INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
Application of Multisensor Fusion to Improve Landmine Detection

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Ajith Harshalal Gunatilaka, M.S., B.Sc.

The Ohio State University

2000

Dissertation Committee:

Prof. E. H. Newman, Adviser
Prof. B. A. Baertlein
Prof. E. K. Walton

Approved by

Edward H. Newman
Adviser
Department of Electrical Engineering
ABSTRACT

Any realistic attempt to clear buried landmines during peacetime (humanitarian demining) will require sensors having nearly perfect detection at very low false alarm rates. Currently no single demining sensor can meet the desired performance level, and in this dissertation sensor fusion is considered for this task.

Techniques for conventional sensor fusion include data-level, feature-level and decision-level fusion. All of these standard techniques are unattractive for reasons that include non-commensurate data (i.e., data of incompatible format), sensor positioning errors, non-coincident sampling, and generally poor performance.

In this dissertation, we develop a practical feature-level fusion algorithm that can accommodate non-coincidently sampled data. Data acquired with ground penetrating radar (GPR), electro-magnetic induction (EMI), and infrared (IR) sensors at two mine test sites are used to compare the performance of individual and fused sensors. These studies confirm that for mine detection feature-level fusion provides the best performance possible for this sensor suite. Soft decision-level fusion also provides significant improvement, but an optimal hard decision-level fusion provides only a marginal improvement over the best individual sensor. The ad hoc decision-level fusion rules AND, OR, and majority voting are ineffective for this application.

To support the fusion studies a new clutter reduction technique is developed, which can significantly reduce the dominant, ground reflection clutter in GPR data.
Unlike previously described clutter reduction techniques, this new method can work successfully even when the target is small and buried at a very shallow depth, even flush to ground. A subspace decomposition procedure is used to reduce additional clutter contributions due to antenna ringing and cable mismatches.

Good features are essential for efficient fusion and considerable effort was devoted to this subject. Using the Fort A. P. Hill data we demonstrate that physical and representation features are ineffective for discriminating mines from clutter. For both GPR and EMI data it is shown that statistical discriminant features provide very good classification.
To my parents
ACKNOWLEDGMENTS

I would like to express my gratitude to my project supervisor, Dr. Brian A. Baertlein, for his guidance, support and patience. I am indebted to him for his comments and suggestions throughout the project and during the dissertation writing, which were invaluable for completing this dissertation. I would also like to thank my academic adviser Prof. Edward H. Newman for his help during my stay at The Ohio State University and for serving in my dissertation committee. I am thankful to Dr. Eric K. Walton for enabling me to join this project, for many useful comments, and for serving in my dissertation committee.

I wish to thank Frank Paynter, Dr. Chi-Chih Chen, Dr. Soumya Nag, Do-Hoon Kwon, Kishor Rama Rao and Kit Wing Lee for their help and useful discussions. I also acknowledge Kursat Sendur for providing me the IR data.

I owe a debt of gratitude to my parents and my sisters for their encouragement, support and the blessings. I cannot thank enough my wife, Shanika, without whose understanding, encouragement, and patience this dissertation would not have been possible.

This project was supported by funds from Duke University under an award from the Army Research Office (ARO) (the Office of Secretary of Defense, Multidisciplinary
University Research Initiative (MURI) program). The findings, opinions and recommendations expressed therein are those of the author and are not necessarily those of Duke University or the ARO.
VITA

March 23, 1964 ........................................................Born in Negombo, Sri Lanka

1990 .................................................................B.Sc.(Eng.) Electronics and Telecommunication Engineering, University of Moratuwa, Moratuwa, Sri Lanka

1992 .................................................................M.S. Electrical Engineering, Cleveland State University, Cleveland, Ohio

1994 .................................................................M.S. Optical Sciences and Engineering, University of Central Florida, Orlando, Florida

1994-present .....................................................Graduate Research Associate, The Ohio State University, Columbus, Ohio

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Electrical Engineering
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Objectives of this Dissertation</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Organization of the Dissertation</td>
<td>4</td>
</tr>
<tr>
<td>2. Landmines</td>
<td>6</td>
</tr>
<tr>
<td>2.1 The Landmine Problem</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Landmine Characteristics</td>
<td>7</td>
</tr>
<tr>
<td>2.3 Mine Detection Sensors</td>
<td>8</td>
</tr>
<tr>
<td>2.4 Summary</td>
<td>14</td>
</tr>
<tr>
<td>3. Multisensor Data Fusion</td>
<td>15</td>
</tr>
<tr>
<td>3.1 Sensor Fusion Preliminaries</td>
<td>15</td>
</tr>
<tr>
<td>3.2 Sensor Fusion Architectures</td>
<td>16</td>
</tr>
<tr>
<td>3.3 Sensor Fusion for Landmine Detection</td>
<td>20</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Description of the objects buried in the MURI mine test grid. For objects indicated by an asterisk, dimensions are width and length.</td>
</tr>
<tr>
<td>6.2</td>
<td>Individual sensor $P_F$ and $P_M$ using &quot;good&quot; thresholds.</td>
</tr>
<tr>
<td>6.3</td>
<td>Discriminatory power of individual features and combinations of features according to the extended Fisher metric.</td>
</tr>
<tr>
<td>7.1</td>
<td>Identities of targets in Fort A. P. Hill Site 71A calibration lane. In the target column &quot;f-clut&quot; and &quot;nf-clut&quot; stand for ferrous clutter and non-ferrous clutter.</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Alternative architectures for identity fusion.</td>
<td>18</td>
</tr>
<tr>
<td>4.1</td>
<td>Hypothetical example showing the lack of discriminatory power in representation features.</td>
<td>45</td>
</tr>
<tr>
<td>4.2</td>
<td>Feature extraction using the Fisher discriminant criterion.</td>
<td>50</td>
</tr>
<tr>
<td>4.3</td>
<td>Discriminating two classes with equal means using the Fukunaga-Koontz (FK) method.</td>
<td>55</td>
</tr>
<tr>
<td>4.4</td>
<td>Discriminating two classes with equal mean using the feature extraction based on Bhattacharyya $\mu_2$ optimization.</td>
<td>59</td>
</tr>
<tr>
<td>5.1</td>
<td>A model of a neuron.</td>
<td>68</td>
</tr>
<tr>
<td>5.2</td>
<td>The typical neural network architecture.</td>
<td>71</td>
</tr>
<tr>
<td>5.3</td>
<td>Receiver operating characteristic (ROC) curves.</td>
<td>73</td>
</tr>
<tr>
<td>6.1</td>
<td>The layout of the ESL mine surrogate field.</td>
<td>79</td>
</tr>
<tr>
<td>6.2</td>
<td>Map of the sensor sampling locations over the minefield. Vertical line segments represent EMI and GPR scan directions. Trapezoids represent the IR images. Circles denote grid points, and the squares indicate the fiducial markers.</td>
<td>81</td>
</tr>
<tr>
<td>6.3</td>
<td>Typical raw EMI waveforms collected over metal-free ground, a mine-surrogate, and a metal disk.</td>
<td>84</td>
</tr>
</tbody>
</table>
6.4 An IR image remapped to ground coordinates. The encircled grey shape corresponds to a buried mine surrogate. The dark square masses are fiducial markers.

6.5 Detection maps for the individual sensors. The dark masses represent detections. Circles give the mine locations. Detection cells are represented by squares.

6.6 ROC curves obtained for individual sensors.

6.7 ROC curves obtained for sensor fused classification.

6.8 Detection map obtained using soft decision-level fusion.

6.9 Detection map obtained using feature-level fusion.

6.10 Detection maps obtained for AND, OR, and majority fusion of hard decisions.

6.11 Detection map obtained using multisensor data fused with an ad hoc hybrid fusion scheme.

6.12 The operating points of AND, OR, majority and the ad hoc fusion based mine detection.

6.13 The effect of sensor reliability estimates $P_{F_i}$ and $P_{M_i}$ on the performance of hard decision-level fusion with the optimal rule. The subscripts $i = 1, 2, 3$ correspond to EMI, IR and GPR respectively.

6.14 A posteriori probability ratio and detection map obtained using hard decision-level fusion assuming $(\text{mines}) = 0.0393$ and $(\text{clutter}) = 0.9607$.

6.15 A posteriori probability ratio and detection map obtained using hard decision-level fusion when the prior probability of mines is overestimated using $(\text{mines}) = 0.2857$ and $(\text{clutter}) = 0.7143$.

6.16 A posteriori probability ratio and detection map obtained using hard decision-level fusion when the prior probability of mines is underestimated using $(\text{mines}) = 0.0036$ and $(\text{clutter}) = 0.9964$.

7.1 Map of the Fort A. P. Hill Site 71A calibration lane.
7.2 Measurement configuration used at Fort A. P. Hill for EMI and GPR sensors. ....................................................... 114

7.3 Plan view of the measurement configuration used at Fort A. P. Hill for IR sensors. Dashed lines represent known distances. IR camera height also is known. ........................................................... 115

7.4 Some example EMI spatial signatures obtained from A. P. Hill calibration lane data. ..................................................... 116

7.5 Classification performance of the peak value of an EMI spatial response. 120

7.6 Classification performance of the peak of an EMI spatial response for separating metal objects from nonmetal. ....................... 120

7.7 Classification performance of the peak of an EMI spatial response for separating large metal objects (> 5 g) from other objects. ........ 121

7.8 Comparison of original and subsampled EMI spatial responses for a VS-50 mine. .......................................................... 123

7.9 The representation and the classification of EMI spatial responses based on KL-coefficients. ............................................... 124

7.10 The representation and the classification of EMI spatial responses using DCT coefficients. ................................................ 126

7.11 The wavelet representation of the EMI spatial responses. ........... 128

7.12 Fisher discriminant feature extraction from the EMI spatial responses. 130

7.13 Fukunaga-Koontz feature extraction from the EMI spatial responses. 132

7.14 Extended Fisher Criterion (EFC) feature extraction from the EMI spatial responses. ...................................................... 133

7.15 Feature extraction based on the covariance difference term for the Bhattacharyya distance. ................................................. 134

xiv
7.16 Feature extraction based on Sammon’s optimal plane mapping. In (b) and (c) the dark areas represent the decision region for clutter and the light areas represent the decision region for mines. ........................................ 136

7.17 Feature extraction based on Fisher with Fukunaga-Koontz method. ........................................ 137

7.18 Raw GPR map generated from data collected over a one inch deep TS-50 mine (B10) at Fort A. P. Hill. ................................................................. 142

7.19 Linear weighting functions used for blending $A_i(x)$ and $B_i(x)$ functions. 145

7.20 The ground reflection clutter profile estimated by applying the polynomial model-based approach to measured GPR data........................................ 150

7.21 The GPR map obtained by removing the estimated ground-reflection from the measured data. This data corresponds to a one inch deep TS-50 mine at location B10 of the Fort A. P. Hill calibration lane. . . 150

7.22 The GPR map obtained by projecting the ground-reflection removed data to reduce the spatially stationary clutter. ........................................ 151

7.23 Comparison of raw and clutter reduced maps generated from data collected over a three inch deep VS-50 mine (E4) at the Fort A. P. Hill calibration lane. ................................................................. 153

7.24 The GPR maps generated from data collected over a surface-flush M-14 plastic anti-personnel mine (E6) at Fort A. P. Hill. ................................. 154

7.25 The correlation maps generated from data collected over a surface-flush M-14 plastic anti-personnel mine (E6) at Fort A. P. Hill. .................... 155

7.26 The GPR map of a 1.75 inch deep M-19 plastic anti-tank mine at location D6. The sub-image used for feature extraction is indicated by the dashed box. ................................................................. 157

7.27 Distribution of the correlation coefficient feature, and the ROC curve obtained by applying a threshold classifier to this feature. ...................... 158

7.28 The GPR sub-image containing the signature of a one inch deep TS-50 mine at grid location B10, and its discrete cosine transform (DCT). . 159
7.29 Representing the GPR sub-image of Figure 7.28(a) using the ten largest DCT coefficients. .......................................................... 160

7.30 Representing the GPR sub-image of Figure 7.28(a) using the 40 largest DCT coefficients. .......................................................... 161

7.31 The distribution of the peak values of GPR sub-images and the ROC curve for classification based on the peak values using a threshold classifier. .......................................................... 162

7.32 The wavelet representation of single look-angle GPR data acquired directly over a target. .......................................................... 164

7.33 The Fisher discriminant feature extracted from subsampled single look-angle GPR data. .......................................................... 165

7.34 The ROC curves obtained for classification using IR features. ......... 167

7.35 The ROC curves obtained for GPR classification based on the Fisher and peak value features taken individually and fused. ............. 169

7.36 A hypothetical distribution of samples in a two-dimensional feature space. .......................................................... 171

7.37 The ROC curves obtained for EMI classification using a nearest-neighbor classifier and the Fisher discriminant feature, Sammon's optimal features and Fisher with Fukunaga-Koontz features. ............. 171

7.38 The ROC curves obtained for classification based on the individual IR features, $a, b, S$ and $E$ and the fused features using a nearest-neighbor classifier. .......................................................... 173

7.39 The ROC curves for classification based on the EMI Fisher feature, the GPR Fisher feature and the fused feature vector. ............. 174

7.40 The ROC curves for classification of the EMI Fisher feature, nine GPR wavelet coefficient features and the fused feature vector. ............. 175

7.41 The ROC curves obtained for classification based on the EMI and GPR Fisher features, IRRIS sensor's $S$ feature and the fused feature vector using a nearest-neighbor classifier. ............. 177
7.42 The performance comparison of feature-level and decision-level fusion of GPR and EMI sensor data. ................................. 178

A.1 The graphical user interface developed for sensor fusion studies. . . . 190
CHAPTER 1

INTRODUCTION

It is estimated that more than 100 million landmines are buried in over 70 countries around the world. Safe, effective detection and removal of landmines is essential to make these mine-infested regions habitable again. The extreme risk posed by mines implies that a mine detecting sensor should not miss a detection. This requirement and the clutter-rich environment in which demining sensors operate lead to high false alarm rates. Nonetheless, very low false alarm rates are required because of both the large land areas where mine clearance is necessary and the high cost of excavating a putative mine. Manual prodding, the most common method for mine removal during peacetime (humanitarian demining), is quite effective but slow and extremely dangerous. The only noncontact mine detection sensors currently in wide use are electromagnetic induction (EMI) sensors. Although some EMI sensors are sensitive enough to detect mines containing very little metal, the rate of false detections caused by metallic clutter is often unacceptably high. Several other sensor technologies have been considered for the detection of mines, including ground penetrating radar (GPR) [1, 2], infra-red (IR) imaging [3, 4], nuclear quadrupole resonance (NQR) [5, 6], and sensors of acoustically induced surface vibrations [7, 8, 9]. At present, none of
these technologies has proven to be acceptably effective and reliable for the wide range of mines and environmental conditions encountered in the field.

Because diverse physical properties are measured by different sensors, multisensor fusion is attractive for improving mine detection. Multisensor fusion, has been successfully used in many military, space, industrial, and medical applications [10, 11]. The raw sensor data, features extracted from the data, or detection decisions derived from the data may be fused. Because both feature extraction and decision making usually result in loss of information, fusing the data at the lowest possible level (i.e., closest to the raw data) is desirable to achieve the highest classification accuracy. Nonetheless, data and feature-level fusion in mine detection are hindered by several factors. First, sensors used for mine detection typically produce noncommensurate data, i.e., data which are not of comparable form. Even when the data are commensurate, the sensors may operate on different platforms, leading to noncoincident sampling and problems with data registration. Since fusion comprises combining information from multiple sources regarding a specified phenomenon (e.g., the presence of a mine) at a specific location, any practical fusion technique should accommodate noncoincident sampling. Decision-level fusion alleviates some of these concerns, but it is suboptimal, in principle, since if a target is not detected by all the sensors, it will not experience the full benefits of fusion.

Many previous works on sensor-fused landmine detection were limited to decision-level fusion, but some works have applied feature-level fusion to coincidently sampled multisensor data collected in controlled environments [12, 13, 14, 15, 16, 4]. This dissertation extends the prior studies by developing a new feature-level fusion technique.
that can accommodate noncoincidently sampled data, and applying it to multisensor data collected under realistic conditions.

1.1 Objectives of this Dissertation

The primary objective of this dissertation is to develop techniques to apply feature-level sensor fusion to the landmine detection problem. In particular, it is our goal to develop a feature-level fusion technique that can accommodate noncoincidently sampled multisensor data. Although data-level fusion is potentially more accurate than feature-level fusion, it is impractical to apply data-level fusion to noncommensurate sensors.

Another objective of this work is to compare the performance of feature-level and decision-level fusion for mine detection. We also compare soft decision-level fusion with hard decision-level fusion. Estimates of prior probabilities and sensor reliabilities are used by many fusion schemes, and we investigate the sensitivity of the fused decisions to the estimates.

Although (in principle) a system can produce perfect performance by fusing sensors that are ineffective individually, in practice the best performance is achieved when the individual sensors work well. Therefore, it is important to obtain the best performances from each sensor through proper signal processing and feature extraction. Identifying and implementing effective feature extraction techniques are additional goals of this research.

A modular, flexible fusion architecture is desirable in both applications and research. The fusion architecture should permit the user to add or delete sensors and/or
features without modifying the software. Developing and implementing such an architecture is the final objective of this work.

1.2 Organization of the Dissertation

A brief summary of the contents of the dissertation follows:

Chapter 2 gives an overview of the landmine problem and demining practice. Mines are characterized, and some statistics illustrating the extent of the global landmine problem are reviewed. Some existing demining sensor technologies are briefly discussed.

The concept of multisensor fusion is introduced in Chapter 3. Different fusion architectures are described and compared. The benefits and limitations of fusion are discussed. A new feature-level fusion algorithm for noncoincidently sampled sensor data is developed. Optimal decision-level fusion algorithms are reviewed.

Feature extraction is an important part of feature-level fusion and many other pattern recognition tasks, and feature extraction techniques are reviewed in Chapter 4. Both physical and nonphysical features are described. Representation and classification feature extraction methods are discussed. Simulated data are used to illustrate the benefits and limitations of these techniques.

Chapter 5 provides a review of pattern classifiers. An optimal Bayes classifier and several examples of parametric and nonparametric classifiers are described. This discussion includes quadratic and linear classifiers, distance classifiers, nearest-neighbor and neural classifiers. Supervised training and classifier performance evaluation are also discussed.
Chapter 6 presents a study conducted using multisensor data collected at a very challenging surrogate mine field. Data in this study were acquired at noncoincident sampling points using GPR, EMI and IR sensors mounted on separate platforms. Empirical models developed for EMI and IR sensors based on experimental data are described. Data are fused using the new feature-level fusion algorithm, a soft decision-level fusion rule developed in Chapter 3, and conventional hard decision-level fusion. Performance comparisons are made among individual sensors and fused suites.

Chapter 7 discusses a study based on multisensor data collected at a mine test field in Fort A. P. Hill, Virginia. Data from GPR, EMI and two IR sensors are fused, and results are analyzed in this study. A powerful, new GPR clutter reduction algorithm, which was developed during this study, is described. Many feature extraction methods are applied to the measured data, and the effectiveness of these features are compared. The performance of different fusion methods are investigated and the results analyzed.

Chapter 8 summarizes the work described in this dissertation and highlights the main contributions. Conclusions are drawn from the results obtained. Suggestions for possible future work are also described.

A modular, flexible fusion architecture and a graphical user interface (GUI) developed during this research are described in Appendix A. Different sensor and feature combinations for fusion are easily selected using the GUI without having to modify the existing software. Adding new sensors, features, classifiers, and fusion algorithms are also simplified by this architecture.
CHAPTER 2

LANDMINES

In this chapter we present an overview of landmines, the global landmine problem and demining technologies. Section 2.1 motivates this research by presenting some statistics that highlight the magnitude of the threat posed by mines. Section 2.2 describes different types of mines and their characteristics. In Section 2.3 some of the existing demining sensor technologies are briefly reviewed. This review of demining technologies illustrates the complementarity of certain sensors, which motivates the use of multisensor fusion.

2.1 The Landmine Problem

Approximately 110 million landmines are estimated to be buried in over 70 countries around the world [17, 18]. Nearly 100 million mines are estimated to be in stockpiles, and another five to ten million new mines are believed to be manufactured annually. It is estimated that approximately two million new mines are buried every year. While the cost to purchase a new anti-personnel mine is between US$ 10 to US$ 30, the average cost of clearing a mine is between US$ 300 to US$ 1000. It is
believed that new mines are laid at a rate that is about 25 times that of mine clearing. According to the United Nations' estimates, at the current rate of demining, it is expected to take 1100 years and US$ 70 billion to clear the existing landmines [17].

Over 1 million people are estimated to have been killed or maimed by mines since 1975. On the average, 70 people are killed or injured by mines every day. Most of the victims are civilians who had no part in the conflicts that led to the use of these mines. Because of the indiscriminate use of landmines in farmlands, roads, waterways, etc., millions of people are denied access to their homes and livelihood, leading to poverty and famine.

2.2 Landmine Characteristics

Landmines are explosive devices placed on or just below the surface of the ground to cause death or injury to personnel or damage to vehicles. Over 200 different types of landmines are known. Mine dimensions cover a wide range; some mines have widths as small as 50 mm, while others are as large as 400 mm. Weights of mines may range from around 20 g to about 15 kg. Mines come in a variety of shapes including flat circular disks, cylindrical, rectangular or square box types, and irregular shapes. Mine explosives include TNT, RDX, Tetryl, Compound B and Amatol. Mine outer case materials can be metal, plastic, Bakelite and wood. This variation in size, shape and composition contributes to the difficulty of mine detection.

Mine explosions may be initiated by direct pressure, by trip wire, or through the use of proximity sensing switches. In most pressure activated mines, pressure on the lid crushes a fuse that contains a friction-sensitive chemical. This chemical causes a spark to shoot into the detonator, which sets off the main explosion. In
trip wire controlled mines, pulling the trip wire releases a spring loaded firing pin, which contacts a percussion cap setting off a spark to initiate the blast. In addition to these conventional triggering mechanisms mines can also be triggered by integral anti-handling devices. These devices obviously increase the dangers of demining.

Landmines are usually classified according to their intended target. Anti-tank mines are those used to destroy enemy tanks and other vehicles. The first landmines, developed and used during World War I, were anti-tank mines. Anti-tank mines are large and have typical diameters of 200-400 mm, heights of 50-150 mm and normally weigh over 10 kgs. Anti-tank mines are usually buried at depths of 5-40 cm. Anti-personnel mines were originally developed to discourage attempts to clear anti-tank mines. Today, however, anti-personnel mines are used regularly for protecting the perimeters in military installations, for slowing enemy advances, and for terrorizing civilians. Because anti-personnel mines need to be activated by the relatively small weight of a human, they are buried at very shallow depths, often less than 10 cm, or even placed on the surface. During the Vietnam war, large quantities of anti-personnel mines were dropped into enemy territory by aircraft. Whether originally buried or dropped on the surface, the depths of mines can change over time due to flooding, soil erosion and ground settling.

2.3 Mine Detection Sensors

The widespread threat of landmines has resulted in an increased interest in demining. Two forms of demining exist, namely, military demining and humanitarian demining. Military demining, also known as breaching, involves rapid clearing of a
path through a mine laden area for troop movement during war. Breaching often utilizes mechanical means such as vehicle-mounted ploughs and flails to quickly displace or destroy mines along a path of vehicle width. Such demining methods are not acceptable for humanitarian demining, because of limited effectiveness. Because of the inherent dangers of landmines, humanitarian demining demands essentially perfect detection. On the other hand, because vast land areas have to be cleared, false alarm rates should be very low. In the remainder of this section we describe several mine detection techniques. Additional reviews of these sensors appear in [19, 1, 20].

**Manual Prodding**

Currently, most humanitarian demining is done by manual prodding, which involves inserting a thin rod into the ground to determine the presence and shape of buried objects. To ensure that the ground is free of mines, each square meter of ground is probed about 400 times. Although the probability of detection using this method is high, it is slow and extremely dangerous. Furthermore, prodding is difficult in hard soils.

**Electromagnetic Induction (EMI) Sensor**

Virtually all land mines in use today contain some metal, which may be detected by an EMI sensor. The quantity of metal used varies widely from several kilograms to a few hundred milligrams. Excessive false alarm rates arise when detecting mines with small metal content as a result of the large amount of metallic clutter of anthropic origin found in most minefields. Yet another source of false alarms for low-metal mines are the small variations in sensor response due to changes in soil moisture content and surface topography.
In a typical EMI sensor, such as the Schiebel AN 19/2 used by US military, a transmit coil is energized by current pulses with duration on the order of milliseconds. These pulses build up a magnetic field, which induces eddy currents in metallic objects located near the sensor. The secondary field generated by these eddy currents is detected by a receiving coil. The rate of decay in the secondary field is slow when a large metallic object is present and fast in a metal-free environment. This difference in the decay rates is utilized to determine the presence of a metal object. For rapid coverage of large areas, arrays of EMI sensors have been demonstrated [20].

**Ground Penetrating Radar (GPR)**

A GPR operates by transmitting an ultra-wideband electromagnetic field into the ground and receiving the scattered field. Frequencies from hundreds of MHz to several GHz are typically used in mine-detection GPR systems. A GPR detects the presence of reflecting boundaries in the soil. Such boundaries occur at the surface of the ground and at most buried objects. A GPR will detect both metallic and metal-free objects and, hence, it is a useful adjunct to an EMI sensor. Unfortunately, voids, changes in soil density and moisture content, and most natural buried objects are also detected, leading to a high false alarm rate.

For a down-looking GPR reflections at the air-to-ground interface pose a problem when detecting shallow mines, and much GPR research has dealt with reduction of surface-reflection clutter [1, 21, 22, 23]. The problem of finding effective GPR features has also received significant attention. Chan et al. [2] suggested complex natural resonances (CNR) as a possible feature for identification. Although these resonances are useful in the identification of unexploded ordnances, their use in mine detection has not been successful because of the very low Q of mine resonances. Nag
and Peters [24] proposed and developed a ramp response feature for GPR target identification, but an experimental demonstration has not yet been presented due to hardware limitations. Azimi-Sadjadi and coworkers [16, 25] reported good results for target identification using principal components and bispectra of GPR data. Because the problem considered was simple, involving only wood and nylon objects, the value of the technique for real mine detection is still unknown.

**Infrared (IR) Sensors**

The natural heating and cooling of soil that occurs over a diurnal cycle causes thermal energy to flow into and out of the earth. The presence of a buried object changes this thermal flow, leading to surface temperature anomalies above the buried object at certain times in the diurnal cycle. These anomalies can be detected in soil temperature maps created by a suitable IR camera [3, 26, 4, 27]. Since a mine's thermal properties are only weakly related to its metal content and electromagnetic properties, passive IR imaging is presumed to be a good complement to EMI and GPR sensors. In addition to buried mines, naturally occurring buried objects (e.g., stones), variations in soil mineral and water content, and variations in solar illumination (due to shadowing by foliage, for example) also produce surface temperature anomalies. Camera sensitivity is of critical importance in IR mine detection, since temperature differences induced by natural solar heating over buried objects are on the order of a few kelvins or less.

**Acoustic Sensors**

Several approaches for detecting landmines using acoustics have been considered. Transmitting ultrasound waves into the ground and detecting the waves reflected from
boundaries is the conventional approach. This method faces problems similar to those encountered in GPR, because of the large reflection at the air-ground interface and undesired reflections from non-mine objects. An acoustic impulse approach based on a single transmitter and multiple receivers also has been demonstrated [28]. Some recent approaches measure vibrations of the ground induced by wideband acoustic sources, lasers, and water-jets to detect mines [7, 8, 9]. Mechanical resonances in the mine are detected by the motion of the overlying ground.

**Trace Explosive Detection**

Dogs have the ability to recognize traces of explosives, and they are sometimes employed in detecting mines. Because these animals get distracted and tire easily, detection using dogs is not suitable for clearing large stretches, but more suited for confirmation of targets detected by other methods. Another drawback in using dogs for demining is that training is costly and time consuming.

Research is being conducted to develop artificial systems (electronic “sniffers”) that can mimic the dogs’ ability to detect trace amounts of explosives. Studies in chemiluminescence [29], biosensors [30], ion mobility spectrometry [31] and antibodies [32] have made some progress in trace explosive detection, but are not yet developed enough to be used for mine detection. One drawback of trace explosive detection is that the explosive vapors released from a mine can spread over a radius of several meters, making it difficult to locate the mine.
**Bulk Explosive Detection**

Significant developments have also occurred in bulk explosive detection. Unlike trace explosive detection, bulk explosive detection can pinpoint the true mine locations. In principle, this technology is the most desirable method to detect mines, because it detects the presence of bulk explosives, which are unique to mines. This characteristic is in contrast to other sensor technologies such as EMI and GPR, which are prone to produce false alarms for harmless objects.

Thermal neutron activation (TNA) is one technique being considered for bulk explosive detection. A TNA sensor uses a radioactive Californium-252 source to generate thermal neutrons, which pass through the soil and interact with a buried object. If the object contains nitrogen, these nuclei are excited and emit gamma rays of a characteristic energy. Because explosives contain a high concentration of nitrogen compared to soil, explosive can be easily detected [33, 34]. Good performance has been reported in detecting large AT mines, but detection of AP mines with small amounts of explosive has not been reliable. System complexity, operator safety issues, limited penetration depth and long data collection time are some of the drawbacks of TNA detectors.

Another approach for bulk explosive detection is nuclear quadrupole resonance (NQR). This method is based on the resonant response of electric quadrupole moments present in some nuclei. Although promising results have been obtained in detecting RDX [5, 6], the signature of TNT, the most commonly used explosive in mines, is much weaker. Research on this technology is continuing.
2.4 Summary

Some statistics regarding the global mine problem were presented in Section 2.1, which makes evident the need for developing effective, safe, low-cost mine detection techniques. Landmines were characterized in Section 2.2. Different types of mines, their triggering mechanisms and compositions were discussed. Mine detection methods, their merits and limitations were described in Section 2.3. Although many demining technologies are available, the wide variety of mines and environmental conditions preclude these technologies from attaining the high probability of detection and low false alarm rate required in humanitarian demining.
CHAPTER 3

MULTISENSOR DATA FUSION

Multisensor data fusion has been the subject of much study for military, space, medical, and industrial applications since the early 1980's, because of the many potential benefits it offers [11, 10]. In this chapter we present the mathematical bases for several forms of fusion. Section 3.1 presents some concepts that underlie sensor fusion and reviews the potential benefits. Section 3.2 describes different sensor fusion architectures. Section 3.3 discusses the application of sensor fusion to landmine detection and highlights some difficulties associated with fusing demining sensor data. Previous work on the fusion of demining sensor data is briefly reviewed. In Section 3.4 is developed a new feature-level fusion algorithm that can accommodate noncoincidently sampled multisensor data. Decision-level fusion techniques are reviewed in Section 3.5, followed by a useful extension to a previously developed hard decision fusion rule.

3.1 Sensor Fusion Preliminaries

Multisensor fusion refers to the combination of data from multiple sensors to achieve improved awareness of the environment. The objective of fusion, the types of sensors used, the sensor suite configuration, and how much the individual sensor data
are processed before combining are all factors that define the fusion strategy. These factors also determine what benefits are achieved by fusion.

If all sensors to be fused are identical and observe the same space, then the fusion process provides redundancy and, hence, more reliability. Sometimes, fusion of identical sensors may provide the system with information that is not available from a single sensor. One such example is the binocular vision in humans, in which two identical sensors are fused, leading to a perception of depth, which does not exist if only one eye is used.

If multiple sensors that sense different physical properties observe the same region of space, then the fusion of these data provides complementarity rather than redundancy. This is the fusion paradigm of greatest interest for landmine detection.

Sensor fusion may be used for achieving different objectives such as the detection and classification of targets, position and/or velocity estimation, and tracking. Sensor fusion for improving the identification or classification of objects has been called identity fusion [35]. Identity fusion has found applications in identifying aircraft, ships, and ground-vehicles in military reconnaissance, identifying industrial parts in autonomous assembly, and diagnosing abnormalities in medical examinations. The application of sensor fusion to improve discrimination of mines from clutter is also an example of identity fusion.

3.2 Sensor Fusion Architectures

Three architectures for identity fusion can be defined based on the amount of processing performed on the sensor data before fusion. These architectures, commonly referred to as data-level fusion, feature-level fusion, and decision-level fusion, are
illustrated in Figure 3.1. Features may or may not be extracted in data-level and decision-level fusion. In the context of this figure “pattern classification” refers to a mapping from features (or data) to a decision. In general, a major portion of the effort in applying multisensor fusion to target classification comprises feature extraction and pattern classification, which are discussed in detail in Chapters 4 and 5.

**Data-Level Fusion**

In the data-level fusion architecture a minimal amount of processing is performed on each sensor’s data before they are combined. Processing is usually limited to calibration, alignment, etc., which are essential to make the data useful. The fusion operation in a data-level fusion architecture may comprise stacking, weighted averaging, or Kalman filtering [11]. The classification decisions can be made on the fused data or on features extracted from the fused data. Data-level fusion is practical only if the sensors are commensurate, i.e., if their data formats and dimensions are identical. As a result, it is most often used with identical sensors. Because data are available at the fusion center with minimal information loss, this architecture has the potential to provide the most accurate fused decisions. Nonetheless, data-level fusion imposes a high computational burden on the fusion processor, because large volumes of high dimensional data may need to be handled. If, in addition, the sensors are in different geographical locations, the need to transmit the raw data to the fusion center may demand a large communication bandwidth.

**Feature-Level Fusion**

The aforementioned drawbacks of data-level fusion can be reduced to some extent by employing feature-level fusion. In this architecture features extracted from the
Figure 3.1: Alternative architectures for identity fusion.
raw data of each sensor are merged to form a fused feature vector. If all sensors coincidentally sample the observation space, then fusion amounts to simply concatenating the feature vectors of all sensors. If the data are noncoincidently sampled, however, feature-level fusion is not straightforward. Developing a feature-level fusion technique that can accommodate noncoincidently sampled data was a primary objective of this dissertation.

Once the feature vectors are combined, classification decisions are made based on this fused feature vector. A feature-level fusion architecture can accommodate fusion of noncommensurate sensors, since feature extraction brings the data to a common format. Because feature vectors generally have a much lower dimension than the corresponding raw data, the computational demand at the fusion center and the communication bandwidth requirements can be significantly lower than those in data-level fusion. The penalty, however, is the possibly lower accuracy of the fused decisions, because of the loss of information occurring during feature extraction at each sensor.

**Decision-Level Fusion**

In the third architecture, classification decisions made separately at the individual sensors are fused at the fusion center. A variety of techniques have been developed for fusing decisions, which may be binary decisions or detection probabilities [35, 36, 37]. Decision-level fusion is, in principle, the least accurate because of the additional loss of information that occurs when sensor data are reduced to decisions. Nonetheless, other considerations such as communication bandwidth and low computational demand at the fusion center make decision-level fusion attractive in many situations, especially those dealing with distributed detection. Decision-level fusion also requires relatively
simple algorithms making it easy to retrofit to an existing suite of sensors. Because fusion in a decision-level architecture takes place after the individual sensor decisions are made, it is also referred to as post-detection fusion. By analogy, data-level and feature-level fusion are called pre-detection fusion.

3.3 Sensor Fusion for Landmine Detection

As noted in Section 3.1, fusion of multiple sensors that observe complementary physical phenomena is attractive for mine detection, but the process is hindered by several factors. Mine-detecting sensors typically produce noncommensurate data, i.e., data are not of comparable form, which eliminates the possibility of data-level fusion. By way of example, three commonly used mine sensors are infrared cameras, which produce surface imagery; ground penetrating radars which produce depth profiles along a line scan; and metal detectors, which produce scalar measurements along a line scan. Even when the data are commensurate, the sensors may operate on different platforms, leading to noncoincident sampling and problems with data registration. Since fusion comprises combining information from multiple sources regarding a specified phenomenon (e.g., the presence of a mine) at a specific location, accommodating noncoincident samples is a necessity.

Some of these concerns can be alleviated by decision-level fusion. Section 3.5 describes the formulation of some well known “hard” decision level techniques. An extension is suggested therein to one hard decision fusion rule, resulting in a “soft” decision-level fusion rule, which is more effective in mine detection. In principle, decision-level fusion is suboptimal, since if a target is not detected by all sensors,
it will not experience the full benefits of fusion. This situation and stringent detection requirements led us to explore other methods of fusion. Section 3.4 develops a new feature-level fusion approach, which can accommodate noncoincidently sampled multisensor data.

**Previous Work on Multisensor Fused Demining**

Fusion of demining sensor data has been investigated previously. Chauduri et al. [12] developed a demining system in which co-registered GPR and EMI sensors were fused using several decision-level fusion approaches. Brusmark et al. [13] demonstrated the decision-level fusion of coincidently sampled GPR and EMI sensor data collected over targets buried in a sand box. Clark et al. [4] fused two IR sensors of different wavelengths, which were registered using a set of fiducial markers. A neural network based feature-level fusion approach was used in that work. Miao et al. considered the fusion of six co-registered IR images corresponding to different wavelengths [16]. Each image was separately classified with neural networks and the resulting decisions were fused using both majority voting and a consensus theory approach. Weisenseel et al. [14] studied decision-level fusion of GPR and EMI data using synthetic data. The problem was greatly simplified by assuming that the mines were detectable by both sensors, metallic clutter by EMI alone, and nonmetallic clutter by GPR alone. Only an AND rule was considered for fusion. Breejen et al. [15] compared several methods for decision-level fusion of EMI, GPR and IR data. They used the same data for both training and testing the classifiers, which yields overly optimistic results. Performance of a vehicle mounted mine detector employing a GPR and EMI arrays, and video and IR cameras, developed by EG&G Management Systems, Inc. was reported in [38]. While surface emplaced anti-tank mines and mines
with significant metal content were easily detected, buried plastic mines were found to be challenging. Development and testing of a vehicle-mounted multisensor demining platform that uses EMI, GPR, and IR for detection and a TNA sensor for target confirmation is underway at the Canadian Defence Research Establishment [39].

3.4 Feature-Level Fusion of Noncoincidently Sampled Data

In this section we consider feature-level fusion of noncoincidently sampled multisensor data. First, the problem is defined and the optimal solution is described. More practical suboptimal forms are then described. Closed-form expressions are derived for the interesting special case of Gaussian features.

3.4.1 Problem Definition

Suppose that a suite of $N_S$ counter-mine sensors are used to acquire data in a region. Sensor $i$ acquires $J_i$ data samples $d_i$ at locations $R_{ij}$, which we denote $d_{ij} = d_i(R_{ij})$. Fusion is to be performed on the data set $\mathcal{D} = \{d_i(R_{ij}), i = 1, \ldots, N_s; j = 1, \ldots, J_i\}$. We assume that different sensors produce samples that differ in number $J_i$, sample positions $R_{ij}$, format, and the dimensionality of their data.

Let $\mathbf{R}$ be a point where we wish to determine the presence of a mine. We refer to $\mathbf{R}$ as the “interrogation point.” We form the $K$ hypotheses $H_k(\mathbf{R})$, $k = 1, 2, \ldots, K$ for the fused suite regarding the presence or absence of various types of mines at $\mathbf{R}$. We may also include in the set $\{H_k(\mathbf{R})\}$ discrete clutter objects, e.g., buried rocks, metallic clutter, etc. The explicit dependence of these hypotheses on the interrogation point $\mathbf{R}$ is omitted hereafter in the interest of brevity.

The decision criterion for the fusion processor is the Bayes risk. At each point $\mathbf{R}$ we wish to determine the hypothesis $H_k$ that minimizes the Bayes risk $\mathcal{R}$ conditioned
on the data $\mathcal{D}$, viz:

$$k = \arg \min_j \mathcal{R}_j(\mathcal{D}) = \arg \min_j \sum_{\ell=1}^{K} C_{j\ell} \Pr(H_\ell|\mathcal{D})$$

(3.1)

where $\mathcal{R}_j(\mathcal{D})$ is the risk of selecting hypothesis $H_j$ given data $\mathcal{D}$, and $C_{ij}$ is the cost of choosing $H_i$ when $H_j$ is true. For the special case $K = 2$ (binary detection) or when all the costs are equal, it is sufficient to compute likelihood ratios and to compare them to a common threshold value. In this work, however, we will use the general formulation given above, since multiple hypotheses may be involved and equal costs are inappropriate for the mine-detection problem.

### 3.4.2 Formulation

We assume that sensor $i$ acquires at a point $\mathbf{R}_{ij}$ the data $d_i(\mathbf{R}_{ij})$ which lies in the vicinity of a putative target of type $k$ at $\mathbf{R}$. This data are modeled as a signal $G_{ik}(\mathbf{R}_{ij}; \Theta_{ik})$ corrupted by additive clutter (and noise) $N_i(\mathbf{R}_{ij})$, viz:

$$d_i(\mathbf{R}_{ij})|H_k = G_{ik}(\mathbf{R}_{ij}; \Theta_{ik}) + N_i(\mathbf{R}_{ij})$$

(3.2)

where $\Theta_{ik}$ is a (possibly random) feature vector (i.e., a set of model parameters) that describes the signature of targets of type $k$ for sensor $i$. We assume that the clutter $N_i(\mathbf{R})$ is independent of the parameter $\Theta_{ik}$. Note that if the clutter discretues are also assigned specific hypotheses $H_j$, the phenomena modeled by $N_i(\mathbf{R})$ can be relatively smooth.

For each sensor the signature of a mine has a finite spatial extent and, hence, only a subset of $\mathcal{D}$ may be relevant to interrogation of the region around $\mathbf{R}$. Let the relevant data for sensor $i$ be given by subsequences $j_1, j_2, \ldots, j_{M_i}$ with length $M_i$. It
is convenient to assemble this reduced data set into the vectors

\[
D = \begin{bmatrix}
d_1(R_{1j1}) & \cdots & d_1(R_{1jM_1}) & d_2(R_{2j1}) & \cdots & d_N(R_{NjM_N})
\end{bmatrix}^T \tag{3.3}
\]

\[
G_k = \begin{bmatrix}
G_{1k}(R_{1j1}; \Theta_{1k}) & \cdots & G_{1k}(R_{1jM_1}; \Theta_{1k}) & \cdots & G_{Nk}(R_{NjM_N}; \Theta_{Nk})
\end{bmatrix}^T \tag{3.4}
\]

\[
N = \begin{bmatrix}
N_1(R_{1j1}) & \cdots & N_1(R_{1jM_1}) & N_2(R_{2j1}) & \cdots & N_N(R_{NjM_N})
\end{bmatrix}^T \tag{3.5}
\]

If a single datum \(d_i(R)\) from sensor \(i\) has length \(N_{Di}\), then the length of these column vectors is \(N_D = \sum_i N_{Di} M_i\). Using this notation, we have for the combined signal model

\[
D|H_k = G_k(\Theta_k) + N \tag{3.6}
\]

where the combined (concatenated) feature vector for all sensors is

\[
\Theta_k = [\Theta_{1k} \Theta_{2k} \ldots \Theta_{Nk}]^T \tag{3.7}
\]

To support both decision-level fusion and feature-level fusion with uncertain sample locations we include in \(\Theta_{ik}\) a position offset \(R_{0,ik}\) that describes the nominal “center” of the mine signature with respect to \(R\). We write

\[
\Theta_{ik} = [\theta_{ik} \ R_{0,ik}]^T \tag{3.8}
\]

In general, we take \(\theta_{ik}\) and \(R_{0,ik}\) to be independent in what follows. For the concatenated feature vector it will be convenient to separate signature parameters from position parameters in the form \(\Theta_k = [\theta_k \ R_{0,k}]^T\) where \(\theta_k = [\theta_{1k} \ldots \theta_{Nk}]^T\) and \(R_{0,k} = [R_{0,1k} \ldots R_{0,Nk}]^T\).

Feature-level fusion of the foregoing data can be formulated along two lines. In both cases we use the Bayesian approach, in which \(\theta\) is assumed to be random, with \(\theta_k \to \theta\). Consider first the direct approach using the a posteriori probabilities, which
avoids the difficult problem of estimating the prior probabilities. Minimizing the risk in equation (3.1) leads us to consider the a posteriori probabilities as an integral over the classifier \( \text{Pr}(H_k|\Theta) \)

\[
\text{Pr}(H_k|D) = \int d\Theta \text{Pr}(H_k|\Theta)f_{\Theta|D}(\Theta)
\]

This expression is exact, but it requires both knowledge of the density \( f_{\Theta|D} \) and that we perform an integral over a feature space of possibly high dimensions. To avoid these issues it is attractive to employ the maximum a posteriori (MAP) approximation. If the data \( D \) strongly imply the value \( \hat{\Theta} \), then \( f_{\Theta|D} \) will be strongly peaked about \( \hat{\Theta} \) and we have

\[
\text{Pr}(H_k|D) \approx \text{Pr}(H_k|\hat{\Theta}) \int d\Theta f_{\Theta|D}(\Theta) = \text{Pr}(H_k|\hat{\Theta})
\]

Conversely, if \( D \) provides no information about \( \Theta \) then \( f_{\Theta|D}(\Theta) = f_\Theta(\Theta) \) and we have \( \text{Pr}(H_k|D) = \text{Pr}(H_k) \), which is just the a priori probability.

Equation (3.10) indicates that to compute the probability of the hypothesis \( H_k \) at the interrogation point conditioned on the set of multisensor data \( D \) we need to extract the features \( \hat{\Theta} \) from \( D \), and then classify the features. Therefore, by making use of the spatial responses of targets, fusion of noncoincidently sampled data is cast into the same form as feature-level fusion of coincidently sampled data, i.e., a feature extractor followed by a pattern classifier.

A further approximation will permit us to separate the effects of features and position offsets. If we take the fused signature features \( \theta_k \) and the detected position \( R_{0,k} \) to be independent when conditioned on \( H_k \), then we obtain

\[
\text{Pr}(H_k|\hat{\Theta}) \approx \text{Pr}(H_k|\hat{\theta}) \frac{f_{R_0|H_k}(R_0)}{f_{R_0}(R_0)}
\]

25
In the absence of a priori information about mine position, it is reasonable to take $f_{R_0}$ to be uniform over some region around each putative sample position.

Feature-fusion can also be formulated using class-conditioned densities. Bayes' rule leads to

$$R_k(D) = \frac{1}{f_D} \sum_{j=1}^{K} C_{jk} \Pr(H_j)f_{D|H_j}(D|H_j) \quad (3.12)$$

The factor $[f_D]^{-1}$ is common to all $R_k(D)$ and, hence, our primary task is to evaluate the class-conditioned probabilities $f_{D|H_j}(D|H_j)$. Again using Bayes' rule, we find

$$f_{D|H_k}(D) = \int d\Theta f_{\Theta|H_k}(\Theta)f_{D|\Theta}(D) \quad (3.13)$$

Since $G$ is presumed known, the data $D$ conditioned on the parameter $\Theta$ has the same density as the clutter $N$, which leads to $f_{D|\Theta}(D) = f_N(D - G_k(\Theta))$ and

$$f_{D|H_k}(D) = \int d\Theta f_{\Theta|H_k}(\Theta)f_N(D - G(\Theta)) \quad (3.14)$$

We approximate this exact result as done in equation (3.10). Assuming that the same model is used for all classes and that $f_N(D - G(\Theta))$ has a well defined maximum for $\Theta = \hat{\Theta}$ leads to

$$f_{D|H_k}(D) \approx f_{\Theta|H_k}(\hat{\Theta}) \int d\Theta f_N(D - G(\Theta))$$

$$\approx (\text{const}) f_{\Theta|H_k}(\hat{\Theta}) f_{R_0|H_k}(\hat{R}_0) \quad (3.15)$$

where the constant term depends only on the data $D$, and we have made the additional assumption that $\theta_i$ and $R_{0,i}$ are conditionally independent.

For the two-class problem, a likelihood ratio formulation is possible. The likelihood ratio for a known signal with unknown parameters is given by

$$\Lambda(D) = \frac{f_{D|H_1}(D)}{f_{D|H_0}(D)} = \frac{\int d\Theta f_{\Theta|H_1}(\Theta)f_N(D - G_1(\Theta))}{\int d\Theta f_{\Theta|H_0}(\Theta)f_N(D - G_0(\Theta))} \quad (3.16)$$
Assume next that the same model is used for both hypotheses. Approximating the results as proposed above we have

\[
\Lambda(D) \approx \frac{f_{\Theta|H_1}(\hat{\Theta}) \int d\Theta f_N(D - G(\Theta))}{f_{\Theta|H_0}(\hat{\Theta}) \int d\Theta f_N(D - G(\Theta))} = \frac{f_{\Theta|H_1}(\hat{\Theta})}{f_{\Theta|H_0}(\hat{\Theta})}
\]

where \(\hat{\Theta}\) is the estimate of \(\Theta\) that maximizes \(f_N(D - G(\Theta))\). Detection is performed on the basis of the class-conditioned densities of \(\Theta\).

It is interesting to compare this formulation to the more conventional generalized likelihood ratio test (GLRT) [40, p. 38]. In the GLRT formulation the model parameters \(\Theta\) (sometimes referred to as “nuisance” parameters) are estimated using a maximum likelihood technique. Beginning with equation (3.16) we assume that \(f_{\Theta|H_k}\) has a well defined maximum at the ML estimate \(\hat{\Theta}_k\). This leads to the approximation

\[
\Lambda(D) \approx \frac{f_N(D - G_1(\hat{\Theta}_1)) \int d\Theta f_{\Theta|H_1}(\Theta)}{f_N(D - G_0(\hat{\Theta}_0)) \int d\Theta f_{\Theta|H_0}(\Theta)} = \frac{f_N(D - G_1(\hat{\Theta}_1))}{f_N(D - G_0(\hat{\Theta}_0))}
\]

Thus, GLRT detection is based primarily on the distribution of the model error. The model parameters are considered indirectly, via their contribution to the model error.

The estimator-correlator method [40, p. 79] uses a similar approach, in which the parameters are estimated and the resulting model signature is correlated with the data. For Gaussian statistics, detection is based on the correlation. The feature-level fusion algorithm described above could be referred to as an “estimator-classifier” technique in that detection (or classification) is based on the parameters \(\hat{\Theta}\) rather than the residual error \(D - G_k(\hat{\Theta})\).

### 3.4.3 Gaussian Approximation

The case of Gaussian statistics is of interest because it provides insight into the problems of feature-level fusion and because it can be used to efficiently simulate the
performance of certain fusion strategies. One finds that with certain assumptions, expressions for $f_{D|H_k}$ in (3.13) can be evaluated in closed form.

Consider first the case in which only the clutter $N$ has a Gaussian density. That is, the $N_s$ random variables $N_i(R_{ij})$ (each of which are evaluated at $M_i$ locations) have a density that is jointly normal with mean $\mu_N$ and covariance matrix $C_N$, viz:

$$N \sim \mathcal{N}(\mu_N, C_N).$$

Thus

$$D|H_k, \Theta \sim \mathcal{N}(\mu_N + G_k(\Theta), C_N)$$

(3.19)

where the mean clutter is given by

$$\mu_N = [\mu_1^T \mu_1^T \cdots \mu_1^T \mu_2^T \cdots \mu_N^T]^T$$

(3.20)

and the covariance matrix $C_N$ has $N_s \times N_s$ block elements $[\hat{C}]_{ij}$, viz:

$$C_N = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} & \cdots & \hat{C}_{1N_s} \\ \hat{C}_{21} & \hat{C}_{22} & \cdots & \hat{C}_{2N_s} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}_{N_s1} & \hat{C}_{N_s2} & \cdots & \hat{C}_{N_sN_s} \end{bmatrix}$$

(3.21)

The blocks define clutter relations among sensors and samples as follows:

$$\hat{C}_{ij} = \begin{bmatrix} \Gamma_{11}^{(ij)} & \Gamma_{12}^{(ij)} & \cdots & \Gamma_{1M_i}^{(ij)} \\ \Gamma_{21}^{(ij)} & \Gamma_{22}^{(ij)} & \cdots & \Gamma_{2M_i}^{(ij)} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{M_i1}^{(ij)} & \Gamma_{M_i2}^{(ij)} & \cdots & \Gamma_{M_iM_i}^{(ij)} \end{bmatrix}$$

(3.22)

In this result $M_i$ is the number of data samples near $R$ from sensor $i$, and

$$\Gamma_{rs}^{it} = \mathcal{E} [N_i(R_{ij},)N_{\ell}(R_{\ell j})^T] - \mu_i\mu_\ell^T$$

(3.23)

Under the Gaussian assumption, the density of $N$ in equation (3.14) involves the factor $\exp(-V^TC_N^{-1}V/2)$ where $V = D - G_k(\Theta_k) - \mu_N$ is the zero-mean model error.
The required matrix product is given by

\[
V^T C_N^{-1} V = \left( D - G_k(\hat{\Theta}_k) - \mu_N \right)^T C_N^{-1} \left( D - G_k(\hat{\Theta}_k) - \mu_N \right) + \\
\left( G_k(\hat{\Theta}_k) - G_k(\Theta_k) \right)^T C_N^{-1} \left( G_k(\hat{\Theta}_k) - G_k(\Theta_k) \right) + \\
2 \left( G_k(\hat{\Theta}_k) - G_k(\Theta_k) \right)^T C_N^{-1} \left( D - G_k(\hat{\Theta}_k) - \mu_N \right)
\]  

(3.24)
in which \( \hat{\Theta}_k \) is the maximum likelihood estimate (MLE) of \( \Theta_k \). This estimate satisfies

\[
\hat{\Theta}_k = \arg\min_{\Theta_k} (D - G_k(\Theta_k) - \mu_N)^T C_N^{-1} (D - G_k(\Theta_k) - \mu_N) 
\]  

(3.25)
and must be found via a numerical search.

Now make the additional assumptions that the mine signature shape matrix \( G_k \) is known (including the underlying mine positions \( R_{0,ik} \)) up to an unknown signature amplitude. Taking \( \theta_k \) to be this amplitude leads to \( G_k(R; \Theta_k) = G_k(R)\theta_k \) in which \( G_k(R) \) is independent of \( \theta_k \). In this case the MLE is well known [41, p. 186]

\[
\hat{\theta}_k = U (D - \mu_N) 
\]  

(3.26)
where

\[
U = (G_k^T C_N^{-1} G_k)^{-1} G_k^T C_N^{-1} 
\]  

(3.27)
The operator \( U \) has the useful property that \( UV = \hat{\theta}_k - \theta_k \), from which it is easy to show

\[
V^T C_N^{-1} V = \left( D - G_k\hat{\theta}_k - \mu_N \right)^T C_N^{-1} \left( D - G_k\hat{\theta}_k - \mu_N \right) + \\
\left( \hat{\theta}_k - \theta_k \right)^T C_{\hat{\theta}}^{-1} \left( \hat{\theta}_k - \theta_k \right) + \\
2 \left( \hat{\theta}_k - \theta_k \right)^T G_k^T C_N^{-1} \left( D - G_k\hat{\theta}_k - \mu_N \right)
\]  

(3.28)
where

\[
C_{\hat{\theta}}^{-1} = G_k^T C_N^{-1} G_k
\]  

(3.29)
If the features $\theta_k$ are also Gaussian distributed with density $\theta_k | H_k \sim \mathcal{N}(\mu_{\theta_k}, C_{\theta_k})$, we obtain the Bayesian linear model [41, p. 326], which leads to

$$f_{D|H_k}(D|H_k) = \frac{(2\pi)^{-MN_D/2}e^{-Q_k/2}}{\det(C_N)^{1/2}\det(C_{\theta_k})^{1/2}\det(C_{\theta_k}^{-1} + C_{\theta_k}^{-1})^{1/2}} \quad (3.30)$$

where

$$Q_k = (D - G_k\hat{\theta}_k - \mu_N)^T C_N^{-1} (D - G_k\hat{\theta}_k - \mu_N) +$$

$$+ (\hat{\theta}_k - \mu_{\theta_k})^T \left[ C_{\theta_k}^{-1} (C_{\theta_k}^{-1} + C_{\theta_k}^{-1})^{-1} C_{\theta_k}^{-1} \right] (\hat{\theta}_k - \mu_{\theta_k}) \quad (3.31)$$

Using the matrix identity $A(A + B)^{-1}B = (A^{-1} + B^{-1})^{-1}$ results in the compact expression

$$Q_k = (D - G_k\hat{\theta}_k - \mu_N)^T C_N^{-1} (D - G_k\hat{\theta}_k - \mu_N) +$$

$$+ (\hat{\theta}_k - \mu_{\theta_k})^T (C_{\theta_k} + C_{\theta_k}^{-1})^{-1} (\hat{\theta}_k - \mu_{\theta_k}) \quad (3.32)$$

the terms of which we interpret as clutter-related signature estimation errors and variability in the signature respectively. Using the approximation in equation (3.15) or the generalized likelihood ratio test formulation in equation (3.18) effectively ignores the second term in this expression.

### 3.5 Decision-Level Fusion Formulation

The formulation of an optimal decision-level fusion algorithm is similar to that defined in Section 3.4 for feature-level fusion. As before, each sensor $i$ collects data $\{d_{i}(R_{i1}), \ldots, d_{i}(R_{iM_i})\}$. On the basis of these data, it makes an intermediate declaration $u_i$ regarding the truth of intermediate sensor-specific hypotheses $h_{iq}(R)$ for $q = 1, 2, \ldots, Q_i$. As an example, for a sensor suite that comprises EMI, GPR and IR sensors we might use $Q_i = 2$ for all $i$ with $h_{11/12} = "metallic\ object\ present/not$
present,” $h_{21/22} =$ “dielectric discontinuity present/not present,” and $h_{31/32} =$ “thermal anomaly present/not present” rather than simply $H_{1/2} =$ “mine present/not present.” In what follows we restrict our attention to the case of equal numbers of intermediate hypotheses $Q_i \equiv Q$.

The goal of fusion is to combine these declarations into a fused declaration $u_0$ that addresses hypotheses $H_k$, $k = 1, 2, ..., K$. To simplify the notation, form the vector of individual declarations given by

$$u \equiv [u_1 \ u_2 \ \ldots \ u_{N_s}]^T$$

(3.33)

The declarations $u$ are assumed to be the only data on which the decision $u_0$ is based. In particular, it is assumed that the fusion processor has no knowledge of the sensor data $D$. It is further assumed that for each sensor $i$ the detection confidences $Pr(u_i|H_k)$ and the a priori probabilities $Pr(H_k)$ are known for all $k$. In addition, the detection $u_i$ involves a threshold $t_i$, which must be specified by the operator.

For optimal decision-level fusion we seek a rule for fusing $u$ into $u_0$ such that we minimize [36]

$$R = \sum_{j=1}^{K} \sum_{k=1}^{K} C_{jk} Pr(H_k) Pr(u_0 = j|H_k)$$

(3.34)

Write the class-conditioned density as

$$Pr(u_0 = j|H_k) = \sum_u Pr(u_0 = j|u) Pr(u|H_k; t_1, t_2, \ldots, t_{N_s})$$

(3.35)

which makes explicit the fusion rule $Pr(u_0 = j|u)$. In this result the summation is over all possible combinations of sensor outcomes, and we have used the fact that $u_0$ is based only on the local decisions $u_i$ and not on $H_k$. Rearranging the summations
and using a vector notation for the thresholds \( t_i \) leads to

\[
\mathcal{R} = \sum_{u} \sum_{j=1}^{K} \Pr(u_0 = j|u) \sum_{k=1}^{K} C_{jk} \Pr(u|H_k; t) \Pr(H_k)
\]  

(3.36)

The individual decisions \( u_i \) address the hypotheses \( h_{iq} \), which are related to \( \Pr(u|H_k; t) \) as follows:

\[
\Pr(u|H_k; t) = \sum_{q=1}^{Q} \Pr(u|h_q; t) \Pr(h_q|H_k)
\]  

(3.37)

where \( h_q = [h_{1q}, h_{2q}, \ldots, h_{N_q}] \) and we have used the fact that decision \( u_i \) depends on \( h_{iq} \) and not on \( H_k \). The minimum of \( \mathcal{R} \) is found by minimizing the following quantity for each \( u \)

\[
\sum_{j=1}^{K} \Pr(u_0 = j|u) \sum_{k=1}^{K} C_{jk} \Pr(u|H_k; t) \Pr(H_k)
\]  

(3.38)

It remains to specify the fusion rule \( \Pr(u_0 = j|u) \) and the decision thresholds \( t \). Some approaches are discussed in the following sections.

### 3.5.1 An Optimal Hard-Decision Fusion Rule

For sensors that produce "hard" decisions (i.e., all-or-nothing declarations) the fusion rule is a mapping from the \( N_s \)-fold product of the space \( \{1, 2, \ldots, Q\} \) to one of \( K \) output hypotheses. There exist \( K^Q \) such mappings, and from this possibly large set we must determine the most effective rule. In general, the most effective mapping will depend on the effectiveness of the individual sensors. Ad-hoc mappings (AND, OR, majority voting, etc.) have been used, some with considerable success.

Ideally, we like to simultaneously determine the thresholds at each of the sensors and at the fusion center, such that the Bayes risk of the system is minimized. Finding a global optimum to this problem is difficult except for modest numbers of sensors,
in which case a global search over all $t$ is possible. The equations corresponding to necessary but not (in general) sufficient conditions to achieve the global optimum have been derived using a sensor-by-sensor optimization methodology [36, §3.4], but even then, obtaining the desired thresholds $t$ is difficult, because it involves the solution of a set of coupled nonlinear equations. For the special case of identical sensors with identical thresholds, this solution reduces to a $k$-out-of-$N_S$ form. That is, the fused output $u_0 = 1$ if $k$ or more sensors have outputs $u_i = 1$. The cases $k = N_S$, $k = 1$ and $k = \lceil (N_S + 1)/2 \rceil$, where $\lceil \cdot \rceil$ is the standard ceiling function, correspond to AND, OR, and majority voting rules, respectively.

Consider next the case when all the sensors are not identical. Because obtaining the true globally optimal solution is difficult, we consider the simpler problem in which the local decision rules (i.e., the mappings from $D$ to the local decisions $u_i$) are fixed, and the sensors output hard decisions. The goal is to obtain the optimal fusion rule at the fusion center for the given set of local decisions [36, §3.3]. For $K = Q = 2$ the minimum in equation (3.38) reduces to a likelihood ratio test, given by

$$\frac{\Pr(u|H_1; t)}{\Pr(u|H_0; t)} \begin{cases} \geq & \eta^{u_0=1} \\ < & \eta^{u_0=0} \end{cases} \quad (3.39)$$

where $\eta$ is the threshold for the fused decision and is independent of $t$. If the declarations $u_i$ are conditionally independent for each sensor\(^1\) then we have

$$\frac{\Pr(u|H_1; t)}{\Pr(u|H_0; t)} = \prod_{i=1}^{N} \frac{\Pr(u_i|H_1; t_i)}{\Pr(u_i|H_0; t_i)} \quad (3.40)$$

If we define miss and false alarm probabilities for each sensor as follows:

$$P_{M_i}(t_i) \equiv \Pr(u_i = 0|H_1; t_i) \quad (3.41)$$

$$P_{F_i}(t_i) \equiv \Pr(u_i = 1|H_0; t_i) \quad (3.42)$$

\(^1\)Some physical phenomena (e.g., a small depression in the soil) can have an effect on several sensors and, hence, this commonly employed assumption can be difficult to justify.
then we can write the optimum test as

\[
\sum_{i=1}^{N_s} \left[u_i \log \left(\frac{1 - P_{M_i}(t_i)}{P_{F_i}(t_i)} \right) + (1 - u_i) \log \left(\frac{P_{M_i}(t_i)}{1 - P_{F_i}(t_i)} \right)\right] \geq \eta^{u_0=1} \quad \eta^{u_0=0}
\]  

(3.43)

If \( u_i = 1 \), then the second term vanishes, and the first term contributes a positive number, which is large if the \( i \)th sensor is reliable and small if it is unreliable. Similarly, if \( u_i = 0 \), then the first term vanishes, while the second term makes a negative contribution, whose magnitude is large if the \( i \)th sensor is reliable and small if it is unreliable. Therefore, the optimal decision fusion rule makes fused decisions that are dominated by the reliable sensors.

Since the Bayes risk can also be written in terms of the a posteriori probabilities \( \Pr(H_k|u) \), we can derive another form of the optimum test. A simple (but not necessarily optimal) method of performing fusion is to combine the declarations \( u_i \) such that the a posteriori probability \( \Pr(H_k|u) \) is maximized [35, §6.3]. We write the a posteriori probability as follows:

\[
\Pr(H_k|u) = \Pr(u_1|H_k; t_1) \ldots \Pr(u_{N_s}|H_k; t_{N_s}) \Pr(H_k)/\Pr(u; t)
\]  

(3.44)

where we have used the fact that the sensors reach their decisions independently. This result can be evaluated using known values for \( \Pr(u_i|H_i; t_i) \) (specified by \( P_{F_i}(t_i) \) and \( P_{M_i}(t_i) \)) using (3.37) and the expression for total probability.

Note that in this formulation \( \Pr(H_k|u) \) depends implicitly on the thresholds \( t \). Optimal fusion requires that we select \( t \) to produce the best performance. For modest numbers of sensors a search over all combinations of \( t \) is possible, but in general the search may be impractical. In such cases it is more attractive to select a value of \( t_i \) that produces "good" sensor performance, and to use that threshold for the calculations described above.
3.5.2 Soft Decision-Level Fusion

The hard decision-level fusion rules in equations (3.43) and (3.44) use global sensor reliability information in the form of \( \Pr(u_i|H_k; t_i) \) and the prior probabilities \( \Pr(H_k) \), but they do not utilize information about the confidence each sensor places in its individual decisions. Such information is quantified by \( \Pr(u_i|\Theta_i; t_i) \). When the sensors output their decisions \( u_i \) and this local confidence information, the outputs are equivalent to "soft" decisions, and we can perform fusion using an approach similar to that described above for feature-level fusion. Treating the \( u \) as features returned by each sensor, a discrete analog of equation (3.9) yields

\[
\Pr(H_k | \mathbf{D}) = \sum_{u_1=1}^{Q} \cdots \sum_{u_{N_s}=1}^{Q} \Pr(H_k | u) \cdot \Pr(u_1 | \Theta_1; t_1) \Pr(u_2 | \Theta_2; t_2) \cdots \Pr(u_{N_s} | \Theta_{N_s}; t_{N_s})
\]

(3.45)

where we have assumed that the fused hypothesis \( H_k \) depends only on the individual sensor decisions \( u_i \), and the \( u_i \) depend only on the features \( \Theta_i \) collected by sensor \( i \). The quantity \( \Pr(H_k|\mathbf{u}) \) is computed as in (3.44) for the hard decision case.

3.6 Summary

Multisensor fusion was described in Section 3.1. Alternative architectures for fusion were presented, and their merits and drawbacks were identified. Section 3.3 considered difficulties associated with the application of sensor fusion to mine detection. Previous work in this area was reviewed. A new feature-level fusion technique for noncoincidently sampled multisensor data was developed in Section 3.4. In Section 3.5, optimal decision-level fusion was described. A widely used hard decision-level fusion rule was extended to give a more effective soft decision-level fusion rule.
CHAPTER 4

FEATURE EXTRACTION

The importance of feature extraction in identity fusion and pattern recognition was mentioned in Chapter 3. This chapter briefly discusses the motivation for feature extraction and reviews feature extraction techniques. Feature selection, a related problem, is also discussed.

Measurement vectors acquired from sensors usually have high dimension. Classifier design for high dimensional data involves estimating a large number of free parameters, which requires a large training set. Because acquiring large, labelled training sets is often difficult, it becomes necessary to reduce the data dimension. Feature extraction is the process of reducing the data dimension, while retaining most of the original information content.

When real, physical attributes of the observed phenomena are extracted as features, they are commonly referred to as physical features. Features extracted by applying various mathematical transforms without consideration of their physical significance have been called nonphysical or mathematical features [42]. This categorization is not strict, however, and some mathematical transformations may result in features that have physical significance. Nonphysical features are usually further categorized into representation and classification (or discriminant) features.
Because a large number of feature extraction methods are available, an exhaustive investigation of them is difficult. During the processing of demining sensor data collected at Fort A. P. Hill (described in Chapter 7), a variety of feature extraction methods were investigated. This chapter describes several different classes of features and presents details of some feature extraction techniques. Simple, hypothetical examples are used to illustrate the capabilities and limitations of some of these techniques.

4.1 Physical Features

As stated above, physical features are quantities that represent physical attributes of the measured object. Whether a physical feature is useful is determined by the type and resolution of the sensor, the properties of the target, and the intended application. The physical features considered here include amplitude statistics and physical model parameters.

4.1.1 Amplitude Statistics

Certain statistical descriptors can be used to concisely summarize the underlying amplitude distribution. Features such as peak value, mean, standard deviation, skewness and kurtosis are widely used [4, 43, 44]. These well known quantities are as follows for a data vector $x = [x_1 x_2 ... x_n]^T$.

$$\text{Peak}(x) = \max_i (x_i) \quad (4.1)$$

$$\text{Mean}(x) = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (4.2)$$
Standard deviation(\(x\)) = \(\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}\) \hspace{1cm} (4.3)

\text{Variance}(x) = \sigma^2 \hspace{1cm} (4.4)

\text{Skewness}(x) = \gamma_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^3}{\sigma^3} \hspace{1cm} (4.5)

\text{Kurtosis}(x) = \beta_2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^4}{\sigma^4} \hspace{1cm} (4.6)

Amplitude statistics were considered for our GPR and EMI sensors. The statistical descriptors defined above were extracted from rectangular subimages of GPR image data, which were chosen to encompass the target signatures in the GPR waterfall plots. Amplitude statistics were also extracted from EMI spatial signatures.

4.1.2 Physical Model Parameters

If the response of a sensor can be accurately modeled by considering the physics of the sensor, target, and environment, then the parameters of such a model may be used as features. A common method of extracting these features involves a least-squared error fitting of measured data to the model. Drawbacks of the physical modeling approach are that (1) a different model needs to be developed for each sensor, and (2) developing sufficiently sophisticated models to account for the observed behavior may be difficult and time consuming. In some situations it may be more convenient to obtain the models empirically. Even when a suitable model is available, accurately estimating the parameters based on noisy measurements can be difficult, especially when the number of parameters is large.

A simple empirical model for IR data was developed in [27] and used to extract parameters from IR images. A similar model was used for EMI measurements [23].
These physical models and the corresponding feature extraction are discussed in Chapter 6.

4.2 Nonphysical Features

Nonphysical feature extraction involves mathematical transformations of the data to achieve various optimality criteria (often related to data compression or class separability). Because these feature extraction techniques are not concerned with the physical significance of the extracted features, they are not tied to any particular sensor. Two different types of nonphysical features exist, namely, (1) representation features, and (2) classification or discriminant features.

Representation features attempt to represent the raw measurement data as accurately as possible using a small number of parameters. This approach is most meaningful when the goal of feature extraction is simply data compression. Techniques for extracting classification features attempt to retain most of the discriminant power of the original data in a small number of parameters. Several representation and classification features were applied to the sensor data and are described next.

4.2.1 Representation Features

Extraction of representation features has received much attention in communication, data storage and image compression. In these applications the goal is to represent a high-dimensional signal or an image using a small number of features, from which the original signal or the image can be recovered as accurately as possible. Discrete Fourier transforms (DFT), discrete cosine transforms (DCT), discrete Karhunen-Loève (KL) transforms, and wavelet transforms are some of the methods used for this purpose and are described below.
Discrete Fourier Transform (DFT)

A one-dimensional discrete signal waveform \( x(n) \) can be represented using the DFT defined by

\[
X(k) = \sum_{n=0}^{N-1} \exp(-j2\pi kn/N) \tag{4.7}
\]

The corresponding inverse transformation is

\[
x(n) = \frac{1}{N} \sum_{k=0}^{N-1} \exp(j2\pi kn/N) \tag{4.8}
\]

Similarly a two-dimensional image \( f(x, y) \) can be represented using the two-dimensional DFT defined as

\[
F(n, m) = \frac{1}{NM} \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} f(x, y) \exp[-j2\pi(nx/N + my/M)] \tag{4.9}
\]

and the inverse transform is

\[
f(x, y) = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} F(n, m) \exp[-j2\pi(nx/N + my/M)] \tag{4.10}
\]

A high-dimensional signal waveform or an image can be compactly represented by using a small number of DFT terms in the summation, with a possible loss of some information. One disadvantage of using the DFT for data representation is the distortion in the coefficients due to edge effects. This occurs in an FFT implementation because of zero padding or implied periodicity. For the DFT the distinction between physical and nonphysical features becomes unclear, because the resultant features (the spectra of the data) are often meaningful physical attributes. Although we extracted DFT spectra from our data, they did not lead to a useful dimensionality reduction or an improvement in classification and, hence, they were not pursued.
Discrete Karhunen-Loéve (KL) Transform

The discrete KL-transform represents an arbitrary n-dimensional random vector, \( \mathbf{X} \), in terms of a set of basis vectors \( \mathbf{\phi}_i \), which are the eigenvectors of the covariance matrix \( \Sigma_X \) [45]. If \( \mathbf{Y} = [y_1 \ldots y_n]^T \) are the coefficients corresponding to the basis vectors \( \mathbf{\Phi} = [\mathbf{\phi}_1 \ldots \mathbf{\phi}_n] \), then the random vector \( \mathbf{X} \) can be expressed as

\[
\mathbf{X} = \sum_{i=1}^{n} y_i \mathbf{\phi}_i = \mathbf{\Phi} \cdot \mathbf{Y}
\]  

(4.11)

If the \( \mathbf{\phi}_i \) vectors are orthonormal, the \( i \)th coefficient \( y_i \) can be computed by projecting \( \mathbf{X} \) onto the basis vector \( \mathbf{\phi}_i \), i.e.,

\[
y_i = \mathbf{\phi}_i^T \mathbf{X} \quad (i = 1, \ldots, n)
\]  

(4.12)

If \( m < n \) terms are summed, \( \mathbf{X} \) is approximated by

\[
\mathbf{X} \approx \sum_{i=1}^{m} y_i \mathbf{\phi}_i
\]  

(4.13)

The mean square error of approximation (4.13) is minimum if the \( \mathbf{\phi}_i \)'s satisfy the eigenequation [45]

\[
\Sigma_X \mathbf{\phi}_i = \lambda_i \mathbf{\phi}_i
\]  

(4.14)

In representing an \( n \)-dimensional random vector \( \mathbf{X} \) with \( m \) (< \( n \)) features, the \( \mathbf{\phi}_i \)'s corresponding to the largest \( m \) eigenvalues of \( \Sigma_X \) are retained in the transformation matrix to minimize the mean square error in the representation. Another useful property of the KL-transform is that the transformed feature vectors are de-correlated (i.e., orthogonal). The need for computing the eigenvalues and eigenvectors of the covariance matrix is a major drawback of the KL-transform in practice, especially when the signal dimension is large and only a few training samples are available. In our data processing, discrete KL features were extracted from the EMI spatial data.
Discrete Cosine Transform (DCT)

The discrete cosine transform is widely used for image compression [43]. The DCT represents an arbitrary signal of interest \( x(n) \) using a weighted sum of cosine signals as

\[
y(k) = w(k) \sum_{n=1}^{N} x(n) \cos \left( \pi k \frac{n + 1/2}{N} \right) \tag{4.15}
\]

where

\[
w(k) = \begin{cases} 
\sqrt{\frac{1}{N}} & \text{for } k = 0 \\
\sqrt{\frac{2}{N}} & \text{otherwise}
\end{cases} \tag{4.16}
\]

The corresponding inverse transform is defined by

\[
x(n) = \sum_{k=0}^{N} w(k)y(k) \cos \left( \pi k \frac{n + 1/2}{N} \right) \tag{4.17}
\]

For images the DCT is defined as

\[
C(n, m) = k_1(n)k_2(m) \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} f(x, y) \cos \left( \pi n \frac{x + 1/2}{N} \right) \cos \left( \pi m \frac{y + 1/2}{M} \right) \tag{4.18}
\]

where \( n = 0, 1, \ldots, N - 1 \) and \( m = 0, 1, \ldots, M - 1 \). In equation (4.18)

\[
k_1(n) = \begin{cases} 
\sqrt{\frac{1}{N}} & \text{for } n = 0 \\
\sqrt{\frac{2}{N}} & \text{otherwise}
\end{cases} \tag{4.19}
\]

and

\[
k_2(n) = \begin{cases} 
\sqrt{\frac{1}{M}} & \text{for } m = 0 \\
\sqrt{\frac{2}{M}} & \text{otherwise}
\end{cases} \tag{4.20}
\]

The inverse DCT transform is defined as

\[
f(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} C(n, m)k_1(n)k_2(m) \cos \left( \pi n \frac{x + 1/2}{N} \right) \cos \left( \pi m \frac{y + 1/2}{M} \right) \tag{4.21}
\]

\(^2\text{DCT is the data representation used in the JPEG image compression standard}\)
Usually, most of the energy is concentrated in a few DCT coefficients, which are used as features.

The DCT has several advantages over the DFT and KL-transform. Unlike the DFT, the DCT has a real kernel and, therefore, produces only real coefficients. The DCT has nearly the same data compression capability as the KL-transform, but without the computational complexity of the latter.

We used the DCT to compactly represent GPR subimages of target signatures. DCT features were also extracted from the EMI spatial response data.

**Wavelet Transform**

The wavelet transform is a powerful tool for analyzing signals with time varying frequency content. The Fourier transform determines the global frequency content of a given time-domain waveform, but it does not describe how frequency content evolves with time. The short-time Fourier transform, which computes the Fourier transform over a sliding window of fixed width, achieves this capability to a degree. Because the window size is fixed in the short-time Fourier transform, the resulting frequency resolution is poor for parts of the signal dominated by low frequencies and the time resolution is poor for parts of the signal dominated by high frequencies [46]. The wavelet transform overcomes this deficiency to some extent by analyzing the signals at different scales. The wavelet decomposition splits the original signal into two orthogonal components, an approximation at coarser scale and the omitted details. The approximation can be recursively decomposed further into another approximation and detail, and so on. In a conventional dyadic wavelet transform the number of
coefficients in the approximation component is halved in each step. After several de­
composition steps the approximation will contain only a small number of coefficients,
which can be used as representation features.

Wavelet packet decomposition, a method in which both the detail components
and the approximations are successively decomposed, provides an even more powerful
analysis tool. In our data processing we extracted wavelet packet coefficients from
metal detector spatial signatures and from time-domain GPR data collected directly
over each target. Wavelet packet decomposition was done using the Matlab Wavelet
Toolbox [46].

4.2.2 Classification Features

Although the features discussed in the previous section are useful for signal rep­
resentation, those features are not necessarily effective for pattern recognition or
discrimination, as demonstrated by the following hypothetical example.

Consider the two normally distributed data sets shown in Figure 4.1(a), in which
the eigenvector direction $\phi_1$ corresponds to the larger eigenvalue and $\phi_2$ corresponds
to the smaller. Using a KL-transform approach, $\phi_1$ is the best representation feature,
but $\phi_1$ is a poor feature for discrimination, because projecting the data onto $\phi_1$
results in a large overlap of the two classes as illustrated in Figure 4.1(b). The
eigenvector direction $\phi_2$ corresponding to the smaller eigenvalue, which is a poor
representation feature, leads to good discrimination as evidenced by the small overlap
in Figure 4.1(c).
Figure 4.1: Hypothetical example showing the lack of discriminatory power in representation features.

(a) Class distributions and the eigenvector directions

(b) Data projected on to $\phi_1$

(c) Data projected on to $\phi_2$
To obtain better class separation we applied several classification feature transformations found in the pattern recognition literature. These transforms are described next.

While orthonormal transformations such as the KL-transform preserve the structure of the data, any nonsingular (invertible) transformation preserves class separability [45]. Singular transformations (e.g., projections) can be used to reduce the data dimension, while losing some discriminatory power. Many feature extraction methods used for classification attempt to find singular transformations such that most of the discriminating power in the original, high-dimensional observational space is retained in the resulting lower-dimensional feature space. These transformations are derived by optimizing various class separability criteria.

The ideal criterion for measuring class separability is the Bayes error. Because a classifier designed using the a posteriori probability functions as features can achieve the performance of the optimal Bayes classifier, these probabilities are ideal classification features. Because the sum of the a posteriori probabilities is equal to unity, these features are linearly dependent and, therefore, for an $L$ class problem, only $L - 1$ a posteriori probability features are necessary for classification. Therefore, for a two class problem such as discriminating mines from clutter, one feature is sufficient. Classification using a posteriori probabilities as features amounts to the trivial task of thresholding and, hence, one finds it difficult to distinguish these features from classifiers.

Because a posteriori probability functions are not easy to estimate in practice, we must consider other suboptimal features, and the number of features required may be more than the minimum cited above. Furthermore, because it is difficult to accurately
estimate the Bayes error in practice, other means are required to quantify the class separability.

Two types of class separability measures are commonly used. The first type involves the within-class \( S_c \), between-class \( S_b \), and mixture \( S_m \) scatter matrices, and the second type is based on upper bounds of the Bayes error. The scatter-matrix based techniques are more general in that they can be applied to classification problems involving more than two classes, but they do not directly relate to the Bayes error. Methods that use criteria based on upper bounds of the Bayes error, such as the Bhattacharyya distance, are limited to two-class problems.

4.2.3 Methods Based on Scatter Matrices

We start by defining the scatter matrices. The within-class scatter matrix \( S_w \) is a measure of how tightly the data are scattered around the mean vectors of each class. It is defined by

\[
S_w = \sum_{i=1}^{L} P_i E \{(X - M_i)(X - M_i)^T | \omega_i \} = \sum_{i=1}^{L} P_i \Sigma_i
\]  

(4.22)

where \( M_i \), \( \Sigma_i \) and \( P_i \) are respectively the mean vector, covariance matrix and the a priori probability of the \( i \)th class (represented by \( \omega_i \)), and \( L \) is the number of classes.

The between-class scatter matrix \( S_b \) indicates how well the individual class means are separated, and is defined by

\[
S_b = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T
\]

(4.23)

where \( M_0 = E \{X\} = \sum_{i=1}^{L} P_i M_i \) is the mean of all data in the mixture. Finally, the mixture-scatter matrix \( S_m \) is defined as

\[
S_m = E \{(X - M_0)(X - M_0)^T \} = S_w + S_b.
\]

(4.24)
It is clear that class separability increases with increasing $S_b$ and decreasing $S_w$. Therefore, class separability criteria such as

$$J = tr(S_w^{-1}S_b)$$  \hspace{1cm} (4.25)

can be defined [45].

Feature extraction in the scatter matrix method is done in a way somewhat similar to the KL-transform. Consider a $n \times m$ matrix $A$ consisting of $m$ linearly independent column vectors. The matrix $A$ is used to perform a linear transformation of an $n$-dimensional $X$ to an $m$-dimensional $Y$ ($m < n$) using

$$Y = A^T X$$ \hspace{1cm} (4.26)

Now $S_w$ and $S_b$ in the $Y$-space can be calculated from $S_w$ and $S_b$ in the $X$-space as

$$S_{wY} = A^T S_{wX} A$$ \hspace{1cm} (4.27)

$$S_{bY} = A^T S_{bX} A$$ \hspace{1cm} (4.28)

Feature extraction in this method involves finding the matrix $A$ that optimizes the selected separability criterion in the transformed space. For a two-class problem using $J_1 = tr(S_w^{-1}S_b)$, it can be shown that only one feature is needed for classification. The feature in this case is given by

$$y_1 = c(M_2 - M_1)^T S_w^{-1} X$$ \hspace{1cm} (4.29)

where $c$ is a constant. This feature extraction process can be extended to $L$-class problems resulting in $(L - 1)$ classification features. The quality of the extracted features is determined by how well the chosen separability criterion measures the actual class separability [45].
Fisher Discriminant Method

The Fisher discriminant method [47] is one of the earliest and most widely used classification features extraction methods. It is also the simplest example of (4.29). Given two data sets with mean vectors \( \mathbf{M}_1, \mathbf{M}_2 \) and covariance matrices \( \Sigma_1, \Sigma_2 \), respectively, this method finds the Fisher direction \( \mathbf{d} \) given by

\[
\mathbf{d} = \alpha \mathbf{S}_w^{-1} (\mathbf{M}_1 - \mathbf{M}_2) \tag{4.30}
\]

where \( \alpha \) is a normalizing constant to make \( \mathbf{d} \) a unit vector. It can be shown that this process maximizes the Fisher criterion defined by

\[
f = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{4.31}
\]

where \( m_1, m_2 \) and \( \sigma_1, \sigma_2 \) are respectively the mean and standard deviation of the two classes when projected onto \( \mathbf{d} \). The Fisher feature value \( y \) of an arbitrary feature vector \( \mathbf{x} \), is found from (4.30), which we write as

\[
y = \langle \mathbf{d}, \mathbf{x} \rangle = \mathbf{d}^T \mathbf{x} \tag{4.32}
\]

Figure 4.2(a) shows the same two data clusters that were considered earlier to demonstrate the lack of discriminatory power in the KL transform. The direction of the Fisher vector computed using (4.30) is also shown on the figure. When the original two-dimensional data is projected onto the Fisher direction, the two reasonably separated feature sets in Figure 4.2(b) are obtained. Because the Fisher criterion is based primarily on a difference in means, it is susceptible to problems such as multimodal data.
(a) Class distributions and the Fisher discriminant directions

(b) Data projected on to the Fisher discriminant direction

Figure 4.2: Feature extraction using the Fisher discriminant criterion.
Foley-Sammon (FS) Method

The Fisher method works by projecting the data onto a single vector or a line. Sammon [48] proposed to find two orthogonal optimum discriminant vectors $d_1$ and $d_2$ that form an optimum discriminant plane for the two-class problem. Here, the first vector $d_1$ is found by solving for the unit vector $d$ that maximizes the discriminant ratio $R$ defined as

$$ R = \frac{d^T S_b d}{d^T S_w d} \tag{4.33} $$

This procedure leads to a generalized eigenvector problem

$$ S_b d = \lambda S_w d \tag{4.34} $$

where only one nonzero eigenvalue exists and the corresponding eigenvector $d_1$ that maximizes equation (4.33) is simply the Fisher vector obtained in equation (4.30) above. The second vector $d_2$ is found by maximizing $R$ with the constraint that $d_2$ must be orthogonal to $d_1$. This is done using the Lagrange multiplier method to maximize

$$ \frac{d_1^T S_b d_2}{d_1^T S_w d_2} - \lambda [d_2^T d_1] \tag{4.35} $$

which leads to

$$ d_2 = \gamma \left[ S_w^{-1} - \frac{\Delta^T [S_w^{-1}]^2 \Delta}{\Delta^T [S_w^{-1}]^3 \Delta} [S_w^{-1}]^2 \right] \Delta \tag{4.36} $$

where $\gamma$ is a normalizing constant to make $d$ a unit vector and $\Delta = M_1 - M_2$.

Once the discriminant vectors $d_1$ and $d_2$ are found, the original multidimensional data are mapped onto the two-dimensional plane spanned by the discriminant vectors. This mapping provides the ability to design two-dimensional classifiers such
as quadratic or piecewise linear classifiers, which are more powerful than the Fisher linear classifier. Mapping onto the two-dimensional optimal plane also provides some ability to visualize the structure and the class separation of the original multidimensional data. Foley and Sammon [49] extended the above idea to an arbitrary number of features. Basically, rather than obtaining just one vector $d_2$ orthogonal to the Fisher vector, $n$ discriminant vectors are found such that $d_i^T d_n = 0$ for $i \neq n$.

Because the Foley-Sammon method is inherently based on the difference of class means (like the Fisher method) it fails when both classes have the same or nearly equal mean vectors. Techniques have been proposed recently [50, 51], which claim to have performance equal to or better than the Foley-Sammon method for multivariate normal data.

**Extended Fisher Criterion (EFC) Method**

Malina [50] proposed an extension to the Fisher criterion, in which the measure of discrimination between two classes is defined as

$$F(d) = \frac{d^T S_b d}{d^T S_w d} + \frac{|d^T \Sigma^{(-)} d|}{d^T S_w d} \tag{4.37}$$

where $\Sigma^{(-)}$ is alternately taken to be $\Sigma_1 - \Sigma_2$ or $\Sigma_2 - \Sigma_1$. The first term of $F(d)$ is the usual Fisher distance, while the second term measures the ratio of the absolute value of the difference of class variances to the weighted sum of the variances along $d$. This extended Fisher distance measure is nonnegative and becomes equal to zero only if the class means and covariances are identical for the two classes. The vector $d$ for which $F(d)$ is maximized is the best discriminant direction in the sense of the EFC.
To compute the best linear discriminant direction, the following two eigenvalue equations are solved.

\[
S_w^{-1}(S_b + \Sigma_{12}^{(-)})d_1 = \lambda_1 d_1, \quad \Sigma_{12}^{(-)} = \Sigma_1 - \Sigma_2
\]  \hspace{1cm} (4.38)

\[
S_w^{-1}(S_b + \Sigma_{21}^{(-)})d'_1 = \lambda'_1 d'_1, \quad \Sigma_{12}^{(-)} = \Sigma_2 - \Sigma_1
\]  \hspace{1cm} (4.39)

The equation corresponding to the larger of \( \lambda_1 \) or \( \lambda'_1 \) is chosen, and the eigenvector \( d_1 \) is computed from it.

**Fukunaga-Koontz (FK) Method**

Realizing that the features extracted with the discrete KL-transform are optimal for representation, but not necessarily for pattern classification, Fukunaga and Koontz [52] proposed a modification that provides more effective discriminating features for two-class pattern recognition problems.

In their method a normalizing transform is first applied to the data such that the joint weighted correlation matrix \( S_1 + S_2 \) is whitened. Here the weighted correlation matrices of each class is defined by

\[
S_i = P_i(\Sigma_i + M_iM_i^T), \quad i = 1, 2
\]  \hspace{1cm} (4.40)

where \( P_i \) is the prior probability of the \( i \)th class, \( \Sigma_i \) is its covariance matrix and \( M_i \) is its mean vector. Let \( A \) be a diagonalizing transform such that

\[
A^T(S_1 + S_2)A = I
\]  \hspace{1cm} (4.41)

It follows that the columns of \( A \) are eigenvectors for both \( S_1 \) and \( S_2 \), and that each eigenvalue lies between 0 and 1, specifically

\[
1 \geq \lambda_1^{(1)} \geq \lambda_2^{(1)} \geq \ldots \geq \lambda_N^{(1)} \geq 0
\]  \hspace{1cm} (4.42)
where $\lambda_j^{(i)}$, $(i = 1, 2$ and $j = 1, 2, \ldots, N)$ is the $j$th eigenvalue of the $i$th class. Further, the sum of eigenvalues of the two classes for each eigenvector is equal to unity, i.e.,

$$\lambda_n^{(1)} = 1 - \lambda_n^{(2)}, \quad n = 1, 2, \ldots, N$$

It is clear that eigenvectors most effective in representing class 1 are least effective in representing class 2 and vice versa, and that eigenvectors corresponding to the largest difference of eigenvalues have the greatest discriminating power. Fukunaga and Koontz recommended that the eigenvectors be ranked according to $|\lambda_j^{(1)} - 0.5|$. Foley and Sammon [49] show a counterexample where this ranking fails to find the best discriminant direction, but if both classes have equal mean vectors and different covariance matrices, then the Fukunaga-Koontz transform provides the best discriminant direction [53].

To illustrate this technique consider the two normally distributed clusters of data shown in Figure 4.3(a). Because these two clusters have nearly identical means, the Fisher discriminant fails to separate them well. The result of the FK normalizing transform is shown in Figure 4.3(b). The optimal discriminant direction based on the FK ranking method is also shown. When the normalized data is transformed in the optimal discriminant direction, the resulting features are distributed as in Figure 4.3(c), from which classes can be separated using two thresholds.

**Fisher with Fukunaga-Koontz Method**

It was pointed out earlier that the Fisher method fails when the mean vectors are equal, but the FK-transform gives the best discriminant direction under those
Figure 4.3: Discriminating two classes with equal means using the Fukunaga-Koontz (FK) method.
conditions. Longstaff [51] used these two ideas to develop a new optimal discriminant plane, which is superior to the optimal plane defined by Sammon for multivariate normal data. To describe how this new optimal plane is obtained, let us first look at the expression for the Fisher vector written with the individual covariances shown explicitly as

\[ d = (S_1 + S_2)^{-1}(M_1 - M_2) \]  

(4.45)

where the normalizing constant has been dropped for convenience. Let \( A \) be the normalizing transform defined in equation (4.41). Now if the original data is standardized by transforming them using \( A \), then the Fisher vector in the standardized coordinate system becomes

\[ d' = A^{-1}d \]

\[ = A^{-1}(S_1 + S_2)^{-1}(A^T)^{-1}A^T(M_1 - M_2) \]

\[ = (M_1' - M_2') \]  

(4.46)

where \( M_j' = (A^T)^{-1}M_j \), \( j = 1, 2 \) are the mean vectors in the new coordinate system. Therefore, after this standardization, the Fisher vector is parallel to the axis intercepting the means of the two class distributions. If the data are now projected onto the subspace normal to \( d' \), the two data sets will have coincident means in that subspace. Any further discriminating information in the projected data must be in the covariance difference. Recall that in Sammon's optimal plane concept the Fisher vector of the raw data is located first and then another Fisher vector is found in the subspace normal to the first. If we apply Sammon's method to the standardized data, it is obvious that the second Fisher vector will be equal to zero, because the class means are coincident in the normal subspace. In this subspace Longstaff proposed to
use the FK-method, which gives the optimal direction for the equal-mean case. The performance of this combination of Fisher and FK-methods equals or exceeds that of either method individually.

4.2.4 Methods Based on Upper Bounds of Bayes Error

Because it is difficult to accurately estimate the Bayes error in practice, some feature extraction techniques attempt to estimate upper bounds on it. One such measure used to quantify the class separability in two-class problems is the Bhattacharyya distance defined by

\[ \mu = \mu_1 + \mu_2 \]  

(4.47)

where

\[ \mu_1 = \frac{1}{8} (M_2 - M_1)^T \left[ \frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (M_2 - M_1) \]  

(4.48)

and

\[ \mu_2 = \frac{1}{2} \ln \frac{\frac{1}{2} |\Sigma_1 + \Sigma_2|}{\sqrt{||\Sigma_1|| ||\Sigma_2||}} \]  

(4.49)

The first term \( \mu_1 \) corresponds to the mean difference of the two classes, while the second term \( \mu_2 \) corresponds to the covariance difference (Note that \( \mu_2 \) vanishes if \( \Sigma_1 = \Sigma_2 \)). If the classes are normally distributed, an upper bound of the Bayes error is given by

\[ \epsilon_u = \sqrt{P_1 P_2 e^{-\mu(1/2)}} \]  

(4.50)

Although the relationship of the Bhattacharyya distance criterion to the Bayes error is attractive, finding a transform that optimizes it is difficult. Only suboptimal solutions are available for the general case, and optimal solutions are available only for
special cases. Fukunaga [45] derives the optimal features for two special cases. First, when $\Sigma_1 = \Sigma_2$, maximizing $\mu_1$ leads to the Fisher discriminant described above. Second, when the mean vectors are identical, only $\mu_2$ needs to be optimized. It can be shown [45] that one needs to compute the eigenvalues $\lambda_i$ of $\Sigma_2^{-1}\Sigma_1$ and to select the eigenvectors corresponding to the largest $(\lambda_i + 1/\lambda_i + 2)$ terms. The application of this feature extraction method to the two class hypothetical data with identical means is shown in Figure 4.4. The feature distribution is similar to that obtained with the FK method.

4.3 Feature Selection

In Sections 4.1 and 4.2 we described several alternatives for extracting features from data. Although in principle two-class pattern recognition can be done using only one optimal feature, optimal features are not available in real applications, and one is led to use a variety of features in the hope that they will provide complementary information. This problem is especially common when dealing with a multisensor data set. One expects the Bayes error to monotonically decrease as more and more features are used for classification, but when a finite number of training samples are available to design a classifier, the classification error initially reduces but then increases as more and more features are added. This nonmonotonic behavior, known as the Hughes phenomena [54], has a significant implication for multisensor fused mine detection. Although one intuitively expects the discriminatory power to improve as more sensors relying on different physical characteristics are fused, the Hughes phenomena may even lead to worse performance than that of the best individual sensor, if too many ineffective features are included.
Figure 4.4: Discriminating two classes with equal mean using the feature extraction based on Bhattacharyya $\mu_2$ optimization.
The process of finding the most effective subset of features from the complete collection of available features is referred to as feature selection. The feature extraction techniques described in Sections 4.1 and 4.2 can be used for this purpose if we replace the data vector in those sections with our feature vector. The only way to guarantee that the optimal feature subset is selected is to do an exhaustive search, but this search is not practical when the total number of features is large. A branch and bound method of feature selection [45], which can find the optimal subset without an exhaustive search, is available, but it works only if the features satisfy the monotonicity property. Two commonly used feature selection methods are forward selection, which successively adds the most effective individual feature, and backward selection, which successively eliminates the least effective feature. The collection of the best individual features, however, is not guaranteed to be the most effective feature subset and, hence, the forward and the backward selection methods are suboptimal.

4.4 Summary

Feature extraction is the process of reducing the dimension of measured data such that information in the observation space relevant to classification is retained in the lower-dimensional feature space. Section 4.1 described physical features, which are measurable attributes of the physical phenomena. Section 4.2 discussed nonphysical features, which are derived by applying certain mathematical transformations. Those features do not necessarily correspond to any particular physical attributes. Nonphysical feature extraction techniques for data representation and classification were described. The importance of the feature subset selection was considered in Section 4.3.
CHAPTER 5

PATTERN CLASSIFICATION

Pattern classification is the task of assigning observed patterns to categories or classes according to some chosen criterion. If the a posteriori and a priori statistical distributions of the pattern classes are known, it is possible to implement the optimal Bayes classifier. For mine detection sensors, however, these distributions are unknown and difficult to estimate. We have implemented a variety of classifiers and evaluated their performance. This chapter reviews the classifiers that were investigated. Section 5.1 describes the optimal Bayes classifier. Section 5.2 discusses how the Bayes classifier can be reduced to simple parametric and distance classifiers for normally distributed data under certain conditions. Nonparametric classifiers are discussed in Section 5.3. Section 5.4 considers methods for classifier performance evaluation.

Essentially all of the material in this chapter is well known and available from other sources. It is provided here for completeness and to introduce our notation.

5.1 Statistical Decision Theory

Suppose we are given an observed pattern vector \( x \) to be assigned to one of \( K \) possible classes \( \{\omega_1, \omega_2, \ldots, \omega_K\} \). Let \( \rho_{x|\omega_i}(x|\omega_i) \) be the class-conditioned density of the \( i \)th class where \( i = 1, 2, \ldots, K \). Let us assume that the prior probability of the
ith class is given by $P_i$. The Bayesian approach to pattern recognition assigns the observed vector $\mathbf{x}$ to the class $\omega_j$ that minimizes the conditional average risk $R_j(\mathbf{x})$ defined by

$$R_j(\mathbf{x}) = \sum_{i=1}^{K} C_{ji} \Pr(\omega_i|\mathbf{x})$$  \hspace{1cm} (5.1)$$

where $C_{ji}$ is the cost of choosing class $j$ when class $i$ is true. Therefore, the class assignment based on minimizing the Bayes risk can be written as

$$k = \arg \min_{j} R_j(\mathbf{x}) = \arg \min_{j} \sum_{i=1}^{K} C_{ji} \Pr(\omega_i|\mathbf{x})$$  \hspace{1cm} (5.2)$$

where $j$ can take the values $\{1, 2, \ldots, K\}$. Using Bayes rule

$$\Pr(\omega_i|\mathbf{x}) = \frac{P_i \rho_{\mathbf{x}|\omega_i}(\mathbf{x}|\omega_i)}{\rho(\mathbf{x})}$$  \hspace{1cm} (5.3)$$

we can rewrite (5.2) as

$$k = \arg \min_{j} \frac{1}{\rho(\mathbf{x})} \sum_{i=1}^{K} C_{ji} P_i \rho_{\mathbf{x}|\omega_i}(\mathbf{x}|\omega_i)$$  \hspace{1cm} (5.4)$$

Because $\rho(\mathbf{x})$ does not depend on the class chosen, the decision rule can be rewritten as

$$k = \arg \min_{j} \sum_{i=1}^{K} C_{ji} P_i \rho_{\mathbf{x}|\omega_i}(\mathbf{x}|\omega_i)$$  \hspace{1cm} (5.5)$$

For the two-class case, i.e., $K = 2$, this decision rule reduces to

$$C_{11} P_1 \rho_{\mathbf{x}|\omega_1}(\mathbf{x}|\omega_1) + C_{12} P_2 \rho_{\mathbf{x}|\omega_2}(\mathbf{x}|\omega_2) \quad \overset{\omega_1}{\underset{\omega_2}{\lor}} \quad C_{21} P_1 \rho_{\mathbf{x}|\omega_1}(\mathbf{x}|\omega_1) + C_{22} P_2 \rho_{\mathbf{x}|\omega_2}(\mathbf{x}|\omega_2),$$  \hspace{1cm} (5.6)$$

which can be manipulated further to yield the Bayes likelihood ratio test (LRT)

$$\Lambda(\mathbf{x}) = \frac{\rho(\mathbf{x}|\omega_2)}{\rho(\mathbf{x}|\omega_1)} \quad \overset{\omega_2}{\underset{\omega_1}{\lor}} \quad \frac{P_1 (C_{21} - C_{11})}{P_2 (C_{12} - C_{22})} \triangleq \lambda$$  \hspace{1cm} (5.7)$$
If the cost functions are chosen as

\[ C_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \]  

(5.8)

then the Bayes risk becomes equal to the probability of classification error. Therefore, the Bayes classifier or the LRT minimizes the probability of error in classification, and no classifier can exhibit a lower error.

By using the costs of (5.8), (5.2) can be reduced to

\[ k = \arg \min_j \sum_{i=1}^{K} \Pr(\omega_i|x) \]  

(5.9)

\[ = \arg \min_j \left[ 1 - \sum_{i=1}^{K} \Pr(\omega_i|x) \right] \]  

(5.10)

\[ = \arg \max_j \Pr(\omega_i|x) \]  

(5.11)

Therefore, the test that minimizes the probability of classification error for the \( K \)-class problem, is to choose the class having the largest a posteriori probability, i.e., the maximum a posteriori (MAP) classifier. This \( K \)-class problem can also be put in the form of \( K \) likelihood ratio tests [55].

Because the data sets acquired in our experimental work were small and, hence, impractical for multiclass pattern recognition, the subsequent data analyses were limited to the two-class problem (\( K = 2 \)) involving mine and clutter classes.

5.2 Parametric Classifiers

The LRT is optimal, but the requisite class-conditioned probability density functions are seldom available. Even when the densities are known, they have complicated forms. It is sometimes convenient to use parametric classifiers designed based on
simplifying assumptions about the densities. Parametric classifiers based on normal density assumptions are popular and, used even when the actual densities are mildly non-normal.

5.2.1 Quadratic Classifier

Let us assume that each class is normally distributed with mean vectors $M_i$ and covariance matrices $\Sigma_i$, $i = 1, 2$. The likelihood $\rho(x|\omega_i)$ of class $\omega_i$ is given by

$$
\rho(x|\omega_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - M_i)^T \Sigma_i^{-1} (x - M_i) \right\}
$$

The LRT of (5.7) becomes

$$
\Lambda(x) = \sqrt{\frac{\Sigma_1}{\Sigma_2}} \exp \left[ -\frac{1}{2} (x - M_2)^T \Sigma_2^{-1} (x - M_2) + \frac{1}{2} (x - M_1)^T \Sigma_1^{-1} (x - M_1) \right] \frac{\omega_2}{\omega_1} \overset{\omega_1}{\sim} \chi^2
$$

Because the logarithm is a monotonically increasing function, an equivalent test is

$$
-(x - M_2)^T \Sigma_2^{-1} (x - M_2) + (x - M_1)^T \Sigma_1^{-1} (x - M_1) + \ln \left\{ \frac{\Sigma_1}{\Sigma_2} \right\} \overset{\omega_1}{\sim} 2 \ln(\lambda)
$$

By expanding (5.14) we can reduce the LRT to

$$
h(x) = x^T A x + b^T y \overset{\omega_1}{\sim} 2 \ln(\lambda) - c
$$

where

$$
A = \Sigma_1^{-1} - \Sigma_2^{-1}
$$

$$
b = 2(\Sigma_2^{-1} M_2 - \Sigma_1^{-1} M_1)
$$

$$
c = M_1^T \Sigma_1^{-1} M_1 - M_2^T \Sigma_2^{-1} M_2 + \ln \left\{ \frac{\Sigma_1}{\Sigma_2} \right\}
$$
This test is optimal for normally distributed features. A classifier defined according to (5.15) is called a quadratic classifier, because the observed pattern vector \( x \) appears as a quadratic term in the equation. Even if the features are not normal, the quadratic classifier is still attractive, because (5.15) requires the estimation of only second order statistics, namely mean vectors and covariance matrices, rather than complete densities.

It is possible to give a different interpretation to the quadratic classifier by introducing the Mahalanobis distance defined by

\[
\| x - M_i \|_{\Sigma_i^{-1}} \triangleq \left[ (x - M_i)^T \Sigma_i^{-1} (x - M_i) \right]^{1/2} \tag{5.19}
\]

into (5.14) as follows.

\[
\| x - M_1 \|_{\Sigma_1^{-1}}^2 - \| x - M_2 \|_{\Sigma_2^{-1}}^2 \overset{\triangle}{=} 2 \ln(\lambda) - \ln \left| \frac{\Sigma_1}{\Sigma_2} \right| \tag{5.20}
\]

Classification according to (5.20) is performed on the basis of the Mahalanobis distance from the observation vector to each class mean vector.

### 5.2.2 Linear Classifier

If the two classes have identical covariance matrices \( \Sigma_1 = \Sigma_2 = \Sigma \), then \( A \) in equation (5.16) vanishes, and the quadratic classifier rule expressed in (5.15) reduces to

\[
h(x) = b^T y \overset{\triangle}{=} \ln \left| \frac{\Sigma_1}{\Sigma_2} \right| 2 \ln(\lambda) - c \tag{5.21}
\]

where

\[
b = 2 \Sigma^{-1} (M_2 - M_1) \tag{5.22}
\]

\[
c = M_1^T \Sigma^{-1} M_1 - M_2^T \Sigma^{-1} M_2 \tag{5.23}
\]
Because the rule is now linear in the observation vector $\mathbf{x}$, this is called a linear classifier. This is equal to the Fisher classifier defined in Chapter 4, which is now seen to be optimal for equal covariance matrices.

If we make the additional assumption that the covariances are identical and equal to the identity matrix $\mathbf{I}$, then the Mahalanobis distance defined in (5.19) reduces to the Euclidean distance

$$
\| \mathbf{x} - \mathbf{M}_i \| = \left[ (\mathbf{x} - \mathbf{M}_i)^T (\mathbf{x} - \mathbf{M}_i) \right]^{1/2}
$$

and the Mahalanobis distance classifier reduces to the Euclidean distance classifier given by

$$
\| \mathbf{x} - \mathbf{M}_1 \|^2 - \| \mathbf{x} - \mathbf{M}_2 \|^2 \overset{\sim_2}{\gtrsim} 2 \ln(\lambda)
$$

If symmetrical cost functions satisfying $C_{21} - C_{11} = C_{12} - C_{22}$ are employed and the classes are equiprobable, i.e., $P_1 = P_2$, then $\ln(\lambda) = 0$. In this case the Euclidean classifier reduces to the rule

$$
\| \mathbf{x} - \mathbf{M}_1 \| \overset{\sim_2}{\gtrsim} \| \mathbf{x} - \mathbf{M}_2 \|,
$$

which chooses the class whose mean vector is nearest to the observation vector. This classifier is sometimes called the nearest mean classifier, and it is optimal in the Bayes sense only if each class is normally distributed with identical covariance matrices of the form $\Sigma = \sigma^2 \mathbf{I}$. It is often used because of its simplicity even when the assumptions about distributions are not satisfied.

### 5.3 Nonparametric Classifiers

It was shown in the previous section how different assumptions regarding the class distributions lead to simple parametric classifiers. If the underlying distributions
agree with the assumed forms, then the parametric classifiers perform optimally, but if the actual distributions are very different from what was assumed, then the performance may be poor. Nonparametric classifiers do not assume any particular form for the densities, and as a result they are more robust and more attractive in many pattern recognition problems. Two nonparametric classifiers, the nearest-neighbor and neural network classifiers, were used in our work. These classifiers are discussed next.

5.3.1 Nearest-Neighbor Classifier

A nearest-neighbor classifier [45, Ch.7] assigns an observation vector to the class of the nearest training sample. The Euclidean distance metric is most commonly used, but other distance metrics may also be used. While parametric classifiers such as the linear and quadratic classifiers have smooth decision boundaries that are dependent on the assumed distributions, the nearest-neighbor classifier boundaries are usually irregular and dependent on the particular set of training samples used.

A useful property of the nearest-neighbor classifier is the bound of its probability of error $\epsilon_{1NN}$ under the large sample case, which is given by

$$\epsilon_{\text{Bayes}} \leq \epsilon_{1NN} \leq 2\epsilon_{\text{Bayes}} \quad \text{as } N \to \infty$$

(5.27)

where $\epsilon_{\text{Bayes}}$ is the Bayes error and $N$ is the number of training samples [42, 56]. It is important to note, however, that the above bounds may not hold when $\epsilon_{1NN}$ is estimated using a finite number of samples.
5.3.2 Neural Network Classifiers

An artificial neural network consists of a collection of interconnected information processing units called neurons. A model of a single neuron, also known as a perceptron, is shown in Figure 5.1 [57, 58]. For an input vector \([x_1, \ldots, x_N]\), the neuron output \(y_k\) is given by

\[
y_k = f \left( \sum_{i=1}^{N} w_{ki} x_i + b_k \right)
\]  

(5.28)

where \(w_{ki}\) are the synaptic weights of the neuron, \(b_k\) is a bias and \(f(\cdot)\) is the activation function. The log sigmoid function

\[
f(v) = \frac{1}{1 + \exp(-av)}
\]  

(5.29)

where \(a\) is a slope parameter and the hyperbolic tangent function

\[
f(v) = \frac{1 - \exp(-v)}{1 + \exp(-v)}
\]  

(5.30)

are the most commonly used activation functions.

Figure 5.1: A model of a neuron.
A typical neural network contains neurons arranged in one or more layers. A network having only input nodes and an output neuron layer is called a single layer network. A multilayer network contains one or more hidden neuron layers, between the input nodes and the output layer. Neural networks can be classified as feed-forward networks, recurrent networks or lattice networks based on the direction of signal flow [57, Section 1.6]. In a multilayer feed-forward network the source inputs are fed to the first hidden layer, whose outputs become the inputs of the second layer and so on. The signals travel only in one direction. Networks in which feedback is present are called recurrent networks. A lattice network is simply a feed-forward network, whose output nodes are arranged in rows and columns.

Values of the synaptic weights and the biases of all the neurons are free-parameters, which must be adjusted to produce the desired input-output mapping. This adaptation of the free-parameters in a neural network may be done through supervised or unsupervised learning.

In supervised learning, input data from known classes are presented to the network. The error between the network output and the true class is computed. The network parameters are then adjusted to minimize a cost function of the output error, usually the mean square error. One cycle of presenting the entire training data set and adapting the weights is called a training epoch. This training process is repeated for a large number of training epochs until the network reaches an acceptable mean square error or a specified number of epochs has elapsed. The backpropagation algorithm and its variants are widely used for supervised learning in multilayer feed-forward neural networks.
An important characteristic of a neural network is generalization, i.e., the ability to correctly respond to previously unseen input patterns. Achieving a low error during training does not guarantee that the network will generalize well. The ability to generalize depends on many factors including the dimensionality of the data, the number of free parameters, the quantity and quality of training data and the length of training. For example, a neural network classifier with a large number of free parameters trained for too long may learn to classify the training set perfectly, but its generalization may be poor.

One way to improve generalization is to use a validation training approach. In this approach the available labeled data are divided into two sets, a training set and a validation set. The training set is used to adapt the free parameters as explained previously, but after a number of training epochs the network's ability to generalize is tested using the validation set. This train-and-validate process is repeated and the validation error is recorded. Usually, as the number of training epochs is increased, the validation error first decreases and then begins to increase again, at which point the training must be stopped. A modified form of validation training was used in this work.

When a feed-forward multilayer perceptron with backpropagation training is used for pattern classification, the number of output neurons is chosen to be equal to the number of classes, and the number of input nodes is chosen to be the same as the dimension of the feature space. A typical neural network architecture is shown in Figure 5.2. The number of hidden layers and the number of neurons in each layer are determined through empirical guidelines and experimentation. In our work two output neurons are used to represent the clutter and mine classes. While a single
hidden layer was always used, the number of neurons in the hidden layer was selected through experimentation. Neurons with hyperbolic tan activation were used for both the hidden nodes and the output nodes.

![Neural Network Architecture](image)

**Figure 5.2: The typical neural network architecture.**

During the training phase of the neural network classifier, training feature vectors extracted over mines and clutter were input to the network. The desired output was specified to be $[1 -1]^T$ for clutter, and $[-1 1]^T$ for mines. Once the network training was completed, the actual value obtained at each output neuron for any test vector is a real number in the range $[-1, 1]$. A useful property of backpropagation classifiers is that the normalized output values are estimates of the a posteriori probabilities of the classes represented by them, provided the network has been properly trained [59].


5.4 Classifier Performance Evaluation

To assess the benefits of different fusion systems it is necessary to quantitatively evaluate the generalization performance of the classifier. Both feature selection and classifier design influence this performance. In the idealized case when the class-conditioned densities are known, the Bayes error can be evaluated, which allows us to isolate the effects of the features. In practice, however, nonoptimal classifiers must be used, and the roles of features and classifier are inextricably linked.

In the two-class problem the probability of classification error consists of two contributions, namely, the probability of false alarm ($P_F$) and the probability of miss ($P_M$). Performance is often quantified by plotting the receiver operating characteristic (ROC) curve, which is a plot of the probability of true detections ($P_D = 1 - P_M$) versus the probability of false alarms as the decision threshold is varied. The combination of a feature set and a classifier that can detect all true targets without any false detections has a ROC curve identical to that shown at the top left corner of Figure 5.3. A completely noninformative feature or an ineffective classifier whose decision is a random guess results in a ROC curve that is a diagonal line as shown in the figure. Any feature-classifier combination that provides some class separability will have a ROC curve within the region bounded by the previous limiting cases. ROC curves that approach the top left corner indicate more effective combinations, and those close to the diagonal line indicate poor feature-classifier combinations. ROC curves were extensively used in our studies to evaluate the discriminatory power of different features extracted from the demining sensor data.

The area under the ROC curve is also a convenient metric to quantify the classification performance. The area under the ideal ROC curve is equal to unity, and the
Figure 5.3: Receiver operating characteristic (ROC) curves.
area under the worst possible ROC curve (the diagonal line) is 0.5. The area under the ROC curve of a real classifier must be a number between 0.5 and 1, with numbers closer to 1 indicating better performance.

To evaluate the performance of a classifier by computing the probability of classification error or the ROC curves, the classifier must be tested with data whose true classes are known. Since it is the generalization (i.e., ability to correctly classify previously unseen data) that is important, testing data should be independent of training data. The allocation of labeled data for training and testing is an important issue in pattern classification. Because of the effort and expense involved in acquiring measured data with labels (i.e., operating sensors over known mines), the number of available labeled data is limited. Using a finite number of design samples tends to bias the classification error estimates, while using a finite number of testing samples predominantly contributes to the variance [45, Section 5.2]. Therefore, to estimate classifier performance reliably, the available labeled samples have to be used efficiently. Three different methods of classifier evaluation, namely, resubstitution, hold-out, and leave-one-out are defined according to how the available samples are utilized for classifier design and testing.

**Resubstitution Method**

The resubstitution method uses all the available labeled samples for designing and testing. The rationale for this approach is that the bias and variance can be minimized by using as many samples as possible for designing and testing. Because the same data set is used for both tasks, this method always provides optimistic estimates of the generalization performance. In general, the resubstitution method gives a lower bound of the classification error [45, Section 5.3]. For some classifiers such as
the nearest-neighbor classifier, the resubstitution method leads to zero classification error.

**Hold-Out Method**

A more reasonable way to evaluate classifier performance is to have independent training and testing sets. In the hold-out method the labeled samples are divided into training and testing sets to guarantee independence. The error estimates obtained with this method are upper bounds of the classification error [45, Section 5.3]. Because the limited number of available samples is split among design and testing sets, the biases and variances of estimates obtained with the hold-out method can be high.

**Leave-One-Out Method**

The leave-one-out method attempts to use the available labeled samples most efficiently, while maintaining the independence of design and testing sets. In this method, one sample is left out and the remaining samples are used to design the classifier. Then the omitted sample is tested with the classifier. Next, a different sample is left out and tested with a new classifier designed with the remaining samples. The process is repeated until all samples have been tested. The leave-one-out method is a limiting case of \(k\)-out-of-\(N\) cross validation, in which \(k\) samples are left out for testing. Because this process guarantees the independence of the tested sample and the design set, it provides an upper bound on the classification error (like the hold-out method) [45, Section 5.3]. Because all the samples are tested, and all but one samples are used for designing the classifier, the bias and variance of the estimates are smaller than those in the hold-out method. The drawback of the leave-one-out method is that to test \(N\) samples, one needs to design \(N\) classifiers, which can be
time consuming. Because the number of available labeled samples was small in our studies, we employed the leave-one-out method for evaluation.

5.5 Summary

This chapter presented an overview of pattern classifiers and methods to evaluate their performance. Section 5.1 reviewed statistical decision theory, which is the basis for the classification methods. Section 5.2 described several widely used parametric classifiers. Section 5.3 discussed nonparametric classifiers, giving particular emphasis to neural network classifiers. Techniques for assessing classifier performance were considered in Section 5.4.
CHAPTER 6

CASE STUDY I: ELECTROSCIENCE LABORATORY
MINE SURROGATE TEST SITE

This chapter (and Chapter 7, to some extent) presents a sensor fusion study that complements and extends the prior works cited in Section 3.3. Each of those investigations showed that sensor fusion has the potential to improve mine detection performance, but the data were acquired on carefully controlled test areas, and the sensors were well registered. Most of the prior studies were limited to decision-level fusion.

The results presented in this chapter are based on multisensor data collected in November 1998. A number of surrogate mines and clutter were buried in native soil near the ElectroScience Laboratory. The sensor suite comprised EMI, IR, and GPR sensors. The resulting data set was fused using several techniques.

This study had two objectives. First we sought to gain insight into the problems faced in acquiring and processing demining sensor data collected under realistic conditions. Therefore, the test site was not specially prepared prior to the emplacement of targets and, hence, had a very irregular topology and contained a large amount of unknown clutter. The mine surrogates buried in the test site also were selected to be quite challenging.
The second objective of this study was to test the feature-level fusion algorithm described in Chapter 3 for noncoincidently sampled multisensor data. As such, the sensors used in this study were allowed to sample the ground at noncoincident locations. Furthermore, the three sensors used in this study were mounted on different platforms, and positioning was not controlled precisely.

The investigations presented herein extend prior studies in several respects. First, we consider both feature-level and decision-level fusion of experimental data from noncommensurate (EMI, GPR and IR) sensors. Second, we investigate the benefits of fusion under more realistic conditions by using a test site containing large amounts of clutter and having an irregular surface. Finally, each sensor was used on a different platform, which led us to consider the effects of noncoincident sampling and positioning errors.

Section 6.1 describes the test site. Section 6.2 discusses the sensors and the signal processing. The results of the study are presented in Section 6.3.

### 6.1 Description of the Test Site

A surrogate mine field was created to test our sensors under conditions that approximate those found in field situations. The site location, adjacent to the ElectroScience Laboratory, is former farm land. A total of 28 mine-like and clutter-like targets were emplaced in a $4 \times 10$ grid as shown in Figure 6.1. An abbreviation identifying the buried object and the depth in inches are shown near each object in the map. The identities of the buried objects are given in Table 6.1.

This site presented a surprising number of challenges to mine detection, and data from it are valuable in that they illustrate problems that could be encountered in real
Figure 6.1: The layout of the ESL mine surrogate field.
demining operations. The target set was deliberately selected to be stressing to our sensors. In contrast to real mines, essentially all of which contain some metal, roughly half the mine surrogates were metal-free. In addition, some of the surrogate mines are quite small (only 1.5 inches in diameter) and undetectable to all of the sensors. Those objects were re-designated as clutter. The site has been intermittently used for various engineering experiments for more than 40 years, which produced a significant amount of metal clutter, primarily in the form of nuts, bolts, screws, and bits of wire.

Consistent with our goal of achieving realistic environmental conditions, the test area was not cleared prior to emplacing the targets. As a result, during two years of testing, roughly 100 metallic fragments and a number of stones were located in (and subsequently removed from) the region. When the targets were initially emplaced, a

<table>
<thead>
<tr>
<th>Abbr.</th>
<th>Object</th>
<th>Dia. (in.)</th>
<th>Ht. (in.)</th>
<th>Metal Content?</th>
<th>Identity</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>Soda can</td>
<td>2.5</td>
<td>5</td>
<td>Y</td>
<td>clutter</td>
<td>4</td>
</tr>
<tr>
<td>CX</td>
<td>Crushed soda can*</td>
<td>3</td>
<td>5</td>
<td>Y</td>
<td>clutter</td>
<td>1</td>
</tr>
<tr>
<td>PT</td>
<td>Pop top*</td>
<td>0.5</td>
<td>1</td>
<td>Y</td>
<td>clutter</td>
<td>2</td>
</tr>
<tr>
<td>M35</td>
<td>Aluminum disk</td>
<td>3.5</td>
<td>5/8</td>
<td>Y</td>
<td>surrogate</td>
<td>1</td>
</tr>
<tr>
<td>P15</td>
<td>Plexiglas disk</td>
<td>1.5</td>
<td>2</td>
<td>N</td>
<td>clutter</td>
<td>4</td>
</tr>
<tr>
<td>R4</td>
<td>Rock</td>
<td>4</td>
<td>2</td>
<td>N</td>
<td>clutter</td>
<td>2</td>
</tr>
<tr>
<td>V4</td>
<td>Void (Styrofoam)</td>
<td>3.5</td>
<td>5/8</td>
<td>N</td>
<td>clutter</td>
<td>2</td>
</tr>
<tr>
<td>H4</td>
<td>Refilled hole</td>
<td>6</td>
<td>2</td>
<td>N</td>
<td>clutter</td>
<td>2</td>
</tr>
<tr>
<td>N35</td>
<td>Nylon disk</td>
<td>4</td>
<td>2</td>
<td>N</td>
<td>surrogate</td>
<td>2</td>
</tr>
<tr>
<td>N3P</td>
<td>Nylon disk with steel pin</td>
<td>3</td>
<td>1</td>
<td>Y</td>
<td>surrogate</td>
<td>4</td>
</tr>
<tr>
<td>T7</td>
<td>Teflon disk</td>
<td>4</td>
<td>2</td>
<td>N</td>
<td>surrogate</td>
<td>2</td>
</tr>
<tr>
<td>N7</td>
<td>Nylon disk</td>
<td>7</td>
<td>1</td>
<td>N</td>
<td>surrogate</td>
<td>2</td>
</tr>
<tr>
<td>-</td>
<td>Blank</td>
<td>-</td>
<td>-</td>
<td>?</td>
<td>clutter</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 6.1: Description of the objects buried in the MURI mine test grid. For objects indicated by an asterisk, dimensions are width and length.
lush grass cover was present in the area, which was subsequently removed to improve the performance of the IR camera. The native Ohio topsoil was not removed. The soil contains a large amount of clay, which has a negative effect on GPR performance. The surface of the mine field contains a number of topological irregularities, but it is planar to within an estimated variation of ±2 inches. The objects in this field had been in situ for more than 18 months at the time these data were acquired.

Data were acquired along the paths indicated by vertical line segments shown in Figure 6.2. The GPR samples were acquired at ~0.5 inch intervals along north-south (N-S) lines spaced every two inches east-west (E-W). The EMI sensor acquired samples at ~1 inch intervals along N-S lines spaced six inches E-W. The field of view of the IR camera was such that a sequence of eight IR images was necessary to cover the complete test grid as indicated by the trapezoids. The square symbols in the
trapezoids represent Styrofoam fiducial markers placed on the region to permit later image warping and registration. The small circles are the target grid points shown previously.

6.2 Sensors and Signal Processing

To test the fusion algorithms, experimental data were acquired with a sensor suite comprising a commercial electromagnetic induction (EMI) sensor, a ground penetrating radar (GPR), and a commercial infrared (IR) camera. In this section we present brief descriptions of these sensors and the signal processing used to derive features and decisions from their data. General descriptions of these sensor technologies appear above in Section 2.3. The discussions here give specifics related to this data collection.

6.2.1 Electromagnetic Induction (EMI) Sensor

Hardware

The EMI sensor used in this work was a commercially available pulsed-induction metal detector, the Schiebel AN-19/2 (Schiebel Instruments, Inc., Washington, D.C.). The default output of this sensor is an audio tone presented to the operator via a pair of headphones. The device was modified slightly for the purposes of this data collection, and the instrument response was sampled with a digital oscilloscope (Tektronix TDS 520) at an internal signal within the sensor ("test point 3"). This internal signal approximates the decaying exponential waveform intrinsic to all pulsed induction EMI sensors. The sensor was fitted to a motor-driven linear scanner, which was used to traverse a 55 inch scan over the ground. The digital oscilloscope acquired 56 waveforms along each scan, and stored the data in a laptop computer under the control
of a Labview program. Scans were performed at regularly spaced intervals (nominally every six inches) over the region of interest to obtain a grid of sample points. A sensor height of approximately two inches was maintained over the region, but irregular surface features caused ±2 inch variations in this elevation. This variation in sensor height produced significant problems in the detection of small targets.

**Signal Processing**

Signal processing for the EMI sensor is a multistep process. Example raw signatures are shown in Figure 6.3(a) for ground containing (1) no metal, (2) a mine surrogate with an embedded metal pin, and (3) a 3 inch diameter metal disk. In these waveforms the large rectangular pulse present between approximately samples 50 and 70 is the transmitted pulse, which is discarded. Next, a sensor response acquired over soil known to be metal-free is subtracted from the waveforms. The resulting difference signal is not well approximated by a decaying exponential (especially for small mines), but its integral is an effective detector of buried metal. When these integrated values from each of the 56 scan positions are plotted, they produce spatial signatures such as those illustrated in Figure 6.3(b). For targets not too near the sensor head and not too large compared to the loop radius, the sensor's spatial response has a mono-polar shape that is reasonably well approximated by the function

\[
G_1(\mathbf{R}; \Theta_1) = B + \frac{S}{1 + (||\mathbf{R} - \mathbf{R}_0||/\alpha)^b}
\]  

where \( \mathbf{R}_0 \) is the location of the target's centroid, \( B \) is the response of the background, \( S \) is the amplitude of the target response and \( \alpha \) and \( b \) are shape parameters. Case Study 2 in Chapter 7 gives situations where this simple model does not hold and describes a more sophisticated physics-based model of the response of a buried sphere.
The second step in EMI signal processing involves estimation of the feature vector \( \Theta_1 \) given by

\[
\Theta_1 = [a \ b \ S \ E \ R_0]^T
\]  

Note that the background level \( B \) is not used in \( \Theta_1 \), and we have added the quantity \( E \), which is the residual error in the estimate normalized to the signal energy.

To compute \( \Theta_1 \) at each sample position \( R \) a spatial window was defined and the data \( D_1 \) within this window was fitted to the model using a nonlinear optimization technique. Since stronger signals tend to persist over larger distances, the size of the window was adjusted on the basis of the signal amplitude.

Figure 6.3: Typical raw EMI waveforms collected over metal-free ground, a mine-surrogate, and a metal disk.

presented by Das et al. [60]. The simple empirical form above, however, was adequate to approximate the limited set of targets of interest in this study.
6.2.2 Ground Penetrating Radar

Hardware

The GPR used in this effort was developed at The Ohio State University (OSU) ElectroScience Laboratory (ESL) [61]. It employs a novel dielectric rod antenna that was scanned horizontally over the earth at a fixed height (nominally two inches). A network analyzer was used to measure the complex reflection coefficient at 51 frequencies between 1 and 6 GHz. Approximately 100 frequency scans were acquired on each spatial scan, which had a length of 55 inches. The antenna and its supporting platform were advanced in two inch intervals, resulting in a grid of samples.

Calibration of the radar was performed via a two step process. A “background” signature (the response of the radar when the antenna points into free space) was acquired and subtracted from all data. The impulse response of the radar was then determined by measuring the field scattered from a reference target (a short cylinder) in free space. The ideal scattered field for this target is known from numerical calculations, and a comparison of the measured and ideal scattering allows the impulse response of the radar to be determined. This response is then deconvolved from the measured data.

Signal Processing

Processing of the GPR data begins with a Fourier transform of the calibrated swept-frequency measurements into the time domain, producing a map of the reflection coefficient as a function of along-scan position and time (or depth). In the sequel we refer to these plots as “GPR maps”. Because the reflection at the air-ground interface is the dominant source of clutter in GPR data, removing this ground-reflected
clutter is an important part of GPR data processing. In processing the data in Case Study 1 a technique developed by Gupta et al. [62] was used to reduce the ground-reflectected clutter. We defer the detailed discussion of clutter reduction until Chapter 7 on Case Study 2, where a new, more effective clutter reduction technique is described.

Once the ground-reflectected clutter is removed, the data are modeled as a point scatterer plus clutter. The response of a point scatterer for a typical mine depth in homogeneous soil is computed analytically and used as a signature template. Target depth has an effect on both the signature shape and time delay, but because the range of mine burial depths is restricted, the effect of depth errors on target shape was found to be insignificant. The GPR feature vector is given by

$$\Theta_2 = [\rho \ d \ R_0]^T$$

where $\rho$ is the peak correlation coefficient for the model signature and data, and $d$ and $R_0$ are respectively the depth and horizontal position indicated by the correlation peak. Although false targets are often seen in GPR scans, they tend not to persist across adjacent scans and, hence, the correlation was computed over a three dimensional data window (depth, along-scan, and cross-scan dimensions). GPR data tends to be highly correlated between adjacent along-scan positions, because of returns from plane-stratified media and antenna ringing not eliminated by calibration. A modest amount of whitening was performed prior to the correlation.

### 6.2.3 Infrared Camera

**Hardware**

The IR sensor used in this study was a commercially available MWIR camera, the IRRIS 160ST, (Cincinnati Electronics, Cincinnati, OH). The sensor comprises an
InSb photovoltaic array of 160 by 120 pixels. The optics provide a field of view of 9.1° by 6.8° for an IFOV of approximately 1 mrad. The noise equivalent temperature difference ($NEΔT$) for this sensor is 0.025 K. The camera was positioned on the roof of a two-story building adjacent to the test site, and data were acquired from this vantage point. The camera’s field of view did not permit sampling of the entire mine field with an acceptable pixel density, and a sequence of images was necessary as described in Section 6.1. Styrofoam fiducial markers were placed on the region to permit later image warping and registration.

**Signal Processing**

IR data collection and signal processing tasks were performed by Mr. I. Kursat Sendur, another graduate student at the ElectroScience Laboratory. The discussion on the IR signal processing procedure is based on [27] and is included for the sake of completeness.

Processing the IR data begins by remapping the imagery to ground coordinates using a standard perspective transformation. This is followed by bilinear interpolation of the distorted pixels and resampling to a uniform pixel size. This processing produces a trapezoidal-shaped image in the ground coordinates. A remapped IR image is shown in Figure 6.4, in which the dark square masses are the fiducial markers, and the encircled grey area is a buried mine surrogate. Other randomly located grey shapes are clutter due to variations in the ground’s thermal properties or emissivity.

An image chip is extracted at each interrogation point $R$. Based on empirical observations, we employ a signature model identical to that used for the EMI sensor,
Figure 6.4: An IR image remapped to ground coordinates. The encircled grey shape corresponds to a buried mine surrogate. The dark square masses are fiducial markers.

namely,

\[ G_3(R; \Theta) = B + \frac{S}{1 + (|R - R_0|/a)^b} \]  \hspace{1cm} (6.4)

The feature vector is

\[ \Theta_3 = [a \ b \ S \ E \ R_0]^T \]  \hspace{1cm} (6.5)

where \( E \) is the normalized residual error energy \(|D - G|^2/|D|^2\).

6.3 Results

6.3.1 Performance of Individual Sensors

For each sensor \( i \), feature vectors \( \Theta_i \) were extracted using the signal processing algorithms discussed above and supplied to sensor-specific classifiers \( Pr(u_i|\Theta_i, t_i) \). Results for the binary detection problem involving hypotheses \( H_1 = \text{"non-mine"} \)
and $H_2 = \text{"mine"}$ are reported here. To permit comparison between sensors we have taken all intermediate sensor-specific hypotheses $h_{ki}$ to be $H_k$, even though some surrogates are undetectable to some sensors. Backpropagation neural networks were used to approximate the required a posteriori probabilities. The number of input and output nodes for these networks were defined by the sizes of $\Theta_i$ and $K$ respectively. The number of hidden nodes were 3, 7, and 6 for the GPR, EMI, and IR sensors respectively. Target detections $u_i$ were produced by thresholding the a posteriori probability ratio at level $t_i$. Leave-one-out training and validation was used in all of the results presented in this chapter. Because networks trained with random weight initializations may converge to spurious local error minima, several trials were conducted and the best results were retained.

Figures 6.5(a), 6.5(b) and 6.5(c) present binary detection maps for each sensor, which were generated by choosing appropriate detection thresholds $t_i$ and discarding regions with small number of adjacent detections. The dark masses on these maps indicate detections, circles represent the locations of the declared mine surrogates, and the squares indicate detection cells used to determine whether a detection declaration made in the vicinity of a mine should be considered as due to that mine or not. Only those cells that contained mine detections are shown in the maps. It is clear from these results that the EMI and IR sensors have a high false alarm rate when all targets are detected. The poor performance of the EMI sensor was anticipated, because most of the mine surrogates in our test grid are nonmetallic and the site contains a large quantity of metallic clutter. The poor performance of the IR sensor was not anticipated, but may be due to the relatively large number of surface irregularities. The GPR sensor shows reasonably good performance. It was found that for each
Figure 6.5: Detection maps for the individual sensors. The dark masses represent detections. Circles give the mine locations. Detection cells are represented by squares.
sensor there exists a class of targets that are very difficult to detect, and these targets control the net sensor performance. As an example, small, deeply buried targets are very stressing to the IR system.

The performance of these sensors can be compared quantitatively using ROC curves, as shown in Figure 6.6. These results confirm the findings shown in the detection maps. To compute $P_D$ and $P_F$ the test area was divided into a large number

Figure 6.6: ROC curves obtained for individual sensors.
<table>
<thead>
<tr>
<th>Sensor</th>
<th>$P_F$</th>
<th>$P_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPR</td>
<td>0.0929</td>
<td>0.0909</td>
</tr>
<tr>
<td>EMI</td>
<td>0.3755</td>
<td>0.4545</td>
</tr>
<tr>
<td>IR</td>
<td>0.5613</td>
<td>0.0909</td>
</tr>
</tbody>
</table>

Table 6.2: Individual sensor $P_F$ and $P_M$ using “good” thresholds.

of square detection cells (280). A cell was considered to contain a detection if at least four adjacent detections fell in that cell. We computed the false alarm probability as the number of cells with false detections divided by the total number of cells not occupied by mines. This approach is reasonable, because we are only comparing the sensors and fusion methods based on the data at this same site.

6.3.2 Performance of Fusion Algorithms

For decision-level fusion we used the “good threshold” approximation described in Section 3.5.1 to avoid the search over all possible sensor thresholds $t_i$ implied in equation (3.43). The EMI and IR sensor thresholds were selected such that all expected targets (i.e., targets with metal for the EMI sensor and targets that present a sizable thermal discontinuity for the IR sensor) were detected while minimizing clutter detections. The criterion used for selecting the GPR threshold was the “knee” in the ROC curve in Figure 6.6. The resulting individual sensor $P_{Mi}$ and $P_{Fi}$ values are shown in Table 6.2. The detections $u_i$ for these thresholds comprise the inputs $u_i$ for the decision-level fusion algorithms and are shown in the foregoing individual sensor detection maps. The local confidences $Pr(u_i|\Theta_i, t_i)$ required for the soft-decision fusion
scheme of equation (3.45) were set to the a posteriori probability estimates produced by the corresponding classifier output.

ROC curves for hard and soft decision-level fused classifiers are shown in Figure 6.7 using dotted and dashed lines respectively. Comparing Figures 6.6 and 6.7 we see that hard-decision fusion based on equation (3.43) offers little benefit over the GPR sensor alone. Based on the form of equation (3.43), which weights most strongly the most
Figure 6.8: Detection map obtained using soft decision-level fusion

reliable sensors, this result was expected. The performance of soft-decision fusion based on our extension to (3.43) in equation (3.45), however, is somewhat better. It is clear that with soft-decision fusion $P_D=100\%$ can be achieved with a smaller $P_F$, which is a significant improvement over the best individual sensor performance. The detection map obtained for soft-decision fusion is shown in Figure 6.8. The detection map for hard-fusion is not shown, because it is nearly identical to the GPR detection map.

For feature-level fusion the sensor feature vectors $\Theta_i$ were concatenated and used in a single classifier (cf. equation (3.10)), resulting in the ROC curve shown in Figure 6.7. As done for the individual sensors, the classifier $Pr(H_k|\hat{\Theta})$ was approximated by a backpropagation neural network. A total of 14 hidden nodes were used in this network.

Using as a figure of merit the smallest $P_F$ at which $P_D = 100\%$, we see that the feature-fused system is significantly better than the individual sensors and marginally
better than the soft-decision fusion case. The detection map produced by the feature-fused system, corresponding to the largest threshold that can achieve \( P_D = 100\% \), is given in Figure 6.9. Comparing the soft-decision case in Figure 6.8 and the feature-level case in Figure 6.9 the reduction in false alarm rate indicated by the ROC curve (roughly 40\%) is evident.

**Alternative Forms of Fusion**

Recall from Section 3.5.1 that nonoptimal fusion rules such as AND, OR, and majority have been used in some applications, especially when dealing with identical and independent sensors. Some of these rules may be successfully used even for non-identical sensors under suitable conditions. For example, as discussed in Section 3.3, a simulation study by Weisenseel et al. [14] showed that AND-fusion of EMI and GPR was effective for demining under some simplifying assumptions. It is instructive to investigate how some of these fusion methods perform with more realistic data sets. When AND, OR, and the majority rule are applied to the binary detection maps
obtained with the individual sensors, the fused detection maps in Figure 6.10(a), 6.10(b), and 6.10(c) are obtained. It is interesting to note that the performance of the AND rule based fusion is extremely poor for these realistic data, which contrasts with the observation in [14]. Fusion based on the OR rule achieves $P_D = 1$, but with a very large increase in the number of false alarms. Majority fusion leads to a reasonably small number of false alarms, but does not detect all the mines.

An ad hoc hybrid fusion scheme involving both feature-level and decision-level fusion was also considered. In that approach, sensor pairs comprising GPR and EMI, and GPR and IR were fused at the feature-level, and the binary decision maps produced by each fused pair were fused using the AND rule. This ad hoc approach produced the final binary decision map in Figure 6.11, which is comparable to that obtained using feature-level fusion of all three sensors. If feature-level fusion of all the sensors is possible, then there is no performance benefit in adopting this kind of hybrid fusion scheme over feature-level fusion. Because of feature dimensionality issues, however, feature-level fusion of all the sensors may sometimes be impractical, while the feature-level fusion of subsets of the sensors may still be possible. Under such circumstances, hybrid fusion schemes similar to the one considered here may be used to retain the benefits of feature-level fusion to some extent. Figure 6.12 shows the operating points obtained for mine detection using AND, OR, majority, and the ad hoc hybrid fusion schemes. The ROC curves shown earlier for the other fusion methods are also shown again for comparison.
Figure 6.10: Detection maps obtained for AND, OR, and majority fusion of hard decisions.
6.3.3 Effect of Sensor Reliability Estimates

Both optimal hard decision-level fusion and soft decision-level fusion require knowledge of sensor reliabilities, which are specified in terms of the $P_{Fi}$ and $P_{Mi}$. These quantities are estimated from the prior performance of the individual sensors. Because sensor reliabilities depend on the type and number of mines and clutter and on the environment, the estimates obtained from prior data may not necessarily be appropriate for new measurements. In our work, we used $P_{Fi}$ and $P_{Mi}$ values obtained from the individual sensor detection maps of Figure 6.5(a), 6.5(b), and 6.5(c), which gave the values shown in Table 6.2. To investigate how changes in these values affect the fused performance, we modified some of the $P_{Fi}$ and $P_{Mi}$ values and obtained the ROC curves displayed in Figure 6.13. The solid curve, which displays the best performance, corresponds to the estimates given in Table 6.2. The dotted curve is obtained when all the sensors are assigned identical reliability factors $P_{Fi} = 0.18$ and
Figure 6.12: The operating points of AND, OR, majority and the ad hoc fusion based mine detection.
Figure 6.13: The effect of sensor reliability estimates $P_{Fi}$ and $P_{Mi}$ on the performance of hard decision-level fusion with the optimal rule. The subscripts $i = 1, 2, 3$ correspond to EMI, IR and GPR respectively.
$P_{Mi} = 0.36$. Because the true sensor reliabilities are far from identical, the fused performance is quite poor. The other two ROC curves correspond to smaller deviations from the actual values and exhibit smaller amounts of degradation. These results demonstrate the importance of selecting correct values for sensor reliabilities in decision-level fusion. A more thorough study of this topic should be done by performing a numerical optimization of the area under the ROC curve using $P_{Mi}$ and $P_{Fi}$ as parameters.

### 6.3.4 Effect of Prior Probability Estimates

Decision-fused results are also affected by the values assumed for the prior probabilities of mines and clutter, as indicated by the maximum a posteriori probability based decision-level fusion using equation (3.44). Figure 6.14 shows the ratio of the a posteriori probability and the detection map obtained for a fixed threshold value assuming the prior probabilities to be $Pr(\text{mines}) = 0.0393$ and $Pr(\text{clutter}) = 0.9607 = 1 - Pr(\text{mines})$. The value of $Pr(\text{mines})$ was obtained by dividing the number of cells occupied by mines (= 11) by the total number of cells (= 280), Figures 6.15 and 6.16 show the a posteriori probability ratios and the detection maps when the prior probability of mines is overestimated and underestimated, respectively. Although the accuracy of prior probability and sensor reliability estimates is an important concern in mine detection, it is not a problem peculiar to sensor fusion. It occurs in target detection even with a single sensor. The only way to reduce these problems is to ensure that the estimates are based on data collected under conditions similar to those present in the field.
Figure 6.14: A posteriori probability ratio and detection map obtained using hard decision-level fusion assuming $Pr(\text{ mines}) = 0.0393$ and $Pr(\text{ clutter}) = 0.9607$. 

(a) Map of a posteriori probability ratios

(b) Detection map corresponding to a fixed threshold
Figure 6.15: A posteriori probability ratio and detection map obtained using hard decision-level fusion when the prior probability of mines is overestimated using $Pr(\text{mines}) = 0.2857$ and $Pr(\text{clutter}) = 0.7143$. 
Figure 6.16: A posteriori probability ratio and detection map obtained using hard decision-level fusion when the prior probability of mines is underestimated using $\Pr(\text{mines}) = 0.0036$ and $\Pr(\text{clutter}) = 0.9964$. 
6.3.5 Feature Discrimination Potential

The detection maps and ROC curves presented in this chapter provide evidence of the benefits of using multisensor fusion for mine detection. We also showed that these results are affected by a variety of issues such as the accuracy of prior probability and sensor reliability estimates, and the effects of dimensionality and training set size.

In this section we considered another approach to verifying the potential advantages of multisensor fusion, namely, by directly computing the discriminatory power of individual sensor features and combinations of them. The extended Fisher criterion (EFC) of (4.37) was used as the criterion to measure discriminatory power. Data sets with identical mean vectors and covariances lead to an EFC of zero, while larger criterion values indicate better class separability. Because the EFC is based only on differences in means and variances, class separability due to higher order statistics is not reflected. Thus, large EFC values are sufficient but not necessary for good class separability.

The criterion values, which were computed using the training data, for individual features and several combinations of features are listed in Table 6.3. We see that the GPR feature has a larger criterion value than the EMI and IR features, which agrees with the previous detection maps and the ROC curves. Another interesting observation is that intra-sensor feature fusion, i.e., the combining of different features of the same sensor, also leads to an increase in the discriminatory power. Combining features from pairs of sensors increases the discrimination, while combining features from all three sensors produce the largest EFC value, indicating the highest discriminatory power. This result is very satisfying, because it confirms the conjecture that fusion of the features from multiple demining sensors can lead to improvements in
Table 6.3: Discriminatory power of individual features and combinations of features according to the extended Fisher metric.

<table>
<thead>
<tr>
<th>Feature Description</th>
<th>Discriminatory Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{\text{EMI}}$ (Amplitude)</td>
<td>2.4178</td>
</tr>
<tr>
<td>$E_{\text{EMI}}$ (Error)</td>
<td>1.1201</td>
</tr>
<tr>
<td>$a_{\text{EMI}}$ (Size)</td>
<td>1.2942</td>
</tr>
<tr>
<td>$b_{\text{EMI}}$ (Power law)</td>
<td>1.7368</td>
</tr>
<tr>
<td>$E_{\text{IR}}$ (Error)</td>
<td>0.5278</td>
</tr>
<tr>
<td>$a_{\text{IR}}$ (Size)</td>
<td>1.2701</td>
</tr>
<tr>
<td>$b_{\text{IR}}$ (Power Law)</td>
<td>0.2454</td>
</tr>
<tr>
<td>$S_{\text{IR}}$ (Amplitude)</td>
<td>1.6066</td>
</tr>
<tr>
<td>$\rho_{\text{GPR}}$ (Correlation coefficient)</td>
<td>4.9869</td>
</tr>
<tr>
<td>all EMI features</td>
<td>2.5441</td>
</tr>
<tr>
<td>all IR features</td>
<td>3.2033</td>
</tr>
<tr>
<td>EMI and IR features</td>
<td>3.5017</td>
</tr>
<tr>
<td>EMI and GPR features</td>
<td>6.5632</td>
</tr>
<tr>
<td>IR and GPR features</td>
<td>5.4176</td>
</tr>
<tr>
<td>EMI, IR, and GPR features</td>
<td>6.8731</td>
</tr>
</tbody>
</table>
discrimination of mines from clutter. A major challenge, however, is obtaining a sufficiently large training data set so that negative effects of the Hughes phenomenon do not degrade the performance of the fused classifier.

6.4 Summary

We tested the fusion algorithms on noncommensurate noncoincidently sampled data acquired by EMI, GPR, and IR sensors over a test field containing mine surrogates and both deliberate and unknown clutter. Both the targets and clutter present in the test site were very challenging. The results presented here indicate that the performance of feature-level fusion and soft-decision fusion can be significantly better than that of the individual sensors. This performance enhancement is encouraging, especially when we recognize that the IR and EMI sensors were only marginally effective due to both the heavy clutter at the test site and our selection of surrogate targets. As expected, hard-decision fusion showed little improvement over the performance of the best individual sensor, since the algorithm emphasizes detections by the most effective sensors. The AND and OR fusion of hard decisions did not produce useful results. A majority fusion scheme for hard decisions produced a better result than the AND and OR fusion methods, but the result was not as good as those obtained with soft decision fusion and feature-level fusion. An ad hoc hybrid form of fusion involving both feature-level and decision-level fusion produced results comparable to feature-level fusion.

Overall, we have shown that feature-level and soft decision-level fusion of noncoincidently sampled EMI, IR and GPR data can result in significant improvements
in mine detection performance. We found an advantage in using feature-level fusion, which provided a 40% reduction in false alarm rate beyond the soft-decision fusion case. The computational costs of decision-level and feature-level fusion algorithms are approximately equal since the computational expense involved in fusion is dominated by features estimation, and all methods of detection must perform this processing. There is, however, a small penalty to be paid in acquiring training data for feature-level fusion, since the classifier must be somewhat more complex to provide this additional performance. This work extends previous studies by others in sensor fusion for mine detection, because it involves fusion of noncoincidently sampled data, comparison of feature-level and decision-level fusion, and a very realistic test site and data collection methods.

The extended Fisher criterion was used to demonstrate the potential of sensor feature fusion for improving the discrimination of mines from clutter. Intra-sensor fusion as well as the multisensor fusion were shown to improve the discriminatory power.

Although these results were encouraging, the small sample size of our data set suggested that a larger study be done to confirm these results. Furthermore, because the target set considered in this study comprised mine surrogates, not real mines, it is useful to investigate the detection performance for real mines. Therefore, another multisensor data collection effort was conducted at Fort A. P. Hill, VA, which has a larger minefield containing real, deactivated mines. This new study is described in Chapter 7.
CHAPTER 7

CASE STUDY II: FORT A. P. HILL, JUXOCO SITE

In this chapter we describe a study based on multisensor data collected at a mine test field in Fort A. P. Hill, Virginia, during June 15-18, 1999. In Case Study I we acquired useful information about the inherent difficulties encountered in processing and fusing demining sensor data collected in irregular, heavily cluttered ground, but that study was limited in several aspects. First, the targets used were mine surrogates, which were homogeneous disks of Nylon, Teflon, or aluminum, rather than real mines. Second, the number of mine surrogates in the test site was very limited. Although the large amount of unknown clutter in the surrogate test site was consistent with our goals for that study, it made performance evaluation difficult. Overcoming some of these limitations was the objective of the data collection effort and the subsequent data analyses presented in this chapter. Case Study II involves measurements collected over real, deactivated mines.

Another important difference between Case Studies I and II was the manner in which the site was surveyed. In Case Study I the sensors examined a contiguous area of roughly 27 square meters in a manner similar to the survey of a true minefield. In Case Study II the sensors were used only at discrete locations where specific mine or clutter targets had been emplaced.
Analysis of the Fort A. P. Hill data resulted in several important contributions. One contribution was the development of a new GPR clutter reduction algorithm, which greatly improves the detection of small, shallowly buried mines and works well even when the ground profile and the electrical properties are not uniform. An extensive investigation of feature extraction for EMI and GPR data showed that the use of statistical discriminant features, rather than the more commonly used representation features, significantly improves the discrimination of mines from clutter. Another useful product of this study was the design and implementation of a modular and flexible multisensor fusion architecture and a graphical user interface (GUI). This GUI, implemented in Matlab, allows one to add or delete sensors and features without modifications to existing source code. Generation of ROC curves and experiments with different combinations of sensors, features, classifiers, and fusion methods are very convenient with this tool.

This chapter is organized as follows. Section 7.1 describes the test site, sensors and the measurement process. Signal processing and feature extraction for EMI data are presented in Section 7.2. New GPR clutter reduction algorithms and feature extraction techniques are discussed in Section 7.3. In Section 7.4 IR signal processing and feature extraction are outlined. The results of Case Study II are presented in Section 7.5. The software implementation of our fusion architecture is discussed separately in Appendix A.

7.1 Description of the Measurement Site and Sensors

A multisensor data collection was conducted in a mine test field at the US Army Fort A. P. Hill, Site 71A, located in Virginia. This mine test site, which is operated
by the Joint Unexploded Ordnance Coordination Office (JUXOCO), was constructed by excavating more than 6 inches of the native topsoil to remove metallic clutter introduced by previous explosives testing, demolition, and military training. Metal detectors were used to locate and remove metallic clutter remaining after the excavations [63]. The test area was leveled using a commercial grader, but wheel tracks of test platforms used in previous tests, and irregularities caused by target emplacement and rain damage were noticeable in some parts of the test field.

Figure 7.1. Two separate test grids are located at the test site. A 25 m × 5 m grid, used as a calibration area, contains targets whose identities and original burial depths are made available to users. A 50 m × 20 m blind test area contains a much larger number of targets, but their identities are not divulged. Each area is divided into 1 m × 1 m squares, and targets are emplaced at the centers of selected squares. Corners of the squares are marked with PVC pipe to make it easy to locate the targets. Although measurements were collected in both the calibration and the test lanes, data collected in the latter area are not considered in this work, since no ground truth data is available. Therefore, all data processing and analyses described in this
study are based on measurements done over the calibration lane illustrated in Target locations in the test site are conveniently referenced using letter-number pairs, which specify the rows and the columns in the grid. The identities of the objects in the calibration lane are given in Table 7.1. In total, 27 mines and 32 clutter objects are present in the calibration lane.

Four sensors were used to acquire data at Fort A. P. Hill. Three of the sensors were previously used in Case Study I and are described in detail in Chapter 6. The other sensor was an Agema long wave IR (LWIR) camera, which operates in the spectral band 8-12 μm. The LWIR band is different from that used in the IRRIS camera, but the signal processing for the Agema camera was identical to that described in Section 6.2.3 for the IRRIS camera.

The main differences in the measurement processes in the two data collection efforts were in the sensor platforms and the way the ground was sampled. In Case Study I the EMI and GPR sensors used different platforms, and the data collections were done separately on different days. For the Fort A. P. Hill data collection effort both sensors were mounted on the same linear scanner (see Figure 7.2) to expedite the measurement process. The large area to be covered at the A. P. Hill site and the slow data collection rate compelled us to collect only a single scan across each target. As a result, the data in this study did not permit us to generate two-dimensional detection maps similar to those obtained in Case Study I. The scans over each target covered a 24 inch length approximately centered on the target. The GPR sensor sampled the ground at 121 points within this length, while the EMI sensor collected 88 samples.

In Case Study I, the IR camera imaged the entire area of interest with a sequence of eight images. In Case Study II the IR camera was mounted on a tripod positioned
<table>
<thead>
<tr>
<th>Cell</th>
<th>Target</th>
<th>Metal (g)</th>
<th>Depth (inch)</th>
<th>Cell</th>
<th>Target</th>
<th>Metal (g)</th>
<th>Depth (inch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>VS-50</td>
<td>18.21</td>
<td>surf-flush</td>
<td>A1</td>
<td>f-clut</td>
<td>44.73</td>
<td>1.5</td>
</tr>
<tr>
<td>B4</td>
<td>VS-50</td>
<td>18.21</td>
<td>1</td>
<td>A3</td>
<td>nf-clut</td>
<td>7.59</td>
<td>1.0</td>
</tr>
<tr>
<td>B6</td>
<td>VS-50</td>
<td>18.21</td>
<td>1.875</td>
<td>A5</td>
<td>nf-clut</td>
<td>12.92</td>
<td>3</td>
</tr>
<tr>
<td>B8</td>
<td>TS-50</td>
<td>4.408</td>
<td>surf-flush</td>
<td>A7</td>
<td>f-clut</td>
<td>306.0</td>
<td>2.7</td>
</tr>
<tr>
<td>B10</td>
<td>TS-50</td>
<td>4.408</td>
<td>1</td>
<td>A9</td>
<td>f-clut</td>
<td>0.43</td>
<td>1.4</td>
</tr>
<tr>
<td>B12</td>
<td>TS-50</td>
<td>4.408</td>
<td>2</td>
<td>A11</td>
<td>f-clut</td>
<td>89.77</td>
<td>1.9</td>
</tr>
<tr>
<td>B14</td>
<td>M-14</td>
<td>0.6</td>
<td>surf-flush</td>
<td>A13</td>
<td>f-clut</td>
<td>16.36</td>
<td>2.0</td>
</tr>
<tr>
<td>B16</td>
<td>M-14</td>
<td>0.6</td>
<td>1.06</td>
<td>A15</td>
<td>f-clut</td>
<td>5.32</td>
<td>2.0</td>
</tr>
<tr>
<td>B18</td>
<td>M-14</td>
<td>0.6</td>
<td>1.875</td>
<td>A17</td>
<td>f-clut</td>
<td>4.45</td>
<td>3.0</td>
</tr>
<tr>
<td>B20</td>
<td>PMA-3</td>
<td>0.35</td>
<td>surf-flush</td>
<td>A19</td>
<td>f-clut</td>
<td>0.12</td>
<td>1.5</td>
</tr>
<tr>
<td>B22</td>
<td>PMA-3</td>
<td>0.84</td>
<td>1.75</td>
<td>A21</td>
<td>f-clut</td>
<td>0.16</td>
<td>2.0</td>
</tr>
<tr>
<td>B25</td>
<td>VAL69</td>
<td>2800</td>
<td>surf-flush</td>
<td>A23</td>
<td>f-clut</td>
<td>2.2</td>
<td>1.3</td>
</tr>
<tr>
<td>D2</td>
<td>VS-2.2</td>
<td>3.29</td>
<td>1.438</td>
<td>A25</td>
<td>f-clut</td>
<td>0.48</td>
<td>1.1</td>
</tr>
<tr>
<td>D4</td>
<td>VS-2.2</td>
<td>3.29</td>
<td>3.25</td>
<td>E1</td>
<td>wood</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>D6</td>
<td>M-19</td>
<td>0.94</td>
<td>1.75</td>
<td>E4</td>
<td>VS-50</td>
<td>18.21</td>
<td>3</td>
</tr>
<tr>
<td>D8</td>
<td>M-19</td>
<td>0.94</td>
<td>2.5</td>
<td>E5</td>
<td>filled hole</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>D10</td>
<td>TMA-4</td>
<td>0.75</td>
<td>1.75</td>
<td>E6</td>
<td>M-14</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>D12</td>
<td>TMA-4</td>
<td>0.75</td>
<td>3.25</td>
<td>E7</td>
<td>stone</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>D14</td>
<td>TM62P3</td>
<td>1.375</td>
<td>E8</td>
<td>VS-50</td>
<td>18.21</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>D16</td>
<td>TM62P3</td>
<td>3.0</td>
<td>E9</td>
<td>stone</td>
<td>0.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>D18</td>
<td>T-72</td>
<td>1.25</td>
<td>E10</td>
<td>M-14</td>
<td>0.6</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>D25</td>
<td>TM-46</td>
<td>2400</td>
<td>3.0</td>
<td>E11</td>
<td>wood</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>f-clut</td>
<td>1.06</td>
<td>1</td>
<td>E12</td>
<td>Sim6</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>C3</td>
<td>f-clut</td>
<td>.85</td>
<td>1.5</td>
<td>E13</td>
<td>plastic</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>C5</td>
<td>nf-clut</td>
<td>.43</td>
<td>1.8</td>
<td>E15</td>
<td>plastic</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>C7</td>
<td>f-clut</td>
<td>.66</td>
<td>1.5</td>
<td>E17</td>
<td>filled hole</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>C9</td>
<td>nf-clut</td>
<td>3.43</td>
<td>0.8</td>
<td>E19</td>
<td>stone</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>C11</td>
<td>f-clut</td>
<td>4.67</td>
<td>1.6</td>
<td>E21</td>
<td>stone</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>C13</td>
<td>f-clut</td>
<td>2.35</td>
<td>1.1</td>
<td>E23</td>
<td>wood</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E25</td>
<td>wood</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 7.1: Identities of targets in Fort A. P. Hill Site 71A calibration lane. In the target column "f-clut" and "nf-clut" stand for ferrous clutter and non-ferrous clutter.
a fixed distance from specific grid locations. The camera’s field of view (FOV) center was aimed a fixed distance from the camera. Plan view of this measurement configuration is shown in Figure 7.3. The camera height and all dimensions were controlled, so that the geometry was known. The FOV of the MWIR camera permitted two grid cells to be imaged simultaneously in this manner. Three grid cells were imaged for each LWIR placement.

Previous work with the IR camera suggested that better data could be acquired after dark. Because of restrictions on the use of the Fort A. P. Hill site, almost all IR measurements had to be done during daylight hours and weather conditions that are not suited for detecting thermal signatures. The only attempt made to collect
IR data at night was hampered by rain, which can adversely affect the temperature differences of the ground surface.

### 7.2 EMI Signal Processing and Feature Extraction

As described in Chapter 6, a two-step sensor signal processing procedure was used for EMI data. The first step converts the family of time-domain waveforms acquired at each scan position into a single spatial signature. The second step extracts a small number of features that concisely describe the high-dimensional spatial signature.

The spatial signatures obtained by applying the first processing step to several calibration lane targets are shown in Figure 7.4. Figure 7.4(a) is the response of a VS-50 anti-personnel mine buried flush to the ground, and Figure 7.4(b) is the response of the same mine type buried at a 3 inch depth. Because the VS-50 mine contains about 18 g of metal, it has a strong response. The response of a surface-flush PMA-3 mine, shown in Figure 7.4(c), is quite small even in the expanded scale used, because it contains only about 0.35 g of metal. Figure 7.4(d) shows the spatial response of a
Figure 7.4: Some example EMI spatial signatures obtained from A. P. Hill calibration lane data.
306 g ferrous clutter object. Figure 7.4(e) corresponds to a TM-46 metal-cased, anti-tank mine buried at a 3 inch depth. Because the spatial extent of the EMI signature of this mine is large, the 24 inch long scan used in the measurement was insufficient to capture the complete spatial profile. Figure 7.4(f) is the response of a Valmara-69 mine buried flush to surface. Both the TM-46 and Valmara-69 have flat regions in the response, which are caused by saturation of the sensor. We next present some observations and results on feature extraction from EMI spatial response data. The feature extraction methods have been described in detail in Chapter 4. In this section we present the details that arise in applying these methods to EMI sensor data and results illustrating their performance.

7.2.1 Feature Analysis
Using Physical Models and Moments

Our initial analysis of this data set used physical models and statistical moments, both of which were found to be ineffective. In contrast to the uni-modal spatial responses obtained for the targets considered in Case Study I, we observe that several targets at Fort A. P. Hill (including the VS-50 mine) have bimodal spatial responses. Significant depth dependence is also observed in the spatial profiles. These observations have important implications for the model-based EMI feature extraction. The simple empirical model used in Case Study I cannot handle bimodal responses and does not accommodate depth dependent changes in the responses. Therefore, the previous empirical model is inadequate and a more sophisticated model is required.

Das et al. [60] developed a physics-based analytical model that describes the spatial response of a square loop scanned over a tilted metallic prolate spheroid. That work gives examples showing the variation of the spatial responses with the
depth, orientation, and metal type of the targets, which indicate that such a model can produce responses similar to those observed in the A. P. Hill data. Because the Schiebel EMI sensor uses a circular loop rather than a square loop, the model in [60] was modified. The spatial distribution of metal in landmines can be complicated, and a tilted spheroid is typically not a good description of the distribution. Nevertheless, because this model can produce spatial profiles similar to those measured, an attempt was made to extract features from the measured data.

Consider an EMI sensor with concentric transmit and receive coils of radii $a_T$ and $a_R$ respectively located in a horizontal plane. Let a metallic prolate spheroid of major and minor axis lengths $a$ and $b$ respectively and orientation $(\phi_0, \theta_0)$ be centered at $(x_0, y_0, z_0)$. If the target spheroid’s major axis direction is $\hat{a}$, the voltage induced in the receive coil is given by

$$E = \mu_0 ab^2 \mathbf{H}^R \cdot \mathbf{M}$$  \hspace{1cm} (7.1)

where

$$\mathbf{M} = (\hat{a} \cdot \mathbf{H}^T)(\beta_L - \beta_T)\hat{a}.$$  \hspace{1cm} (7.2)

The induced voltage $E$ is shown to depend on the object depth $z_0$ and on the horizontal offset $\sqrt{x_0^2 + y_0^2}$. The lengthy expressions that describe these relations are available in [60] and will not be repeated here. Since only a single linear scan was made across the center of each target, it is reasonable to fix $y_0 = 0$ and to vary only $x_0$ to produce the desired spatial responses.

Both conventional nonlinear optimization and genetic algorithms were used to extract features by fitting this model to the measured data, but it was not possible to obtain unique, meaningful parameter estimates. We believe the main reason for
this failure is that the metallic composition of the measured targets is more complex than that assumed in the model. Because different mines have quite different metal compositions, a simple fixed parametric model is not a useful basis for feature extraction. Mines may contain multiple metallic components (detonator tubes, springs, balls, etc.) at different positions and orientations, and developing a general model applicable to all mines is difficult if not impossible.

**Amplitude Statistics as Features**

In its normal mode of operation, the operator of an EMI sensor determines the presence and the location of a target based on variations in the amplitude of the audio tone as the sensor sweeps the ground. We noted in Chapter 4 that amplitude statistics can be used to characterize these variations. The amplitude of the EMI signal is important in metal detection, and it is pertinent to investigate the discriminating power of the amplitude statistics of EMI data.

The distribution of the peak value of both mine and clutter objects in the calibration lane is shown in Figure 7.5(a). The data indicate little class separability, which is confirmed in the ROC curve obtained using a simple threshold classifier on this data (Figure 7.5(b)). Mean, variance, skewness, and kurtosis also demonstrated no discriminatory power.

To gain further insight to the problem we considered classifying the data into metal and nonmetal classes rather than into mine and clutter. The distribution of the peak values and the classifier ROC curve are shown in Figure 7.6.

Although the EMI sensor is a metal detector, the signature peak value does not provide satisfactory discrimination between metal and nonmetal objects. This poor performance occurs because some of the "metal" objects contain extremely small
Figure 7.5: Classification performance of the peak value of an EMI spatial response.

(a) Feature distribution 
(b) The ROC curve

Figure 7.6: Classification performance of the peak of an EMI spatial response for separating metal objects from nonmetal.

(a) Feature distribution 
(b) The ROC curve
Figure 7.7: Classification performance of the peak of an EMI spatial response for separating large metal objects (> 5 g) from other objects.

amounts of metal (less than 1 g) which are almost undetectable. The problem is exacerbated when the sensor-to-target separation is large, such as for deep targets. To confirm this reasoning the peak amplitude feature was again used to classify the data into large metallic objects (containing more than 5 g of metal), and objects containing less than 5 g of metal. The distribution of the feature values and the threshold classifier ROC curve in Figure 7.7 now indicate very good performance.

The preceding sequence of results highlight the difficulty of mine detection with EMI sensors, and point out the inadequacy of reliance on signal amplitude for target discrimination. Some promising mine detection results have been reported by using a singularity expansion of the decay curves [64] and a frequency domain characterization [65], but those results were based on measurements done using different EMI sensors. The Schiebel sensor did not permit the use of these techniques.
7.2.2 Representation Features

The difficulties encountered in extracting physical model parameters as features, and the ineffectiveness of amplitude statistics prompted us to investigate the alternative feature extraction techniques described in Chapter 4. The physical and statistical modeling approaches discussed above attempt to represent each spatial response using a few model parameters. A counterpart to these approaches is the use of representation features, in which the original data is mapped into another domain where most of the signal energy becomes concentrated in a few coefficients.

Subsampling of Spatial Responses

Recall from Chapter 4 that many statistical feature extraction methods require computation of the within-class covariance matrices. The covariance matrices computed from our limited data set are noninvertible, because each EMI spatial response has a dimension of 88, but there are only 28 mine samples and 27 clutter samples in the data set. To mitigate this problem we reduced the dimensionality of the raw data by subsampling. The original spatial response and the subsampled versions obtained by retaining every 4th and every 8th sample are shown in Figure 7.8. subsampling by a factor of eight causes excessive distortion, while the response subsampled by a factor of four is not significantly different from the raw data. The 1:4 subsampled data has a dimensionality of only 22, making the covariance matrices invertible.
Karhunen-Loéve (KL) Transform Features

The eigenvalues of the mixture covariance matrix of subsampled EMI spatial responses arranged by decreasing magnitude, are shown in Figure 7.9(a). Let $\mathbf{A}$ be a matrix whose columns are the eigenvectors corresponding to the first $n$ dominant eigenvalues. An arbitrary spatial response $\mathbf{x}$ can be projected onto the $n$ dominant eigenvector directions using $\mathbf{y} = \mathbf{A}^T \mathbf{x}$, and the $n \times 1$ vector $\mathbf{y}$ is defined as the KL-coefficient feature vector representing $\mathbf{x}$. The representation accuracy of these features can be determined using the inverse KL-transformation $\mathbf{x} = \mathbf{A} \mathbf{y}$. Figure 7.9(b) shows the original spatial response and the approximations corresponding to $n = 2, 4, \text{and } 6$ for a VS-50 mine (E8). The figure illustrates that six KL-coefficients are sufficient to represent the original 22-dimensional response.
Figure 7.9: The representation and the classification of EMI spatial responses based on KL-coefficients.
For comparison with one-dimensional features such as the peak, we show in Figure 7.9(c) the ROC curve generated by applying the first (dominant) KL-coefficient to a threshold classifier. The first KL-coefficient is the value obtained when the data are projected onto the eigenvector corresponding to the dominant eigenvalue. To improve the poor performance obtained for one coefficient, additional terms were considered. Because Figure 7.9(b) shows that at least six KL-coefficients are needed for good representation, ROC curves were also obtained using the first six KL-coefficients in a nearest-neighbor classifier and a distance classifier. These ROC curves, shown in Figure 7.9(d), also indicate poor classification.

**Discrete Cosine Transform (DCT) Features**

The coefficients obtained by applying the discrete cosine transform to the response of a VS-50 mine (E8) are shown in Figure 7.10(a). Because only about the first 10 (out of a total of 22) coefficients are significant, the original response can be represented with very low distortion by retaining only these significant coefficients and forcing the others to zero. The responses reproduced by an inverse DCT of 8 and 9 nonzero DCT coefficients are compared to the original response in Figure 7.10(b), from which it can be seen that at least 9 DCT coefficients are required for accurate representation.

The ROC curve obtained by applying only the first DCT coefficient to a threshold classifier is shown in Figure 7.10(c). Figure 7.10(d) shows the ROC curve obtained by applying the nine largest DCT coefficients to nearest-neighbor and distance classifiers. Like KL-transform features, DCT features also demonstrate poor discriminatory power.
Figure 7.10: The representation and the classification of EMI spatial responses using DCT coefficients.
Wavelet Features

Wavelets are known to be effective for detecting local phenomena and for data compression or analysis and, hence, wavelet analysis was explored for feature extraction. A level three wavelet packet decomposition of spatial EMI responses was performed using a biorthogonal wavelet of order four for both decomposition and reconstruction. The complete 88-point response was used, and nine wavelet coefficients around the center of the waveform were retained as the feature values. The wavelet coefficients obtained for a VS-50 mine at location E8 are shown in Figure 7.11(a). The ROC curves obtained for the classification based on seven, eight, and nine wavelet coefficients using a nearest-neighbor classifier and a nearest mean distance classifier are shown in Figure 7.11(b) and Figure 7.11(c), respectively. The best classification performance is observed with eight coefficients in these figures. This figure also shows the nonmonotonic behavior discussed in Section 4.3. It is also interesting to note how the different classifiers lead to significant differences in the ROC curves obtained for identical feature sets. As mentioned in Section 5.4, these ROC curves measure the effectiveness of the classifier-feature combination, not the inherent effectiveness of the features themselves.

7.2.3 Classification Features

Extraction of classification features begins by splitting the available training data into classes for clutter (class 1) and mines (class 2). These data are used to compute class mean vectors $\mathbf{M}_1$ and $\mathbf{M}_2$, class covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, within-class covariance $\mathbf{S}_w$, between-class covariance $\mathbf{S}_b$, mixture mean $\mathbf{M}_0$, and mixture-scatter $\mathbf{S}_m$, according to the expressions in Chapter 4. Feature extraction is performed
(a) The wavelet-coefficient values for a VS-50 mine (E8)

(b) ROC curves obtained using different numbers of wavelet-coefficients in a nearest-neighbor classifier

(c) ROC curves obtained using different numbers of wavelet-coefficients in a distance classifier

Figure 7.11: The wavelet representation of the EMI spatial responses.
by projecting the data onto discriminant directions specified by various optimality criteria.

**Fisher Discriminant Feature**

When the training data are projected onto the Fisher discriminant direction defined by (4.30), the projected values are distributed as shown in Figure 7.12(a). In this figure the horizontal axis corresponds to the sample number in each class, and the vertical axis represents the Fisher feature value. Not only does the original 22-dimensional data (88 points decimated by 4) now reduce to a single dimension, but significant separability between mine and clutter classes is apparent.

This separation can be quantified by computing the classification errors and generating the ROC curve for the Fisher linear classifier. Figure 7.12(b) shows the number of false alarms, the number of missed mines, and the total number of errors as a function of the decision threshold. The optimum decision threshold of the classifier can be determined from this plot to satisfy an appropriate criterion such as minimum classification error or any other cost-weighted combination of errors. The ROC curve in Figure 7.12(c) shows that the Fisher discriminant feature is much more effective than the representation features considered earlier.

**Fukunaga-Koontz (FK) Transform Features**

To extract the Fukunaga-Koontz features, a transform that whitens the within-class covariance matrix must be found. After whitening, the covariance matrices of the transformed data share the same eigenvectors, but have eigenvalues $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$ that lie between 0 and 1 and sum to unity. The directions corresponding to the largest covariance difference are identified by ranking the eigenvectors according
Figure 7.12: Fisher discriminant feature extraction from the EMI spatial responses.
to $|\lambda_1^{(1)} - 0.5|$ as recommended in [52]. The eigenvalues and ranking obtained are shown in Figures 7.13(a) and 7.13(b), respectively. The eigenvectors corresponding to the six largest ranking values were chosen to form the FK-transformation matrix, resulting in six-dimensional feature data. The EMI data projected onto the optimum FK-direction, shown in Figure 7.13(c), demonstrates a very limited ability to classify the data. Only those targets that produce large variances in data are easily separable.

The ROC curve for this classifier is shown in Figure 7.13(d). The initial rise in the ROC curve is due the correct detection of the few strong targets, which have well separated data values. Because the variances of the other targets are nearly identical for both classes, the remainder of the ROC curve shows poor performance.

**Extended Fisher Criterion (EFC) Feature**

The feature values obtained by projecting the mine and clutter data onto the optimal discriminant direction defined by the extended Fisher criterion are shown in Figure 7.14(a). Significant class separability is apparent. The ROC curve obtained for this feature using a threshold classifier is given in Figure 7.14(b).

Although we expected the EFC-based classification to perform better than the conventional Fisher discriminant classification, inspection of the ROC curves in Figures 7.12(c) and 7.14(b) shows the opposite. According to [50], the EFC-method must perform at least as well as the Fisher approach for normally distributed data. We suspect that the poor performance of the EFC-feature based classifier is due to outliers in the data, which make the distributions nonnormal. The few large metal targets in the training data tend to bias the class covariance matrices. Because the EFC-technique uses a covariance dependent term not present in the standard Fisher technique, the optimal EFC discriminant direction tends to be affected by bias in the
Figure 7.13: Fukunaga-Koontz feature extraction from the EMI spatial responses.
covariance estimate. Specifically, EFC attempts to account for the strong influence of a few large targets. This process causes a larger number of weaker targets to be misclassified.

Bhattacharyya Distance Based Features

Features extracted by optimizing the covariance difference term ($\mu_2$) in the Bhattacharyya distance have the spread shown in Figure 7.15(a). Figure 7.15(b) shows the ROC curve generated using a threshold classifier on these features. The performance is similar to that obtained with the FK method. This finding is reasonable, because both methods rely on covariance information.
Sammon's Optimal Plane

Sammon's optimal discriminant plane method projects high-dimensional observed data onto a plane defined by Fisher's discriminant vector and another vector orthogonal to it. The second vector is obtained by projecting the original data onto the plane orthogonal to the Fisher vector, and then finding another Fisher vector in this orthogonal plane. Sammon's optimal plane performs a many-to-two feature mapping, which provides two benefits. First, it enables visualization of the structure of the original high-dimensional data. Second, it provides an additional feature, which becomes an extra degree of freedom in the classifier design. As a result, rather than using a simple threshold classifier, a more powerful classifier can be used to reduce the classification error.
The measured data projected onto Sammon's optimal plane are shown in Figure 7.16(a). The classification boundary obtained using a backpropagation-trained two-layer feed-forward neural network with only three hidden nodes is shown in Figure 7.16(b). Another neural network consisting of ten hidden nodes produced the classification boundary in Figure 7.16(c). Although the latter boundary provides perfect classification of the training data, its intricate shape is unlikely to provide good generalization and must be avoided. This example also serves to illustrate the over-fitting problem discussed in Chapter 5. ROC curves obtained using nearest-neighbor and distance classifiers are shown in Figure 7.16(d).

**Fisher with Fukunaga-Koontz (FFK) Features**

When the FFK approach of Section 4.2 is applied to EMI spatial data, the two-dimensional feature plane in Figure 7.17(a) results. The \( d_1 \) direction in this plane is the Fisher discriminant direction, while \( d_2 \) is the FK-optimal direction of the data projected onto a plane orthogonal to the Fisher vector. Because the FK-method makes the most important eigenvalues of one class the least important for the other, we see in Figure 7.17(a) that values corresponding to mines are concentrated near the origin, while those of clutter are spread out in the vertical direction. Like Sammon's optimal plane, the FFK-feature plane also indicate good separation between mine and clutter classes. ROC curves obtained using nearest-neighbor and distance classifiers are shown in Figure 7.17(b).
Figure 7.16: Feature extraction based on Sammon’s optimal plane mapping. In (b) and (c) the dark areas represent the decision region for clutter and the light areas represent the decision region for mines.
Figure 7.17: Feature extraction based on Fisher with Fukunaga-Koontz method.

Summary of EMI Features

Results of several feature extraction methods applied to EMI data were presented in this section. Extracting physical model parameters as features was difficult. Amplitude statistics were convenient to obtain, but lacked discriminating power. Representation features based on KL, DCT, and wavelet transforms were also found ineffective for discriminating mines from clutter. Fisher, EFC, Sammon’s optimal plane, and the FFK features produced good classification. Two-dimensional features based on Sammon’s plane and FFK plane have slightly better discriminating power than the Fisher feature alone. These results suggest that classification features are more effective than representation features for mine detection.
7.3 GPR Sensor Signal Processing and Feature Extraction

This section describes GPR signal processing and feature extraction. GPR clutter reduction is discussed in detail. The clutter reduction algorithm used in Case Study I was found to be inadequate for processing the A. P. Hill data, and a new clutter reduction algorithm was developed. This new clutter reduction algorithm is an important contribution of this research.

Subsection 7.3.1 describes the new clutter reduction algorithm. Subsection 7.3.2 discusses feature extraction using sub-images of data displayed as depth-position maps. Subsection 7.3.3 considers feature extraction from single waveforms acquired directly over the targets.

7.3.1 New GPR Clutter Reduction Technique

The procedures for calibration of the measured data and for producing GPR maps (i.e., a display of reflection coefficient as a function of time (vertically) and antenna position (horizontally)) were described in Chapter 6. We noted previously that for downward looking antennas and near-surface targets (e.g., anti-personnel mines) the ground reflection is the dominant source of clutter. When the antenna is scanned across the surface, this reflection produces a strong band in the image which obscures shallowly buried targets. Imperfections in the system impulse response (e.g., antenna ringing and cable reflections) can produce similar bands. If the GPR system consists of separate transmit and receive antennas, coupling between them also contributes to clutter. Removing these clutter contributions is an essential preprocessing step in obtaining useful GPR data. Applying conventional imaging algorithms to this data
does not significantly reduce this clutter, since refocusing the ground reflection does not reject this form of clutter.

Several GPR clutter reduction techniques, including time gating, ensemble average subtraction [66], early-time subtraction [67], and frequency-domain Prony estimation of the reflection [62] are discussed in the GPR literature, but none of these methods perform satisfactorily when applied to GPR data collected over shallowly buried mines in nonuniform ground. We will present a brief overview of these existing techniques and discuss their limitations. Then a new more effective clutter removal technique will be described. Results obtained by applying the proposed technique to data from Fort A. P. Hill are presented.

The concept of time gating is well known in radar, and it needs no explanation here. Time gating has been used to remove the ground reflection from GPR measurements of deep targets such as buried pipes, unexploded ordnances and tunnels, but time gating is impractical when the targets are very shallow.

For ensemble average subtraction one averages the time-domain signatures acquired as the antenna is scanned over the (mostly target-free) ground and subtracts it from the data. This approach is reasonable only if the ground is flat and uniform over the antenna path. Another disadvantage of ensemble average subtraction is that it tends to remove parts of the target if the ensemble average includes the target response.

In the early-time subtraction technique [67] the band-limited specular reflection is synthesized in the time-domain using a sinc approximation and subtracted from the data. The early-time subtraction technique was reformulated as a frequency-domain method [62] to improve its performance. In that method frequency-domain GPR
measurements are represented by a parametric model consisting of a sum of damped exponential terms. The unknown parameters are estimated using a total-least-squares (TLS) Prony technique.

Van der Merwe improved on the method in [62] by introducing an iterative estimation approach [21], which considers GPR measurements collected at individual spatial locations and removes all components that are different from a reference target signature. If the measured target signature matches the reference signature, the algorithm will successfully remove clutter components both in early-time and late-time. The need for a priori knowledge of the target signature is unattractive but not unrealistic, since there usually exist a priori knowledge regarding the mine types used in a given region. Successful clutter reduction with this algorithm requires the use of a data base of reference signatures of all possible mines, since even when the GPR data contain a real mine signature, this algorithm will remove it as clutter unless that signature is used as a reference. Further, Van der Merwe has shown that even when the reference signature corresponding to the measured mine is utilized, the algorithm can inadvertently remove it if the measured signature is sufficiently distorted due to tilting of the mine, interactions with the ground interface, or differences in the ground properties.

Brunzell [1] proposed a method, in which the time delay of each GPR scan is estimated by using a nonlinear least square error optimization. Then, by stacking a sequence of system impulse responses with the estimated delays, he synthesized a ground reflection, which was subtracted from the measured data. This method works well, even with GPR data taken over soil with an uneven profile, if the magnitude of the ground reflection is uniform across the GPR map. Factors such as variations
in soil moisture content often produce variations in the magnitude of the ground reflection over real ground and, hence, Brunzell's method is unsatisfactory.

Because of limitations in the existing clutter reduction techniques, an approach was developed in which the spatial variations of both the ground reflection magnitude and delay are simultaneously estimated using nonlinear optimization. Because these variations are relatively smooth functions of position, we modeled them using low-order Chebyshev polynomials. Low-order polynomials cannot model the relatively rapid spatial variations due to buried anti-personnel mines and, hence, the estimated ground reflection profiles are relatively unaffected by the targets. When the estimated ground reflection is subtracted from the measured data, it removes the ground clutter leaving the target signatures largely intact. While this new approach removes the ground reflection clutter, which is the dominant source of clutter in typical GPR data, it does not reduce late time clutter contributions due to antenna ringing, cable mismatches, etc. These clutter contributions, which are spatially invariant, are removed using a generalized singular value decomposition (GSVD) based subspace decomposition technique described later.

**Estimating the Ground Reflection**

To illustrate the procedure, consider the GPR map generated from data collected over a one inch deep TS-50 mine at Fort A. P. Hill is shown in Figure 7.18(a). The reflection from the air-ground interface near -0.5 ns dominates the data, while the characteristic hyperbolic arc due to the mine is barely visible. The smooth spatial variations in the ground reflection magnitude and the time delay observed in Figure 7.18(a) are typical of GPR data collected over real ground. If we split the ground
Figure 7.18: Raw GPR map generated from data collected over a one inch deep TS-50 mine (B10) at Fort A. P. Hill.
reflection profile into $N_{\text{seg}}$ spatial segments as indicated in Figure 7.18(b), it is possible to approximate the spatial variation of both the amplitude and the time delay of the ground reflection within each segment by weighted sums of low-order polynomials. The smaller the segment length the better is the model fit, but if the segment width is comparable to the dimension of a mine, then a shallow mine in the measurement will bias the model. Because we want to model only the ground clutter, the segment width should be restricted to dimensions larger than a mine. To minimize redundancy in the estimate it is preferable to use orthogonal polynomials. In this work Chebyshev polynomials are used. Because Chebyshev polynomials are orthogonal within the interval [-1, 1], the horizontal axis of each segment is mapped to [-1, 1].

Let $A_i(x)$ denote the spatially varying ground reflection amplitude and $B_i(x)$ the time delay of the ground reflection peak over the $i$th segment. We approximate $A_i(x)$ and $B_i(x)$ by the following summations

$$A_i(x) = \sum_{n=0}^{4} a_{in} T_n(x)$$

$$B_i(x) = \sum_{n=0}^{4} b_{in} T_n(x)$$

where $T_n(x)$ are the Chebyshev polynomials defined via the recursion relation

$$T_0(x) = 1, \quad T_1(x) = x, \quad \text{and} \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n \geq 1$$

Because the reflection coefficient at the air-ground interface can be complex, the magnitude factor $A_i(x)$ may take complex values and, hence, the coefficients $a_{in}$ in (7.3) are actually defined as

$$a_{in} = a_{in}^{\text{real}} + j a_{in}^{\text{imag}}$$
The coefficients
\[ a_i = [a_{i0}^{\text{real}}, a_{i0}^{\text{imag}}, a_{i1}^{\text{real}}, a_{i1}^{\text{imag}}, \ldots] \]  
\[ b_i = [b_{i0}, b_{i1}, \ldots] \]  
are the unknown real parameters to be estimated.

We estimate the frequency-domain ground reflection in spatial segment \( i \) as
\[ G_i^{\text{est}}(x, f) = A_i(x)H(f) \exp(-j2\pi f B_i(x)) \]  
where \( H(f) \) is the windowed frequency response of the radar, i.e.,
\[ H(f) = H_{\text{Avg}}(f)H_w(f) \]
where \( H_{\text{Avg}}(f) \) the average frequency spectrum of raw GPR data within the segment and \( H_w(f) \) is a window function applied in forming the time-domain data. In our work a Hanning window was used. The time-domain ground bounce estimate \( g_i^{\text{est}}(x, t) \) becomes
\[ g_i^{\text{est}}(x, t, a_i, b_i) = A_i(x)h(t - B_i(x)) \]
where \( h(t) \) is the inverse Fourier transform of \( H(f) \). The measured ground reflection over the \( i \)th segment can then be expressed as
\[ g_i^{\text{meas}}(x, t) = g_i^{\text{est}}(x, t, a_i, b_i) + n_i(x, t) \]
where \( n_i(x, t) \) represents the modeling error.

If the modeling errors are assumed to be normally distributed, then the maximum likelihood estimates of \( a_i \) and \( b_i \) are found by a least square process. Thus, the model parameter values \([\hat{a}_i, \hat{b}_i]\) for the best fit are given by
\[ [\hat{a}_i, \hat{b}_i] = \arg\min_{a_i, b_i} \|g_i^{\text{meas}}(x, t) - g_i^{\text{est}}(x, t, a_i, b_i)\|^2 \]
Figure 7.19: Linear weighting functions used for blending $A_i(x)$ and $B_i(x)$ functions.

Once the weighting coefficients $a_i$ and $b_i$ are estimated, the amplitude term $A_i(x)$ and the delay term $B_i(x)$ of the $i$th segment are evaluated using equation (7.3).

To reduce discontinuities in the estimated reflection across the GPR map, adjacent segments are selected to overlap each other as shown in Figure 7.18(b). The final amplitude function $A(x)$ across the full GPR map is computed by blending the $A_i(x)$ functions obtained for all $N_{seg}$ segments using a set of linear weighting functions shown in Figure 7.19. The final delay function $B(x)$ is similarly computed by blending the individual delay functions $B_i(x)$.

In our GPR data processing the above parameter estimation was performed using the nonlinear least square error minimization function $fminu$ in the Matlab Optimization Toolbox. We observed that the optimization routine could not converge to the desired optimum solutions for several data sets. Performance was improved by using a recursive approach in which the approximation was initially set to zero order. The parameters $a_{i0}^{real}$, $a_{i0}^{imag}$ and $b_{i0}$ were estimated for this zero order case and then,
these estimates were used as the initial guesses for $a_{i0}^{\text{real}}$, $a_{i0}^{\text{imag}}$, and $b_{i0}$ in a first order approximation where zero initial values were used for the remaining parameters, $a_{11}^{\text{real}}$, $a_{11}^{\text{imag}}$, and $b_{11}$. This procedure was repeated until the parameters at the desired order are estimated.

Because our goal is the accurate estimation of the ground bounce term, the above procedure was applied only to a vertical region of the GPR map that contained the ground bounce. To find the data block of interest, the ensemble sum of energy in the GPR map was computed by summing the energy in the horizontal direction. Data lying within time indices that encompass 99% of the energy were retained and used for the estimation process.

**Subspace Decomposition Technique for Removing Stationary Clutter**

The polynomial-model based clutter reduction algorithm described above removes most of the ground reflection clutter, but this algorithm does not eliminate other clutter, which appears at later times (greater depths). These late-time clutter components due to antenna ringing, cable mismatches, etc. usually appear as uniform horizontal bands across the GPR map. Although in principle, these clutter sources can be eliminated, by proper system design and calibration, residual amounts are usually present in practice. GPR data acquired at Fort A. P. Hill also showed weak clutter bands, which we believe to be due to shaking and twisting of connecting cables due to wind during calibration and measurements. Even though these clutter contributions are small compared to the ground reflection, they are nevertheless a problem in mine detection, because they may mask weak returns from deeper mines.
To remove these clutter bands, which are nearly spatially independent, we adopt a subspace decomposition technique proposed by Marinovich [68] for detecting nonstationary signals in stationary noise. We present the important steps in that formulation below for completeness.

Let $W_N (N \times M)$, $M \leq N$ be a stationary noise (clutter) matrix generated by stacking in the horizontal direction $M$ replicas of a GPR impulse response waveform. $W_N$ consists of one or more uniform horizontal bands that mimic stationary clutter. A signal matrix $W_S (N \times M)$ is formed by the GPR response when an antenna scans over a buried point scatterer. The matrix $W_S$ has the familiar hyperbolic arc. Because the noise is assumed stationary, $W_N$ has unit rank, while $W_S$ has rank greater than unity. Now consider the generalized singular value decomposition (GSVD) of the signal with respect to the noise. In general, the GSVD yields a matrix representation of the form

$$W_S = U_S D_S X^{-1}$$

$$W_N = U_N D_N X^{-1}$$

The matrices $D_S$ and $D_N$ are diagonal with the elements

$$D_S = \text{diag}(\alpha_i), \quad 0 \leq \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_M$$

$$D_N = \text{diag}(\beta_i), \quad \beta_1 \geq \ldots \geq \beta_q \geq \beta_{q+1} = \ldots = \beta_M = 0$$

where $q = \text{rank}(W_N)$. The columns of $X$, denoted by $x^{(i)}$, are the generalized singular vectors of $W_N$ and $W_S$ satisfying

$$\beta_i^2 W_S^T W_S x^{(i)} = \alpha_i^2 W_N^T W_N x^{(i)}$$

147
In this application \( \text{rank}(W_N) = q = 1 \), and only \( \beta_1 \neq 0 \). Now isolate the noise subspace in \( X \), \( U_S \), \( U_N \) and \( D_S \) as follows

\[
X = \begin{bmatrix} x^{(1)} & X_1 \end{bmatrix}, \quad U_S = \begin{bmatrix} u_S^{(1)} & U_{S_1} \end{bmatrix}, \quad U_N = \begin{bmatrix} u_N^{(1)} & U_{N_1} \end{bmatrix}
\]

(7.19)

and

\[
D_S = \begin{bmatrix} \alpha_1 & 0 \\ 0 & D_1 \end{bmatrix}, \quad D_N = \begin{bmatrix} \beta_1 & 0 \\ 0 & 0 \end{bmatrix}
\]

(7.20)

Post-multiplying (7.14) and (7.15) by \( X \) and using the expressions in (7.19) and (7.20) we obtain

\[
W_S X = U_S D_S = \begin{bmatrix} \alpha_1 u_S^{(1)} & U_{S_1} D_1 \end{bmatrix}
\]

(7.21)

\[
W_N X = U_N D_N = \begin{bmatrix} \beta_1 u_N^{(1)} & 0 \end{bmatrix}
\]

(7.22)

which can be expanded to produce

\[
W_S x^{(1)} = \alpha_1 u_S^{(1)}
\]

(7.23)

\[
W_S X_1 = U_{S_1} D_1
\]

(7.24)

\[
W_N x^{(1)} = \beta_1 u_N^{(1)}
\]

(7.25)

\[
W_N X_1 = O
\]

(7.26)

According to eqns. (7.23)-(7.26) all of the stationary noise energy together with a small fraction of signal energy lie along the first generalized singular vector \( x^{(1)} \). There is no noise energy in the subspace determined by the remaining generalized singular vectors (i.e. the columns of \( X_1 \)). Because \( \alpha_1 \) is small, \( W_S \) is mapped in (7.24) with minimum loss of signal power.

Because \( X^{-1} \) can be ill-conditioned, Marinovich suggests using a QR-decomposition

\[
(X^{-1})^T = QR
\]

(7.27)
with which (7.14) and (7.15) can be put in the form

\[ W_S Q = U_S D_S R^T \]  \\
\[ W_N Q = U_N D_N R^T \]  

(7.28)

(7.29)

If \( Q \) is expressed as \( Q = [q^{(1)} \ Q_1] \), then a projection matrix \( P = Q_1 Q_1^T \) can be defined such that

\[ W_S P \approx W_S \]  \\
\[ W_N P = 0 \]  

(7.30)

(7.31)

We evaluated the projection matrix \( P \) according to the above procedure and applied it to the clutter reduced data. This projection is a useful method for enhancing the detection of small anti-personnel mines, which usually have weak, localized signatures that are easily masked by clutter. The projection reduced the stationary clutter contributions appearing in the form of constant horizontal bands, but did not significantly affect the signatures of small mines, which appear as localized hyperbolic arcs. Mine signatures with relatively slow spatial variation (for example, signatures of large anti-tank mines such as the TM-46) are affected by this nonstationary clutter removal step. Because large anti-tank mines have strong returns, they can often be detected easily and do not require this processing step.

**Example Clutter Reduction Results**

When the foregoing clutter-reduction process is applied to the measured GPR data of Figure 7.18(a), corresponding to a one inch deep TS-50 mine (B10), it produces the estimated ground reflection shown in Figure 7.20. Subtracting this estimate from the raw data gives the map in Figure 7.21. The ground reflection term, which was present
Figure 7.20: The ground reflection clutter profile estimated by applying the polynomial model-based approach to measured GPR data.

Figure 7.21: The GPR map obtained by removing the estimated ground-reflection from the measured data. This data corresponds to a one inch deep TS-50 mine at location B10 of the Fort A. P. Hill calibration lane.
around -0.5 ns in the raw data, is no longer visible in the clutter removed data. The most important observation to make from Figure 7.21 is how the signature of this very shallow mine is preserved and enhanced, something that is difficult to achieve (for such shallow targets) with the existing clutter removal techniques. In Figure 7.21 we observe a weak band of clutter between -1.5 ns and -1.0 ns. Next, we apply the subspace-decomposition technique described above to the map in Figure 7.21, which produces the map shown in Figure 7.22. The clutter band noted above is diminished but not eliminated, because this clutter is not entirely spatially invariant as assumed in the formulation.

Next, we compare the clutter reduction technique and ensemble average subtraction for a VS-50 mine buried at a three inch depth at location E4 of the Fort A. P.
Hill calibration lane. Figure 7.23(a) is the raw data map, and the mine signature is barely visible. Figure 7.23(b) is the map obtained by applying ensemble average subtraction. Because the ground reflection profile is nonuniform, the clutter reduction is unsatisfactory in this case. When the raw GPR data of Figure 7.23(a) is processed by applying the new clutter reduction technique, the GPR map in Figure 7.23(c) is obtained. The improvement in signal to clutter ratio with this approach is clear.

As a final example of the new clutter reduction procedure we present the result of processing data collected over an M-14 anti-personnel mine buried flush with the surface at location E6. This small \((56\text{ mm diameter})\) plastic anti-personnel mine is a stressing target for GPR. Both because it has a very weak GPR response and because when buried flush to the surface, it is overwhelmed by the strong ground reflection. Figure 7.24(a) is the raw data, which shows no obvious indication of a target. Figure 7.24(b) shows the result of applying the new clutter reduction procedure. A very weak target signature appears at position \(0.075\text{ m}\) near the surface. Effective clutter reduction can significantly improve target detection based on correlating a characteristic hyperbolic signature with the GPR maps, which we demonstrate for the case of the M-14 mine. If the signature is correlated directly with the raw data, the ground reflection produces very strong values as shown in Figure 7.25(a), and the mine is essentially undetectable. If the signature is correlated with the clutter reduced data shown in Figure 7.24(b), it produces the correlated image in Figure 7.25(b), which shows a high intensity at the target location.
Figure 7.23: Comparison of raw and clutter reduced maps generated from data collected over a three inch deep VS-50 mine (E4) at the Fort A. P. Hill calibration lane.
Figure 7.24: The GPR maps generated from data collected over a surface-flush M-14 plastic anti-personnel mine (E6) at Fort A. P. Hill.
Figure 7.25: The correlation maps generated from data collected over a surface-flush M-14 plastic anti-personnel mine (E6) at Fort A. P. Hill.
7.3.2 Feature Extraction for GPR Sub-Images

A number of features are extracted from the clutter-reduced GPR maps. As done for the EMI sensor, the effectiveness of these features was investigated by generating ROC curves. A data set comprising 32 clutter samples and 27 mine samples collected at the Fort A. P. Hill calibration lane were used in this study.\(^3\)

We noted above that GPR data acquired over targets contain characteristic hyperbolic signatures having an apex at the target location. Several examples of such signatures were shown previously. GPR maps of larger targets such as anti-tank mines may contain multiple hyperbolic signatures, which appear in the example map shown in Figure 7.26 corresponding to a 1.75 inch deep M-19 mine. If the GPR data contain sufficient information to identify a target, most of that information will be present within a region centered on the hyperbola such as that identified by the dashed rectangle in Figure 7.26. This region captures most of the spatial and temporal target information. Although an experienced human observer might develop the ability to visually identify targets directly from such GPR sub-images, designing an automatic classifier for this task is difficult, because a large training set is needed to deal with the high-dimensionality of the data.

We investigated several approaches to extracting features from GPR sub-images. Because several statistically-based classification features were quite effective for the EMI sensor, we sought to extract similar features from the GPR data. Due to the high-dimensionality of the GPR sub-images, however, the number of training samples in hand was insufficient to compute invertible within-class covariance matrices, which \(^3\)A slightly smaller data set was used for EMI feature extraction because of computer failures during the EMI data acquisition.
are a prerequisite for those extraction techniques. We considered two alternative approaches. One alternative was to abandon classification feature extraction from the GPR sub-image data and to apply amplitude statistics and representation feature extraction techniques, which require no covariance matrix. The other alternative was to extract features from single look-angle data (i.e. GPR waveforms measured at a single antenna position) collected directly over the targets. Identification based on single look-angle data has been attempted by other researchers in the past. For example, Van der Merwe [21] correlated reference waveforms of known mines with single look-angle measured data. Nag et al. [24, 61] derived a ramp response feature for target identification from single look-angle data. The use of single look-angle
data, however, ignores the spatial variations of the signature, which contain additional information.

**Correlation Coefficient Feature**

As in Case Study 1, the maximum correlation coefficient produced in correlating a synthetic GPR signature with clutter-removed GPR maps was used as a feature. In Case Study I the synthetic signature was correlated with a three-dimensional volume of data to locate the targets. Because only a single scan was acquired across each target at Fort A. P. Hill, the signature was correlated with the two-dimensional data measured over each target. Correlation coefficient was found to be ineffective in discriminating mines from clutter, which is illustrated clearly by the feature values, shown in Figure 7.27(a) and the virtually diagonal ROC curve shown in Figure 7.27(b). Because the hyperbolic arc response used for the correlation is not a
Figure 7.28: The GPR sub-image containing the signature of a one inch deep TS-50 mine at grid location B10, and its discrete cosine transform (DCT).

unique characteristic of mines, but a common characteristic of any small target that is GPR detectable, this poor discrimination must be expected. This feature may, however, be useful in detecting mines buried in homogeneous ground, for example in desert sand, which is relatively free from other GPR detectable objects such as rocks, roots, and voids. It is also a good method for cueing sub-image subtraction.

**Discrete Cosine Transform (DCT) Features**

The DCT can produce very high data compression without covariance matrix computation, and it is attractive for feature extraction from GPR sub-images. Consider the approximately 50 × 60 pixel sub-image shown in Figure 7.28(a), which contains the signature of a one inch deep TS-50 mine. This sub-image was selected from
Figure 7.29: Representing the GPR sub-image of Figure 7.28(a) using the ten largest DCT coefficients.

clutter-reduced data corresponding to grid location B10 shown previously in Figure 7.22. The absolute value of the two-dimensional DCT of Figure 7.28(a) is shown in Figure 7.28(b). It is clear from this figure that most of the energy is concentrated in a few coefficients around the lower left. If we retain only the ten largest DCT coefficients indicated by the white blobs in Figure 7.29(b) and set all other coefficients to zero, the inverse DCT results in the image of Figure 7.29(a). Comparison of the original and the compressed images show that ten DCT coefficients are not adequate for representing the image. We find that the 40 largest DCT coefficients can provide a reasonably good representation, which is illustrated in Figure 7.30(a) and Figure 7.30(b). Because the significant coefficients occur at arbitrary indices (frequencies), the feature vector must contain the indices of these coefficients as well as
Figure 7.30: Representing the GPR sub-image of Figure 7.28(a) using the 40 largest DCT coefficients.

their values, thus doubling the effective feature dimension. Although a data compression (dimensionality reduction) ratio of approximately $50 \times 60/80 = 37.5$ is achieved, the DCT features are still too numerous to be useful for classification, because the training data set is small.

**Amplitude Statistics**

As noted previously, moments of the amplitude distribution reduce high-dimensional sub-images to a single dimension, thereby overcoming the training set size problem. The distribution of the peak values of GPR sub-images corresponding to the calibration lane targets are shown in Figure 7.31(a). A threshold classifier was applied to the
Figure 7.31: The distribution of the peak values of GPR sub-images and the ROC curve for classification based on the peak values using a threshold classifier.

peak values to generate the ROC curve in Figure 7.31(b), which indicates some discriminatory power. Classifiers based on the variance produced a similar ROC curve, but mean, skewness, and kurtosis resulted in ROC curves that fall on the diagonal line indicating that these features are ineffective.

7.3.3 Feature Extraction from Single Look-Angle GPR Data

When extracting features from single look-angle data, waveforms acquired with the antenna positioned directly over the target are desirable. For those targets for which a hyperbolic arc signature was discernible, the time-domain waveform taken through the apex was selected. For other targets the waveform at the center of the scan was chosen.
Wavelet Coefficients

A wavelet packet decomposition of the absolute values of single look-angle GPR data was performed using a biorthogonal wavelet of order four for both decomposition and reconstruction. The approximation coefficients at level three were computed, and coefficients near the peak were retained as features. As an example, the absolute value of the waveform collected over a VS-50 mine at location E8 is shown in Figure 7.32(a) and the corresponding approximation is shown in Figure 7.32(b). The ROC curves shown in Figure 7.32(c) were obtained when different numbers of coefficients were used. A marginal amount of classification information is present in these coefficients.

Fisher Discriminant Feature

The Fisher discriminant feature was found to be quite effective for EMI data, and we applied the Fisher method to single look-angle GPR waveforms also. The plot of the absolute value of the single look-angle data for a VS-50 mine is shown in Figure 7.33(a). If we choose the portion of the waveform that encompasses the target signature, the dimensionality is still too high to form a covariance matrix. As done for the EMI features, the GPR waveforms were subsampled and twenty one samples covering the target signature were retained. The subsampled waveform and the retained samples for the VS-50 examples are shown in Figure 7.33(b). The distribution of the data projected onto the Fisher discriminant direction is shown in Figure 7.33(c), which indicate excellent class separation. Accordingly, the ROC curve generated by a threshold classifier, shown in Figure 7.33(d), indicates very good
(a) The absolute value of the GPR waveform acquired directly over a surface-flush VS-50 mine at grid location E8

(b) The wavelet-coefficient values

(c) ROC curves obtained using different numbers of wavelet-coefficients in a nearest-neighbor classifier

Figure 7.32: The wavelet representation of single look-angle GPR data acquired directly over a target.
Figure 7.33: The Fisher discriminant feature extracted from subsampled single look-angle GPR data.
performance. Just as we found for the EMI data, classification features are superior to physical features and representation features for GPR data.

7.4 IR Sensor Signal Processing and Feature Extraction

IR data processing and feature extraction were done by Mr. I. Kursat Sendur of the OSÜ ElectroScience Laboratory. The data processing and feature extraction procedure for Case Study II was identical to that described in Section 6.2.3 for Case Study I. Figure 7.34(a) shows the ROC curves obtained using a Euclidean distance classifier and a nearest-neighbor classifier on the feature set \([a, b, S, E]\) extracted from the IRRIS camera data collected at Fort A. P. Hill. The ROC curves corresponding to the same features extracted from the Agema data are shown in Figure 7.34(b). These ROC curves indicate that the IR features considered are not effective. Part of the poor performance of the IR features may be due the adverse conditions under which the IR data were collected, but more study of the problem is required.

7.5 Performance of Fusion Algorithms

Several forms of fusion were performed on the features described in Sections 7.2-7.4. As mentioned in Section 4.3, identifying the optimal feature subset is difficult when the number of features is large. The feature selection problem is further complicated by the small sample size and imperfections in the classifiers required to assess the feature performance. In this dissertation we do not attempt an exhaustive search for an optimal feature set. Rather, we take the simpler approach of eliminating features that are ineffective when used singly, and then consider fusion of various combinations of the effective features. As pointed out in Section 4.3 this approach is
Figure 7.34: The ROC curves obtained for classification using IR features.
not guaranteed to result in the optimal feature subset, but it is commonly used. A comparison of all possible combinations of features and classifiers is time consuming, and we consider only a few cases to gain some insight and to identify useful trends. All the results shown are obtained using the leave-one-out testing approach described in Section 5.4. For the small data set available here, neural network classifiers introduce additional complications including convergence to different local minima for different random initializations and over-fitting or under-fitting of training data if an improper number of training epochs are used. As we found in Case Study I, varying the neural network size and training under different initializations is also required to obtain good results for neural network classifiers, which is both inconvenient and time consuming for our investigation. For reasons of simplicity and time, attention in this study is restricted to distance classifiers and nearest neighbor classifiers.

7.5.1 Single Sensor Detection Using Multiple Features

In the preceding discussion we have primarily considered detection by each sensor using one feature. In this section we consider the use of multiple features from individual sensors. This could also be considered intra-sensor feature fusion. If the features carry useful, independent information about the phenomena of interest, then intra-sensor fusion is expected to improve performance. Several cases of intra-sensor feature fusion were considered for the Fort A. P. Hill data. The ROC curves obtained for some of these cases appear below.

In Section 7.3 it was found that the Fisher discriminant feature extracted from GPR data had the highest discriminant power. Figure 7.35 considers intra-sensor fusion of the Fisher feature with the peak value feature. Figure 7.35(a) shows the
Figure 7.35: The ROC curves obtained for GPR classification based on the Fisher and peak value features taken individually and fused.
ROC curves obtained using a nearest-neighbor classifier, while Figure 7.35(b) shows the ROC curves obtained using a distance classifier. These results show that intra-sensor fusion of an ineffective feature with a highly effective feature does not provide a significant improvement. In fact, Figure 7.35(a) shows that the fused ROC curve is slightly worse than that obtained with the Fisher feature alone. This performance degradation due to the addition of a feature seems counter-intuitive, but it can be caused by the finite number of samples and the use of nonideal classifiers. The hypothetical example in Figure 7.36 can be used to explain this behavior. In this example, feature 1 alone can lead to perfect classification based on either a nearest-neighbor or distance classifier. In the 2-D space, however, samples with widely scattered values for feature 2 (the less effective feature) may be located closer to the cluster of the wrong type. If the classifier is imperfect (as is often the case for small training data sets) this can lead to misclassification.

Figure 7.37 compares ROC curves obtained with (1) the one-dimensional Fisher feature, and with the two-dimensional features produced by (2) Sammon's method and (3) the Fisher with Fukunaga-Koontz technique. Recall that for both of the two-dimensional features, one feature vector is the Fisher vector. The second feature vector is a discriminant feature selected in a plane orthogonal to the Fisher vector. In effect, these two-dimensional feature vectors represent examples of intra-sensor fusion for the EMI sensor. The ROC curves in Figure 7.37 illustrate that both fused features perform better than the Fisher feature alone.
Figure 7.36: A hypothetical distribution of samples in a two-dimensional feature space.

Figure 7.37: The ROC curves obtained for EMI classification using a nearest-neighbor classifier and the Fisher discriminant feature, Sammon’s optimal features and Fisher with Fukunaga-Koontz features.
The ROC curves obtained for classification based on the individual IR features \( a, b, S \) and \( E \) extracted from the Agema sensor and the fused feature using a nearest-neighbor classifier are shown in Figure 7.38(a). Figure 7.38(b) shows the corresponding ROC curves obtained for the IRRIS sensor. These ROC curves indicate poor performance.

### 7.5.2 Examples of Multisensor Feature Fusion

Feature-level fusion of data from multiple sensors is performed in a manner identical to the intra-sensor fusion described above: the features are concatenated into a single feature vector, which is then classified. Accordingly, the major findings of this section mirror those of Section 7.5.1.

Fisher features extracted from EMI and GPR data are effective. Feature-level fusion of these two features improves the performance as illustrated by the ROC curves in Figure 7.39(a) generated using a distance classifier. ROC curves generated for the same features using a nearest-neighbor classifier are shown in Figure 7.39(b). The performance improvement is obvious in both cases. Next, consider fusion of an effective EMI feature, the Fisher feature, and an ineffective GPR feature vector comprising nine wavelet coefficients extracted from the GPR data. The ROC curves generated by a distance classifier for these features, shown in Figure 7.40(a), indicates a marginal performance improvement due to fusion. Classification using a nearest-neighbor classifier does not show an improvement as illustrated by the ROC curves in Figure 7.40(b). This result highlights the importance of effective features for achieving successful fusion, a finding that is also relevant to intra-sensor fusion.

172
Figure 7.38: The ROC curves obtained for classification based on the individual IR features, \(a, b, S\) and \(E\) and the fused features using a nearest-neighbor classifier.
Figure 7.39: The ROC curves for classification based on the EMI Fisher feature, the GPR Fisher feature and the fused feature vector.
Figure 7.40: The ROC curves for classification of the EMI Fisher feature, nine GPR wavelet coefficient features and the fused feature vector.
Because the IR features are ineffective, fusion of IR with EMI and GPR sensors did not produce a benefit. As an example, Figure 7.41(a) shows the ROC curves obtained for classification based on the GPR Fisher feature, the IRRIS sensor's $S$ feature and the fused feature vector using a nearest-neighbor classifier. Figure 7.41(b) shows the ROC curves for classification based on the EMI Fisher feature, the IRRIS sensor's $S$ feature and the fused feature vector using a nearest-neighbor classifier.

7.5.3 Comparison of Feature-Level and Decision-Level Fusion Performance

For normal statistics, classification features are closely related to the log-likelihood ratio and to a posteriori probabilities, which are in turn related to soft decisions. In this section we present numerical comparisons of feature-level fusion and decision-level fusion. In the latter, the features are explicitly mapped to decisions prior to fusion. We noted above that including the IR features caused the performance to degrade. In this study we consider the fusion of GPR and EMI data only. If improved IR features become available in the future, these studies should be repeated with the addition of the IR features. Figure 7.42 compares the classification performance of the Fisher discriminant features extracted from EMI and GPR sensors individually and under different forms of fusion. For decision-level fusion, binary decisions are generated from the individual sensor features by selecting suitable decision thresholds corresponding to the "knee" of the respective ROC curves. The sensor reliabilities $P_{F_i}$ and $P_{D_i}$ are defined by these selected thresholds. The prior probabilities of the clutter and mine classes, which are required for hard and soft decision-level fusion, were assigned the respective values 0.51 and 0.49 based on the population of the test site.
Figure 7.41: The ROC curves obtained for classification based on the EMI and GPR Fisher features, IRRIS sensor's S feature and the fused feature vector using a nearest-neighbor classifier.
Figure 7.42: The performance comparison of feature-level and decision-level fusion of GPR and EMI sensor data.
ROC curves computed for both the nearest-neighbor and distance classifiers show that feature-level fusion provides the best performance followed by soft and hard decision-level fusion. For the distance classifier, OR fusion leads to reasonably good performance, but OR fusion for the nearest-neighbor case has excessive false alarms. On the other hand, AND fusion has very few false alarms, but the probability of detection is unsatisfactory. Because only two sensors are fused, majority fusion was not considered in this study. As we saw above, the addition of other ineffective features extracted from the EMI and GPR data negatively impacts the fused results and, therefore, is to be avoided. The trends regarding the performance of the different fusion schemes in this Case Study agree well with those found earlier in Case Study I (Chapter 6).

7.6 Summary

Multisensor data were collected at the JUXOCO mine test lanes located at Fort A. P. Hill, Site 71A, using GPR, EMI, LWIR, and MWIR sensors. Data were acquired in a calibration lane where the identities of the buried targets are known and in a blind test lane where the target information is unknown. Only the calibration lane data were used in the feature extraction and fusion studies reported here, because ground truth information is required to quantify classification performance. The target set in the calibration lane comprised 32 clutter objects and 27 deactivated mines. Because effective features are essential for successful sensor fusion, much effort was devoted to developing good feature extraction techniques.

EMI sensor feature extraction was performed on spatial signatures derived from the time-domain sensor data acquired over each target. Various difficulties prevented
the extraction of physically based features. Amplitude statistics such as the peak, mean, and the variance were extracted, but lacked any discriminatory power. Representation features derived from the Karhunen-Loève, discrete cosine, and wavelet transforms also were ineffective for classification. Features based on statistical discriminants such as the Fisher discriminant and Sammon's optimal discriminant were found to be quite effective.

To process the GPR data it was necessary to develop an effective technique to remove the ground reflection clutter, which dominates GPR data from our sensor. Unlike many existing clutter reduction methods, which operate on individual waveforms, this new technique describes the spatial variation of the ground reflection by a polynomial model whose parameters are estimated using a least square technique. Subtracting the clutter estimate from the measured data significantly reduces the ground reflection. A subspace decomposition technique was then used to reduce other clutter contributions. Several examples were shown to demonstrate the power of these techniques.

Representation features and amplitude statistic features were extracted from GPR sub-images. Because of dimensionality issues, statistical discriminant features could not be extracted from the sub-images, but Fisher discriminant features extracted from individual waveforms collected directly over the targets were found to be extremely effective.

Physically based features extracted from IR data acquired by both MWIR and LWIR cameras turned out to be ineffective in discriminating mines from clutter. The unfavorable conditions under which the IR data were collected is suspected as the primary reason for the poor performance, but it is also possible that these physical
parameters are simply not effective. The success achieved by discriminant features for the EMI and GPR data, suggest that one should apply similar methods to the IR data, but this task was not attempted during the present study.

Examples of both multisensor fusion and single sensor detection (using one or more features) were presented. Nearest-neighbor and mean-distance classifiers were used to generate ROC curves for performance comparison. It was shown that the fusion of an ineffective feature with an effective feature is not beneficial and can even degrade the performance for small data sets. Feature-level fusion of the Fisher discriminant features extracted from EMI and GPR data was shown to be significantly better than the individual sensors. Soft and hard decision-level fusion also improved the performance but to a lesser extent than feature-level fusion. AND and OR rule based decision-level fusion were not found to be useful.
CHAPTER 8

SUMMARY AND CONCLUSIONS

8.1 Summary and Key Contributions

This dissertation describes research into multisensor fusion to enhance the detection of buried landmines. In this section we summarize the results of each chapter noting significant research contributions.

Chapter 2 described the landmine problem and mine detection sensor technologies.

Chapter 3 presented an overview of multisensor fusion and described a new, practical feature-level fusion algorithm that can accommodate noncoincidently sampled multisensor data. The relation of this algorithm to an optimal fusion algorithm was noted. Closed-form expressions for optimal fusion in the special case of normally distributed features were presented. Chapter 3 also extended a classical hard decision-level fusion rule to a more effective soft decision-level fusion rule.

Chapter 4 reviewed feature extraction techniques used in this research. Hypothetical data were used to illustrate the capabilities and limitations of some features. Feature subset selection was briefly discussed.
Chapter 5 reviewed pattern classification methods. Several parametric and non-parametric classifiers were described. Methods for classifier performance evaluation were discussed.

Chapter 6 described a case study involving multisensor data acquired at a surrogate minefield near the ElectroScience Laboratory using EMI, GPR, and IR sensors. Targets were buried in native soil with the objective of gaining insight into the problems faced in acquiring and processing sensor data under realistic conditions. Consequently, the amount of unknown clutter was very large and the soil surface was irregular. Sensors were used on different platforms and sampled the site at noncoincident sampling points. Simple empirical models were developed to describe the sensor data and to extract features. The fusion algorithms presented in Chapter 3 were applied to the data. The results demonstrate that feature-level fusion significantly improves detection performance. Soft decision-level fusion also improved performance to a lesser extent. The performance of optimal hard decision-level fusion was not significantly different from that of the best individual sensor. Fusion of hard decisions using AND, OR and majority rules was ineffective.

Chapter 7 described a case study based on multisensor data acquired at a mine test site in Fort A. P. Hill, Virginia. The target set at this site comprised real, deactivated mines and clutter objects. A new technique for reducing the dominant ground reflection in our GPR sensor data was developed during this study. Unlike previously applied clutter reduction methods, this new technique works well for small mines that are shallowly buried or even flush to surface, as long as the ground surface profile and the soil properties are reasonably smooth. It was also shown that a subspace decomposition technique can be used to remove some of the late time clutter.

183
due to antenna ringing and cable mismatches, which appear as bands across the GPR data.

Application of different fusion techniques to the Fort A. P. Hill data confirmed the findings of Chapters 6. Namely, feature-level fusion was significantly better than the individual sensors. Soft decision-level fusion also provided a large improvement, while the benefit of hard decision-level fusion was marginal. Fusion using AND and OR rules were again found ineffective.

The results presented in this dissertation consistently demonstrate that fusion can significantly improve mine detection. It was shown that feature-level fusion is both practical and effective for this application, even when the sensor data are noncommensurate or acquired with position errors. The results also suggest that, although decision-level fusion is more convenient, feature-level fusion is more effective. Based on our tests, the additional effort required to develop a feature-level fusion system is warranted.

Achieving good performance for both individual sensors and feature-fused sensors requires good features. Indeed, fusion of poor features can actually degrade system performance for small data sets. A significant effort was made during this work to identify effective features. Using the Fort A. P. Hill data we demonstrated that physical and representation features, which have been used in prior mine detection studies, are ineffective for discriminating mines from clutter. For both the EMI and GPR data we showed that statistical discriminant features are more promising. Only physical model parameter features were extracted from the IR data, and those proved to be ineffective.
The foregoing conclusions should be treated cautiously for several reasons. First, the data set used for the study was quite small. Small data set size leads to errors in class means and covariances. Because statistical feature extraction techniques rely on estimates of means and covariance, errors in those estimates affect feature extraction. Small training and testing sets also affect classifier boundaries and performance evaluation. Second, the test sites were not actual minefields, but prepared plots. This mismatch between test and true conditions violates a fundamental requirement in statistical pattern recognition that training set be representative of the real distribution. Third, the prior probabilities of the mine and clutter classes, which were estimated to be 0.49 and 0.51 for the Fort A. P. Hill calibration lane, do not represent those of a real minefield. For example, prior demining efforts have shown that for every real mine detected with a metal detector, several hundred pieces of metallic clutter are found. Therefore, the disparity in the class prior probabilities encountered in real mine fields and prepared test fields is evident. Finally, the types, numbers, and burial depths of different mines in the test grid are not necessarily representative of the global distributions. Therefore, the class means and covariances that would be estimated with the data from a real minefield could be quite different from those found in Case Study 2. The importance of acquiring valid training data, which are representative of the true distributions cannot be overemphasized.
8.2 Suggested Future Research

The results obtained in the present study suggest several desirable areas for future research.

Many feature extraction methods implicitly assume that each class has a unimodal distribution. Real data often have multimodal distributions, which could be accommodated if feature extraction and classification were preceded by cluster analysis. Due to the limited data set available, it was not meaningful to consider cluster analysis during the present work. If larger training data sets become available, future research should consider cluster analysis as a means to improve classifier performance.

In the present work the effectiveness of the individual features and selected feature combinations were judged from ROC curves computed using simple classifiers. Because only a few features were available, it was adequate to identify the effective features and to demonstrate the improvement due to fusion. Future research should consider more systematic approaches to optimal feature selection, such as genetic algorithm based selection.

The features used for the IR sensors must be improved. Discriminant features were found to be effective for EMI and GPR sensors, and these features should also be investigated for the IR data.

The sensor suite should be expanded. For example, other researchers have shown that the temporal or spectral information from more sophisticated EMI sensors can be useful for discriminating mines from clutter. NQR sensors and acoustic sensors are also becoming available, and their contribution to a fused sensor suite should be explored.
Studies of the statistical variability of data acquired over specific mine and clutter targets should also provide useful information and must be undertaken. For example, data can be acquired over a particular mine, for various values of depth, tilt, surrounding soil type, moisture, overburden, and clutter distribution in the vicinity. Such data can supplement the limited data set of Case Study 2. Because these measurements are meant to provide statistical scatter, none of the variables such as depth need to be controlled precisely and, hence, measurement is simple. In contrast, if the measurements are intended for developing or validating a physical model, they can be difficult, because precise control of variables is important.
APPENDIX A

SOFTWARE IMPLEMENTATION OF A FLEXIBLE FUSION ARCHITECTURE

A secondary objective of this dissertation was to develop and implement a modular, flexible sensor fusion architecture. It is also desirable for the architecture to be easily extensible, i.e., to accommodate new sensors, features, classification methods, and fusion techniques without having to modify the software. This appendix describes a sensor fusion architecture and a graphical user interface (GUI) with these desirable properties. The fusion architecture, which was implemented in Matlab, was used extensively to analyze the multi-sensor data collected at Fort A. P. Hill.

Central to this flexible architecture is a structure array that holds the features extracted from each sensor at all the interrogation points. If all sensors sample the ground at coincident sampling points, interrogation points may be identical to the sampling points. otherwise they may be arbitrarily selected. Storing features corresponding to interrogation points rather than sensor sampling points lets the architecture handle both coincidently and non-coincidently sampled multi-sensor data. Each element in this structure array comprises the fields

\[ \text{[position, sensor}_1, \ldots, \text{sensor}_N\text{s]} \]  

(A.1)
where $N_S$ is the total number of sensors. Each element $\text{sensor}_i$ is itself a structure having the fields $[\text{feature}_{i1}, \text{feature}_{i2}, \ldots, \text{feature}_{il_i}]$ where $l_i$ is the number of features extracted from the $i$th sensor's data.

When the GUI, shown in Figure A.1, is initialized, it loads the structure array and recognizes the available sensors, features, and the dimensionality of each feature. If a new feature extracted from the $i$th sensor's data needs to be made available for fusion, it is only necessary to add to structure $\text{sensor}_i$ a new field $\text{feature}_{il_i+1}$ containing the feature values. Similarly, adding a new sensor amounts to adding a new $\text{sensor}_{N_S+1}$ structure to the structure array. It is possible to have more than one data structure for a given data collection. For example, one structure might be used for training data and another for testing data. The DATA SET button allows the user to select the appropriate data structure for classification and fusion. If the selected data set has known ground truth information, the GUI generates ROC curves after each classification, otherwise it provides only the binary decisions at the interrogation points.

The user can dynamically select one or more sensors and features for fusion with the SENSOR and FEATURE buttons. Because some features, such as the KL-coefficients, are multi-dimensional, the GUI provides a DIMENSION button to specify the number of coefficients for the selected feature. Each time a feature is selected by the user, the name and the dimensionality of the feature and the corresponding sensor name are appended to a feature selection array. With the CLASSIFIER button, one of many classifiers including nearest neighbor, mean distance, and neural network, can be selected. When the CLASSIFY button is activated, all features specified in the feature selection array are fused and applied to the classifier. The classifier decisions
Figure A.1: The graphical user interface developed for sensor fusion studies.
are generated over a range of threshold values, so that ROC curves can be generated if the ground truth is known. Pressing the CLASSIFY INDIVIDUAL SENSORS button on the GUI causes the individual sensors in the feature selection array to be classified separately. These individual sensor decisions can then be used to perform various forms of decision-level fusion.

All classifier functions use a uniform naming convention, so that the GUI can recognize the available classifiers and include them automatically in the CLASSIFIER selection menu. Furthermore, each classifier function uses the same input-output interface. Adding a new classifier requires only that the new classification routine adhere to the standard interface and the same naming convention. The decision-fusion functions follow a similar standard. Modularity is achieved by defining each of the tasks (sensor and feature selection, classification, decision-level fusion, and ROC curve generation) as separate call-back modules.
BIBLIOGRAPHY


