Advanced feature learning and representation in image processing for anomaly detection

By

Stanton Robert Price

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Electrical and Computer Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

May 2015

UMI Number: 1586997

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

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

UMI 1586997

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Copyright by

Stanton Robert Price

2015

Advanced feature learning and representation in image processing for anomaly detection

By

Stanton Robert Price

Approved:

Derek T. Anderson
(Major Professor)

John E. Ball
(Committee Member)

Nicolas H. Younan
(Committee Member)

James E. Fowler
(Graduate Coordinator)

Jason Keith
Interim Dean
Bagley College of Engineering

Name: Stanton Robert Price

Date of Degree: May 8, 2015

Institution: Mississippi State University

Major Field: Electrical and Computer Engineering

Major Professor: Dr. Derek T. Anderson

Title of Study: Advanced feature learning and representation in image processing for anomaly detection

Pages of Study: 96

Candidate for Degree of Master of Science

Techniques for improving the information quality present in imagery for feature extraction are proposed in this thesis. Specifically, two methods are presented: soft feature extraction and improved Evolution-COnstructed (iECO) features. Soft features comprise the extraction of image-space knowledge by performing a per-pixel weighting based on an importance map. Through soft features, one is able to extract features relevant to identifying a given object versus its background. Next, the iECO features framework is presented. The iECO features framework uses evolutionary computation algorithms to learn an optimal series of image transforms, specific to a given feature descriptor, to best extract discriminative information. That is, a composition of image transforms are learned from training data to present a given feature descriptor with the best opportunity to extract its information for the application at hand. The proposed techniques are applied to an automatic explosive hazard detection application and significant results are achieved.

Key words: image processing, soft features, object detection, pattern recognition

DEDICATION

To my family.

ACKNOWLEDGEMENTS

I would like to thank my closest friend and brother, Steven Price. His continuous guidance and support are an immeasurable blessing. Nothing quite like our coffee discussions or "getaway" trips. Those are times that I will forever cherish.

I would like to thank my family. You all have provided me with an incredible support system for as long as I can recall. I appreciate each and everyone of you; you are the inspiration of my work. To my # 1 buddy, Pops, who always saw and believed the best in me. He is the driving force behind my desire to work hard and put forth my best effort towards my graduate studies and research. Though he is no longer with us in the physical realm, his unwavering optimism towards life and cheerful heart will forever inspire me to be a better person. My family, whether knowingly or not, you all provided me with much needed relief and comfort along this journey, and I am thankful for each of you. For me, you all are the perfect equation.

I would like to thank my major advisor and mentor, Dr. Derek Anderson. Boy was I lucky when Steven mentioned to me that you were looking to hire a graduate student. Mutually important, that you gave me a chance to work under your guidance. Such a decision has had a far greater impact than I could have fathomed. The quality of your mentoring is rivaled by very few. Beyond this, you always allowed me to dream up new

ideas and avenues for approaching a research problem; never limiting or disregarding them, but rather helping to sculpt them. Thank you for this, and I look forward to our future work.

To my committee members, Dr. John Ball and Dr. Nicolas Younan. Thank you for agreeing to help guide me through this part of my graduate career. I appreciate your comments, suggestions, and guidance on this thesis.

This work was supported in part by the Army Research Office grants numbered W911NF-14-1-0114 and 57940-EV to support the U.S. Army RDECOM CERDEC NVESD. The findings and opinions in this thesis belong solely to the author, and are not necessarily those of the sponsor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

The work presented in this thesis is conducted in the context of advancing an automatic buried *explosive hazard detection* (EHD) system using imagery captured by a *forward-looking* (FL) *long-wave infrared* (LWIR) camera. The work herein could easily be extended to handle a vast array of object detection applications. Nevertheless, there is great need for a buried EHD system as such devices are often used by terrorists threat networks due to their relatively simple assembly and deployment. Furthermore, explosive hazard devices allow insurgents to inflict their desired damage from afar while remaining clear of any counterstrike that is possible when engaging in face-to-face combat.

Development of an automatic EHD system has the potential to save a number of lives from potentially fatal injury. Such a system also allows for route clearance; be it for military personnel moving from one site to the next, or for roads potentially traversed by civilians. Thus, there is a great need for an EHD system to detect buried explosive hazards so that appropriate measures can be taken, e.g., safe removal of the device. Due to the expense required to move such hazards, it is important for the system to have a low *false alarm rate* (FAR). The FAR is defined herein as the number of *false alarms* (FAs) per square meter. In the context of this work, a FA is the event in which the system identifies

an object as an explosive hazard, yet in reality it is not. Further, *weapons technical intelligence* (WTI) strategies for extracting forensic evidence can be utilized when the hazards are successfully detected and neutralized. The intelligence gained from thorough examination of these devices is very beneficial [2] in identifying those who make them and improving the understanding of the devices which leads to better identification techniques and route clearance.

There are a number of approaches currently being explored for EHD, i.e., sensor(s) used, features extracted, ground versus air vehicle-based, etc. Common sensors used include IR cameras, *ground-penetrating-radar* (GPR), and *electromagnetic induction* (EMI) sensors. This work was conducted in collaboration with the Army Research Office grants numbered W911NF-14-1-0114 and 57940-EV to support the U.S. Army *Research, Development, and Engineering Command* (RDECOM) *Communications-Electronics Research, Development and Engineering Center* (CERDEC) *Night Vision and Electronic Sensors Directorate* (NVESD) with a focus being on the applicability of IR for EHD. This thesis is concerned with the development of a highly robust, computationally efficient EHD system using FL-LWIR imagery. IR sensors allow for EHD of devices that are buried, i.e., does not require direct signatures of the explosive devices. Thus, IR imagery can be processed and used as an anomaly detector. As a result, IR is an appealing technology for use in combination with a FL approach to EHD. The biggest advantage (and attraction) for using FL cameras is that it enables EHD from an advanced standoff distance, i.e., distance between the vehicle and area captured by the camera's field of view.

In this thesis, the goal is to develop new image/signal processing techniques that help us solve the EHD problem (via anomaly detection). The (common) general framework used by the computer vision community for such a task is to perform feature extraction on the imagery and subject this information (encoded in the form of a feature descriptor) to a classifier, where each instance is labeled as one of M classes, i.e., class 1, class 2, ..., or class M. While there are many different avenues that can be explored to improve this general framework, I chose to focus my efforts on improving feature extraction. This was the result of my initial expectation, and later realization, that a classification system's performance is heavily (if not completely) dependent on the information (features) it is tasked with understanding and discriminating. Features attempt to provide a mechanism for characterizing/describing what is present in a given image. Therefore, if one can develop a method that will focus on features that are specific to an object of interest for some task, then one can expect that classification should benefit as a result. Development of such a methodology is the focus of this thesis.

1.2 Contributions

Herein, two main contributions to the applications of object detection via image processing in the context of an automatic buried EHD system are presented. These are briefly introduced in Sections 1.2.1 and 1.2.2, but will be covered in much greater detail in later Chapters.

1.2.1 Soft Features

In [45], I put forth an approach, coined *soft features*, that emphasizes features that directly describe an object of interest (i.e., buried explosive hazard devices). This is achieved by performing a per-pixel weighting, defined by an *importance map*, on each image chip that was extracted for further analysis by a pre-screener. Initially, the importance map was derived using a bank of *Gabor energy filters* (GEFs). However, in [46] the more sophisticated and appealing shearlet filters were implemented as a means for deriving the importance map. It was found that, for EHD in FL-LWIR imagery, the bank of GEFs could be replaced by a single shearlet filter. Soft features lead to the extraction of information that are more likely target specific. This allows a classifier to focus more on pertinent information in identifying explosive hazard devices versus being relied on to make sense of target and non-target (i.e., all) information. Publications on this contribution include:

[45]: **S. R. Price**, D. T. Anderson, R. H. Luke, K. Stone, and J. M. Keller, "Automatic detection system for buried explosive hazards in FL-LWIR based on soft feature extraction using a bank of Gabor energy filters," *SPIE Defense, Security, and Sensing*, 2013.

[46]: **S. R. Price**, D. T. Anderson, K. Stone, and J. M. Keller, "Investigation of context, soft spatial and spatial-frequency domain features for buried explosive hazard detection in FL-LWIR," *Proc. SPIE 9072, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets*, 2014.

1.2.2 iECO feature descriptors

The *improved Evolution-COnstructed* (iECO) features framework [44] enhances one of the state-of-the-art feature learning methods currently published, Lillywhite et al.'s ECO features [34]. This is accomplished by addressing two key shortcomings of the ECO framework (discussed in detail in Chap. 4). First, feature descriptors are implemented to extract actual features– the so-called ECO features are simply an unrolled image that has undergone a series of image transformations. It is my hypothesis that each feature descriptor employed has its own unique set of image transforms that can be learned using training data and these unique transforms would lead to the descriptors extracting maximal discriminative information for a given problem domain. Second, constraints are introduced on each individual's chromosome to promote population diversity and prevent infeasible solutions. It is shown through experiments that the proposed iECO framework results in, and benefits from, a unique series of transforms for each descriptor being learned and maintaining population diversity.

[44]: **S. R. Price**, D. T. Anderson, and R. H. Luke, "An improved evolution-constructed (iECO) features framework," *Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2014 IEEE Symposium on,* Dec 2014, pp. 1-8.

To some extent, these two methods are pursuing similar end-goals. That is, soft features are employed to extract an object of interest's defining characteristics and to "ignore" background, i.e., non-target information. Similarly, the iECO framework seeks to find an optimal composition of image transforms such that a given feature descriptor can best extract information for discriminating target from non-target. Therefore, soft features can be

viewed somewhat as a precursor to the iECO framework. Additional contributions made over the course of this Masters degree, but not discussed in this thesis are the following:

- Indices for Introspection on the Choquet Integral:
  [47] **S. R. Price**, D. T. Anderson, C. Wagner, T.C. Havens, and J. M. Keller, "Indices for Introspection on the Choquet Integral," *Advance Trends in Soft Computing*, Springer, 2014, pp. 261-271.

- Regularization-Based Learning of the Choquet Integral:
  [7] D. T. Anderson, **S. R. Price**, and T. C. Havens, "Regularization-based learning of the Choquet integral," *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, July 2014, pp. 2519-2526.

- Comparing Fuzzy, Probabilistic and Possibilistic Partitions Using the Earth Mover's Distance:
  [10] D. T. Anderson, A. Zare, **S. Price**, "Comparing Fuzzy, Probabilistic, and Possibilistic Partitions Using the Earth Mover's Distance," *Fuzzy Systems, IEEE Transactions on*, vol. 21, no. 4, 2013, pp. 766-775.

- **(Accepted)** Fusion of iECO image descriptors for buried explosive hazard detection in forward-looking imagery:
  S. R. Price, and D. T. Anderson, "Fusion of iECO image descriptors for buried explosive hazard detection in forward-looking imagery," *SPIE Defense + Security*, 2015.

- **(Accepted)** Design of a buried explosive hazard pre-screener in forward looking imagery based on shearlet filtering and image post-processing:
  S. R. Price, D. T. Anderson, and J. M. Keller, "Design of a buried explosive hazard pre-screener in forward looking imagery based on shearlet filtering and image post-processing," *SPIE Defense + Security*, 2015.

- **(Accepted)** Learning of an explosive hazard detection pre-screener in forward looking imagery based on adaptive mutation:
  R. Singh, **S. R. Price**, and D. T. Anderson, "Learning of an explosive hazard detection pre-screener in forward looking imagery based on adaptive mutation," *SPIE Defense + Security*, 2015.

The remainder of this thesis is organized as follows. Soft features and their impact are discussed in Chapters 2 and 3. In Chapter 4, the iECO features framework along with performance analysis is presented. Finally, a conclusion for this thesis is made in Chapter 5.

CHAPTER 2

SOFT FEATURES - GABOR

2.1 Introduction

Buried explosives are commonly used by insurgents because they can be relatively inexpensive and simple to make. In addition, these devices allow threat networks to inflict their desired damage from a distance while remaining clear from a possible counterstrike that exists in face-to-face combat. The development of an automatic buried *explosive hazards detection* (EHD) system can help save many lives that would otherwise be subject to injury, sometimes fatal. With such a system in place, these hazards can be removed and the threat negated. However, due to the expense required to remove such hazards, it is important for the system to have a low FAR. In the case of a ground vehicle, a FA results in the need to dig up an area of earth, which in return means lost time and ultimately a lower route clearance rate. Further, *weapons technical intelligence* (WTI) strategies for extracting forensic evidence can be utilized when the hazards are successfully *detected* and neutralized. The intelligence gained from thorough examination of these devices is very beneficial[2] in identifying those who make them and improving the understanding of the devices which leads to better identification techniques and route clearance. The prevalence of these buried explosives is on the rise as their use has "*increased 42 percent, from 9,300 in 2009 to 16,000 in 2011*"[1] in Afghanistan alone. The growing threat of explosive haz-

ards deployed not only overseas but also potentially in our homeland provides our group with the motivation to *develop a highly robust, dependable, and low-cost EHD system*.

There are a number of approaches currently being explored for EHD, i.e., sensor(s) used, features extracted, ground versus air vehicle-based, etc. The most common sensors used include IR cameras, GPR, and EMI sensors. The focus of this Chapter is the development of a highly robust, computationally efficient EHD system for FL-LWIR imagery. FL has its advantages as well as disadvantages: two key advantages are that explosive hazards can be detected at greater distances from the vehicle and a system also gets, in general, multiple looks on targets. Multiple looks on targets allows for the accumulation of target evidence, potentially leading to a more dependable/robust EHD system. A disadvantage is it sees everything. Of course, this depends on a sensors *field of view* (FOV), focal length, location on earth, etc. As a result, most pre-screeners typically have a higher FAR. A goal of our research group is to develop new algorithms within and fusion across FL-*electro-optical* (EO)/IR and FL-GPR for direct detection or *region of interest* (ROI) detection for another technology like *downward looking* (DL) GPR.

2.2 Related Work

A number of technologies and algorithms are being investigated for EHD. In this section, a few related approaches are discussed. In Anderson et al.[9], a new method was put forth to combine evidence from multiple anomaly detection algorithms with different image space and *universal transverse Mercator* (UTM) space features for EHD in FL-LWIR. Specifically, they explored the use of the *maximally stable extremal region* (MSER) al-

gorithm, *Gaussian mixture models* (GMMs), and an ensemble of trainable size-contrast detectors. These anomaly detection algorithms operate in very different respects. Their fusion represents a combination of per-frame and temporal (down-track) change detection. Next, the anomaly algorithm evidence is temporally aggregated using clustering (weighted mean shift clustering) and features are extracted using a combination of their results and the corresponding raw LWIR imagery. Extremely encouraging results were demonstrated and it was found that significant FAR reduction was the main benefit over any one particular algorithm used.

Yuksel et al. [66] recently put forth an approach for the classification of ambiguous time-series data using *multiple instance hidden Markov models* (MI-HMMs) for EHD from GPR. Their approach is the combination of two concepts. Specifically, HMMs for time series data and MI learning for learning from ambiguous data. Yuksel proposed that MI-HMMs are capable of modeling the *temporal* properties of their data. Training data was selected using a Markov Random Field (MRF) "bounding box" (used to find the target). The MI-HMM was trained by placing five evenly spaced sequences from within the MRF bounding box into positive bags and five randomly selected non-target sequences were placed into negative bags. Classification results from testing the MI-HMM showed improvements over the standard HMM approach.

In Anderson et al.[8], the aggregation of an ensemble of MSERs and its fusion with the output of a GMM was explored. The motivation is that different MSERs, i.e., different sets of parameters, are useful for detecting targets of different difficulty. Too *weak* of conditions (parameters) result in the detection of a large number of things. Specifically, weak

conditions detect a large amount of clutter, but more importantly, the hard-to-detect targets. A *stronger* set of parameters however results in less clutter but also fewer actual targets. The investigation was how to identify and aggregate evidence across a set of MSERs to improve the overall robustness of the system. In addition, they also explored the use of *contrast limited adaptive histogram equalization*[41] (CLAHE) for local image enhancement to aid the MSER. Once again, in comparison to [9], the main benefit was found to be significant FAR reduction versus the identification of new targets.

In addition to the works discussed above, a number of other noteworthy methods exist (for GPR [4, 21, 24, 27, 26, 28, 48, 50, 60, 61, 62, 30], IR [6, 43, 42, 51, 53], and acoustic technologies [56, 64]). The point is, there are a number of ways to sense buried explosive hazards and no single source is perfect. Multi-sensor fusion and/or context-dependent identification and classification will likely be needed to see EHD systems flourish. While many potential sources of evidence exist, this thesis focuses on just EO/IR.

2.3 Sensor and Data Sets

The FL-LWIR sensor used is a U6000 (DVE) Camera by DRS Infrared Technologies. The camera has an uncooled core and it was mounted with a fixed mast on a *high mobility multipurpose wheeled vehicle* (HMMWV). The camera was focused on a distance of approximately 10 meters from the mast. Detection of spectral responses from 8-12 μm are captured at a frame rate of fifteen frames per second and stored as 8-bit images with a resolution of 640$\times$480. Mounted next to the camera is a differential GPS receiver and

inertial measurement unit, which in conjunction with the times at which each image was captured, allowed accurate geo-referencing to be achieved[52].

For real-world deployment, an EHD system first-and-foremost needs to be highly accurate and reliable, but is made more *practical* if the targets are detected at a further distance away from the end-user (i.e., the vehicle). In addition, it needs to be robust and applicable to various terrains rather than designed for a specific environment. IR can be used for FL imaging and it is well known that all objects with a temperature above absolute zero emit IR radiation in the LWIR (also known as thermal) portion of the electromagnetic spectrum. The amount of radiation emitted by an object increases with temperature and the amount of said radiation that is thermal radiation depends on an object's emissivity. An object's emissivity, which varies with wavelength, is a value representing its ability to emit thermal radiation. Thermal cameras capture not only the emitted radiation of the object, but also transmitted radiation, radiation from an external source which passes through the object toward the camera, and/or reflected radiation, radiation from an external source which reflects off the object toward the camera. However, detection of buried explosives is not critically impaired by these factors, as they complicate assigning absolute temperature values to objects and this is not of great importance to the detection of buried explosives. A large hindrance is the so-called diurnal cross-over period in which the soil and buried objects are in thermal equilibrium. Another disadvantage is the difference in emitted radiance seen at the soil surface (even for the same soil composition and object) varies based on factors such as the amount of incident thermal radiation, which depends on the time of day, time of year, and current weather conditions.

The data set used in this Chapter consists of 20 runs from two lanes at an arid United States Army test site: 10 runs on *lane A* and 10 runs on *lane B*. The data were collected over a nine day period and includes runs during two times of the day and in two directions, morning/afternoon and east/west, respectively. Of the 10 runs per lane, 5 are morning and the other 5 are afternoon and 5 are for when the vehicle was heading east and the other 5 heading west. Each run was captured continuously as the vehicle traversed the entire length of the lane. Weather conditions were clear and typical for the region during the data collection, though it did rain in the middle of the collection period. Lane A has 44 buried explosive hazards, i.e., *targets*, and lane B has 50. The targets have various levels of metal content and burial depths ranging from shallow to deep. Table 2.1 is a summary of the data collection for lane A and lane B. Table 2.2 is a summary of the distribution of the targets for this data collection based on its metal content and burial depth. All targets are buried in the road, within approximately 2 meters cross-track either side of the vehicle. A pre-determined road mask corresponding to approximately 3 meters to either side of the vehicle and 5-30 meters down-track is used when processing this data collection. However, see [33] for our recent work on road detection which will be used to replace this manually selected road mask in the near future. In addition, the method for calculating lane area is described in Stone et al.[54]. It is necessary for normalizing the FAR so that the *receiver operating characteristic* (ROC) curves generated are in terms of FAs per meter squared.

Table 2.1

Data collection summary for lane A and lane B.

			Lane A	Lane B
Collection Day	Time of Day	Direction	Area (m^2)	Area (m^2)
1	Afternoon	West	7189.73	7232.17
2	Morning	East	7275.86	7277.15
2	Afternoon	East	7252.11	7270.67
3	Morning	West	7211.82	7256.75
3	Afternoon	West	7182.21	7195.96
7	Morning	West	7215.70	7263.49
7	Afternoon	West	7217.15	7258.06
8	Morning	East	7282.94	7237.67
9	Morning	East	7247.02	7280.96
9	Afternoon	East	7261.31	7276.04

Table 2.2

Distribution of targets according to metal content and burial depth.

	Lane A	Lane B
Metal Content		
Low metal	$\approx 57\%$	$\approx 56\%$
High metal	$\approx 43\%$	$\approx 44\%$
Burial Depth		
Shallow	$\approx 61\%$	$\approx 60\%$
Middle	$\approx 12\%$	$\approx 12\%$
Deep	$\approx 27\%$	$\approx 28\%$

2.4 Methods

Figure 2.1 is an illustration of the proposed EHD system.

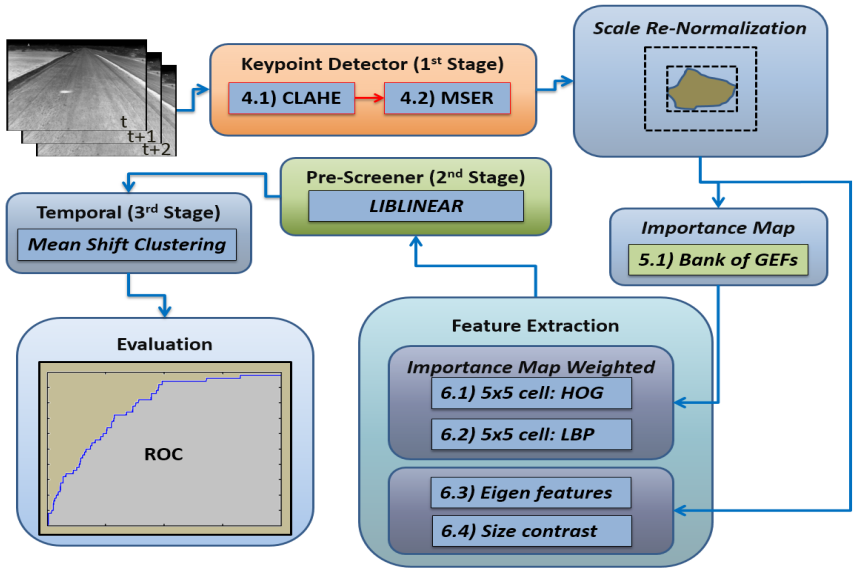


Figure 2.1

Proposed FL-LWIR EHD system.

Framework consists of low-level keypoint detection, feature extraction and classification (per-frame), and temporal pattern analysis (i.e., multi-look aggregation).

This system consists of three stages.

- *(Stage 1) Keypoint identification*: This stage is responsible for finding interesting image space regions across translation, rotation and scale. It should do this without making any assumptions regarding the shape and size or orientation of a target. Its goal is to find all possible targets while minimizing the number of FAs as much as possible but not at the expense of the *positive detection rate* (PDR).

- *(Stage 2) Pre-screener (classifier)*: The goal of this stage is to reduce stage 1 hits based on the use of advanced image space features and pattern classifiers. This stage is performed at each image.

- *(Stage 3) Temporal aggregation (clustering)*: The goal of this stage is to combine evidence across frames, e.g., time. Linking hits helps both in terms of accumulating target evidence (i.e., multiple strong detections of the same object over time) as well as further FAR reduction.

2.4.1 Image Pre-Processing

In Anderson et al.[8], our group explored the inclusion of the CLAHE technique for EHD to enhance local contrast (i.e., detail) in LWIR imagery. The primary purpose for adding CLAHE in that work was to enhance keypoint detection. Herein, I explore the benefit of extracting features from the CLAHE enhanced imagery. CLAHE is used to emphasize image content for keypoint detection of difficult targets and to also aid in the extraction of useful features. The dynamic range of 8-bit uncooled LWIR is limited for EHD. A goal of the first stage is to find all potential targets, at the expense of a relatively high FAR (but generally better than a brute force or random search strategy). Global histogram techniques are overly sensitive to scene-level detail and often wash away local detail needed to find keypoints. Instead, I use a local image enhancement technique to find all keypoints. Full details regarding the theory and implementation of CLAHE can be found in[39]. Next, the keypoint detector for feature extraction is discussed.

2.4.2 Maximally Stable Extremal Regions: Keypoint Detector

Due to the large amount of candidate image regions to explore in a FL-EHD system, i.e., all possible translations, rotations and scales, I make use of a low-level image processing technique to identify ROIs from each image frame and I subsequently investigate only these regions for EHD. Note, no assumptions regarding size, shape or orientation are

made here. In contrast to our groups previous work, which implemented complex pre-screeners (see Anderson et al.[9]), I propose a relatively simple, but effective and efficient, pre-screener that uses a single MSER keypoint detector. The goal of this single MSER keypoint detector is to identify 100% (or as high as possible in FL-LWIR) of the *true targets* (TTs) while limiting the number of instances moved further through the EHD pipeline. This strategy reduces computational costs and time required, complexity of the classifiers, as well as the cost of equipment required for the system to be deployed in a real-world scenario. MSERs, in layman terms, are nothing more than regions, referred to as *blobs*, in an image that are either darker or brighter than their surroundings (exhibit sufficient contrast) and meet a stability criterion (and possibly other additional constraints and/or heuristics) selected by the user or learned from data. It is important to note the following. Due to the definition of a MSER, multiple detections (instances) of approximately the same region (e.g., nested blobs) can be (and often are) detected. A detailed overview of MSERs can be found in [8, 23, 59]. For the remainder of this Chapter, attention is focused on the application of MSERs as it pertains to EHD rather than its theory and implementation.

Once keypoints are found, focus shifts to feature extraction. Herein, a number of shape and texture features are extracted from each blob (referred to as an *instance* herein). These features are combined into a single long concatenated feature vector. During training, the UTM location of each instance's centroid is compared to a ground truth. If the hit is within 0.5m of a ground truth location, that instance is labeled as a TT. Otherwise, the point is labeled as a false hit (i.e., clutter). Because of factors such as inaccuracies in image-to-world projection (especially when based on the use of projections that assume

a flat-earth assumption, such as the experiments reported herein), difference between a computed centroid of a blob and the recorded ground truth location, and number of pixels on target (which drops off quickly in the down-track direction), I only keep blobs in an ideal detection range to the vehicle (range of [4-10]m used herein). Note, this range depends on the resolution of the camera and its center of focus. Thus, a better camera looking 30 meters out will use a different meter interval. The point is that this monitor region is associated with image quality (e.g., number of pixels on target) and image-to-world correspondence.

It is important that I now establish how a MSER blob is prepared for feature extraction. As previously stated, the goal of the pre-screener is to identify as high of a PD as possible. To this end, I empirically select MSER parameters that are not very *strict* (allowing for the detection of weakly expressed targets). At the same time, I obviously desire a set of MSER parameters that ignore as much clutter as possible (which reduces the FAR). While empirically selected here, these parameters could be learned from data. However, the goal of this work was to see if I could empirically pick a set of parameters initially to find weakly expressed blobs vs resorting to sophisticated machine learning techniques (which are all too often subject to overfitting and likely suffer from a lack of desirable generalizability).

For each MSER blob, the minimal *axis aligned bounding box* (AABB) is calculated. It is specified in terms of upper left, (UL_x, UL_y), and lower right, (LR_x, LR_y), blob coordinates. This box is expanded (empirically chosen ratio) by

$$\Delta_h = \left\lceil \frac{h * 6}{13} \right\rceil, \tag{2.1}$$

$$\Delta_w = \left\lceil \frac{w * 6}{13} \right\rceil, \tag{2.2}$$

where Δ_h and Δ_w is the amount of pixels to expand (UL_x, UL_y) and (LR_x, LR_y) by in each x and y direction, respectively. The resulting (expanded) box, $\widehat{AABB}$, was obtained in order to get a ROI that contains the supposed target, its boundary, and its immediate surrounding context.

While the MSER is useful for detecting keypoints across scale, rotation, and translation, affine invariance at the feature extraction level is not as simple. In order to place features on a level playing field, I re-scale $\widehat{AABB}$ to size 125×125 via bi-linear interpolation. This is due to the following reasons. First, I use cell-structured image descriptors to preserve spatial context of features. Re-sampling ensures an even samples rate for histogram population. Second, when extracting information such as image gradients, it is of benefit to have the gradients encode very similar (in scale) features. Algorithm 1 details the keypoint detector and how it prepares *each instance* for feature extraction. At this point, $\widehat{AABB}$ is ready for feature extraction.

Algorithm 1 Per-frame LWIR image keypoint detector.

Determine (pick or learn) set of MSER parameters, $(\Delta_{MSER}, a_+, a_-, v_+, d_-)$, where, for a candidate MSER, a_+ and a_- are the max and min areas respectively, v_+ is max variation, and d_- is min diversity

for each image, I_t, **do**

    Obtain road mask R_t ▷ note: a single road mask is used here

    Find all MSERs in I_t in R_t ▷ note: I use VLFeat's MSER implementation

    for each MSER *blob*, x_i **do**

        if x_i centroid is within a desired range ($[4 - 10]$m used herein) of the vehicle **then**

            Find the minimal axis aligned bounding box, $AABB$ ▷ described by (UL_x, UL_y) and (LR_x, LR_y)

            Compute $w = LR_y - UL_y$ ▷ where w is the width of $AABB$

            Compute $h = LR_x - UL_x$ ▷ where h is the height of $AABB$

            Compute Δ_h and Δ_w via Eqns. (2.1)-(2.2)

            Compute $(\widehat{UL_x}, \widehat{UL_y}) = (UL_x, UL_y) - [\Delta_h, \Delta_w]$

            Compute $(\widehat{LR_x}, \widehat{LR_y}) = (LR_x, LR_y) + [\Delta_h, \Delta_w]$

            Re-size $\widehat{AABB}$ to size 125×125 via bi-linear interpolation

        end if

    end for

end for

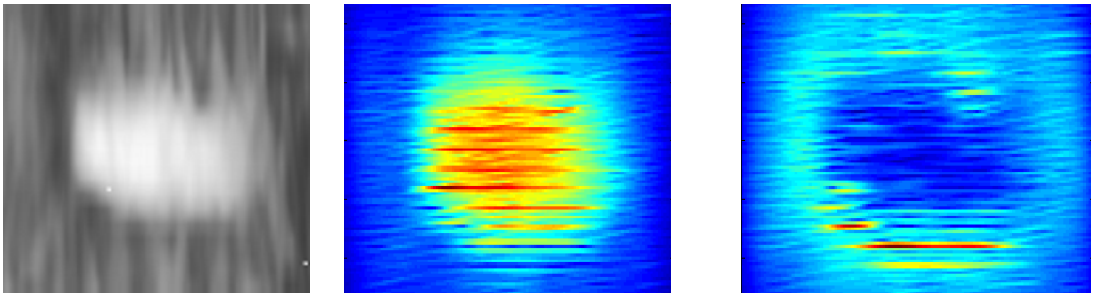
2.5 Importance Map

In Scott et. al.[49], it was shown that importance map weighted features can significantly outperform un-weighted features in automatic target recognition of aerial vehicles in remotely sensed data. Herein, I explore this technique for EHD. The motivation is that I want to focus on learning just features in an image chip, i.e., $\widehat{AABB}$, that are more likely target specific and not background/context. Thus, I desire the identification of image structure and features that transfer better to other new scenes than what was included in training. Herein, I provide evidence that shows that this approach leads to improved performance of an EHD system. However, this is difficult to truly do herein as the range within the data set is rather limited (i.e., not for two different continents as was the case in the remotely sensed work). Specifically, through the use of a bank of *Gabor energy filters* (GEFs), an importance map is derived for an image chip that is used for weighted feature extraction. In prior work, differential morphological profiles and fuzzy integrals were used to acquire an importance map. However, that approach, in calculation not concept, is insufficient for the EHD case studied here. In the next sections I detail importance map calculation and associated soft feature extraction.

2.5.1 Gabor Energy Filter Bank

In FL perspective imagery, targets generally appear elongated in the cross-track direction. The degree of enlongatedness depends in part on a targets distance from the vehicle, its shape, the terrain, and factors such as the imaging sensor (e.g., FOV, focal plane, lensing, etc.). Herein, I make use of orientation and spatial frequency (still relatively weak

conditions) of desired targets in FL perspective imagery. Specifically, a set of GEFs are used to find desired target evidence at various spatial frequencies and orientations. Namely, those that support the concept longer in the cross-track or near circular in shape (but not longer in the down-track direction). Figure 2.2 shows the use of a GEF bank to detect target evidence.



(a) CLAHE chip (b) GEF bank return for (a) (c) GEF bank return for inverse of (a)

Figure 2.2

GEF bank aggregation example.

GEF aggregation for (a) is shown in (b). The GEF aggregation for the inverse of (a) is also shown, in (c), which aids in the detection of both light-on-dark and dark-on-light blobs. Note, (b) and (c) are normalized (individually) between min and max for display.

The Gabor filter is the product of a Gaussian kernel (i.e., envelope) times a complex sinusoid (i.e., carrier). The Gabor filter is complex valued, thus it has both a real and imaginary part,

$$g_{Re}(x,y) = exp(-(\frac{\tilde{x}^2 + \gamma^2\tilde{y}^2}{2\sigma^2}))cos(2\pi\frac{\tilde{x}}{\lambda} + \psi), \tag{2.3}$$

$$g_{Im}(x,y) = exp(-(\frac{\tilde{x}^2 + \gamma^2\tilde{y}^2}{2\sigma^2}))sin(2\pi\frac{\tilde{x}}{\lambda} + \psi), \tag{2.4}$$

$$\tilde{x} = xcos\theta + ysin\theta, \tag{2.5}$$

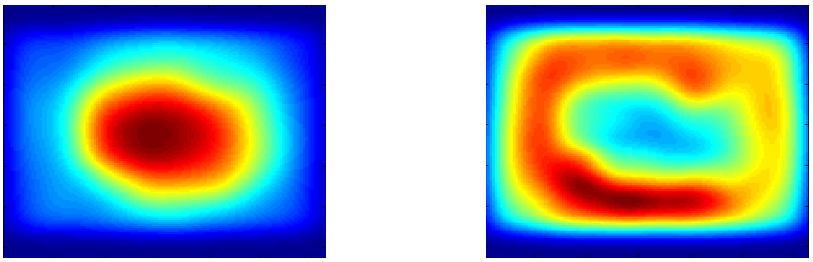
$$\tilde{y} = -xcos\theta + ysin\theta, \tag{2.6}$$

where σ is the standard deviation of the Gaussian function, λ is the wavelength of the sinusoidal function, θ is the orientation of the normal to the parallel stripes of the Gabor function, γ is the spatial aspect ratio, and ψ is the phase shift. To obtain a response that is phase insensitive, I use the GEF, denoted as E,

$$E = \sqrt{g_{Re}(x,y)^2 + g_{Im}(x,y)^2}. \tag{2.7}$$

The GEF bank is applied to both the image chip and its inverse to account for time of day and its effects on the relative thermal signature returned for the blob; that is, the cases of bright-on-dark and dark-on-bright relative contrast. The bank is aggregated by taking the maximum return at each pixel across all of the filters in the bank. The aggregated GEF return is then normalized by dividing each pixel by the maximum value of the GEF. Hits are almost always center aligned in the image chips (i.e., the center of a MSER blob). Therefore, a patch (21×21 used herein) of the center of the GEF return is summed to select a "winner." The "winner" tells which GEF return (i.e., original vs. inverse) to use as the importance map. To ensure that I do not lose the edges of targets in the importance map (a

feature that can be of vital importance), a Gaussian blur is applied to the GEFs to spread out the evidence and capture edge regions. Figure 2.3 illustrates this technique using the same image chip as Figure 2.2a.



(a) Filtered GEF bank return (b) Filtered GEF bank return for inverse

Figure 2.3

Gaussian blur applied to GEF bank return for the image chip shown in Figure 2.2a.

This illustration shows how the blurring helps spread out evidence and places importance (weight) on both target and target edge regions, but not surrounding context/background.

2.6 Feature Extraction

In this section, I describe the image features extracted from each instance found by the keypoint detector. The features extracted include *histograms of oriented gradients* (HOG) [37], *local binary patterns* (LBP) [40], principle components and Bhattacharyya distance on a size contrast filter. The HOG and LBP characterize texture, principal components describe approximate shape and orientation, and the Bhattacharyya distance is a measure of divergence between relative intensity amounts of a size contrast filter centered on the image chip. Of these features, the HOG and LBP are computed using a 5×5 (non-overlapping)

cell-structured configuration (which preserves spatial context) and they are weighted via the image chip's importance map. To be clear, *only* the HOG and LBP features are importance map weighted.

2.6.1 Importance Map Weighted Histogram of Oriented Gradients

The HOG descriptor captures the texture of an image according to its image gradients, see [37, 36] for the HOG calculation. I sub-divide an image chip into a 5×5 non-overlapping set of equally sized squares. This results in 25 cells, each with an 8-bin histogram, resulting in a concatenated descriptor with total length 200 (note, the David Lowe approach used a 4×4 cell-structured descriptor[37, 36]). I picked the configuration through experimentation. From the importance map, the pixel-weighted magnitude of each gradient is proportionally contributed to a primary and secondary bin for each gradient using the alpha independent scheme of Lowe. Distances between the primary bin center and the secondary bin center determines how the gradient magnitude is distributed with its direction determining the primary bin. Thus, only image gradients around strong areas of target evidence are used to populate the descriptor, and background has a lesser effect.

2.6.2 Importance Map Weighted Local Binary Pattens

The LBP is another feature that roughly captures a notion of image texture. A benefit of the LBP is that it is relatively good at distinguishing edge curvature characteristics of an object (i.e., target, clutter) and its surroundings. The LBP is also invariant to uniform scaling in intensity. The LBP code is

$$LBP_n = \sum_{k=0}^{n} (s(i_k - i_c)2^k), \tag{2.8}$$

where i_c is the center value, i_k is the value of the k^{th} neighbor, and function $s(x)$ is

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

$LBP^u_{n,r}$ represents the uniform, circular neighborhood version of the LBP which samples n points at radius r and has no more than u 0-1 transitions (i.e., uniform pattern). I compute $LBP^2_{8,3}$ for each cell as a summed pixel-weight normalized histogram, resulting in 57 features per cell. Note, there are typically 59 features but I remove the last two patterns as they contribute little-to-nothing to the task at hand. I use the same 5×5 non-overlapping boxes (cells), which results in a 1425 length concatenated descriptor.

2.6.3 Eigen Features

The next two features approximately characterize a targets shape and orientation. This is useful for exploiting the general cross-track and elliptical shaped behavior of targets in FL perspective imagery for ground vehicles. First, the eigenvalues and corresponding eigenvectors are extracted from the MSER. Let Z be the set of pixels in the current MSER (hit) i. Specifically, Z is the set of respective image space coordinates. First, I estimate the covariance of Z,

$$C = \frac{\sum_{k=1}^{|Z|}(z_k - \bar{z})(z_k - \bar{z})^\intercal}{|Z| - 1}. \tag{2.9}$$

The eigen information is extracted from C, which for image space coordinates (two dimensional) is 2 eigenvalues and 2 eigenvectors, i.e., λ_1, λ_2, $\vec{\mu_1}$, and $\vec{\mu_2}$ (where $\lambda_1 \geq \lambda_2$). I

calculate the similarity of the eigenvector with maximum eigenvalue (λ_1) with the image width, cross-track, dimension

$$v_1 = |(1,0) \cdot \vec{\mu_1}|. \tag{2.10}$$

The absolute value is taken as the eigenvector could be either $(1,0)$ or $(-1,0)$. Next, I compute the ratio of the two eigenvalues,

$$v_2 = \frac{\lambda_1}{\lambda_2}. \tag{2.11}$$

Thus, if the target has a circle shape then $v_2 = 1$ and if it is elongated $v_2 > 1$. Together, these two features inform us about the approximate shape and orientation of a potential target.

2.6.4 Bhattacharyya Distance

The last feature discussed is the Bhattacharyya distance. This measure expresses the divergence between two distributions. For this application, an inner box (AABB on the MSER blob) and outer box proportional to the inner AABB is used (the outer box is the area I expanded from $AABB$ to $\widehat{AABB}$ in Algorithm 1). Image chips, whether the target is strongly or weakly expressed, show some amount of contrast between its surroundings. The Bhattacharyya distance measures the contrast between these two distributions[25]. The Bhattacharyya calculation used here is

$$D_B = \frac{(\mu_a - \mu_b)^2}{4(\sigma_a^2 + \sigma_b^2)} + 0.5 * ln(\frac{\sigma_a^2 + \sigma_b^2}{\sqrt{\sigma_a^2 \sigma_b^2}}), \tag{2.12}$$

where (μ_a, σ_a) and (μ_b, σ_b) are the means and variances of the outer and inner boxes respectively.

2.6.5 Feature Concatenation

In order to combine the different features and perform classification, I take a conventional approach of concatenating all features into a single vector. Let $\vec{F_i} = \{f_{i,1}, ..., f_{i,N}\}$ denote a $1 \times N$ feature vector (which corresponds to a single instance, or blob, x_i) with N features (i.e., HOG, LBP, eigen and Bhattacharyya). The feature vector is the following:

$\vec{F_i} = \{f_{i,1}, ..., f_{i,5}\}$ where

$f_1 \rightarrow v_1$ (feature length:1) ,

$f_2 \rightarrow v_2$ (feature length:1) ,

$f_3 \rightarrow$ Weighted HOG (feature length:200) ,

$f_4 \rightarrow$ Weighted LBP (feature length:1425) ,

$f_5 \rightarrow$ Bhattacharyya Distance (feature length:1) ,

which has a combined total length of 1628. Note, most of the feature dimensionality is the LBP feature.

For each run per lane, a set $X_{j,p} = \left\{\vec{F_1}, ..., \vec{F_{k_{j,p}}}\right\}$ is created where j represents the lane processed (i.e., A or B), p represents the run number (i.e., 1,2,...,10 for this data set), and $k_{j,p}$ is the total number of $\vec{F}$ for the particular run. Before moving on to classification, it should be noted that during the process of creating the set $X_{j,p}$, I also create the label vector, $L_{j,p}$. Instances that are a TT are assigned a label of 1 and FAs are assigned a label of -1.

2.7 Classification

The material described up to this point has been the identification and characterization of image features. However, the goal is to realize an automatic EHD system so a decision must now be made with respect to each hit, i.e., target or not target. While a vast number of classifiers have been proposed to date, herein I utilize *support vector machine* (SVM) based classification. With respect to a SVM, one, in general, must separate their data into different classes (two herein). One must also pick a kernel function, and depending on the type of SVM, a range of different parameters. For example, much effort has recently been placed on the topic of so-called multiple kernel learning for SVMs as a method for improved classification and feature selection[29]. Due to the nature of the pre-screener, I incur a relatively large number of training data-points. To handle this large data set, I use the linear classification package LIBLINEAR[20], an efficient linear classifier[20, 65]. Training sets, depending on a number of factors (i.e., MSER parameter selection, range from vehicle, etc.) could have anywhere from 52,000 to upwards of 130,000 or more instances. What further complicates this problem is the following. In general, the number of FAs dwarf the number of TTs. This can (and does) affect a number of classifiers (gross class size imbalance). In addition, one generally gets a relatively small number of rich samples that likely do not cover a range of possibilities such as different soil types, different atmospheric conditions, different times of day, burial depth, etc. EHD data sets also typically contain a relatively small number of instances, if not just one, of target types. This affects a great deal, such as sparse patterns (*clusters*) buried in a sea of FAs. Therefore, the rationale for selecting particular features and training data is very important. Machine

learning algorithms are only as good as the features and training examples. As a result, I detail a rather lengthy experimental section with some sensitivity tests to illustrate the quality of the proposed system. In particular, the SVM is applied to the keypoint hits and it is used as a second stage filter to find candidate targets per-frame. For the experiments discussed, only two LIBLINEAR parameters are selected: the type of solver, -s, and the cost parameter, -c. The solver used is -s 4, the multi-class support vector classification by Crammer and Singer, and the cost parameter used is -c 10 (both were empirically chosen).

2.8 Clustering

A so-called advantage and selling point of FL imagery technologies is that they get multiple looks on target. Thus, a system is not limited to any one look on target. Instead, one can image a target from multiple viewpoints and that evidence can be combined. Herein, I make use of multi-look in a rather simple, but effective regard. Clustering is used herein to group hits across frames and link decisions. An advantage of my approach is its robustness to scale (i.e., ability to detect and make decisions over a distance range) and computational efficiency. I use mean shift clustering to "clump" (cluster) instances together based on its mapped UTM location. With mean shift clustering, I am looking for modes in the data, not partition it (e.g., k-means, fuzzy c-means, etc.). An advantage of clustering is that it prevents redundant counting of both TTs and FAs, providing a true PDR and FAR when testing the data (i.e., does not count the same target or the same FA multiple times, which would "corrupt" (incorrectly increase) the PDR and FAR). The radius used by mean-shift

in my approach is 0.5 meters. In its current state, this is non-causal, but is easily made causal by (re)running the clustering algorithm at each frame (i.e., iterative updating).

2.9 Preliminary Results

In this section, preliminary results are presented that are extremely encouraging for FL-EHD. Let $X_A = \{X_{A,1}, ..., X_{A,10}\}$ and $X_B = \{X_{B,1}, ..., X_{B,10}\}$ be the sets that encompass all of the instances from each run for lane A and B respectively, found by the keypoint detector. Testing was performed using lane-based *cross validation* (CV), more specifically, 2-fold CV in which X_A is used for training the SVM and testing is performed on each run of lane B individually. This process is repeated with training on X_B and testing on each individual run of lane A. Instead of co-plotting multiple ROCs on the same figure, which is very difficult to view, I vertically average the different individual ROCs for each lane tested.

An important FL-EHD question is, how far away from the vehicle can one detect buried explosive hazards at an acceptable PD/FAR? Herein, a number of experiments are performed to investigate system performance by varying the standoff distance (specifically, a set or range of different standoff distances). This helps one better understand the sensitivity of a system with respect to inclusion of multiple *scales*. Specifically, is it better to use a small standoff range interval, for which targets are at approximately the same scale, or can one instead perform detection over a wide standoff distance? The first helps to fix scale of hits, while the latter has the advantage of providing more looks on target. Table 2.3 details the range intervals used in these experiments. Note, while these specific intervals are tied

to my particular setup, what is of use is the following. Should one, in general, use a small range interval (thus similar scale) or wider range interval? Moreover, is there a point of diminishing return? The vertically averaged ROC curves for lane A and lane B are shown in Figure 2.4 and Figure 2.5. Note, a human performed manual inspection of this data set and an expert was able to find approximately 85% of the targets in FL-LWIR.

Table 2.3

Different standoff distance ranges investigated herein.

Identifier	min distance (m)	max distance (m)
Exp_{0-10}	0	10
Exp_{4-10}	4	10
Exp_{4-7}	4	7
Exp_{8-14}	8	14

The ROC curves in Figures 2.4 and 2.5 tell the following story. First, both lanes have better performance when the standoff distance interval is relatively large, e.g., $[0, 10]$m and $[4, 10]$m (vs. $[4, 7]$m). This reinforces what one might suspect. A system prefers to have multiple looks on target. However, the specific interval is important. That is, $[4, 10]$m and $[8, 14]$m are both 6 meters wide; however, the target is much harder to detect at further distances. This is not to say that $[4, 10]$m is the *best* interval. For this camera and experimental setup, it just happens to be the correct standoff distance in which the targets are in focus and there are enough pixels on target. Second, the system appears to not vary much, if at all, between $[0, 10]$m and $[4, 10]$m. This shows that there is indeed a point,

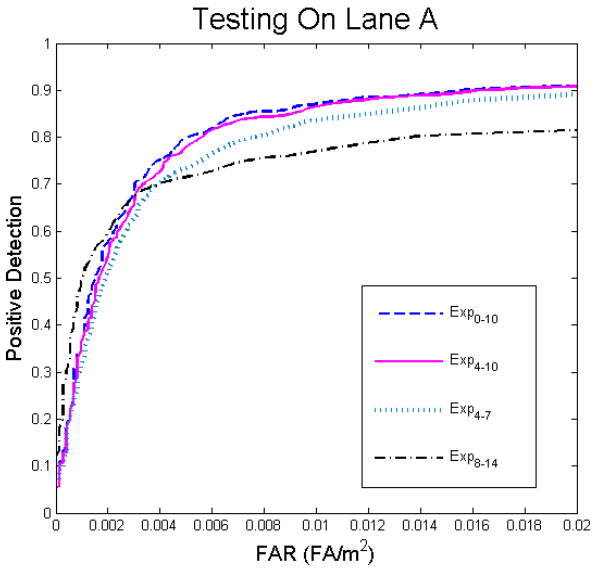


Figure 2.4

ROCs for lane-based CV with respect to different stand off distance intervals.

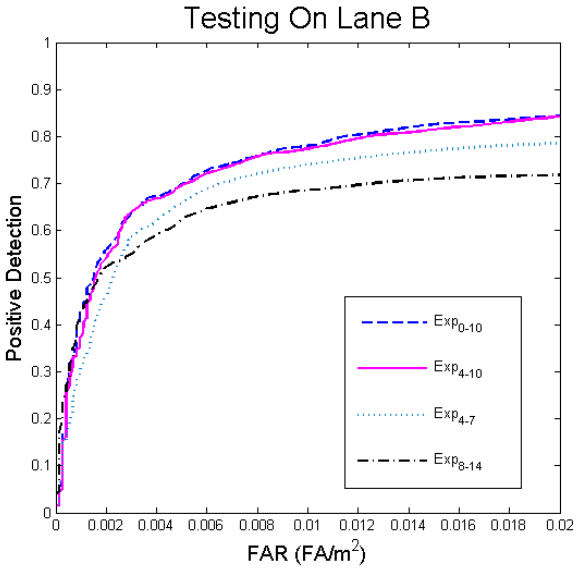


Figure 2.5

ROCs for lane-based CV with respect to different stand off distance intervals.

in particular a standoff distance interval width, where performance plateaus. That is, the system already gets enough looks on target at $[4, 10]$m. The additional 4 meters of tracking, i.e., $[0, 10]$m, does not improve performance (however, it likely results in increased system complexity and computational time). Third, one can tell that lane B is harder than lane A. That is, training on lane B and testing on lane A is easier than training on lane A and testing on lane B. This is confirmed by looking at the data. Lane B has more diversity and is similar enough to lane A, but lane A does not generalize to lane B so well. Lastly, one can see that the system is indeed limited in standoff capability. Inclusion of far off distances, e.g., $[8, 14]$m, is far worse in performance. Note, the performance does go up to above 90%. I just report performance for the given FAR interval, as higher values will not lead to a realistic system. However, with additional processing, feature extraction and algorithm design, I feel confident that this FAR can be further reduced to generate a detection rate above that of a human in FL-LWIR.

Table 2.4

Analysis of the impact of different algorithmic components on system performance.

Algorithm	Performance Increase
Importance map weighted features	PD increase of up to 10% at a given FAR
CLAHE for feature extraction	PD increase of up to 35% at a given FAR

Table 2.4 tells the following story. Overall, both proposed algorithms (importance map weighting and CLAHE-based feature extraction) improve system performance. Specifically, CLAHE-based feature extraction gives rise to the greatest system improvement. The importance map weighted features also lead to a noticeable improvement, however not as much as CLAHE. Also, importance map weighted features take substantially more computation (firing of a bank of Gabor energy filters, its aggregation and post-processing).

2.10 Future Work

In summary, Figures 2.4 and 2.5 show that the system achieves a detection rate of approximately $[80, 90]\%$ in the reported FAR interval for the lanes studied. That is, the FL-EHD system is more-or-less discovering what a human is able to find in LWIR. Also, Table 2.4 shows that the different proposed algorithmic stages do indeed lead to better system performance. Two big questions remain. First, is it possible to do better than the human in FL-EO/IR? Results indicate it may be a possibility, but further experimentation is required. Specifically, my experiments indicate that approximately 60% of the targets are relatively simple to find. There is then another $[20, 25]\%$ that is more challenging (e.g., buried in clutter and/or weakly expressed targets). The remaining set of targets is extremely challenging to identify and it is not certain yet if a system that can detect such targets is "lucky" per se. One thing is for certain, the system put forth herein has to accept a higher FAR to identify such targets. Second, the system put forth herein is able to identify a great number of targets. A logical next step is the identification of additional features and classifiers to help reduce the systems FAR.

In addition, I will investigate strategies for optimizing system parameters currently selected by humans. Though I have done some sensitivity analysis during the development of this system, a more exhaustive approach is needed. Additional future work will include testing this system on new data, which will lead to a deeper understanding of the system and addressing any weaknesses it may express. I also want to take a more formal approach to addressing the temporal aspect of clusters and the usefulness they potentially hold. Lastly, I am interested in the exploration of the interaction and fusion of FL-LWIR and DL-GPR as it is possible that I need a second information source such as, but not limited to, DL-GPR.

CHAPTER 3

SOFT FEATURES - SHEARLETS

3.1 Introduction

Buried explosives remain a major threat to both military personnel and civilian bystanders as well as regions targeted by terrorist groups. This is because buried explosives are relatively inexpensive and easy to make using commonly available materials. Additionally, buried explosives are, unfortunately, quite effective at inflicting devastating damage without the perpetrator even needing to be present for detonation. The aforementioned reasons help explain why buried explosives are "a predominant weapon of choice for terrorists, criminal organizations and extremists[3]."

The detection of buried explosive hazards is an extremely difficult and unsolved task. Herein, I explore the use of FL-LWIR imaging. Thermal sensing modalities enable us to capture/image the infrared portion of the electromagnetic spectrum, which is approximately $8-14$ micrometer wavelength. Infrared radiation is emitted at the surface of objects whose temperature is above absolute zero. As temperature increases, so does the amount of emitted thermal radiation. The amount of emitted thermal radiation by a given object is dependent upon that object's emissivity, which is a value that represents a material's ability to emit thermal radiation at a given wavelength. For explosive hazard detection, I can exploit the fact that buried objects will likely possess different thermal properties than

their surrounding soil, resulting in a detectable thermal signature due to the temperature differential between the object and its surrounding soil. With that said, one issue, natural phenomena really, that arises when using solely LWIR for buried hazards detection is the diurnal cross-over. The diurnal cross-over is a period (of time) in which the buried object comes to near thermal equilibrium with its surroundings making the device virtually unidentifiable using thermal imaging techniques. Another factor is the difference in emitted radiance seen at the soil surface (even for the same soil composition and object) varies based on factors such as the amount of incident thermal radiation, which depends on the time of day, time of year, and current weather conditions.

In contrast to its limitations, FL-LWIR does have many attractive attributes. For example, in the case of route clearance, a unit wants to clear an area as quickly as possible. Using FL technologies, an automatic explosive hazards detection system can identify targets at greater standoff distances versus other sensing solutions (e.g., DL-GPR). Additionally, FL systems obtain multiple looks on target, which results in more information gathered to either support or dispute the hypothesis of the ROI being an explosive hazard.

There are a number of approaches currently being explored for explosive hazard detection, i.e., sensor(s) used, features extracted, ground versus air, vehicle-based, etc. Common sensors used include IR cameras[45, 43, 42, 51, 53, 52, 54], GPR[24, 21, 26] and EMI sensors[55, 32, 57]. From these papers, we realize that there are a number of ways to sense buried explosive hazards and no single sensing modality is perfect. To see explosive hazard detection systems flourish, it is likely that multi-sensor fusion and/or context-dependent identification and classification will be needed. However, it is important to advance each

to gain insight of their abilities– enabling a strong foundation for multi-sensor fusion. This thesis focuses on just the advancement of IR for explosive hazard detection.

Clutter, as referred to herein, consists of a number of factors such as the environment, for example rocks, bushes, tire tracks, etc., all of which can convolute the thermal radiation detected by the sensor. As a result, imagery can be extremely challenging to discriminate, e.g., explosive hazards appear to blend in with its surroundings or clutter can sometimes look very similar to a target. Thus, there is a need for methods in which the surrounding environment information is minimized and information characterizing the object of interest (explosive hazard) is highlighted. Herein, I put forth a method coined *soft* feature extraction to address this. Another contribution of this Chapter the investigation into features in the spatial-frequency domain. A third contribution is the investigation into context extraction and different feature grouping strategies. I explore an approach that utilizes multiple cell-structured configurations during feature extraction. This concept has proven useful on occasion in computer vision. Further, different cell-structured configurations on a tight ROI (i.e, target and just enough surrounding to allow feature techniques to capture adequate edge/contour information) provides various feature groupings, resulting in localization of features for spatial context preservation in an image.

The remainder of this Chapter is structured as follows. Section 3.2 provides a brief overview of the sensor and data set used in my experiments. My proposed methods are detailed in Section 3.3. Last, I discuss the results obtained, which are characterized in the context of ROC curves in Section 4.3.

3.2 Sensors and Data Set

The FL-LWIR sensor used is a Selex L20 Camera. The camera has a cooled core and was mounted with a fixed mast on a *high mobility multipurpose wheeled vehicle* (HMMWV). The camera's center of view was focused on a distance of 30 meters from the mast and its cross-track coverage distance was 20.2 meters. The Selex camera has a spectral range of 8-12 μm at a frame rate of 30 frames per second and stored as 16-bit images with a resolution of 640$\times$512. A differential GPS receiver and inertial measurement unit are mounted next to the camera, which in conjunction with the times at which each image was captured, allowed accurate geo-referencing to be achieved[52]. The data set used in this work consists of 16 runs across two lanes at an arid United States Army test site. Of the 16 runs, 7 were from lane A and 9 from lane B. The targets are buried at different depths, ranging from shallow to deep, and also vary in terms of metal content, i.e., heavy metal to devices with low-to-no metal content but only plastic. This data set was collected during the morning and afternoon to account for the thermal variations that occur at different times of the day.

3.3 Methods

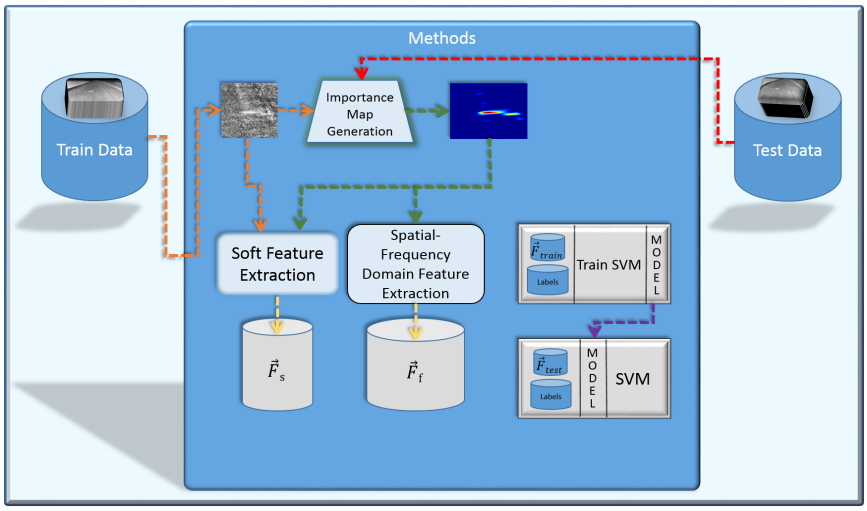


Figure 3.1

Visual illustration of the processing stages put forth in this Chapter.

For each sample (candidate target detected by pre-screener) in the training data set, an importance map is generated and used for both soft spatial and spatial-frequency domain feature extraction. An SVM model based on the set of training data is obtained and used to classify each instance of the test data set, which goes through the same feature extraction pipeline as the training data.

The focus of this Chapter is a second stage processing step which further investigates ROI detected by our pre-screener [54]. Our pre-screener is an ensemble of trainable size-contrast filters with mean-shift clustering[54, 52]. The goal of a pre-screener is typically to identify 100% (in reality as many as possible) of the targets in an efficient way. This candidate set is typically drastically fewer in number than a brute force approach. However, a pre-screener still has too many *false alarms* (FAs). Further stages of processing exist to refine this candidate set (not improve the *positive detection rate* (PDR) but the *FA rate*

(FAR)). Specifically, the focus of this Chapter is the enhancement of the quality of features extracted from the candidate image chips (i.e., ROI), which, as a result of the correlation between quality of features and system performance, is expected to enhance the output of the classifier. With respect to this work, system performance can be enhanced two ways: improving the PDR, and reducing the FAR (the latter being the focus of this work). Figure 3.1 is an overview of this work and 3.2 is an example of our pre-screener's output.

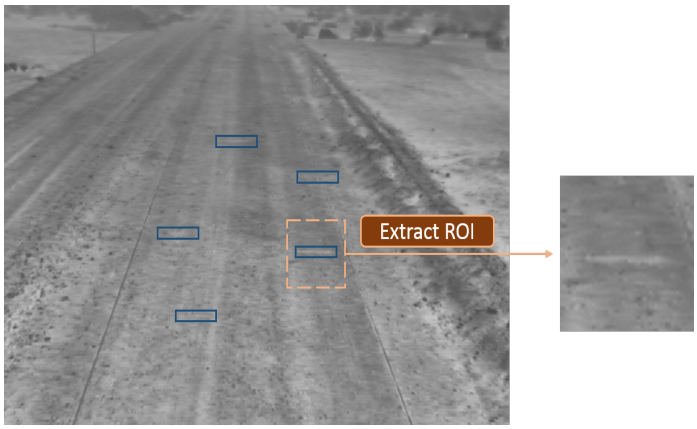


Figure 3.2

Example ROI identified by our pre-screener.

Blue rectangles indicate the target and the dashed orange box is the image chip extracted for subsequent processing.

In laymans terms, the pre-screener detects ROIs and extracts them as image chips for image space feature extraction and classification (second stage processing). In the remainder of this section, I investigate three methods: soft spatial features, feature grouping strategies, and stats of spatial-frequency domain features. All of these can be thought of as methods designed to focus on capturing target context.

3.3.1 Capturing Target Context

In this subsection I put forth a method to improve the quality of features extracted in a ROI based on the desire to only analyze (what is believed to be) target features versus target and surrounding environment features. I further develop one of my hypotheses put forth in [45] that, under our explosive hazard detection system's current state (i.e., object detection and classification), the key information in detecting a buried explosive hazard in LWIR lies in the contour of the target. That is, targets are expected to be ellipse-shaped and elongated in the cross-track direction as a result of forward looking perspective imagery. I propose that, while in some applications additional image context is beneficial (i.e., the inclusion of surrounding non-target information), it does not appear to provide noticeable benefits in buried explosive hazard detection in the road. The only "context" I have encountered is pressure plates (triggering devices) and possibly the wire connecting the pressure plate to the buried explosive hazard device. Furthermore, I do not feel it is best to simply increase the size of the ROI to detect such context as the objects will likely always be in a different image (thus world) position. Some secondary linking step may be more appropriate. In Price et al.[45], preliminary work was discussed for a filtering approach to reduce the importance of features that describe non-target regions (e.g., soil texture, tire tracks, etc. and highlight those describing a target. Herein, I extend my previous work in the following ways:

1. Soft feature extraction (Sec. 3.3.1.1)

   Soft feature extraction is a technique I put forth to address the notion that all spatial features extracted are not of equal importance in target detection. For example, what value do tire tracks provide that helps to truly distinguish targets? Is the contour of a target not of more importance than that of a bush next to it? For reasons such as these,

I generate *importance maps* via shearlet filtering to weight the spatial features. The goal of the importance maps is to highlight target information and neglect irrelevant task-specific information.

2. Multiple cell-structured configurations (Sec. 3.3.1.2)

   To address the shifting of the alignment of a target in a ROI as identified by the pre-screener, I investigate the use of multiple cell-structured configurations. Cell-structured configurations are used to preserve the relative spatial properties in an image. Using multiple cell-structured configurations, thus different feature grouping strategies, lets us address challenges that arise from target shifting and/or target size variation (illustration shown in Figure 3.3).

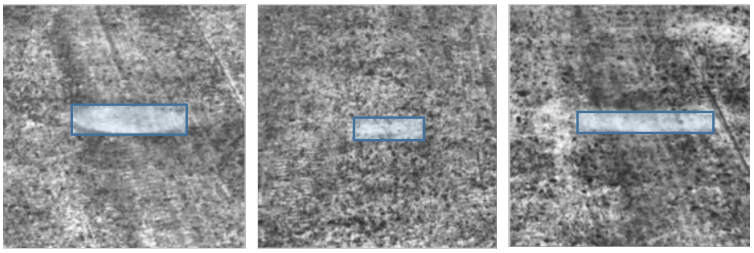


Figure 3.3

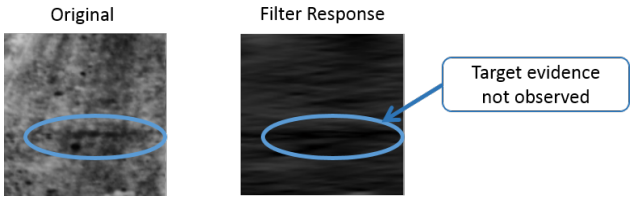
The blue rectangle is superimposed to highlight target area.

3.3.1.1 Shearlets

In my previous work, [45], a bank of Gabor energy filters was deployed to generate a candidate chip's importance map (which drives soft feature extraction). The importance map is a per-pixel weighting that defines the "importance" of each pixel to the task of buried explosive hazard detection. The impact of soft feature extraction is highly related to the quality of the importance map generated for each candidate chip. Herein, *shearlets*[35] are used for importance map generation because they are well localized and have high directional sensitivity. I implemented shearlets using the ShearLab toolbox[31]. Specifically,

the *Fourier-based shearlet transform* is used. In Easley et al.[18], it was shown that shearlets constructed in the Fourier domain express good frequency localization and directional selectivity, both of which is exploited herein.

First, I address the effects time of day has on LWIR imagery. At different times of day, targets appear as bright or dark elongated ellipses (in the cross-track direction) in relation to its surroundings. If I perform shearlet filtering to the image shown in Figure 3.4a, the (dark) target is missed. To address this, both the original candidate image chip and its inverse (i.e., 1 - original, where pixel values are in $[0, 1]$) are filtered. Figure 3.4 shows the shearlet filter response for the original candidate image chip and its inverse and we see that the target is observable when filtering the inverted candidate image chip.



(a) Original

(b) Inverted

Figure 3.4

Shearlet response on both the original and inverted candidate image chip. Ellipses are superimposed to identify where the target is for the reader.

Additionally, I exploit the shearlets directional selectivity by designing a single shearlet filter that highlights target evidence and this also filters out frequencies and, more specifically, orientations that are not expressed in the cross-track direction. As mentioned above, targets are elongated and ellipse-shaped in the cross track direction, therefore I utilize this property to design the shearlet filter to be highly selective of direction. A Gaussian blur ($\sigma = 3$) is applied to the shearlet response to smooth out the evidence. Finally, a contrast stretching transform is applied to the filter response to increase the weighting of (potential) target evidence and further reduce weakly expressed non-target information. I use Matlab's imadjust function to perform the contrast stretching transform such that pixel intensity values less than 65% of the maximum pixel value are mapped to 0 and the maximum pixel intensity is mapped to 1. Figure 3.5 illustrates this technique.

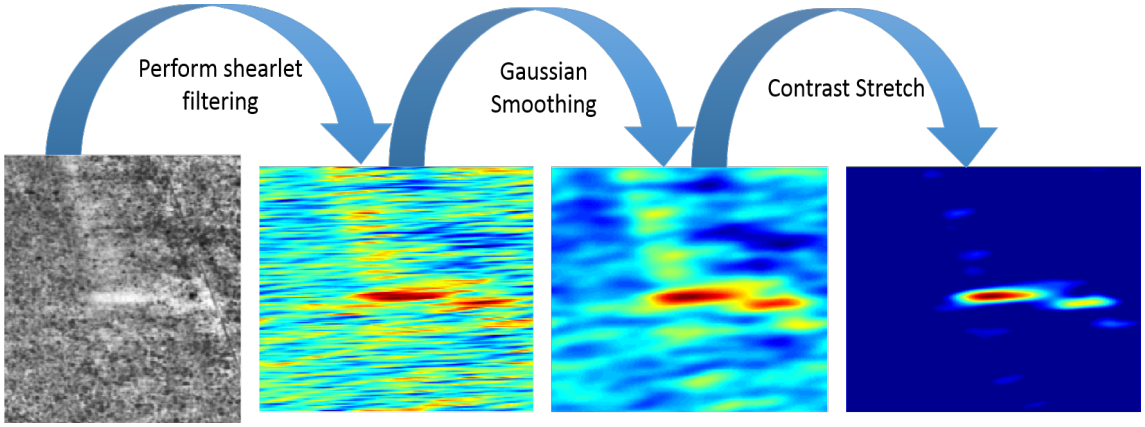


Figure 3.5

Proposed filtering process to generate importance map from start to finish.

3.3.1.2 Multiple Cell-Structured Configurations

As mentioned above, the pre-screener detects ROI from a single image. These image chips are subjected to a more thorough analysis with the goal of FA reduction. Prospective targets are expected to be centered in the image chip received by the pre-screener; however, this is not always the case as targets are, at times, slightly shifted from the image center. To help preserve the spatial context of features, cell-structured image descriptors are used. A cell is a sub ROI. Herein, I use non-overlapping rectangular shaped cells and all cells are the same size. Preserving spatial context is important as it helps one group features in the relative spatial configurations they appear in. For example, it is often used in computer vision for tasks such as face detection. In that example it is important for features that encode the eyes, mouth, nose, etc., to be located in a specific relation to each other. However, it is not always simple to know what cell configuration to use. I explore multiple cell configurations to address the robustness of this approach. However, this strategy does delay the decision making and pushes the challenge back in the pipeline for the classifier. Target shift occurs as a result of not having perfect alignment of detection on the target. Targets are also not the same size. Some are longer, others shorter; some thin, others wide. This is a result of the terrain surface, object buried, burial depth, soil type, etc. Figure 3.3 shows a few occurrences of targets that are encountered. It is easily seen that their size varies (relatively) dramatically.

Instead of using just one cell-structured configuration (e.g., 3×3), I use a set of cell-structured configurations (e.g., 3×3, 5×5, and 7×7). Let $\boldsymbol{\mathcal{T}} = \{(T_{11}, T_{21}), (T_{12}, T_{22}), \ldots, (T_{1t}, T_{2t})\}$, denote t different cell-structured configurations where (T_{1k}, T_{2k}) corresponds

to the k^{th} configuration's T_{1k} by T_{2k} non-overlapping windows (i.e., cells) centered on a pixel (3×3, 5×5, and 7×7 cell-structured example shown in Fig. 3.6). To ensure clarity, T_{1k} is the number of cells in one direction, say, along the x-axis, and T_{2k} is the number of cells in the other direction, from this example, along the y-axis (also illustrated in Fig. 3.6).

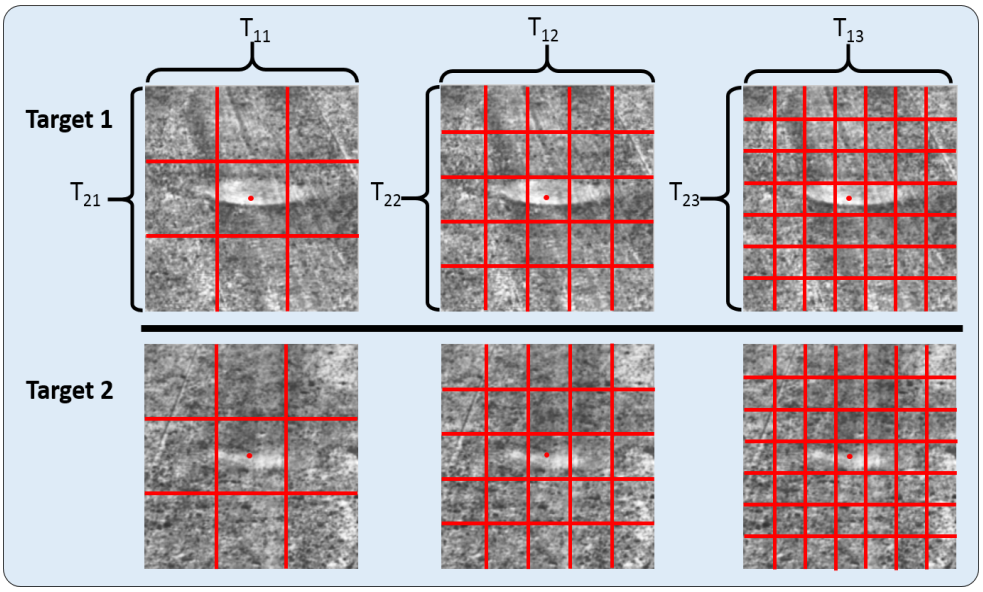


Figure 3.6

Illustration of multiple non-overlapping cell-structured configurations.

Specifically, this is shown for two different targets and the following configurations: 3×3, 5×5, and 7×7. T_{1k} and T_{2k} are the number of cells in the vertical and horizontal directions.

Before discussing the specific features extracted herein, some basic notation is introduced. Let $I = \{I_1, \ldots, I_n\}$ denote n candidate target instances (i.e., image chips) returned by the pre-screener. Let $\vec{F_i} = (\vec{f}_{i,1}, \ldots, \vec{f}_{i,J})$ represent the feature vector for the i^{th} instance

of I, where $\vec{f}_{i,1}, \ldots, \vec{f}_{i,J}$ are J different extracted features (e.g., histogram of oriented gradients, local binary pattern, etc.) concatenated to form a single vector. Herein, the features extracted can be split into two categories: spatial (Sec. 3.3.2) and spatial-frequency (Sec. 3.3.3). Thus, I will distinguish between the two using $\vec{F}_{i,s}$ and $\vec{F}_{i,f}$ to denote the spatial and spatial-frequency features, respectively, extracted from the i^{th} instance of I. The final feature vector, $\vec{F}_i$, which will be subjected to a *support vector machine* (SVM) for classification, is formed by concatenating $\vec{F}_{i,s}$ and $\vec{F}_{i,f}$ to form a single feature vector. This is the simplest and most common way of fusing features (also referred to as arraying). I will now discuss the features extracted in this work.

3.3.2 Spatial Domain Features

Three state-of-the-art features are used to characterize spatial domain information. Spatial domain features are typically designed to exploit the pixel level information in an image. Specifically, the spatial domain features extracted herein include my soft versions of *histogram of oriented gradients*[19] (HOG), *local binary patterns*[40] (LBP), and *edge histogram descriptor*[24] (EHD). These histogram-based image descriptors encode the frequency of occurrence of different patterns with an image. However, the soft versions provide a way to de-emphasize non-target information and lessen their contribution in describing the object of interest. Illustrations are given in each section that makes use of my soft feature extraction technique to show how the approach results in the features extracted being improved target focused.

3.3.2.1 Soft Histogram of Oriented Gradients

The first feature explored is the Edelman HOG[19] descriptor. Image texture is captured by the HOG through the calculation of image gradients. The HOG is a commonly used image descriptor and appears in a vast number of publications; therefore, its implementation is briefly discussed and I refer the reader to [19, 38, 36, 37] for detailed references. An 8-bin histogram is obtained for each cell, thus a concatenated descriptor of length $8 * T_{1k} * T_{2k}$ is extracted (recall T_{1k} and T_{2k} is the cell-structured used for the k^{th} configuration.). To generate the soft HOG, the image chip's pixels are weighted by its importance map. The soft HOG is calculated as the pixel-weighted magnitude of each gradient being proportionally contributed to a primary and secondary bin for each gradient using the alpha independent scheme of Lowe. I reiterate that through the use of the soft HOG, its image descriptor emphasizes image gradients that are indicative of target evidence and the (unimportant) background's impact is reduced. To illustrate its effects, Figure 3.7 displays the image descriptor obtained for both the HOG and soft HOG with respect to a 5×5 cell-structured configuration. The top right image in Figure 3.7 is the original image chip and the histogram to its left is its HOG descriptor; the bottom right image is the importance map generated for the image chip and its corresponding soft HOG descriptor is shown to its left. From this, it becomes very obvious the impact this technique (i.e., soft HOG) can have on formulating an image descriptor that is more focused on the target evidence and ignores much of the background information.

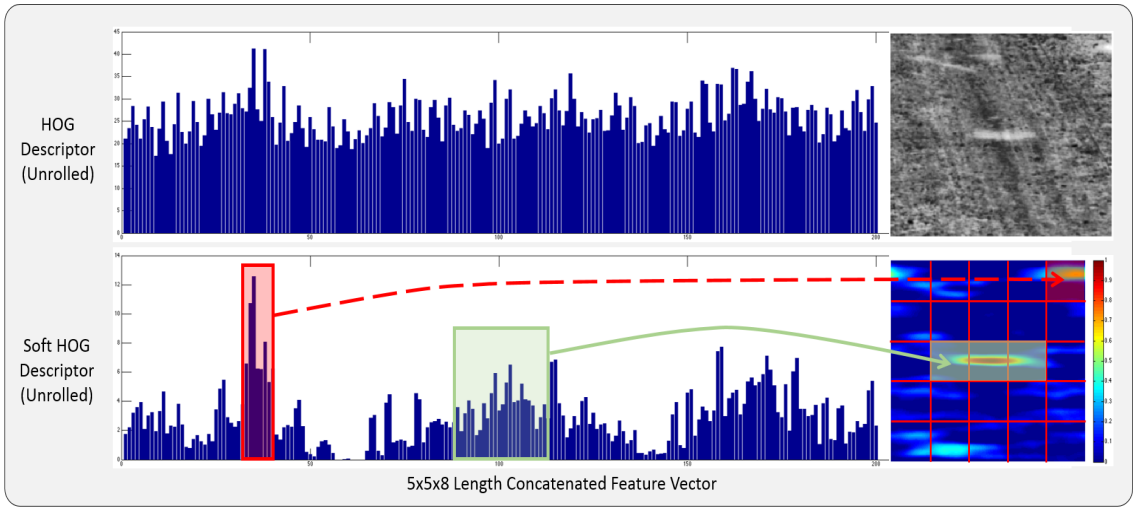


Figure 3.7

Example of the difference between the HOG and soft HOG descriptor.

The HOG descriptor is the top histogram; the soft HOG descriptor is the lower histogram.

3.3.2.2 Soft Local Binary Pattern

The second feature explored is the LBP descriptor, first proposed by Ojala et al.[40]. The LBP is another image descriptor that has been widely used in publications, e.g., human recognition[15, 63, 5]. It is relatively good at distinguishing edge curvature characteristics of an object and its surroundings. Additionally, the LBP has the favorable trait of being invariant to uniform scaling in intensity. The LBP code is

$$LBP_n = \sum_{k=0}^{n} (s(i_k - i_c)2^k), \tag{3.1}$$

where i_c is the center value, i_k is the value of the k^{th} neighbor, and function $s(x)$ is

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

$LBP^{u}_{n,r}$ represents the uniform, circular neighborhood version of the LBP which samples n points at radius r and has no more than u 0-1 transitions (i.e., uniform pattern). I compute $LBP^{2}_{8,3}$ for each cell as a summed pixel-weight normalized histogram, resulting in 57 features per cell. Note, there are typically 59 features but I remove the last two patterns as they contribute little-to-nothing to our task at hand. Thus, the concatenated descriptor is of length $57 * T_{1k} * T_{2k}$. Figure 3.8 is an example of the impact softening the LBP has on its resulting image descriptor with respect to a 5×5 cell-structured configuration. The top right image in Figure 3.8 is the original image chip and the histogram shown to its left is its LBP descriptor; the bottom right image is the importance map generated for the image chip and its corresponding soft LBP descriptor is shown to its left.

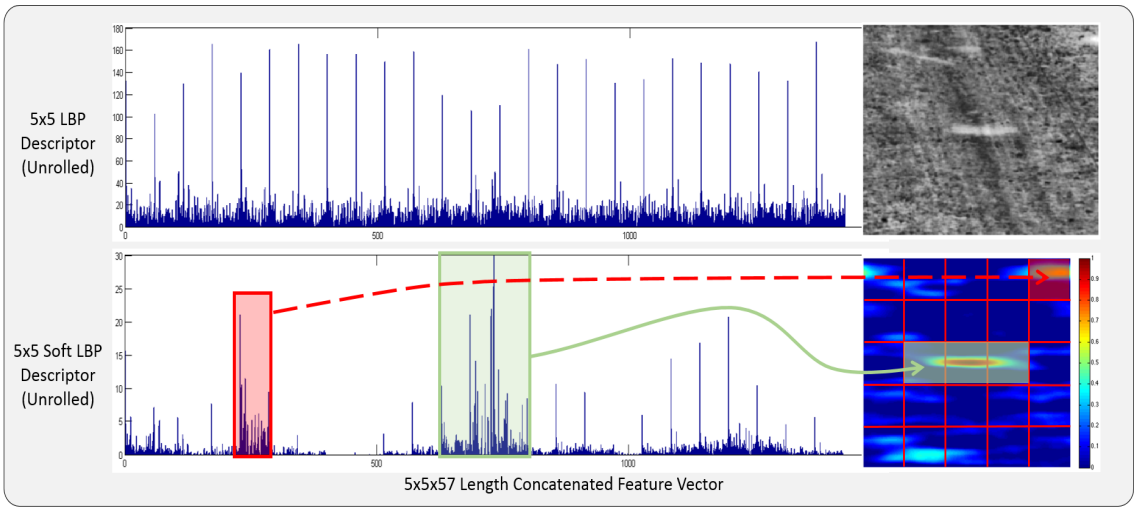


Figure 3.8

Example of the difference between the LBP and soft LBP descriptor.

The LBP descriptor is the top histogram; the soft LBP descriptor is the lower histogram.

3.3.2.3 Soft Edge Histogram Descriptor

The final spatial feature explored herein is the EHD and our implementation is based on the EHD put forth in [24]. The method in [24] was inspired by the MPEG-7 EHD, of which used five convolution operators to represent vertical, horizontal, diagonal, anti-diagonal, and non-directional edge classes; however, it does not uses a non-edge class and and edge threshold instead of a non-directional class. Using a non-edge class in conjunction with an edge threshold, pixels whose maximum response is less than the edge threshold for any of the four directional edge masks is assigned to the non-edge class. In Stone et al.[52], improvements were made to the method in [24] to make the EHD more robust (e.g., less sensitive to noise). For full details of the improved approach, see [52]. I do point out herein that the improved approach in [52] introduced two additional (new) edge masks, resulting in the EHD being of length 7. Herein, I extract two EHDs for each cell, with thresholds of 15 and 35. This results in a concatenated feature descriptor of length $7 * 2 * T_{1k} * T_{2k}$.

I improve the performance of the our EHD with the introduction of the soft EHD. The importance map generated for each image chip serves two main purposes: emphasize regions that are likely a target, and remove much of the background information that is present. Thus, it seemed intuitive to use the EHD on the resulting pixel-weighted image. That is, I pixel-weight the original image with its importance map and compute the EHD on the pixel-weighted image. As a result, the soft EHD provides information that is drastically more target evidence focused (an example is given in Figure 3.9).

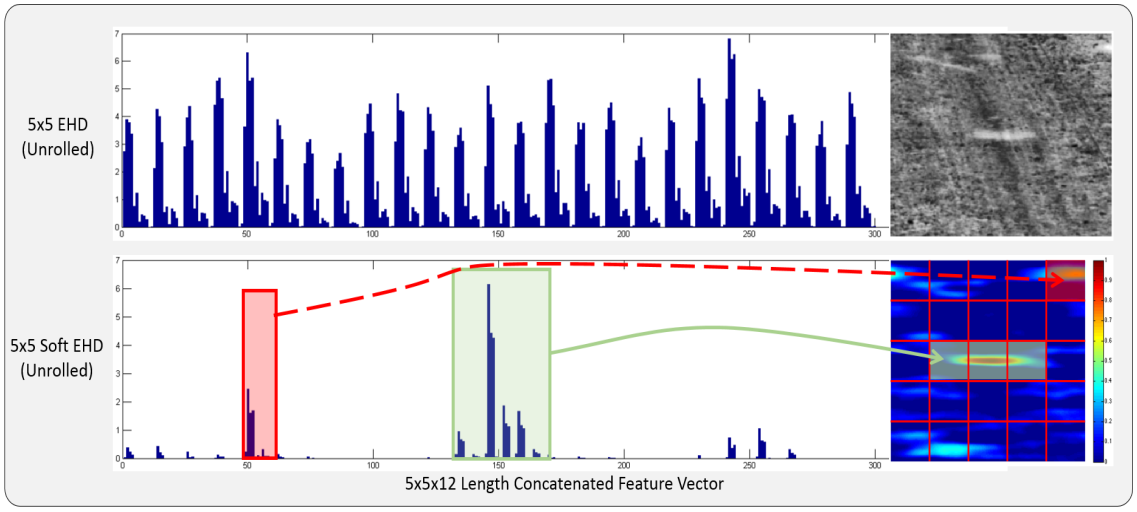


Figure 3.9

Example of the difference between the EHD and soft EHD feature.

The EHD feature is the top histogram; the soft EHD feature is the lower histogram.

3.3.3 Spatial-Frequency Domain Features

In our groups previous works, we have relied on the extraction of just spatial domain features. Herein, I begin an exploration in capturing and utilizing spatial-frequency domain features. The goal is to find robust methods for extracting (beneficial) information that becomes observable (or, whose pattern is more easily distinguishable) in the spatial-frequency domain. I perform spatial-frequency domain processing in the Fourier spatial-frequency domain, specifically, I used the two-dimensional discrete Fourier transform (2D-DFT). For a function $f(x, y)$ of size $M \times N$, its 2D-DFT is

$$\mathcal{F}(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})}, \tag{3.2}$$

where $u = 1, 2, \ldots, M-1$ and $v = 1, 2, \ldots, N-1$.

One motivation behind the use of spatial-frequency domain features is it provides a mechanism to capture descriptive target information that exists in a state similar to the human visual system. There has been a number of studies on the human visual system which suggest it transforms an image (more specifically, retinal image) into a local spatial-frequency representation[13, 16, 22, 17]. We can emulate this behavior through convolution of an image with a filter whose frequency and orientation is tuned for the application (note: a bank of filters with different frequencies and orientations is commonly used in practice to cover a larger range of phenomena). The resulting convolved image is our spatial-frequency response. That is, I modify the spectrum with the intent to filter out non-target information, which that in itself has the effect of highlighting target-like information. Aside from its motivation stemming from the human visual system, I argue that spatial-frequency features are appealing because they give us a powerful way to characterize and separate spatial-frequency and orientation information (i.e., analyze more than just the pixel level).

Spatial-frequency domain features can be calculated through analysis of the 2D-DFT coefficients of the shearlet filter response, $\mathcal{S}(u, v)$. Herein I use a single filter but multiple could be used if desired/needed. Feature extraction is performed on the magnitude of shearlet filters response. Spatial-frequency domain features extracted herein include: mean, standard deviation, L_2-Norm, and bias-corrected kurtosis (4^{rd} order central moment). Kurtosis measures the peakedness expressed by a given distribution (herein, I unroll the shearlet filter response to define the distribution). Filter response returns, as designed in this work, typically have a cleaner (i.e., less noisy) return when target-like objects are

present in the candidate chip than responses from non-target chips. Therefore, I employ kurtosis as a feature to capture this (commonly expressed) property. Specifically, I use the inverse shearlet transform and the spatial-frequency domain features are calculated in each cell to spatially localize the information. This is repeated for each cell-structured configuration.

3.4 Experiments

In this section, encouraging results are presented for FL-explosive hazard detection. The results support my conjecture of target-specific context being a key contributor to proper classification of candidate chips and little (if anything) appears to be gained from utilization of the region not *directly* surrounding the target. Herein, I utilize SVM based classification. In general, SVM based classification requires that the data be separated into different classes (two herein) and that a kernel function is chosen. I use the LibSVM[14] package and its linear kernel function for experimentation. Experiments were also conducted using the *radial basis function* (RBF) under a few different values for its σ parameter; however, the linear kernel produced the best performance (i.e., ROC curves).

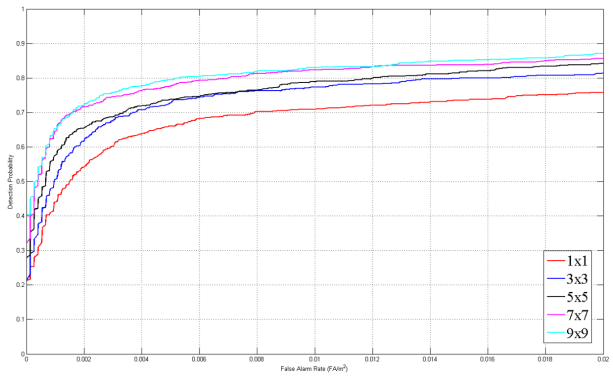
Experiments were performed using lane-based *cross validation* (CV), specifically, 2-fold CV in which X_A is used for training the SVM and each run from lane B is tested individually. This is repeated with training on X_B and testing on each run from lane A individually. Here, X_A and X_B are the sets the encompass all instances from their respective lane output by the pre-screener for 2^{nd} stage processing and classification. I vertically average the individual ROCs for each lane to ease the viewing of multiple ROCs, which

can become quite difficult when all runs are co-plotted on the same figure. Also, manual inspection of this data set was performed by a human expert and approximately 85% of targets were found.

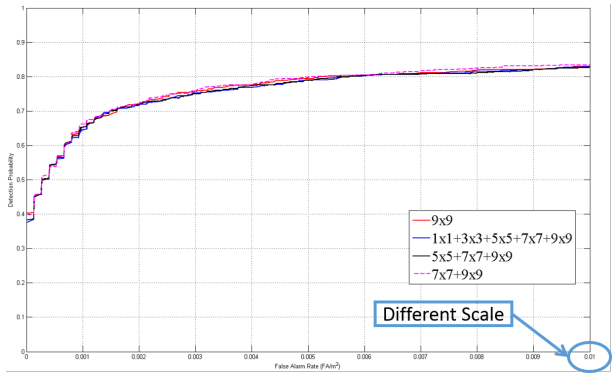
3.4.1 Experiment 1: Comparison of Individual and Multiple Cell-Structured Configurations

I begin my analysis by exploring the impact of different spatial cell-structured configurations. Note, there are a large number of experimental combinations that I could run, e.g., different configurations of multiple cell-structured strategies, different feature subsets, etc. I have selected the most salient set that illustrates the concepts discussed in this Chapter. Figure 3.10a reports the impact of individual cell-structured configurations whereas Figure 3.10b shows the impact of various sets of multiple cell-structured configurations.
These ROCs help us understand the importance of spatial context. Furthermore, I can observe the sensitivity of parameter selection on PD and FAR. It also gives us a baseline, i.e., no soft features or spatial-frequency domain features.

First, Figure 3.10a shows that different individual cell-structured configurations does indeed have a big impact with respect to spatial domain features for explosive hazard detection. I stress that it is not the exact 7×7 or 9×9 configuration that is universally best. Obviously this will likely change for different targets, window sizes and applications. Second, I found that, relative to these experiments, inclusion of multiple cell-structured configurations did not turn out to really help nor hinder detection. Meaning, there is not a big noticeable jump in ROC curves observed. While that strategy has been beneficial on occasion in the computer vision community, I simply do not observe similar benefits herein.



(a) Different single cell-structured configurations for the HOG, LBP, EHD, and pre-screener confidence. No soft or spatial-frequency domain features are extracted..



(b) Different multiple cell-structured configurations for the HOG, LBP, EHD, and pre-screener confidence. No soft or spatial-frequency domain features are extracted. Note, scale has been adjusted to help see the ROC curves.

Figure 3.10

Impact of different single and multiple cell-structured configurations.

Figure 3.10b is for all intents and purposes the same ROC curves. Overall, I found some, not much though, benefit in combining the best 7×7 and 9×9 configurations. It helps

in some instances, and does not hinder performance. Inclusion of other cell-structured configurations often (slightly) degraded system performance.

3.4.2 Experiment 2: Soft Spatial Domain Features

Next, I explore the impact of soft features on classification performance. Figure 3.11 shows the vertically averaged ROC curves for combining (concatenating) the feature vectors obtained using a 7×7 and 9×9 cell-structured configuration with soft features and without.

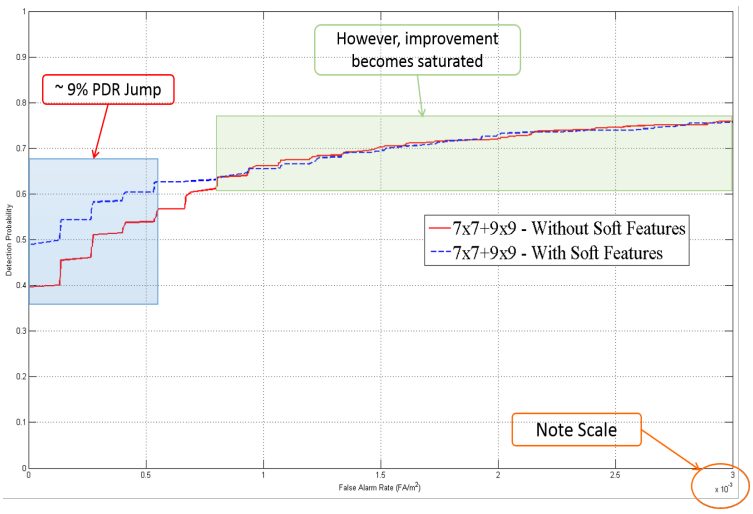


Figure 3.11

Results achieved using soft feature extraction.

Scale of the x-axis is adjusted to emphasize the improvement gained. Little-to-no difference in the ROC curves are obtained at higher FARs (with respect to this data set).

Here, we see that performance is improved at lower FARs through the use of soft features. Specifically, performance is improved on the subset of relatively easier targets

but soft features does not appear to raise the PD (for a given FAR) with respect to more difficult targets. It is possible that I have saturated the information already in these features. That is, something more powerful than the HOG, LBP or EHD is needed to discriminate targets that have very weak signatures or clutter that looks extremely like a target.

3.4.3 Experiment 3: Combination of All Features and Windowing Techniques

Last, I investigate the inclusion of spatial-frequency domain features (Figure 3.12). Specifically, performance improved some, but not much, over the use of soft features. However, performance was never hindered when the two were combined. The take away from these experiments is that a few appropriate multiple cell-structured configurations, soft spatial domain features, and spatial-frequency domain information gave rise to performance gain. However, the biggest gain to date has been in low FAR areas, which are hard but not near impossible to detect targets. Figure 3.13 shows some targets in IR that it is hard to believe any soft or spatial-frequency domain features will get without calling a lot of clutter target.

3.5 Conclusion/Future Work

In summary, the experiments performed show that system performance is improved through utilization of the methods investigated herein. Namely, the biggest improvements are at low FARs. Furthermore, the use of multiple cell-structured configurations coupled with soft feature extraction led to the most noticeable jump in performance. Additionally, system performance was improved some when features were extracted from the spatial-frequency domain. This further supports the utility of soft features as they are driven by

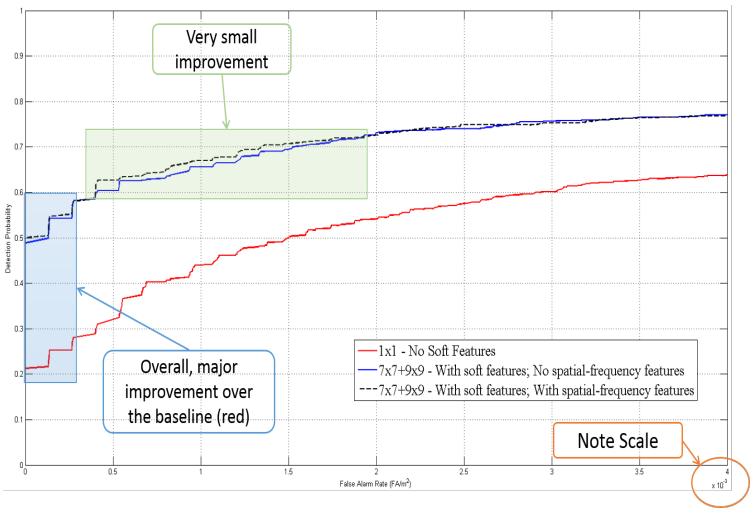


Figure 3.12

Experimental results achieved by including spatial-frequency domain features.

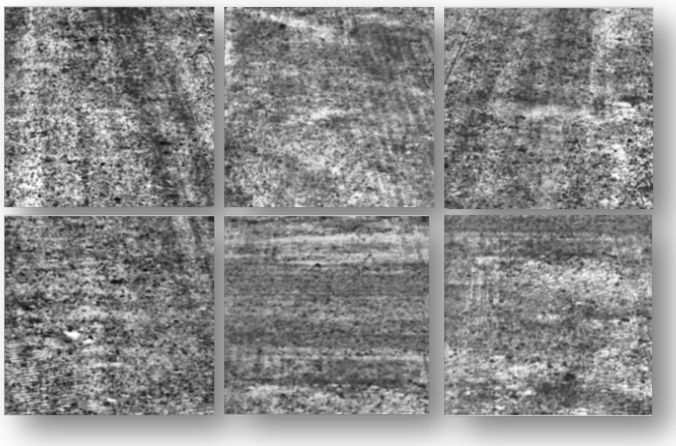


Figure 3.13

Extremely difficult to detect/distinguish targets in IR. Can you find the targets?

the candidate chip's generated importance map, which is what is used to extract the spatial-frequency domain features.

In future work, a more intelligent approach is needed to select which importance map should be used (i.e., the one generated from the original image chip or its inverse) for each candidate image chip to see soft feature extraction become more robust and useful. Additional future work will include a deeper exploration of the information extracted from the shearlet filter response. I believe there is rich information contained in the filter response yet to be extracted. Lastly, I will test my system on more data from different locations, times of day, target type, and burial conditions and explore the interaction and fusion of FL-LWIR and FL-MWIR.

CHAPTER 4

IMPROVED EVOLUTION-CONSTRUCTED (IECO) FEATURES

4.1 Introduction

Object recognition is a very challenging task, but one that has garnered increased interest in recent years. In 2013, Lillywhite et. al [34] put forth the Evolution-COnstructed (ECO) features framework for general object recognition. Their approach, while novel and shown to produce state-of-the-art results, left a number of areas asking for further investigation and improvement. Herein, I present a novel approach to address two major shortcomings of the ECO features framework. Specifically, speaking in high-level terms, I employ feature descriptors and introduce diversity promoting constraints to the genetic algorithm (GA).

Feature descriptors are commonly used in the computer vision and image processing fields to robustly capture pertinent information for object recognition. However, they are typically taken at face-value. That is, for a given recognition task, a feature descriptor, or more commonly, a set of feature descriptors are extracted from the original imagery. In a sense, it is assumed that, if using the proper feature descriptor(s) for a given task and there is some amount of class separability, the feature descriptor(s) will adequately capture this information. It is then the job of the classifier, e.g., Minimum Risk Bayes Decision Theoretic Classifier, k-nearest neighbor (kNN), SVM, etc., to find the discriminatory in-

formation for proper classification. Herein, I propose a novel ideology for approaching feature descriptors in a non-traditional way.

The ECO features framework uses a standard GA to learn a series of transforms that leads to the discovery of their so-called ECO features (a more thorough description is given in Section 4.2). There are no direct mechanisms incorporated into the GA, outside of mutation, that have been put in place to promote diversity within the population. This is a major potential shortcoming. For one, this framework has a massive search space. Referring to [34], a total of 27 image transforms where available, with all but 6 having at least one parameter associated with it (one had as many as 6). Additionally, there are four parameters used to select sub-regions in which the transforms were to be performed on. It is rather easy to recognize that the search space is incredibly large and that the GA is very likely to get stuck searching only a relatively small portion of the space if no specialized diversity promoting method is implemented. Thus, I introduce constraints on each individual's chromosome to ensure population diversity.

I do acknowledge that in [34], speciation was implemented with the intention of obtaining a diverse solution. Therein, speciation was incorporated by learning ECO features in multiple small population sizes rather than using one large population. For further details on their implementation, I refer the reader to [34].

The remainder of this Chapter is organized as follows. In Section 4.2, I put forth my improved ECO (iECO) framework. Specifically, I present a detailed discussion of the approach to using feature descriptors in Section 4.2.1, and I formally introduce the diversity promoting constraints for the GA in Section 4.2.2. An analysis of the proposed

iECO framework is presented in Section 4.3. Finally, I conclude this Chapter and provide insight to my future work in Section 4.4.

4.2 iECO Framework

I begin this section with a brief overview of the ECO features framework (for a detailed discussion, I refer the reader to [34]). The ECO features framework is an interesting method for approaching object recognition. It attempts to fully automate the process of feature construction (to an extent– if I am being sensitive, I could argue that potential solutions were imposed through the selection of transforms made available to the learner in addition to the assumption that this is the best way to recognize discriminative information from imagery). This is achieved using a GA to learn a series of image transforms and the region in the imagery that is best suited for such an ordering of transforms. In regards to this work, the image transforms that are available to the GA, their number of free parameters, and their gene identifier is shown in Table 4.1.

Table 4.1

List of image transforms available to the GA, their number of free parameters, and a gene identifier for each transform.

Gene ID	Image Transform	$\|\phi\|$
0	Harris Corner Detector	1
1	Gradient	1
2	Square Root	0
3	Gaussian Blur	1
4	Hough Circle	1
5	Median Blur	1
6	Canny	0
7	Rank Transform	0
8	Log	0
9	Sobel	0
10	Difference of Gaussian	2
11	Erode	1
12	Dilate	1
13	Contrast Limited Adaptive Histogram Equalization	4
14	Distance Transform	0
15	Histogram Equalization	0
16	Laplacian Edge	1
17	Maximally Stable Extremal Regions	6
18	Shearlets	3
19	Gabor	7

Table 4.2 is the notation used hereafter.

Table 4.2

List of notation.

$\mathcal{T}$	Ordered set of image transforms
T_i	Single image transform i
ϕ_i	Parameters for image transform i
N	Population size
τ	GA termination criteria
$\mathbf{x}_j$	Individual j
n_j	j^{th} individual's number of genes
p_c	Crossover probability
p_m	Mutation probability
Θ	Diversity promoting constraints
f	Feature descriptor
F_j	Fitness for individual j

As in the ECO framework, iECO allows individuals chromosome's be of variable length. Herein, a chromosome is the segment of genes (i.e., series of image transforms) that represents a potential solution to the optimization task at hand. It is important to note that the ordering of genes does matter. Dilating an image followed by a distance transform will produce a different output than if conducted in the reverse order. In [34], the maximum allowed chromosome length was limited to 8 and this convention is also used for this work. Furthermore, ECO allows sub-regions of the image to be learned (if desired, as stated in [34], this is not required). For this work, experiments are performed on a data set composed of image chips that have been identified by a pre-screener as regions of interest (I refer the reader to [52, 54] for details). The image chips are sub-regions of the original image, this is one factor that lead us to designing iECO to be performed on the "entire" image. Thus, the resultant transformed image produced by an individual will be the same size as the original image chip. Adaboost and perceptron are implemented in [34] (and herein) to combine chromosomes and formulate a fitness score for each individual. As in [34], the fitness score is computed as

$$F = \frac{tp * 500}{tp + fn} + \frac{tn * 500}{tn + fp}, \tag{4.1}$$

where tp, tn, fp, and fn is the number of true positives, true negatives, false positives, and false negatives, respectively. Therefore, the fitness score will be a rational number in the range $[0, 1000]$, with higher values being better.

4.2.1 Unique Feature Descriptor Approach

The iECO framework first advances the ECO features framework by using the ECO pipeline as a preprocessing stage, followed by the use of well known feature descriptors. The benefits of doing this are many. For one, feature descriptors have been heavily studied by the computer vision community, and many are founded on robust statistical properties. Furthermore, feature descriptors can be more easily interpreted than the ECO features. This is desirable as it allows more knowledge to be gained about a given problem and potentially further advanced. Additionally, there are a number of methods put forth that preserve the spatial context in imagery, resulting in localization of the features extracted (e.g., deploying feature descriptors on each cell/patch of an image partitioned in a cell-structured configuration– see [45] for an example). I expect the ECO framework to be highly susceptible to data sets in which there is much variation. Meaning, the bulk of imagery in the data sets used in [34] are centered on the object of interest. The ECO features learn specific sub-regions in the imagery; what happens when a new instance comes in where the object of interest is in the upper-left corner of the image, i.e., not exactly where the ECO feature is focused. My conjecture is that it would fail under such scenarios.

Beyond replacing ECO features with feature descriptors, and perhaps the more novel aspect of this method, is the ideology that different feature descriptors require different transforms to enable their full potential in extracting pertinent information. Herein, I propose that one should learn the series of image transforms, $\mathcal{T}$, that gives a particular feature descriptor the best chance at capturing discriminative information for the given domain's

problem. That is, different feature descriptors represent information differently. As a result, each requires a unique $\mathcal{T}$ such that the data is best presented to the descriptor for it to exploit its unique method of extracting discriminative features. It is important to emphasize that the learning of $\mathcal{T}$ and their parameters is done for *each* feature descriptor being used. Algorithm 2 summarizes this process.

Algorithm 2 Learn $\mathcal{T}$ for each feature descriptor.

for each feature descriptor, f_i, **do**
    Create and initialize population;
    while stopping condition not true **do**
        for each individual, $\mathbf{x}_j$, $j = (1, 2, \ldots, N)$ **do**
            for each training image, I_k, **do**
                Process I_k with $\mathcal{T}_j$;
                Compute f_i on transformed image;
            end for
            Train perceptron;
            for each image in holding set **do**
                Process image with $\mathcal{T}_j$;
                Compute f_i on transformed image;
            end for
            Compute and assign fitness score for $\mathbf{x}_j$;
        end for
    end while
end for

The need for such an approach is further supported when looking at the transformed imagery that was learned for the different feature descriptors used herein. First, I briefly mention these feature descriptors. Three different descriptors were implemented: *histogram of oriented gradients* (HOG) [19, 36, 37], *edge histogram descriptor* (EHD) [24, 52], and a statistical-based descriptor (SD). The SD is a simple descriptor I put forth and is composed of the following: local mean, standard deviation, kurtosis, and L_2-Norm, and the difference between the local values and their corresponding global value. Here, "local" refers to each cell of the cell-structured configuration.

Figure 4.1 provides an example of each feature descriptor's resulting transformed image for an instance of the data set in which a target, i.e., object of interest, is present. In Fig. 4.1, it is quite obvious that each feature descriptor has a unique $\mathcal{T}$ that is preferred by the descriptors for extracting features on this domain. It is important to stress that the learned $\mathcal{T}$ is *problem domain dependent*. That is, when moving from one problem domain to the next, $\mathcal{T}$ should be re-learned so that it is optimized for that domain.

At this point, I have discussed the portion of my iECO features framework that advances ECO features in two ways. One, feature descriptors are used instead of the ECO features, which (ECO features) are an unrolled image patch that has undergone a series of learned image transforms. And second, I propose a novel ideology to approaching feature descriptors for object recognition: condition the original data (through learned image transforms) for each feature descriptor. This allows each descriptor to extract better features, i.e., they are more discriminative. Now, I will discuss my approach to ensuring that a

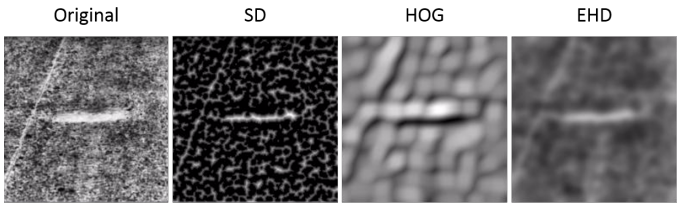


Figure 4.1

Visualization of each feature descriptor's learned $\mathcal{T}$.

From left to right: the original, $\mathcal{T}_{SD}$, $\mathcal{T}_{HOG}$, and $\mathcal{T}_{EHD}$ images are shown. Individually scaled for visual display.

diverse population is maintained throughout the lifetime of the GA through the implementation of constraints.

4.2.2 GA: Diversity Promoting Constraints

The importance of GAs having a diverse population has been well documented by the Evolutionary Computing community [12, 67, 11, 58]. First and foremost, a diverse population indicates that much of the search space is being explored. It is rather intuitive to realize that less diversity within a population will result in more regions of the search space being neglected. In standard GAs, mutation, which has a (typically low) probability of occurring, is the component that aims to add diversity to a population. However, the ECO search space is a rather unique and large optimization problem. Mutation alone is very unlikely to result in an adequate exploration by the population, i.e., pre-mature convergence to a sub-optimal solution is highly probable. As a result, a sophisticated

approach needs to be incorporated into the GA for the search space to be more sufficiently explored.

Herein, I introduce diversity promoting constraints that attempt to consider the uniqueness and complexity of the ECO's search space. Two main questions were considered when designing constraints for this problem.

4.2.2.1 How much gene overlap allowed between individuals?

Further complicating this problem, the solution must be mindful of the path traversed, i.e., the ordering of the genes used that composes the individual. An illustration is provided in Fig. 4.2 to help with understanding. Additionally, there should be some amount of gene overlap allowed within the population so as to give regions of the space a fair chance to be adequately explored. This is illustrated in Fig. 4.3.

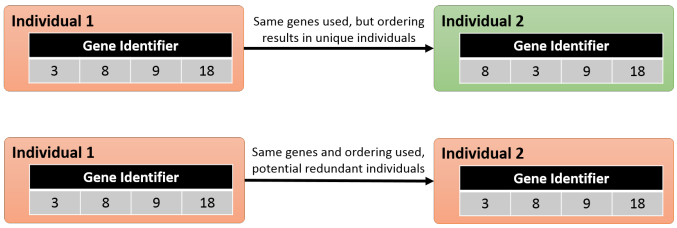


Figure 4.2

Diversity promoting constraints need to account for gene ordering.

4.2.2.2 How to address consecutive uses of the same gene?

It is my conjecture that the same gene occurring back-to-back is unwanted, an inefficient search of the same space (i.e., an unnecessary gene). Could the two genes be col-

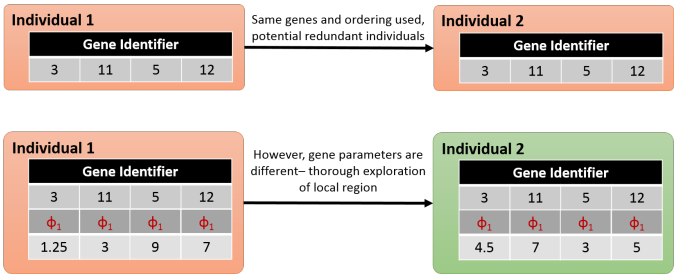


Figure 4.3

Showing the need to allow some amount of gene overlap to occur.

lapsed into one by removal of the latter occurrence? Perhaps some middle-ground would be better, such as taking the mean between the repeated genes parameter values.

Considering the questions above, I propose the following solution. Designing a set of diversity promoting constraints, Θ that define what percentage of the population is allowed overlapping genes at each layer of the individual's gene segment. For example, say that the maximum number of genes any one individual can have is 4. Then Θ would be a vector of length four, with the value at each index representing the population percentage that is allowed to overlap at the corresponding gene layer. It is important to understand that the constraint at the i^{th} layer of the gene segment is with respect to the sub-population in which all gene types (i.e., same image transform, not required to also have the same parameters) leading up to the i^{th} layer are the same. Figure 4.4 provides a graphical illustration to help with the understanding of this concept.

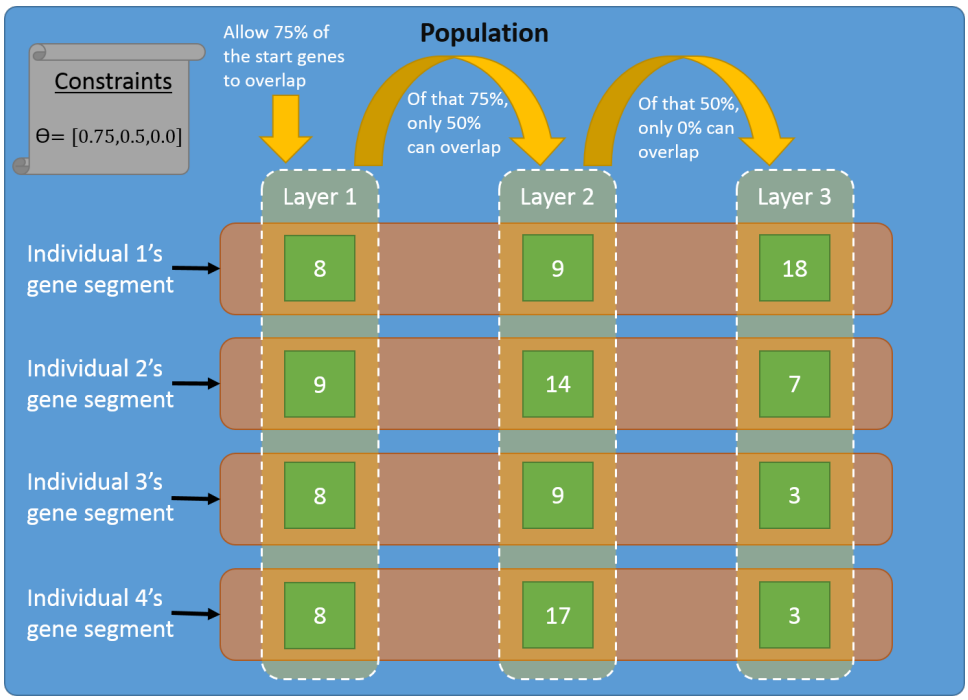


Figure 4.4

Illustration of the implementation of diversity promoting constraints.

Based on Θ, its first value, 0.75, ensures that no more than 75% of the population has the same start gene (represented by Layer 1). Its next value of 0.5 translates to the following. At most, 50% of the population with the *same* start gene can have the same second gene (represented by Layer 2). Finally, 0% ensures that no two individuals have the exact same chromosome.

Next, I address the issue of having the same gene occur back-to-back in $\mathcal{T}$. As noted above, I believe that such a scenario is undesirable for most applications of this framework. For example, what sense does it make to take a rank transform twice, one right after the other? Additionally, repetitive genes in this manner increases the computational complexity of the system as a consequence of the unnecessary image transforms. In this work, I combat this by collapsing down consecutive uses of the same gene type. That is,

if any particular gene occurs two or more times, consecutively, only the first occurrence is retained and the remaining consecutive genes of that type are removed. This results in the iECO framework also having the benefit of improving the efficiency of what is already a computationally expensive (ECO) system.

4.3 Experiments

To assess the impact of the proposed iECO features framework, a comparative analysis between it and the original ECO features framework is given. This is done in the following two ways: $1)$ population diversity (Section 4.3.1), and $2)$ system performance expressed by *Receiver Operating Characteristic* (ROC) curves (Section 4.3.3). Additionally, I show that learning a unique $\mathcal{T}$ for each descriptor is both beneficial and necessary (Section 4.3.2). Experiments are performed using the following set of parameters (which were empirically defined, but will be the subject of future work):

τ**:** 50 iterations

N: 50

max n**:** 8

p_c**:** 95%

p_m**:** 2%

Θ**:** $[0.15, 0.9, 0.8, 0.9, 0.8, 0.9, 0.7, 0.0]$

4.3.1 Experimental Analysis: Population Diversity

Table 4.3 is provided to help the reader realize how little diversity is achieved under the original ECO features framework, and is interpreted as follows. In the same manner as that in Fig. 4.4, as one traverses the layer's of the gene segment, the population $\%$ at the i^{th}

layer is with respect to the sub-population from the preceding layer. All three descriptors' populations had more than 90% of its final population exploring relatively similar regions of the search space (i.e., same start gene). Of the population with the same start gene, the same phenomenon exists throughout the entire length of the gene segment (as shown in Table 4.3). There is indeed a need for a method to not only promote, but also ensure population diversity. Such a method is given in the iECO features framework in the form of Θ. It would be a moot point to give a similar table for the iECO features framework as it is guaranteed to abide by the constraints enforced by Θ.

Table 4.3

ECO features framework has major shortcoming in the lack of population diversity.

Traversing Gene Segment: % Overlap Across Population								
Population	*Gene Layer*							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
HOG	92%	78%	56%	60%	58%	71%	100%	100%
SD	92%	74%	56%	74%	50%	86%	83%	100%
EHD	96%	73%	49%	76%	35%	100%	50%	0%

Next, I look at the gene segments of the individuals representing the top 20% of each population at the completion of the learning algorithm for both the ECO and iECO features frameworks. This, along with each individual's fitness score is given in Tables 4.4 and 4.5 for the ECO and iECO features framework, respectively. I only show this for the EHD feature descriptor; however, the same phenomenon holds true for the other two descriptors used herein. It is imperative to understand that the top 10% of the population are the elitist,

therefore, there was no alteration of their genes. This is why most, if not all, of the first five individuals for each descriptor has gene segments that lack diversity– showing a sign of convergence. However, the next five individuals, from the iECO framework (Table 4.5), have been subjected to the diversity promoting constraints and this is reflected in their gene segments. It becomes very obvious that iECO features framework results in a much more diverse population than the ECO features framework and thus, will have performed a more thorough exploration of the search space. This is a highly desirable property for GAs to possess as they are very susceptible to pre-mature convergence.

Table 4.4

ECO features framework - f_{EHD}: Top 10 Individuals.

Individual ID	*Gene Layer*								*Fitness*
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
1	18	8	8	7	13	7	-	-	840
2	18	8	8	7	13	7	-	-	840
3	18	8	8	7	13	7	-	-	840
4	18	8	8	7	13	7	8	7	840
5	18	8	8	7	13	7	7	-	840
6	18	8	7	8	7	8	8	7	840
7	18	8	8	7	13	8	7	8	840
8	18	8	7	7	8	7	7	-	840
9	18	8	7	8	8	7	13	18	840
10	18	7	-	-	-	-	-	-	840

Table 4.5

iECO features framework - $\boldsymbol{f_{EHD}}$: Top 10 Individuals.

Individual ID	*Gene Layer*								*Fitness*
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
1	3	11	5	-	-	-	-	-	886
2	3	11	5	-	-	-	-	-	886
3	3	11	5	-	-	-	-	-	886
4	3	11	5	-	-	-	-	-	886
5	3	11	5	-	-	-	-	-	886
6	8	5	11	5	4	5	11	-	886
7	4	5	3	11	5	-	-	-	863
8	1	11	5	-	-	-	-	-	863
9	13	3	11	5	8	11	5	-	818
10	12	11	5	-	-	-	-	-	818

Furthermore, Table 4.4 provides a great example of how the ECO features framework allows for nonsensical ordering of genes. To elaborate, look at the first individual's second, third, and fourth gene. The second and third genes are consecutive Log image transforms, which has no parameter. This is followed by a rank transform– why the need for the third gene? The repeated use of the third gene has no impact on the fourth gene's result. Such phenomena occurred consistently in all of the experiments conducted using the ECO features framework (including those not reported herein). Also, for this problem at least, the necessity for using feature descriptors over unrolling an image patch is realized as eight of the top ten iECO individuals result in higher fitness scores than does the best ECO individual. Lastly, comparing Tables 4.4 and 4.5, it is obvious that the iECO framework

tends to produce individuals of shorter length than does the ECO approach (i.e., better computational efficiency).

4.3.2 Experimental Analysis: Unique $\mathcal{T}$ Impact

In Section 4.2.1, it was proposed that a unique $\mathcal{T}$ should be learned for each feature descriptor being used. Additionally, it was shown in Fig. 4.1 that each descriptor does, visually, appear to be learning something entirely different. I now give a more analytical approach to supporting the need for such an approach. To do this, I computed the fitness score for the top 10 individuals from each population, but having them extract the other feature descriptors used herein. Their resulting fitness score and corresponding $\%$ change (Δ) over using its own feature descriptor is given in Tables 4.6-4.8.

Table 4.6

Extract HOG using Top Individuals from SD and EHD Populations.

Individual ID	***SD***		***EHD***	
	F	% Δ	*F*	% Δ
1	818	-3%	818	-3%
2	818	-3%	795	-5%
3	818	-3%	795	-5%
4	818	-3%	795	-5%
5	818	-3%	795	-5%
6	795	-5%	795	-5%
7	795	-5%	795	-5%
8	795	-5%	773	-8%
9	795	-5%	773	-8%
10	795	-5%	773	-8%

Table 4.7

Extract SD using Top Individuals from HOG and EHD Populations.

Individual ID	***HOG***		***EHD***	
	F	% Δ	*F*	% Δ
1	705	-18%	773	-11%
2	705	-18%	773	-11%
3	705	-18%	773	-11%
4	705	-18%	750	-13%
5	705	-18%	727	-16%
6	705	-18%	682	-21%
7	705	-18%	659	-24%
8	682	-21%	659	-24%
9	682	-21%	659	-24%
10	659	-24%	659	-24%

Table 4.8

Extract EHD using Top Individuals from HOG and SD Populations.

Individual ID	***HOG***		***SD***	
	F	% Δ	*F*	% Δ
1	818	-8%	795	-10%
2	750	-15%	773	-13%
3	727	-18%	773	-13%
4	727	-18%	773	-13%
5	727	-18%	750	-15%
6	705	-21%	750	-15%
7	705	-21%	727	-18%
8	705	-21%	727	-18%
9	705	-21%	705	-21%
10	682	-23%	705	-21%

In all instances, the fitness score dropped. From this, we can infer that each descriptor's learned $\mathcal{T}$ is indeed unique and beneficial to that descriptor.

4.3.3 Experimental Analysis: System Performance

I now present results for implementing the iECO framework on a real-world data set with application for automatic detection of buried explosive hazards in forward looking-long wave infrared (FL-LWIR) imagery. This data set consists of 16 runs across two lanes at an arid United States Army test site. Of the 16 runs, 7 were from lane 1 and 9 from lane 2. Targets are buried at different depths, ranging from shallow to deep, and also vary in terms of metal content, i.e., heavy to low to no metal content. Additionally, this data set was collected during the morning and afternoon to include the thermal variations that occur at different times of the day. Lane-based cross validation (CV) was used for testing, e.g., train an SVM using runs from lane 1 and test using runs from lane 2. Because co-plotting multiple ROCs on the same figure can be very difficult to view, I vertically average each individual runs ROC belonging to the lane being tested.

To assess the impact of the iECO framework, I compare its results with the ECO framework in two ways. First, a direct comparison between iECO and ECO are given. As will be shown, iECO vastly outperforms ECO features in this setting. Second, I show that the iECO framework also outperforms the ECO framework even if using the feature descriptors instead of ECO features (i.e., unrolling the transformed image). I believe that this is likely a result/indicative of the ECO framework's poor searching of the solution space. Learning for both methods was performed on a very small subset of the training data (roughly 1%),

half of which were targets (class 1) and half non-target (class 2). Lastly, results are given for experiments using only the best individual as well as using the top 5 individuals. This was done to explore the idea that the different individuals could each be learning something independent from each other that, when combined, improves system performance. To ensure clarity, the top y individuals from *each* descriptor is used, not the globally best y individuals and their corresponding descriptors.

4.3.3.1 iECO Outperforms

First, an assessment of the need for using feature descriptors instead of ECO features is given. Figure 4.5 shows the vertically averaged ROC curves along with the 95% confidence intervals resulting from the two approaches.

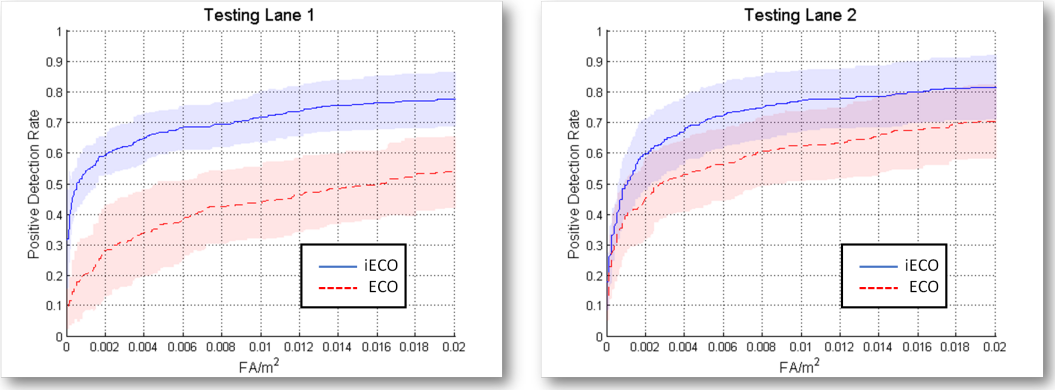


Figure 4.5

Vertically averaged ROC curves with 95% confidence intervals.

iECO is shown as the solid blue line, and ECO as the dashed red line. For each, 95% confidence interval is overlaid in corresponding color.

Here we see iECO drastically outperforms ECO features. This supports my hypothesis that using cell-structured feature descriptors provides a much more robust method for capturing pertinent image data than does the ECO features. Additionally, this data set's vast collection of varying target signatures (i.e., size, shape, texture, etc.) potentially exploits the ECO features vulnerability to intra-class variation. That is, the ECO features' need for objects of interest to be presented in a relatively static/repetitive setting (e.g., images of faces that are centered and the focus of the imagery).

From this, the following question may arise. A major part of this paper's contribution is in the diversity promoting constraints, how do we know that they are contributing to performance improvement?– Is improvement simply the result of using feature descriptors on the $\mathcal{T}$ transformed imagery instead of ECO features? To investigate such a scenario, experiments were performed using the ECO framework; however feature descriptors were applied to ECO's $\mathcal{T}$ transformed imagery.

4.3.3.2 iECO– Is it Just the Descriptors?

I now provide preliminary results on the investigation into the impact iECO's diversity promoting constraints has on a real-world application. It was already shown that Θ does have a positive impact in regards to ensuring population diversity is sustained, and thus, a more thorough search of the space is achieved. However, how does this translate in terms of system performance. Figures 4.6 and 4.7 show the vertically averaged ROC curves and 95% confidence intervals for experiments using the best individual and the top 5 individuals from each descriptors population, respectively.

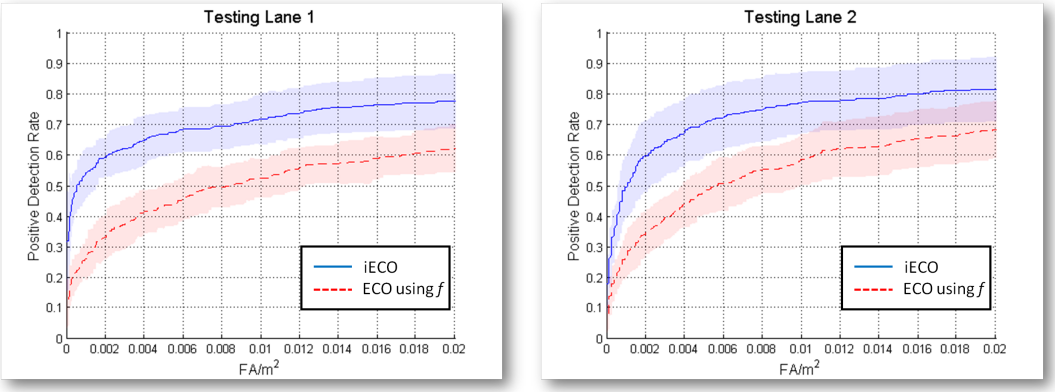


Figure 4.6

ROC curves with 95% confidence intervals produced using each f's best individual. iECO is shown as the solid blue line, and ECO as the dashed red line. For each, 95% confidence interval is overlaid in corresponding color.

Interestingly, iECO largely outperforms the ECO using feature descriptors approach. Noting that the only difference between the two methods in iECO and ECO using feature descriptors (really, a hybrid of ECO and iECO) is that iECO includes the diversity promoting constraints. Seeing that ECO using f (best individual from each only– Fig 4.6) does in fact perform better than the ECO features reported in 4.5, support is gained for the introduced Θ being an key contributor to the overall improvement in system performance.

Lastly, a potentially unexpected result comes from comparing iECO's performance in Figs. 4.6 and 4.7. That is, performance is actually better if only the best individual from each f is used instead of a collection of top individuals. I do remind the reader that a very simply and straight-forward approach to feature fusion was used herein (feature vector

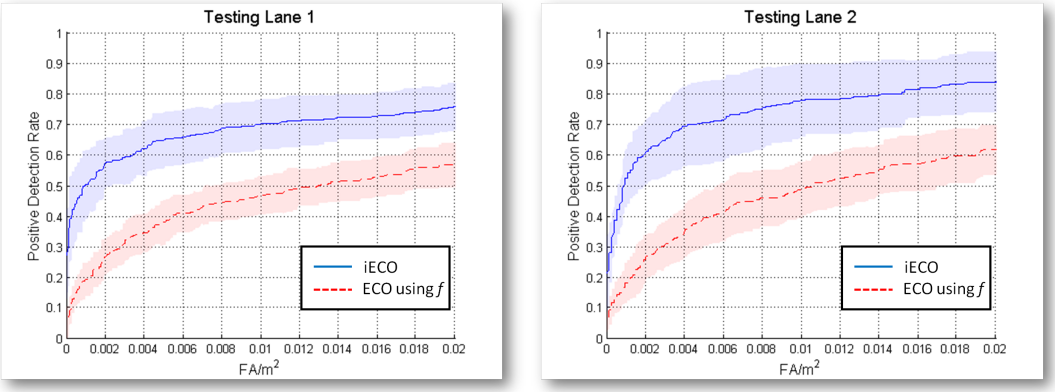


Figure 4.7

ROC curves with 95% confidence intervals produced using each f's top 5 individuals.

iECO is shown as the solid blue line, and ECO as the dashed red line. For each, 95% confidence interval is overlaid in corresponding color.

concatenation)– future work is planned for investigating more sophisticated feature level fusion, which is likely required to see the full potential of the iECO framework.

4.4 Conclusion

Herein, I proposed the iECO features framework. This advances the ECO features by using feature descriptors rather than a simple unrolling of a transformed image patch and also through the introduction of diversity promoting constraints. It was shown that iECO has a number of benefits and improvements to the ECO framework. Experimental results suggest that my constraints lead to a better searching of the solution space. Additionally, iECO tends to find individuals whose chromosomes are smaller in length, and therefore less computationally complex then the ECO algorithm. Furthermore, feature descriptors

were shown to provide great performance improvement over ECO features. In future work, I plan to look into developing novel fusion techniques for improving the ability to identify discriminatory information for the different individuals and across the different feature descriptors. Additionally, I am developing methods that will track which volumes of the search space have been explored and ensure that the population is thoroughly diverse.

CHAPTER 5

CONCLUSIONS

In this thesis, two new methods were presented for solving anomaly detection in signal/image processing and were demonstrated on FL-LWIR imagery for EHD. From this research, I found that anomaly detection benefits from the extraction of features that are focused on target-like information versus considering both target and non-target information (e.g., background) equally important. The iECO framework provided a method that, to some extent, generalizes soft features. That is, while its main focus is to extract features that provide information to best discriminate target from non-target, it automatically learns a series of image transforms to condition the data for optimal feature extraction (similar to the pre-defined steps for generating an importance map for soft feature extraction).

Soft features are very useful when an algorithm designer or field expert can "define" their object of interest. Meaning, they have a deep understanding of the problem, or the problem is very specific to a particular object (e.g., "four-door car" is much more defined/specific than "vehicle"), and as a result, are able to make assumptions. It is important to elaborate on the use of the term assumptions here as it can be quite ambiguous, and this is best done via examples. As in this work, a certain problem may be concerned with identifying an object from imagery that has a specific shape (e.g., square, ellipse, triangle,

etc.). Under such a scenario, an algorithm designer can *assume* that if there is no expression in the imagery that resembles their object's shape, then the object is absent from the image (i.e., cannot "force" something to be present that is absent). To give another example, if we are concerned with the detection of a commercial jet using satellite imagery, we can *assume* a general geometry (much more complex than a simple square or circle) for commercial jets and design the algorithm to focus its efforts on objects that express similar geometries. Thus, in both examples mentioned, it can be easily rationalized that each would be good candidates for soft feature extraction to emphasize ROIs and ignore background/clutter information. While soft features provide a great method for improving the quality of information extracted, they do require (at least to some extent) prior domain knowledge and the ability to make such assumptions. For problems in which there is little knowledge, or assumptions cannot be made, the more flexible and powerful iECO framework was introduced.

The iECO framework improved ECO features through promotion of population diversity and the replacement of ECO features with descriptors. iECO provides us with a tool that resembles the motivation behind soft features, i.e., emphasize target/important information, but without the need for an expert to project their idea of important information into the algorithm. The discriminative information is exposed by iECO through a learned composition of image transforms. As a result, iECO attempts to learn the best way to transform imagery such that the quality of features extracted are optimal.

5.1 Future Work

Going forward, there are further advancements that will be the subject of future research. Specifically, the focus will be on the iECO framework. One avenue that will be investigated is iECO's search space. It is my belief that we can better understand and realize a near approximation of the underlying space that the learning algorithm is exploring. The reason there is a need for such an approach is the following. While, yes, part of the benefit of the iECO framework is that it ensures a diverse population (via its diversity promoting constraints) which leads to greater exploration of the search space, there is still a need to ensure that the entire space is explored (to an extent, this is not a brute force strategy– but rather we desire that each "unique" pocket in the space be given a chance). There are two main contributors to this need. First, the iECO search space is incredibly vast and complex. For example, in this thesis there are 20+ image transforms, most with accompanying parameters; all of which is learned (i.e. very complex). Second, depending on the problem domain and other factors, two different compositions of image transforms could, for all intents and purposes, result in the "same" output (transformed) image. Thus, I am currently working on a method that will allow us to model, or approximate, iECO's search space such that we can identify (ahead of time) that certain compositions of transforms leads to the (near) same solution. As a result, the learning algorithm is poised to benefit greatly (e.g., true search of the entire space, reduces the number of seemingly different, yet redundant solutions, potentially reduces the number of individuals and/or time until convergence needed, etc.).

Another way that I am focusing on advancing iECO is through fusion. In this thesis, I empirically chose to use the top five individuals from each learned descriptor and concatenated their features into one long feature vector. However, if one better combined the different individuals from each descriptor, could they have used fewer individuals? Do we need to use more? Can performance be improved by more intelligently fusing each individual's information? Addressing this is concern of future work and will be explored in conjunction with better combining the different individuals via some fusion strategy.

REFERENCES

[1] *IEDs will remain 'weapon of choice' for decades*, Sept. 2012, https://www.jieddo.dod.mil/news_story.aspx?ID=1488.

[2] *Terrorist Explosive Device Analytical Center (TEDAC)*, Dec. 2012, http://www.fbi.gov/about-us/lab/terrorist-explosive-device-analytical-center-tedac/tedac.

[3] 2013, http://www.un.org/en/terrorism/ctitf/pdfs/The%20Beam%20Vol%207.pdf.

[4] A. Abdallah, H. Frigui, and P. Gader, "Adaptive Local Fusion With Fuzzy Integrals," *Fuzzy Systems, IEEE Transactions on*, vol. 20, no. 5, Oct. 2012, pp. 849–864.

[5] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, 2006, pp. 2037–2041.

[6] D. Anderson, J. M. Keller, and O. Sjahputera, "Algorithm fusion in forward-looking long-wave infrared imagery for buried explosive hazard detection," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 801722–801722.

[7] D. T. Anderson, S. R. Price, and T. C. Havens, "Regularization-based learning of the Choquet integral," *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, July 2014, pp. 2519–2526.

[8] D. T. Anderson, K. Stone, J. M. Keller, and J. Rose, "Anomaly detection ensemble fusion for buried explosive material detection in forward looking infrared imaging for addressing diurnal temperature variation," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83570T–83570T.

[9] D. T. Anderson, K. E. Stone, J. M. Keller, and C. J. Spain, "Combination of Anomaly Algorithms and Image Features for Explosive Hazard Detection in Forward Looking Infrared Imagery," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 1, Feb. 2012, pp. 313 –323.

[10] D. T. Anderson, A. Zare, and S. Price, "Comparing Fuzzy, Probabilistic, and Possibilistic Partitions Using the Earth Mover's Distance," *Fuzzy Systems, IEEE Transactions on*, vol. 21, no. 4, 2013, pp. 766–775.

[11] J. Arabas, Z. Michalewicz, and J. Mulawka, "GAVaPS-a genetic algorithm with varying population size," *Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on*. IEEE, 1994, pp. 73–78.

[12] E. Burke, S. Gustafson, and G. Kendall, "Diversity in genetic programming: an analysis of measures and correlation with fitness," *Evolutionary Computation, IEEE Transactions on*, vol. 8, no. 1, Feb 2004, pp. 47–62.

[13] F. W. Campbell and J. Robson, "Application of Fourier analysis to the visibility of gratings," *The Journal of Physiology*, vol. 197, no. 3, 1968, p. 551.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, 2011, p. 27.

[15] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," *Computer Vision–ECCV 2006*, Springer, 2006, pp. 428–441.

[16] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision research*, vol. 20, no. 10, 1980, pp. 847–856.

[17] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A*, vol. 2, no. 7, 1985, pp. 1160–1169.

[18] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, 2008, pp. 25–46.

[19] S. Edelman, N. Intrator, and T. Poggio, "Complex cells and object recognition," 1997.

[20] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A Library for Large Linear Classification," *J. Mach. Learn. Res.*, vol. 9, June 2008, pp. 1871–1874.

[21] J. Farrell, T. C. Havens, K. Ho, J. M. Keller, T. T. Ton, D. C. Wong, and M. Soumekh, "Detection of explosive hazards using spectrum features from forward-looking ground penetrating radar imagery," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 80171E–80171E.

[22] J. M. Foley and G. E. Legge, "Contrast detection and near-threshold discrimination in human vision," *Vision research*, vol. 21, no. 7, 1981, pp. 1041–1053.

[23] P.-E. Forssen and D. Lowe, "Shape Descriptors for Maximally Stable Extremal Regions," *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, Oct. 2007, pp. 1 –8.

[24] H. Frigui and P. Gader, "Detection and Discrimination of Land Mines in Ground-Penetrating Radar Based on Edge Histogram Descriptors and a Possibilistic -Nearest Neighbor Classifier," *Fuzzy Systems, IEEE Transactions on*, vol. 17, no. 1, Feb. 2009, pp. 185–199.

[25] F. Goudail, P. Réfrégier, and G. Delyon, "Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images," *J. Opt. Soc. Am. A*, vol. 21, no. 7, July 2004, pp. 1231–1240.

[26] T. C. Havens, K. Ho, J. Farrell, J. M. Keller, M. Popescu, T. T. Ton, D. C. Wong, and M. Soumekh, "Locally adaptive detection algorithm for forward-looking ground-penetrating radar," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 76642E–76642E.

[27] T. C. Havens, J. M. Keller, K. Ho, T. T. Ton, D. C. Wong, and M. Soumekh, "Narrow-band processing and fusion approach for explosive hazard detection in FLGPR," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 80171F–80171F.

[28] T. C. Havens, C. J. Spain, K. Ho, J. M. Keller, T. T. Ton, D. C. Wong, and M. Soumekh, "Improved detection and false alarm rejection using FLGPR and color imagery in a forward-looking system," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 76641U–76641U.

[29] T. C. Havens, K. Stone, D. T. Anderson, J. M. Keller, K. Ho, T. T. Ton, D. C. Wong, and M. Soumekh, "Multiple kernel learning for explosive hazard detection in forward-looking ground-penetrating radar," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83571D–83571D.

[30] 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, no. 4, April 2008, pp. 1177–1191.

[31] G. Kutyniok, M. Shahram, and X. Zhuang, "ShearLab: A rational design of a digital parabolic scaling algorithm," *SIAM Journal on Imaging Sciences*, vol. 5, no. 4, 2012, pp. 1291–1332.

[32] M. Laffin, M. A. Mohamed, A. Etebari, and M. Hibbard, "Fusion techniques for hybrid ground-penetrating radar: electromagnetic induction landmine detection systems," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 76641R–76641R.

[33] D. Lewis, J. Keller, M. Popescu, and K. Stone, "Dirt road segmentation using color and texture features in color imagery," *Computational Intelligence for Security and Defence Applications (CISDA), 2012 IEEE Symposium on*, July 2012, pp. 1–6.

[34] K. Lillywhite, D.-J. Lee, B. Tippetts, and J. Archibald, "A feature construction method for general object recognition," *Pattern Recognition*, vol. 46, no. 12, 2013, pp. 3300–3314.

[35] W.-Q. Lim, "The discrete shearlet transform: A new directional transform and compactly supported shearlet frames," *Image Processing, IEEE Transactions on*, vol. 19, no. 5, 2010, pp. 1166–1180.

[36] D. Lowe, "Object recognition from local scale-invariant features," *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, vol. 2, pp. 1150 –1157 vol.2.

[37] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, Nov. 2004, pp. 91–110.

[38] R. H. Luke, J. M. Keller, and J. Chamorro-Martinez, "Extending the scale invariant feature transform descriptor into the color domain," *Proc ICGST Int J Graph Vis Image Process, GVIP*, vol. 8, 2008, pp. 35–43.

[39] MathWorks, Jan. 2013, http://www.mathworks.com/help/images/ref/adapthisteq.html.

[40] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, 1996, pp. 51–59.

[41] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, and J. B. Zimmerman, "Adaptive histogram equalization and its variations," *Comput. Vision Graph. Image Process.*, vol. 39, no. 3, Sept. 1987, pp. 355–368.

[42] M. Popescu, K. Stone, T. Havens, D. Ho, and J. Keller, "Anomaly detection in forward looking infrared imaging using one-class classifiers," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 76642B–76642B.

[43] M. Popescu, K. Stone, and J. M. Keller, "Detection of targets in forward-looking infrared imaging using a multiple instance learning framework," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 80171Z–80171Z.

[44] S. R. Price, D. T. Anderson, and R. H. Luke, "An improved evolution-constructed (iECO) features framework," *Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2014 IEEE Symposium on*, Dec 2014, pp. 1–8.

[45] S. R. Price, D. T. Anderson, R. H. Luke, K. Stone, and J. M. Keller, "Automatic detection system for buried explosive hazards in FL-LWIR based on soft feature extraction using a bank of Gabor energy filters," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013, pp. 87091B–87091B.

[46] S. R. Price, D. T. Anderson, K. Stone, and J. M. Keller, "Investigation of context, soft spatial and spatial-frequency domain features for buried explosive hazard detection in FL-LWIR," *Proc. SPIE 9072, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets*, 2014, pp. 907217–907217–14.

[47] S. R. Price, D. T. Anderson, C. Wagner, T. C. Havens, and J. M. Keller, "Indices for Introspection on the Choquet Integral," *Advance Trends in Soft Computing*, Springer, 2014, pp. 261–271.

[48] C. Ratto, P. Torrione, and L. 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, May 2011, pp. 1689–1700.

[49] G. Scott and D. Anderson, "Importance-weighted multi-scale texture and shape descriptor for object recognition in satellite imagery," *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, July 2012, pp. 79 –82.

[50] 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, no. 1, Jan. 2011, pp. 451–461.

[51] C. J. Spain, D. T. Anderson, J. M. Keller, M. Popescu, and K. E. Stone, "Gaussian mixture models for measuring local change down-track in lwir imagery for explosive hazard detection," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 80171Y–80171Y.

[52] K. Stone, J. Keller, D. Anderson, and D. Barclay, "An automatic detection system for buried explosive hazards in FL-LWIR and FL-GPR data," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83571E–83571E.

[53] K. Stone, J. Keller, M. Popescu, T. Havens, and K. Ho, "Forward looking anomaly detection via fusion of infrared and color imagery," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 766425–766425.

[54] K. Stone, J. M. Keller, M. Popescu, and C. J. Spain, "Buried explosive hazard detection using forward-looking long-wave infrared imagery," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 801725–801725.

[55] S. L. Tantum, K. D. Morton Jr, L. M. Collins, and P. A. Torrione, "Frequency domain electromagnetic induction sensor data feature extraction and processing for improved landmine detection," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 80170B–80170B.

[56] N. T. Thanh, H. Sahli, and D. N. Hao, "Infrared Thermography for Buried Landmine Detection: Inverse Problem Setting," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 12, Dec. 2008, pp. 3987–4004.

[57] A. C. Turlapaty, Q. Du, and N. H. Younan, "A Partially Supervised Approach for Detection and Classification of Buried Radioactive Metal Targets Using Electromagnetic Induction Data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 1, 2013, pp. 108–121.

[58] R. K. Ursem, "Multinational evolutionary algorithms," *Proceedings of the Congress on Evolutionary Computation*. 1999, pp. 1633–1640, IEEE Press.

[59] VLFeat, Jan. 2013, http://www.vlfeat.org/overview/mser.html.

[60] T. Wang, J. Keller, P. Gader, and O. Sjahputera, "Frequency Subband Processing and Feature Analysis of Forward-Looking Ground-Penetrating Radar Signals for Land-Mine Detection," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 3, March 2007, pp. 718–729.

[61] T. Wang, O. Sjahputera, J. M. Keller, and P. D. Gader, "Feature analysis for forward-looking landmine detection using GPR," *Defense and Security*. International Society for Optics and Photonics, 2005, pp. 1233–1244.

[62] T. Wang, O. Sjahputera, J. M. Keller, and P. D. Gader, "Landmine detection using forward-looking GPR with object tracking," *Defense and Security*. International Society for Optics and Photonics, 2005, pp. 1080–1088.

[63] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 32–39.

[64] N. Xiang and J. Sabatier, "Laser Doppler vibrometer-based acoustic landmine detection using the fast M-sequence transform," *Geoscience and Remote Sensing Letters, IEEE*, vol. 1, no. 4, Oct. 2004, pp. 292–294.

[65] H.-F. Yu, C.-J. Hsieh, K.-W. Chang, and C.-J. Lin, "Large linear classification when data cannot fit in memory," *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2010, KDD '10, pp. 833–842, ACM.

[66] S. Yuksel, J. Bolton, and P. Gader, "Landmine detection with Multiple Instance Hidden Markov Models," *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*, Sept. 2012, pp. 1 –6.

[67] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary computation*, vol. 8, no. 2, 2000, pp. 173–195.