Image Processing Methods Applied to Landmine Detection in Ground Penetrating Radar

by

Rayn Sakaguchi

Department of Electrical and Computer Engineering
Duke University

Date: _____

Approved:

_____
Leslie M. Collins, Supervisor

_____
Loren W. Nolte

_____
Guillermo Sapiro

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in the Department of Electrical and Computer Engineering
in the Graduate School of Duke University
2013

Abstract

Image Processing Methods Applied to Landmine Detection in Ground Penetrating Radar

by

Rayn Sakaguchi

Department of Electrical and Computer Engineering
Duke University

Date: _____

Approved:

_____
Leslie M. Collins, Supervisor

_____
Loren W. Nolte

_____
Guillermo Sapiro

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Electrical and Computer Engineering in the Graduate School of Duke University
2013

Abstract

Recent advances in statistically based ground penetrating radar (GPR) landmine detection have utilized 2-D slices of data to recognize the hyperbolic shapes caused by a sub-surface landmine. The objective in this research is to identify these shapes using methodology found in the field of image processing. Three different recognition methods were considered; (1) instance matching, which aims to recognize occurrences of a specific object; (2) object detection, which aims to find objects belonging to a class of objects; and (3) category recognition, which aims to categorize entire images based upon the contents of each image. This research consists of the adaptation and evaluation of these methods applied to GPR landmine detection. The results from this work illustrate the additional information and performance increases that these methods provide to the GPR detection system. In addition, this work shows promise for the application of additional methods from the image processing and computer vision fields.

Contents

Abstract **iv**

List of Tables **vii**

List of Figures **viii**

Acknowledgements **xv**

1 Introduction **1**

2 Background **6**

2.1 GPR System and Data 6

2.2 Current GPR Processing Methods 7

2.2.1 Pre-processing and Pre-screening Methods 8

2.2.2 Inversion Methods 12

2.2.3 Statistical Methods 13

2.3 Image Processing Methods 16

2.3.1 Instance Matching 17

2.3.2 Object Detection 25

2.3.3 Image Category Recognition 33

3 Instance Matching **37**

3.1 Dataset and Performance Metrics 37

3.1.1 Pre-screener Fusion 38

3.2 Instance Matching Applied to GPR Data 38

3.2.1 Feature Extraction . . . 39
3.2.2 Classification . . . 42
3.2.3 Image Matching Results . . . 43
3.3 Modified Instance Matching - Classification of Feature Points . . . 46
4 Object Detection **49**
4.1 Object Detection Methods . . . 49
4.1.1 Training Data . . . 50
4.1.2 Descriptor Representation . . . 52
4.1.3 Classifiers . . . 53
4.2 Object Detection Results . . . 53
4.2.1 Descriptor and Classification Results and Analysis . . . 53
4.2.2 Automated Training Data . . . 57
5 Image Category Recognition **60**
5.1 Category Recognition Methods . . . 60
5.1.1 Patch Representation . . . 62
5.1.2 Patch Localization . . . 62
5.2 Category Recognition Results . . . 63
5.2.1 Patch Size and Descriptors . . . 64
5.2.2 Patch Localization . . . 68
6 Conclusions **71**
Bibliography **75**

List of Tables

3.1 This table compares the top performers utilizing instance matching between B-scans and the classification of feature point regions. Large differences in PD exist when not fused with the pre-screener. However, when fused with the pre-screener, most of this performance gap is recovered by the instance matching method and in the case of $0.003FA/m^2$, surpassed 48

5.1 Probability of detection at each FAR operating point using different patch sizes and both whitened patch data and HOG features. All detection rates include fusion with the pre-screener confidences. . . . 65

6.1 The above results show the probability of detection at various false alarm rates. At each operating point, the highest probability of detection achieved is in bold. For all trials, methods presented in Chapter 4 had the highest probability of detection. 74

List of Figures

1.1 An example recorded reflected signal at one antenna. The initial high energy response due to the air/ground interface, as well as a later response due to a buried object are highlighted 4

1.2 An example 2-Dimensional slice of GPR data around a target known as a B-scan. B-scans can be either all sensors at one location (cross-track B-scan), or one sensor as the vehicle moves in the down-track direction. 4

2.1 The flow of the HMDS on-board processing architecture. The pre-screener identifies GPS locations to be further analyzed by the feature extractor and classifier. The feature extractor processes features from the GPR data around each pre-screener location. The classifier uses the features to make a determination if the location is a landmine or a pre-screener false alarm. 8

2.2 Figure 2.2(a) shows the original data as received from the HMDS hardware. Figure 2.2(b) shows the data after it has been aligned by the maximum energy peak. Figure 2.2(c) shows the data after the ground bounce has been removed and the data has been normalized to account for the low energy signal of deeper targets. 11

2.3 B-Scan of a target response and associated landmine HMM states. A confidence value for a given B-scan would be based on the probability that the HMM progressed through the states shown above in the order shown left to right. 14

2.4 Algorithm flow for instance matching between an example image shown in and a cluttered scene where the object is matched. 19

2.5 Figure 2.5(b) shows the corner response at each pixel given the original image in 2.5(a). Figure 2.5(c) shows the final selected points overlaid on the original image. 21

2.6 Figure 2.6(b) shows the response to the blob detector at each pixel given the original image shown in in 2.6(a). Figure 2.6(c) shows the final selected points overlaid on the original image after non-maximal supression. 21

2.7 Example generation of SIFT descriptors on a patch of GPR data. The patch of data taken from around a feature point location is shown in 2.7(a). The region split up into the 4x4 grid of sub-patches. Plot 2.7(b) shows a quiver plot representing the gradients at each pixel location. Plot 2.7(c) shows the result after binning the gradients within each sub-patch. Each arrow in this plot originating from the center of the sub-patch represents the magnitude of the bin in that angular direction 23

2.8 Example generation of a SURF descriptor on a patch of data. Plot 2.8(a) shows the patch of data around a feature point. The statistics of each sub-patch are shown in 2.8(b). The statistics of all 16 sub-patches will be concatenated to form a 64 dimensional feature vector. 24

2.9 The original patch of data to be processed is shown in 2.9(a). The same data is smoothed and shown in 2.9(b). A small number of comparison features are also displayed. The gray arrows represent location pairs where the first location has lower intensity than the second. The black arrows represent pairs where the first location has higher intensity. . . 25

2.10 Example flow of object detection method applied to face detection in images. Example training samples are generally obtained via hand labeling. The classifier is trained on the descriptor representation of all example images. Locations within new images are converted to their descriptor representation and classified using the trained classifier. 27

2.11 Patches #1 and 2 both exhibit the same geometric structure. However the gradient directions along each edge are reversed. The HOG descriptor will bin the gradients in the 90° and 270° in the same bin. The SIFT descriptor will bin the gradients in separate bins as the range for the sift angular bins goes from 0-360°. 28

2.12 Figure 2.12(a) shows the possible decision lines that classify the training data without error. Many of these lines however, are not desirable given new testing data. The solid line shown in Figure 2.12(b) is the desired robust decision boundary that is better suited to classify unseen data given these distributions 31

2.13 Processing steps for the Bag-of-Words model applied to visual images. The patch vocabulary is generated based off a random sample of patches within the training data. Each image is in turn represented by the frequencies of each patch type. These frequency histograms are then used to represent the image for both training and testing stages. 35

3.1 A comparison of all three feature point detectors applied to a landmine response. The Harris corner detector tends to locate dots on the outer edges of the high energy areas. The blob detector finds small high energy areas, or around the edges of larger high energy masses. The 2-D energy method focuses purely on areas of high energy regardless of the surrounding area. Many of the points from the 2-D energy method lie on the central A-scan as that is typically the highest energy column. 41

3.2 The above comparison shows the ROC curves obtained when utilizing only the minimum distance to targets, and summing the labels of the 15 closest matches across all positive and negative examples of training data. Utilizing the non-target areas helps to create low distance matches between non-target areas that results in an increase in performance. 42

3.3 The results above illustrate the performance of various descriptor methods over the best performing parameters utilized in instance matching. The results on the left show the performance of the descriptors without fusing with the pre-screener confidence. The results on the right include the pre-screener fusion that shows additional information is being introduced by the fusion with instance matching. These results show that without the pre-screener information, the SIFT descriptor method outperforms other descriptors. In contrast, BRIEF features, perform very badly without pre-screener information perform much better after fusion with the pre-screener. 44

3.4 The results above evaluate the performance of various point localization methods utilizing the best performing descriptors. The results on the left show the performance without fusion with the pre-screener confidence. The results on the right include the pre-screener fusion that shows additional information is being introduced by utilizing the instance matching confidences. These figures show the large disparity between performance with and without the fusion with the pre-screener, especially in relation to the BRIEF descriptors. 45

3.5 This figure illustrates matching between two different instances of the same landmine target type. The target on the right is the closest match over all training alarms available. The lines represent a match was found between the given SIFT descriptor whose patch center is represented by the white dots. 45

3.6 This figure shows the down-track matching between a target and its top match, a false alarm. The second matching image in 3.6(b) shows the feature point matching between the same target and an instance of the same landmine type at the same depth at a different location. The feature points are not located similarly and the overall responses do not share all of the same visual characteristics. Thus the two B-scans were not matched well during testing 46

3.7 These results show the best performance of the modified instance recognition method. The performance of the SIFT on the results not fused with the pre-screener remain high. The BRIEF descriptor also continues to do well only after pre-screener fusion is introduced. Both sets of results utilize 20×20 pixel patches around keypoints generated by the 2-D energy feature point locater. These results are compared numerically to using the initial method of instance matching in Table 3.1 48

4.1 The above figure shows a target at a pre-screener alarm by its B-scan as well as the A-scan at the pre-screener alarm location. There is an initial energy response that begins at about 100 time samples. A second, higher energy response also appears at about the 200th time sample. It is unclear whether one or both of these responses is more important to characterize. Choosing either of the responses could be overlooking important information characteristic of this landmine. Choosing both requires that a method be developed that determines what constitutes a secondary landmine response or a piece of clutter, such as a rock. 51

4.2 These results illustrate the performance across a number of patch sizes utilizing the SVM classifier. It was found that the larger patch sizes have better performance. This is likely due to the utilization of late time structure as seen in Figure 4.3. 54

4.3 This example illustrates the structure observed after the initial higher energy response from a buried landmine. The data shown in Figure 4.3(b) is scaled differently allowing the lower energy areas to become more pronounced. 54

4.4 These results illustrate the differences between the various descriptor methods. The best performing patch size is used in conjunction with the SVM classifier. The SIFT descriptor consistently performed the best among all methods tested despite the similarities to the HOG descriptor. The BRIEF descriptor performed poorly in both trials, especially when not fused with the pre-screener. 55

4.5 The results presented above show the performance of the SIFT descriptor as well as the modified HOG descriptor. The HOG descriptor was modified to have the same sub-patch size as the SIFT descriptor which results in an increase in performance. 56

4.6 This figure shows the comparison between different classifiers utilizing the modified HOG descriptor results as shown above. The SVM classifier performed the best on all trials. 57

4.7 The above response is caused by a landmine with a non-metal casing. The initial reflection due to the soil/casing interface is located around time sample 210, whereas the highest energy response is located around time sample 250. This is a case where the hand labeled row index would differ greatly from the max energy method 58

4.8 This figure shows the results of attempts to automatically generate isolated training examples from the training data. It can be observed that the hand labeled training data continues to perform better at certin operating points, most notably at 0.003 FA/m^2. In the fused results, the hand-labeled data would be the clear choice. However the method utilizing three different training examples per training image does have better performance at some operating points in the unfused results. It is important to note that this is not necessarily a fair comparison. By having 3 training samples per positive image, there are three times the number of positive training samples. 59

5.1 This figure illustrates three different methods of locating image regions on a GPR B-scan. Each white dot represents the center of a region extracted from around that pixel. Patch locations are constrained to remain near enough to the central A-scan so that the entire patch is localized within the B-scan. The leftmost image shows the dense sampling method where regions are extracted at regular intervals. The center image shows the random sampling method that extracts regions from uniformly random sampled locations across the image. The number of sampled regions using this method is the same as the number extracted using the dense method. The rightmost image shows the 2-D energy method that extracts regions from points that are local maxima of the 2-D energy map of the B-scan. This method generally produces substantially fewer features points than the other two methods. 63

5.2 This figure shows the performance of the descriptors over a range of square patch sizes independent of the pre-screener confidence. Performance is measured by the probability of detection at a given false alarm rate (FAR). The darker bars represent the normalized patch's performance as a descriptor. The normalized patch performance was found to be more consistent across various patch sizes than the modified HOG method. 64

5.3 B-scans of the same target visualizing the likelihood ratios of the centroids representing each region. Each dot's intensity is determined by the rank of the centroids likelihood ratio. These results were obtained using a 16×16 patch size, with a sampling rate of 4 in both dimensions. 66

5.4 Visualizing the likelihood ratios of centroids that represent regions of a false alarm B-scan. Note that the energy range in these images is much smaller than the images of a buried target. The HOG patches have a much higher response to these areas where the energy is low, but structure in the data remains. 66

5.5 The figures above illustrate the differences in centroids for different sized patches. The smaller patch centroids can only represent parts of the landmine response. The larger patch centroids are able to represent both the rising and falling edge of the landmine response allowing it to represent the entire landmine structure. 68

5.6 Probabilities of detection for the different methods of localization of regions within individual B-scans. Note that the number of random points chosen is the same as the number of points used in dense sampling. 69

5.7 B-scans showing the region centers when utilizing different region localization techniques. The intensity of each dot is determined by the rank of the centroid's likelihood ratio that represents the region around that dot. While the regions with high intensity dots are mostly consistent with the dense processing method shown in Figure 5.3, the density of center regions in both methods is variable and notably lower in the 2-D energy map method. 69

Acknowledgements

This thesis would not have been possible without the guidance and support of Dr. Leslie Collins and the SSPACISS research group. In particular, Dr. Peter Torrione and Dr. Kenneth Morton have had a large impact on the direction of this work as well as my progress as a researcher. I would also like to thank the US Army Night Vision and Electronic Sensors Directorate (NVESD) which generously funds this research and allows us to work on a program with real world implications. Finally I would like to thank my family and friends, both near and far, who have supported and encouraged me throughout this process.

1

Introduction

Around the world, both state and non-state entities continue to add to the devastating problem of victim operated buried explosives. These hidden killers indiscriminately cause harm to those who unknowingly encounter them. While these devices are generally placed intending to target combat forces, in 2011 an estimated 72% of the over 4,000 mine-related injuries or deaths were civilians, with 40% of those being children [1]. In addition to the loss of life and permanent injury, uncleared landmine areas can impede economic progress in areas already affected by conflict. Land not cleared of landmines cannot be farmed, or otherwise utilized. Roads contaminated with a single anti-vehicular landmine can restrict access to communities for years. Desperately needed humanitarian aid for these areas is often impeded by the suspicion of buried explosives on access roads. Often humanitarian agencies will not even consider aid to areas where cleared passage is not guaranteed [2].

While global efforts have succeeded in banning the use, manufacturing, stockpiling and transfer of anti-personnel landmines in most countries around the world, anti-vehicular mines are not currently restricted. In the year 2011, there were 663 confirmed injuries or deaths associated with anti-vehicular landmines; an 85% in-

crease over the average for the previous 4 years [1]. These types of threats are especially of interest in present times as they are the leading cause of injuries and deaths to the American forces currently in Afghanistan [3]. This research focuses specifically on improving a system that is designed for anti-vehicle landmines and Improvised Explosive Devices (IEDs) aimed at destroying vehicles. Removing these threats helps keep soldiers and civilians safe from unnecessary harm and assures that the roads are safe for civilian and humanitarian traffic.

The aim of this research is to develop algorithms that can improve the detection of buried explosive devices while reducing the number of false alarms. Along with ensuring that explosives are detected, the number of false alarms is especially important in this application. False alarms are costly as each location flagged as a potential alarm requires cautious excavation and neutralization – something ideally reserved only for dangerous buried explosives. Therefore the reduction of false alarms can significantly increase the rate of clearance and decrease costs associated with clearing suspected areas.

Traditionally landmine detection has been performed using an Electro Magnetic Induction (EMI) sensor to flag underground targets by inducing a current in the metal content of a buried threat. While this method is simple and detects many different landmine types, there are two significant issues which cannot be overcome by EMI sensors. The first problem is the manufacturing of low-metal landmines and the use of low metal components for IEDs. While most of these devices generally have a small amount of metal that can be detectable, the detector's threshold must be lowered so much such that very small metal clutter will cause a false alarm [4]. The second, closely related problem is metallic clutter – often in areas of conflict large pieces of shrapnel, casings, and other high metal waste is distributed throughout the region. For example, according to data from the Cambodian Mine Action Center, of the approximately 6 million pieces dug up during land clearing in 2011, only 104,464

were actual explosives [5]. Due to problems such as these, a significant research effort has been focused on methods to mitigate the high number of false alarms caused by EMI sensors [6, 7]. However, these methods are typically either designed for offline processing or utilize high-fidelity sensors that are not widely deployed.

Due to these challenges, ultra-wideband Ground Penetrating Radar (GPR) has been introduced as a complimentary technology to EMI sensors for landmine detection [8, 9, 10, 11]. GPR subsurface detection works by first projecting an electronic radar pulse into the ground. As the pulse travels downwards through the ground, discontinuities in dielectric constants between materials cause a partial reflection of the pulse. The resulting reflected signal is collected near the transmitting antenna. An example of a received signal can be seen in Figure 1.1. Notable features in this example are the two peaks highlighted below. The first, largest, peak is caused by the air/ground interface which generally causes the strongest reflection as it is the first major reflection which reduces the energy of the pulse as it progresses downwards. The second reflection in this example is caused by a buried metallic target. The magnitude of a targets response is dependent on a multitude of factors including target depth, antenna height, soil dielectric constant, soil moisture content, homogeneity of soil, and the physical makeup of the target.

While characteristics of landmines can be seen in single A-scans [12], most methods make use of the spatial characteristics of A-scans near an object [11, 13, 14]. These spatial characteristics can be seen by viewing GPR data as a 2-D B-scan as seen in Figure1.2. The hyperbolic shape is generated by the changing time it takes for the reflected signal to reach the antennae as the sensor location moves with respect to the buried object. Most algorithms, with a few exceptions[12, 15] specifically extract features which are designed to respond to hyperbolic-like shapes seen in B-scans containing targets.

The goal of this research is to adapt and apply methods specifically developed for

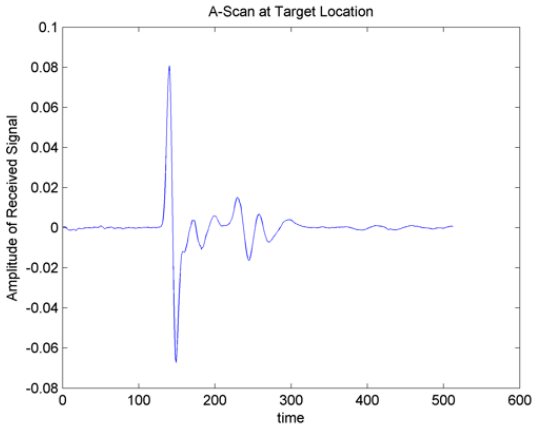


FIGURE 1.1: An example recorded reflected signal at one antenna. The initial high energy response due to the air/ground interface, as well as a later response due to a buried object are highlighted

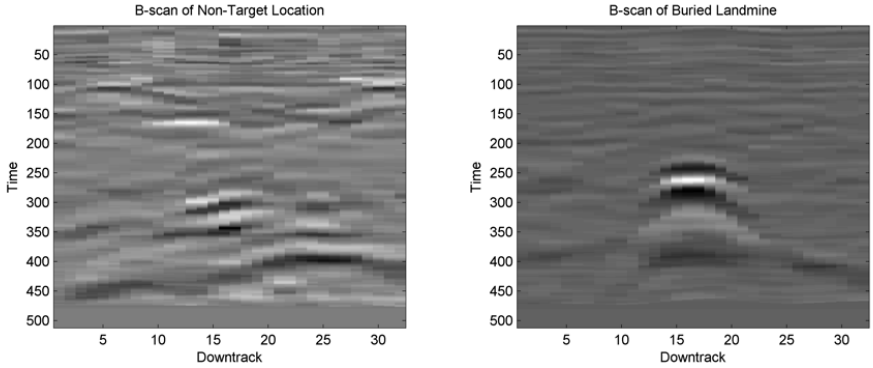


FIGURE 1.2: An example 2-Dimensional slice of GPR data around a target known as a B-scan. B-scans can be either all sensors at one location (cross-track B-scan), or one sensor as the vehicle moves in the down-track direction.

computer vision applications to GPR data to detect the shapes in B-scans caused by a landmines presence. Due to the strong variation in how landmine responses appear across various conditions [16], this research focuses on computer vision techniques which were designed to account for various types of uncertainty. Three different methods were studied during the course of this research. The first, Instance matching, is used in image processing to find instances of the same or identical objects within image scenes [17, 18]. This method has been shown to be highly invariant to a multitude of transformations including, rotation, intensity, scale, and partial oc-

clusions [19]. However, it is not designed to identify classes of objects, only identical instances. In contrast, the second method explored in this work, object detection, is specifically designed to find instances of a class of objects within images [20, 21]. Unlike instance matching, this method is not specifically designed to account for a large number of transformations. Instead, object matching relies heavily on a diverse training set and a more complex classifier. Unfortunately, obtaining a large training dataset of isolated examples can be a problem in GPR data since the exact location of a target response in GPR data as a function of time is unknown. There is currently no fast solution to determine exactly where within the image the actual landmine response is located. It is for this reason that the third method, category recognition was studied. Category recognition methods were developed to determine the class of an image by processing the whole image [22, 23]. This can vary from "What type of animal is present in this image" to "Is this image from a city, or suburban landscape." This method evaluates all parts of the image, and attempts to label the image based upon frequencies of certain types of features.

This research has found that these methods used for various types of image processing are applicable and beneficial for GPR threat detection. It has been found that shapes within the 2-dimensional B-scans can be recognized as landmine responses using these algorithms and can outperform methods currently used on fielded systems. Chapter 2 will go more into depth on the current methods used for landmine detection as well as the image processing methods that will be used. Chapters 3-5 describe the work necessary to tailor these methods to GPR data and provide results showing the effectiveness of each method. Chapter 6 will discuss the implications of these results and what new avenues of research are possible given these results.

2

Background

This section will provide background on the data and system involved in landmine detection using the NIITEK GPR sensor as well as an overview of the current algorithms implemented in fielded systems. Current image processing methods will also be introduced with special attention to the methods used in this research.

2.1 GPR System and Data

The data used for this research was collected with a vehicle-mounted array of GPR sensors manufactured by NIITEK known as the Husky Mounted Detection System (HMDS). The sensing system consists of a 51-channel array of sensors mounted perpendicularly to the vehicle's direction of travel. Channels are set to have approximately 6cm spacing between them. Sensors take readings of the subsurface at 5cm intervals as the vehicle moves forward. The direction of vehicle travel is referred to as the down-track direction [24]. The reflected signal at each physical location is recorded temporally for a total of 512 time samples. The resulting data available for processing at each down-track location is 512 time samples by 51 channels. Given these parameters, the amount of data collected grows quickly; the total amount of

data collected from 50 meters of travel results in a matrix of numerical doubles of size 512x51x1000. Since the detection system is mounted on a moving vehicle, landmines need to be detected quickly as there is very limited time between when the panel passes over an object and when the wheels could potentially trigger the object to detonate. Near real time processing of the data is therefore a necessity in order to keep the operator safe. Thus, in order to be able to process more complex features, the amount of data sent to higher level processing algorithms needs to be reduced. This constraint is addressed through the use of a pre-screening algorithm that detects anomalies in subsurface energy content [25]. Locations of interest on the ground surface are identified by the pre-screening algorithm; the data associated with the location of interest is then processed by a feature processing algorithm. Locations not flagged by this pre-screening algorithm are not reviewed by any other processing methods. Therefore, the pre-screening algorithm is designed to operate with high sensitivity to ensure it achieves a high probability of detection while potentially allowing a large number of non-target areas to be flagged as well. This allows for the more computationally expensive feature processing methods to process smaller amounts of data which allows the system to operate within the time constraints. While there is still a time requirement for the algorithms to complete, the two step process allows more flexibility in the feature processing computational requirements. A diagram of the system processing flow is shown in Figure 2.1. Numerous methods have been proposed and evaluated for use within the framework shown in Figure 2.1. Section 2.2 will review current methods of GPR processing, several of which are currently employed on the HMDS on-board processing system.

2.2 Current GPR Processing Methods

This section reviews several methods that have been shown to be successful in processing GPR data for landmine detection. This also section introduces standard pre-

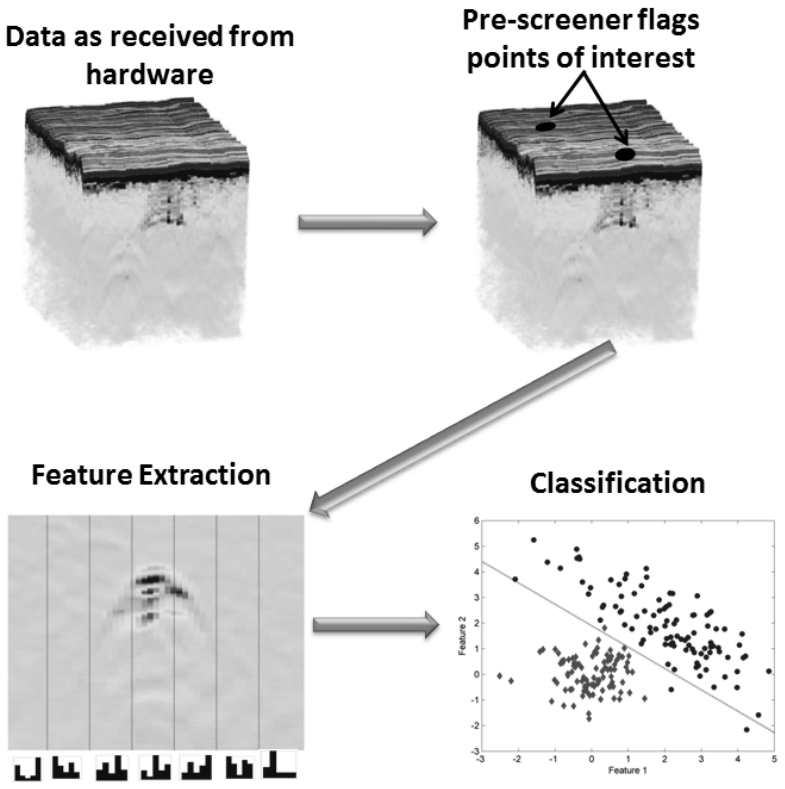


FIGURE 2.1: The flow of the HMDS on-board processing architecture. The pre-screener identifies GPS locations to be further analyzed by the feature extractor and classifier. The feature extractor processes features from the GPR data around each pre-screener location. The classifier uses the features to make a determination if the location is a landmine or a pre-screener false alarm.

processing and pre-screening methods which have been utilized in this research. In addition, a variety of current feature processing systems are introduced with emphasis on those that have shown good performance and subsequently been implemented on the HMDS system.

2.2.1 Pre-processing and Pre-screening Methods

Pre-processing Methods

Pre-processing the GPR data serves to reduce the effect of certain GPR characteristics that make it difficult to discriminate between targets and clutter. Three

major issues are introduced and addressed in this section. The first, and one of the most challenging features of GPR is the ground-bounce, or the initial spike in energy caused by reflected signal generated by the air ground interface. The energy of the ground-bounce is generally the highest energy peak in a GPR A-scan. As most methods of GPR processing respond strongly to areas of high energy and the ground contour can often appear hyperbolic, false alarms are commonly generated in locations where the ground-bounce is not removed properly. As a result, most GPR based systems employ some method to remove the ground-bounce from the data [26].

Another issue, one that affects the shape of a target in 2-D B-scans, is the potentially changing height of the antenna above the ground. If the antenna height changes as a sensor passes over an object, the number of time samples before the reflected signal is received at the antenna will change. Due to the GPR antenna being extended in front of a vehicle and this vehicle's use on often unpaved roads, the change in antenna height can be sudden and dramatic. This will cause variation in the time sample at which the landmine reflection is received. This leads to inconsistencies in the hyperbolic shape generated by a sub-surface object.

The last issue that pre-processing addresses is the attenuation caused by the propagation through the ground and the loss of energy due to reflections caused by heterogeneous soils [25] . This causes deep target responses to have much lower magnitude than targets found near the surface. In order to use features that are dependent on the energy of the landmine reflection at all depths, a landmine signal at a later time sample needs to be comparable in energy to one near the ground surface.

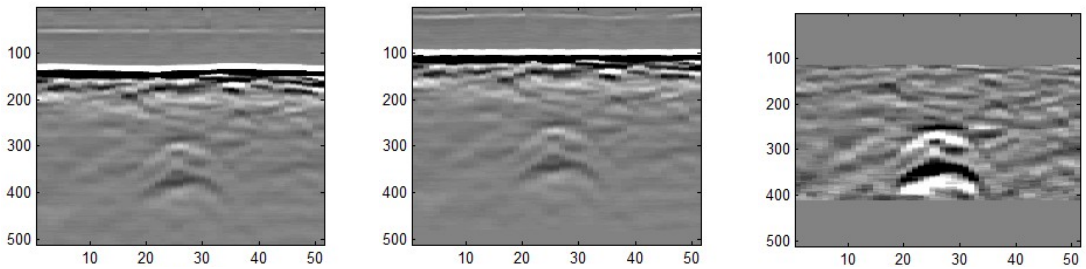
To remove the ground-bounce from a section of GPR data, it must be first identified. In downward looking GPR, it is usually a reasonable assumption that the time index associated with the highest energy within each A-scan is the ground-

bounce since it is the first major dielectric discontinuity encountered by the GPR signal. In ground-bounce removal, all A-scans are aligned so that the maximum amplitude index is on the same time sample. Then the all data above and including the ground-bounce is discarded. This process leaves the latest time samples of consecutive A-scans to be mis-aligned. The time samples where misalignments are present are not considered in processing and removed. The resulting trimmed data is then treated as the ground-bounce removed data.

The issue of changes in antenna height is also solved by aligning the data to the ground-bounce. When the antenna changes height, the signal propagation time between the antenna and the ground changes. However, the signal propagation time between the top of the ground and the subsurface is minimally affected by the changes in antenna height. Therefore, using the aligned data with the ground-bounce removed ensures the time indexes are in relation to the first contact with the air-ground interface. This time delay is going to be more consistent over a target as the distance between the air ground-interface and the object will not change noticeably due to antenna height.

To solve the issue with signal strength at later time samples, the data must be normalized so that targets have approximately the same characteristics. In the NITEK system this is done via a process termed guard-band whitening. Independently at each time sample below the ground-bounce, the background mean and standard deviation are determined by data surrounding the area being whitened. To normalize the data at each point, the mean is subtracted and then the sample is divided by the standard deviation. After normalization, the data's signal strength is more consistent throughout all depths as the data strength is based on its relation to the statistics at the same depth below the air-ground interface [27, 25].

The output of these pre-processing steps is data that has been: ground-bounce aligned; ground-bounce removed; and normalized by depth. An example of the pre-



(a) Original data (b) Data aligned to ground-bounce (c) Data haircut and normalized

FIGURE 2.2: Figure 2.2(a) shows the original data as received from the HMDS hardware. Figure 2.2(b) shows the data after it has been aligned by the maximum energy peak. Figure 2.2(c) shows the data after the ground bounce has been removed and the data has been normalized to account for the low energy signal of deeper targets.

processing progression is shown in Figure 2.2 where a deep target that is barely visible in the raw data has been pre-processed. Most pre-screeners and algorithms utilize these pre-processing steps so that features extracted can be easily compared across depths.

Pre-screener [25]

Due to the necessity of real time detection and the benefit of computationally expensive feature extraction techniques, a low complexity pre-screener is used to flag locations of anomalous responses and thus decrease the amount of data necessary for other algorithms to process. The pre-screener identifies anomalies and records the positions, which are termed pre-screener alarms. This pre-screener is designed to achieve a high probability of detection while allowing for a moderate false alarm rate in favor of simplified processing. The only sections of data that other algorithms will process are the ones located near pre-screener alarm locations. Using a pre-screener in this manner is similar to cascading classifiers [21] that aims to eliminate the maximum number of observations in each step using simple classifiers. This method is often used when processing time is an important consideration and data needs to be

eliminated from consideration at each step.

The pre-screener employed by the HMDS system is based on a least mean squares algorithm [25]. The algorithm operates independently at various signal response time bins and utilizes the aligned data with the ground-bounce removed. The LMS-based algorithm operates by linearly predicting the signal data at a given location and time using knowledge of previous data. The difference between the estimate and the actual data is the output of the LMS-based algorithm. Smoothing, processing by depth, and thresholding are also applied to the output of the LMS-based algorithm. Non-maximal suppression is then performed as a final step to identify the specific location of the anomaly. Further information about specific parameters can be found in [25] where this method is introduced.

2.2.2 Inversion Methods

While numerous approaches to generating and processing GPR data exist, one of the most potentially accurate methods of landmine detection is via electromagnetic inversion. By inverting the signal, it is possible to reconstruct the actual physical shape of the object and its dimensions under certain assumptions [28, 29, 30, 31]. In a landmine task, this would involve inverting the received data over time so that instead of a hyperbolic shape, the actual shape of the landmine could be imaged. Trials involving buried landmines have shown good performance under homogeneous soils [31]. These methods rely on models that assume a homogeneous medium; when faced with a medium that is impossible to predict, the resulting inverted image degrades substantially [31]. In addition, when the soil parameters are not known, these 3-D model inversions are also very computationally intensive. While there have been methods that aim to speed up the process for landmine detection [30, 28], these methods work to focus the image. These methods however, still make assumptions of soil homogeneity which degrades the landmine image rendering it impossible to

infer the dimensions of the object without statistical methods. Therefore, using inversion methods still requires using statistical methods to model the uncertainty in the signal after inversion. This further adds to the processing time necessary to classify a pre-screener alarm. It is for this reason that the most successful methods of fast GPR processing have been statistical methods designed to identify landmine responses in non-inverted data.

2.2.3 Statistical Methods

In order to develop methods that account for the variability and uncertainty commonly seen in landmine responses, researchers have developed many methods that rely on statistical modeling. These statistically based methods search for indications of landmines in the data without attempting to invert the signal. These methods are generally composed of two steps, feature extraction, and classification based upon the feature vectors. The following sections will describe a number of statistical methods developed for the HMDS system.

Hidden Markov Models [13]

Hidden Markov Models (HMMs) have shown to be a powerful tool when modeling non-stationary random processes in a variety of applications [32, 33, 34]. The HMM model assumes that data is generated according to the current state of the system, and by the distributions associated with that state. This state information can be exploited to determine a common sequence of hidden states that can signify an event. HMMs used in landmine data attempt to identify a sequence of hidden states that signify a hyperbolic landmine response as illustrated in target B-scans. The states associated with a landmine response are shown in Figure 2.3.

Transition probabilities between states as well as the distributions associated with each state are learned from example B-scans in the training dataset. In evaluating a

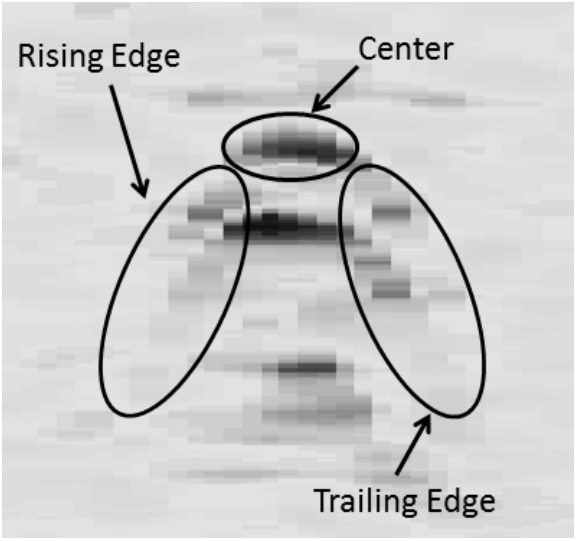


FIGURE 2.3: B-Scan of a target response and associated landmine HMM states. A confidence value for a given B-scan would be based on the probability that the HMM progressed through the states shown above in the order shown left to right.

given B-scan, the algorithm moves from the left side of the image to the right side. In testing, the confidence assigned by the HMM algorithm is the probability, given the features extracted, that the model progressed from background to rising edge, center, trailing edge, and finally back to background. This is known as the mine model. An important constraint on the model is that in order for the model to reach the trailing edge state, the model must have already progressed through the rising edge and center states. The final confidence of each B-scan is then the difference between the log-likelihood of the observations given a mine model and a background model.

Edge Histogram Descriptors[14]

Edge Histogram Descriptors (EHD) were originally used as a visual texture descriptor for images under the MPEG-7 standard. A texture descriptor aims to numerically quantify the types of edges within an area of an image. EHD features have also been utilized as a descriptor in image matching tasks [35, 36]. Similar to these

applications, the EHD algorithm also uses GPR 2-D B-scans to detect landmines.

Calculating an EHD value for an image region involves assigning an edge type to each pixel within the region. Pixels can be classified as being a diagonal, anti-diagonal, vertical, horizontal, or non-edge. Each pixel represents one vote based upon the edge type designation. The descriptor for the region is simply the 5 dimensional histogram created by binning each pixels edge type vote. In order to determine the edge type of each pixel, the region is independently filtered by 4 simple operators, one to detect each of the edge directions. The highest filter response for each pixel, if it surpasses a given threshold, receives a vote. Any values that fail to surpass the threshold instead vote for the non-edge bin.

Similarly to the HMM method of landmine detection, using EHD also relies on the detection of the hyperbolic shapes in a B-scan at target locations. For a given pre-screener alarm, multiple B-scan regions are extracted from the sensors nearby an alarm location. EHD features are generated for 7 overlapping columns of data within each B-scan section. The goal of this approach is to observe the differences in edge types across each alarm location flagged by the pre-screener. Since a typical landmine response has edges that change direction as a sensor passes over an object, the change in EHD descriptors across the response will reflect this. Training of the classifier is performed by collecting regions of B-scans that contain landmine responses and regions that contain background data. In order to avoid utilizing background sections of the data as training examples, it is useful to know the time at which the landmine signature begins to appear in the data. For training data, this localization is performed manually. Each training example must be hand labeled where the indications of a landmine response first appear in the data. At each alarm during testing, the data is broken up into 10 overlapping time regions and each is separately given a confidence value by the classifier. Order statistics [37] are then used to determine the final confidence for each pre-screener alarm.

Spectral Characteristics Feature [12]

Spectral Characteristic Features (SCF), unlike HMM and EHD methods, are not designed to leverage the shape characteristics visible in a B-scan signal as the GPR is moved spatially over a buried object. Instead, SCF considers the individual A-scans for indications of a landmine. Through experiments using finite-difference time-domain (FDTD) simulations it is possible to observe the characteristics in Energy Density Spectrum (EDS) generated by different kinds of shapes and materials under varying conditions. It was observed that the EDS generated from a given landmine type varied little under conditions of varying depth and soil dielectric constant.

Using the results of these simulations, SCF was tested on real data taken from various testing sites. Similar to results seen in the the simulations, it was observed that the correlation between targets in varying conditions remained high. Since the EDS calculation removes the uncertainty in depth and soil type, the classification of landmines is done simply by correlating matched filters generated from known landmines in the training dataset, with the test data. In testing, to reduce the effect of noise, the EDSs from A-scans near a pre-screener alarm are averaged together. The output of the averaged matched filter responses from around a pre-screener alarm determine the final confidence given from the SCF method.

2.3 Image Processing Methods

The desire for computers to understand the physical world has driven researchers to develop methods of visual image processing. One goal of this research is to leverage these image-based applications that identify shapes and objects and apply them to find shapes and objects within GPR B-scans. Three common types of object understanding are common within the image processing field: (1) Instance recognition is the process of identifying occurrences of a single identical object, (2) object

detection attempts to find a particular class of objects within images, and (3) category recognition attempts to classify whole images as having certain properties [19]. This research aims to evaluate all three methods and their applicability to landmine detection using GPR data.

2.3.1 Instance Matching

Instance matching aims to recognize instances of a known object within a scene. An example of this would be attempting to find all instances of a particular book within a visual scene. The specific title text and the cover design of the book can be used for recognizing it in a scene. It is assumed in this type of processing that the algorithm has an isolated image example of the known object. While there are applications where this problem is simply a template matching problem, modern research in this field is focused on accounting for various types of transformations common in visual images. Scale, intensity, rotation, and occlusion are all common types of transformations that affect how an object appears within an image [38]. Successful methods for instance matching must be able to identify instances of the object under these types of transformations.

The main contribution that allowed algorithms to reliably detect these transformations was the use of a part based model where matching occurs between small discriminable patches of the images instead of whole objects [19]. This method not only allows for more efficient processing of the data, but also allows for partial matches and accurate detection in the case of occlusion. The patches on an object and within a scene that are selected for processing are based upon their ability to be located repeatedly and their ability to be distinguished from other patches of data. These patches are located at points known as feature points, that are designed to identify these areas. Research has shown that corner detectors and blob detectors find the most informative image patches used for processing [39]. Corner detectors

are designed to find areas where two edges in an image meet [40] while blob detectors look for areas within an image that are unlike the immediate area around them [17, 18, 41]. Both of these methods provide feature points with excellent localization, and generally find areas within an image that are indicative of the object they are attempting to identify.

Although patches are selected for their ability to repeatedly localize on the same physical point on an object, errors of a few pixels are still common. This is one of the reasons why something called a descriptor is used to describe each patch of data instead of the vectorized pixel intensity values. The descriptor's function in instance matching is to provide a relatively low-dimensional representation of the data within a patch while being invariant to intensity and minor pixel differences in localization of the patch.

In most modern techniques for image instance matching the matching process begins with interest points located on the example image as well in the image of the scene in which the object is to be detected. The descriptors are calculated for each patch where a feature point is generated. To evaluate a scene, descriptors at each feature point within the scene are compared to the descriptors generated from training examples. A match is made if the distance between an example descriptor and one found in the scene is lower than a set threshold. These steps are shown in Figure 2.4 where the process of instance matching is shown with an example.

In this research, a number of feature point detection and descriptor methods were evaluated for instance matching in GPR data. Harris corner points [40] and a blob detector based on the determinant of the Hessian [18] were both used since their performance is generally good across a variety of applications [39]. Descriptors evaluated for this method were adapted from the Scale Invariant Feature Transform (SIFT) [17], Speeded Up Robust Feature (SURF) [18], and the Binary Robust Independent Elementary Features (BRIEF) [42]. The SIFT descriptor was chosen for its good

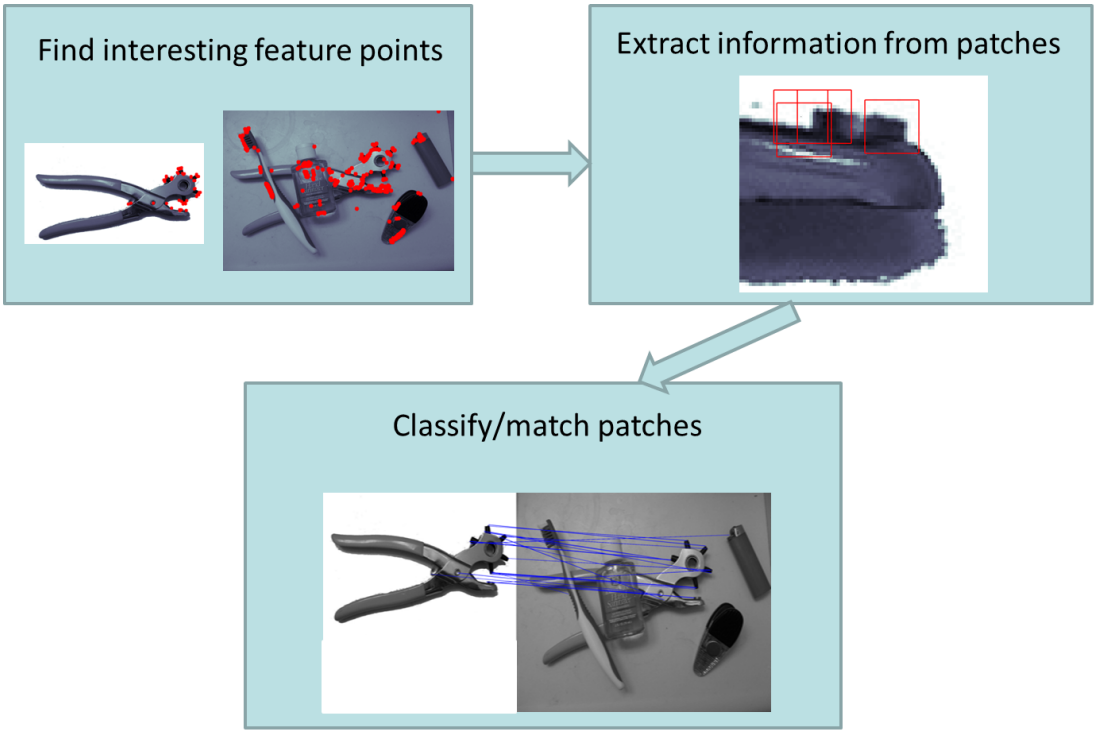


FIGURE 2.4: Algorithm flow for instance matching between an example image shown in and a cluttered scene where the object is matched.

performance in tests [43], while SURF is intended to represent a computationally less expensive version of SIFT. BRIEF was chosen since it provides a complimentary approach and its strong invariance to variation in pixel intensity differences.

Feature Point: Harris Corner Detector [40]

The Harris corner detector is based on previous work performed by Hans Moravec [44] where a corner is found within an image by shifting a window around each pixel and noting the intensity changes. If a corner is present, the intensity should change in all directions. If the pixel lies on a single edge, or a constant area, at least one shift direction will produce a low intensity difference. Harris and Stephens proposed a method that works similarly, but uses the gradients around each pixel to approximate the sum of squared differences (SSD) surface that models the change in intensity as a window is shifted around a pixel. The equation below gives the SSD

given a shift of $\bar{\phi}$ using a Taylor series approximation

$$SSD(\bar{x}_o, \bar{\phi}) \approx \bar{\phi}^T \sum_{\bar{x} \in w(\bar{x}_o)} [g(\bar{x} - \bar{x}_o) A(\bar{x})] \bar{\phi} \tag{2.1}$$

where $w(\bar{x}_o)$ is the window around a point location $\bar{x}_o$; and g is a Gaussian function. Note here that $\bar{x}_o$ is a two dimensional vector to indicate location within the image space. At each location, A is defined as

$$A(\bar{x}) = \begin{pmatrix} I_x(\bar{x})^2 & I_x(\bar{x}) I_y(\bar{x}) \\ I_y(\bar{x}) I_x(\bar{x}) & I_y(\bar{x})^2 \end{pmatrix}$$

where $I_x(\bar{x})$ is the derivative of the image, I, in the x-direction. The corner response at each pixel is then defined by Harris and Stephens [40] as:

$$R(\bar{x}) = \det(A(\bar{x})) - k(\mathrm{tr}(A(\bar{x})))^2$$

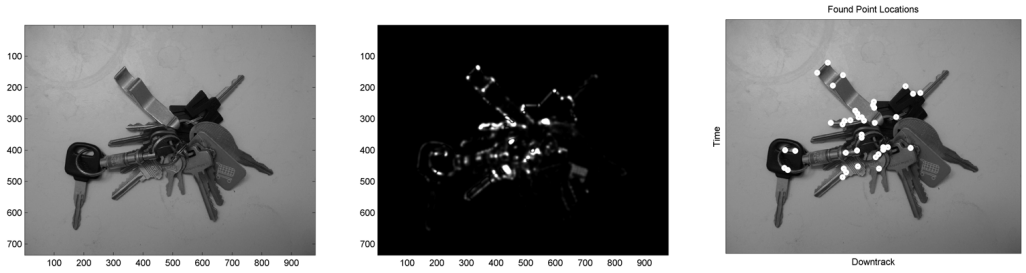
, where k is small constant.

Given the corner response for each pixel across the image, only local maxima are considered possible corner points. Those points that remain must then be above a given threshold to be recognized as a corner point by the algorithm. Figure 2.5 shows the process of detecting corner points on an example image.

Feature Point: Blob Detector [18]

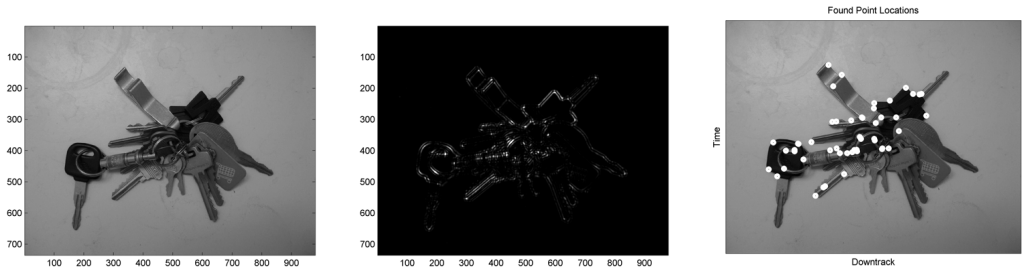
A blob detector is designed to identify areas where the intensity within the region differs greatly from the surrounding area. The blob detector that was utilized for this research is based on the determinant of the Hessian method that was used in the implementation of SURF [18]. This algorithm operates by calculating the Hessian matrix at each pixel location after a 2-dimensional Gaussian smoothing of the image.

$$Hessian(\bar{x}) = \begin{pmatrix} I_{xx}(\bar{x}) & I_{xy}(\bar{x}) \\ I_{yx}(\bar{x}) & I_{yy}(\bar{x}) \end{pmatrix}$$



(a) Original image (b) Corner response across image (c) Corner points identified

FIGURE 2.5: Figure 2.5(b) shows the corner response at each pixel given the original image in 2.5(a). Figure 2.5(c) shows the final selected points overlaid on the original image.



(a) Original image (b) Blob detector Response Across Image (c) Blob points identified

FIGURE 2.6: Figure 2.6(b) shows the response to the blob detector at each pixel given the original image shown in in 2.6(a). Figure 2.6(c) shows the final selected points overlaid on the original image after non-maximal supression.

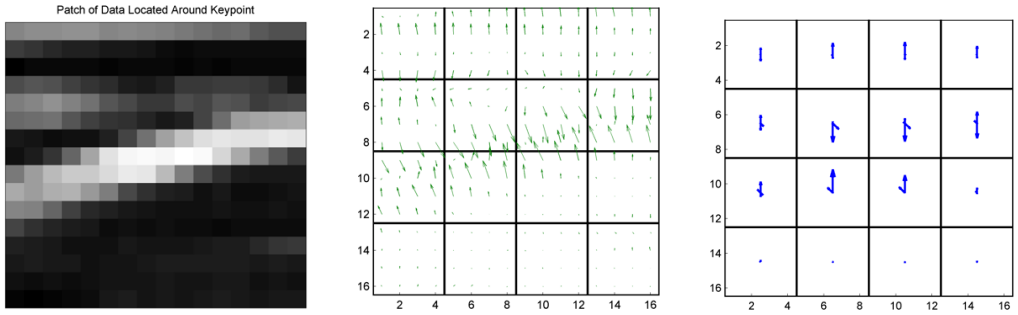
where $I_{xx}(\bar{x})$ is the second derivative in the horizontal direction at the image pixel $I(\bar{x})$ within the image I. If the determinant of the Hessian is positive, the location is either a local minimum, or a local maximum of the function depending on the sign of I_{xx} and I_{yy}. Each pixel's response to the blob detector is represented by the determinant of the Hessian matrix at each pixel. Local maxima and minima of the blob detector that surpass a given threshold are then identified as blob locations. Figure 2.6 shows the result of the blob detector on an example image. Note that a blob detector will also produce feature points on corners as corner pixels share characteristics that blob detectors respond to.

Descriptor: SIFT[17]

The Scale Invariant Feature Transform has proven to be a very robust and powerful feature descriptor for object matching in images [43]. This descriptor utilizes a 16x16 patch of pixels around each feature point. Within this 16x16 patch, the pixels are subdivided into 4x4 pixel sub-patches. Each pixel's gradient magnitude and angle is calculated using simple gradient filters. The magnitudes are weighted according to their distance from the feature point using a Gaussian mask. The pixel's magnitudes are then binned within each sub-patch according to their calculated angle's distance to 8 equiangular between 0° and 360°. Each sub-patch generates an 8 dimensional vector where each value is the sum of magnitudes from pixels that shared a similar gradient angle. The vectors from all sub-patches are concatenated to create a 128 dimensional vector representing the SIFT descriptor for one feature point. In addition, [17] found that a particular normalization method was most effective for the SIFT descriptors matching abilities. Each 128 dimensional SIFT descriptor is converted to a unit vector. All values over 0.2 are then set to 0.2 and the vector is again converted to a unit vector. This was found to provide invariance to intensity differences between images as the same edges can have vastly different gradient strengths between images depending on lighting or camera factors. Figure 2.7 shows the transformation of a data patch into a set of binned gradient vectors.

Descriptor: SURF[18]

The descriptor used in SURF was intended to be a faster alternative to the SIFT descriptor with no loss in performance. The SURF descriptor implementation found that a 20x20 patch of pixels around each feature point performed best for matching. Similar to the SIFT descriptor, SURF descriptors subdivide the patch of data and aggregate within each sub-patch. The SURF implementation subdivides each patch into 16 - 5x5 pixel sub-patch regions. Within each sub-patch x and y-direction gra-



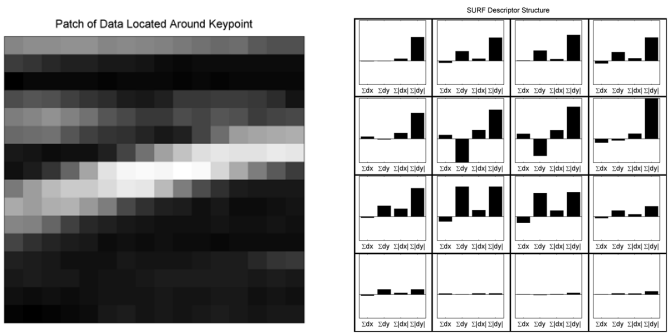
(a) Image patch (b) Gradients at each pixel (c) Binned Gradients in sub-patch

FIGURE 2.7: Example generation of SIFT descriptors on a patch of GPR data. The patch of data taken from around a feature point location is shown in 2.7(a). The region split up into the 4x4 grid of sub-patches. Plot 2.7(b) shows a quiver plot representing the gradients at each pixel location. Plot 2.7(c) shows the result after binning the gradients within each sub-patch. Each arrow in this plot originating from the center of the sub-patch represents the magnitude of the bin in that angular direction

dient information is calculated using two simple Haar wavelet filters. The gradients are aggregated to obtain 4 statistics that represent the sub-patch. The four gradient statistics are: $\sum dx, \sum dy, \sum |dx|, \sum |dy|$. These statistics were chosen as they are very fast to calculate and provided comparable performance to the SIFT descriptors. An example descriptor is shown in Figure 2.8. This calculation of a descriptor is fast as a result of using simple box filters that can be calculated quickly using an integral image [18]. Normalization for the SURF detectors simply consists of converting the 64 dimensional vector to a unit vector. As with the SIFT descriptor this is to provide invariance to changes in intensity between images.

Descriptor: BRIEF[42]

The BRIEF descriptor was selected in this research for its invariance to pixel intensity changes and its alternative method to account for feature point localization inconsistencies. Unlike previous descriptors introduced, the BRIEF feature is not calculated based upon the gradients within the patch. Instead BRIEF calculates comparisons

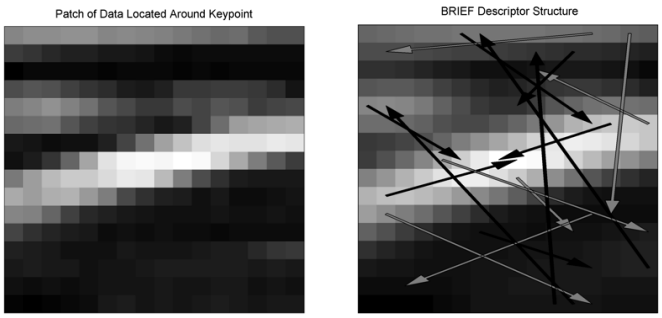


(a) Patch of Data (b) Statistics within each sub-patch

FIGURE 2.8: Example generation of a SURF descriptor on a patch of data. Plot 2.8(a) shows the patch of data around a feature point. The statistics of each sub-patch are shown in 2.8(b). The statistics of all 16 sub-patches will be concatenated to form a 64 dimensional feature vector.

between locations within the smoothed patch around a feature point. These comparisons, when dense enough, represent the relative intensity changes within a patch, thus representing structure within the patch. As the only information retained from each comparison is the location that has a higher intensity, no knowledge of absolute pixel intensity is incorporated. By smoothing the patch with a Gaussian filter prior to any comparisons, smaller differences in localization are not as detrimental to performance. The locations utilized in BRIEF descriptors are chosen at random with replacement and kept throughout training and testing. Figure 2.9 shows an example of how a small number of brief features would be generated within a patch.

In this experiment 256 different comparisons were used per patch. Matching between patches is performed with a simple Hamming distance, the number of comparisons that do not match between the two binary vectors. In the binary case, the matching between two vectors is then simply an XOR operation that can be optimized for very high speeds [42]. The resulting descriptor for each patch is a 256 dimensional binary vector.



(a) Patch of Data (b) Example Comparisons

FIGURE 2.9: The original patch of data to be processed is shown in 2.9(a). The same data is smoothed and shown in 2.9(b). A small number of comparison features are also displayed. The gray arrows represent location pairs where the first location has lower intensity than the second. The black arrows represent pairs where the first location has higher intensity.

2.3.2 Object Detection

In contrast to instance matching where an identical version of an object is identified, object detection is designed to find instances of a class of objects within images. An example of this relating to books would be trying to find books in a scene with no constraints on which specific book is to be found. Common examples of applications of this technology are face detection [21], pedestrian detection [20], and car detection [45]. For these types of detection problems, the challenges can be both the image transformation challenges seen in instance matching, as well as the uncertainty associated with interclass variations. In the example of pedestrian detection, interclass variations include the issues such as the size/weight of the person, the colors and patterns of their clothes, their poses, and all other differences that distinguish one human from another. Due to the amount of variation possible, most algorithms designed for this task tend to require more training data to effectively model the interclass differences. Classifier choice is also important in this application as over-fitting can significantly reduce classifier performance.

Training data for object detection generally consists of a large number of example

images of the desired object as well a sample of background images. Naturally, the dataset used for training will also contain some portion of the expected statistical inter-class variations typical of the problem under consideration. Instead of using a patch based approach, object detection generally uses a patch size that encompasses the entire object that is being detected. While a patch based method could work well for objects that have similar small scale characteristics, generally within a class, these small-scale characteristics vary too much to provide reliable detection. However, as in instance matching, descriptors are used to represent the data for the same advantages of localization and intensity differences.

The classifier for object detection is trained using the descriptor representation of all training example images. Detection of objects in images is then done by computing the descriptor and classifying each location across an entire image at a sampling rate that is acceptable given the descriptor aggregation size and expected object size. These steps can be seen in Figure 2.10 which outlines the steps that object detection generally follows.

Most of the methods utilized in the implementation of object detection are very similar to that of instance matching. Descriptor methods from both fields are generally inter-changeable; however the size of the patch is usually optimized for specific applications. In addition to the descriptors used previously, this research also utilizes a descriptor originally applied to pedestrian detection - the Histogram of Oriented Gradients (HOG) descriptor [20]. As classification takes a larger role in object matching, this research also evaluates three different types of classifiers. The first is a linear classifier that also performs dimensionality reduction known as Partial Least Squares Discriminant Analysis (PLS-DA). The Support Vector Machine (SVM) was studied due to its ability to create a decision boundary that maximizes the margins between classes [46]. Finally the Random Forest (RF) classifier is used due to its strong performance in recent comparisons between classifiers [47].

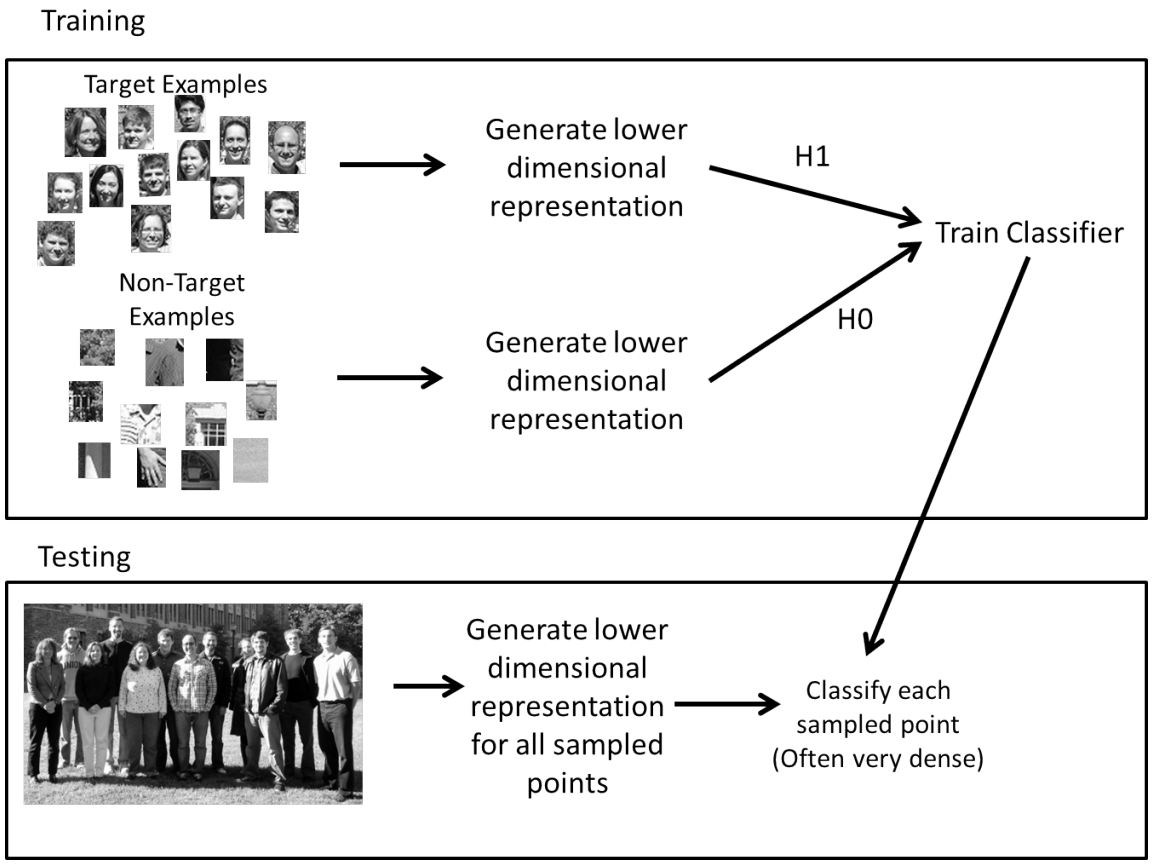
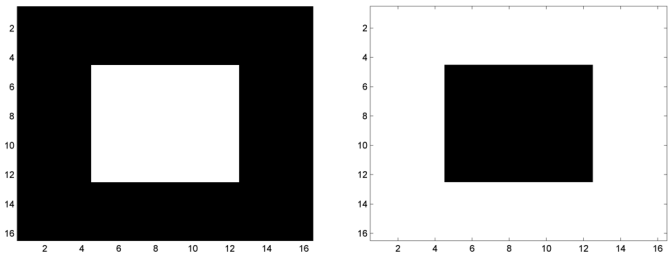


FIGURE 2.10: Example flow of object detection method applied to face detection in images. Example training samples are generally obtained via hand labeling. The classifier is trained on the descriptor representation of all example images. Locations within new images are converted to their descriptor representation and classified using the trained classifier.

Descriptor: HOG

The Histogram of Oriented Gradients (HOG) shares many similarities with both the SIFT and SURF descriptors. As the HOG descriptor was originally developed for pedestrian detection, it was designed to detect objects with the same orientation. This allows for the the requirement of a square descriptor that most rotation invariant descriptors maintain. Like the SIFT descriptor, the HOG descriptor takes the patch of data around a given area and subdivides it into sub-patches where gradients are binned according to orientation. However, instead of 8 equiangular bins from 0-360°,



(a) Patch #1 (b) Patch #2

FIGURE 2.11: Patches #1 and 2 both exhibit the same geometric structure. However the gradient directions along each edge are reversed. The HOG descriptor will bin the gradients in the 90° and 270° in the same bin. The SIFT descriptor will bin the gradients in separate bins as the range for the sift angular bins goes from 0-360°.

the HOG descriptor as presented in Dalal and Triggs [20], uses a number of bins between 0-180°. This difference means that while SIFT will identify a difference between the two images in Figure 2.11, the HOG descriptor will be the same for both.

Normalization of the HOG descriptor is necessary to account for intensity changes for the same reasons as other gradient based descriptors used in instance matching. HOG mainly differs from other approaches in that it does not have a set number of sub-patches within which normalization occurs. Instead the number of sub-patches included in normalization is referred to as a block size, and can be adjusted depending on the application. The original implementation of HOG found that between the methods of normalization they tried, the method of normalization generally made little difference. This research utilizes the L2-norm method, where the HOG vector for the appropriate blocks, $\mathbf{v}$, is normalized by

$$\mathbf{v}_{norm} = \frac{\mathbf{v}}{\sqrt{\mathbf{v}^2 + \epsilon^2}} \tag{2.2}$$

Where ϵ is a small non-zero number that tends to have little effect on the descriptor. The resulting normalized HOG sub-patches are then concatenated to form the descriptor representing the area desired.

Classifier: PLS-DA [48]

The Partial Least-Squares Discriminant Analysis (PLS-DA) classifier is used in this research because of its ability to perform dimensionality reduction to a space where all features are uncorrelated. While this type of dimensionality reduction is often carried out using an unsupervised method like Principal component Analysis (PCA), PLS is advantageous because it also leverages the class labels. This is especially important if the between-class separation is not as dominant as the interclass variation [49]. The models for both Principal Component Analysis (PCA) coupled with least squares optimization, and PLS-DA are the same. The data matrix $\mathbf{X}$ of dimensions $n \times m$, contains n observations, each with m features. For the i^{th} observation, there is a corresponding class label, y_i. The goal of these dimensionality reducing methods is to project $\mathbf{X}$ onto a new subspace that maintains the desired information about $\mathbf{X}$ while reducing the dimensionality. This can be modeled using

$$\mathbf{T} = \mathbf{XW} \tag{2.3}$$

Where $\mathbf{T}$ is the lower dimensionality representation of $\mathbf{X}$, and $\mathbf{W}$ is the transformation matrix that projects $\mathbf{X}$ into a lower dimensional space. In this model, our class labels, $\mathbf{y}$ are a linear function of $\mathbf{U}$. $\mathbf{C}$ dictates the relationship between $\mathbf{U}$ and $\mathbf{y}$ while $\mathbf{G}$ represents the error in the model.

$$\mathbf{y} = \mathbf{UC}^T + \mathbf{G} \tag{2.4}$$

When training a classifier for either model, it is reasonable to assume $\mathbf{y}$ is dependent on our representation of the data, $\mathbf{T}$. To learn the relation between the representation of the data and the labels, $\mathbf{C}$ must be calculated. This is done using least squares regression where $||\mathbf{G}||^2$ is minimized. The solution for $\mathbf{C}$ is then

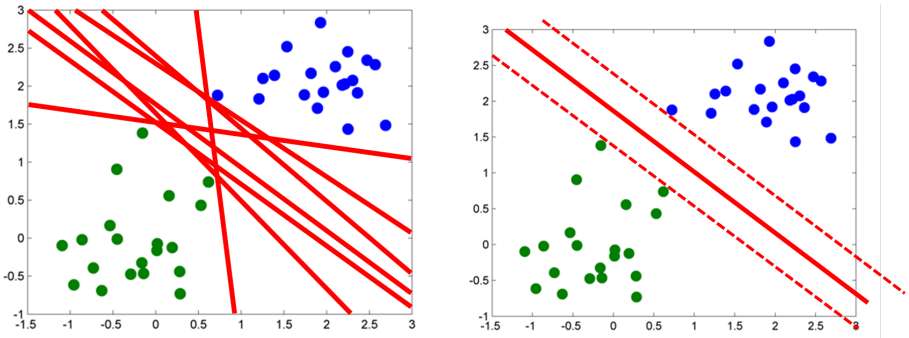
$$\mathbf{C} = (\mathbf{TT}^T)^{-1}\mathbf{Ty} \tag{2.5}$$

Given projected data, $\mathbf{T}$, both PCA regression and PLS regression utilize the same method of training the model. The difference however is in the forming of the transformation matrix, $\mathbf{W}$. PCA involves maximizing the variance in the final projected data with no regard to class labels. Therefore the orthogonal directions, in which PCA selects to project the original data, are the directions with the highest variance within the observation data. In PLS however, the goal is to find the direction vectors that maximize the covariance between observations, $\mathbf{X}$ and labels $\mathbf{y}$. There are a number of different algorithms designed to complete this task. The method used in this research is known as SIMPLS [50]. SIMPLS leverages the knowledge that the direction that maximizes covariance between $\mathbf{X}$ and $\mathbf{y}$ is the first left singular vector of $\mathbf{Xy}$. This method is iterative in that it finds a single direction vector per iteration. The residual information of $\mathbf{X}$ and $\mathbf{y}$ are calculated after each iteration and used as the input for the successive iterations. When the weights are generated, they are applied to $\mathbf{X}$ to obtain a $\mathbf{T}$ for the training data. Least squares is then used to obtain a value for the linear model variable $\mathbf{C}$ that explains the relationship between $\mathbf{T}$ and $\mathbf{y}$.

New test data that is received for classification is first reduced using the weights found during training, $\mathbf{W}$. The lower dimensional representation is then multiplied by $\mathbf{C}$ to obtain estimated labels for the new data. These estimated labels are the confidences output by the PLS-DA classifier.

Classifier: SVM [51]

The Support Vector Machine (SVM) is a statistical classifier developed to maximize the margin between the decision boundary and the two classes. This decision boundary is referred to as the maximum-margin hyperplane. In a linearly separable dataset, this decision boundary can be found heuristically by expanding the width of the decision boundary line until it contacts data points from either class. When the



(a) Possible decision boundaries (b) Maximum margin lines

FIGURE 2.12: Figure 2.12(a) shows the possible decision lines that classify the training data without error. Many of these lines however, are not desirable given new testing data. The solid line shown in Figure 2.12(b) is the desired robust decision boundary that is better suited to classify unseen data given these distributions

decision boundary can no longer expand, or rotate to expand further, it has achieved the maximum margin. An example of this can be seen in Figure 2.12. The label, y, of a test data point, $\mathbf{x}$, is determined by which side of the decision boundary it lies in. This is done using a linear model where

$$y = \mathbf{w}^T\mathbf{x} + b. \tag{2.6}$$

In this model, $\mathbf{w}$ is a vector whose direction is perpendicular to the decision boundary line. The data points are projected onto the $\mathbf{w}$ subspace, then the decision is determined by this projection and the offset, b. In training the linear SVM classifier, finding the best decision boundary involves maximizing the margins until it reaches the nearest data points. Training of a linear SVM with completely separable data is then an optimization problem as the width of the margins can be expressed in terms of $\mathbf{w}$. The minimization required for a linearly separable dataset is performed by minimizing the following

$$\text{argmin}_{\mathbf{w},b}\, \mathbf{w}^T\mathbf{w} \quad s.t.\ \mathbf{w}^T\mathbf{x}_i + b \geqslant y_i\ \forall\ I = 1, 2, ..., n. \tag{2.7}$$

This optimization is possible by introducing Lagrange multipliers that reduces to optimizing the following for the Lagrange multipliers,

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\mathbf{w}^T\mathbf{w}, \ subject\,to \ \ \alpha_i \geqslant 0 \forall i. \tag{2.8}$$

Where because of the α_i terms, $\mathbf{w}$ can be calculated as

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}. \tag{2.9}$$

Constraints on this minimization ensure that every point is classified correctly. However, this minimization problem fails if the data is not separable which is often the case with noisy data. Therefore a new term, that penalizes the misclassification of a training point, is introduced. The minimization problem is then

$$\text{argmin}_{\mathbf{w},b} \ \mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n} \epsilon_i \quad s.t. \ \ y_i(bw^T\mathbf{x}_i + b) \geqslant 1 - \epsilon_i \ , \ \ \epsilon_i \geqslant 0 \ \ \forall \mathbf{x}_i. \tag{2.10}$$

This allows for classification errors when $\epsilon_i \geqslant 0$ to have an effect on the minimization. The constant, C, weights how much a misclassification affects the decision boundary. The higher the value of C, the more important that training points lie on the side of the decision boundary that their label corresponds to.

This research utilizes a kernelized version of the SVM implemented in MATLAB by [52] utilizing a radial basis kernel. By performing the SVM learning and classification using a kernel function that maps into a higher dimensional space, the relationship between observations can be non-linear allowing for more complex relationships to be realized. The radial basis kernel is chosen for its flexible nature and wide acceptance in SVM applications.

Classifier: Random Forest [53]

The Random Forest (RF) classifier differs with respect to other classifiers in that it does not attempt to classify the observation by its entire set of features at once. The approach often taken to classify using all features generally leads to complex classifiers giving little intuition of how changes affect classification. Random forests operate using an ensemble classification method, where the decision is based off the mode of a large number of simple classifiers. The simple classifiers are decision trees and the whole system of trees is appropriately known as a forest.

For each decision tree, a subset of observations is randomly selected with replacement in a process referred to as bootstrapping. This is used as the training data for the tree, while the remaining data is used to test the error of the decision tree. For each decision node, a small subset of features is selected that are used to calculate the split between classes. This is repeated for all nodes in the tree. The resulting decision tree is one of many used in the ensemble.

To classify a new sample, the observation is run through all decision trees that have been trained on different bootstrap samples of the data. In a binary classification setting, the number of individual trees that classified the observation as a positive instance is used to assign a confidence to the observation.

2.3.3 Image Category Recognition

Image category recognition tasks require the algorithm to classify entire images into a small number of categories. An example of this would be the categorization of pictures from the inside of a library (containing books) from a school's image database. The kinds of categories can range from city vs. suburban images, to what kind of animal is present in the image. As these algorithms must classify the entire image, one of the main challenges is the multiple instance problem where parts of the image consist of background data, which is not indicative of the class label. However

due to the variety of possible realizations of each class of image, there is no way to distinguish easily between indicative and non-indicative patches within an image. Therefore a method that takes into account this multiple instance problem must be employed.

The algorithm that this research utilizes is a categorization method known as the bag-of-words approach. This approach was selected due to its simplicity and ability to process the entire image, where many part based models [54] focus on particular areas and are more akin to object detection. The bag-of-words method is based off a method originally developed for text categorization. Instead of words however, images are classified based upon patches that are sampled densely across the entire image. In order use this model with images, a discreet set of possible types of patches, analogous to a vocabulary, must be developed. This visual vocabulary is created by randomly sampling a large number of patches from the training data and clustering them based off of their descriptor representation. This research utilizes K-means as it has shown to cluster image patches reliably and performs well on recognition tasks [55]. Patches in an image are selected using a variety of methods. These patches can be extracted densely at given intervals, or at certain interest points such as corner and blob locations [19]. Each extracted patch in an image is then represented by the cluster that its descriptor representation is most similar to. The entire image is in turn represented by the frequencies of these clusters being present. These frequency histograms are then utilized for training a classifier using both positive and negative examples. Similarly to object detection the requirement for large amounts of training data exists in order to cover the statistical variation possible within the image categories. The final confidence of each image is the classier output given the frequency histogram vector. The steps involved in category recognition can be seen in Figure 2.13.

Many different methods of patch representation when clustering, patch localiza-

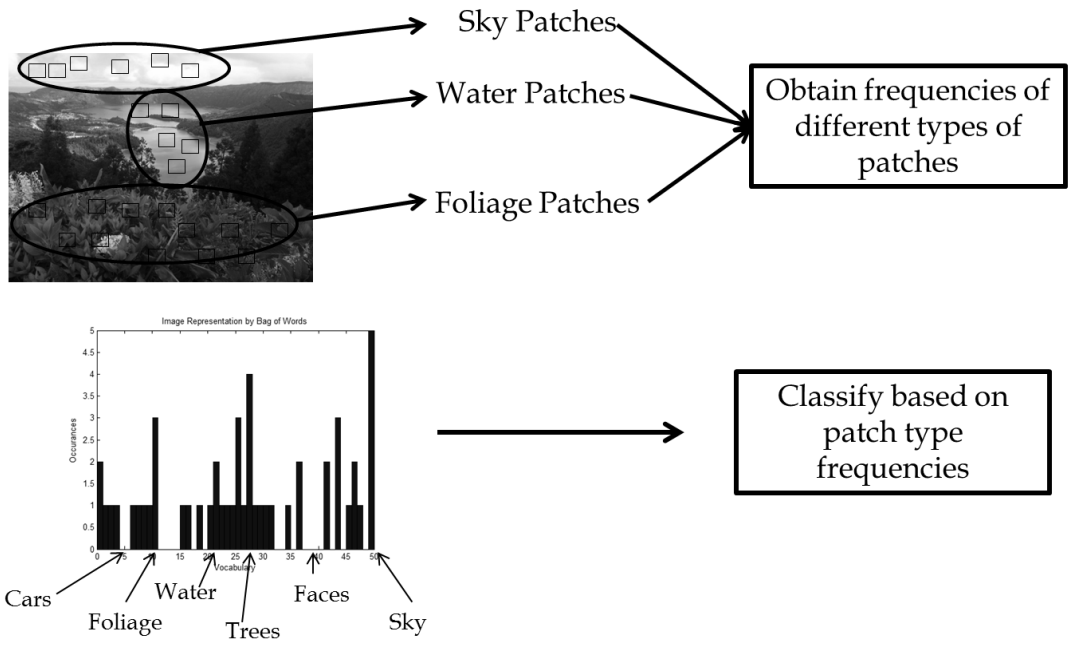


FIGURE 2.13: Processing steps for the Bag-of-Words model applied to visual images. The patch vocabulary is generated based off a random sample of patches within the training data. Each image is in turn represented by the frequencies of each patch type. These frequency histograms are then used to represent the image for both training and testing stages.

tion, and patch size parameters can be used in this method. This research evaluates the use of two descriptors, A modified HOG descriptor and the raw pixel values which been normalized and whitened. The modified HOG descriptor is the same as normal HOG descriptors, except the block size is changed in order to match the sub-patch size of a SIFT descriptor given the patch size. Therefore the HOG descriptor block size will be such that a gride of 4x4 blocks will fit into each image patch.

Whitened Patches

Utilizing normalized patches for clustering to determine a visual vocabulary has been found to perform poorly in image recognition tasks [55]. By whitening in addition to normalization as a pre-processing step, improvements were found in the ability to cluster the patches using K-means into centroids that provided better results in

recognition tasks [56]. By whitening the patches, the correlation between pixels is removed. This allows the whitened patch centroids to be more evenly distributed across the data sub-space [56]. The whitening process used in this research is based off the method used by [56]. The training data, $\mathbf{X}$ is a $n \times m$ matrix where each row is a random patch take from the training set of B-scans. The data is first normalized so that each row has a mean of 0 and a standard deviation of 1.

Given the decomposition of the covariance of the data matrix, $\mathbf{V}\mathbf{D}\mathbf{V}^T = \text{cov}(\mathbf{X})$, the whitened data is computed as

$$\mathbf{X}_{white} = \mathbf{V}(\mathbf{D} + \epsilon\mathbf{I})^{-\frac{1}{2}}\mathbf{V}^T\mathbf{X} \tag{2.11}$$

where ϵ is a regularization term to ensure that the eigenvalues, $\mathbf{D}$ are not too close to 0. The $\mathbf{V}$ and $\mathbf{D}$ matrices used in to whiten the training data prior to clustering are preserved to whiten the testing data in the same manner.

3

Instance Matching

This chapter presents the results and analysis obtained from implementing instance matching algorithms as a method for landmine detection in GPR data. An introduction to the dataset and scoring methods utilized throughout this document is also provided in this section.

3.1 Dataset and Performance Metrics

The data used for this experiment was collected over inert targets buried at known locations marked by Global Positioning Satellite (GPS) locations. Data was collected by driving the data collection system over the targets at operational speeds multiple times from different directions. The dataset used for this research contains a total of 788 target pre-screener alarms and 1635 pre-screener alarms that do not contain targets. Algorithms are tested using 10 fold cross validation where the classifier training is performed on 9/10ths of the data in each fold. As multiple data collections were obtained over each target, the folds were generated so that all instances of the same target at one location are contained in one fold to avoid incestuous testing and training.

Algorithm performance is presented utilizing a variant of the receiver operator characteristic (ROC) curve. Instead of viewing performance in terms of the probability of false alarm (FA), this research instead utilizes the false alarm rate (FAR) with units FA/m^2. This does not change the shape of the ROC curve but allows us to evaluate performance at certain operating points that are of interest. The probability of detection (PD) at set values of FAR will allow the comparison between algorithms at operating conditions acceptable to HMDS operators. The FAR levels that will be presented are at the rates 0.001 FA/m^2, 0.003 FA/m^2, and 0.005 FA/m^2.

3.1.1 Pre-screener Fusion

When creating a new type of discriminating algorithm for the detection of landmines, it is important to consider the algorithms capability to provide information that is at least partially orthogonal to the information provided by the pre-screener. The pre-screening algorithm assigns a confidence based on the energy content of an A-scan in relation to the statistics of nearby GPR data. The pre-screener is extremely effective at flagging subsurface objects as they generally produce a reflection with enough energy to statistically stand out from nearby data. Therefore an algorithm that produces correlated features, while performs well, may not necessarily provide additional information to the system. It is for this reason that fusing algorithm results with a pre-screener can offer better insight into the amount of new information being provided to the system.

3.2 Instance Matching Applied to GPR Data

Instance matching in visual images consists of attempting to find occurrences of a specific object within scene images. Generally matching occurs between one example of the exact object attempting to be found. An example of this would be attempting to identify the Eiffel Tower in albums of pictures taken in Paris. The object in ques-

tion is identical, however its appearance in images can be occluded, under different lighting conditions, and seen from different angles. The research presented in this chapter aims to utilize methods designed for instance matching to identify landmines by matching the shapes observed in 2-D B-scans.

B-scans from pre-screener alarm locations in the training data are used as the example images that will be found in the test samples. Certain modifications are necessary to both the feature extraction and matching in order for this method to be applied successfully to GPR B-scans. The following sections outline the methods used for feature extraction and classification that includes slight modifications to traditional instance matching approaches.

3.2.1 Feature Extraction

Due to the presence of background data within each B-scan, examples of GPR targets are not isolated examples. Instance matching algorithms generally assume that all areas extracted from a training example are located on the object that is meant to be matched. When this assumption does not hold in GPR data, extracted areas located in background data of landmine B-scans will reduce the distance to B-scans that only contain background data. This will lead to higher confidence false alarms if the landmine response is included in the training set, or lower the confidence of the landmine if it is located in the testing set. Therefore in GPR data, it is necessary to ensure processing takes place on data regions that are affected by a landmine's presence.

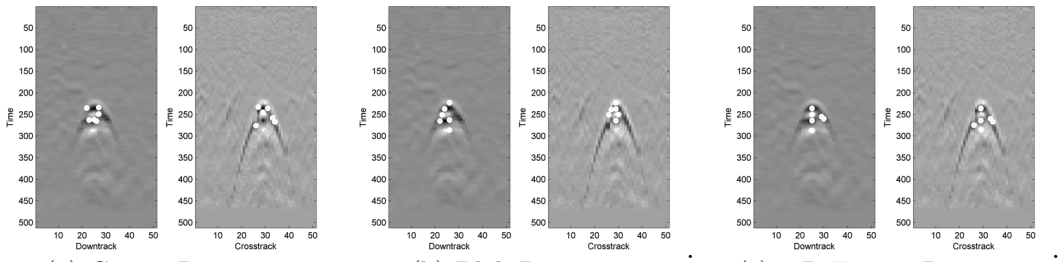
In this research, this is accomplished by utilizing feature point methods which respond strongly to landmine responses. Feature points in images are locations that can be repeatedly identified by a detector so that matching can occur between parts of an object. In this research the number of points located on each B-scan is limited to ensure that points are found on the landmine response. This diverges from instance

matching that typically utilizes a threshold to limit the feature points. However, low energy landmines often have difficulty generating enough feature points above the threshold to be matched. In addition, if thresholds are used, B-scans that do not contain landmine responses often will not produce a strong enough response to produce feature points above the threshold. These areas are valuable to process as they provide negative examples of pre-screener alarms. The value of processing negative examples can be seen in Figure 3.2 where the performance of only utilizing targets as training data is evaluated. By forcing these B-scan images to generate feature points, it allows non-target B-scans to match with each other based on their characteristics.

Feature Points

As mentioned previously in Section 2.3.1, this research utilizes both a blob detector and a corner detector developed for the purposes of finding feature points in visual images. The blob detector, taken from the SURF implementation [18], identifies areas that are unlike those surrounding it. The size of the area is dependent on the scale on which the blob detector operates. The corner detector, developed by Harris and Stephens, identifies areas where two edges in an image meet [40]. While these methods were not designed to operate on GPR data, many of the characteristics in visual images that these points react strongly to are present on a GPR landmine response. Figures 3.1(a) and 3.1(b) show the results of the application of these two detectors on a landmine response.

In addition to the feature point methods utilized from visual images, a new feature point method was developed for use on GPR data. Instead of finding corners or blobs, this method aims to identify peaks of high energy. High energy areas are desirable as any strong reflection after the initial ground-bounce is typically due to a change in conducting materials, such as soil to metal. Therefore, by identifying areas of high



(a) Corner Detector (b) Blob Detector (c) 2-D Energy Detector

FIGURE 3.1: A comparison of all three feature point detectors applied to a landmine response. The Harris corner detector tends to locate dots on the outer edges of the high energy areas. The blob detector finds small high energy areas, or around the edges of larger high energy masses. The 2-D energy method focuses purely on areas of high energy regardless of the surrounding area. Many of the points from the 2-D energy method lie on the central A-scan as that is typically the highest energy column.

energy in a 2-D B-scan, data caused by subsurface anomalies are identified. In order to find these areas, the B-scan energy is calculated from the data, then smoothed with a Gaussian filter. Non-maximal suppression is performed to find the local peaks. As with other methods, only the top number of peaks are kept based on their energy value. Therefore this method finds the top global peaks within a B-scan as feature points. Feature points extracted using each of the three methods is shown in Figure 3.1.

Descriptors

The variety of descriptors used in this research were chosen for performance, speed, and invariance to energy differences. The descriptor designed in the SIFT algorithm is used because of its strong performance across varied tests [43]. The descriptor presented in the SURF algorithm is used as it performs comparably to SIFT on certain tasks while reducing the computational complexity. BRIEF, as introduced previously was chosen for its alternative methods and invariance to intensity changes.

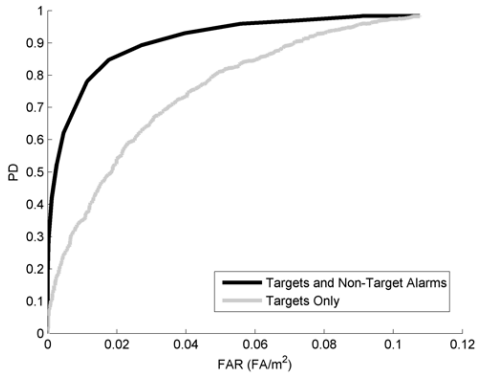


FIGURE 3.2: The above comparison shows the ROC curves obtained when utilizing only the minimum distance to targets, and summing the labels of the 15 closest matches across all positive and negative examples of training data. Utilizing the non-target areas helps to create low distance matches between non-target areas that results in an increase in performance.

3.2.2 Classification

Instance matching methods generally do not attempt to characterize the unwanted data as part of the training process [17, 18, 42]. Only a single example image of the desired object is considered when performing matching. While single target matching is possible to implement in GPR data, it was found to perform poorly. These results where only the distance between test samples and positive training samples were used can be seen in Figure 3.2.

This research instead utilizes both positive and negative training examples when performing matching. For each test sample, the confidence is determined by the number of H1 instances out of 15 training alarms that have the smallest B-scan distance. This is equivalent to a K-nearest neighbor (KNN) classifier where distance is determined by the match between B-scans.

The distance between two B-scans is determined by the sum of distances between descriptor pairs. Each descriptor extracted at a feature point in the first B-scan is paired with one descriptor in the second B-scan. The B-scan distance is the sum of Euclidean distances between all descriptor pairs between the two images. In this

application of matching, it is necessary to prevent descriptors from background areas matching multiple descriptors. Therefore, this method constrains each descriptor to only be matched once. When conflicts arise when a descriptor provides the closest match to multiple descriptors, the pair with the lowest distance is always taken. This matching can be visualized in Figure 3.5

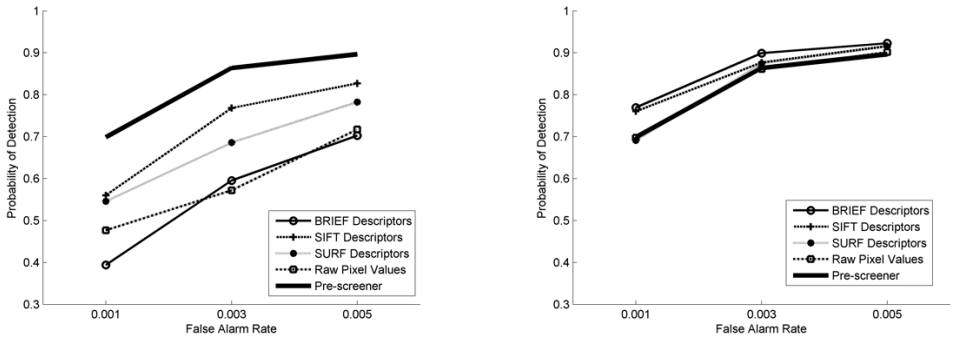
3.2.3 Image Matching Results

The initial results of landmine instance matching are shown below in Figure 3.3 at the FAR levels specified in Section 3.1. These results are shown with the top performing patch sizes utilizing 2-D energy points. Further analysis into the performance of various feature point methods is presented in Figure 3.4. Both of these methods show the improved performance with the addition of the pre-screener fusion. In certain cases, especially in Figure 3.4, the performance with the addition of the pre-screener can cause otherwise low scoring methods to achieve performance better than other methods that perform well without pre-screener fusion. This is due to the correlation between pre-screener confidences and matching confidences.

In the case of the BRIEF descriptors, it is especially apparent that the performance is greatly improved by the fusion with the pre-screener. As the BRIEF feature is simply represented as a binary comparison between pixels, it has absolutely no information regarding the absolute energy in the image. In comparison, the normalized SIFT descriptor can still be affected by local changes in signal energy; e.g., a number of bins with very large values, while normalized, still indicates the presence of large gradients as other gradient directions will be significantly smaller if no large edge exists.

Due to the poor performance of raw pixel data, further analysis is not included in this research. Raw pixel values have a high reliance on energy levels remaining constant across B-scans. While performing as well as BRIEF descriptors without

fusion, the addition of the pre-screener shows raw descriptors add little information to the system.



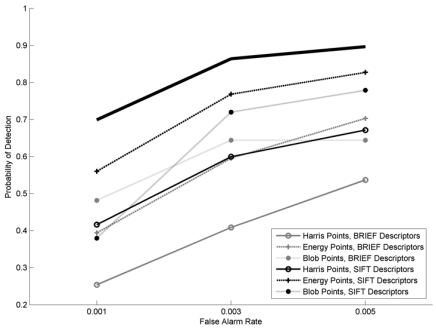
(a) Results independent of pre-screener confidence

(b) Results with confidences fused with pre-screener

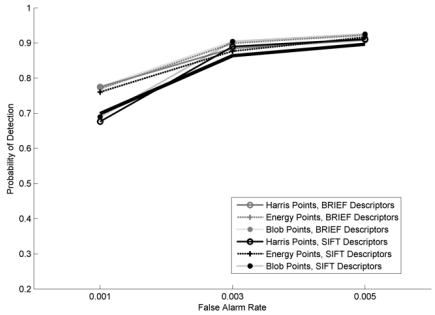
FIGURE 3.3: The results above illustrate the performance of various descriptor methods over the best performing parameters utilized in instance matching. The results on the left show the performance of the descriptors without fusing with the pre-screener confidence. The results on the right include the pre-screener fusion that shows additional information is being introduced by the fusion with instance matching. These results show that without the pre-screener information, the SIFT descriptor method outperforms other descriptors. In contrast, BRIEF features, perform very badly without pre-screener information perform much better after fusion with the pre-screener.

To further analyze these results, it is useful to be able to visualize which B-scans are matched between the training and testing data. Figure 3.5 is an example of a very good match between two targets of the same type in different ground locations. While some points are not matched exactly across these two B-scans, this example shows good feature point localization and most points match to similar locations within both responses.

Figure 3.5 shows an example where this method works well. In contrast, Figure 3.6(a) shows a match utilizing the same parameters where matching fails to occur between alarms of the same type. The target on the left matches most closely with the false alarm on the right. To understand why this alarm did not match with a training example of the same target type, Figure 3.6(b) is presented. This figure



(a) Results independent of pre-screener confidence



(b) Results with confidences fused with pre-screener

FIGURE 3.4: The results above evaluate the performance of various point localization methods utilizing the best performing descriptors. The results on the left show the performance without fusion with the pre-screener confidence. The results on the right include the pre-screener fusion that shows additional information is being introduced by utilizing the instance matching confidences. These figures show the large disparity between performance with and without the fusion with the pre-screener, especially in relation to the BRIEF descriptors.

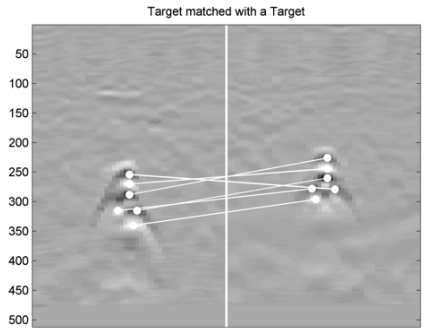
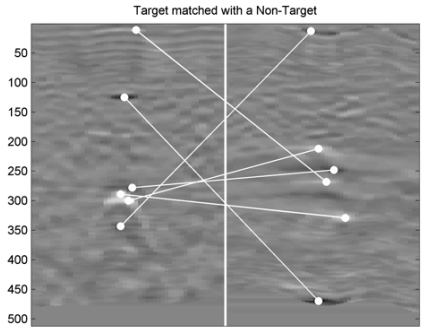


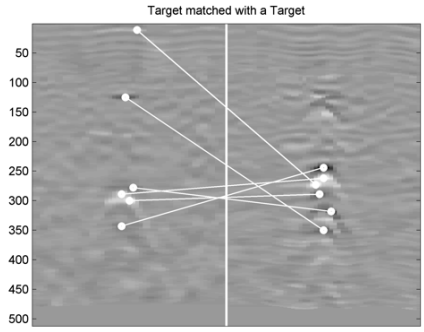
FIGURE 3.5: This figure illustrates matching between two different instances of the same landmine target type. The target on the right is the closest match over all training alarms available. The lines represent a match was found between the given SIFT descriptor whose patch center is represented by the white dots.

shows the points matched between the two targets of the same type but different ground locations.

The responses in Figure 3.6(b) look significantly different due to the soil properties around the alarm. This difference in appearance causes the points to be localized in different areas of the responses. As the points are located in different areas of the



(a) Target with top match, a false alarm.



(b) Target match between target of same type and depth

FIGURE 3.6: This figure shows the down-track matching between a target and its top match, a false alarm. The second matching image in 3.6(b) shows the feature point matching between the same target and an instance of the same landmine type at the same depth at a different location. The feature points are not located similarly and the overall responses do not share all of the same visual characteristics. Thus the two B-scans were not matched well during testing

alarm response, the patch data surrounding these points will not be similar, causing large distances between descriptors. This makes it difficult to match whole landmine responses between B-scans. While matching between entire B-scans is difficult due to points being located on different parts of a landmine response, it is likely within the training data that another point is located similarly. It is for this reason that a method that compares the data around a feature point to the entire set of points individually is developed.

3.3 Modified Instance Matching - Classification of Feature Points

While the addition of matching B-scan responses contributes additional information to the system, the performance of matching between B-scan images of landmines is often poor due to the variability in signal as illustrated in Figure 3.6(b). It has been found that in landmine data, the localization of feature points is not consistent over landmines of the same type as the conditions vary the signal appearance. It is for this reason that a modified version of instance matching is explored to offer flexibility

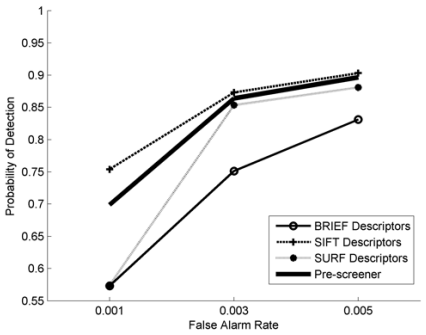
in matching patches.

Instead of attempting to match entire B-scan images, this method aims to classify the individual patches around each feature point. Descriptors are extracted from feature points in the same way as the previous experiment. All descriptors extracted from a B-scan that contain a landmine response are treated as positive (H1) instances, while all descriptors extracted from B-scans without targets are treated as negative (H0) instances. A SVM classifier is trained using the H1 and H0 descriptor vectors obtained from training data. Descriptors extracted from testing samples are classified and given confidences. The overall confidence for a given testing alarm is the average of confidences from all the extracted descriptors in both the cross track and down-track B-scan.

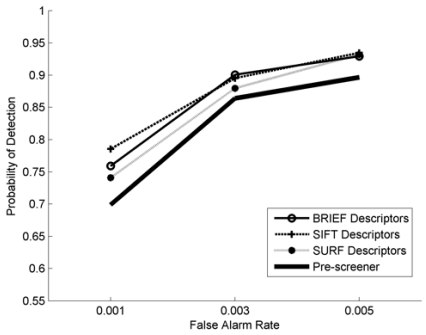
Instead of looking for instances of specific landmines in the data, this method is now classifying landmine like patches at various feature points. This is very similar to object detection as described in Section 2.3.2. The major difference is that only select parts around keypoints are being recognized, not the entire response. This method is looking for landmine characteristics at certain points within the image.

The best performing results of this method are displayed in Figure 3.7 for both the independent and fused results. The performance gain utilizing this method is especially apparent in the results that are not fused with the pre-screener. The parameters utilized for this trial that yielded the best results was a window size of 20×20 and utilizing 2-D energy points without fusion. With fusion, there were many methods that performed comparably, generally involving a patch size near 20×20 and either the blob detector or 2-D energy point method.

The results of utilizing the patches independently significantly boosted the performance of the detector in most cases. While instance matching was able to add additional information to the detection of landmines, the challenges of GPR processing prevented the model from working as it was designed. This chapter has shown



(a) Results independent of pre-screener confidence



(b) Results with confidences fused with pre-screener

FIGURE 3.7: These results show the best performance of the modified instance recognition method. The performance of the SIFT on the results not fused with the pre-screener remain high. The BRIEF descriptor also continues to do well only after pre-screener fusion is introduced. Both sets of results utilize 20×20 pixel patches around keypoints generated by the 2-D energy feature point locater. These results are compared numerically to using the initial method of instance matching in Table 3.1

Table 3.1: This table compares the top performers utilizing instance matching between B-scans and the classification of feature point regions. Large differences in PD exist when not fused with the pre-screener. However, when fused with the pre-screener, most of this performance gap is recovered by the instance matching method and in the case of $0.003\text{FA}/\text{m}^2$, surpassed .

False Alarm Rate	$0.001/\text{m}^2$		$0.003/\text{m}^2$		$0.005/\text{m}^2$	
Pre-Screener	0.6990%		0.8639%		0.8966%	
	Alone	Fused	Alone	Fused	Alone	Fused
Instance Matching	0.5602%	0.7696%	0.7683%	0.8992%	0.8272%	0.9228%
Feature Classification	0.7539%	0.7853%	0.8730%	0.8953%	0.9031%	0.9346%

that the within-class variance of landmine responses is too high for instance matching to be used effectively. It is for this reason that the next chapter utilizes methods that have been designed to account for greater amounts of within-class variation.

4

Object Detection

This chapter introduces the use of object detection methods on 2-D GPR B-scans. The goal in object detection is to identify instances of a class of objects. As described in Chapter 2, object detection is very different from instance matching, discussed in the previous chapter, where the goal is to identify instances of the same object. Object detection methods are designed to take into account the variation within a subjective class. These classes can be items such as bicycles, people, vehicles or phones. The objects within a class often have substantial variation in appearance. For example, with a vehicle detector, the object detection method needs to be able to identify both a small, white two-person car, as well as a dark blue passenger van. This challenge is similar to that of identifying objects in GPR data as there exists a large amount of variation in the shape, size, and energy of landmine responses in B-scans as seen in Chapter 3.

4.1 Object Detection Methods

The data required for object detection consists of multiple training examples, and the images that are to be tested. Each positive training example is an image region

that encompasses an example of the object class. Negative examples are generally taken from random locations within images that are representative of the background expected to be found in the testing set. All training examples and testing image regions are converted into a vector that represents the data. This representation is generally a descriptor that is extracted from the raw pixel data. A test sample can then be classified using a classifier trained on labeled descriptors in the training data. To detect objects within a test image, the image is sampled at regular, generally overlapping intervals. A descriptor is calculated for each image patch extracted at each of these locations. The classifier then calculates a confidence value based upon each descriptor. Any confidences that are above a certain threshold are identified as occurrences of the object class.

4.1.1 Training Data

In this research, all data extracted from B-scan images will be centered at the image column corresponding to the pre-screener alarm location. For non-target training examples, image regions are taken at regularly sampled depths in order to obtain a diverse set of negative training examples. To select positive training examples for the first part of this chapter, hand labeled row indexes where the landmine response begins was used to locate the correct image region on positive training examples. While this method works well and provides a good baseline for performance, it is not desirable to require hand labeling of data for a number of reasons.

The first issue is the subjective nature of where a landmine signature begins. Figure 4.1 shows a normalized B-scan at the location of a buried target. In this case, it is not easy to determine the exact spot where the training region should begin.

A second drawback is the time it takes to hand label a training set. In object detection applications, classifier robustness can be improved by incorporating additional training data. Additional data allows for the statistical variation in target

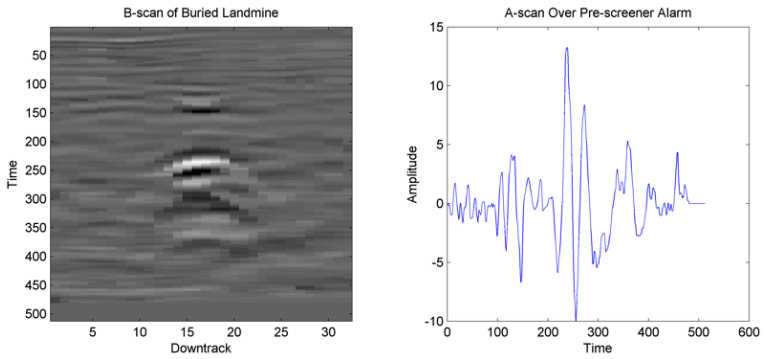


FIGURE 4.1: The above figure shows a target at a pre-screener alarm by its B-scan as well as the A-scan at the pre-screener alarm location. There is an initial energy response that begins at about 100 time samples. A second, higher energy response also appears at about the 200th time sample. It is unclear whether one or both of these responses is more important to characterize. Choosing either of the responses could be overlooking important information characteristic of this landmine. Choosing both requires that a method be developed that determines what constitutes a secondary landmine response or a piece of clutter, such as a rock.

type, orientation, soil conditions, and other factors to be represented during the training of the classifier. Datasets that encompass all of these variations are therefore very large, and have multiple passes over different configurations in order to encompass the sources of random variation.

The third drawback is the difficulty in the reproduction of results. Due to the subjective nature of hand labeling where the GPR target response begins, it would be difficult to ensure that any attempts to recreate the results are accurate. This makes it difficult for other researchers to accurately compare results between methods on different datasets as they may not have the same expertise in labeling data.

It is for these reasons that automatic training data identification will be attempted utilizing methods similar to those used in instance matching. Given the spatial location of an anomaly center (e.g., from the pre-screener alarm), all that remains is to localize the target response in depth. This research evaluates depth estimation using local maxima in the A-scan at the alarm location. Just as the 2-D energy feature points identified areas of a landmine in 2-D, this method aims to identify the

landmine response depth by its energy as well.

4.1.2 Descriptor Representation

The descriptor representations utilized for this research area come from both the instance matching and object detection fields of study. As both SIFT and BRIEF performed well in instance matching trials, they will be tested as descriptors for object detection as well. However, since the image region extracted for object detection is supposed to encompass the entire object, the constraint of square descriptors is no longer used. Instead, various image regions will be tested to determine the best performing size for detecting landmines. As the size of the image patch being processed is expected to be much larger, the number of BRIEF comparisons will need to be adequately scaled. The number of BRIEF comparisons used throughout this chapter is 512 per image patch.

In addition to the SIFT and BRIEF descriptors, the Histogram of Oriented Gradients (HOG) descriptor will also be used. The HOG descriptor was developed specifically for pedestrian detection but has also achieved success as a local image representation method in other tasks as well.[57] The HOG descriptor is very similar in calculation to the SIFT descriptor. Both methods aggregate the gradient magnitudes within local regions based on the gradient direction.

As discussed in Chapter 2, the differences between the HOG and SIFT descriptor are the sizing of the local regions (sub-patches), the directions which gradients are binned, and the normalization method. HOG descriptors use a fixed sub-patch size that vary in number depending on the size of the image patch being processed. SIFT descriptors use a fixed number of patches, varying the size of the sub-patches to achieve a total of 16 sub-patches. In addition, the HOG descriptor angular bins represent 9 equiangular directions between 0° and 180°. SIFT utilizes 8 equiangular bins between 0° and 360°. Normalization between HOG and SIFT are very similar

in that the descriptors are first converted to unit length. The difference is that SIFT then reduces any value which is above 0.2, to 0.2 and renormalizes.

4.1.3 Classifiers

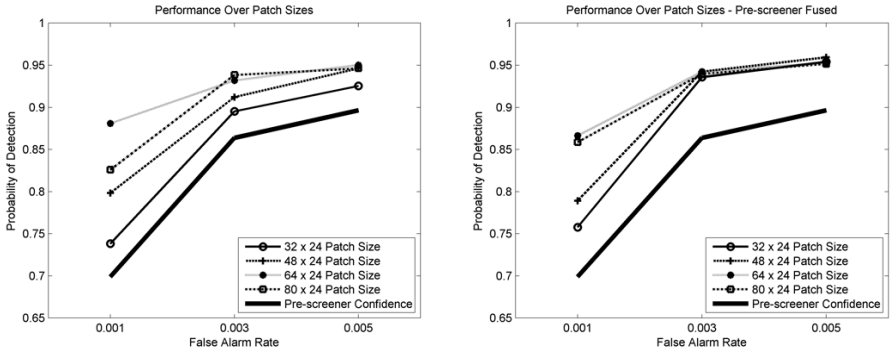
In this implementation of object detection, the ability to account for intraclass variation lies in the dataset variation, and the ability of the classifier to learn the structures within in the feature vectors that identify landmines. Three very different approaches to learning will be evaluated in this research. PLSDA, a linear classifier that performs dimensionality reduction that takes into account the class labels; the kernelized SVM classifier that aims to maximize the non-linear decision boundary between classes; and the Random Forest method that relies on ensembles of simple classifiers to learn relationships between features and labels.

4.2 Object Detection Results

This section provides the results and analysis of the application of object detection methods on GPR B-scans. Comparisons between descriptors and classifiers is given utilizing hand labeled training data in Section 4.2.1. Section 4.2.2 evaluates the use of a method to automatically extract training regions from target B-scans to avoid the need for hand-labeling.

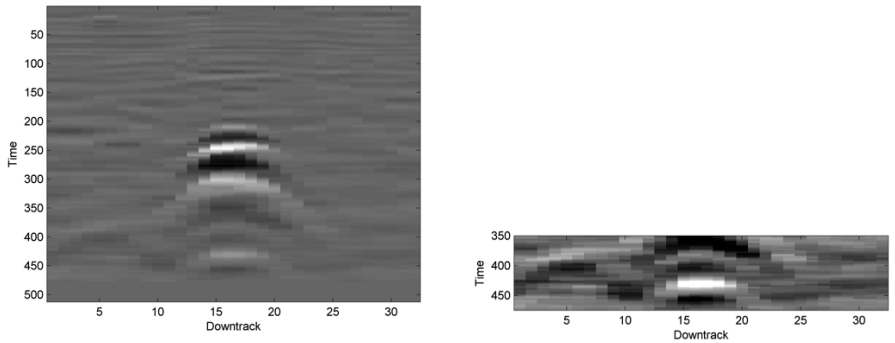
4.2.1 Descriptor and Classification Results and Analysis

Utilizing the hand labeled training data, a variety of image patch sizes were tested utilizing the SVM classifier. These results are shown in Figure 4.2. In general our results show that patch sizes that encompassed the largest landmine responses performed better. Although the inclusion of larger patch sizes mean that more background information is processed on B-scans with smaller responses, the information provided by the larger widths compensates for this. The number of rows included



(a) Non-fused Results (b) Results fused with pre-screener

FIGURE 4.2: These results illustrate the performance across a number of patch sizes utilizing the SVM classifier. It was found that the larger patch sizes have better performance. This is likely due to the utilization of late time structure as seen in Figure 4.3.

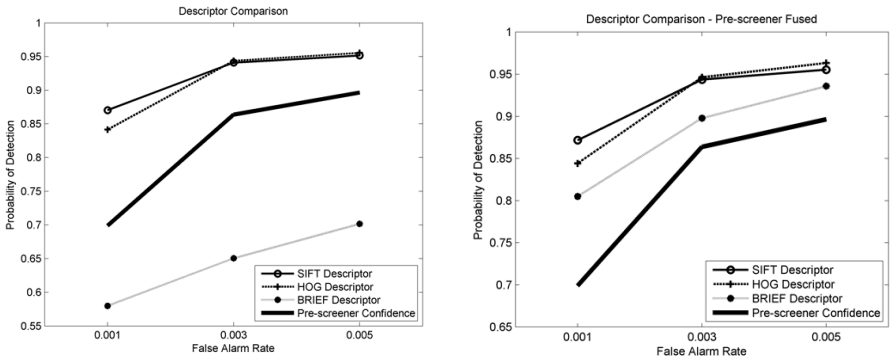


(a) Full B-scan (b) Late-Time Response

FIGURE 4.3: This example illustrates the structure observed after the initial higher energy response from a buried landmine. The data shown in Figure 4.3(b) is scaled differently allowing the lower energy areas to become more pronounced.

in each region that provide the best performance spans a large part of each B-scan image. This is due to the structure in the data located after the initial high energy response. This is shown further in Figure 4.3 which emphasizes the late time signals.

In this application it is important for a descriptor method to be invariant to intensity changes so that the late time structure shown in Figure 4.3(b) do not have similar descriptors as low energy background areas. If the descriptor relies too much on the energy content, the descriptors will match based upon energy similarity



(a) Non-fused Results (b) Results fused with pre-screener

FIGURE 4.4: These results illustrate the differences between the various descriptor methods. The best performing patch size is used in conjunction with the SVM classifier. The SIFT descriptor consistently performed the best among all methods tested despite the similarities to the HOG descriptor. The BRIEF descriptor performed poorly in both trials, especially when not fused with the pre-screener.

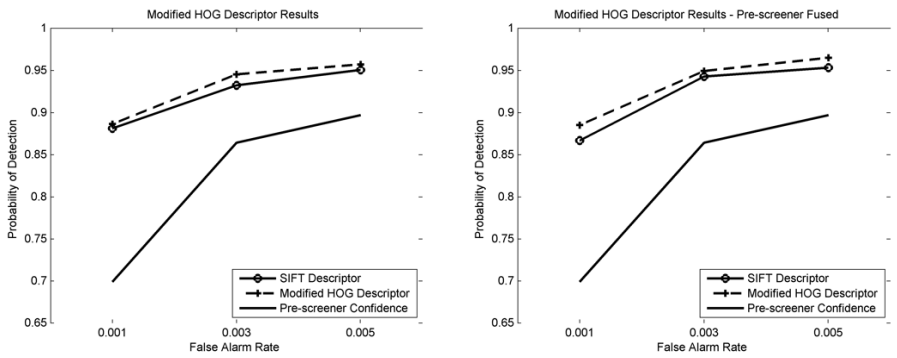
regardless of the structure present.

The results comparing the descriptors under consideration are shown in Figure 4.4. These results show very good performance in object matching methods in both the SIFT and HOG descriptors.

The results in Figure 4.4 show that BRIEF features continue to perform poorly without fusion with the pre-screener. In these experiments unlike trials in the previous chapter, even after fusing with the pre-screener confidences BRIEF features did not perform well. In addition, the SIFT descriptors outperformed the HOG descriptors in all patch sizes tested.

To better understand the factors causing higher performance for SIFT descriptors, a modified HOG descriptor was also tested in this research. Instead of dividing up the image patch region into sub-patches of a set size regardless of patch size, modified HOG descriptor sub-patches are set to be the same size as in SIFT descriptors. The results comparing the two descriptors are shown in Figure 4.5.

The performance of the modified HOG method shows that a large part of the performance difference between SIFT and HOG was the sub-patch size within which

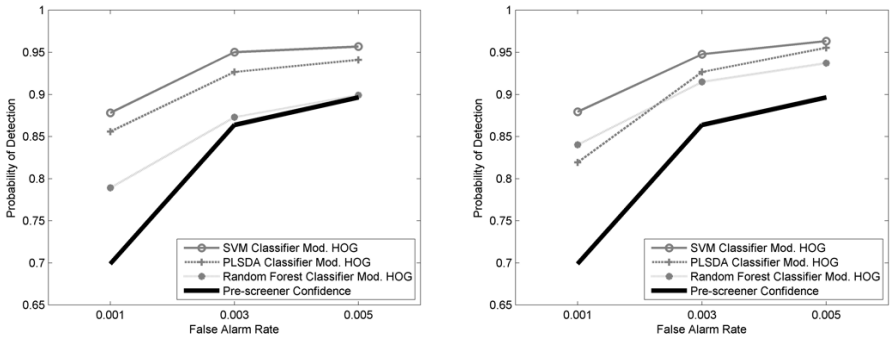


(a) Non-fused Results (b) Results fused with pre-screener

FIGURE 4.5: The results presented above show the performance of the SIFT descriptor as well as the modified HOG descriptor. The HOG descriptor was modified to have the same sub-patch size as the SIFT descriptor which results in an increase in performance.

aggregation occurred. When that parameter is equalized, the HOG descriptors actually perform better. The differences between the two methods are reduced to only the bin orientations and the normalization method. The bin orientations differ in that the eight SIFT descriptor orientation bins are evenly spaced between 0° and 360°. HOG on the other hand utilizes 9 bins between 0° and 180°. This difference was likely the most significant as the normalization scheme between the methods is only different in that SIFT constrains all values above 0.2. In addition, [20] found that little difference existed when utilizing the different normalization methods using HOG descriptors.

Figure 4.6 shows the performance across the various classifiers tested in this research. Utilizing the SVM classifier provided the best results across all trials. The random forest classifier however, was the only classifier that benefited from fusion with the pre-screener. Both the SVM and PLSDA classifier has the same or worse performance when fused with the pre-screener.



(a) Non-fused Results (b) Results fused with pre-screener

FIGURE 4.6: This figure shows the comparison between different classifiers utilizing the modified HOG descriptor results as shown above. The SVM classifier performed the best on all trials.

4.2.2 Automated Training Data

Previously, experiments were performed using training data that was extracted based on a hand labeled row index where the first occurrence of a landmine could be seen. This section provides results and analysis utilizing row indexes found automatically. To determine the depth of the landmine, this method utilizes the smoothed A-scan at the pre-screener alarm location. As a high energy received signal after the initial ground bounce typically indicates a change in subsurface electromagnetic properties, it is assumed that the buried landmine is the cause of these large reflections. We can therefore attempt to identify the row at which the landmine response is present by the largest peak after the ground bounce.

While utilizing the single largest peak in the central A-scan is effective at identifying the landmine response within the data, it is shown in Figure 4.7 that this is not always the start of the landmine response. For example, the landmine pictured in Figure 4.7 is a non-metal landmine. In this case, the dielectric discontinuity between the soil and the outside surface of the landmine is not the largest discontinuity (time sample 210). Instead, the transition from materials within the landmine actually produces the highest energy response (time sample 250). It is for this reason that

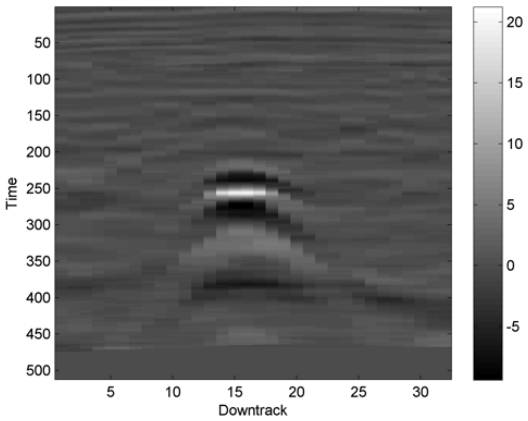
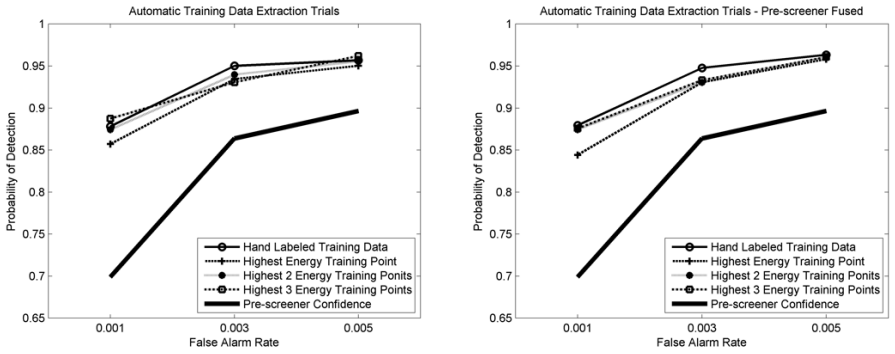


FIGURE 4.7: The above response is caused by a landmine with a non-metal casing. The initial reflection due to the soil/casing interface is located around time sample 210, whereas the highest energy response is located around time sample 250. This is a case where the hand labeled row index would differ greatly from the max energy method

between 1 and 3 multiple local maxima from each A-scan are utilized as positive instances for training.

The results shown in Figure 4.8 test the effectiveness of the automatic training data extraction utilizing between one and three of the top energy peaks in each A-scan.

As seen in Figure 4.8, methods of automatic training data extraction do not always work as well as selecting the index manually. While further research could be devoted to re-creating the hand labeled points automatically for this dataset, there are no guarantees that this method would translate to other datasets. It is therefore advantageous to consider methods that do not require isolated training examples, but still perform well on detection tasks. Research into a method that is aimed at solving this problem is presented in the next chapter.



(a) Non-fused Results (b) Results fused with pre-screener

FIGURE 4.8: This figure shows the results of attempts to automatically generate isolated training examples from the training data. It can be observed that the hand labeled training data continues to perform better at certin operating points, most notably at 0.003 FA/m^2. In the fused results, the hand-labeled data would be the clear choice. However the method utilizing three different training examples per training image does have better performance at some operating points in the unfused results. It is important to note that this is not necessarily a fair comparison. By having 3 training samples per positive image, there are three times the number of positive training samples.

5

Image Category Recognition

In the previous chapter it was shown that the performance of object detection methods far surpassed methods previously considered. However, a significant challenge continues to exist in the automatic selection of data used for training. Utilizing the peaks of the smoothed energy at the central A-scan is reasonably effective, but could not consistently out perform the use of hand labeled training points. It is for this reason that this research continues to explore models that do not require the processing to take place in only select regions of the image. In the image processing field, this is referred to as image categorization, or visual categorization. This problem, as explained in Chapter 2, requires the processing of the entire image to determine the correct category of the image. This chapter will evaluate the use of an image categorization model applied to the categorization of GPR 2-D B-scans and evaluate its impact on performance.

5.1 Category Recognition Methods

The category recognition model used in this research is the bag-of-words model [22]. This model is based on a text categorization method that categorizes documents

based upon the frequency of individual words found in each document. To extend this method to images, where individual image patches represent words, a closed set of possible image "words" is required. This closed set is referred to as a visual vocabulary, or a visual codebook. This research utilizes the commonly used method of K-means clustering to find centroids, and the centroids will be used to represent all image patches extracted [22, 19].

To create a visual vocabulary, a large number (hundreds of thousands) of random patches are extracted from the training data B-scans. The patches are then converted into a descriptor representation, or pre-processed and vectorized. All vectors are clustered utilizing K-means and the centroids are kept as the visual vocabulary. During feature extraction of a B-scan, each image patch is represented by the centroid that it is closest to by Euclidean distance. Each B-scan image is then represented by the frequencies of the centroids present. Therefore a vector with the same length as the number of centroids is used to represent each B-scan.

As each B-scan is represented by a vector, a classifier is trained using the vectors from all B-scans in the training set and associated labels. This classifier then calculates confidences based upon the vectors generated from test sample B-scans.

Within the bag-of-words model used in this research, a number of different methods can be used for the image patch descriptors as well as the image patch localization. Patch descriptors considered in this chapter are limited to the modified HOG descriptor and pixel data which has been contrast normalized and whitened. Different patch localization methods were also be tested as part of this research. The bag-of-words model has been applied to visual images utilizing a feature point based method [22] as well as a dense sampling method.[58] The results presented in this chapter will evaluate the use of these different methods as applied to GPR B-scan data.

5.1.1 Patch Representation

In applying methods of category recognition, only two descriptors were tested. These descriptors represent methods which show good performance in this research and from image processing research. The first method used for patch representation is the modified HOG descriptor that was introduced in Chapter 4. As previously discussed, it combines the benefits of a constant number of sub-patches within each patch region along with invariance to positive and negative region reversals.

The second method used for patch representation utilizes the pixels themselves as the features after they have been pre-processed and whitened. This method not only provides good performance on similar tasks in image processing [56], it also allows for the visualization of clusters so that the variety of features that are being recognized can be analyzed.

5.1.2 Patch Localization

When localizing patches, the conventional wisdom learned from instance matching methods is that extracting patches from feature-points, such as corners and blobs, will allow for more accurate matching as feature points tend to be well-localized on images, even under various image transformations [22]. In addition these points are less likely to be found on areas with low contrast since these are generally of little interest in both image and GPR data processing. However, other research has found that the number of patches extracted is more valuable than the actual localization in categorization tasks [58]. This research evaluated both methods by using a feature point variant and extracting a large number of patches both densely and at random locations.

The feature point method utilized in this research is the energy based method that was introduced in Chapter 3. Given the phenomenology of a subsurface object response, this feature point extractor identifies regions within the data that are

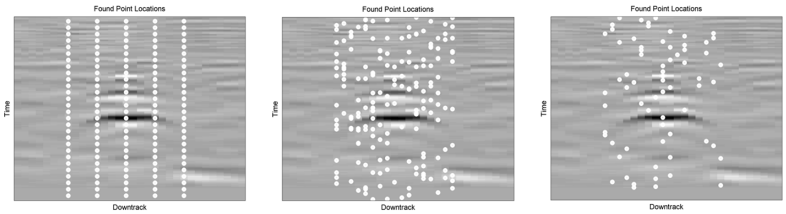


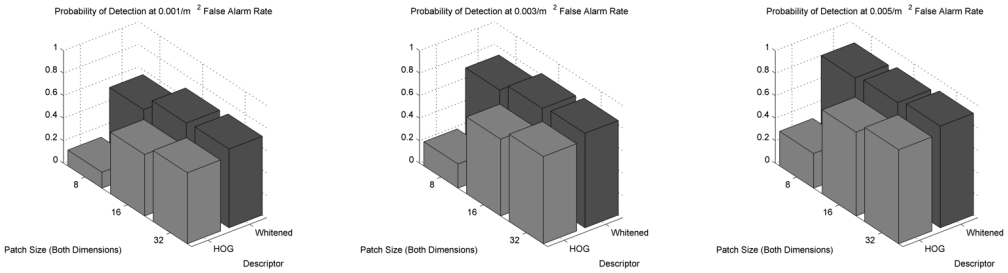
FIGURE 5.1: This figure illustrates three different methods of locating image regions on a GPR B-scan. Each white dot represents the center of a region extracted from around that pixel. Patch locations are constrained to remain near enough to the central A-scan so that the entire patch is localized within the B-scan. The leftmost image shows the dense sampling method where regions are extracted at regular intervals. The center image shows the random sampling method that extracts regions from uniformly random sampled locations across the image. The number of sampled regions using this method is the same as the number extracted using the dense method. The rightmost image shows the 2-D energy method that extracts regions from points that are local maxima of the 2-D energy map of the B-scan. This method generally produces substantially fewer features points than the other two methods.

indicative of a landmine response. Using local maxima of the smoothed B-scan's energy ensures that most regions to be extracted should be located near sub-surface object responses. In this implementation no thresholds are enforced, allowing local maxima from both high and low energy areas to be processed. This allows for target signal strength variance but assumes that the bag-of-words method will properly handle the background, low energy areas.

In addition to the feature point method, this research also evaluates the use of densely sampled and randomly sampled points within the B-scan. Figure 5.1 shows the examples of the different methods of region center localization applied to the GPR data.

5.2 Category Recognition Results

Section 5.2.1 will present the performance and analysis of different patch sizes and descriptors applied to category recognition. Section 5.2.2 will evaluate the differences in performance between patch localization techniques and discuss how they affect the



(a) Probability of detection at a FAR of $0.001/\text{m}^2$ across patch sizes and descriptors

(b) Probability of detection at a FAR of $0.003/\text{m}^2$ across patch sizes and descriptors

(c) Probability of detection at a FAR of $0.005/\text{m}^2$ across patch sizes and descriptors

FIGURE 5.2: This figure shows the performance of the descriptors over a range of square patch sizes independent of the pre-screener confidence. Performance is measured by the probability of detection at a given false alarm rate (FAR). The darker bars represent the normalized patch's performance as a descriptor. The normalized patch performance was found to be more consistent across various patch sizes than the modified HOG method.

performance of the bag-of-words model in GPR data.

5.2.1 Patch Size and Descriptors

The first series of results shown in Figure 5.2 examines the use of the two methods of patch representation, the modified HOG descriptor, and the whitened patches. The first parameter evaluated is a range of patch sizes. The results presented are independent of the pre-screener fusion. This is due to the poor performance of the HOG descriptors causing the pre-screener confidence to drive ROC performance.

These results show that utilizing a larger region size improves the classification performance across both descriptor methods. The increase in region size however is much more beneficial to the HOG descriptor. To evaluate the performance differences between descriptors further, it is beneficial to look at performance when fused with the energy pre-screener. In addition, for the following results, since larger sizes perform better across both methods, only larger patch sizes are evaluated.

The results presented in Table 5.1 show the detection performance of multiple width patches utilizing the two descriptor methods. The modified HOG descriptors,

Table 5.1: Probability of detection at each FAR operating point using different patch sizes and both whitened patch data and HOG features. All detection rates include fusion with the pre-screener confidences.

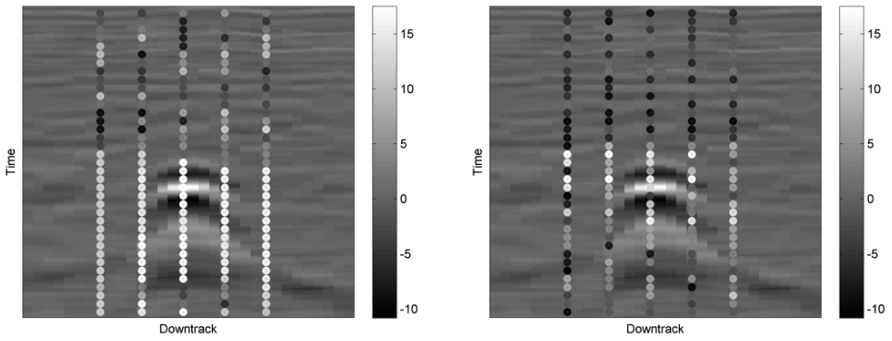
False Alarm Rate →	$0.001/m^2$		$0.003/m^2$		$0.005/m^2$	
Pre-Screener →	0.6990%		0.8639%		0.8966%	
Patch Size ↓	Whitened	HOG	Whitened	HOG	Whitened	HOG
32x16	0.7435%	0.7932%	0.8809%	0.8953%	0.9071%	0.9188%
32x20	0.7474%	0.8285%	0.8835%	0.9058%	0.8992%	0.9293%
32x24	0.7539%	0.7670%	0.8874%	0.9018%	0.9175%	0.9254%
32x28	0.7474%	0.7552%	0.8730%	0.8940%	0.9071%	0.9319%
32x32	0.7749%	0.7605%	0.8901%	0.9136%	0.9254%	0.9332%

while not performing well in the un-fused data, outperformed the whitened patches across all patch sizes. This was not the case in the performance results shown in Figure 5.2 where the whitened patches performed much better without the pre-screener fusion. It is reasonable to assume that this is due to the correlation between the confidences generated when using whitened patches and the pre-screener confidence. Since the pre-screener confidences are based heavily on the energy content of a landmine response, the whitened patches likely have a high sensitivity to areas of high energy. Figures 5.3 and 5.4 illustrate this by showing the two descriptor methods patch region performance side by side. The dots in each image represent the center of an image region extracted. The intensity of the dot corresponds to the centroid that represents the region in the classification of the image. The dot intensities are determined by ordering centroids, c_i by the ratio:

$$\frac{p(c_i|H_1)}{p(c_i|H_0)}, \tag{5.1}$$

so that light colored dots correspond to centroids that have a higher likelihood ratio score.

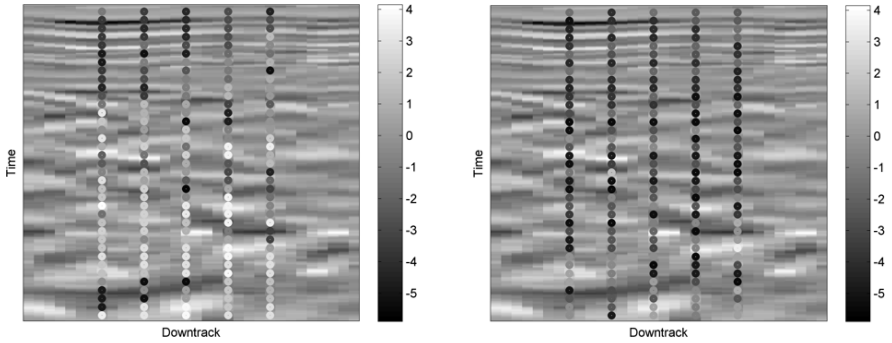
The images shown in Figure 5.3 support the hypothesis that the whitened patches react strongly to areas of high energy. In both methods, areas of high energy have lighter colored dots representing centroids that appear often on landmines. In the



(a) HOG Patches (b) Whitened Patches

FIGURE 5.3: B-scans of the same target visualizing the likelihood ratios of the centroids representing each region. Each dot's intensity is determined by the rank of the centroids likelihood ratio. These results were obtained using a 16×16 patch size, with a sampling rate of 4 in both dimensions.

area below the high energy however, the dots from the whitened patches method are no longer indicative of landmines. This is due to the lower energy levels below a high energy response. The modified HOG descriptor however, still recognizes the low energy areas as landmine like patches because of the late time ringing effect. This illustrates the modified HOG descriptor's ability to observe structure in low energy data.



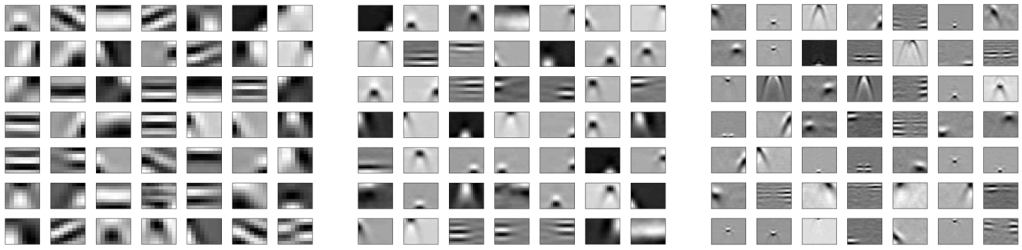
(a) HOG Patches (b) Whitened Patches

FIGURE 5.4: Visualizing the likelihood ratios of centroids that represent regions of a false alarm B-scan. Note that the energy range in these images is much smaller than the images of a buried target. The HOG patches have a much higher response to these areas where the energy is low, but structure in the data remains.

In addition, the images shown in Figure 5.4 explain why the performance obtained utilizing the modified HOG descriptors is worse without the pre-screener fusion. Many of the dots in the false alarm shown in Figure 5.4(a) are represented by centroids that appear landmine like. This is due to the blob like structure seen in the B-scan data. While performance without fusion with the pre-screener is poor, with the addition of the energy information, false alarms such as these have much lower confidences.

In addition to the reduced energy dependence, the modified HOG descriptor is more consistent within the oscillations of the hyperbola. This is because the HOG descriptor angles are binned by their angle between 0° and 180°. This means that a center point located on a negative region surrounded by positive regions will look the same as a positive region surrounded by negative regions. This is not the case for whitened images because large, positive magnitude signal data will correlate poorly with large, negative magnitude signal data. Both this and the fact that HOG performs aggregation over sub-patches is why HOG descriptor centers remain more consistent within a landmine response's ringing effect as shown in Figure 5.3.

The images shown in Figure 5.4 also provides insight into why performance with larger patches is generally better. This is particularly evident in the HOG descriptor image in Figure 5.4(a). This particular false alarm was shown because it contains significant amounts of texture in the later time samples where the image regions are represented by centroids commonly found on landmine responses. As the classifier is not given the spatial relation between these parts, it simply classifies based on the frequencies of centroids present in the B-scan. This can be further explained by considering Figure 5.5 which displays centroids from different sized patches. The larger patch sizes contain much more complete landmine structure while the smaller patches only identify parts of the landmine structures. Using smaller patches, the spatial information is lost because the order of the parts is not taken into account in



(a) 8x8 Patch Centroids (b) 16x16 Patch Centroids (c) 32x32 Patch Centroids

FIGURE 5.5: The figures above illustrate the differences in centroids for different sized patches. The smaller patch centroids can only represent parts of the landmine response. The larger patch centroids are able to represent both the rising and falling edge of the landmine response allowing it to represent the entire landmine structure.

the bag-of-words model.

5.2.2 Patch Localization

This section considers the use of different methods of patch localization. Only the modified HOG descriptor will be used as the whitened patches reliance on energy levels degrades performance when fused with the pre-screener confidences. Due to the constraint of keeping region center locations in areas where the entire patch can be extracted, a patch size of 32×16 was used for analyzing patch localization. As discussed in the previous section, the largest, 32×32, patch performed the best. However given that the data used in this experiment is 32 pixels wide, this would allow for only one column of region centers to be extracted from each image's central A-scan. This would give little insight into which localization method's performance would be the most effective given different patch sizes.

Figure 5.6 illustrates the differences obtained across localization techniques. Note that these methods are fused with the pre-screening algorithm. Without fusion, the results would be difficult to extrapolate from as the 2-D energy points would naturally be located in areas of higher energy than the surroundings. This could lead to performance differences due to energy based information that is already acquired by

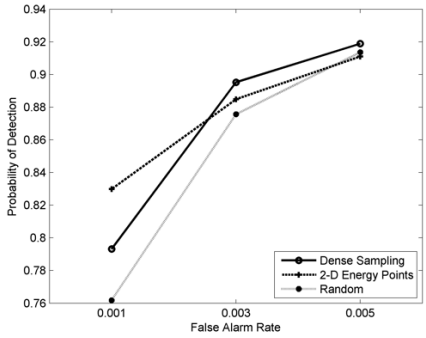
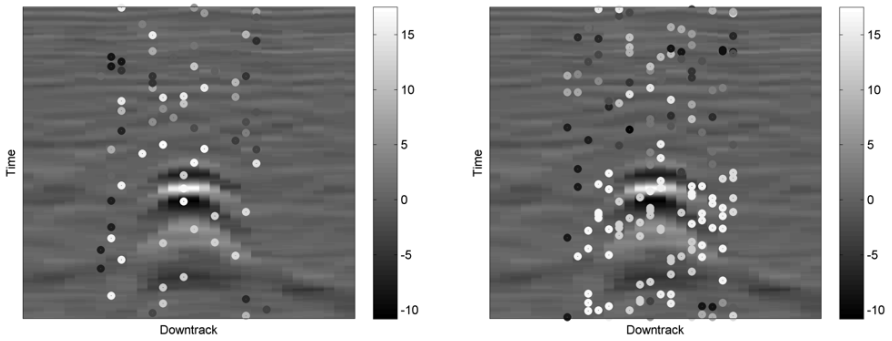


FIGURE 5.6: Probabilities of detection for the different methods of localization of regions within individual B-scans. Note that the number of random points chosen is the same as the number of points used in dense sampling.



(a) 2-D Energy Map (b) Random Sampling

FIGURE 5.7: B-scans showing the region centers when utilizing different region localization techniques. The intensity of each dot is determined by the rank of the centroid's likelihood ratio that represents the region around that dot. While the regions with high intensity dots are mostly consistent with the dense processing method shown in Figure 5.3, the density of center regions in both methods is variable and notably lower in the 2-D energy map method.

the pre-screener.

It is difficult to determine which method would be the most robust based on the images shown in Figure 5.7 alone. While it appears as if the 2-D energy map has a much more random scatter of high intensity dots, it does not appear to affect the overall final ROC performance. However, these images do provide insight into certain localization issues with 2-D energy feature points. The insight is that areas close to

the high energy potion of the landmine are sparsely populated aside from the few maxima locations. This is undesirable as the rising and falling edges are then only characterized by the regions centered at the peaks of the hyperbolas. This suggests that any part-based extension of the bag-of-words model [54, 57] will not be able to reliably utilize the rising and falling edges as a distinct part of the landmine response. In addition, the unreliable nature of localization means that any part based model will struggle to have consistent relative spatial relationships between region centers.

The random sampling method, which has a much higher sample count than the the 2-D energy method, appears to more effectively identify areas of a landmine response by the higher intensity centroids seen around the hyperbola. The number of regions extracted is the same as dense sampling to be able to accurately compare the two methods. The performance difference in testing seen in Figure 5.6 however, does not provide evidence that random sampling would provide better performance than a dense sampling method.

6

Conclusions

The previous chapters have presented the application of image processing methods to the detection of landmines in GPR data. It was found that methods for detecting real world objects in visual images are also effective at detecting landmines present in GPR data.

Chapter 3 considered the use of instance matching to match test B-scans with known landmine B-scans. It was shown that shapes in GPR responses were often similar between landmines of the same type. Instance matching methods were able to match these shapes and successfully identify landmines based upon these matches.

It was also observed in this work that descriptor methods contributed a significant portion to the overall algorithm performance as compared to the point localization methods. SIFT and BRIEF both provided good performance in instance matching applications. In addition, it was found that descriptors with greater invariance to changes in energy benefited the most when confidences were fused with the pre-screener.

In addition to instance matching, a modified method was also tested to account for the variation in point localization. By relaxing the requirement that patches were

only used to match between two B-scans, individual patches were instead used as positive and negative examples. A classifier was trained using the descriptor representation of all patches extracted from training alarms. In testing, the individual patches are classified and confidences summed across patches from each B-scan. This method attempts to detect parts of a landmine response independently. Performance utilizing the classification of individual patches outperformed instance matching between B-scans. This is due to the ability of this method to compare landmine parts across more than two B-scan images. Good performance utilizing this method indicated the potential for object detection methods to identify full landmine responses.

Chapter 4 introduced the use of object detection methodology on GPR B-scans. Object detection utilizes a patch size that is meant to encompass the entire landmine response. Utilizing hand labeled training examples, this method provided the best performance across all methods tested in this work.

Similar to the results found in instance matching trials, the choice of descriptor had significant impact on the overall algorithm performance. In addition, after analyzing the differences between SIFT and HOG, it was found that using feature descriptors with sub-patches whose size is dependent on the overall patch size performs better. This lead to the use of a modified HOG descriptor that produced the best performance by utilizing the HOG method of gradient binning and the SIFT method of sub-patch aggregation.

As the need for hand labeled training data is not desirable for both research and adaptability reasons, methods to generate training data locations automatically were also evaluated. These methods were able to achieve very good performance by utilizing multiple A-scan energy peaks as locations that indicate a landmine response. However, performance achieved by selecting only a single hand labeled training example was still higher at certain operating points. This motivated research into methods that do not require hand labeling of training data such as image category

recognition methods.

Chapter 5 presents research evaluating the use of category recognition methods. These methods are desirable as they are designed to take into account the background data present within each image. The bag-of-words model is utilized in order to generate confidences based upon the frequency of patches that are categorized by a visual vocabulary. The bag-of-words model applied to GPR images gave the best performance when coupled with the modified HOG descriptor. This was especially evident in the late time reflections caused by a landmine response. However, it was observed that the performance of HOG without pre-screener fusion was degraded in false alarm images that displayed unordered, low energy structure. While the performance of this method could not match the performance of object detection methods, it demonstrates the importance of utilizing a model that takes into account spatial information across a B-scan. This motivates future work into category recognition models that are able to utilize the relative spatial locations of the parts of a landmine response.

Table 6.1 shows the comparison of the best top results from the different image processing methods applied in this research. These results present the best set of parameters separately for each independent approach, and fusion with the pre-screener.

These results clearly show that the object detection methods perform the best across all methods. Given the challenges of recognizing landmine response shapes and the variability seen in the data, object detection clearly appears to be the more appropriate match for this problem. However, the object detection method requires updating of the expected target population changes. This requires hand labeling large amounts of data. Therefore, methods that can automatically extract training data are desirable. It is for this reason that further research into methods of object detection and image category recognition are necessary.

Future work will consist of the consideration of new methods to detect objects

Table 6.1: The above results show the probability of detection at various false alarm rates. At each operating point, the highest probability of detection achieved is in bold. For all trials, methods presented in Chapter 4 had the highest probability of detection.

	Independent			Fused with pre-screener		
Pre-Screener	0.6990%	0.8639%	0.8966%	0.6990%	0.8639%	0.8966%
Instance Matching	0.5602%	0.7683%	0.8272%	0.7749%	0.9045%	0.9188%
Modified Instance Matching	0.7539%	0.8730%	0.9031%	0.7853%	0.8953%	0.9346%
Hand Labeled Object Detection	0.8783%	**0.9503%**	0.9568%	**0.8796%**	**0.9476%**	**0.9634%**
Automatically Labeled Object Detection	**0.8874%**	0.9306%	**0.9620%**	0.8757%	0.9332%	0.9607%
Category Recognition	0.6990%	0.8390%	0.9031%	0.6387%	0.9254%	0.9437%

within cluttered images without the need to hand label training data. Specifically this research will evaluate methods that utilize the spatial organization of object parts within images [54, 57]. It is believed that this will benefit the performance of discrimination in the targets evaluated in this research. In addition, the understanding of spatial relationships between GPR response parts, can lead to the detection of more complex targets. As a growing number of anti-vehicle threats emplaced are improvised devices, the spatial relationship between components is no longer constrained to manufactured standards. Models that utilize the spatial relationship between object components within visual images therefore present a possible solution to the constantly evolving challenge of finding explosive threats in GPR data.

Bibliography

[1] International Campeign to Ban Landmine, "Landmine monitor 2012," November 2012.

[2] Geneva International Centre for Humanitarian Demining, "Mines other than anti-personnel mines," March 2012.

[3] Joint Improvised Explosive Device Defeat Organization, "Annual report," 2010.

[4] J. MacDonald, "Alternatives for landmine detection," tech. rep., RAND Corporation, 2003.

[5] Cambodia Mine Action Centre, "Annual report 2011," 2011.

[6] *Improving detection of low-metallic content landmines using EMI data*, vol. 4, 2000.

[7] S. L. Tantum, W. R. Scott, K. D. Morton, L. M. Collins, and P. A. Torrione, "Target classification and identification using sparse model representations of frequency-domain electromagnetic induction sensor data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. PP, no. 99, pp. 1 –18, 2012.

[8] J. Peters, L.P., J. Daniels, and J. Young, "Ground penetrating radar as a subsurface environmental sensing tool," *Proceedings of the IEEE*, vol. 82, pp. 1802 –1822, dec 1994.

[9] C. Bruschini and B. Gros, "A survey of current sensor technology research for the detection of landmines," in *Proc. International Workshop on Sustainable Humanitarian Demining*, vol. 6, pp. 18–27, 1997.

[10] E. Rosen, "Detecting improvised explosive devices," 2010.

[11] J. Wilson, P. Gader, W.-H. Lee, H. Frigui, and K. Ho, "A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection

and discrimination," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, pp. 2560 –2572, aug. 2007.

[12] K. Ho, L. Carin, P. Gader, and J. Wilson, "An investigation of using the spectral characteristics from ground penetrating radar for landmine/clutter discrimination," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, pp. 1177 –1191, april 2008.

[13] P. Gader, M. Mystkowski, and Y. Zhao, "Landmine detection with ground penetrating radar using hidden markov models," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, pp. 1231 –1244, jun 2001.

[14] H. Frigui and P. Gader, "Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic k-nearest neighbor classifier," *Fuzzy Systems, IEEE Transactions on*, vol. 17, pp. 185 –199, feb. 2009.

[15] D. Potin, P. Vanheeghe, E. Duflos, and M. Davy, "An abrupt change detection algorithm for buried landmines localization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, pp. 260 – 272, feb. 2006.

[16] C. R. Ratto, P. A. Torrione, and L. M. Collins, "Exploiting ground-penetrating radar phenomenology in a context-dependent framework for landmine detection and discrimination," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 5, pp. 1689–1700, 2011.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[18] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision ECCV 2006* (A. Leonardis, H. Bischof, and A. Pinz, eds.), vol. 3951 of *Lecture Notes in Computer Science*, pp. 404–417, Springer Berlin / Heidelberg, 2006.

[19] R. Szeliski, *Computer Vision: Algorithms and Applications*. Texts in Computer Science, Springer London, 2011.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.

[21] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[22] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *In Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.

[23] J. Willamowski, D. Arregui, G. Csurka, C. R. Dance, and L. Fan, "Categorizing nine visual classes using local appearance descriptors," in *In ICPR Workshop on Learning for Adaptable Visual Systems*, 2004.

[24] "Personal communication, niitek inc.," 2008.

[25] P. Torrione, C. Throckmorton, and L. Collins, "Performance of an adaptive feature-based processor for a wideband ground penetrating radar system," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 42, pp. 644 – 658, april 2006.

[26] R. Wu, A. Clement, J. Li, E. Larsson, M. Bradley, J. Habersat, and G. Maksymonko, "Adaptive ground bounce removal," *Electronics Letters*, vol. 37, no. 20, pp. 1250–1252, Sep.

[27] P. Gader, W.-H. Lee, and J. Wilson, "Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 11, pp. 2522–2534, Nov.

[28] F. Soldovieri, O. Lopera, and S. Lambot, "Combination of advanced inversion techniques for an accurate target localization via gpr for demining applications," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, pp. 451 –461, jan. 2011.

[29] L.-P. Song and Q. H. Liu, "Ground-penetrating radar land mine imaging: Two-dimensional seismic migration and three-dimensional inverse scattering in layered media," *Radio Science*, vol. 40, no. 1, pp. n/a–n/a, 2005.

[30] J. Song, Q. H. Liu, P. Torrione, and L. Collins, "Two-dimensional and three-dimensional nufft migration method for landmine detection using ground-penetrating radar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, pp. 1462 – 1469, june 2006.

[31] T. Counts, A. Gurbuz, W. Scott, J. McClellan, and K. Kim, "Multistatic ground-penetrating radar experiments," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, pp. 2544 –2553, aug. 2007.

[32] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257 –286, feb 1989.

[33] P. Scheet and M. Stephens, "A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase," *The American Journal of Human Genetics*, vol. 78, no. 4, pp. 629–644, 2006.

[34] M. Mohamed and P. Gader, "Handwritten word recognition using segmentation-free hidden markov modeling and segmentation-based dynamic programming techniques," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, pp. 548 –554, may 1996.

[35] D. K. Park, Y. S. Jeon, and C. S. Won, "Efficient use of local edge histogram descriptor," in *Proceedings of the 2000 ACM workshops on Multimedia*, MULTIMEDIA '00, (New York, NY, USA), pp. 51–54, ACM, 2000.

[36] S. Rahman, S. Naim, A. Al Farooq, and M. Islam, "Performance of mpeg-7 edge histogram descriptor in face recognition using principal component analysis," in *Computer and Information Technology (ICCIT), 2010 13th International Conference on*, pp. 476 –481, dec. 2010.

[37] P. D. Gader, R. Grandhi, W.-H. Lee, J. N. Wilson, and D. K. Ho, "Feature analysis for the niitek ground-penetrating radar using order- weighted averaging operators for landmine detection," in *Proceedings of SPIE*, vol. 5415, pp. 953–962, 2004.

[38] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, pp. 325–376, Dec. 1992.

[39] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, pp. 43–72, 2005.

[40] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Proceedings of the Alvey Vision Conference*, pp. 147–151, 1988.

[41] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, pp. 79–116, 1998.

[42] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Computer Vision ECCV 2010* (K. Daniilidis,

P. Maragos, and N. Paragios, eds.), vol. 6314 of *Lecture Notes in Computer Science*, pp. 778–792, Springer Berlin / Heidelberg, 2010.

[43] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, 2005.

[44] H. P. Moravec, *Obstacle avoidance and navigation in the real world by a seeing robot rover*. PhD thesis, Stanford, CA, USA, 1980. AAI8024717.

[45] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, pp. 15–33, 2000.

[46] V. Vapnik, *The nature of statistical learning theory*. springer, 1999.

[47] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, ICML '06, (New York, NY, USA), pp. 161–168, ACM, 2006.

[48] H. Wold, "Estimation of principal components and related models by iterative least squares," in *Multivariage Analysis* (P. Krishnaiah, ed.), New York: Academic Press, 1966.

[49] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[50] S. de Jong, "Simpls: an alternative approach to partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 3, pp. 251–263, 1993.

[51] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[52] P. Torrione, S. Keene, and K. Morton, *PRT: The Pattern Recognition Toolbox for MATLAB*, 2011. Software available at `http://newfolderconsulting.com/prt`.

[53] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[54] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, IEEE, 2008.

[55] A. Coates, H. Lee, and A. Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," *Ann Arbor*, vol. 1001, p. 48109, 2010.

[56] A. Coates and A. Ng, "Learning feature representations with k-means," *Neural Networks: Tricks of the Trade*, pp. 561–580, 2012.

[57] R. B. Girshick, *From rigid templates to grammars— Object detection with structured models*. PhD thesis, The University of Chicago, 2012.

[58] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Computer Vision ECCV 2006* (A. Leonardis, H. Bischof, and A. Pinz, eds.), vol. 3954 of *Lecture Notes in Computer Science*, pp. 490–503, Springer Berlin Heidelberg, 2006.