrive at the solution. EM is one such efficient iterative technique. The conversion of incomplete data problem into a complete data problem is discussed more in Section 2.5.

2.2. MOTIVATION FOR EM

Many attempts have been made to estimate the parameters using the training data. In literature, this has been referred to as a *supervised* approach. One major drawback of this type of method is that it is unrealistic because many times the training data with reliable class association is unavailable. Therefore, attempts have been made to estimate the parameters using some *unsupervised* technique, i.e. the one that does not require the training data. EM is one such technique. Statisticians have used EM to estimate

parameters for the incomplete data problem because EM works very well for this type of practical problems. In recent times, there has been considerable interest in stochastic model-based image segmentation. Here the image is separated into a set of disjointed regions, and each region is associated with one of a finite number of classes. Each class is assumed to have been modeled as a random field. Because these random fields are often parametric models, an important problem that one is faced with is regarding parameter estimation. Clearly, the parameter estimation problem here is an incomplete data problem, because the observed image is a mixture of several data classes with the class status of each pixel unknown, which means that the correct segmentation is not known.

2.3. EM—A BRIEF HISTORY

The name ‘EM’ was coined by Dempster, Laird and Rubin in a paper in 1976 and was published in the *Journal of Royal Statistical Society* in 1977 [48]. Because the idea behind the EM algorithm is very general, algorithms like it were formulated and applied in a variety of problems even before the paper was presented. However, it was in the paper presented by Dempster, Laird and Rubin that various ideas were synthesized and a general theory was developed. The various references to literature on an EM-type of algorithm can be found in [1].

2.4. EM—THE CONCEPT

The EM algorithm estimates the parameters of the mixture model iteratively, starting from some initial guess. Each iteration consists of the following two steps:

Step 1: (Expectation): This step tries to find the distribution of the complete data given the known values of the observed data and a current estimate of the parameters. The estimation step basically involves the formulation of a ‘*Q*’ function (to be described later), which is basically the estimation of the likelihood function of the complete data given the observed data and the current fit of parameters (i.e. the parameters obtained in the maximization step of the previous iteration).

Step 2: (Maximization): This step re-estimates the parameters to be those with maximum likelihood under the assumption that the distribution found in the estimation step is correct. The maximization step is so called because it maximizes the '*Q*' function formulated in the estimation step to obtain a new set (fit) of parameters.

Each step is carried out in the above order until the terminating condition is reached. The terminating condition can be assumed to have been reached when the log-likelihood function of the complete data does not show significant improvement over its previous value. It can be shown that each successive iteration either improves the true likelihood or leaves it unchanged (if a local maximum has already been reached) [1].

2.5. MATHEMATICAL FORMULATION OF EM

Let us suppose that for any practical situation:

x = Complete Data (observed plus unobserved),

y = Observed (incomplete) data,

z = Additional data which is missing (or is unobservable).

Also $f(y,\varphi)$ is the probability density function (pdf) of the observed incomplete data, where 'φ' is the set of parameters that characterize the pdf.

The problem is to estimate 'φ' based on the incomplete information represented by the observed data, 'y.' The likelihood function, $£(\varphi \mid y)$, for the parameter, 'φ' given 'y' can be formed as:

$$£(\varphi \mid y) = f(y,\varphi). \tag{2.1}$$

Log-likelihood function, $\ln[£(\varphi \mid y)]$ can be formed where '*ln*' is the natural logarithm.

$$\ln[£(\varphi \mid y)] = \ln[f(y,\varphi)]. \tag{2.2}$$

A log-likelihood function is considered because it makes the analysis and calculations easier without changing the optimization problem at all.

The maximum likelihood estimation problem is complicated by the fact that only the incomplete data is at hand. Thus in order to ease the problem, the *incomplete* log-likelihood function is converted to the *complete* log-likelihood function by adding the missing information, 'z.' The log-likelihood of the complete data, $\{y, z\} = \{x\}$, is defined as:

$$\ln[£_c(\varphi \mid y, z)] = \ln[£_c(\varphi \mid x)] = \ln[f(x \mid \varphi)] \,, \tag{2.3}$$

where $£_c(\varphi|x)$ is the likelihood function of the *complete* data.

Therefore the EM algorithm approaches the problem of solving the incomplete data likelihood function indirectly by proceeding iteratively in terms of the complete data log-likelihood function, $£_c(\varphi \mid x)$. As the complete data is unobservable, it is replaced by its conditional expectation given the observed data, and the current fit of parameters, i.e

$$\ln[f(x \mid \varphi)] = \ln[f(y \mid \varphi).f(z \mid y, \varphi)]. \tag{2.4}$$

This step is discussed in more detail in specific context of Gaussian mixture model for image segmentation in section 3.4.

Starting from some initial guess of the parameters, 'φ^0,' the EM algorithm iterates between the two steps, known as the *'E'* and *'M'* steps. These are described as follows:

<u>*E (Estimation) Step:*</u> This step forms the 'Q' function by estimating the log-likelihood of the complete data given the observed data and current fit of parameters. The current fit of parameters is the set of parameters that is obtained from the maximization step of the previous iteration.

The 'Q' function is formulated as:

$$Q\left(\varphi,\varphi^{k}\right)=E_{\varphi^{k}}\left[\ln\{\pounds_{c}\left(\varphi|x\right)\}\mid y,\varphi^{k}\right], \tag{2.5}$$

where 'φ^{k}' represents the current set of parameters and '$E_{\varphi^{k}}$' represents the expectation of the log-likelihood of the complete function over 'φ^{k}.'

M (Maximization) Step: This step maximizes the 'Q' function in Equation (2.5), to get a new set of parameters, so that,

$$E\left[\frac{\partial\left[Q\left(\varphi,\varphi^{k}\right)\right]}{\partial\varphi}\right]=0. \tag{2.6}$$

The new value of the parameter 'φ' is given by:

$$\varphi^{k+1}=\varphi^{k}+\left(I_{m}\right)^{-1}E\left[\frac{\partial\left[Q\left(\varphi,\varphi^{k}\right)\right]}{\partial\varphi}\right], \tag{2.7}$$

where 'I_{m}' is the information matrix, calculated at 'φ^{k}.' Information matrix for different mixture models is discussed in more detail in appendix-A.

Once the parameters, 'φ^{k},' are obtained in the k^{th} iteration, the 'E' step is carried out again, and the 'Q' function is updated with these new parameters to form the new estimate of the log- likelihood of the complete data. The 'M' step is then carried again for the new updated 'Q' function to yield a more updated set of parameters, 'φ^{k+1},' and the process is carried on until the difference, $\pounds_{c}\left(\varphi^{k+1}\mid x\right)-\pounds_{c}\left(\varphi^{k}\mid x\right)$, changes by an arbitrarily small amount. At this point, the most likely set of parameters is reached. This condition is often used as the terminating condition.

2.6. A VARIANT OF EM—STOCHASTIC EM

One major weakness of EM is its vulnerability to the initial values, 'φ^0.' If these are "far" from the actual values, then there may be cases when the algorithm does not converge to the actual value. This happens because the EM algorithm sometimes gets trapped into local maxima, (if it finds one) before the actual global maxima. These are known as saddle points. If this happens, then the parameter values might get stuck to these saddle points, and therefore the parameters may be incorrectly estimated.

The weakness is partially removed with SEM that resembles EM. In case of standard EM, the indicator variables are obtained using maximum likelihood and the class probabilities, 'φ^k.' In case of SEM the value of these indicator variables is obtained based on a draw using the current class probabilities, 'φ^k.' This assigns each sample to one of the classes probabilistically in proportion to the class probabilities, 'φ^k.' Because of the random nature of SEM, it does not remain confined to a local maxima and instead is more likely to progress towards global maxima.

In image modeling, another form of EM, known as SEM-EM, is often used. This is so called because SEM is used to run for the initial major part of the iterations (say 75%). After this initial "warm up", EM takes over. It is assumed that after using SEM for 75% of the iterations, the results are close to the actual values. The EM then runs until the end, converging to global maxima. The set of parameter values obtained in the last certain number of iterations (for example 100) are averaged to give the values of final converged parameters.

2.7. CONVERGENCE PROPERTIES OF EM

As seen in the previous section, it is possible for the parameters to converge to a saddle point rather than a global point. This depends a lot on the type of log-likelihood function. If the log-likelihood function is uni-modal, then the convergence of the likelihood function and the parameters is unique. If the likelihood function is not uni-

modal, then in that case the likelihood function and the parameters might converge to some saddle point.

The situation can be analyzed by using the analogy of a round bowl. The likelihood surface can be thought to be a bowl. The bowl can have a steep or flat bottom. If the bowl has a steep bottom, then any round ball that is put at a certain position inside the bowl finally reaches the steep bottom after certain time and remains there. However, if the bottom is flat, then the ball might reach the flat bottom and circle around the flat bottom surface rather than reaching a particular point at the bottom.

Similarly, it has been observed in some cases that if the number of parameters in a distribution is large, then the parameters tend to undergo periodic oscillations after a certain number of iterations. This happens because the likelihood surface tends to become flat as the number of parameters in a distribution is large. Thus rather than converging to a steady value, the parameters of such type of distribution tend to converge to a range. When this happens, EM is said to converge to a circle rather than a single point [1].

2.8. SELECTION OF NUMBER OF CLASSES

Choosing the number of classes is an important issue in the EM-based applications. However no fully general and satisfactory solution seems available. The most common approach of selecting the number of classes is based on the log-likelihood of the samples given the number of classes [41]. Let 'x' be the data, 'g' be the number of classes and '$\hat{\varphi}$' be the parameters of the mixture model that are estimated from the data. Then a criteria based on log-likelihood has been given in [42] using the Bayesian Information Criteria (BIC) approximation. This is given by:

$$BIC = 2\log p\left(x \mid g, \hat{\varphi}\right) - N_p \log(n) \tag{2.8}$$

Here, N_p is the number of parameters estimated from the data, and n is the number of samples.

The larger the value of *BIC*, the better the model is according to this criterion. However in order to incorporate this criterion, one has to run the algorithm for certain number of times (each time with a different number of class) and then has to select the number of classes that gives the maximum value of *BIC*. This is time consuming and therefore the number of classes is generally pre-specified. The *BIC* criterion for selecting the number of classes is discussed in detail in [42].

2.9. APPLICATIONS OF EM

The EM algorithm has seen wide applications in different fields. Some of the applications where the EM algorithm and its variants are widely used are the following:

2.9.1. Background Modeling. The applications of the EM algorithm in ATR systems, echolocation systems and minefield detection have already been discussed in Section 1. In these systems, the EM algorithm helps to characterize the background, which then helps to conduct different types of analysis. Modeling of detection statistics gives an insight to the nature of the background. The concepts and applications of background modeling and modeling of detection statistics are discussed in Sections 3, 4, 5 and 6 of this thesis.

2.9.2. Speech Recognition. The Automatic Speech Recognition (ASR) systems have been making significant progress, but an important consideration is in its robustness to different speaking styles and environments. ASR systems trained in one environment perform poorly in the other environments due to a mismatch between testing and training conditions. These mismatches are caused by different speaking styles, the presence of noise and insufficient characterization of the speech signal [22].

To reduce this mismatch, mappings and transformations are done between the test utterances and original models. Let 'Λ_x' be a set of trained Hidden Markov Models (HMMs), where the subscript 'x' denotes that the models are trained on given set of training data, $\{X\}$. Let the test utterance, 'Y,' be given as: $Y = \{y_1, y_2, ..., y_j, ...y_n\}$. The

problem is to recognize the word sequence, '*W*,' from '*Y*,' where '*W*' is given by $W = \{w_1, w_2, ...w_L\}$.

Any mismatch between '*X*' and '*Y*' results in error in recognizing the word sequence, '*W*.' In order to reduce this mismatch, '*Y*' is mapped to '*X*' using the transformation function, 'F_v' such that,

$$X = F_v(Y). \tag{2.9}$$

Here, 'v' are the parameters of the transformation function, 'F_v.' The mismatch can be reduced by finding the parameters 'v' such that the likelihood, $p(Y \mid v, \Lambda_x)$ is maximized. Therefore 'v' can be obtained as:

$$v' = \arg\max_{v} p(Y \mid v, \Lambda_x). \tag{2.10}$$

Thus '*v*' can be obtained using the EM algorithm, and the transformation function, 'F_v,' can be obtained. This decreases the distortion caused by the mismatch. A discussion on this approach can be found in [43] and [44].

2.9.3. Medical Imaging. Magnetic Resonance (MR) imaging is an advanced medical imaging technique providing rich information about the human soft tissue anatomy. For brain MR images, the only method developed to date for statistical segmentation of brain MR images is based on the finite mixture (FM) model, in particular the finite Gaussian mixture (FGM) model where the Gaussian distribution is assumed because of its simple mathematical form and the piecewise constant nature of ideal brain MR images [40]. The mixture model consists of different tissue classes, especially gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). The EM algorithm is used to estimate the parameters of the mixture model, and assign class labels to all the pixels. It then segments the image by representing all samples that belong to a particular

class as a distinct region in the image. The discussion on EM model and image segmentation in MR imaging can be found in [21].

2.9.4. Dairy Science. Quantitative variation in traits that change with age is important to animal breeders. Often biological traits such as body size, weight or growth are measured at various times or ages. These traits are highly correlated. A (co)variance function is a continuous function that represents the variance and covariance of such traits measured at different points of time. The (co)variance function is useful when spatial or temporal data are modeled and can be linked to time-dependent phenomena such as growth of lactation. Using these (co)variance functions, information such as milk-yield etc. can be predicted for different points.

Let '$\sum$' denote the covariance matrix calculated at different times and 'φ' represent the matrix of polynomial functions evaluated at these times then the observed covariance matrix can be given by:

$$\sum = \varphi.K.\varphi^{T}, \tag{2.11}$$

where 'φ^{T}' denotes the transpose of 'φ.' Here 'K' denotes the coefficients of the (co)variance function. 'K' can be estimated as:

$$K = \varphi^{-1} \sum \left(\varphi^{-1}\right)^{T}. \tag{2.12}$$

Once 'K' is obtained, it is used to get updated covariance matrix, $\sum^{\circ}$. This covariance matrix is then used to give a better estimate of 'K.' Thus the EM algorithm is used to estimate the (co)variance function coefficients for different measurements such as milk, fat test-day yield etc. Estimation of the (co)variance function using EM has been discussed in [20], [38] and [39].

2.9.5. AIDS Epidemiology. The forecasting of future AIDS incidences using the estimates of past and present HIV incidences is very important for public health planning. Data relating to HIV infection is incomplete due to missing information and delayed reporting etc. Also there is an unknown time interval between the infection and diagnosis. All these aspects lead to the incompleteness of the problem and therefore in order to forecast the AIDS incidence, the EM algorithm is used [1], [18].

Let us suppose that:

A_t = Number of cases diagnosed as AIDS during the month '*t*.'
N_t = Number of individuals infected with HIV in month '*t*.'
P_k = probability that duration of infection is '*k*' months.

Clearly, '*k*' denotes the period between infection and diagnosis, in months.

$$A_t = \sum_{c=1}^{t} N_c P_{t-c+1}\,. \tag{2.13}$$

Let,

$$\mu_t = E(A_t),$$
$$\lambda_c = E(N_c).$$

where, '*E*' denotes the expectation.

The prediction of AIDS incidences for any future month, '*t*,' is given by:

$$\mu_t = \sum_{c=1}^{t} \lambda_c P_{t-c+1}. \tag{2.14}$$

The parameter, 'λ_c' is estimated using EM techniques. Thus this helps in the forecasts of the future AIDS incidences. The methods for estimating 'λ_c' are discussed in [37].

2.9.6. Census Surveys. Record linkage, or computer matching, is a means of creating and updating information that may be used in surveys. It serves as a means of linking individual records via name and address information from differing administrative files. Various types of personal information such as name and address, is linked using mathematical models and this helps in the proper analysis of the records.

The record linkage classifies pairs in product space '*AxB*' from two files '*A*' and '*B*' into a set of true matches, '*M*' and set of non-matches, '*U*.' In order to find if the given pair forms a match (or a link), the following ratio is evaluated [19]:

$$R = \frac{p(\alpha \mid M)}{p(\alpha \mid U)}\ , \tag{2.15}$$

where 'α,' represents the pair under consideration.

To find if 'α' forms a valid link, the following decision rule is applied on '*R*.'

If $R > T_u$, then designate pair as a link.

If $T_l \leq R \leq T_u$, then designate pair as a possible link and hold for review

If $R < T_l$, then designate pair as a non-link.

Here, 'T_u' and 'T_l' are the thresholds determined from prior information. The ratio

'*R*' is known as the matching parameter. It has been presented in [36] that these matching parameters can be estimated directly from the data using EM.

2.10. CONCLUSIONS

This Section discussed the underlying concepts of the EM algorithm. It also showed the mathematical formulation of the algorithm. The application of the EM algorithm in estimating the mixture model is discussed. Applications of the EM algorithm in various fields are also presented.

3. IMAGE SEGMENTATION USING EM

3.1. INTRODUCTION

Image segmentation is the process of division of an image into distinct region. The level to which this division is carried depends on the type of details needed. Image segmentation algorithms are generally based on the following criteria [24]:

(i) Segmentation based on similarity
(ii) Segmentation based on discontinuity

In segmentation based on similarity, all regions that are similar as per some criteria are classified as one region, whereas in the case of segmentation based on discontinuity, certain discontinuity criteria such as edges are used to divide images into regions. This thesis presents a technique for segmentation based on similarity using EM approach. The criterion of similarity is statistical in nature, i.e. regions that are statistically similar are classified as one region. Single pixel level segmentation and segmentation based on spatial distribution are performed. In case of segmentation based on spatial distribution, spatial correlation of the data is captured.

In the case of a minefield detection system, the sensor is selected so that they have good signal for the objects of interest. Thus infrared imaging is used to detect mines with good thermal signatures. Image segmentation can also be used to detect anomalies. It is because the anomalies, due to their statistical nature, would most likely belong to a class that is well separated from the background class. This would help in their detection using the concept of pixel membership. This is the topic of discussion of Section 4.

3.2. SEGMENTATION—SINGLE PIXEL

In stochastic model-based image segmentation, an image is divided into regions. Each region is associated with one of a finite number of classes and each class comes from a pre-defined number of classes that are modeled as random distributions. Because

all these distributions are parametric models, a problem of parameter estimation often arises. The EM algorithm is an excellent technique to estimate parameters in problems that have some information missing. Once the parameters are estimated, all samples can be associated with classes. Therefore the EM algorithm can be used to segment the image.

Let the observed data vector be, $y = \{y_1, y_2, ..., y_j, ...y_n\}$ assuming '*n*' observations. In the image segmentation model, the pixels are considered as samples. Therefore 'y_j' denotes the pixel intensity of the j^{th} pixel. In the process of image segmentation, as previously described, each pixel intensity belongs to a particular class that is unknown. Additional unobserved variable 'z_{ij}' is added to the unknown variable list which is an indicator variable that is one or zero according to whether 'y_j' (j = 1, 2... n) arose from the i^{th} (i = 1, 2... g) class.

It is assumed that each class has a Gaussian distribution with the unknown mean and variance so that the probability density function, pdf, for the ith class is given by:

$$p_i(y) = N(y;\, \mu_i,\, \sigma_i^{\,2}), \text{ where } i = 1, 2...g. \tag{3.1}$$

Then the pdf of the observed data, 'y,' can be written as:

$$f(y \mid \varphi) = \sum_{i=1}^{g} \pi_i N(y;\mu_i,\sigma_i^2), \quad \varphi = \{\pi_i, \mu_i, \sigma_i^2\}, \tag{3.2}$$

where 'π_i' is a coefficient that denotes proportion of the sample due to the i^{th} class and

$$\sum_{i=1}^{g} \pi_i = 1. \tag{3.3}$$

Following the lines of the prior discussion in Section 2.5, the log-likelihood function of the complete data is formed as.

$$\ln\left[f\left(x \mid \varphi\right)\right] = \ln\left[£_c\left(\varphi \mid x\right)\right] = \sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij} \ln\left[\pi_i N\left(y_j;\mu_i,\sigma_i^2\right)\right]. \tag{3.4}$$

This log-likelihood function can also be written as:

$$\ln\left[£_c\left(\varphi \mid x\right)\right] = \sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij} \ln\left[\pi_i\right] + \sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij} \ln\left[N\left(y_j;\mu_i,\sigma_i^2\right)\right]. \tag{3.5}$$

As the above likelihood function is linear in ‘ z_{ij} ,’ the ‘*E*’ step on the *(k+1)*th iteration requires the calculation of the expectation of ‘ z_{ij} ,’ given the observed data, ‘ y .’ This expected value of ‘ z_{ij} ’ is obtained as:

$$z_{ij}^k = E_{\varphi^k}\left[z_{ij} \mid y\right] = \frac{\pi_i^k N\left(y_j;\, \mu_i^{(k)}, \sigma_i^{2(k)}\right)}{f(y_j, \varphi^k)}\,, \tag{3.6}$$

where E_{φ^k} is the Expectation operator over the current parameter ‘ φ^k ,’ and

$$f\left(y_j, \varphi^k\right) = \sum_{i=1}^{g} \pi_i^k N\left(y_j;\, \mu_i^{(k)}, \sigma_i^{2(k)}\right), \tag{3.7}$$

gives the distribution of the observed data, ‘ y_j ,’ for the k^{th} iteration. The ‘Q’ function is then written as:

$$Q\left(\varphi, \varphi^0\right) = E_{\varphi^0}\left[\ln\left\{£_c\left(\varphi|x\right)\right\} \mid y, \varphi^0\right]. \tag{3.8}$$

On maximizing the 'Q' function, the update equations are obtained for the various parameters. These equations are provided in appendix—A.1.1.

Therefore with the iteration of the expectation and maximization step, the EM algorithm not only gives a good estimate of the parameters of the class distribution but also tells about the proportion in which these classes are mixed. This is then used to model the data. The above model that represents the observed data as a mixture of classes is also sometimes referred to as a mixture model because it represents a mixture of 'g' classes, each with a pdf of its own. In the above problem all classes are assumed to have a Gaussian distribution with each class having a mean and variance of its own. The distribution can be any parametric distribution. For the modeling of the background, pixel-level classification has been done using Gaussian mixture model. In case of modeling of the RX statistic, the classification has been done on the RX detection statistic using Beta and Gamma distributions as shown in Section 5.

3.3. SEGMENTATION—SPATIAL DISTRIBUTION

In order to capture spatial correlation, 5-dimensional data is generated from the spatial neighborhood at a distance of two. Figure 3.1 shows a scheme for sampling the data to capture the local spatial variations of the data. Pixels at a distance of two are considered since 4-neighborhood pixels are expected to be highly correlated and do not capture useful spatial characteristics of the local neighborhood. Taking a large neighborhood with more samples increases the dimensionality of the model which is computationally expensive.

Since in this case the segmentation is based on spatial distribution, segmentation reflects spatial correlation in the data. The class assignment depends not only on the gray value of the pixel but also on the gray values of the nearby positions with respect to the center pixel. The Gaussian mixture model discussed in Section 3.2 is still valid with the only difference that a multivariate normal distribution is used to define the mixture

model. The update equations for multivariate Gaussian mixture model are provided in appendix—A.1.1.

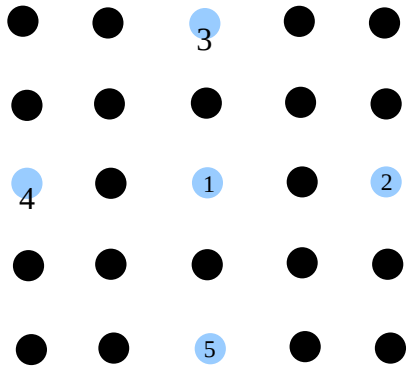


Figure 3.1. Spatial Distribution of the Pixels.

3.4. PIXEL LEVEL CLASS ASSIGNMENT

Assuming there is a *512x640* image, it has *327,680* samples, (y_1 to y_{327680}). If the given image is modeled by '*g*' classes, then each one of these samples can be assumed to belong to one of the '*g*' classes. Before the start of the algorithm, each pixel is assigned a class randomly with equal probability, since there is no estimate of the initial parameters. It is to be noted here that the information known to the algorithm are the observed data, the initial guessed values and the number of classes. After the algorithm converges, array of indicator variables, ' z_{ij} ,' can be used to assign classes to all the samples. If there are '*n*' samples and '*g*' classes then the ' z_{ij} ' array has '*g*' rows and '*n*' columns. Consider the j^{th} column (j = 1 to n). The i^{th} row (i = 1 to g), of the j^{th} column represents the probability of the j^{th} sample belonging to the i^{th} class. The row that has the maximum probability for a given column, j, represents the class assigned to the j^{th} sample. The segmentation is achieved by representing all samples that belong to a particular class as a distinct region in the image.

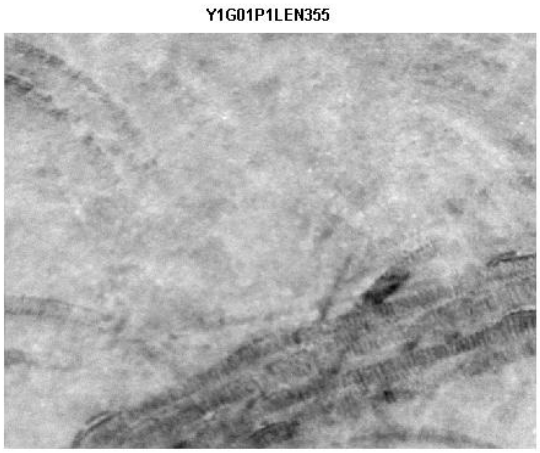
3.5 RESULTS—SEGMENTATION USING SINGLE PIXEL

Figures 3.2-3.4 show the results of segmentation by two, three and four classes. In these figures, part (a) shows the image that has been segmented by the given number of classes, part (b) shows the segmented image and part (c) shows the probability density function (pdf) of the constituent classes. The sum of these pdf's models the histogram of the samples. The color on the segmented image and the line color of the pdf's are matched for convenience of interpretation. The histogram of the samples has been obtained by sampling the image (row-wise and column-wise) with a sampling rate of five. The sample pdf has been plotted using the red dash-dot line in part (c) of the Figures 3.2-3.4. A good match between the sum of class pdf's and the sample pdf shows the accuracy of segmentation and modeling with the given number of classes.

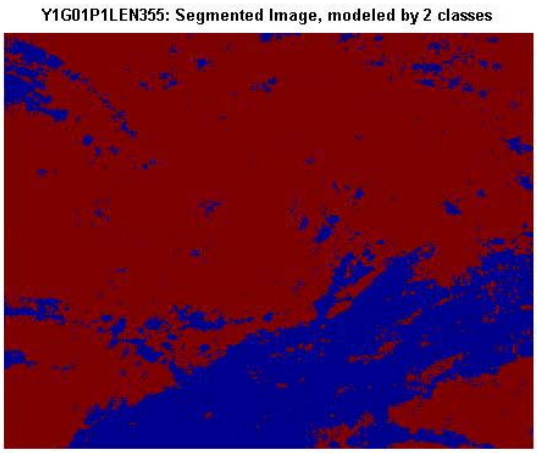
The legend of the figure shows the fractional membership of each mixture component. For example, in Figure 3.2 (c) the proportion of Class-1 is 0.32 whereas the proportion of Class-2 is 0.68. The two regions can be clearly seen in the image. One is the background and the other is region formed by the track marks of a vehicle. As can be seen from the pdfs of the classes, the region formed by the tracks, forms the minor class that has the proportion of 0.32. In the case of three classes, the image can be divided into three regions. The first region consists of the background, the second consists of the vegetation and the third region is the slightly bright region surrounding the vegetation. The pdfs clearly show that the proportion of the classes consisting of the second and third region is less than that of the major class that consists of the background.

In the case of four classes, the background is divided into two regions based on the type of soil, i.e. dark and light. There are some bright regions in the image that most likely represent trees. These bright regions are further divided into the other two classes, which have smaller memberships.

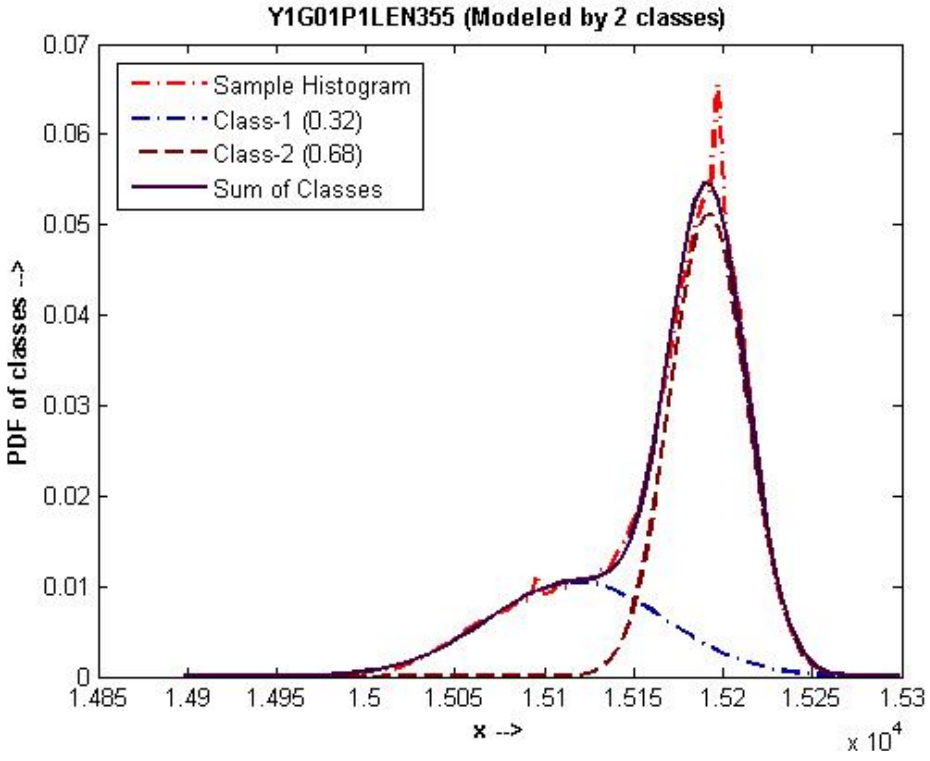
The results (Figures 3.2-3.4) show that the EM algorithm was able to segment the image well for all the three cases. It can also be seen from these figures that the sum of the constituent pdfs very well modeled the sample pdfs.



(a) Actual Image



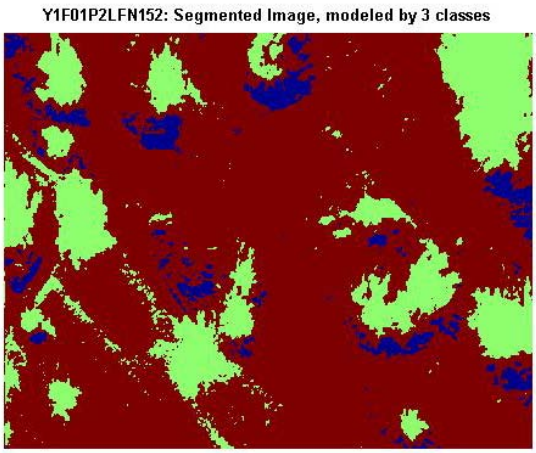
(b) Segmented Image



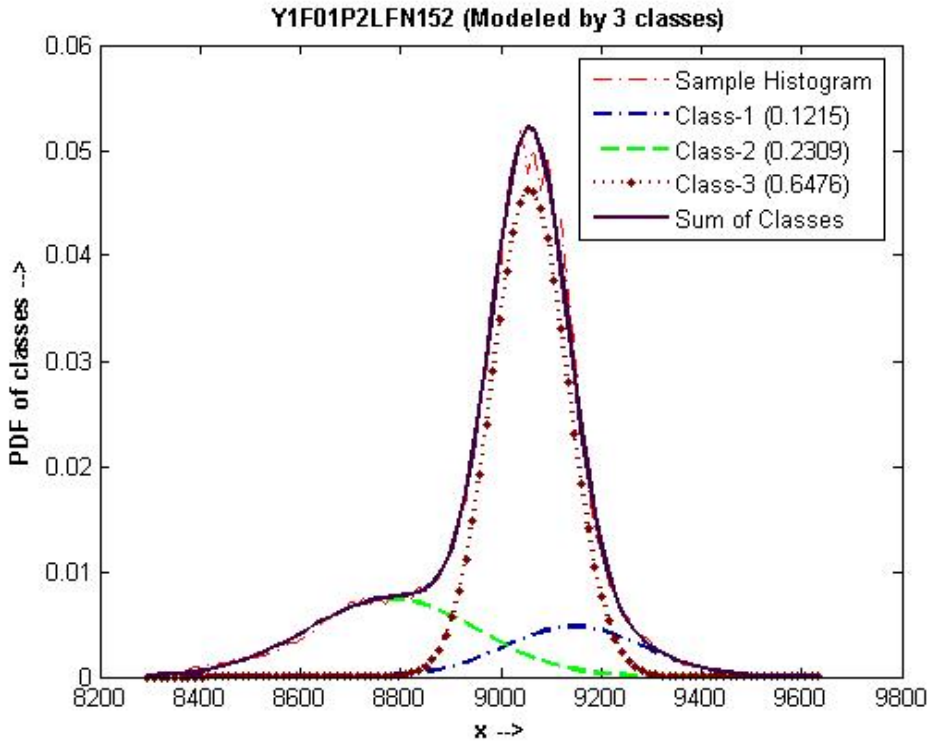
(c) Modeling by Two Classes

Figure 3.2. Example of Image Segmentation Using Two Classes.

(a) Actual Image

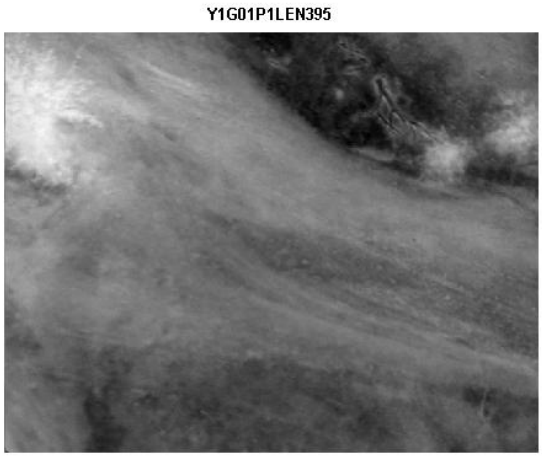


(b) Segmented Image

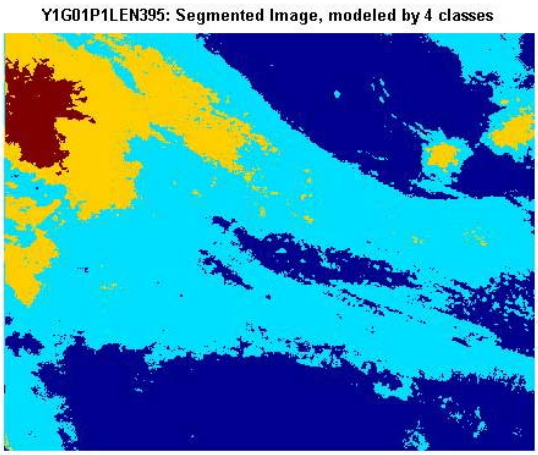


(c) Modeling by Three Classes

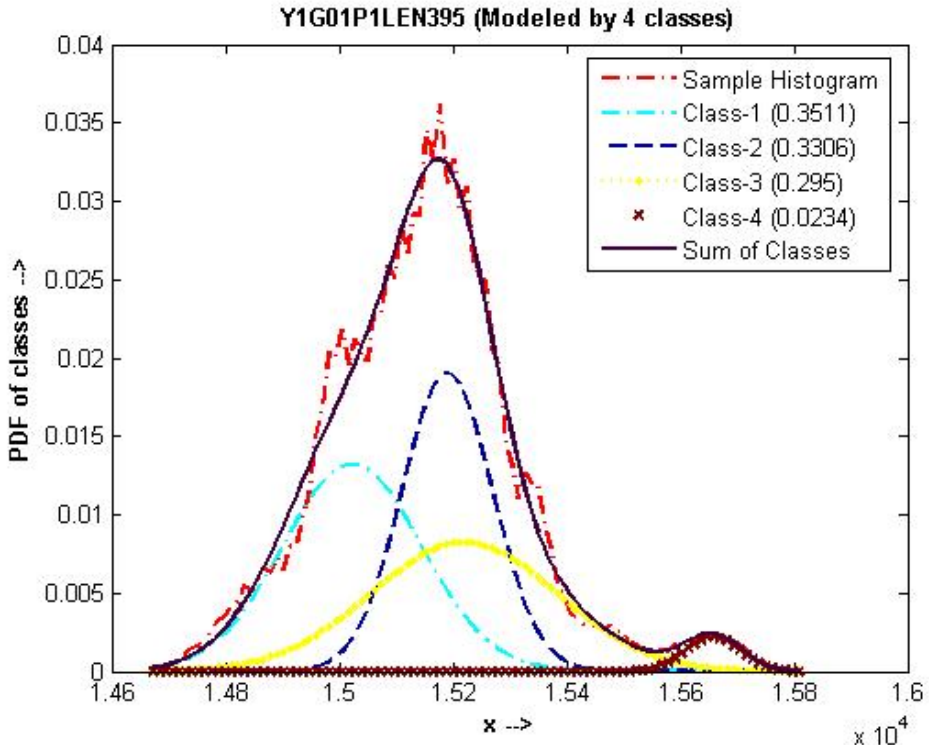
Figure 3.3. Example of Image Segmentation Using Three Classes.



(a) Actual Image



(b) Segmented Image



(c) Modeling by Four Classes

Figure 3.4. Example of Image Segmentation Using Four Classes.

3.6. RESULTS—SEGMENTATION USING SPATIAL DISTRIBUTION

Figure 3.5 shows the results for the segmentation based on spatial distribution from two representative frames of the May 2003 data. Figure 3.5 (a) and (c) show the image frames whereas Figure 3.5 (b) and (d) show the corresponding segmented images.

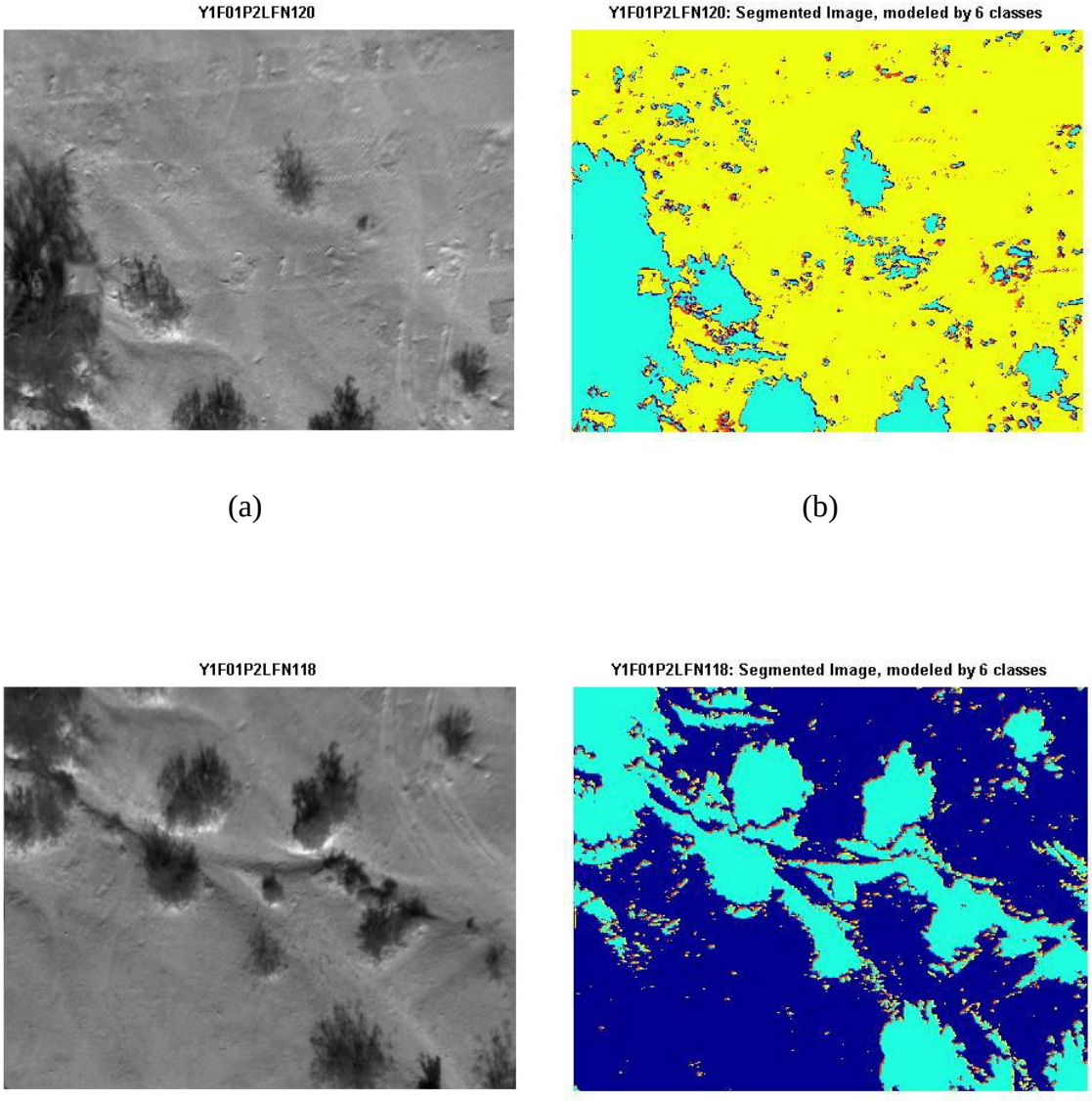


Figure 3.5. Image Segmentation Using Spatial Distribution.

In this case the segmentation captures both gray value features (like vegetation soil etc) as well as the transition from one type of region to another. Because of this reason the edges formed due to the transitions in the background areas (for example due to the color of the soil) are captured and shown as small patches in the segmented image.

3.7. CONCLUSIONS

In this section, single pixel level segmentation of the MWIR images is discussed. Multi (five) band data is generated from spatial neighborhood and segmentation based on spatial distribution is performed. The results of these two types of segmentation are shown and discussed.

4. ANOMALY DETECTION USING EM

4.1. INTRODUCTION

In this section, various methods of anomaly detection are presented. First of all, a brief literature of anomalies or outliers (as they are called) is presented. This is followed by the study of some of the recent EM-based techniques that have evolved to detect anomalies. All of these require a mixture model. This section uses the results of EM-based segmentation that have been obtained in the previous section. In other words, the anomaly detection stage comes after the pixel-level class assignments have been made. Therefore once the background data has been modeled with a Gaussian or some other mixture model, the concept of pixel membership is employed to define different EM-based detection statistics.

4.2. ANOMALIES AND OUTLIERS

The values that are away from the mainstream data are called as anomalies. Thus anomalies or outliers can be broadly described as observations (or subset of observation) that “appear to be inconsistent” with the remainder of that set of data [12]. The crucial point is that the phrase “appear to be inconsistent” is subjective because it is a matter of subjective judgment on the part of the observer to declare a particular observation as inconsistent. In the case of anomaly detection using EM-based algorithms, the outliers are said to be inconsistent because they are different from the background in a statistical sense.

The outlier problem is two-fold. First, it is necessary to determine whether the observation is really an outlier. Once this is decided, the next step is to find an efficient way to deal with these outliers. Rejection of outliers may not be an efficient way in certain situations. Thus the methods for processing the outliers take an entirely relative form and differ from situation to situation. For the current discussion, these outliers are of interest as possible indications of presence of mine-like targets.

Anomaly detection is extensively used within the field of target detection and computer security [13]. In the case of target detection, the process of anomaly detection is comprised of finding the data samples that are different from the other samples in its neighborhood in some way. In the case of intrusion detection (in computer security), the call traces are observed for sequences that are different from the others, and these are then categorized as anomalous or intrusions [13]. Anomaly detection for hyper spectral imagery (HSI) is a useful technique for detecting objects of military interest [14]. Once the anomalies are detected, they can be passed to a classifier in order to determine the exact nature of the unusual object.

The remaining part of this section presents different EM-based anomaly detectors that use the concept of pixel membership. In all these detectors, it is assumed that the anomalies, being different from the background, would form a class that is well separated from the background. All these EM-based detectors are compared with the RX anomaly detector. Therefore the next section gives a brief review of the RX anomaly detector.

4.3. THE RX ANOMALY DETECTOR

The RX algorithm as suggested by Reed and Yu [25] can be used for multi-band images with a zero mean, uncorrelated Gaussian background. For a '*J*' band image '*I*,' the RX statistics, r_X , at any location is given as:

$$r_X = N_T\left(\mu_S^T M^{-1} \mu_S\right), \tag{4.1}$$

where 'M' is the locally estimated covariance matrix of dimension 'J x J' given by:

$$M = \frac{1}{N_C}\sum_{l \in W_C} I_l^T I_l, \qquad I_l = I(i, j) \tag{4.2}$$

‘ μ_S ’ is the mean target signature and is given by:

$$\mu_S = \frac{1}{N_T} \sum_{l \in W_T} I_l \ . \tag{4.3}$$

Here, ‘ N_C ’ is the number of clutter pixels and ‘ N_T ’ is the number of target pixels which are given as:

$$N_C = \pi\left(r_u^2 - r_l^2\right), \tag{4.4}$$

$$N_T = \pi r_T^2 \ . \tag{4.5}$$

This thesis considers only single band data, and therefore J = 1. The RX statistic in this case is given as:

$$r_X = N_T \left(\frac{\left(\frac{1}{N_T} \sum_{j \in W_T} I_j \right)^2}{\frac{1}{N_C} \sum_{i \in W_C} I_i^2} \right) = N_T \left(\frac{\mu_s^2}{M} \right). \tag{4.6}$$

4.4. SEM-BASED ANOMALY DETECTORS IMPLEMENTED

This section will discuss three SEM-based anomaly detectors. All these detectors are based on the Gaussian mixture model of the image. The mixture model is estimated using the SEM algorithm. Once the SEM divides all the image pixels into classes and estimates the class parameters, one is ready to enter the SEM-based anomaly detection stage. The detectors that have been implemented are discussed in the following sub-sections.

4.4.1. Detector Using Pixel Membership. Sometimes the target pixels are not clearly resolved from the background, but they do appear somewhat atypical relative to the densities that define the mixture model. This fact is exploited by the SEM detector for detecting poorly resolved targets.

In this method, the likelihood of pixels is evaluated for the presence of anomalies. If the target pixels are separated out from the background, then their membership with respect to the background is low. Thus the pixels for which the likelihood is sufficiently low can be declared as targets. To test for the presence of anomalies, the metric used over a target window, 'W' is given as [17]:

$$T = -\sum_{j \in W} \ln\left[\sum_{i=1}^{g} \pi_i \, p_i\left(y_j\right)\right]. \tag{4.7}$$

where, y_j, is the pixel value of the j^{th} sample, '*ln*' is the natural logarithm, 'g' is the number of classes, 'π_i' is the probability of i^{th} class and $p_i\left(y_j\right)$ is the class membership for pixel 'y_j.' The size of the window 'W' should be matched to the typical size of the expected anomaly.

The test statistic *'T,'* gives a high value at the locations of the anomalous pixels.

4.4.2. Detector Using Locally Optimum Bayes Detection (LOBD). In this detector, a locally optimum detection algorithm is applied for detection of random signals that occur in the mixture noise. The LOBD statistic for a discrete zero-mean signal in mixture noise is derived under the assumption that the signal and noise are independent. The detection statistic is used to distinguish between the following null and non-null hypotheses:

$$\begin{aligned} H_0 &: y_j \in N_j \\ H_1 &: y_j \in \sqrt{as_j} + N_j, \end{aligned} \tag{4.8}$$

where $j \in W$ are the pixels in some neighborhood window 'W', 's' is the unit variance signal template, 'a' is the unknown variance of the received signal and 'N_j' is the j^{th} noise sample.

If the mixture noise is Gaussian, then a locally optimum test statistic has been derived in [16]. This is given as:

$$T(y_j) = \frac{1}{2} \sum_{j \in W} \sum_{i=1}^{g} \left(\frac{-2}{\sigma_i^2} + \frac{\| y_j \|^2}{\sigma_i^4} \right) p(i \mid y_j, H_0) , \tag{4.9}$$

where,

$$p(k \mid y_j, H_0) = \frac{\frac{\pi_k}{2\pi\sigma_k^2} \exp\left(-\frac{\| y_j \|^2}{2\sigma_k^2} \right)}{\sum_{i=1}^{g} \frac{\pi_i}{2\pi\sigma_i^2} \exp\left(-\frac{\| y_j \|^2}{2\sigma_i^2} \right)} . \tag{4.10}$$

Here, $p(k \mid y_j, H_0)$ is the probability of class 'k' for null hypothesis, given the input samples. The detection statistic, $T(y_j)$, gives a high value for the anomalies.

It can be noted that the statistics discussed in section 4.4.1 and 4.4.2 are based on the segmentation of the data representing the whole image. As a result they are not locally adaptive. In the next section, a locally adaptive anomaly detector, based on multi-class RX is discussed. This algorithm has been adapted from the one proposed by Dr. Fries for the airborne program [15].

4.4.3. Detector Using Multi-Class RX. It is known that SEM segments an image into homogeneous regions based on the intensity at each pixel. The intensity at each pixel is then tested against the regions that surround it to measure the degree to which it stands out. This can be operated on a window that moves down through the image. This can be

done by means of a likelihood ratio test [15] where the numerator is likelihood of the observed spectrum assuming the target is present and denominator is likelihood of the observed spectrum assuming the target is not present.

Let the probability of the j^{th} pixel belonging to the i^{th} class be given by:

$$P_j\left(H_{Bi}\right)=\frac{P\left(y_j \mid \sigma_{Bi},\mu_{Bi}\right)}{\sum_m P\left(y_j \mid \sigma_{Bm},\mu_{Bm}\right)}. \tag{4.11}$$

Likelihood of the i^{th} class in a neighborhood under the window, 'W_c' is given by:

$$P_{W_c}\left(H_{Bi}\right)=\sum_{j\in W_c}\frac{P_j\left(H_{Bi}\right)}{N_j}. \tag{4.12}$$

Then, the likelihood of 'y_j' under null and non-null hypothesis is given by:

$$P\left(y_j \mid H_0\right)=\sum_i P_{W_c}\left(H_{Bi}\right)P\left(y_j|\sigma_{Bi},\mu_{Bi}\right). \tag{4.13}$$

$$P\left(y_j \mid H_A\right)=\sum_i P_{W_c}\left(H_{Bi}\right)P\left(y_j \mid \sigma_{Bi},\mu=y_j\right). \tag{4.14}$$

The likelihood ratio, 'Λ' is given as:

$$\Lambda=\frac{P\left(y_j \mid H_A\right)}{P\left(y_j \mid H_0\right)}. \tag{4.15}$$

The detection statistic, 'Λ,' gives a high value for the anomalous pixels. If the anomaly target region is known to be of some size 'W' then the test statistic for the log-likelihood can be written as:

$$\Lambda_W=\sum_{j\in W}\log(\Lambda)=\sum_{j\in W}\left[\log\left\{P\left(y_j|H_A\right)\right\}-\log\left\{P\left(y_j|H_0\right)\right\}\right]. \tag{4.16}$$

4.5. THE AIRBORNE IMAGERY AND THE MINE DISTRIBUTION

This section presents an overview of the data processed. The image data in airborne minefield detection is collected by a low flying aircraft over different terrain during four different times of the day. Each flight of data collection is called a run which consists of a certain number of images (called frames) captured at approximately *8* frames per second. Each image is of *512 x 640* pixels in dimension. The data is collected for different mid wave infra red (MWIR) bands and over four different times of the day, typically—early morning, afternoon, evening and night.

The October 2002 data consists of *15* runs. The *15* runs contain data from three missions over different times of the day and bands. There are a total of *2143* frames (*143x14 + 141x1*). This particular collection of data consists of daytime data only. The primary terrain for the data collection can be classified as rocky arid areas with some vegetation. Also, the data present contains only surface mines in it. The ground truth is available for *371* mine targets, consisting of *190* large surface mine targets and *181* small surface mine targets. Ground Truth is also available for *430* fiducials. These are rejected from the false alarms at the time of analysis. Only daytime data was available under this data set.

The dataset for May 2003 is the biggest collection of data available for analysis. The primary terrain with mines can be classified as arid with little or no vegetation. In this data collection, data showing background alone has also been collected in addition to data showing both background and minefields. The collection consists of *37765* image frames from *83* runs over mine areas from ten missions covering different times of the day and bands. It also contains *55* runs over background only areas from nine missions covering different times of day for full band (Band 0 and Band 1). The ground truth is available for *12391* mine targets of which there is *1748* large surface mine targets, *844* small surface mine targets, *3946* medium surface mine targets and *5853* buried mine targets [23], [35].

These two data collections together encompass a number of mine and minefield modalities such as terrain type, band, mine signatures and minefield performances. Some of the characteristics of the airborne imagery data [23] that spans across all the data collections are listed as follows:

- Mine Types: Different types of mines can be classified based on size (large, medium, small), composition (metal, plastic) or method of placement (surface, buried):
    - Large surface mines : LP_B (plastic) and LM_A (metal)
    - Medium surface mines : MP_A (plastic)
    - Small surface mines : SM_A (metal)
    - Large buried mines : LM_A_B (large metal), LM_C_B (large metal) and LP_D_B (large plastic).
    - Medium buried mines : MP_A_B (medium plastic)

- Spectral Bands: The four spectral bands in the MWIR spectrum range for which the data is collected are identified as:
    - Full band (3-5 μm)—Band 0, Band 1
    - Solar band (3-4 μm) —Band 2
    - Thermal band 1 (4-4.5 μm) —Band 3
    - Thermal band 2 (4.5-5 μm) —Band 4

The various detection statistics discussed in section 4.4 were performed on the two daytime datasets from May 2003 data. The mine distribution for surface and buried mines for these datasets is shown in Table 4.1.

Table 4.1. Mine Distribution for Surface and Buried Mines.

Mine Name	Y1F01P2LFN	Y1F01P2LDN	Total
LM_A	12	12	24
MP_A	47	50	97
LP_B	12	9	21
SM_A	12	12	24

(a) Surface Mines

Mine Name	Y1F01P2LFN	Y1F01P2LDN	Total
LM_A_B	12	12	24
MP_A_B	12	13	25
LM_C_B	40	41	81
LP_D_B	11	8	19

(b) Buried Mines

Figure 4.1 shows the mine frames from this data (daytime). The various mines and clutter are marked using '×' marks. The red color is used to represent the mines whereas the cyan color is used to represent the clutter or fiducial markers. Figure 4.1 (a) shows the frame with only surface mine signatures. Figure 4.1 (b) shows the frame with only buried

mine signatures. Figures 4.1 (c) and (d) show the frames with both surface and buried mine signatures.

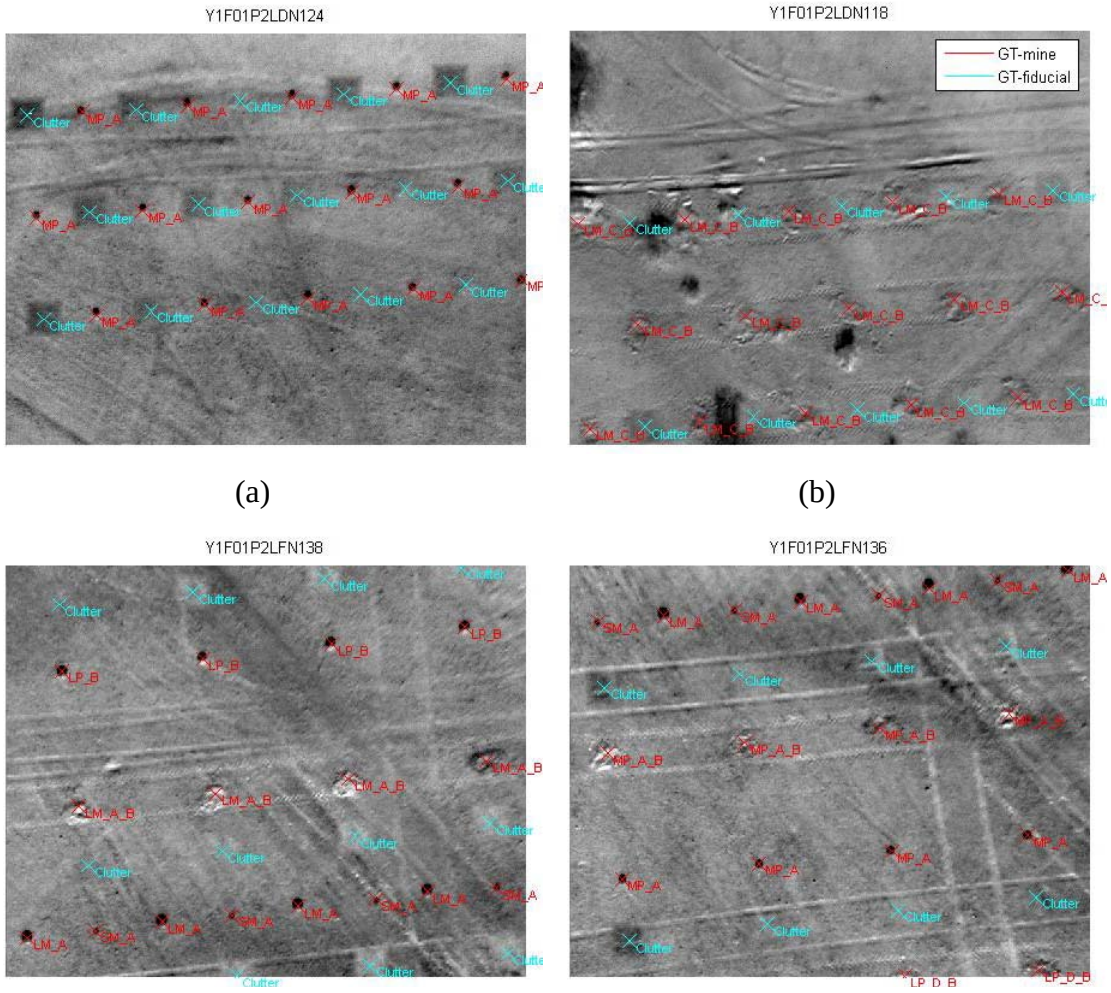
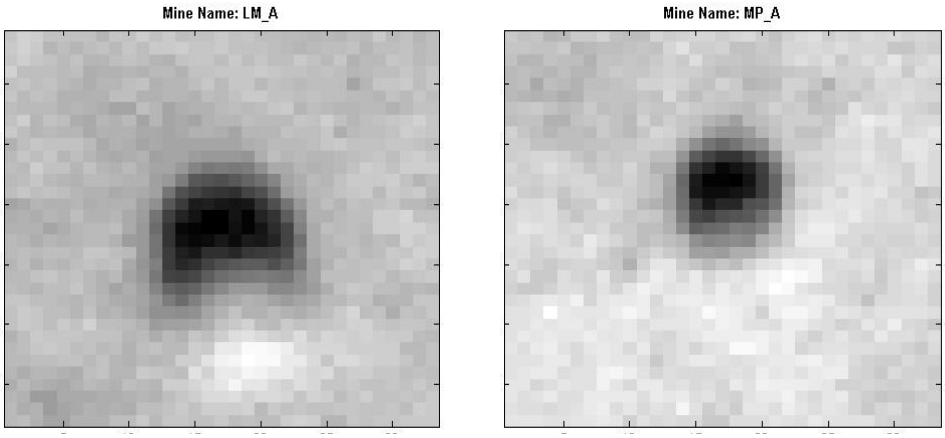
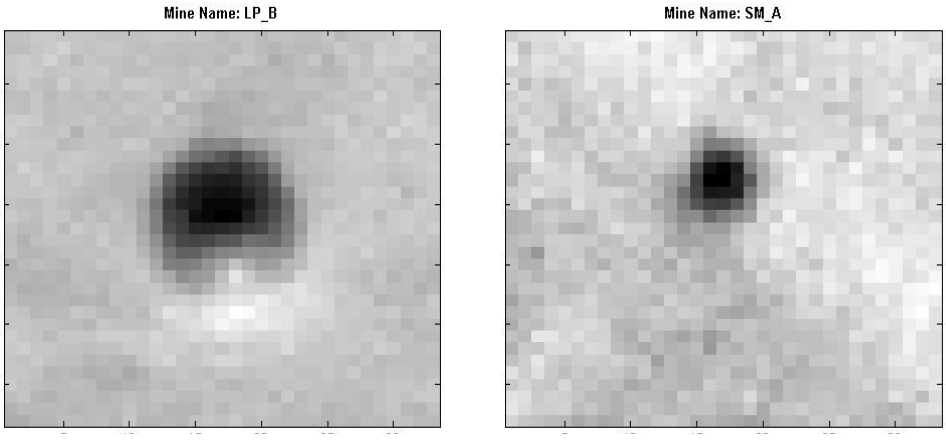


Figure 4.1. Mine Frames Showing Surface and Buried Mines from the May 2003 Data.

Figure 4.2 shows some of the typical mine signatures of various surface mines. Typical surface mine signature shows a bright reflection and a prominent shadow feature. In case of MP_A and SM_A mines the reflection is not prominent. It is the signature due to shadow that is captured by different anomaly detectors.



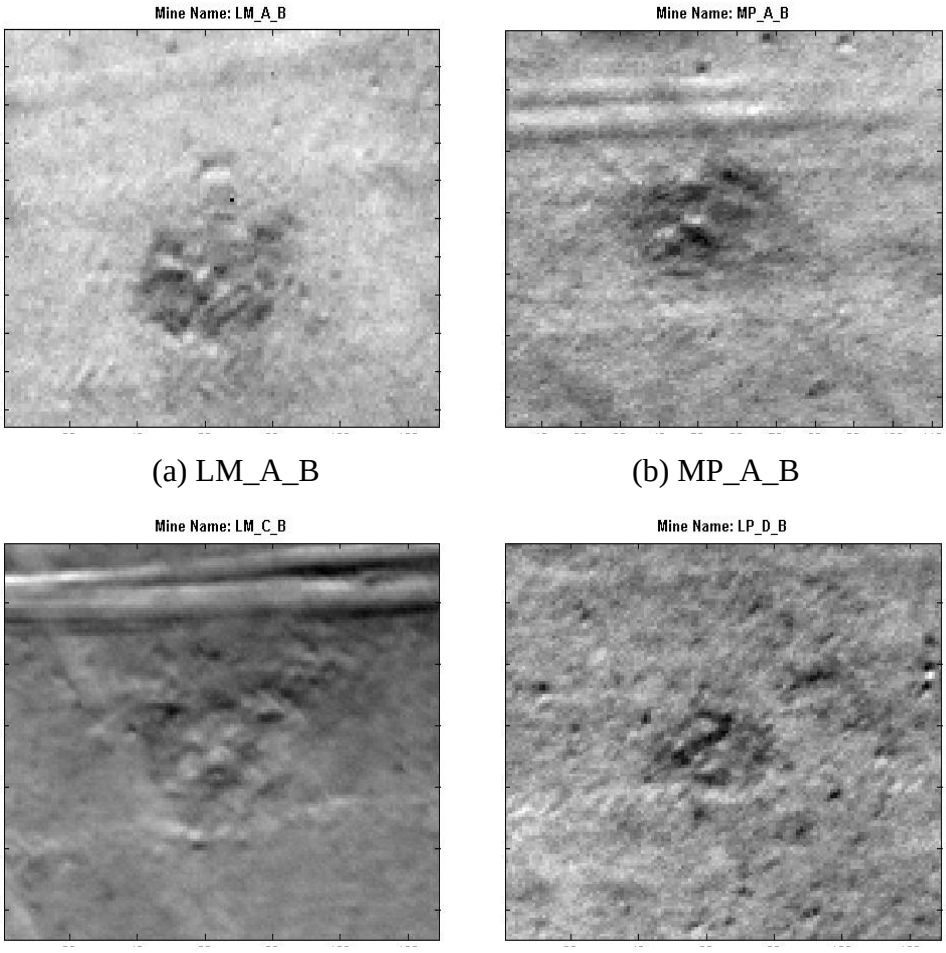
(a) LM_A (b) MP_A



(a) LP_B (b) SM_A

Figure 4.2. Individual Mine Signatures of Surface Mines.

Figure 4.3 shows some of the typical mine signatures of different buried mines. It can be seen that the buried mines have a weak signature. The mine signature of the buried mines is bigger than the physical size of the mine. This is because the signature mainly represents the disturbed soil. Therefore for the detection of these mines a larger target window is required for better detection performance.



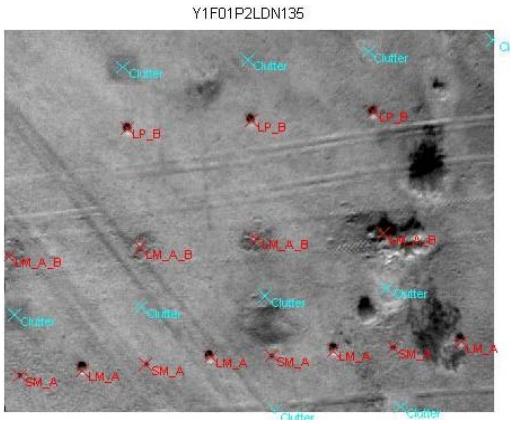
(a) LM_A_B (b) MP_A_B

(c) LM_C_B (d) LP_D_B

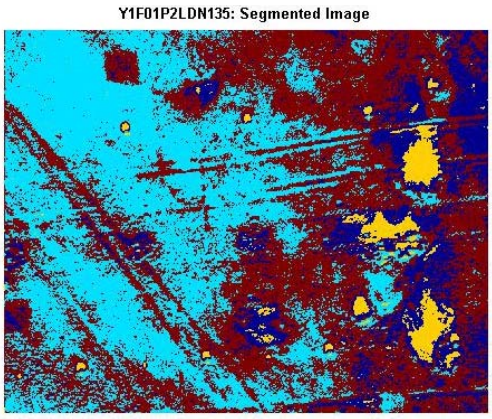
Figure 4.3. Individual Mine Signatures of Buried Mines.

4.6. RESULTS

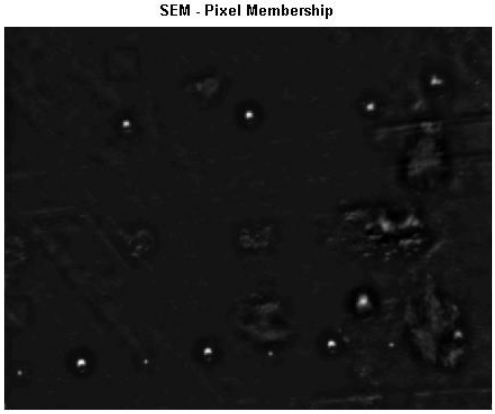
Figure 4.4 shows the MWIR image from one of the datasets of May 2003 data. It also shows the segmented image and the output of the various detectors discussed in Section 4.3 and 4.4.



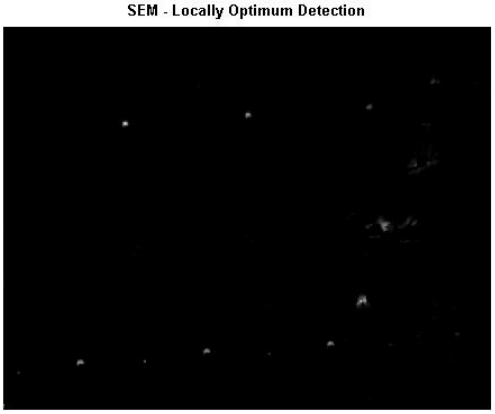
(a) Raw Image Data



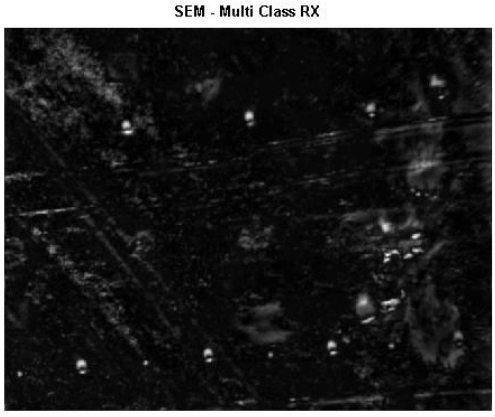
(b) Segmented Image



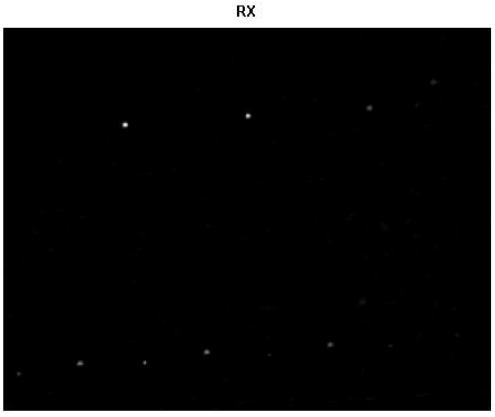
(c) Detector Output for Pixel Membership



(d) Detector Output for LOBD



(e) Detector Output for Multi-Class RX



(f) Detector Output for RX Detector

Figure 4.4. Raw Image Data, Segmented Image and Outputs of Various Detectors.

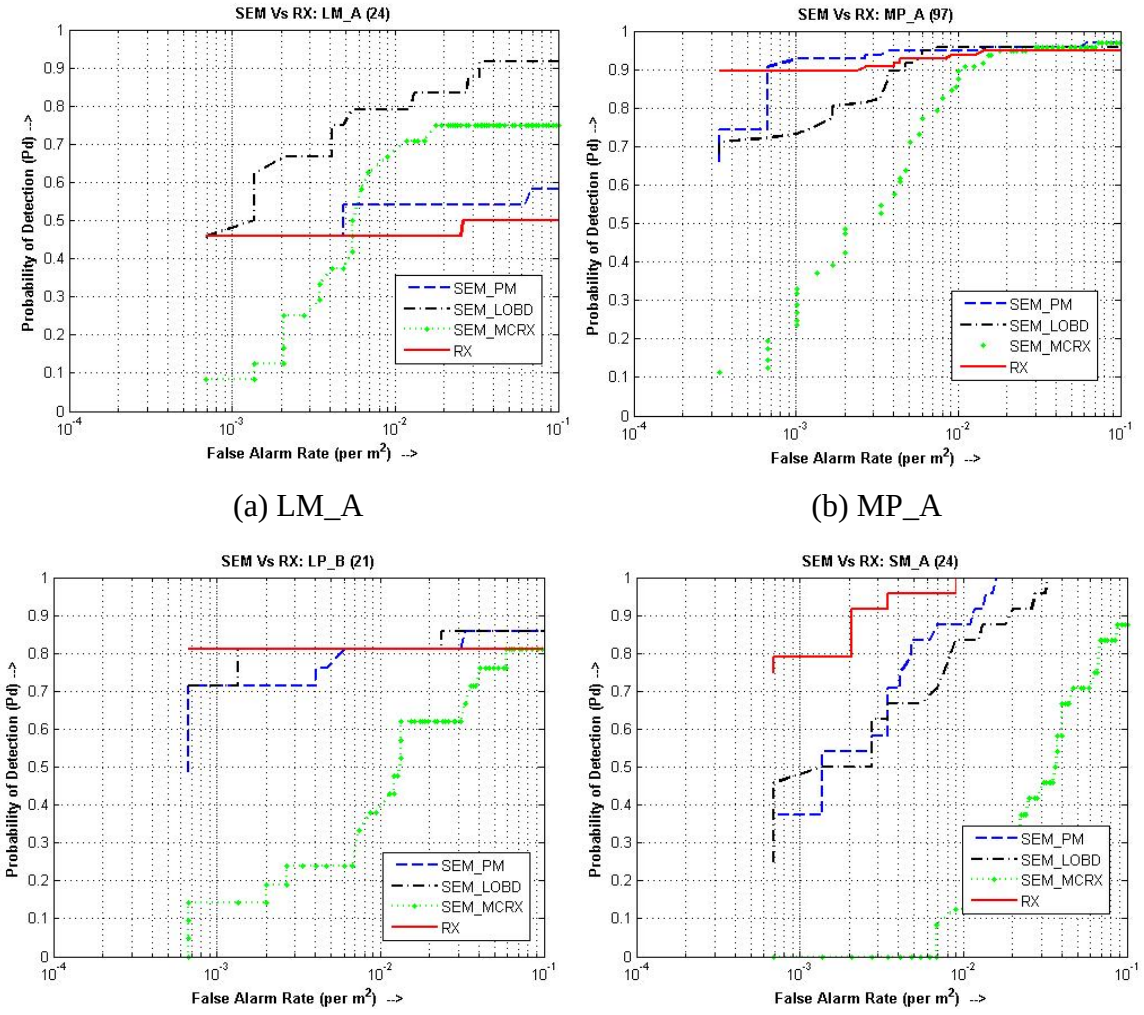
Figure 4.4 (a) shows one of the mine frames from the May 2003 data. The frame shows three surface mine types (LM_A, LP_B and SM_A), and one buried mine type, LM_A_B. It can be seen that the surface mine signatures consists mainly of the shadows. It can also be seen that in case of LM_A and LP_B the signatures partly consist of reflections as well. On the whole the signatures of the surface mines have a high contrast. On the other hand, the signatures of the buried mines are weak.

Figure 4.4 (b) shows the segmented image using four classes. The class containing surface mines is basically of shadow and is a minority class. Also this class is reasonably separated from the background class. The signatures of the buried mines are weak. These signatures basically consist of the disturbed soil and therefore they belong to a class that is not very well separated from the background classes.

Figure 4.4 (c)-(e) shows the outputs of different SEM-based detectors, whereas Figure 4.4 (f) shows the output of the RX anomaly detector. The outputs of the EM detectors using pixel membership and multi-class RX (parts (c) and (e) respectively) are on a log scale and therefore show a large variation in intensity of different regions. It can be noted that the statistics based on pixel membership and Bayes detection (parts (c) and (d) respectively) are based on the segmentation of the data representing the whole image, whereas the statistic based on multi-class RX (part (e)) is operated over a window. Hence the detector based on multi-class RX tends to detect local anomalies. This can be seen by observing the output of this detector (part (e)) of Figure 4.4.

Figure 4.4 (f) shows the output of the RX anomaly detector. The RX detection statistic is basically the signal-to-clutter ratio as discussed in section 4.3. This statistic has a high value if the anomalies have a high contrast signature (high signal strength) and are surrounded by a homogeneous region (low clutter variance and low clutter strength) in its neighborhood. As can be seen from part (a) of the figure that the surface mines have a high contrast signature and they are mostly surrounded by a relatively homogeneous neighborhood. As a result, surface mines are well detected by the RX detector.

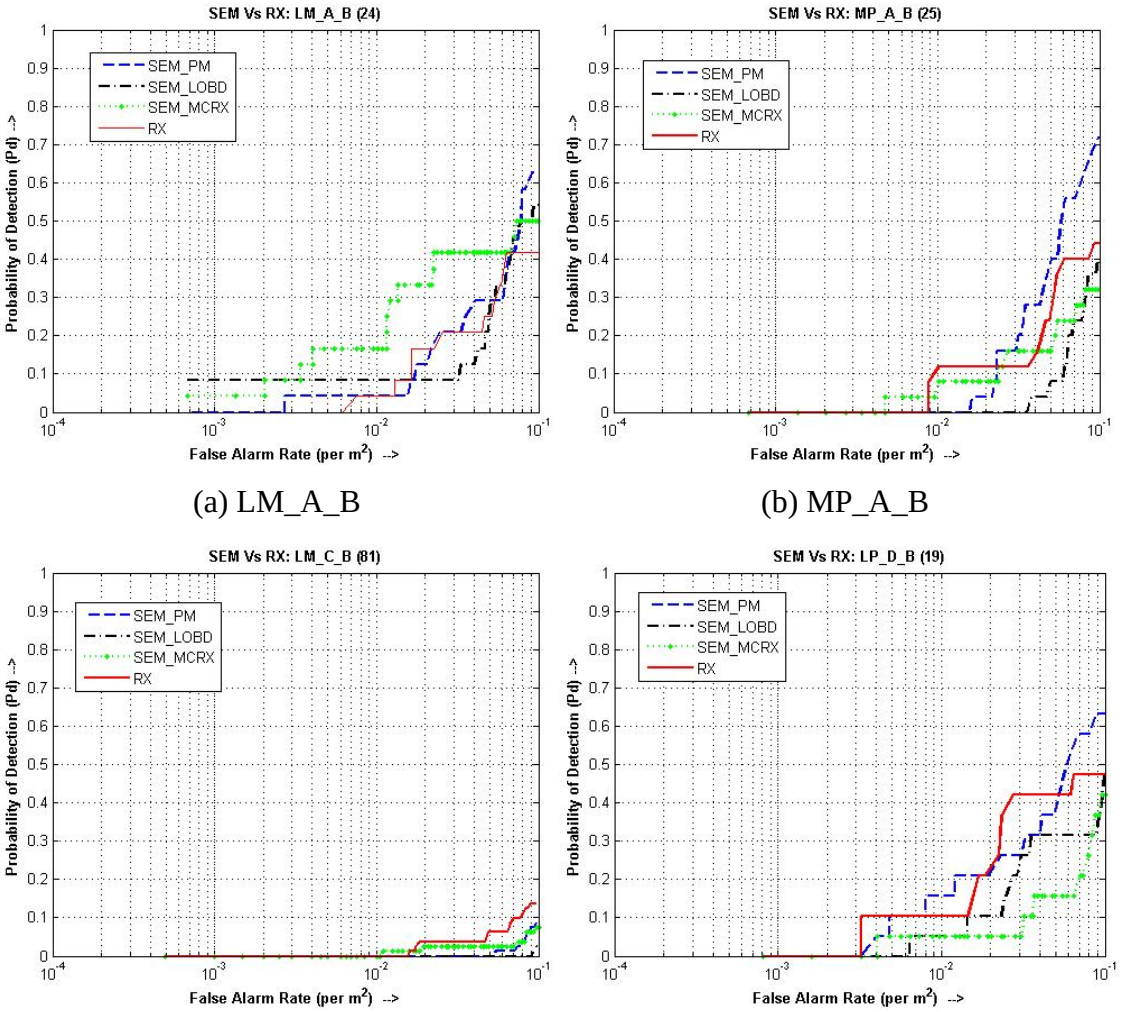
In order to compare the performance of the detectors, ROCs have been generated using the different SEM-based detectors discussed in Section 4.4. These detectors are compared with the RX anomaly detector. All the ROCs show the probability of detection, P_d, achieved until the False Alarm Rate (FAR) of 0.1/m^2. There are eight sub-plots, each corresponding to one of the eight types of mines listed in Section 4.5. The title of these figures shows the mine type and the number of such mines in the data. For example, LM_A (24) in Figure 4.5 (a) tells that there are 24 LM_A mines in the data. The ROCs are calculated over the frames with mines only. Figures 4.5 shows ROCs calculated for surface mines. Figure 4.6 shows the ROCs calculated for the buried mines.



(a) LM_A (b) MP_A

(c) LP_B (d) SM_A

Figure 4.5. ROC Curves for Different Surface Mines.



(a) LM_A_B (b) MP_A_B

(c) LM_C_B (d) LP_D_B

Figure 4.6. ROC Curves for Different Buried Mines.

The observations of this section follow from the discussion in section 4.6. As can be seen from the mine frames, the surface mines produce a high contrast signature. The target mask used in RX [23], [35] covers the signature very well and produces a high value of the signal strength. The background is generally homogeneous and therefore the clutter variance is low. Thus, the RX anomaly detector performs very well for the surface mines. Because of the high contrast signature, the surface mine pixels are well separated from the background classes. Therefore SEM-based detectors perform well for these mines as well. On the other hand, the detector based on multi-class RX detects lots of

false alarms because the detector operates over a window and therefore detects local anomalies. Overall performance of the RX detector for surface mines is better than (or comparable to) that of the EM-based detectors because of a relatively homogeneous background.

The buried mines have a very weak signature that is not well separated from the background. Due to this weak signature, the signal strength is very low and for this reason the performance of the detectors is poor for the buried mines. In case of the SEM-based detectors, the class containing buried mines is not well separated from the background class and there are many regions in the background that are in the same class as the mine area. For example, on observing parts (a) and (b) of Figure 4.4 it can be seen that the regions of dark soil belong to the same class that contains buried mines. Thus, the performance of the SEM-based detectors also goes down.

It is clear from the mine signatures of the buried mines that a bigger target radius is required to capture the mine signatures of these mines. In case of RX, a large clutter radius introduces non-homogeneities of the background which increases the clutter variance and thus reduces the detection statistics. However, the SEM-based detectors do not suffer from the disadvantages of the higher mask size because target pixels are evaluated against a set of background class. The mask size can be increased to accommodate the complete mine signature and correspondingly larger background area can be used to evaluate the detection statistics.

Thus, in the case of buried mines, the performance of the SEM detectors is likely to be better than the RX anomaly detector in couple of cases. In the above results for the case of LM_A_B mines (Figure 4.6 (a)) SEM_MCRX is better performing than RX. In case of MP_A_B mines (Figure 4.6 (b)) SEM_PM performs better than RX. Overall, the results are comparable.

4.7. CONCLUSIONS

In this section, the problem of anomaly detection using image segmentation has been presented. The various SEM-based anomaly detectors that use the concept of the anomaly class to separate the anomalies from the background have also been presented. The SEM-based anomaly detectors are compared with the RX anomaly detector. The performance of the surface and buried mines has also been analyzed for these detectors.

5. MODELING OF DETECTION STATISTICS USING EM

5.1. OVERVIEW

In the previous section, segmentation of the background data into distinct classes was discussed. This section shows that the EM algorithm can also be used to model the output of detection statistics such as RX. The modeling of the detection statistic is very important since it represents the spatial correlations and the local non-homogeneities in the data. Also in order to compare and quantify the performance of an anomaly detector such as RX over different terrain, it is necessary to model the detection statistic into probabilistic models. This modeling also helps in the adaptive Constant False Alarm Rate (CFAR) threshold selection to locate potential targets as discussed in Section 6.

Various parametric distributions such as Beta and Gamma distributions are used for modeling the detection statistic. These distributions have been used in modeling due to their modeling flexibility. Once the background data has been modeled with various mixture models, the model that fits the data most accurately can be determined. This is done by using various test statistics. In the end, the performance of various mixture models is compared based on the test statistics.

5.2. THE RX STATISTIC

The RX test statistic for a single-band image data has been discussed in Section 4.4. The RX statistic follows an F-distribution for clutter and is given by:

$$r_x = N_T \left(\frac{\left(\frac{1}{N_T} \sum_{j \in W_T} I_j \right)^2}{\frac{1}{N_C} \sum_{i \in W_C} I_i^2} \right) = \left(\frac{N_C}{1} \right) \frac{\left(\frac{\sqrt{N_T}}{N_T} \sum_{j \in W_T} I_j \right)^2}{\left(\sum_{i \in W_C} I_i^2 \right)}. \tag{5.1}$$

The above equation is of the form, $\left(\frac{v_2}{v_1}\right)\left(\frac{X_1}{X_2}\right)$, where X_1 and X_2 are the Chi-Square variables with degrees of freedom $v_1 = 1$ and $v_2 = N_C$. Let, $r = \frac{X_1}{X_2} = \frac{r_X}{N_C}$.

It has been shown in [46] and [47] that the output of the RX anomaly detector can be modeled by a Gamma mixture model. This thesis uses the two parameter Gamma mixture model to model the detection statistic, $'r'$. The Gamma model used is given by:

$$G(r:\lambda,k) = \frac{\lambda^k r^{k-1} e^{-\lambda r}}{\Gamma(k)} \quad , \quad 0 \le r < \infty \; ; \; k,\lambda > 0 \, , \tag{5.2}$$

where, $\Gamma(k)$ is the complete Gamma function.

The Gamma model has received widespread importance in the modeling because it represents the distribution of sum of squares of independent normal variables. It is used extensively in the modeling of the SAR data [27]. One of the most important models of the SAR data is the multiplicative model. Here the SAR data, '*Z*,' is modeled as the product of two independent random variables, '*X*' and '*Y*.' Therefore:

$$Z = X.Y \tag{5.3}$$

The random field '*X*' models the backscatter that depends on the area each pixel belongs to, while '*Y*' takes into account the fact that the SAR images are the result of a coherent imaging system [26]. The Gamma distributions have been proposed for '*X*' and '*Y*' because of their excellent modeling capability of heterogeneous areas such as forests. The modeling of SAR images using the multiplicative model of Gamma distributions is also done in [28]. Gamma mixture models have been used for target recognition by Webb [32], who uses the EM algorithm to estimate the parameters of a Gamma mixture model. The Gamma mixture model approach has also been used for target recognition in [33].

Now consider the following transformation on '*r*':

$$x = \frac{r}{r+1}, \qquad r \in [0,\infty), \qquad x \in [0,1]. \tag{5.4}$$

With this transformation, the probability density function for the test function value ('*x*') under null hypothesis is a Beta density function given by [45]:

$$f\left(x \mid H_0\right) = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-J}{2}\right)\Gamma\left(\frac{J}{2}\right)}(1-x)^{\left(\frac{N-J-2}{2}\right)} x^{\left(\frac{J-2}{2}\right)}; 0 < x < 1. \tag{5.5}$$

Here 'N' denotes the number of samples and 'J' denotes the dimension of the image, which is one for a single-band case. In standard form the expression in Equation (5.3) can be written as:

$$B(x:\gamma,\eta) = \frac{\Gamma(\gamma+\eta)}{\Gamma(\gamma)\Gamma(\eta)} x^{(\gamma-1)}(1-x)^{(\eta-1)}, \quad \gamma > 0,\ \eta > 0 \text{ and } 0 < x < 1, \tag{5.6}$$

where $\gamma = \frac{J}{2}$ *and* $\eta = \frac{N-J}{2}$. (5.7)

Another widely used Beta model in literature, [11], is the so called three parameter Beta given by:

$$B(x:\gamma,\eta,\lambda) = \frac{\lambda^{\gamma} x^{\gamma-1}(1-x)^{\eta-1}}{B(\gamma,\eta)[1-(1-\lambda)x]^{\gamma+\eta}}, \quad 0 \le x \le 1; \gamma,\eta,\lambda > 0, \tag{5.8}$$

where, $B(\gamma,\eta)$ is the complete Beta function.

Equation (5.6) reduces to the two parameter Beta distribution for λ = 1. The parameter, 'λ,' allows the three parameter Beta distribution to have a wider variety of shapes than the standard two parameter Beta. This is an important property in modeling because the three parameter Beta has an additional flexibility. For example, if in the two parameter Beta, $\gamma = \eta$, then the distribution is symmetric with a mean value at 0.5. However, the three parameter Beta can be skewed depending on 'λ' because the kurtosis, mode and skewness depend on 'λ.'

5.3. THE NON-MAX SUPPRESSION

The output of the RX algorithm is basically the "signal-to-clutter" image. This list of detections consists of the coordinates of the potential targets along with the RX test statistic value. This is then subjected to the non-max suppression in order to get a list of targets that have been highlighted by the anomaly detector. Non-maximal suppression is a processing algorithm that suppresses (makes zero) all the targets in a pre-defined *R*-pixel radius neighborhood except the local maximum.

Let $f(x|H_0)$ be the probability density function for the RX clutter statistic x under null hypothesis after non-max suppression. Let the function $g(l)$ represent the mapping function performed by the non-max operation and 'W_R' be the *R*-pixel neighborhood. Thus,

$$g(l) = \begin{cases} 1 & if \quad x_l = X_l \equiv \text{local maximum in } \mathrm{W_R} \\ 0 & elsewhere \end{cases} \tag{5.9}$$

The probability density function used to model the background clutter statistics after non-max suppression is given as:

$$f(x \mid g(l) = 1) = \frac{f(x \mid H_0).e^{-N(1-F(x))}}{\int_0^\infty f(x \mid H_0).e^{-N(1-F(x))}dx}, \tag{5.10}$$

where $f(x \mid H_0)$ is the probability due to RX statistics under null hypothesis, $N = \theta A$ is the expected number of independent targets present in the neighborhood 'W_R', 'A' is the area of the neighborhood 'W_R' and 'θ' is the density of the potential targets. Also,

$$F(x) = CDF\ of\ x = f(\bar{x} < x \mid H_0) = \int_0^x f(\bar{x} \mid H_0) d\bar{x}\,, \tag{5.11}$$

where *CDF* is the Cumulative Distribution Function.

Thus for example in the case of three parameter Beta distribution,

$$f_3\left(x \middle| H_0\right) = B(x:\gamma,\eta,\lambda) = \frac{\lambda^{\gamma} x^{\gamma-1}(1-x)^{\eta-1}}{B(\gamma,\eta)\left[1-(1-\lambda)x\right]^{\gamma+\eta}} \quad , \quad 0 \le x \le 1; \gamma,\eta,\lambda > 0\,. \tag{5.12}$$

Then, the modified three parameter Beta, after the non-max suppression has the distribution given by:

$$B(x:\gamma,\eta,\lambda,N) = \frac{f_3(x \mid H_0).e^{-N(1-F_3(x))}}{\int_0^{\infty} f_3(x \mid H_0).e^{-N(1-F_3(x))} dx}\,, \tag{5.13}$$

where, $F_3(x)$ is the CDF of $f_3(x \mid H_0)$.

In this thesis, $B(x:\gamma,\eta)$, $B(x:\gamma,\eta,\lambda)$ and $B(x:\gamma,\eta,\lambda,N)$ are the Beta distributions used to model the detection statistic. The Gamma model used for modeling is $G(r:\lambda,k)$.

5.4. TEST STATISTICS TO MEASURE GOODNESS OF FIT

5.4.1 Various Test Statistics. A list of various test statistics that measure goodness of fit is presented along with their description in Appendix—B. These tests include the following:

(i) Chi-Square Test
(ii) Cramer Von-Mises (CVM) Test
(iii) Kolmogorov-Smirnov (KS) or Kuiper Test

The main limitation of the CVM, KS or Kuiper tests is that in literature the critical values for the test statistics are given only for some given distributions, such as Exponential, Weibull and Normal. Thus these test statistics cannot be used for applications where different distributions are used. If the distribution is completely specified, then these test perform well and have a high power. But if the parameters are estimated from the data, then the power of these tests reduces. This is because new critical values need to be formulated and these must be obtained for the specific parametric form that is to be tested.

On the other hand, this does not pose a problem for the Chi-Square test because the number of degrees of freedom is adjusted in accordance with the parameters that are estimated from the data. It is because of the flexibility of the Chi-Square test that it was used for testing the modeling results. The specifics of this test are discussed in the following section.

5.4.2. The Chi-Square Test. The Chi-Square test is described in detail in Appendix—B. This section briefly presents the test and the way it is applied in this study. For the Chi-Square test, the samples are grouped in bins in such a way that there are at least a certain fixed number of samples per bin. Also, when the test statistic is evaluated for the frame '*k*', the samples (RX values) from the three frames, '*k-1,*' '*k*' and '*k+1*' are considered.

To evaluate the modeling performance of the background, it is necessary to make sure that the samples are not anomalies. There are about 400 samples per frame, and about 10 samples per frame are considered anomalous (i.e. these don't represent the background). Therefore, a threshold is set so as to remove the top 30 samples (assuming that they are anomalous), and so the top 30 samples are not considered while evaluating the test performance. Depending on the number of parameters that are to be estimated, the degrees of freedom is calculated. The test statistic is then calculated for a particular frame. After calculating the test statistic and the degrees of freedom for a particular frame, the threshold is found for the given confidence level. The confidence level generally taken is in the range of 0.90 to 0.95. For the current results, a confidence level of 0.95 is used.

5.4.3. Performance Evaluation. After calculating the threshold, one of the following two cases arise:

Case – 1 (Test Statistic < Threshold): In this case one says that the frame *passes* the test. For a test at a confidence level of 0.95 it can be said with 95% confidence level that the background represented by the frame *can* be modeled by the given distribution.

Case – 2 (Test Statistic >= Threshold): In this case one says that the frame *fails* the test. For a test at a confidence level of 0.95 it can be said with 95% confidence level that the background represented by the frame *cannot* be modeled by the given distribution.

Once the decision has been made for the frame, the test is performed on all the frames of the given dataset and the percentage of frames that have passed the test is found. Thus on an average it is expected that 95% of the frames will pass the test and 5% would fail. If the fail percentage is much higher than 5% one can say that the modeling is bad otherwise the modeling is good. It is clear from the above procedure that the pass percentage is a good performance measure to judge the modeling ability of the given distribution to model the background represented by the frames.

5.5 THE KULLBACK-LEIBLER (KL) DIVERGENCE

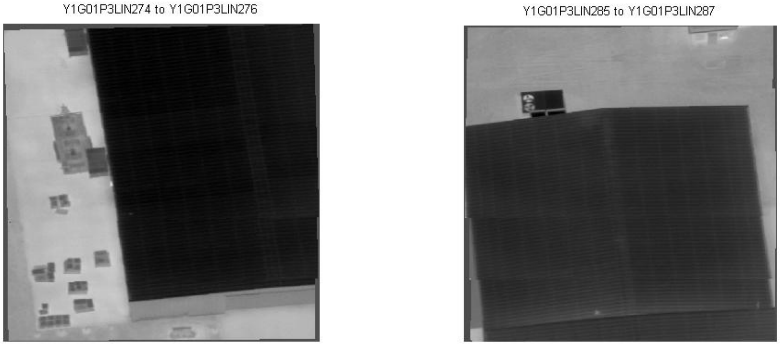
The KL divergence [30] is a way to measure the closeness between the two pdfs. It can also be seen as a measure of relative entropy between two pdfs [31]. Generally the KL divergence is used to measure the closeness between the actual distribution and the modeled distribution. The closer the pdfs are to each other, the smaller is the divergence. The KL divergence between two pdf's $p(x)$ and $q(x)$ is given as:

$$D(p(x),q(x)) = \int_x p(x)\log\left[\frac{p(x)}{q(x)}\right]dx + \int_x q(x)\log\left[\frac{q(x)}{p(x)}\right]dx. \tag{5.14}$$

Clearly if the pdfs are equal, then the logarithmic terms become zero and the divergence between them is zero. Also, the divergence between the two pdfs is always non-negative. In all the modeling results that are shown, the divergence between the modeling distribution and the sample histogram is given in the legend of the figure. In the legend, the divergence value is given in brackets, next to the modeling distribution.

5.6. REMOVAL OF BAD DATA

Certain frames in the data don't represent the background. Therefore, these frames are not considered for the test. Some of these frames are shown in Figure 5.1.



(a) Frame Nos. 274-276 (b) Frame Nos. 285-287

Figure 5.1. Frames Showing Data Not Representing the Background.

5.7. PARAMETER ESTIMATION USING LEAST SQUARES

The results of EM estimation are compared to another method of estimating the parameters of a given distribution. This method is the method of Least Squares (LS) as discussed in [23], where it is used for estimating the parameters of the modified three parameter Beta distribution. In this method, to estimate the values of the parameters of the modified three parameter Beta distribution, an arbitrary sample space for the set of parameters $(\gamma, \eta, \lambda, N)$ is assumed. This sample space forms the initial search space and a brute force search is done across the sample space, to identify a set of values for γ, η, λ and *N* that is at a minimum distance from the measured pdf. In order to do this, for every set of parameters, the error between modeled pdf and the observed pdf is calculated by taking the Euclidean distance between the two pdfs. Amongst the error values obtained, the minimum error is determined and the set of parameters $(\gamma, \eta, \lambda, N)$ corresponding to this minimum error is taken as the first estimate of the parameters. After this coarse estimate of the parameters is obtained, a finer resolution sample space around the coarse estimates is assumed. The method of least squares is reiterated. This second iteration results in a finer resolution on the estimated parameters.

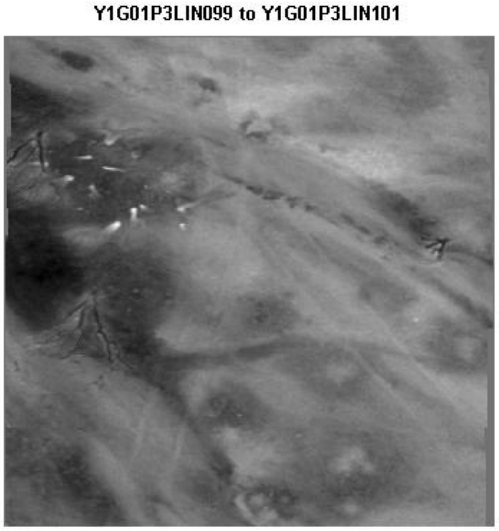
In the next section, the modeling results of the parameter estimation techniques using the EM algorithm and the method of the Least Squares are discussed.

5.8. MODELING RESULTS

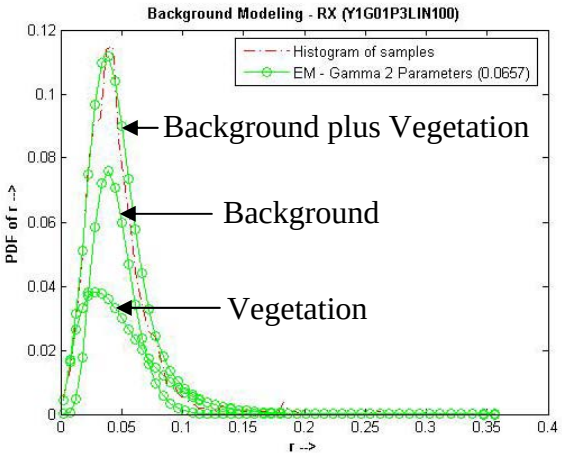
This section includes a visual inspection to get an idea of the fit of the distributions with the data. A more rigorous quantitative analysis has been done after that when the results for the Chi-Square test have been reported on the given data. As mentioned before, for evaluating the performance for frame '*k*', the samples from the three frames, '*k-1,*' '*k*' and '*k+1*', are considered. Therefore, while displaying the modeling results, the images have been registered to show all three consecutive frames.

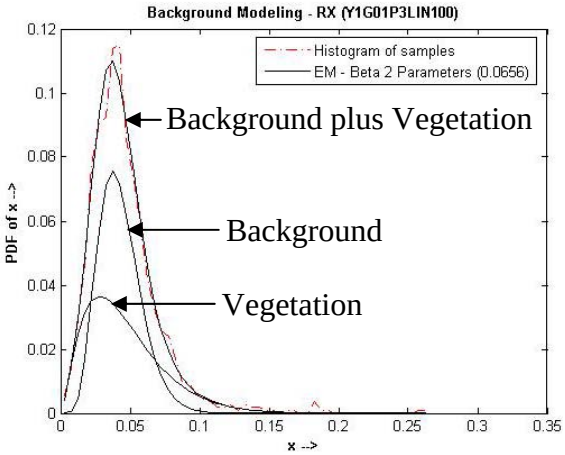
Figure 5.2 shows a sample modeling result for the Beta and Gamma distribution. It shows the pdf of the two constituent classes along with the pdf of their sum. It is to be

noted that the Gamma distribution is plotted on the '*r*' domain whereas the Beta distribution is plotted on the '*x*' domain. In part (a) of the figure, it can be seen that there are two prominent regions in this image. One region is the background where as the other region is some vegetation like trees. The background containing the two regions has been modeled by the two classes as shown in parts (b) and (c).



(a) Registered Image—May 2003 Data (Nighttime)





(b) Gamma 2-Parameters (c) Beta 2-Parameters

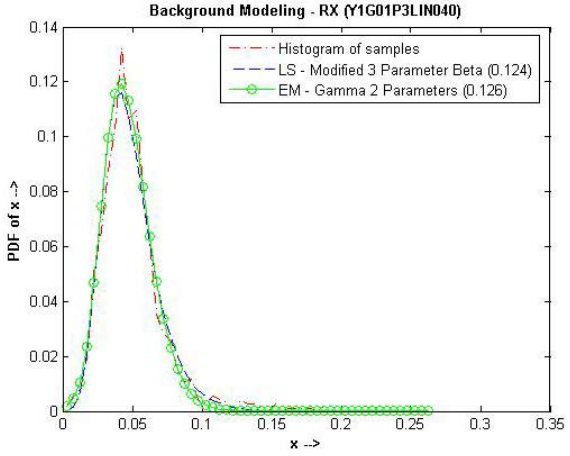
Figure 5.2. Background Modeling Showing the Mixture of Two Classes.

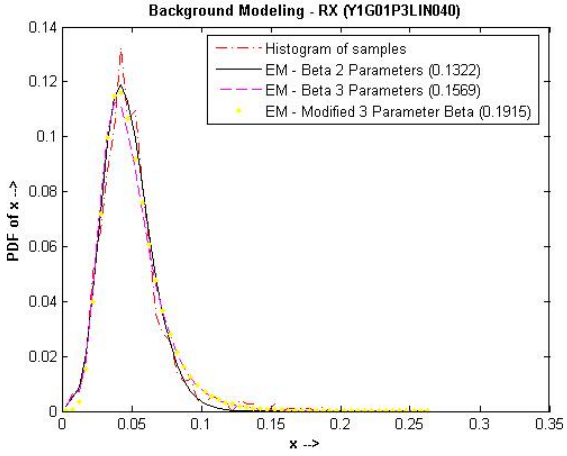
Figures 5.3 (a) and 5.4 (a) show the registered image of the frames under consideration for May 2003 nighttime data. The dataset name for the result is mentioned at the top of every figure. In all the figures showing the modeling results of both Beta and Gamma distributions, the histogram are drawn in common 'x' domain ($x \in [0,1]$) for easy comparison. Part (b) of Figures 5.3 and 5.4 shows the fit of modified three parameter Beta distribution (Estimated using Least Squares) and the two parameter Gamma distribution (Estimated using EM), with the histogram of the samples. Part (c) of Figures 5.3 and 5.4 shows the fit of the two parameter Beta, three parameter Beta and the modified three parameter Beta distributions (all estimated using EM), with the histogram of the samples.

The pass percentage of various models for RX samples for the dataset, Y1G01P3LIN, is reported in Table 5.1. It can be clearly seen that the performance of the two parameter Gamma and Beta mixture models is the best at 96.06% and 95.14% respectively. This is followed by the performance of the modified three parameter Beta model obtained using least-squares that has a pass percentage of 94.68%. It can be noted that these percentages are quite close to 95% as was expected at a confidence level of 95%.

As mentioned in Section 2.7, as the number of parameters to be estimated increases, the log-likelihood surface is no longer steep. The surface becomes flat, and all the parameters do not converge to a steady value. In fact, the parameters tend to converge to a range of values. Because the log-likelihood surface is flat, the convergence is not very stable, and the performance deteriorates as the number of parameters increase. This is the main reason for the poor performance of the modified three parameter Beta model obtained using EM.

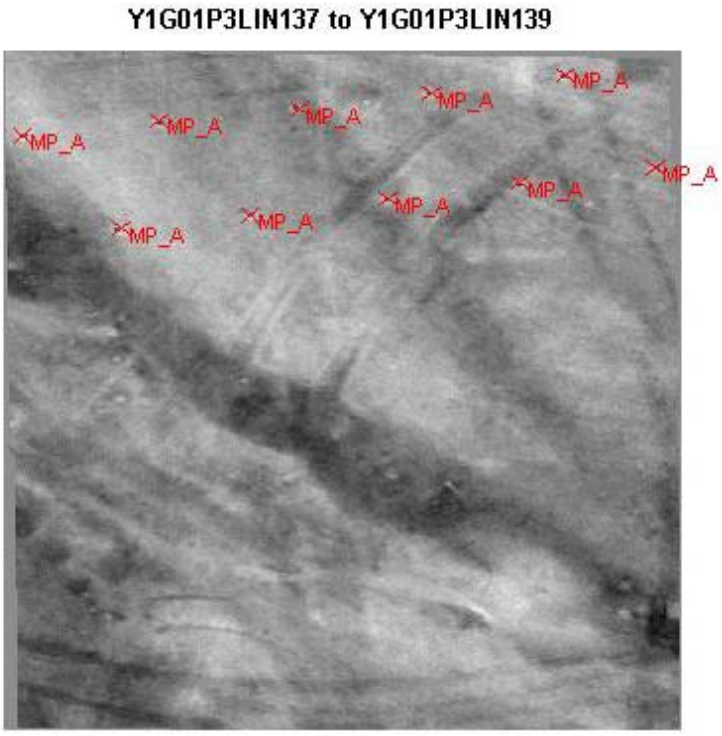
(a) Background Only Registered Image—May 2003 Data (Nighttime)



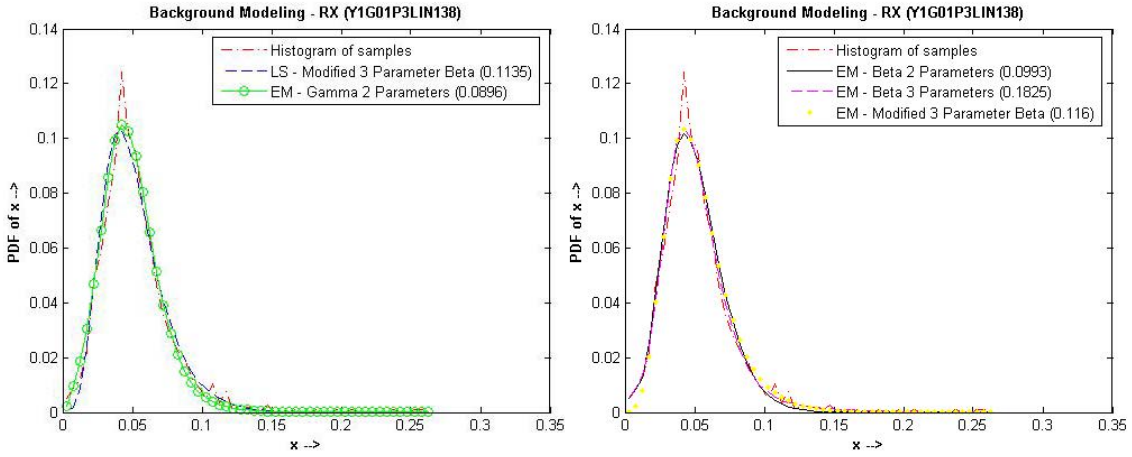


(b) Least-Squares and Gamma 2-Parameters (c) Beta 2, 3 and Modified 3-Parameters

Figure 5.3. Background Modeling for Different Distributions—Nighttime.



(a) Registered Image With Mines—May 2003 Data (Nighttime)



(b) Least-Squares and Gamma 2-Parameters (c) Beta 2, 3 and Modified 3-Parameters

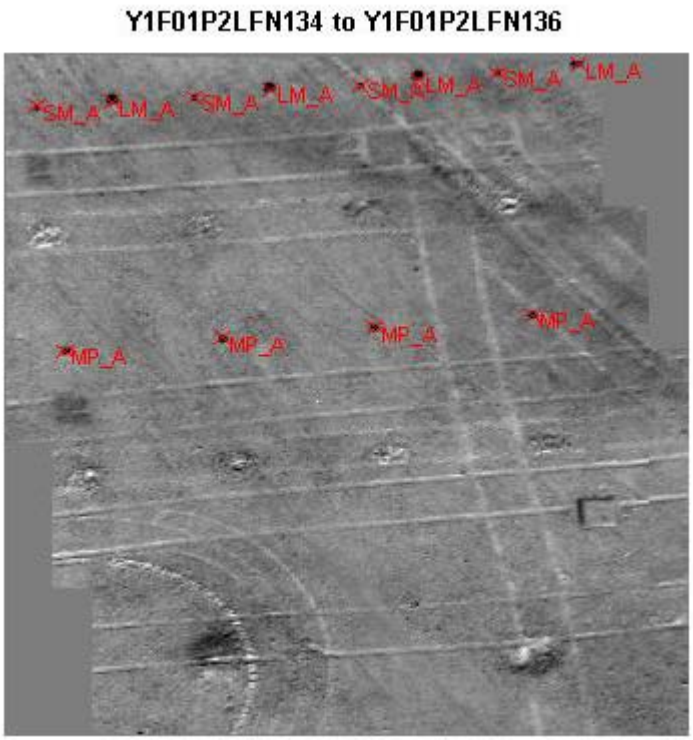
Figure 5.4. Background Modeling With Mines for Different Distributions—Nighttime.

Table 5.1. Pass Percentages of Various Mixture Models.

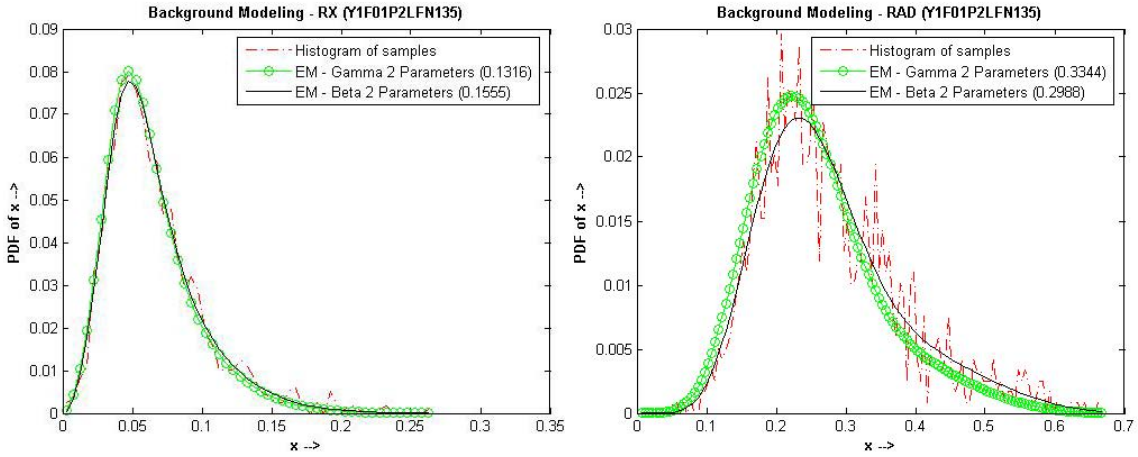
Modeling Distribution	**Pass Percentages (%) for 95% confidence level for RX**
Gamma 2-Parameters	96.06
Beta 2-Parameters	95.14
Beta 3-Parameters	90.05
Modified 3 Parameter Beta	68.75
Modified 3 Parameter Beta (Least Squares)	94.68

After considering the performances of different models, the further analysis was carried extensively on various datasets using the two parameter Beta and Gamma distributions only, because of their superior performance. The analysis was done for both RX and Radial Anomaly Detector (RAD) [35] samples. Figures 5.5-5.10 show the modeling results for these two distributions for the May 2003 data (daytime and nighttime) and October 2002 data (daytime). Part (a) of the Figures 5.5-5.10 shows the registered image of the frames. Part (b) of the Figures 5.5-5.10 shows the modeling results of the two parameter Gamma and Beta distributions for RX samples. Part (c) of the Figures 5.5-5.10 shows the results of the same distributions for RAD samples.

It can be seen that the histogram of the RAD samples is noisier than that of the RX samples. Part of the reason is that in case of RX, the samples are distributed in a range of 0-0.2, whereas in case of RAD, the samples are distributed in a wider range of 0-0.6, which makes the observed histogram for RAD seem noisier.

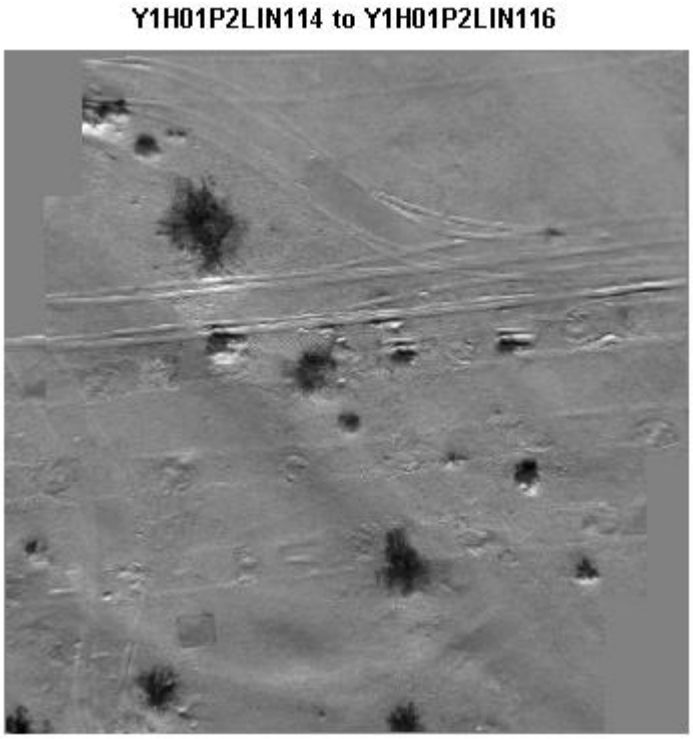


(a) Registered Image—May 2003 Data (Daytime)

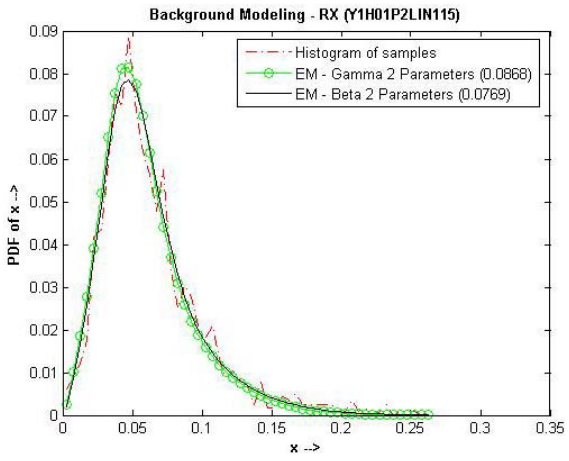


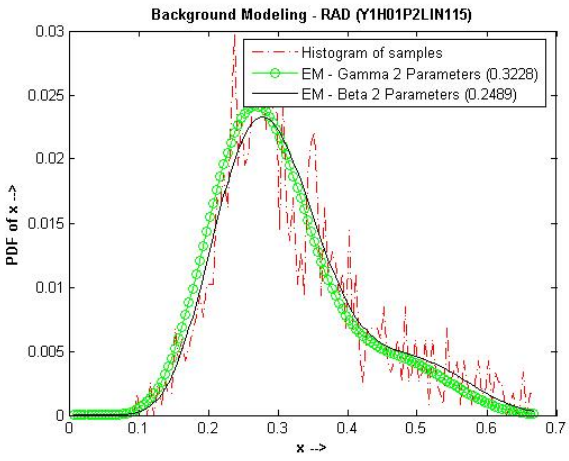
(b) RX (c) RAD

Figure 5.5. Background Modeling for May 2003 Data for Daytime.



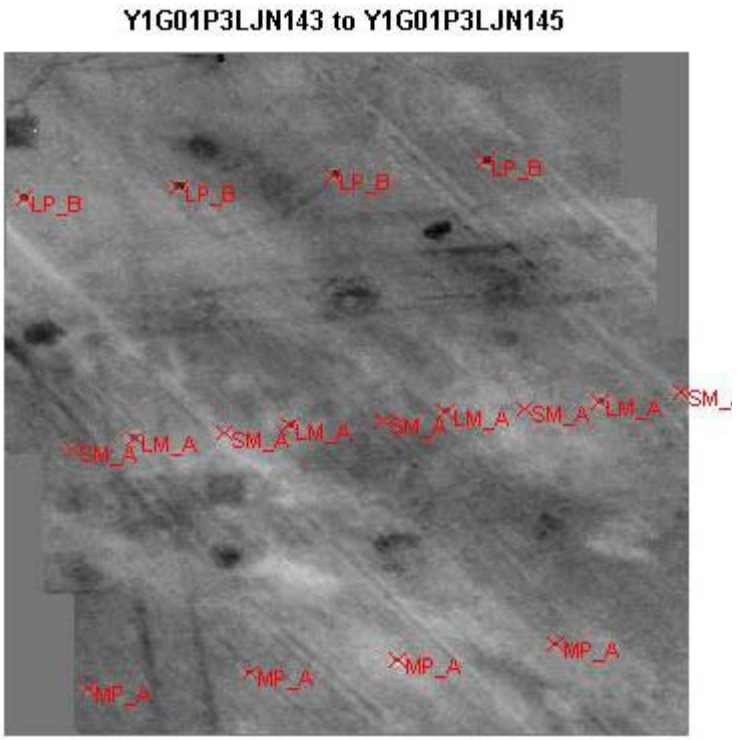
(a) Registered Image—May 2003 Data (Daytime)



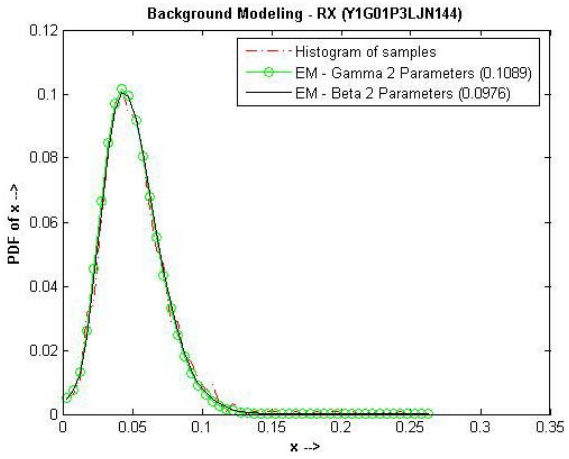


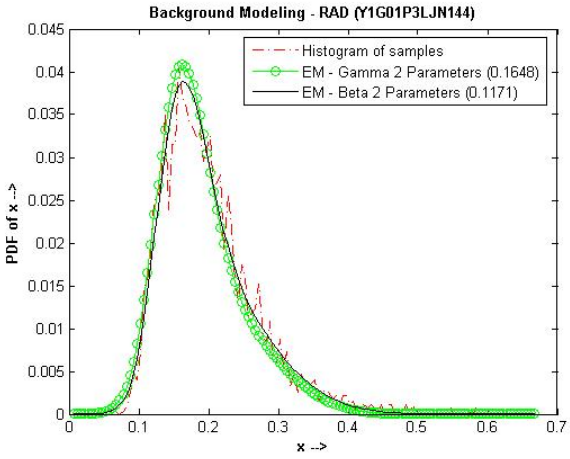
(b) RX (c) RAD

Figure 5.6. Background Modeling for May 2003 Data for Daytime.



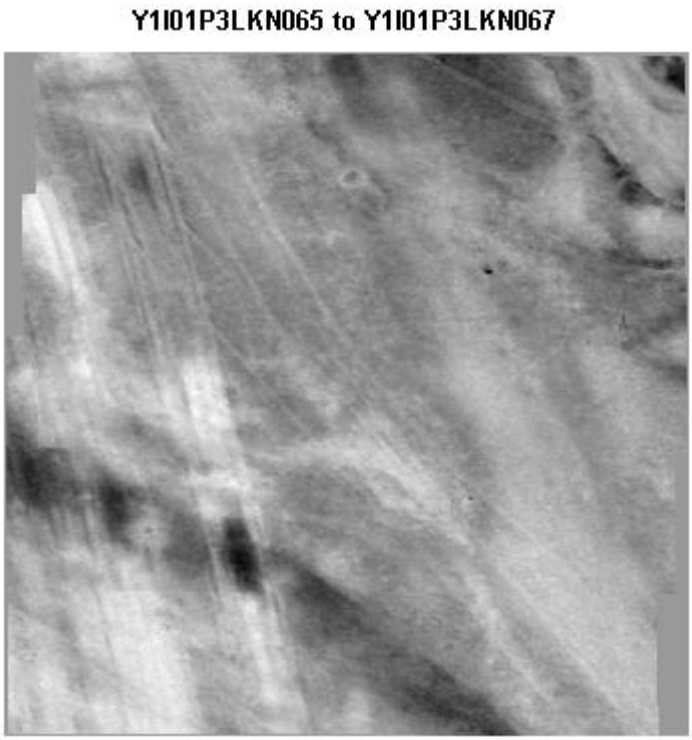
(a) Registered Image—May 2003 Data (Nighttime)



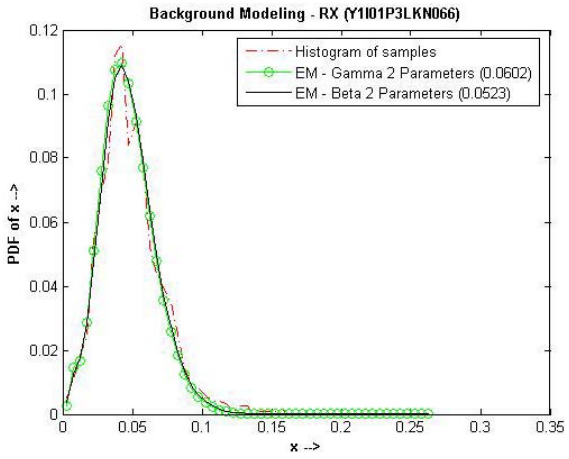


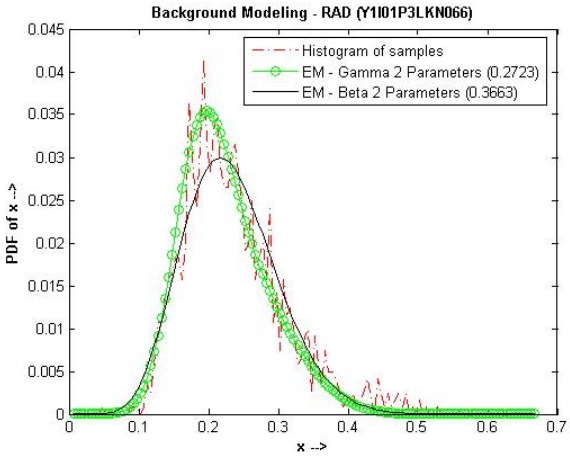
(b) RX (c) RAD

Figure 5.7. Background Modeling for May 2003 Data for Nighttime.



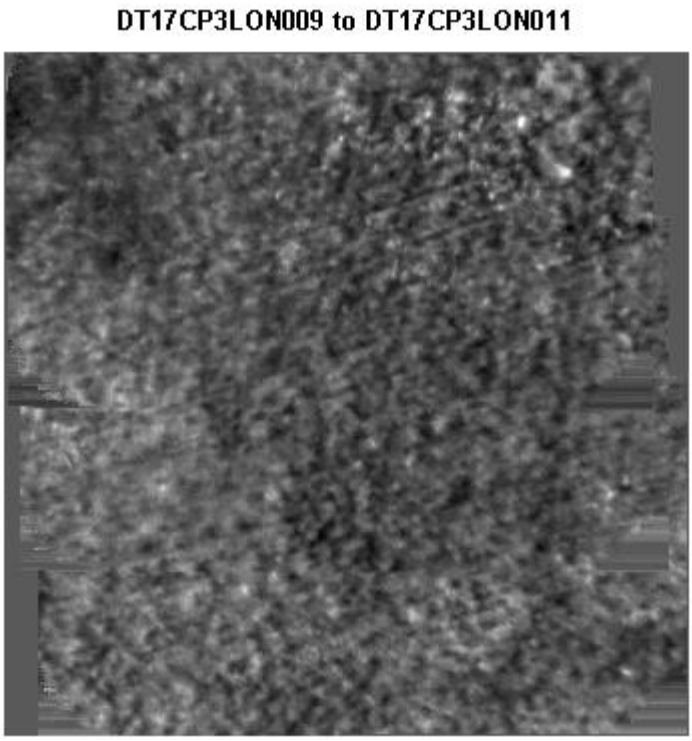
(a) Registered Image—May 2003 Data (Nighttime)



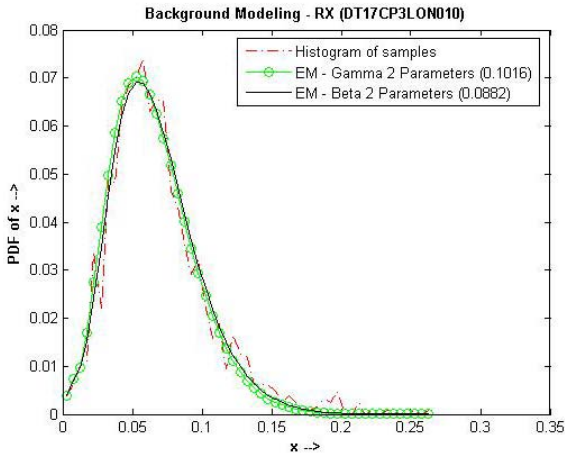


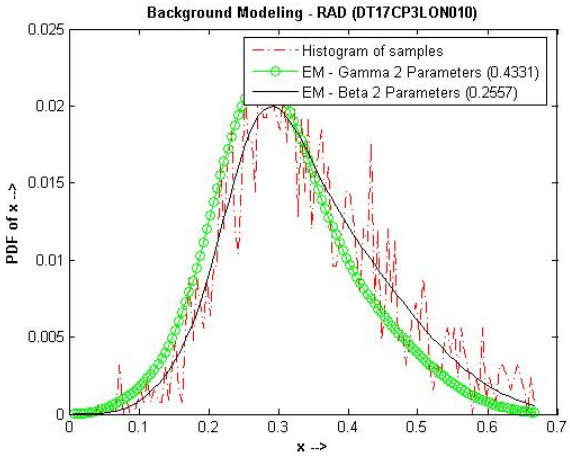
(b) RX (c) RAD

Figure 5.8. Background Modeling for May 2003 Data for Nighttime.



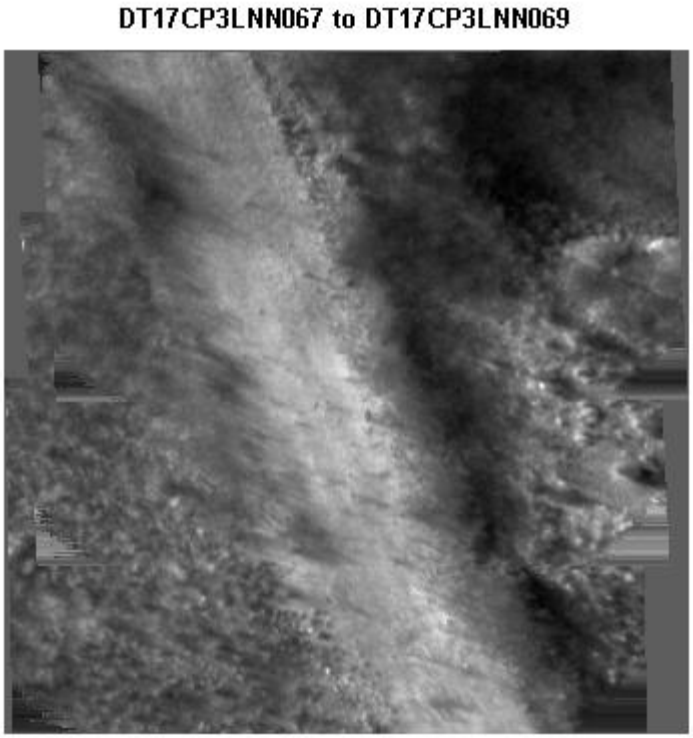
(a) Registered Image—October 2002 Data



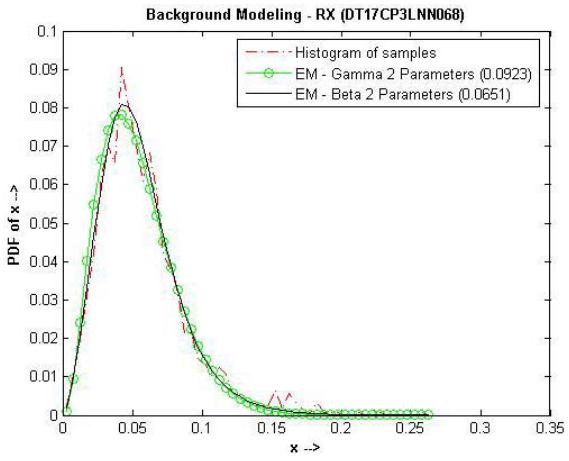


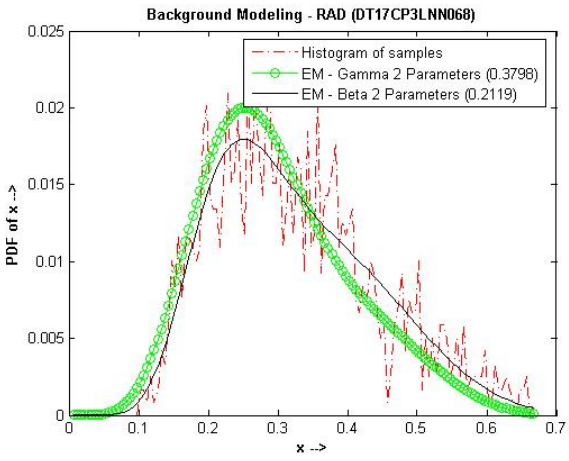
(b) RX (c) RAD

Figure 5.9. Background Modeling for October 2002 Data.



(a) Registered Image—October 2002 Data





(b) RX (c) RAD

Figure 5.10. Background Modeling for October 2002 Data.

Because the histogram of the RX samples is smoother in comparison to that of the RAD samples, the average divergence between the actual and modeled pdf in case of RAD samples is more (≈ 0.3) than the average divergence in case of the RX samples (≈ 0.1). From the results it can be seen that the performance of the two models is quite comparable for RX and RAD detection statistics. The modeling performance of the two mixture models is nearly the same for the two types of data in the case of RX. However, in the case of RAD, the modeling performance of the Gamma mixture model goes slightly down for the October 2002 data. This can also be seen from Figure 5.9 (c), where the modeling by the two parameter Gamma model is not good for RAD samples. Tables 5.2-5.7 show the pass percentage of these distributions for RX and RAD samples.

Table 5.2. Pass Percentages for RX—May 2003 Daytime.

May 2003 (Day Time)	**Pass Percentages (%) for 95% confidence level, RX**	
	Beta 2-Parameters	**Gamma 2-Parameters**
Y1F01P2LFN	95.36	93.82
Y1F01P2LDN	90.53	94.71
Y1F01P2LEN	92.53	93.41
Y1H01P2LIN	90.64	93.15

Table 5.3. Pass Percentages for RX—May 2003 Nighttime.

May 2003 (Night Time)	**Pass Percentages (%) for 95% confidence level, RX**	
	Beta 2-Parameters	**Gamma 2-Parameters**
Y1I01P3LKN	94.48	94.70
Y1G01P3LIN	94.44	94.21
Y1G01P3LJN	92.66	94.66
Y1G01P3LKN	94.97	91.10

Table 5.4. Pass Percentages for RX—October 2002.

Oct 2002	Pass Percentages (%) for 95% confidence level, RX	
	Beta 2-Parameters	**Gamma 2-Parameters**
DT17BP2LEN	95.10	97.20
DT17CP3LNN	99.30	100.00
DT17CP3LON	95.80	98.60
DT17AP1LAN	88.81	93.00

Table 5.5. Pass Percentages for RAD—May 2003 Daytime.

May 2003 (Day Time)	Pass Percentages (%) for 95% confidence level, RAD	
	Beta 2-Parameters	**Gamma 2-Parameters**
Y1F01P2LFN	99.34	98.89
Y1F01P2LDN	98.89	95.60
Y1F01P2LEN	98.68	97.80
Y1H01P2LIN	99.54	98.40

Table 5.6. Pass Percentages for RAD—May 2003 Nighttime.

May 2003 (Night Time)	Pass Percentages (%) for 95% confidence level, RAD	
	Beta 2-Parameters	**Gamma 2-Parameters**
Y1I01P3LKN	98.23	98.23
Y1G01P3LIN	96.53	97.90
Y1G01P3LJN	94.00	97.33
Y1G01P3LKN	94.97	98.63

Table 5.7. Pass Percentages for RAD—October 2002.

Oct 2002	Pass Percentages (%) for 95% confidence level, RAD	
	Beta 2-Parameters	**Gamma 2-Parameters**
DT17BP2LEN	99.30	94.40
DT17CP3LNN	100.00	92.28
DT17CP3LON	97.20	85.31
DT17AP1LAN	100.00	89.44

It can be recalled from section 5.4.3 that for good modeling it is expected that on an average 95% of the frames should pass the test. From the results in Tables 5.2-5.7, it can be observed that the pass percentages of the two parameter Beta and Gamma model are close to 95%. Therefore the modeling by the above models can be considered to be good. Also by observing the overall performance across the samples and the data, it is evident that the performance of the Beta mixture model is comparable to that of the Gamma mixture model for the May 2003 data. However for the October 2002 data, the performance of the Beta model is better than that of the Gamma model for RAD samples.

On observing the October 2002 data, it can be seen that many frames have images of vegetation such as trees and bushes. In fact, a good portion of the dataset is covered by frames having images of trees. Thus, these frames don’t represent a homogeneous background. This, along with the fact that the RAD samples are quite noisy, brings down the performance of the Gamma model for this data. The fit of the Beta model is good in spite of the non-homogeneity of the data. This shows that the Beta model is robust in modeling these non-homogeneous areas.

5.9. FRAME CHARACTERIZATION

This section shows how to characterize a frame based on the mixing proportions of the classes. This gives some information about the homogeneity of the data. The characterization based on the detection statistic is important since it represents the spatial

correlations and the local non-homogeneities in the data. Therefore the detection statistic represents the true nature of the background. After investigating the class coefficients of the frames, Table 5.8 is obtained. The data has been modeled by two classes by the Gamma mixture model with two parameters.

Table 5.8. Frame Distribution.

Distribution of the Coefficient of the Major Class Across Frames Modeled by Gamma 2-Parameters using Two Classes (for RX)			
Coefficient range (Major Class)	**Run Names**		
	DT17CP2LMN (Total Frames/Frames Failed)	**Y1F01P2LFN (Total Frames/Frames Failed)**	**Y1G01P3LIN (Total Frames/Frames Failed)**
0.98 to 1.00	68/6	139/18	5/1
0.95 to 0.98	12/1	21/2	7/0
0.90 to 0.95	9/0	16/1	35/2
0.80 to 0.90	12/0	20/0	147/8
0.70 to 0.80	21/0	48/1	148/8
0.60 to 0.70	12/0	79/0	62/5
0.50 to 0.60	9/0	130/6	28/1
Total	**143/7**	**453/28**	**432/25**

Based on the Table 5.8, three main categories are considered. They are as follows:

Category-1 (1 > Major class coefficient > 0.98): This category corresponds to the frames that essentially comprise the background. The background represented by these frames can be called benign in comparison to the background represented by the frames from the other categories. In the case of the dataset, Y1G01P3LIN, most of the frames represent a background consisting of some type of vegetation, i.e. grass, bushes and trees. Due to

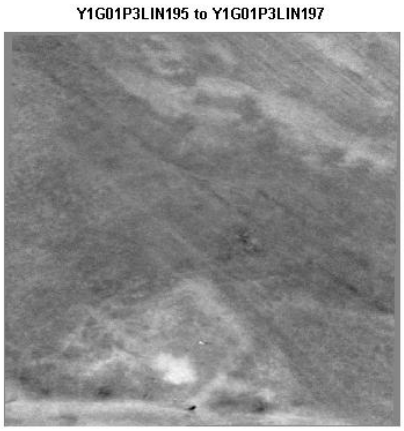
this, a majority of frames have a major coefficient in the range of 0.7 to 0.9. In the case of the datasets DT17CP2LMN and Y1F01P2LFN, the frames mainly represent the background alone. Therefore, most of the frames belong to this category.

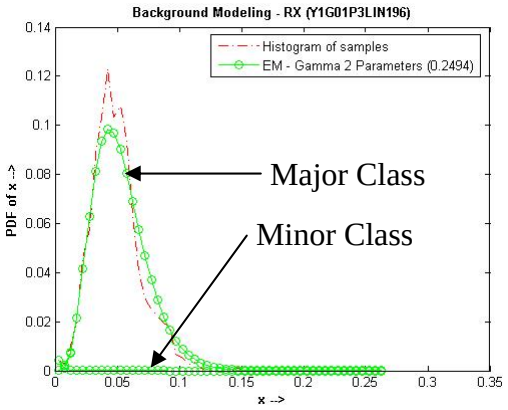
Category-2 (0.98 > Major class coefficient > 0.70): Generally, these are the frames that consist of some clusters of features that are essentially not the part of the background. For the dataset Y1G01P3LIN, the data show that most of the frames contain a background that has mixing coefficients that belong to this category. However in case of the other two datasets not many frames belong to this category.

Category-3 (0.70 > Major class coefficient > 0.50): These are the frames that are highly non-homogeneous. In these frames, the background is mainly comprised of clusters of features that are not the part of the background.

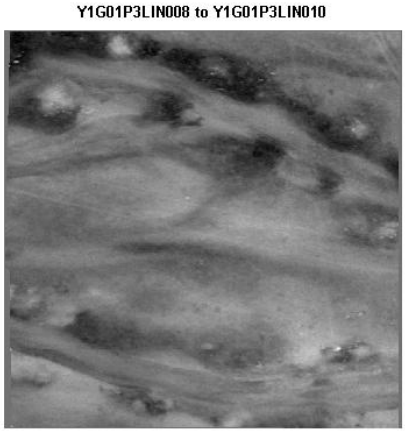
Some representative frames from these three categories are shown in Figure 5.11. Parts (a), (c) and (e) of the figure show the registered frames. Parts (b), (d) and (f) show the modeling using two parameter Gamma mixture model. The modeling is done using two classes. The pdf of the constituent classes along with their sum is also shown. The major and minor classes have been shown in the figure. The sum of the pdfs of the classes is used to model the histogram of the samples.

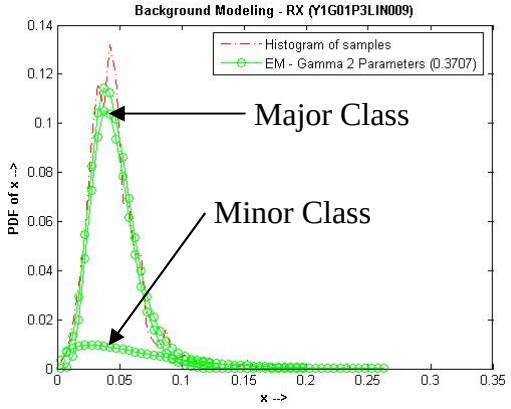
It can be clearly seen that for the second and third category, the minor class is quite prominent. For the first category the minority class has a very insignificant contribution to the overall fit.



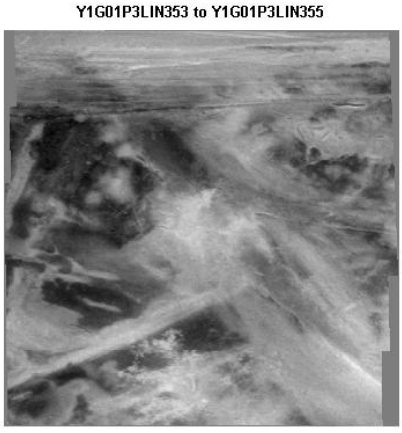


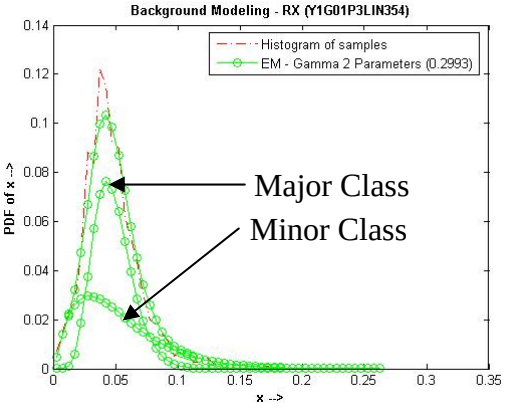
(a) Coefficient of Major Class = 0.99 (b) Modeling





(c) Coefficient of Major Class = 0.82 (d) Modeling





(e) Coefficient of Major Class = 0.58 (f) Modeling

Figure 5.11. Background Modeling for Different Categories Using Gamma Distribution.

5.10. CONCLUSIONS

This section has shown the application of the EM algorithm in modeling of the RX and RAD statistics. The modeling performance of the various mixture models was also evaluated, using the Chi-Square test. It can be clearly seen that the two parameter Beta and Gamma mixture models have the best performance. A brief characterization of the background based on the coefficient of the majority class was also done.

6. THRESHOLD SELECTION

6.1. OVERVIEW

In ATR systems [7], often there is a front-end detection stage or anomaly detection stage. This is also known as the prescreening stage. The aim of this stage is to select the targets at a given False Alarm Rate (FAR) and then pass these targets to the next stage. The threshold value is the minimum value of the detection statistic at which a particular false alarm rate is achieved. In order to detect a high percentage of targets, the prescreening stage often uses a low threshold that may result in a large number of false alarms. A large number of false alarms increase the burden on the later stages and therefore the prescreening stage should be designed in such a way so that it reduces the number of false alarms while maintaining a high probability of detection.

In many cases, the process of threshold selection simply involves selecting the highest '*T*' targets based on their detection statistic. This method of pre-specifying the number of targets per frame, though simple in a functional sense, fails to provide any insight into the nature of the background or in providing the mechanism of comparing thresholds across different datasets and anomaly detector algorithms. This is basically because the detection statistics by themselves are affected by the spatial correlations and local non-homogeneity in the image. Also, threshold selection based on desired number of targets per frame does not statistically quantify the true anomalies. Therefore, the need for a more sophisticated threshold selection scheme is evident.

This section addresses the application of background modeling in the adaptive Constant False Alarm Rate (CFAR) threshold selection. Here the distribution of the RX statistics of the background is modeled using the two parameter Beta and Gamma model along with the modified two parameter Beta model discussed in the next section. The threshold that confirms to the desired FAR is then selected by inverse mapping of the Cumulative Distribution Function (CDF) of this model. The mapping onto a probabilistic model helps select a threshold that is invariant to the background that the detector is

operating in. This mapping onto a distribution also helps map the anomaly detection statistics across different algorithms onto a common distribution, thus facilitating comparison among their statistics for sensor fusion and algorithm level fusion. This threshold which depends on the specified CFAR value, determines the number of selected detections above the threshold. This number of detections for background frames gives a good evaluation of the modeling capability of the distribution used for background modeling.

6.2. THE MODIFIED TWO PARAMETER BETA MODEL

From the discussion of Section 5.2 and 5.3, it is known that if $f(x \mid H_0)$ is the probability distribution of the RX statistics under null hypothesis, then after the non-max suppression, the probability distribution of the locally maximum detection statistics under null hypothesis becomes:

$$f(x \mid g(l) = 1) = \frac{f(x \mid H_0).e^{-N(1-F(x))}}{\int_0^{\infty} f(x \mid H_0).e^{-N(1-F(\bar{x}))}dx}. \tag{6.1}$$

For two parameter Beta model, the distribution under null hypothesis is given by:

$$f_2\left(x \middle| H_0\right) = B(\gamma, \eta), \tag{6.2}$$

where,

$$B(x:\gamma,\eta) = \frac{\Gamma(\gamma+\eta)}{\Gamma(\gamma)\Gamma(\eta)} x^{(\gamma-1)}(1-x)^{(\eta-1)}, \quad \gamma > 0,\ \eta > 0 \text{ and } 0 < x < 1. \tag{6.3}$$

Then the modified two parameter Beta distribution is given by:

$$f_3(x) = \frac{1}{\mathrm{K}} f_2(x) e^{-N(1-F_2(x))} \ , \tag{6.4}$$

where $F_2(x)$ is the cumulative distribution function of $f_2(x)$, the generalized two parameter Beta distribution. '*K*' is a normalizing constant and is given by:

$$\mathrm{K} = \int_0^1 f_2(x) e^{-N(1-F_2(x))} dx \ . \tag{6.5}$$

Parametric estimation of this modified two parameter Beta model $f_3(x)$ involves obtaining estimates 'γ°,' 'η°' and 'N°' of the three parameters 'γ,' 'η' and '*N*' for the input image. The parameters are estimated using the EM algorithm. As mentioned in the previous section, to ensure a good number of sample points for modeling, the targets from the current frame and ± 1 frame about it are used for modeling the current frame. A typical frame consists of about 380 targets, and thus about 1140 (380x3) targets are used to model every frame. This distribution is compared with the two parameter Beta and Gamma models for comparison. The update equations for the modified two parameter Beta distribution are given in appendix—A.1.4.

6.3. THRESHOLD CALCULATION FOR A GIVEN CFAR

In this section, the working procedure for calculating the threshold from the model is discussed.

Let 'N_r' and 'N_c' be the number of rows and columns in the image. The *GSD* is the Ground Sampling Distance for the airborne imagery in inches. If there are 'N_t' targets per image segment (frame), then the probability of false alarm, 'P_f' is given as:

$$P_f = \frac{(CFAR).A}{N_t} \tag{6.6}$$

where '*A*' of the image segment represents the area in meter^2 and is given as:

$$A = (N_r.N_c)(GSDx0.0254)^2 \tag{6.7}$$

Thus once the CFAR is determined, the 'P_f' gets fixed. Corresponding to this 'P_f,' the inverse mapping of the CDF, that would give the threshold, can be found as follows.

Let *F(x)* be the CDF of the modeling distribution $f(x)$, and 'P_f' be the probability of false alarms obtained from the Equation 6.6, then the threshold, '*T*,' can be obtained as:

$$T = \arg\left[F(x) = (1 - P_f)\right] = F^{-1}(1 - P_f). \tag{6.8}$$

Here, $f(x)$ can be any modeling distribution. In this thesis, two parameter Beta, two parameter Gamma and modified two parameter Beta distributions have been used for determining the threshold by inverse mapping of the CDF. In case of the Beta distribution, it is necessary to apply the reverse transformation [i.e. $\frac{T}{1-T}$] to get the correct threshold for the RX detection statistic. This transformation is not required for the Gamma distribution that is represented by the RX statistic '*r*.' Please see Section 5.2 for the discussion of detection statistic, '*r*' and 'x.'

6.4. PERFORMANCE EVALUATION

The Chi-Square test has been used to evaluate the performance of the distributions for the threshold analysis. The Chi-Square test employed for the threshold analysis is a little different from the one used for obtaining the results in the previous section (Section 5.4.2). The low value of CFAR (10^{-1} to 10^{-3}) corresponds to the tail of the pdf. For the threshold analysis, only the targets that correspond to a CFAR of 0.1 or lesser are considered, since the threshold analysis is performed in this low CFAR region.

This Chi-Square test was performed at the confidence level of 95% for the three distributions shown in the Table. The results were as shown in Table 6.1.

Table 6.1. Pass Percentages for Threshold Analysis for CFAR ≤ 0.1

Distribution	**Pass Percentages (%) for RX samples**
Gamma 2-Parameters	28.00
Beta 2-Parameters	19.44
Modified Beta 2-Parameters	67.36

The results show that the performance of the modified two parameter Beta model is superior to the other two models in the tail region. This is basically because of the introduction of the parameter '*N*.' However the pass percentage is still much lower than 95% which would be expected for a 95% confidence level. The reason for this is the fact that the number of targets at the CFAR of 0.1 FA/m^2 is quite small and the observation is noisy. Better results may be expected if a bigger area is considered for background modeling as would be available from newer stair-step collection.

6.5. SELECTION OF NUMBER OF CLASSES

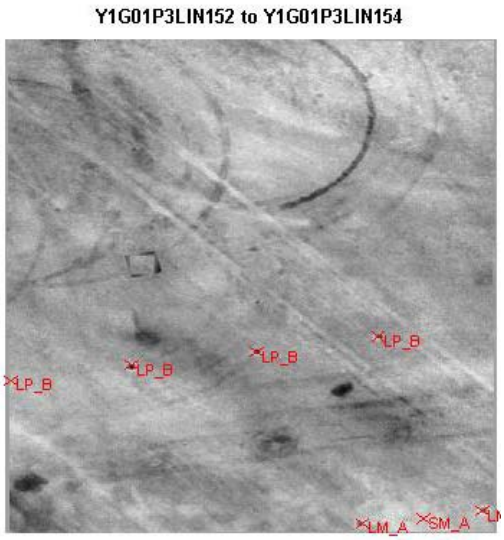
The following approach has been adopted for the selection of the number of classes. The modeling is started with one class and the Chi-Square test is performed on the model. If the test passes, the next frame is modeled otherwise the modeling is repeated for the current frame with two classes. After modeling with two classes, the parameters are stored along with the number of classes.

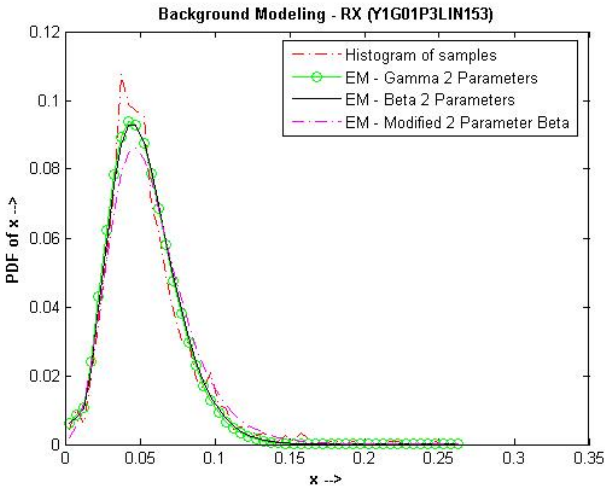
6.6. RESULTS—MODELING

This section shows the modeling of the detection statistics. In Figures 6.1 and 6.2, the histogram of the samples is shown in red color. The samples are the values of the detection statistic, 'x.' Please see Section 5.2 for the discussion of the detection statistic 'x.' Figure 6.1 shows the fit of the three distributions discussed. Figure 6.2 shows a closer look at the tail region of the pdf in Figure 6.1. The model that models the tail best would perform the best for the adaptive CFAR threshold selection in the low CFAR region. In Figure 6.1, parts (a) and (c) show the registered image from the May 2003 data whereas parts (b) and (d) show the modeling using two parameter Beta, two parameter Gamma and the modified two parameter Beta distributions.

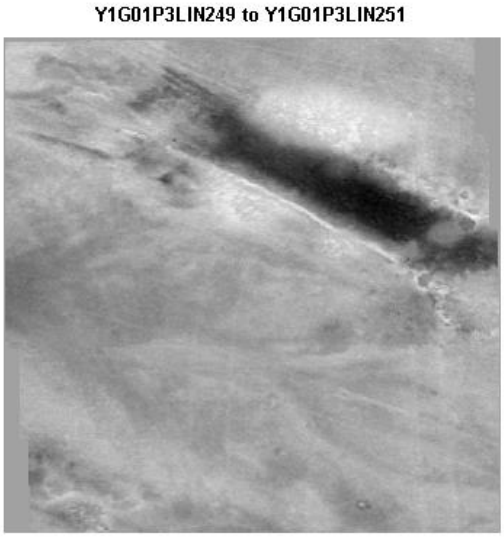
From the figures showing the modeling results, it is very clear that the modeling performance of the modified two parameter Beta distribution is superior to that of the two parameter Beta and Gamma distribution in the tail region. This is mainly because of the introduction of the parameter 'N,' that needs to be considered because of the modification due to the non-max suppression. The modified two parameter Beta distribution takes the parameter 'N' into consideration whereas the two parameter Beta and Gamma distributions do not. Please see Section 5.3 for the discussion on the parameter 'N.' This is also the reason for poor performance of the two parameter Beta and Gamma models in table 6.1.

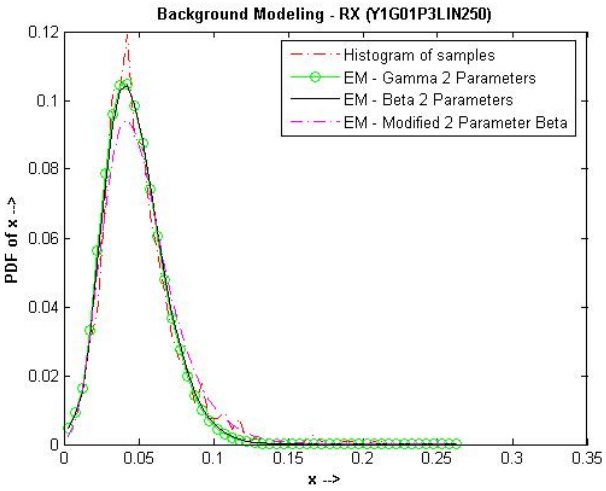
Note that a low CFAR in the range of 10^{-2} to 10^{-3} is of interest and it corresponds to the tail region of the pdf. Because the modeling performance of the modified two parameter Beta model is superior in the low CFAR region, this model is best for the adaptive CFAR threshold selection analysis. This is also shown by the results of the next section, where the results using the modified two parameter Beta model agree well with the predicted results.





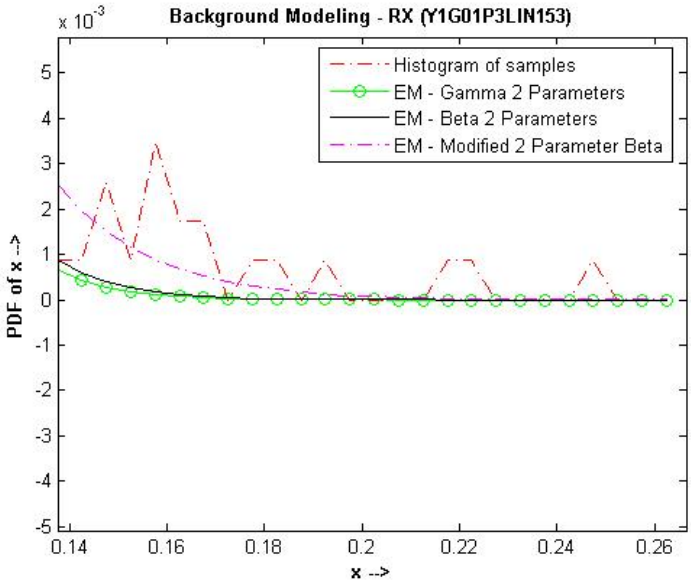
(a) Registered Image with Mines (b) Modeling



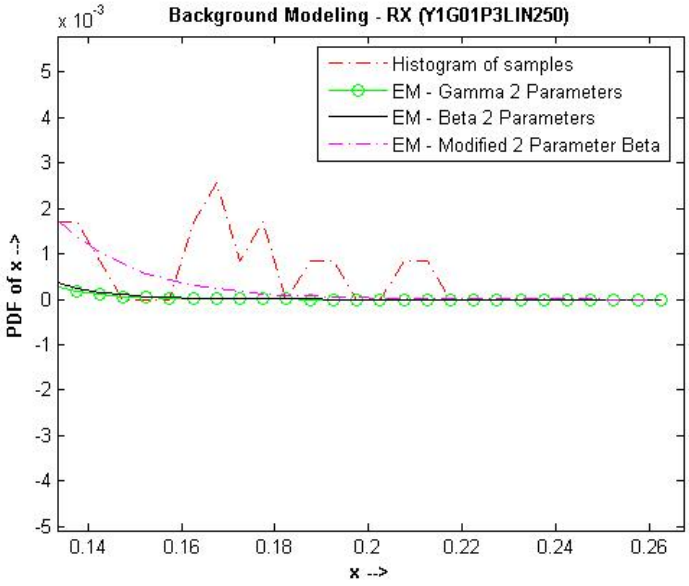


(c) Background Only Registered Image (d) Modeling

Figure 6.1. Modeling of Detection Statistic By Different Distributions—May 2003 data.



(a) Tail Region for Modeling in Figure 6.1 (b)



(b) Tail Region for Modeling in Figure 6.1 (d)

Figure 6.2. Expanded view of the tail region of modeling in Figure 6.1

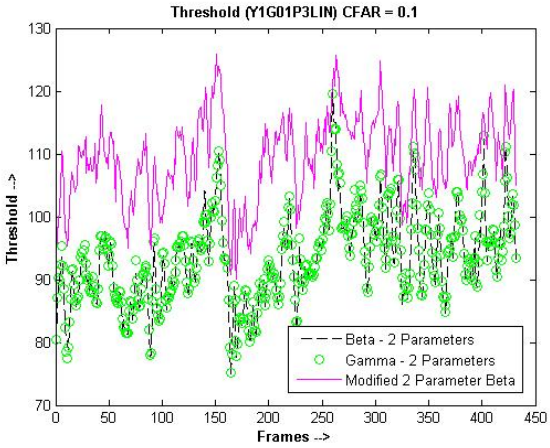
6.7. RESULTS—CFAR THRESHOLD SELECTION

This section shows the results of the threshold analysis that has been performed for the May 2003 data. The result has been calculated on the dataset Y1G01P3LIN, and the results are shown for following three values of CFAR:

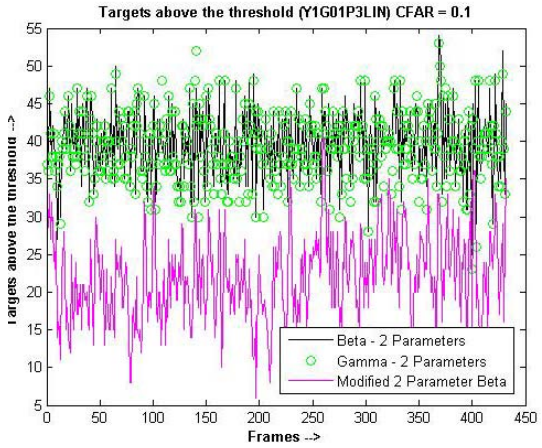
(1) 0.1
(2) 0.01
(3) 0.004

Figures 6.3-6.5 (a) show the threshold obtained for a given value of CFAR. The number of targets that are above this threshold are shown in Figures 6.3-6.5 (b). Figures 6.3-6.5 (c) and (d) show the histogram of the targets for the two parameter Beta and the modified two parameter Beta model respectively.

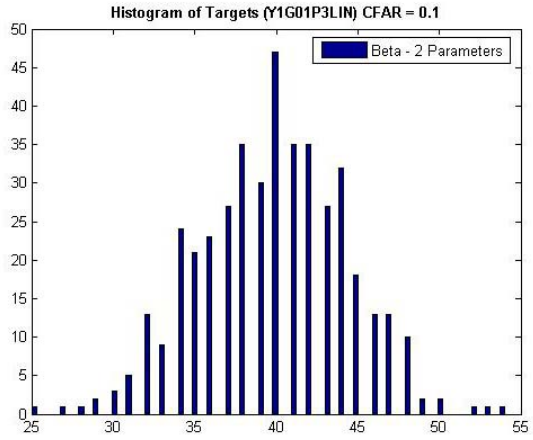
The histogram of the targets is a good measure to judge the modeling performance of the distributions for a given CFAR value. For example in case of the CFAR value of 0.1, the predicted number of false alarms per frame is 255x0.1 ≈ 25. (Here *255* is the area per frame, '*A*,' as shown in Equation 6.7) From Figure 6.3, it can be seen that the histogram of the two parameter Beta model is centered near 40 whereas the histogram of the modified two parameter Beta model is centered near 25 which is very close to the expected value of about 25. This tells that the performance of the modified two parameter Beta model is better than that of the two parameter Beta model for the CFAR of 0.1.



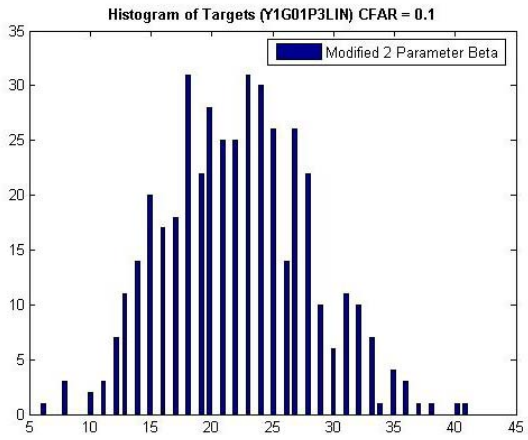
(a) Threshold Determination



(b) Targets Above the Threshold



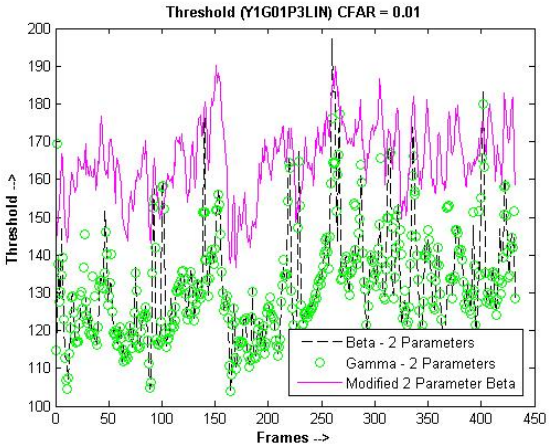
(c) Histogram (Beta 2-parameters)



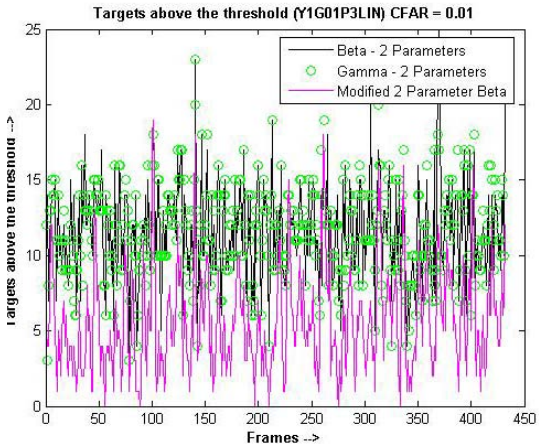
(d) Histogram (Modified 2 Parameter Beta)

Figure 6.3. Adaptive Threshold Determination for CFAR = 0.1

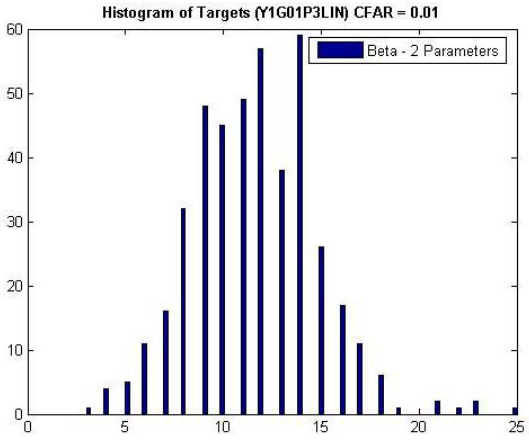
Ideally, the RX threshold across a run is expected to be almost constant, but due to the non-homogeneities of the data, there are significant variations in the threshold. This can be seen from the results. For example it can be seen that the threshold obtained for the frame nos. 160 to 190 is somewhat different from the threshold obtained for other frames in the dataset. On observing the dataset, it can be seen that these frames represent the regions that are highly non-homogeneous as compared to the regions represented by the other frames of the dataset.



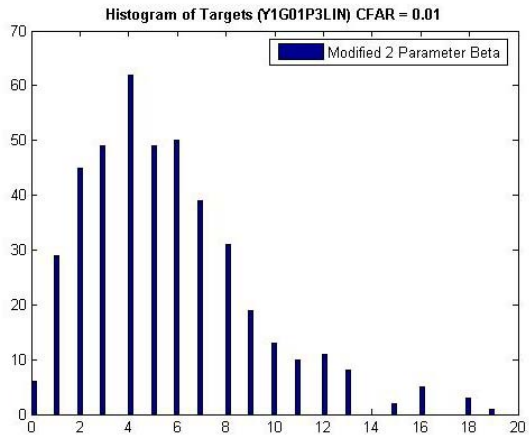
(a) Threshold Determination



(b) Targets Above the Threshold



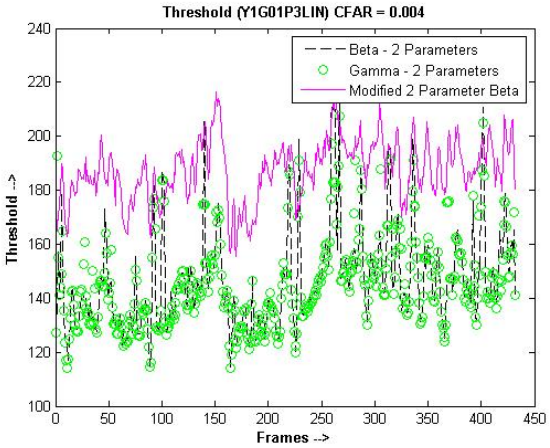
(c) Histogram (Beta 2-parameters)



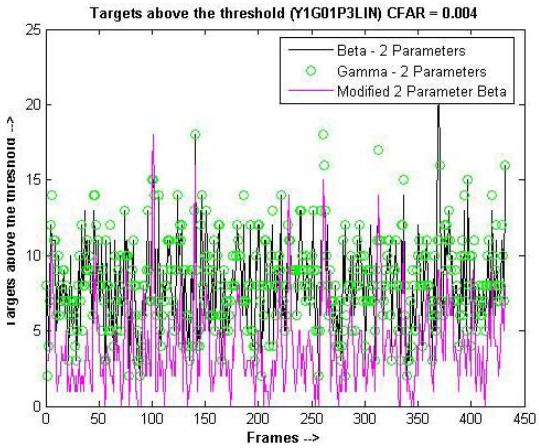
(d) Histogram (Modified 2 Parameter Beta)

Figure 6.4. Adaptive Threshold Determination for CFAR = 0.01

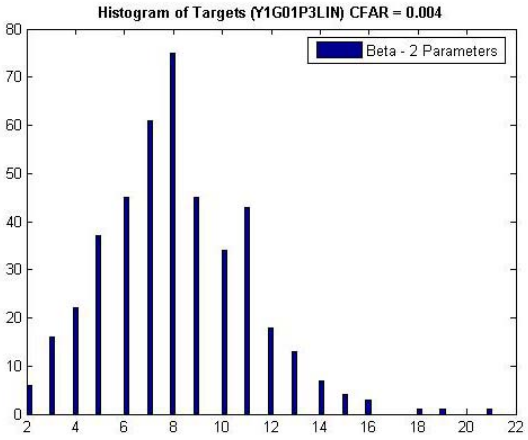
It can be seen from the Figures 6.3-6.5 (a) and (b) that as the CFAR value decreases, the threshold obtained increases and the number of targets above the threshold decreases. This is because a low value of CFAR corresponds to the tail region of the pdf. As one approaches the tail region, the CDF value of the distribution approaches one. Because the threshold is obtained from the inverse mapping of the CDF, the threshold value increases as the CDF value approaches one.



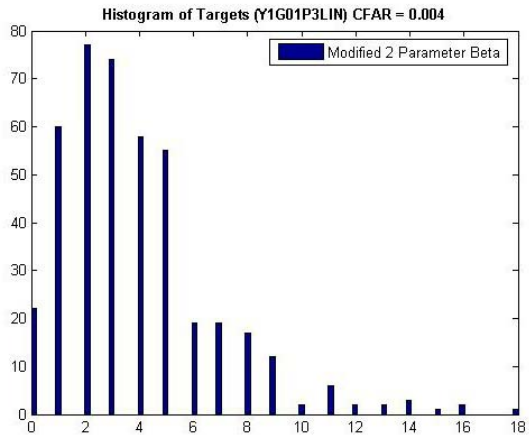
(a) Threshold Determination



(b) Targets Above the Threshold



(c) Histogram (Beta 2-parameters)



(d) Histogram (Modified 2 Parameter Beta)

Figure 6.5. Adaptive Threshold Determination for CFAR = 0.004

The CFAR of 0.1 corresponds to about 25.5 targets (255x0.1). The histogram of the modified two parameter Beta model is centered very close to this value. The CFAR of 0.01 corresponds to about 2.55 targets (255x0.01). The histogram of the modified two parameter Beta model is centered at 4 which is about 1 target higher. Finally, the CFAR of 0.004 corresponds to about 1 target (255x0.004) and the histogram of the modified two parameter Beta model is centered at 2 targets which is again 1 target higher. For the low value of CFAR (10^{-2} to 10^{-3}), the modeling results are away by just 1 target. This is fine

because for such a low value of CFAR, it is possible to have 1 extra target in the given region. In terms of the anomaly detector, the CFAR is of the order of 10^{-2} to 10^{-1}. The targets detected at the threshold corresponding to this CFAR would be passed to a False Alarm Mitigation (FAM) scheme, the aim of which is to reject the false alarms.

From the above plots, it is evident that the predicted number of targets corresponding to a given CFAR agrees well with the modeling results obtained from the modified two parameter Beta model. The superior performance of this model is because of the fact that it takes the parameter '*N*' into consideration, while the two parameter Gamma and Beta models do not.

6.8. CONCLUSIONS

This section discussed the need for the adaptive CFAR threshold selection and shows the application of the modeling of the detection statistic in performing adaptive CFAR threshold selection. For the adaptive CFAR threshold selection, good modeling in the low CFAR (10^{-2} to 10^{-3}) region is required. It has been shown that the modified two parameter Beta distribution has the best performance in the low CFAR region and therefore it is best suited for performing the adaptive CFAR threshold selection. Results obtained are in good agreement with the predicted results.

7. CONCLUSIONS AND FUTURE WORK

In this thesis, the EM algorithm is investigated and used for various applications. In Sections 3 and 4 the algorithm is used to model the background data and segment the image into classes. The concept of class membership is used to implement the SEM-based anomaly detectors. The results for segmentation were good. The SEM-based anomaly detectors performed well and in the case of buried mines their performance was slightly better than that of RX.

In Sections 5 and 6, the EM algorithm was used to model the detection statistic of the anomaly detector. This modeling was then used to perform the adaptive CFAR threshold selection to determine the optimum number of targets for the given CFAR. Here also the modeling results were very good. The results showed that the detection statistic can be very well modeled with the two parameter Gamma and Beta distributions. The results of the adaptive CFAR threshold selection performed using the modified two parameter Beta model were in good agreement with the expected number of targets.

While implementing the EM algorithm, it has been observed that sometimes a large number of iterations are needed before the algorithm converges to a steady-state value. In literature, different methods are mentioned to speed up the convergence of the algorithm. Some of these are mentioned in [1]. In future, these methods can be employed to speed up the convergence of the EM algorithm. This would directly reduce the amount of time needed to process the applications that use the EM algorithm to model the data.

The EM-based segmentation could be tested on multi-band data. This would help segmenting the image based on spatial and spectral correlation. Practical issues such as consistency of class assignment over consecutive frames need to be addressed. Better methods for selecting the number of classes must be studied.

APPENDIX—A
DISTRIBUTIONS AND UPDATE EQUATIONS

This appendix has two sections. The first section presents the various distributions and the corresponding update equations. The second section explains how these update equations are used to estimate the parameters of the mixture model.

A.1. UPDATE EQUATIONS

In this section, the distributions of the various models are given. The partial derivatives that are used to obtain the update equations are also presented for the various mixture models.

Please refer to the following terminology for all the equations presented in this section.

n = Number of samples

g = Number of classes

$$\Psi(u) = \frac{\partial \log(\Gamma(u))}{\partial u} \text{ where } \Gamma \text{ is the Gamma function.} \tag{A.1}$$

$$B(\gamma,\eta) = \frac{\Gamma(\gamma)\Gamma(\eta)}{\Gamma(\gamma+\eta)} \tag{A.2}$$

$$I_1 = \int_0^x \frac{(1-x)^{n-1}x^{\gamma-1}}{[1-(1-\lambda)x]^{\gamma+\eta}}dx \tag{A.3}$$

$$I_2 = \int_0^x \frac{(1-x)^{n-1}x^{\gamma}}{[1-(1-\lambda)x]^{\gamma+\eta+1}}dx \tag{A.4}$$

$$I_3 = \int_0^x \frac{(1-x)^{n-1}x^{\gamma-1}\log(1-x)}{[1-(1-\lambda)x]^{\gamma+\eta}}dx \tag{A.5}$$

$$I_4 = \int_0^x \frac{(1-x)^{n-1}x^{\gamma-1}\log[1-(1-\lambda)x]}{[1-(1-\lambda)x]^{\gamma+\eta}}dx \tag{A.6}$$

$$I_5 = \int_0^x \frac{(1-x)^{n-1}x^{\gamma-1}\log(x)}{[1-(1-\lambda)x]^{\gamma+\eta}}dx \tag{A.7}$$

where $\log(x)$ denotes natural logarithm and z_{ij} is the probability that the j^{th} (j = 1, 2... n) sample arose from the i^{th} (i = 1, 2... g) class. For all the mixture models, the coefficient of the i^{th} (i = 1, 2... g) class, π_i, is given by:

$$\pi_i = \frac{\sum_{j=1}^{n} z_{ij}}{n} \tag{A.8}$$

Let the partial derivative of the cost function, 'Q', with respect to some parameter 'φ' (that is to be estimated from the data) be given by $\frac{\partial Q}{\partial \varphi}$, then the update equation for this parameter is given by:

$$E\left[\frac{\partial Q}{\partial \varphi}\right] = 0$$

where, 'E' is the expectation operator.

A.1.1. Gaussian Mixture Model, $N(x : \mu, \Sigma)$

Here, 'μ' is the mean, 'Σ' is the covariance matrix and 'd' is the dimensionality of the data, 'x.' The multivariate Gaussian distribution is given as:

$$f(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\} \tag{A.9}$$

The ML estimates of the parameters of a Gaussian mixture model are given by:

$$\mu = \frac{\sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}[x]}{n} \tag{A.10}$$

$$\Sigma = \frac{\sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\left[(x-\mu)^T(x-\mu)\right]}{n} \tag{A.11}$$

A.1.2. Two Parameter Beta Model, $Beta(x:\gamma,\eta)$.

$$f(x) = \frac{x^{\gamma-1}(1-x)^{\eta-1}}{B(\gamma,\eta)} \quad , \quad 0 \le x \le 1; \gamma,\eta > 0 \tag{A.12}$$

$$\frac{\partial \log[f(x)]}{\partial \gamma} = \sum_{i=1}^{g} z_{ij}\left[\Psi(\gamma+\eta) - \Psi(\gamma) + \log(x)\right] \tag{A.13}$$

$$\frac{\partial \log[f(x)]}{\partial \eta} = \sum_{i=1}^{g} z_{ij}\left[\Psi(\gamma+\eta) - \Psi(\eta) + \log(1-x)\right] \tag{A.14}$$

A.1.3. Three Parameter Beta Model, $Beta(x:\gamma,\eta,\lambda)$.

$$f(x) = \frac{\lambda^{\gamma} x^{\gamma-1}(1-x)^{\eta-1}}{B(\gamma,\eta)[1-(1-\lambda)x]^{\gamma+\eta}} \quad , \quad 0 \le x \le 1; \gamma,\eta,\lambda > 0 \tag{A.15}$$

$$\frac{\partial \log[f(x)]}{\partial \gamma} = \sum_{i=1}^{g} z_{ij}\left[\Psi(\gamma+\eta) - \Psi(\gamma) + \log(\lambda) + \log(x) - \log(1-(1-\lambda)x)\right] \tag{A.16}$$

$$\frac{\partial \log[f(x)]}{\partial \eta} = \sum_{i=1}^{g} z_{ij}\left[\Psi(\gamma+\eta) - \Psi(\eta) + \log(1-x) - \log(1-(1-\lambda)x)\right] \tag{A.17}$$

$$\frac{\partial \log[f(x)]}{\partial \lambda} = \sum_{i=1}^{g} z_{ij}\left[\frac{\gamma}{\lambda} - \frac{(\gamma+\eta)x}{\log(1-(1-\lambda)x)}\right] \tag{A.18}$$

A.1.4. Modified Two Parameter Beta Model, $Beta(x:\gamma,\eta,N)$**.**

$$\text{Let, } f_2(x) = \frac{x^{\gamma-1}(1-x)^{\eta-1}}{B(\gamma,\eta)} \quad , \quad 0 \le x \le 1; \gamma,\eta > 0 \tag{A.19}$$

Then the modified two parameter Beta model is given by:

$$f_2(x) = \frac{1}{\mathrm{K}} f_2(x) e^{-N(1-F_2(x))} \qquad (N > 0) \tag{A.20}$$

where $F_2(x)$ is the cumulative distribution function of $f_2(x)$.

K is given by:

$$K = \int_0^1 f_2(x) e^{-N(1-F_2(x))} dx. \tag{A.21}$$

$$\frac{\partial \log[f(x)]}{\partial \gamma} = \sum_{i=1}^{g} z_{ij} \left[\Psi(\gamma+\eta) - \Psi(\gamma) + \log(x) + \frac{N}{B(\gamma,\eta)} \left\{ I_5 + \frac{\eta I_1}{\gamma(\gamma+\eta)} - I_4 \right\} \right] \tag{A.22}$$

$$\frac{\partial \log[f(x)]}{\partial \eta} = \sum_{i=1}^{g} z_{ij} \left[\Psi(\gamma+\eta) - \Psi(\eta) + \log(1-x) + \frac{N}{B(\gamma,\eta)} \left\{ I_3 - I_4 + \frac{\gamma I_1}{\eta(\gamma+\eta)} \right\} \right] \tag{A.23}$$

$$\frac{\partial \log[f(x)]}{\partial N} = \sum_{i=1}^{g} z_{ij} [1 - F_3(x)] + \frac{e^{-N}(N+1)-1}{KN^2} \tag{A.24}$$

Here all the integrals 'I_u' are same as defined previously with the parameter $\lambda = 1$

A.1.5. Modified Three Parameter Beta Model, $Beta(x:\gamma,\eta,\lambda,N)$.

$$\text{Let, } f_3(x) = \frac{\lambda^{\gamma}x^{\gamma-1}(1-x)^{\eta-1}}{B(\gamma,\eta)[1-(1-\lambda)x]^{\gamma+\eta}} \quad , \quad 0\le x\le 1;\gamma,\eta,\lambda>0 \tag{A.25}$$

Then the modified three parameter Beta model is given by:

$$f(x)=\frac{1}{\mathrm{K}}\, f_3(x)e^{-N(1-F_3(x))} \qquad (N>0) \tag{A.26}$$

where $F_3(x)$ is the cumulative distribution function of $f_3(x)$ and 'K' is given by:

$$\mathrm{K}=\int_0^1 f_3(x)e^{-N(1-F_3(x))}dx \tag{A.27}$$

$$\begin{aligned}\frac{\partial \log[f(x)]}{\partial\gamma} &= \sum_{i=1}^{g} z_{ij}\left[\Psi(\gamma+\eta)-\Psi(\gamma)+\log(\lambda)+\log(x)-\log(1-(1-\lambda)x)\right] \\ &\quad + z_{ij}\left[\frac{N\lambda^{\gamma}}{B(\gamma,\eta)}\left\{\log(\lambda)I_1+I_5+\frac{\eta I_1}{\gamma(\gamma+\eta)}-I_4\right\}\right]\end{aligned} \tag{A.28}$$

$$\begin{aligned}\frac{\partial \log[f(x)]}{\partial\eta} &= \sum_{i=1}^{g} z_{ij}\left[\Psi(\gamma+\eta)-\Psi(\eta)+\log(1-x)-\log(1-(1-\lambda)x)\right] \\ &\quad + z_{ij}\left[\frac{N\lambda^{\gamma}}{B(\gamma,\eta)}\left\{I_3-I_4+\frac{\gamma I_1}{\eta(\gamma+\eta)}\right\}\right]\end{aligned} \tag{A.29}$$

$$\frac{\partial \log[f(x)]}{\partial\lambda} = \sum_{i=1}^{g} z_{ij}\left[\frac{\gamma}{\lambda}-\frac{(\gamma+\eta)x}{\log(1-(1-\lambda)x)}+\frac{N\lambda^{\gamma-1}}{B(\gamma,\eta)}\{\gamma I_1-\lambda(\gamma+\eta)I_2\}\right] \tag{A.30}$$

$$\frac{\partial \log[f(x)]}{\partial N} = \sum_{i=1}^{g} z_{ij}\left[1 - F_3(x)\right] + \frac{e^{-N}(N+1)-1}{KN^2} \tag{A.31}$$

A.1.6. Two Parameter Gamma Model, $Gamma(x : k, \lambda)$**.**

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} \quad , \quad 0 \le x < \infty ; \mathrm{k}, \lambda > 0 \tag{A.32}$$

$$\frac{\partial \log[f(x)]}{\partial k} = \sum_{i=1}^{g} z_{ij}\left[\log(\lambda) + \log(x) - \Psi(k)\right] \tag{A.33}$$

$$\frac{\partial \log[f(x)]}{\partial \lambda} = \sum_{i=1}^{g} z_{ij}\left[\frac{k}{\lambda} - x\right] \tag{A.34}$$

A.2. PARAMETER ESTIMATION FROM THE UPDATE EQUATIONS

In this section, the application of the update equations to estimate the set of parameters of the mixture model is shown. Let us consider the estimation problem of the three parameter Beta distribution, Beta($x : \gamma, \eta, \lambda$)

Once the update equations are obtained, the information matrix is formed. The information matrix, 'I_m,' is a square matrix with the dimension 'p,' where 'p' is the number of parameters to be estimated. In the case of Beta($x : \gamma, \eta, \lambda$), p = 3. Each diagonal element of the information matrix is the derivative of the log-likelihood of the complete data with respect to the estimated parameter. Thus, in the case of Beta($x : \gamma, \eta, \lambda$), this matrix is obtained as follows:

$$I_m = \begin{bmatrix} \sum_{j=1}^{n}\left[\frac{\partial \log[f(x)]}{\partial \gamma}\right]^2 & 0 & 0 \\ 0 & \sum_{j=1}^{n}\left[\frac{\partial \log[f(x)]}{\partial \eta}\right]^2 & 0 \\ 0 & 0 & \sum_{j=1}^{n}\left[\frac{\partial \log[f(x)]}{\partial \lambda}\right]^2 \end{bmatrix} \quad \text{(A.35)}$$

where $\log[f(x)]$ is the log-likelihood of the complete data.

The value of the off-diagonal elements of the information matrix, 'I_m' gives the dependence of one parameter over the other [49]. In case of the EM algorithm, since the parameters are often assumed to be independent, the off-diagonal elements can be taken to be zero. Let $\gamma^k, \eta^k, \lambda^k$ be the estimates of the parameters in the k^{th} iteration, then the estimate of these parameters in the $(k+1)^{\text{th}}$ step is given by:

$$\begin{bmatrix} \gamma \\ \eta \\ \lambda \end{bmatrix}^{k+1} = \begin{bmatrix} \gamma \\ \eta \\ \lambda \end{bmatrix}^{k} + (I_m)^{-1} \begin{bmatrix} \sum_{j=1}^{n}\frac{\partial \log[f(x)]}{\partial \gamma} \\ \sum_{j=1}^{n}\frac{\partial \log[f(x)]}{\partial \eta} \\ \sum_{j=1}^{n}\frac{\partial \log[f(x)]}{\partial \lambda} \end{bmatrix}^{k} \quad \text{(A.36)}$$

This new estimate of the parameters $'\gamma^{k+1}', '\eta^{k+1}', '\lambda^{k+1}'$ is used to calculate the log-likelihood function again in the next iteration. This new log-likelihood function is again maximized to get the new set of parameters $'\gamma^{k+2}', '\eta^{k+2}', '\lambda^{k+2}'$ in the $(k+2)^{\text{th}}$ iteration. The process continues until the parameters converge to a steady state value. Please see section 2.5 for further details on the calculation of the log-likelihood function.

APPENDIX—B

TEST STATISTICS TO MEASURE GOODNESS OF FIT

This appendix discusses some of the tests to measure goodness of fit.

B.1. CHI-SQUARE TEST

After the visual inspection of the modeling results, it is worthwhile to have a quantitative analysis of the performance of the different models. One such comprehensive test is the Chi-Square test. It has the following test statistic [45]:

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right] \tag{B.1}$$

where,

χ^2 = Test Statistic

O_i = Observed value for ith observation

E_i = Estimated value for ith observation

n = Number of observations

Sometimes a correction known as the “Yates correction” is also applied to account for the quantization error. The Yates correction is introduced as a correction for discontinuity. In that case, the test statistic is modified as:

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(|O_i - E_i| - c)^2}{E_i} \right] \tag{B.2}$$

where ‘c’ is the Yates correction and generally c = 0.5

The degree of freedom required to calculate the threshold is calculated as follows. First the samples are grouped in bins in such a way that there is at least a certain number of samples per bin. If the number of bins are ‘n_b’ and there are ‘p’ parameters that have been estimated from the data, then the degrees of freedom, ‘v,’ is given as:

$$v = n_b - (p + 1) \tag{B.3}$$

The hypothesis $H_0 : X \sim F$ is rejected if $\chi^2 \geq \chi^2_{1-\alpha}$

B.2. CRAMER VON-MISES (CVM) TEST

The test statistic for this test is given as [45]:

$$CM = \frac{1}{12n} + \sum_{i=1}^{n}\left[F\left(x_{1:n};\theta\right) - \frac{i-0.5}{n} \right]^2 \tag{B.4}$$

where,

n = Number of observations.

$F\left(x_{1:n};\theta\right)$ = CDF of the ordered observations, given θ.

θ = Parameter vector of the given distribution.

Here, CDF is the Cumulative Distribution Function. An approximate size 'α' test of $H_0 : X \sim F$ is to reject H_0 *if* $CM \geq CM_{1-\alpha}$. The critical values, $CM_{1-\alpha}$, are given in [45] for several values of 'α' and censoring levels. If the parameters are estimated from the data then the value of 'θ' is replaced by its maximum likelihood estimate.

B.3. KOLMOGOROV-SMIRNOV (KS) OR KUIPER TEST

In order to study this statistic let us assume that [45],

n = Number of observations.

$F\left(x_{1:n};\theta\right)$ = CDF of the ordered observations, given 'θ.'

θ = Parameter vector of the given distribution.

$$D^+ = \max_i \left[\frac{i}{n} - F\left(x_{i:n};\theta\right) \right] \tag{B.5}$$

$$D^- = \max_i \left[F\left(x_{i:n};\theta\right) - \frac{i-1}{n} \right] \tag{B.6}$$

$$D = \max\left(D^{+}, D^{-}\right) \tag{B.7}$$

$$V = D^{+} + D^{-} \tag{B.8}$$

The statistic given by 'D' is known as the "Kolmogorov-Smirnov" or "KS" statistic and the statistic given by 'V' is known as the "Kuiper" statistic. Here also, the 'α' test of $H_0 : X \sim F$ is to reject H_0 *if* $KS \geq KS_{1-\alpha}$. The critical values for this statistic are given in [45] for several values of 'α' and censoring levels.

BIBLIOGRAPHY

[1] McLachlan, G., Krishnan, T. (1997). The *EM Algorithm and Extensions*. Wiley series in probability and statistics. John Wiley & Sons.

[2] Zhang, J., Modestino, J.W., Langan, D.A. (1994). *Maximum-Likelihood Parameter Estimation for Unsupervised Stochastic Model-Based Image Segmentation.* IEEE Transaction on Image Processing Vol. 3, No. 4, pp. 404-420.

[3] Moon, T.K. (1996). *The Expectation-Maximization Algorithm*. IEEE Signal Processing Magazine.

[4] Dellaert, F. (2002). *The Expectation Maximization Algorithm*. College of Computing, Georgia Institute of Technology, Technical Report number GIT-GVU-02-20.

[5] Bilmes, J. A. (1998). *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.* International Computer Science Institute, Berkeley CA and Computer Science Division, Department of Electrical Engineering and Computer Science, U.C. Berkeley, TR-97-021.

[6] Copsey, K., Webb, A. (2003). *Bayesian Gamma Mixture Model Approach to Radar Target Recognition.* IEEE Transactions on Aerospace and Electronic Systems Vol. 39, No. 4 pp. 1201-1217.

[7] Kim, M., Fisher III, John., Principe, J.C. *A new CFAR stencil for target etections in Synthetic Aperture Radar (SAR) imagery.* Computational NeuroEngineering Laboratory, CSE 447, Electrical and Computer Engineering Department, University of Florida, Gainesville, FL 32611. SPIE Vol. 2757, pp. 432-442.

[8] Gnanadesikan, R., Pinkham, R.S., Hughes L. P. (1967). *Maximum Likelihood Estimation of the Parameters of the Beta Distribution from Smallest Order Statistics.* Technometrics, Vol. 9, No. 4, pp. 607-620.

[9] Harter, H. L. (1967). *Maximum Likelihood Estimation of the Parameters of a Four-Parameter Generalized Gamma Population from Complete and Censored Samples.* Technometrics. Vol. 9, No. 1. , pp. 159-165.

[10] Pham, T., Almhana, J. (1995). *The Generalized Gamma Distribution: Its Hazard Rate and Stress-Strength Model.* IEEE Transactions on Reliability, Vol. 44, No. 3, pp. 392-397.

[11] Pham-Gia, T., Duong, Q. P. (1989). The Generalized Beta and F-Distributions in Statistical Modelling. Mathl. Comput. Modelling, Vol. 12, pp. 1613-1625.

[12] Barnett, V., Lewis, T. (1994). *Outliers in statistical data*. New York: John Wiley and Sons.

[13] Eskin, E. *Anomaly Detection over Noisy Data using Learned Probability Distributions*. Computer Science Department, Columbia University, New York.

[14] Stein, D. (2001) *Modeling Variability In Hyperspectral Imagery Using A Stochastic Compositional Approach*. IEEE International Vol. 5, pp. 2379-2381.

[15] Fries, R. (2003). *SEM*. PAR Government Systems Corporation.

[16] Stein, D. (1995). *Detection of Random Signals in Gaussian Mixture Noise.* IEEE Transactions on Information Theory, Vol. 41, No. 6, pp. 1788-1801.

[17] Stein, D. (2002) *Stochastic compositional models applied to subpixel analysis of hyperspectral imagery*. Proceedings of SPIE Vol. 4480, pp. 49-56.

[18] Downs A.M., Heisterkamp, S.H., Jean-Baptiste, B., Hamers, F.F. (1997) *Reconstruction and prediction of the HIV/AIDS epidemic among adults in the European Union and in the low prevalence countries of central and eastern Europe.* AIDS 1997 Vol. 11 No. 5, pp. 649-662.

[19] Winkler, W.E. *Advanced Methods for Record Linkage*. Bureau of the Census, Washington DC 20233-9100.

[20] Genglar, N., Tijani, A., Wiggans, G.R., Misztal, I. (1999). *Estimation of (Co)variance Function for Test Day Yield with a Expectation-Maximization Restricted Maximum Likelihood Algorithm.* J. Dairy Sci. (Aug. 1999).

[21] Zhang, Y., Brady, M., Smith, S. (2001). *Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm.* IEEE Transactions on Medical Imaging, Vol. 20, No. 1, January 2001. , pp. 45-57.

[22] Sankar, A., Lee, C. H. (1995). *Robust Speech Recognition Based on Stochastic Matching*. AT&T Bell Laboratories, Murray Hill, NJ 07974, USA.

[23] Ramachandran, H. (2004). *Background Modeling And Algorithm Fusion For Airborne Landmine Detection.* University of Missouri-Rolla, USA.

[24] Gonzalez, R.C., Woods, R.E. (2002). *Digital Image Processing (Second Edition).* Prentice Hall, New Jersey 07458, USA.

[25] Reed, I.S., Yu, X. (1990). *Adaptive multi-band CFAR detection of an optical pattern with unknown spectral distribution* IEEE Trans. On Acoustics, Speech and Signal Processing, Vol. 38, No. 10, pp 1760-1770.

[26] Chomczimsky, W., Mejail, M., Jacobo-Berlles, A. V., Kornblit, F., Frery, A. C. (1998). *Classification of SAR images based on estimates of the parameters of the* ς_A^0 *distribution.* Part of the SPIE Conference on Mathematical Modeling and Estimation Techniques in Computer Vision, San Diego, California. SPIE Vol. 3457, pp. 202-208.

[27] Huiyan, Z., Yongfeng, C., Wen, Y. *SAR Image Segmentation Using MPM Constrained Stochastic Relaxation.* Civil Engineering Department, Wuhan Polytechnic University, Wuhan, China.

[28] Liu, G., Huang, S., Torre, A., Rubertone, F. *Statistical analysis of multi-look polarimetric SAR data and terrain classification with adaptive distribution.*

[29] Lew, H., Drumheller, D.M. (2002). *Estimation of non-Rayleigh clutter and fluctuating-target models.* IEE Proc. -Radar Sonar Navig., Vol. 149, No. 5, pp. 231-241.

[30] Rosenberg, C., Hebert, M., Thrun S. (2001). *Color Constancy Using KL-Divergence.* Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213. IEEE Transactions, pp. 239-246.

[31] Zhai, C. X. (2003). *Notes on the KL-divergence retrieval formula and dirichlet prior smoothing.*

[32] Webb, A. R. (2000). *Gamma mixture models for target recognition*. Pattern Recognition, 33, 2045-2054.

[33] Copsey, K. D, Webb, A. R. (2001). *Bayesian Gamma mixture models for target recognition.* Proceedings of CIMA' 2001, Bangor, Wales, June 2001.

[34] Menon, D., Agarwal, S., Ganju, R., Swonger, C.W. (2004). *False-alarm mitigation and feature-based discrimination for airborne mine detection.* Proceedings of the SPIE, Volume 5415, pp. 1163-1173.

[35] Menon, D. (2005). *A Knowledge Based Architecture for Airborne Minefield Detection.* University of Missouri-Rolla, USA.

[36] Fellegi, I. P., Sunter, A. B. (1969), *A Theory for Record Linkage.* Journal of the American Statistical Association, 64, 1183-1210.

[37] Bacchetti, P., Segal, M. R., Jewell, N.P. (1993). *Backcalculation of HIV infection rates (with discussion).* Statistical Science, 8, 82-119.

[38] Meyer, K., Hill, W.G. (1997). *Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by Restricted Maximum Likelihood*. Institute of Cell, Animal and Population Biology, Edinburgh University, West Mains Road, Edinburgh, Scotland.

[39] Groeneveld, E., Kovac, M. (1990). *A Note on Multiple Solutions in Multivariate Restricted Maximum Likelihood Covariance Component Estimation.* Department of Animal Sciences, University of Illinois, Urbana, 61801.

[40] Kilian, M., William, M., Alexandre, G., Kiyoto, K., Shenton, M. E., Ron, K., Grimson, W. E. L., Warfield, S. K. (2002). *Incorporating Non-rigid Registration into Expectation Maximization Algorithm to Segment MR Images*. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA, USA.

[41] Dasgupta, A., Raftery, A. E. (1995). *Detecting Features in Spatial Point Processes with Clutter via Model-Based Clustering.* Technical Report No. 295, Department of Statistics, University of Washington.

[42] Kass, R. E., Raftery, A. (1995). *Bayes Factors*. Journal of the American Statistical Association, 90:773-795.

[43] Sankar, A., Lee, C. H. (1996). *A Maximum-Likelihood Approach to Stochastic Matching for Robust Speech Recognition.* IEEE Transactions on Speech and Audio Processing, Vol. 4, No. 3.

[44] Arthur. N., David, N. (1989). *Speech Recognition Using Noise-Adaptive Prototypes.* IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 37, No. 10.

[45] Bain, L., Engelhardt, M. (1991). *Introduction to Probability and Mathematical Statistics*. Duxbury Press, Belmont, California (Second Edition).

[46] Stein, D., Beaven, S. G., Hoff, L., Winter, E., Schaum, A. P., Stocker, A. D. (2002). *Anomaly Detection from Hyperspectral Imagery*. IEEE Signal Processing Magazine, Vol. 19, No. 1, pp. 58-69.

[47] Press, S. J. (1966). *Linear Combinations of Noncentral Chi-Square Variates*. Annals of Mathematics and Statistics, No. 37, pp. 480-487.

[48] Dempster, A. P., Laird, N. M., Rubin, D.B. (1977). *Maximum Likelihood from incomplete data via the EM algorithm (with discussion).* Journal of the Royal Statistical Society.

[49] Harrison, H., Denny, J. L., Robert, F. (1995). Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance. J. Opt. Soc. Am. A, Vol. 12, No. 5, pp. 834-852.

VITA

Ritesh Ganju in Meerut, India. In June 2000, he received his Bachelor's degree in Electronics and Communication Engineering from Regional Engineering College—Calicut (R.E.C.—Calicut, India), now known as the National Institute of Technology—Calicut (N.I.T.—Calicut, India). He joined the Master of Science program in Electrical Engineering at the University of Missouri—Rolla in Fall 2003. While pursuing his graduate degree, he was supported by the Department of Electrical Engineering as a Graduate Research Assistant. He received his master's degree in Electrical Engineering from the University of Missouri—Rolla in May 2006.