

A Large-Scale Multi-Institutional Evaluation of Advanced Discrimination Algorithms for Buried Threat Detection in Ground Penetrating Radar

Jordan M. Malof¹, *Member, IEEE*, Daniël Reichman, Andrew Kareem, Hichem Frigui, *Member, IEEE*,
K. C. Ho², *Fellow, IEEE*, Joseph N. Wilson, *Member, IEEE*, Wen-Hsiung Lee,
William J. Cummings, and Leslie M. Collins, *Senior Member, IEEE*

Abstract—In this paper, we consider the development of algorithms for the automatic detection of buried threats using ground penetrating radar (GPR) measurements. GPR is one of the most studied and successful modalities for automatic buried threat detection (BTD), and a large variety of BTD algorithms have been proposed for it. Despite this, large-scale comparisons of GPR-based BTD algorithms are rare in the literature. In this paper, we report the results of a multi-institutional effort to develop advanced BTD algorithms for a real-world GPR BTD system. The effort involved five institutions with substantial experience with the development of GPR-based BTD algorithms. In this paper, we report the technical details of the advanced algorithms submitted by each institution, representing their latest technical advances, and many state-of-the-art GPR-based BTD algorithms. We also report the results of evaluating the algorithms from each institution on the large experimental data set used for development. The experimental data set comprised 120 000 m² of GPR data using surface area, from 13 different lanes across two U.S. test sites. The data were collected using a vehicle-mounted GPR system, the variants of which have supplied data for numerous publications. Using these results, we identify the most successful and common processing strategies among the submitted algorithms, and make recommendations for GPR-based BTD algorithm design.

Index Terms—Buried threat detection (BTD), ground penetrating radar, landmine detection.

Manuscript received June 7, 2018; revised October 24, 2018 and February 22, 2019; accepted March 12, 2019. Date of publication May 2, 2019; date of current version August 27, 2019. This work was supported by the U.S. Army CCDC Night Vision and Electronic Sensors Directorate through the Army Research Office under Grant W911NF-06-1-0357, Grant W911NF-13-1-0065, Grant W911NF-14-1-0589, and Grant W911NF-13-1-0002. (Corresponding author: Jordan M. Malof.)

J. M. Malof and D. Reichman are with the Electrical and Computer Engineering Department, Duke University, Durham, NC 27708 USA (e-mail: jmmalo03@gmail.com).

A. Kareem and H. Frigui are with the Electrical and Computer Engineering Department, University of Louisville, Louisville, KY 40292 USA.

K. C. Ho is with the Electrical Engineering Department, University of Missouri, Columbia, MO 65211 USA.

J. N. Wilson is with the Computer and Information Science and Engineering Department, University of Florida, Gainesville, FL 32611 USA.

W.-H. Lee and W. J. Cummings are with Chemring Sensors & Electronic Systems, Inc., Dulles, VA 20166 USA.

L. M. Collins is with the Electrical Engineering Department, Duke University, Durham, NC 27708 USA.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2019.2909665

I. INTRODUCTION

IN THIS paper, we consider the development of algorithms for the automatic detection of buried threats in ground penetrating radar (GPR) data. GPR is one of the most well studied and successful modalities for buried threat detection (BTD), and a large variety of BTD algorithms have been proposed in the literature for GPR-based BTD [1]–[18]. For example, GPR-based BTD algorithms have employed techniques from fields as varied as statistics [19]–[21], computer vision [22]–[24], and machine learning [6], [25]–[27].

Prolific development within the research community has advanced the effectiveness of GPR-based BTD systems; however, most modern studies focus on proposing new algorithms, and they often compare their results against just one or two other algorithms [2], [22], [23], [28]–[31]. Systematic comparisons of modern algorithms are rare, and therefore it can be difficult to discern which algorithms, and more generally which processing practices, are best.

A. Multi-Institutional Comparison of Algorithms

In this paper, we report the results of a recent multi-institutional effort to develop, and compare, advanced BTD algorithms. The effort involved five institutions with substantial GPR-based BTD experience: Duke University, University of Louisville, University of Missouri, University of Florida, and Chemring Sensors and Electronic Systems (CSES). A major objective of this effort was to identify the best processing approaches, and evaluate them in an unbiased manner, for potential inclusion in a real-world BTD system. As a result, each institution was provided with the same period of time for algorithm development; the same experimental data set; and advanced knowledge of the scoring criteria.

The institutions were specifically tasked with developing *discriminators*. Discriminators must accept a small cube of GPR data (e.g., centered at a suspicious spatial “alarm” location), and return a decision statistic. The decision statistic indicates the relative likelihood that a buried threat is located at that location. An example cube of GPR data is shown in Fig. 1. The discriminators produced in this effort were compared using a large GPR data set collected using

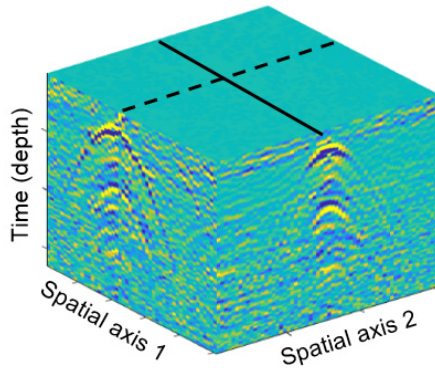


Fig. 1. Illustration of a cube of GPR data of the kind considered in this paper. A slice of the data (i.e., an image) indicated by the solid line is projected onto Spatial axis 1, and the slice of data indicated by the dashed line is projected onto Spatial axis 2, for visualization of the contents of the GPR cube. Discrimination algorithms must distinguish between data cubes corresponding to threats and nonthreats. This particular example corresponds to data collected at the location of a buried threat.

a vehicle-mounted GPR-based BTM system, the variants of which have been involved in numerous publications over the preceding years [18], [23], [25], [32]–[36]. The GPR data set collected using this system, and used to compare the discriminators, was comprised of 120 000 m² of surface area, collected over 13 lanes at two different U.S. test sites, and encompassing 4552 buried threat encounters.

B. Contributions of This Paper

The first major contribution of this paper is a technical description of the discrimination algorithms submitted by each institution. While some algorithms represent the latest advances to previously proposed GPR-based BTM discriminators, others are novel GPR-based BTM discriminators. In either case, these algorithms incorporate the latest technical advances from each institution, and therefore represent state-of-the-art algorithms within the GPR-based BTM community.

The second major contribution of this paper is a presentation of the results of the algorithm performance comparison, including a breakdown of the performance by burial depth. A simple fusion of the algorithms is also presented to assess the complementarity of the approaches.

Our final contribution is an analysis of the discriminator designs. We discuss the algorithm design practices that our group of collaborators has found to be effective over the last decade. We also discuss the conceptual differences between the discriminators in this paper, and based on the experimental results, we draw conclusions about the effectiveness of these different design concepts.

This paper builds on a previous discriminator comparison published by our consortium of research institutions approximately ten years ago [18]. That paper received significant interest in the research community, motivating us to share our latest findings and insights in this paper. This paper builds on our previous findings in many respects and, where appropriate, we will draw comparisons with [18] to help clarify any new insights.

C. Omissions of Proprietary Information

The GPR system employed in this investigation is proprietary, and therefore we were required to omit certain details about the 1) technical operation of the system and 2) the system's precise detection performance. Information about these omissions is provided directly in the text when they occur, and here we explain the broad impact of these omissions on our results and conclusions. Regarding 1), we believe that our GPR system is sufficiently similar to others that readers can readily utilize our conclusions in their own work. For example, our GPR system is the same as, or fundamentally similar to, systems that have been employed in numerous recent publications [18], [23], [25], [32]–[36]. Furthermore, our GPR uses the same fundamental sensing mechanism (which we do describe) as most others in the broader literature. Regarding 2), we provide sufficient detail about the system's performance so that it is possible to effectively compare the detection capabilities of the discrimination algorithms—the main goal of this paper. Our scoring procedures related to this are described in Section III-C.

D. Paper Organization

Section II describes the GPR system used to collect data for this paper, and provides details regarding the experimental data set. In Section III, we discuss the experimental design that was used to evaluate the discriminators: the full detection processing pipeline for the GPR BTM system, data handling (e.g., how the data were used for training and testing supervised algorithms), and the scoring criteria. Section IV provides a broad overview of the prescreening algorithm, which identifies suspicious locations for processing by the discrimination algorithms. The technical details of the prescreeners are provided in Appendix I. Section V presents the technical details of the four discrimination algorithms submitted for the algorithm comparison. Section VI presents the experimental results, and Section VII presents some additional analysis of the results, including fusion of the discriminator outputs, and a breakdown of their performance by target burial depth. Section VIII presents our analysis of the discriminator designs, and the implications of the experimental results on those designs. Section IX presents our conclusions, and our recommendations for major areas of future work improving discriminators.

II. GPR SENSOR AND DATA SET

In this section, we describe the vehicle-mounted radar system that was used to collect the GPR data used in our experiments, as well the GPR data set that was collected using this system.

A. Vehicle-Mounted Radar System

The radar system employed to collect GPR data is similar to those that have been involved in numerous publications over the preceding years [18], [23], [25], [32]–[36]. The system is comprised of an array of radar antennas that are mounted on the front of a vehicle. The array is oriented perpendicular to the direction of travel (i.e., cross-track), and the antennas are

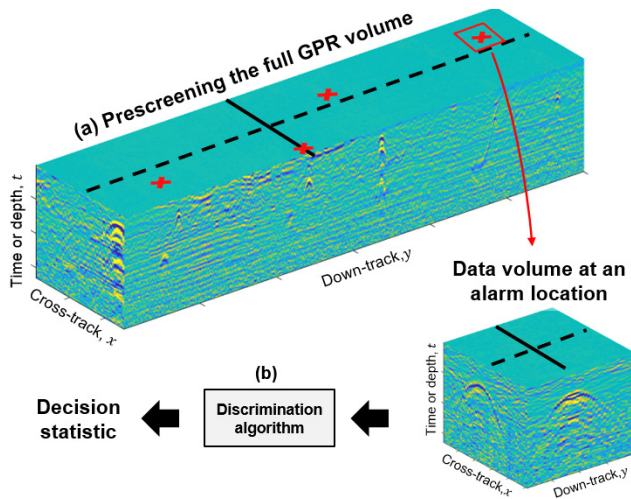


Fig. 2. Illustration of the two-stage detection processing approach employed in the BTM system in this paper. (a) First stage of processing is “prescreening,” in which a relatively fast algorithm (a prescreener) is applied to the raw incoming volume of GPR data to identify a small set of suspicious locations, called alarms, for processing by the more computationally intensive discrimination algorithms (or discriminators). The prescreening operation reduces the amount of data considerably, making it possible to apply the discriminators in real time (e.g., on a moving vehicle). (b) In the second stage of processing, discriminators are applied to cubes of GPR data extracted at each alarm location identified by the prescreener. The goal of the discriminators is to assign a “decision statistic” to each alarm location indicating the relative likelihood that the data correspond to a true buried threat. Note that the discriminators operate only on the locations identified by the prescreener, and therefore any buried threats that are missed by the prescreener cannot be identified by the discriminators.

equally spaced along the array. During data collection, each antenna emits an ultra-wideband radar signal, consisting of a differentiated Gaussian pulse, and then measures the energy reflected back toward the array from the subsurface.

The result of this procedure is a timeseries of signal strengths, known as an A-scan [12], [23], from each antenna. As the vehicle moves forward down a path or lane, referred to as the “down-track” direction, each antenna records A-scans at regular down-track (spatial) intervals. The recording of cross-track A-scans (across the array), and down-track A-scans (along the direction of travel) results in a volume of GPR data. An example of a large GPR data volume is illustrated in Fig. 2.

B. GPR Data Set Details

The data in this experiment were collected at two distinct U.S. testing sites, designated “Site A” and “Site B.” A series of testing lanes (or pathways) was constructed at each site. The lanes were similarly smooth, and absent of major rough terrain with only a few exceptions. The lanes are predominantly dirt and gravel. A mixture of government manufactured and improvised threats were buried at known locations throughout each lane in order to evaluate the detection capabilities of the BTM system. The objects were buried at various depths, and contain varying levels of metal content. The burial depths and composition of the threat population is summarized in Table I. Deliberately placed man-made clutter objects were excluded from our data sets after data collection. This is because these objects can have GPR signatures similar to threat objects and

TABLE I

METAL CONTENT AND BURIAL DEPTH OF THE EXPERIMENTAL THREAT POPULATION. THE PRECISE DEPTH RANGES FOR EACH BURIAL DEPTH CATEGORY HAVE BEEN OMITTED TO OBSCURE THE PRECISE PERFORMANCE CHARACTERISTICS OF THE PROPRIETARY GPR-BASED BTM SYSTEM. THE “STANDARD BURIAL DEPTHS” CATEGORY HAS A MEDIAN BURIAL DEPTH OF APPROXIMATELY FOUR INCHES. “DEEP BURIAL DEPTHS” CORRESPONDS TO THE DEEPEST 10 PERCENT OF THREATS, AND HAS A GREATER MEDIAN DEPTH

	Metal	Low Metal	Non-metal	Total
Standard burial depths	1441	2121	465	4027
Deep burial depths	308	0	217	525

were emplaced in the lanes for different purposes. Naturally occurring clutter (e.g., rocks, plant roots) was not excluded. If these are detected by the discriminator algorithms as threat objects, they are treated as false alarms.

Site A was located in a temperate geographic location, and was comprised of 449 unique threats, distributed over seven lanes. The GPR system was typically driven over each individual lane (referred to as a “pass” or “run” over a lane) multiple times, often in different directions, or on different days, in order to capture common variations in real-world GPR data. A total of 56 runs were made at Site A, resulting in a total of 48 000 m² of scanned lane area, and 3368 threat encounters. Site B was located at an arid geographic location, and was comprised of 210 unique threats, distributed over six lanes. The 34 total passes were made over the lanes, resulting in a total of roughly 72 000 m² of scanned lane area, and 1184 threat encounters. In total, the testing data consisted of 90 runs over 13 unique lanes, encompassing roughly 120 000 m² of lane area.

III. EXPERIMENTAL DESIGN

In this section, we present details of the experimental design employed in this paper to evaluate and compare the performance of the discrimination algorithms submitted by each institution. Importantly, this section also lays out the design specifications and/or constraints for the algorithms that were provided to each institution during development.

A. Two-Stage Detection Processing Pipeline

The full detection algorithm, or processing pipeline, employed on our BTM system is comprised of two sub-processes: prescreening and discrimination (i.e., classification). This processing pipeline is illustrated and described in Fig. 2, and has been applied in numerous previous studies that considered GPR-based BTM [1], [18], [23], [37]–[39]. The prescreener employed in our experiments was developed by the CSES Corporation, while the discriminators were developed by at least one of the other institutions (all of which are Universities). We will describe both systems, but the focus of this paper is on the development of the discriminators.

As discussed in Fig. 2, each discriminator must accept a cube of GPR data as its input, and return a real number

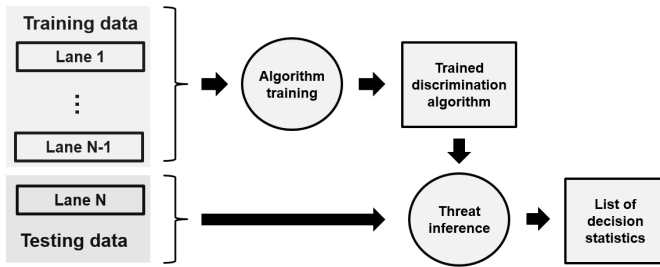


Fig. 3. Illustration of the lane-based cross-validation procedure used to evaluate the discrimination algorithms. Training of the supervised portions of the discrimination algorithms (i.e., parameter inference) is performed on prescreener alarms from $N - 1$ lanes. This process yields a trained discrimination algorithm that can be deployed for threat inference on new, previously unseen, GPR data. The trained algorithm is applied for threat inference on prescreener alarms on the remaining lane that was withheld from training. This yields a list of decision statistics, indicating the relative likelihood that each of the prescreener alarms corresponds to a true buried threat. This process is performed N times, so that each lane is employed exactly once for threat inference. The decision statistics from each lane are then aggregated and used to compute performance metrics for the algorithm.

indicating the relative likelihood that the location under consideration contains a buried threat. These cubes are centered at suspicious spatial locations, called alarms, that are identified by the prescreener. The cube of GPR data has a predetermined spatial extent, imposed by the need to process the data in real-time during system operation. Similarly, the radar system collects A-scans of a predetermined, and fixed, (temporal) length. All discriminators were required to operate within these constraints.

B. Data Handling: Discriminator Training and Testing

All of the discriminators submitted for this comparison contain some type of supervised statistical, or machine learning, model. Supervised models have parameters that must be inferred using labeled examples of the classes of data we wish to categorize (e.g., threats and nonthreats in this case). This process of parameter inference is often referred to as *training*, and it results in a trained algorithm that can then be applied to new, unlabeled, data in order to infer whether it is a threat or not (i.e., *threat inference*). Supervised classification models obtain excellent performance when sufficient training data is available, and they have now become popular in the GPR BTM literature [2], [18], [25], [38], [40], [41].

A common approach to assess the performance of any supervised algorithm (i.e., an algorithm containing a supervised model) is to employ cross-validation. This process involves training the models on some subset of the data, and then evaluating their ability to discriminate between the desired data classes on data that was excluded from training. In this paper, the discrimination algorithms were evaluated using a lane-based cross-validation approach, which we illustrate and describe in Fig. 3.

C. Scoring the Algorithms

The output of the discriminators (and prescreener) is a list of spatial alarm locations and their associated decision statistics. Before scoring these alarms, we must establish which of them

should be considered correct alarms (i.e., they occur over real threats), and which of them are false alarms. In this paper, any alarm located within a radius of 0.25 m of a buried threat was considered to be a correct alarm, while all other alarms were considered false. This is a popular criterion that has been employed in numerous previous GPR-based BTM studies [30], [33], [42]–[45].

Given the true identity of each alarm (threat or nonthreat), we can score the performance of the discriminators and the prescreener. In this paper, we employ receiver operating characteristic (ROC) curves to measure performance. ROC curves quantify the tradeoff between the *correct* alarm rate (or probability of detection), P_d , and the *false* alarm rate (or probability of false detection), FAR, as we vary the sensitivity of the algorithm. The ROC curve is a popular metric in the GPR BTM literature [1], [6], [18], [23], [39], [46], where it is common to scale the FAR metric so that it corresponds to the number of false alarms per square meter of scanned surface area. This representation of the BTM system's FAR is often more interpretable and operationally relevant than false alarm probabilities. Unless otherwise stated, any reference to *FAR* in this paper refers to false alarms per square meter.

The vehicle-mounted BTM system considered in this paper is proprietary, and therefore it was necessary to obscure its precise performance capabilities. In order to meet this need, while still effectively comparing the algorithms, we omitted the upper FAR values from the ROC curves reported in this paper (i.e., all x -axis begin at 0, but all other values are omitted). Although FARs were omitted, all of the ROC curves in this paper (except Fig. 9 involving the prescreener) use exactly the same range of FAR values on the x -axis. This makes it possible to compare the relative performance capabilities of the algorithms under each scoring scenario—the main goal of this paper. The presented range of FAR values corresponds to those FARs that were considered most operationally relevant for the discrimination algorithms by our sponsors (i.e., the organizations funding this paper—see acknowledgment). Consequently, the algorithm designers were tasked with developing discrimination algorithms to achieve the highest possible P_d over this range.

IV. PRESCREENING ALGORITHM

In this section, we provide a brief overview of the prescreening algorithm, which is comprised of the fusion of two prescreeners: F2 and CCY, respectively. Since discriminators are the focus of this paper, this section is brief and the technical details of the prescreeners are provided in Appendix I. Furthermore, the major design details and parameter settings of the prescreeners were established by prior to the discriminator comparison, and they were assumed to be fixed by the discriminator designers. However, because prescreening is an important part of the overall detection system, we provide its major technical details.

A. F2 Prescreener

The F2 prescreener is an updated version of F1 (Fast One) which in turn was derived from a least-mean-squares-based

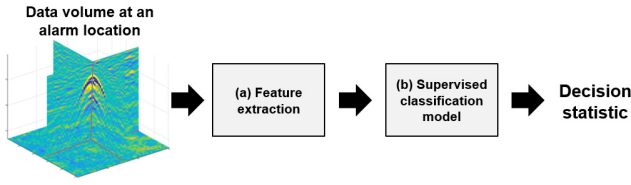


Fig. 4. Although not a design constraint, each discrimination algorithm submitted for the algorithm comparison adopted the same processing pipeline, illustrated here. (a) Feature extraction is the first step of processing, and consists of extracting measures or statistics from the GPR data, with the aim of concisely (e.g., in a low dimensional vector) summarizing the important characteristics of the data. For example, many features attempt to encode the strength and orientation of edges in GPR slices. (b) Features extracted in (a) are provided to a supervised classification model that has been trained to distinguish between features corresponding to threats and nonthreats, respectively. The classifiers assign a decision statistic to the input feature, indicating the relative likelihood that it corresponds to a threat.

prescreener developed by Torriane *et al.* [42] to reduce its run time while maintaining comparable performance. The overall strategy of F2 is to identify locations in the GPR data with high signal energy, relative to the surrounding data. This strategy is implemented through a series of smoothing and energy-based anomaly filtering.

B. CCY Prescreener

To provide prescreening information that is complementary to F2, CSES developed a shape-based prescreener called concavity (CCY). Given that the signal returns from real buried threats manifest as hyperbolas in the GPR data, the estimated convexity, or CCY of signals in the data serve as a useful metric for identifying threats. This insight forms the basis for the CCY prescreener.

C. Fusion of the Prescreeners

The F2 and CCY prescreeners were developed independently, and can both function as standalone prescreeners. However, the two prescreeners identify threats using fundamentally different criteria: F2 searches for energy-based anomalies, while CCY searches for hyperbolic patterns. Investigations within our group revealed that the two prescreeners identify somewhat different subsets of threats (which is corroborated by experimental results in Section VI-A), and a set of decision rules was developed to fuse the alarms produced by each prescreener while still maintaining real-time operability. This process yields a new set of alarms, which are subsequently treated by the discriminators as though they were generated from a single prescreener.

V. DISCRIMINATION ALGORITHMS

This section presents the technical details of each of the discrimination algorithms submitted for the comparison. Although it was not a design constraint, all of the submitted discriminators follow the same basic processing pipeline, involving feature extraction and classification. This pipeline is illustrated and described in Fig. 4. For simplicity, we will refer to each discriminator by its feature extraction approach. For example, the first algorithm is based on the

edge histogram descriptor (EHD) feature and we will therefore refer to it as “the EHD algorithm.” Using the aforementioned nomenclature, a total of four discriminators were submitted for the comparison: EHD, log Gabor (LG), the histogram of oriented gradients for GPR (gprHOG), and spatial edge descriptors (SED). The description of each algorithm is broken down into three parts: feature extraction, classifier and training, and threat inference (i.e., how predictions are made on new data).

As described in Section III, all of these discriminators were required to operate on 3-D cubes of GPR data. All of the discriminators apply two initial preprocessing steps: 1) alignment of all A-scans so that the ground response in each A-scan occurs at the same time (depth) index and 2) the removal of all GPR data at, and above, the ground time index. LG only applies 1) for preprocessing.

All of the discriminators submitted to the comparison apply some form of depth-based calibration of the data. Although the precise approaches varied slightly, this calibration procedure always consisted of normalizing, or whitening, each pixel by removing a locally computed mean from the data, and dividing by a locally computed standard deviation. For example, at a given time index (depth), the mean and variance can be computed using a set of leading and trailing GPR samples, and subsequently used for whitening all pixels at that time index.

A. Edge Histogram Descriptors

The EHD uses translation invariant features that are based on the histogram of edges in the GPR signature [32]. It is an adaptation of the MPEG-7 EHD feature [47] which encodes image texture information. It has been adapted to capture the spatial distribution of the edges within a 3-D GPR volume. To keep the computation simple, 2-D edge operators are used, and two types of edge histograms are computed. The first one, called EHD^{DT} , is obtained by fixing the cross-track dimension and extracting edges in the (time, down-track) plane. The second edge histogram, called EHD^{CT} , is obtained by fixing the down-track dimension and extracting edges in the (time, cross-track) plane.

1) *Feature Extraction:* The EHD^{DT} and EHD^{CT} features are extracted from a GPR volume located at the prescreener alarm location with sizes $(T, X, D) = (60, 15, 15)$. Let $S(t, x, y)$ denote this volume, and let $S_{ty}^{(x)}$ denote the x th plane of S . First, for each $S_{ty}^{(x)}$, four categories of edges are computed using 3×3 Sobel filters: vertical, horizontal, 45° diagonal, and 135° antidiagonal. If the maximum of the edge strengths exceeds a preset threshold, the corresponding pixel is considered to be an edge pixel and is labeled according to the direction of the maximum edge. Otherwise, it is considered a nonedge pixel. Next, each $S_{ty}^{(x)}$ image is vertically subdivided into seven overlapping subimages $S_{ty,i}^{(x)}, i = 1, \dots, 7$. For each $S_{ty,i}^{(x)}$, a five-bin edge histogram, $H_{ty,i}^{(x)}$, is computed. The bins correspond to the four-edge categories and the nonedge pixels. The EHD^{DT} is defined as the concatenation of the seven five-bin histograms. That is, $\text{EHD}^{\text{DT}}(S(t, x, y)) = [\bar{H}_{ty,1} \dots \bar{H}_{ty,7}]$, where $\bar{H}_{ty,i}$ is the cross-track average of

the edge histograms of subimages $S_{ty}^{(x)}$ over the seven middle channels, i.e., $\bar{H}_{ty,i} = (1/7) \sum_{x=5}^{11} H_{ty,i}^{(x)}$. To compute the EHD output in the cross-track direction, i.e., EHD^{CT} , the y plane of $S(t,x,y)$ is fixed and edges are extracted from $S_{tx}^{(y)}$ in a similar way.

2) *Classifier and Training*: Support vector machine (SVM) [48] classifiers, with the radial basis function, were used for class prediction. One SVM for the EHD^{DT} features (SVM^{DT}) is trained and a second SVM for the EHD^{CT} (SVM^{CT}) is also trained. Both classifiers were implemented using the libSVM package [49]. All parameters were set to their default values.

At each spatial location indicated by the prescreener, EHD^{DT} and EHD^{CT} features are extracted at multiple depths down the temporal axis by sampling one $S(t,x,y)$ every 25 temporal indices. That is, we extract features from $S(t = 1..60, x, y)$, $S(t = 25..84, x, y)$, ... This results in a total of 14 EHD features, f_i , $i = 1, \dots, 14$, per spatial location identified by the prescreener. For nontarget alarms, any of the 14 features could be included in the training data. To maintain a balance between the number of training samples from both classes, we randomly select 5 of the 14 features. For targets, we use a kernel density estimator (KDE) to identify the few features that correspond to the most likely temporal location of the buried threat signal. Let $P^- = \{p_1^-, p_2^-, \dots, p_k^-\}$ be a set of k prototypes that summarize all of the nontarget training alarms. Then, for each target training alarm, we estimate the KDE of its 14 EHD feature using

$$\text{KDE}(f_i) = \frac{1}{Z} \sum_{j=1}^k \exp(-\beta \|f_i - p_j^-\|) \quad (1)$$

where β is a resolution parameter (learned during summarization of the nontarget training alarms) and Z is a normalization factor. The f_i features with very low KDE (close to zero) are selected as the most likely temporal locations that correspond to the actual target signatures and will be used for training.

3) *Threat Inference*: For threat inference on new data, 14 EHD^{DT} features and 14 EHD^{CT} features are extracted from each prescreener alarm at multiple temporal locations. Then, the trained SVM^{DT} and SVM^{CT} classifiers are used to assign confidence values to each EHD^{DT} and EHD^{CT} feature, respectively. Next, we fuse the decision statistics from both directions by taking their geometric mean at each temporal location. The final decision statistic is computed by summing the three top fused values.

B. Log Gabor

The Gabor filter, which is essentially a modulated Gaussian function at some frequency f_0 , has been useful for many filtering tasks in signal processing. The Gabor filter bank is a series of Gabor filters created by imposing the constraint that the standard deviation governing the spread of the Gaussian function is inversely proportional to the modulation frequency f_0 . Allowing the modulation frequency to increase in a dyadic manner creates the Gabor wavelets that are common for time–frequency signal analysis [50]. A distinct

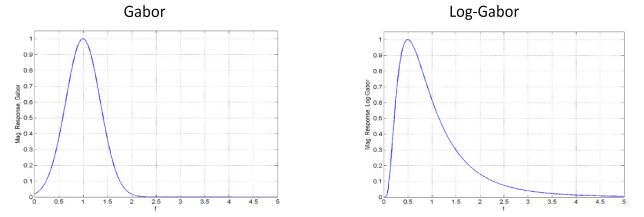


Fig. 5. Frequency response of the Gabor and the LG filter.

property of the Gabor filter is that its Fourier transform follows a Gaussian shape as well. Thereby, the Gabor wavelets define a filter bank with each bandpass filter having a Gaussian shape frequency response.

In the processing of GPR data, empirical evidence has shown that (for GPRs employing differentiated Gaussian pulses) the frequency spectrum of the radar echo reflected by threats is often asymmetric and has a long tail toward the high-frequency region [4], [51]. To better capture the characteristics of the threat signal to aid detection, we apply the LG wavelets instead. The LG filter has a Gaussian frequency response in the log-frequency axis, thereby having a long tail response in frequency. Fig. 5 illustrates the difference in frequency responses between the Gabor and the LG filter.

1) *Feature Extraction*: As noted previously, the threat signature often appears with a hyperbolic pattern in the B-scan. To capture the frequency as well as spatial responses, we shall apply the LG filters in 2-D for extracting the features for threat detection [52]. The 2-D LG filter is defined in the frequency domain through the polar coordinates. Let ρ be the radius from the center and θ the angle from the x -axis. The frequency response of the LG filter response is given by

$$G_L(\rho, \theta) = \exp\left(-\frac{1}{2[\log(\sigma_\rho)]^2} \left[\log\left(\frac{\rho}{\rho_0}\right)\right]^2\right) \times \exp\left(-\frac{1}{2\sigma_\theta^2}(\theta - \theta_0)^2\right). \quad (2)$$

Parameters ρ_0 and θ_0 control the location and the orientation of the filter response, and σ_ρ and σ_θ determine the spreads in the frequency and angle. We create a set of 36 filters to represent the LG curvelets, by using four values of ρ_0 that decompose the frequency range and nine values of θ_0 that provide an angle resolution of 20° . These 36 LG filters cover the frequency plane of a threat's GPR signal with four different filters at each of the nine orientations. Fig. 6 shows the frequency as well as the spatial responses of the filters. The left part of the filters extract the rising edge behavior in the B-scan, the right part the trailing edge and the middle portion the horizontal.

The B-scan at the prescreener alarm location is separated into three spatial regions, with overlap: the left, middle, and right. The left part of the image is processed by the filters in the first three columns in Fig. 6, the right part by the filters in columns 6 to 8 and the middle part by the filters in columns 4, 5, and 9. The filtering process is performed in the frequency domain. Each filter output is separated into 15 depth bins with 50% overlap. The element of the feature vector is the

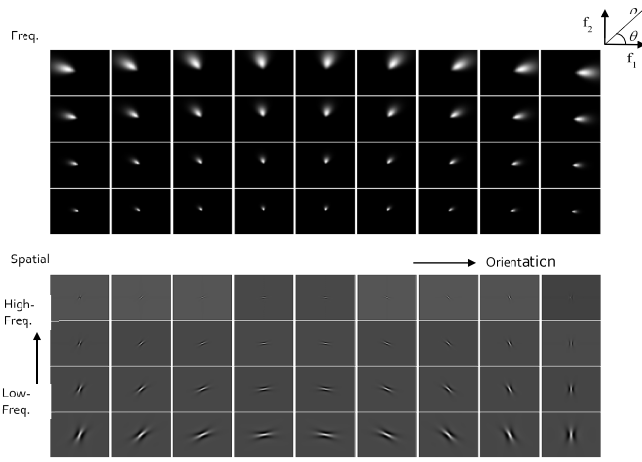


Fig. 6. Frequency and spatial responses of the 36 LG filters for feature extraction.

maximum of the energies over the 15 depth bins of each LG filter output.

In addition to applying the filters to the B-scan in the down-track, we also apply the LG filters to the 2-D image collected over the cross-track, as well as those from the positive-diagonal and antidiagonal in the 3-D data cube at the alarm location. Each 2-D image results in 36 features. Since there are four total images, this yields a final feature vector of 144 total feature values.

2) *Classifier and Training*: The feature vectors are used to train two SVM classifiers with an RBF kernel to perform target versus nontarget classification. The first SVM is trained on the feature vector associated with each alarm. The second SVM is trained on a transformation of the feature vector. Before the max operation is performed depth wise, each alarm is represented by a matrix of feature vectors of dimension 15×144 . The feature matrix is sorted in descending order of magnitude and the top four rows are used as individual feature vectors with the same label as that of the corresponding alarm. To avoid overtraining either SVM, a subset of the data is used for training. For the data set of feature vectors for each SVM, the feature vectors of deeply buried threats are replicated once (because there are fewer deeply buried threats). Second, only 5% of the resulting samples are selected randomly to train the SVM classifier.

3) *Threat Inference*: For threat inference, each SVM is applied to the feature vector transformation that corresponds to the data on which it was trained. Thus, the first SVM is tested on the feature vector that is the maximum value depth-wise for each of the 144 features for that alarm. The second SVM is tested on each of the 15 feature vectors for an alarm, resulting in 15 output confidences. The output confidences are sorted, and the sum of the top three confidences is the final confidence for this classifier. The final confidence is the sum of the confidences of the two SVM classifiers.

C. Histogram of Oriented Gradients for GPR

Histogram of oriented gradients (HOG) is an image descriptor that was originally developed in the computer

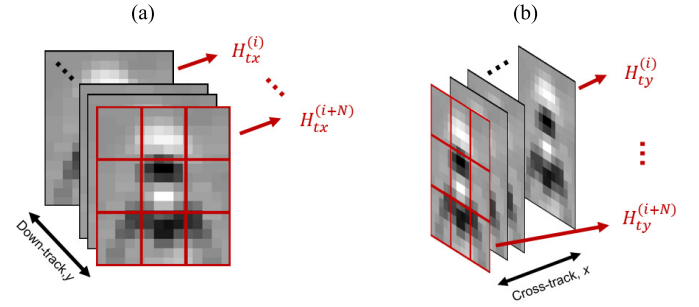


Fig. 7. Illustration of the extraction of the (a) cross-track gprHOG feature, denoted H_{tx} , and (b) the down-track gprHOG feature, denoted H_{ty} . $H_{ty}^{(i)}$ represents an HOG feature (with no block normalization) extracted on a t - x GPR slice located at the i th down-track location within the GPR cube provided for feature extraction. The final cross-track gprHOG feature is given by the average of the HOG features in each slice: $H_{tx} = (1/N) \sum_i H_{tx}^{(i)}$. Similarly, $H_{ty}^{(i)}$ represents an HOG feature extracted on a t - y GPR slice located at the i th cross-track location within the GPR cube, and the final down-track feature is given by an average of the individual HOG features.

vision literature [53]. HOG was first applied for GPR-based BTD in [23] and has since become a very popular feature in the literature for BTD [2], [7], [28], [33], [54], [55]. Here, we employ gprHOG, which incorporates several improvements to the original HOG implementation for GPR-based BTD in [23]. Further details and experimental justification for gprHOG can be found in the original paper [56].

1) *Feature Extraction*: Before feature extraction, the data are downsampled in the temporal axis by a factor of 2, in accordance with [23] (although our numbers are slightly different due to differences in the radar system). The gprHOG feature is extracted on a cube of the preprocessed GPR data, denoted $S(t, x, y)$, that is $[T, X, Y] = [18, 18, 18]$ in size. Similar to [23], we extract two gprHOG features: one cross-track feature, denoted H_{tx} , and one down-track feature, denoted H_{ty} . In [23], a single down-track B-scan was used to compute H_{tx} and H_{ty} , respectively. With gprHOG, however, H_{tx} is an average of features computed over several neighboring cross-track B-scans, as illustrated in Fig. 7. This averaging step reduces noise in the feature and improves performance. Another important modification of HOG is the removal of the histogram (or block) normalization step. It was demonstrated in [33] that this substantially improves the effectiveness of the HOG descriptor for GPR-based BTD.

2) *Classifier and Training*: In order to create a training data set, we extract four image patches at threat locations using the MSEK algorithm [38]. MSEK identifies locations along GPR time axis that exhibit high levels of signal energy. At nonthreat locations, 24 patches are extracted at regular intervals along the time axis. gprHOG features are extracted on all of the aforementioned patches and used for training. Another important improvement of the gprHOG algorithm (compared to [1]) is the use of two Random Forest classifiers [57]: one trained on down-track gprHOG features, and one trained on cross-track gprHOG features. Both classifiers are trained with 100 trees.

3) *Threat Inference*: At each prescanner alarm location, down-track and cross-track gprHOG features are extracted, respectively, at small regular intervals down the time axis.

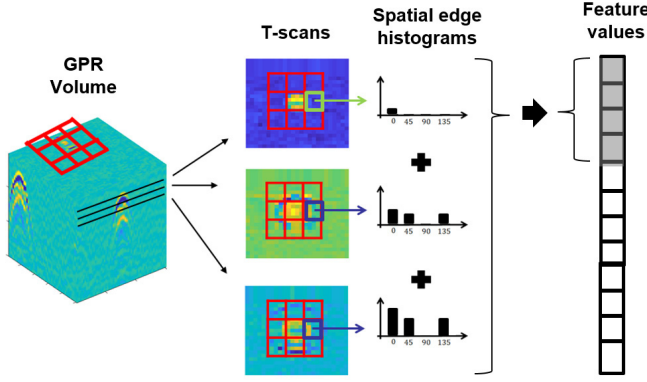


Fig. 8. Illustration of the process for computing the SED feature. Around the spatial coordinates of the prescreener alarm location and at the time-index estimated using MSEK, a 15×15 pixel patch is extracted which is divided into a 3×3 grid of cells. In each cell, a histogram of gradients is computed using four angle bins. The final descriptor vector supplied to the classifier is the average of computing SED on 50 temporal scans, with the first one computed at five temporal locations above the MSEK temporal location and the last one is 44 time samples lower.

A final decision statistic is computed for each track by summing the top 12 classifier decision statistics. The two resulting statistics, one from cross-track and one from down-track, are multiplied to obtain a final statistic for the alarm.

D. Spatial Edge Descriptors

The SED algorithm is based upon extracting shape information, via gradient histograms, in 2-D GPR images. Unlike most descriptors proposed for GPR-based BTBD, SED operates on spatial imagery: images comprised of GPR returns collected at the same instance in time. In these images, referred to as T-scans, buried threat signals appear circular rather than hyperbolic. This is illustrated in Fig. 8. The SED algorithm is designed to capture this shape to provide a descriptor of buried threats to the classifier.

1) *Feature Extraction*: The SED feature is extracted from a GPR volume located at the prescreener alarm location, $S(x, y, t)$, with sizes $(T, XT, DT) = (50, 15, 15)$. The process for computing SED is illustrated in Fig. 8. For the i th temporal sample in the volume, t_i , we extract an image $I(x, y) = S(x, y, t = t_i)$. The image is divided into a 3×3 grid of cells. In each cell, a histogram of gradients is computed using four angle bins, resulting in a 36-D descriptor for the image, $H_{xy}^{(i)}$. The final descriptor is constructed by averaging the descriptors over time, $H_{xy} = (1/T) \sum_i H_{xy}^{(i)}$. The averaging step is intended to increase the signal-to-noise ratio of the descriptor, due to uncertainty in the temporal location of the threat signal, and the tendency of the threat signal to appear over many time samples.

The gradients at each pixel are computed using four 3×3 pixel Sobel filters, each rotated by 45° . An additional bin is added in this step corresponding to “no-edge” if the response to all templates is less than a specified threshold. In that case, a count is maintained of the number of pixels in the cell whose gradient response was less than the threshold. The threshold we use is 3, and this parameter has been relatively insensitive

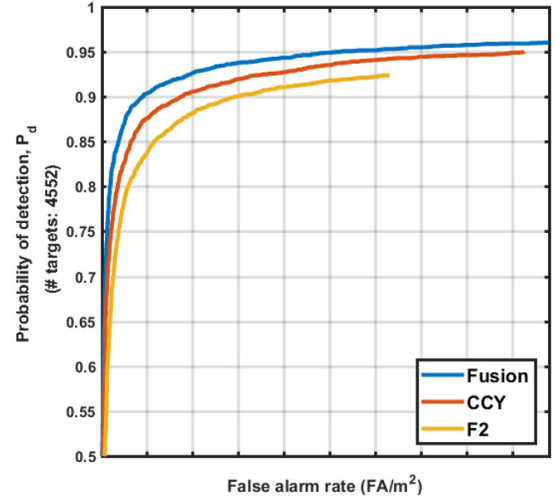


Fig. 9. ROC curves for the two individual prescreeners, F2 and CCY, as well as their fusion. Note that the y-axis has been truncated to the range $P_d = [0.5, 1]$. The x-axis labels have been omitted—see Section III-C for details.

to data set changes. We note that both using edge templates and a “no-edge” bin are similarly implemented in the EHD algorithm [32].

2) *Classifier and Training*: A SVM [48] with the radial basis function was used for class prediction. This was implemented using the libSVM package [49], with parameters $\gamma = 0.001$ and SVM cost parameter $C = 15$, which were chosen in cross-validation. At each spatial location indicated by the prescreener, the time index of the buried threat signal is estimated using an energy-based procedure referred to as MSEK [38]. The MSEK method computes the signal energy along an A-scan (i.e., over the GPR temporal dimension, smooths the energy timeseries, and then identifies local maxima. For both threat and nonthreat data, we extract two local maxima for training.

3) *Threat Inference*: During threat inference, SED features are extracted at regular intervals down the temporal (or depth) axis. The temporal locations are obtained, beginning with the first temporal index, and then by sampling one location every 25 temporal indices. This results in a total of 14 temporal locations. At each of these locations, we extract SED features at each spatial location within a 5×5 spatial grid. This results in a total of 350 full SED feature vectors for each prescreener alarm. This extraction can be done very efficiently by reusing gradient computations between neighboring SED feature vectors. A final decision statistic for a given prescreener alarm location is computed by applying the classifier to all 350 SED feature vectors, and then summing the top 25 resulting decision statistics. We have found the performance to be largely insensitive to the number of decision statistics in the summation.

VI. EXPERIMENTAL RESULTS

A. Prescreener Results

The performance results obtained by applying the prescreeners to the experimental GPR data set are presented in Fig. 9.

Each prescreener was applied to the same total area of lane, but obtained a different total number of false alarms, and so their respective ROC curves end at different FAR values. The FAR range of the x -axis in Fig. 9 has been extended so it includes the maximum FAR value across all of the prescreeners. Similarly, each prescreener missed a different total number of the true buried threats, and therefore each ROC curve reaches a different maximum P_d value (which is obtained at its corresponding maximum FAR value).

The results in Fig. 9 indicate that CCY substantially outperforms F2 at all shared values of FAR. However, the fusion of the two prescreeners (referred to as the “Fusion” prescreener in Fig. 9) obtains a much greater P_d value than either F2 or CCY at all shared values of FAR. The relative advantage of CCY over F2 suggests that it is important to leverage shape information in threat detection algorithms. CCY relies primarily (but not exclusively) on shape-based cues in the GPR data, while F2 relies primarily (though not exclusively) on signal energy. The large performance improvement yielded by their fusion demonstrates that, while a shape-focused approach has an overall advantage in our experiments, both energy and shape content appear to be important to obtain the best performance. This is implied by the substantial performance gain when fusing the two prescreeners.

Another important finding in Fig. 9 is that the Fusion prescreener obtains a greater maximum P_d than either F2 or CCY. This implies that the F2 prescreener identified some threats that were not identified by CCY, further corroborating the complementarity of CCY and F2. The discrimination algorithms, even if they perform perfectly ($P_d = 1$ with no false alarms) can never identify buried threats that were not already identified by the prescreener. As a result, the maximum P_d of a prescreener can be an important performance criterion, and one which is improved via the fusion of CCY and F2.

B. Individual Discriminator Results

This section presents the results of evaluating the discrimination algorithms using lane-based cross-validation (Section III-B) on the alarms extracted by the fusion prescreener. The decision statistics of the (fusion) prescreener alarms were thresholded at a predetermined operating sensitivity threshold, yielding 4372 threat locations (96.1% of the total present). The total number of nonthreat locations (i.e., false alarms) is omitted to obscure the precise performance capabilities of the system (see Section III-C for details). The cross-validation results for all discrimination algorithms that were applied to this data set are presented in Fig. 10.

The results indicate that no single algorithm consistently performs best (i.e., provides the best detection rate P_d) across all of the FAR values. In the lower FAR range SED performs the best, while in the higher FAR range LG performs best. EHD never provides the best P_d but it provides comparable performance to SED and LG over most of the considered FAR range. The gprHOG algorithm briefly performs best at very low FAR, but then performs poorly for the remaining FAR range, providing lower P_d than even the prescreener at most FARs.

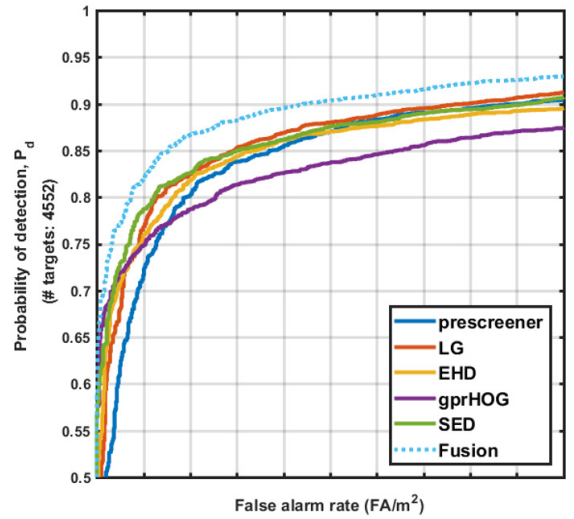


Fig. 10. ROC curves of all discrimination algorithms computed using lane-based cross-validation on the (fusion) prescreener alarms. We also include the fusion of the discriminators using a simple geometric mean. Note that the y -axis has been truncated to the range $P_d = [0.5, 1]$. Also, the x -axis labels have been omitted—see Section III-C for details. Note also that the P_d reported for each discriminator here is based upon the entire threat population, rather than the subset found by the prescreener (i.e., both Fig. 9 and Fig. 10 are computed as a proportion of 4552 total threats, as indicated on the y -axis labels).

It is not surprising that the algorithms obtain (relatively) similar performances, due to several shared fundamental processing strategies that our group members have each independently found to be effective. These shared processing strategies, as well as key differences, are elaborated in greater detail in Section VIII.

VII. FURTHER EXPERIMENTAL ANALYSIS

A. Fusion Experiment

In this section, we attempt to assess the relative complementarity of the discrimination algorithms by measuring the performance of a simple fusion of their outputs. Let t_i be the decision statistic of the i th discriminator; we use a fusion comprised of a simple unweighted multiplication of *all* four of the discriminator decision statistics: $t_{\text{fusion}} = \prod_{i=1}^4 t_i$. Before the fusion, we apply the popular Platt scaling [58] to the decision statistics of each discriminator. The Platt scaling applies a logistic regression (two parameters) to the statistics of each discriminator, after which $t_i \in [0, 1]$, and the statistics approximate a class posterior probability. Unique scaling parameters were inferred for each algorithm using the decisions statistics generated by cross-validation. In order to minimize positive bias in the results, only the statistics from a single lane were used for parameter inference (i.e., training), and we note that the results were insensitive of the lane chosen for training.

After this scaling, t_{fusion} can be interpreted as an “AND” operation between the discriminators, in which an object is only labeled as a threat if all of the individual discriminators label it as a threat. The results of this fusion are presented in Fig. 10. The results indicate that this simple fusion yields

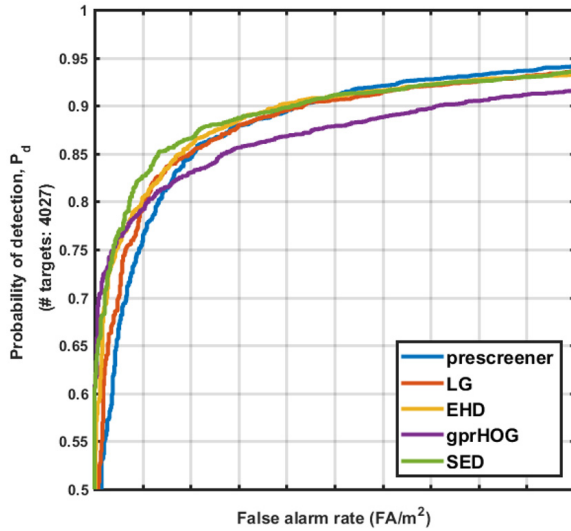


Fig. 11. ROC curves of all the individual discrimination algorithms when applied (but not trained) to threats buried at standard burial depths (i.e., not deep). Note that the y-axis has been truncated to the range $P_d = [0.5, 1]$. Also, the x-axis labels have been omitted—see Section III-C for details.

a substantial improvement in P_d across all values of FAR. The substantial benefit of fusion, without any subselection or weighting of the individual discriminators, suggests that the algorithms possess some complementary decision characteristics, despite utilizing seemingly similar processing approaches.

B. Algorithm Performance by Burial Depth

In this section, we evaluate the performance of the individual discrimination algorithms when measuring their performance on two disjoint populations of buried threats: threats that are buried at standard burial depths, and those with relatively deep burial depths. The definition of these categories is described in Table I. Note that the supervised portions of the algorithms were not retrained in each case; instead we computed ROC curves on the subsets of buried threats with the designated burial depths. These results are presented in Fig. 11 (standard burial depths) and Fig. 12 (deep burial depths).

As would be anticipated, the results indicate that the deeply buried threats are substantially more difficult to detect than those with a more shallow burial depth. This result is consistent with findings in [1] and [7], and is likely caused by the lower expected signal-to-noise ratio among deeply buried threats. It may also be exacerbated by the poor representation of deeply buried threats in the training data set.

Interestingly, however, the LG algorithm offers a substantial performance advantage over the other discrimination algorithms on deeply buried threats. This implies that in addition to lower signal-to-noise ratios, deeply buried threats may also exhibit different signal characteristics that are captured more effectively by the LG features. One unique characteristic of the LG feature is that it encodes shape content at multiple scales, which may make it well suited to the smaller and/or weaker signals typical of deeply buried threats. Further investigation is needed to confirm this hypothesis, however.

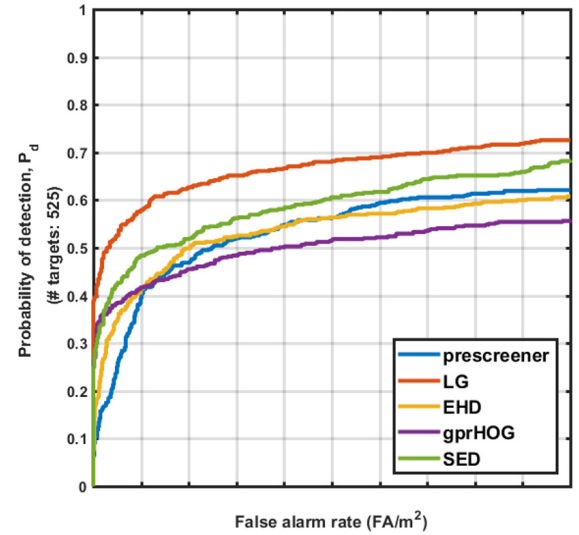


Fig. 12. ROC curves of all the individual discrimination algorithms when applied (but not trained) on threats buried at deep burial depths. Note that the y-axis has not been truncated to the range $P_d = [0.5, 1]$. Also, the x-axis labels have been omitted—see Section III-C for details.

Among threats buried at common depths, the SED algorithm performs best over most of the ROC curve, and gains a noticeable performance advantage over LG. EHD also gains performance relative to LG in this regard. This is consistent with the hypothesis that threats may require different processing depending upon their burial depth.

VIII. ANALYSIS OF DISCRIMINATOR DESIGNS

The authors and institutions that contributed to this paper—termed “the group”—are part of an ongoing collaboration that has existed for roughly two decades. As part of this effort, we conduct regular meetings to share experimental results, research insights, and coordinate research projects such as the algorithm comparison presented here. Through these longstanding efforts, our group has developed many insights. In this section, we present the perspectives of the group on various aspects of discriminator design, with a focus on those opinions that have received a consensus among the group members.

We begin by discussing algorithm design practices that have been found by our group to be generally effective, and which have been adopted voluntarily by all of the group members. Subsequently, we focus on the major conceptual differences between the discriminators presented in this paper. Last, we present the implications of the experimental results on the design concepts employed by the discriminators.

A. Group Trends in Algorithm Design

Several algorithm design strategies and practices have been found by our group to be generally effective. These are practices that are currently used by all members of the group for algorithm development, including the discriminators presented in this paper.

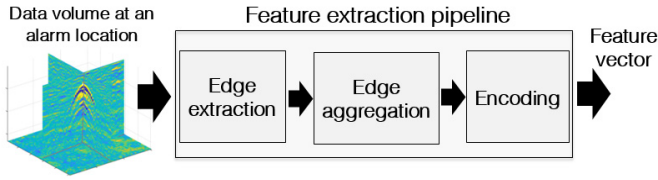


Fig. 13. Conceptual processing pipeline for feature extraction used by our group. All features presented in this paper represent special cases of this pipeline. The pipeline begins with edge extraction, in which filters are used to identify the locations, strengths, and orientations of edges in the GPR imagery. For example, LG uses LG filters at multiple scales and orientations; while gprHOG, EHD, and SED use simpler single-scale filters. The next step of processing aims to pool, or “aggregate,” the edge information within predefined spatial or temporal regions. Examples of aggregation regions can be found in Fig. 7 (gprHOG) and Fig. 8 (SED). The edge information within each aggregation region is summarized using an “encoding,” with the aim of compressing the information and reducing the dimensionality of the resulting feature. Encodings may involve counting (EHD, SED), summing (gprHOG), or taking the max-order statistic (LG) of the edges within each pooling region.

1) *Two-Stage Processing Pipeline*: We have found that using a standardized two-stage processing pipeline for discriminators, comprised of feature extraction and a supervised classifier, to be highly effective for discriminator development. All algorithms in this paper employ this pipeline, which is illustrated and described in Fig. 4. This pipeline leverages modern supervised classification models, which we have found to be highly effective for this problem. The modularity of this pipeline has also been a major advantage for collaboration, facilitating the development of a standard input–output formulation for discriminator components, and thereby allowing us to easily exchange modules in controlled experiments.

Approximate versions of the two-stage processing have been employed in numerous previous studies, including our group’s previous discriminator comparison [18], and is therefore largely consistent with the literature. One point of contrast, however, may be our increasing utilization of standard (non-linear) supervised classifiers, such as the nonlinear SVM and random forest. We have found these approaches to be competitive with hand-crafted or customized decision-making models, such as those in [18], but with the advantage of being much simpler to employ or tune.

2) *Image-Based Processing*: Although GPR data are collected as a 3-D volume, each group member has developed algorithms that process 2-D slices of the data rather than directly processing the 3-D volumes (e.g., with 3-D filters).

3) *Feature Processing Pipeline*: We have found that effective features focus on succinctly encoding the locations, strengths, and orientation of edges in the GPR imagery. Over the last decade, all members of the group have converged on the same conceptual feature extraction pipeline, which implements this general strategy, and which is described in Fig. 13.

B. Conceptual Differences Between Discriminators

Modern supervised classification models are well-developed and often require little modification to work effectively. As a result, most discriminator research within our group has focused on improving the feature extraction process

TABLE II
CONCEPTUAL DIFFERENCES BETWEEN THE FEATURES IN THIS WORK. WITH RESPECT TO THESE QUALITIES, NO TWO FEATURES IN THE GROUP ARE IDENTICAL

Quality	Feature vector			
	LG	gprHOG	SED	EHD
Feature dimensionality	Medium	High	Low	Low
Multi-scale shape encoding?	Yes	No	No	No
Hyperbolic threat model?	Yes	Yes	No	Yes
Encodes shape changes over the temporal axis of B-scans?	No	Yes	No	No

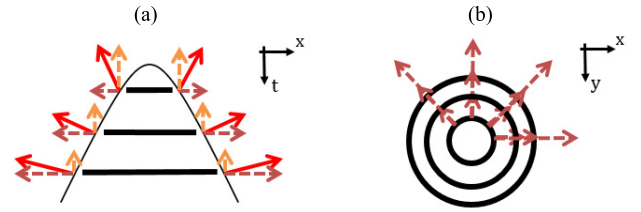


Fig. 14. (a) Hyperbolic model of buried threats assumes that threats appear as strong hyperbolic patterns in the (t, x) or (t, y) planes of GPR data. Feature extraction approaches based upon this model typically focus on encoding the gradients (red arrows) across the spatial or temporal dimensions of GPR B-scans. (b) Circle model of threats makes a weaker assumption about the shape of threats, assuming only that they exhibit circular gradients in the (y, x) plane of the GPR data. Note that this circular pattern can arise from many shapes in the (t, x) and (t, y) plane, including, but not limited to, hyperbolas.

(or “features”). Similarly, features serve as the major differentiator between different discriminators within the group.

As described in Section VIII-A, all features within the group employ the same basic processing pipeline (see Fig. 13). However, each feature employs somewhat different modules in the pipeline, reflecting different hypotheses, or concepts, each group has about the visual appearance of threats in GPR imagery. Next, we will briefly identify and discuss these conceptual differences. We focus primarily on the conceptual differences that: 1) make each individual feature unique and 2) that are most consequential to their detection behavior, particularly with respect to the experimental results presented in this paper. The feature differences we consider are summarized in Table II.

1) *Circular Versus Hyperbolic Threat Models*: Perhaps the most fundamental difference between the features is their underlying model for the visual manifestation of a threat in the GPR imagery. Historically most features have relied upon the “hyperbolic model” shown in Fig. 14(a) [2], [23], [59]–[61], including the previous large-scale study by this group [18]. The gprHOG, EHD, and LG features share this modeling assumption, motivating their shared strategy of extracting edges from GPR B-scans. The SED feature is unique because it assumes a circular threat model, as shown in Fig. 14(b). As a result, it focuses on extracting edge information from T-scans (GPR imagery constructed from slices of the xy data plane). A full discussion of this model and its implications is beyond the scope of this paper, however, a few observations are

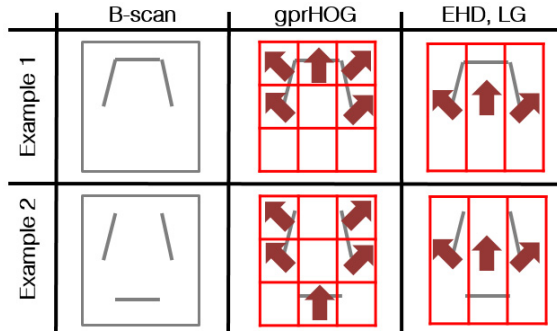


Fig. 15. Illustration of how gprHOG encodes changes in shape along the temporal axis of B-scans, compared to the other features. Two cartoon B-scans are shown in the first column. Black line segments indicate strong edges in the B-scan. The middle row shows the approximate aggregation regions of gprHOG (red lines), and the arrows visualize the gprHOG feature encoding for the shape. A similar illustration is presented for EHD, LG, and SED in the third column. Notice that EHD and LG would make a similar encoding for both examples, although one is a hyperbola and the other is not. gprHOG does not have this limitation, but at the cost of having many more aggregation regions, and consequently a higher dimensional feature encoding.

important to mention here. A B-scan hyperbola will manifest itself as a circular pattern in T-scans [as shown in Fig. 14(b)], but many other objects will as well, including very irregular nonhyperbolic patterns. Therefore, this model cannot resolve hyperbolas from many other patterns; however, as we discover, this assumption appears to be reasonable, and yields a relatively low-dimensional feature representation in SED.

2) *Encoding Temporal Shape Changes*: Although EHD, LG, and gprHOG all assume a hyperbolic model, they differ in their precise conception of the hyperbola, resulting in somewhat different processing strategies. The biggest conceptual difference relates to the configuration of their pooling regions. All three of these features utilize multiple pooling regions along the spatial dimension, allowing them to encode whether the edges of the GPR imagery gradually change orientation (across the spatial dimension) in a manner that is consistent with a hyperbola (e.g., diagonal edge, flat edge, and antidiagonal edge, consecutively). However, gprHOG uses multiple pooling regions along the temporal axis as well, whereas EHD and LG do not. Therefore, gprHOG features can encode whether edge variations along the temporal dimension are also consistent with a hyperbola, whereas the other features cannot. This difference is illustrated in Fig. 15. This capability, while potentially useful, comes at the cost of a much higher feature dimensionality.

3) *Multiscale Edge Encoding*: LG is the only feature that employs a multiscale hyperbola model. LG employs LG filters with different sizes, ensuring that a strong edge response will be encoded even if the hyperbolas vary in size (within some limits). In contrast, the fixed-size filters of the other features will not match well when hyperbolas are unusually large or small, resulting in weaker estimation of their edges, and potentially poorer detection of them.

C. Implications of the Experimental Results

In this section, we describe the implications of the experimental results from Sections VI and VII on the design of

discriminators. In particular, we will analyze the experimental results to distill which feature concepts tend to work well.

1) *Circular Threat Model Is a Viable Alternative to the Hyperbolic Model*: One notable outcome of the experiments was that the SED discriminator almost always achieved the best detection rate (P_d) for the standard-depth threats (roughly 90% of the total) as shown in Fig. 11. This suggests that the circular threat model, upon which the SED features is derived, may be a viable alternative to the hyperbolic model that has been popular in the GPR-based BTM community for the last two decades [2], [23], [59]–[61], including within our group's last discriminator comparison paper [18].

2) *Multiscale Edge Extraction Is Beneficial for Deep Threats*: Another notable outcome of the experiments was the substantial performance advantage of the LG discriminator on deeply buried threats, as shown in Fig. 12. We find that deeply buried threats often appear smaller than others, and we hypothesize that LG's unique multiscale filters are advantageous for encoding edges from these smaller deeply buried threats.

3) *Encoding Temporal Shape Changes Is Not Strongly Beneficial*: One potentially surprising finding of the experiments is the relatively poor performance of the gprHOG feature. Since its introduction to GPR-based BTM the HOG feature has become quite popular and these results cast some doubt on the effectiveness of unique design concepts of HOG and gprHOG: this concept suggests that encoding changes in shape along the temporal dimension of B-scans is beneficial. However, these results are consistent with a more comprehensive analysis that was recently published [62].

4) *Discriminators Are Complementary*: Although we found performance differences between the individual discriminators, the substantial performance improvements offered by a simple fusion (see Fig. 10) suggests the features are complementary—they encode somewhat different threat cues. Similar findings have been made in the past with GPR-based BTM features [63], including in our group's last comparison paper [18], but this fusion provides evidence that the specific conceptual differences between the discriminators here may be highly complementary. Future work is needed to disentangle which features are most important, or complementary.

IX. CONCLUSION

In this paper, we report the results of a recent multi-institutional effort to develop, and compare, BTM algorithms. The effort involved five institutions with an established track record in GPR-based BTM: Duke University, University of Louisville, University of Missouri, University of Florida, and Chemring Sensors and Electronic System (CSES). A major objective of this effort was to identify the best processing approaches, and evaluate them in an unbiased manner, for potential inclusion in a real-world BTM system. To achieve this goal, each institution was provided with the same period of time for algorithm development; the same experimental data set; and advanced knowledge of the experimental design, and performance metrics.

The institutions (excluding CSES) were specifically tasked with developing discriminators. Discriminators must accept a

small cube of GPR data (e.g., centered at a suspicious spatial location), and return a decision statistic, indicating the relative likelihood that a buried threat is located at that location. The discriminators produced in this effort were compared using a large GPR data set collected using a vehicle-mounted GPR BTM system comprised of 120 000 m² of surface area and 4552 buried threat encounters.

Four discriminators were submitted: LG, EHD, gprHOG, and SED and the results reveal similar performance among them, but with the SED and LG algorithm providing the best detection rates over most of the considered range of FAR. Further analysis revealed that the LG discriminator had a substantial advantage over the others on deeply buried threats, while SED almost always achieved the best performance on threats with common burial depths (i.e., all nondeeply buried threats). A simple fusion of the discriminators, involving no discriminator subselection or weighting, yielded substantial performance improvements.

A. Major Conclusions Regarding Discriminator Design

Our conclusions are divided into two parts. The first does not directly involve the experimental results and instead focuses on discriminator design practices that our group has found to be beneficial over the last decade. These are practices that were employed by all members of the group in the design of their discriminators and thereby form a consensus of opinion. These practices are described in more detail in Section VIII-A, but we summarize our findings as follows.

- 1) We employ two-stage discriminator processing pipelines, comprised of feature extraction and a conventional supervised classification model.
- 2) Processing GPR imagery (2-D slices) has been more effective than processing 3-D GPR data directly, although we suggest this for future work (see Section IX-B).
- 3) We have found that effective features focus on succinctly encoding the locations, strengths, and orientations of edges in the GPR imagery. All of our features employ the conceptual processing pipeline in Fig. 13, which relies upon this strategy.

Our second set of conclusions focuses on the experimental results presented in this paper, and their implications regarding discriminator design concepts. A detailed discussion of these implications is provided in Section VIII-C, but we summarize them as follows.

- 1) The circular threat model, upon which the SED discriminator is based, is a viable alternative to the popular hyperbolic model (see Fig. 14).
- 2) Multiscale feature representations (e.g., in LG) are likely to be beneficial for the detection of threats across multiple burial depths.
- 3) Encoding shape change along the temporal axis of B-scans (e.g., as gprHOG does) provides relatively little useful decision-relevant information, while substantially increasing feature dimensionality, and thereby it tends lower overall detection performance.
- 4) A simple fusion of the discriminators in this paper substantially increased detection performance,

suggesting that some of the feature concepts are highly complementary.

Algorithm 1 Algorithm to Compute Concavity Measures, c^+ and c^-

Input: A GPR slice, $S(t, z)$, where t is the temporal axis and z is the spatial dimension. Denote the spatial center of the patch by z_0 .

Parameters:

ω = Size of the search window for maxima (positive integer)

γ = The threshold to retain maxima values

Output: Two concavity measures, c^+ and c^-

1. Compute $S'(t, z) = |\max(S(t, z), 0)|\forall(t, z)$

2. Compute $t^* = \max_i S'(i, z_0)$

3. **If** $S'(t^*, z_0) \geq \gamma$ set $c^+ = 0$ and go to step (9), **else**, initialize the set of coordinates $X = \{(t^*, z_0)\}$

4. **For** $j = 1 \dots 5$

(a) Set $i^* = \arg \min_{i \in [-\omega, \omega]} |t^* - i| s.t. S'(t^* + i, z_0 + j) \geq \gamma$

(b) **If** no $i^* = \emptyset$, go to step (5),

else add $(t^* + i^*, z_0 + j)$ to X and set $t^* = t^* + i^*$

End

5. Repeat step (4), with $j = -1 \dots -5$. The set C now contains a sequence of (spatially) neighboring points, with a maximum potential length of 11 points. See Fig. 6 for examples on real data.

6. Construct X' as the set of all possible sequences of consecutive points that can be constructed from the points in X .

7. $c^+ = 1/|X'| \sum_{x \in C'} f(x)$, where $f(\cdot)$ is a function that measures concavity as the difference between a sequence's midpoint and the average of its end-points.

8. Set $S = -S$ and repeat steps (1)-(7) to obtain c^-

B. Future Challenges With Discriminator Design

There are many open questions regarding the design of GPR-based BTM algorithms, and opportunities to improve performance. We describe a few that we believe are most promising for future work.

1) *Feature Learning and Deep Learning:* All of the algorithms presented in this paper, and many in the literature, employ a conventional image-based object recognition pipeline comprised of hand-crafted features and a conventional supervised classification model. Deep learning algorithms, such as convolutional neural networks, have the capability to learn features directly from the data, and they yielded groundbreaking results in other application domains. Some papers have investigated applying these methods to GPR-based BTM [24], [64], [65]; however, our group has not (yet) found these approaches to outperform conventional methods. Nonetheless, we see great potential in deep learning techniques, and consider their investigation an important line of inquiry.

2) *3-D Processing:* All of the approaches here, and most in the literature, perform detection by analyzing GPR imagery (2-D slices of 3-D GPR data). There have been several

attempts within our group to process 3-D GPR data directly (e.g., using 3-D filtering). None of these approaches have yet achieved competitive results, but 2-D processing ignores obvious patterns that exist in 3-D GPR data; however, designing features for 3-D data is challenging. Therefore, we see this as another important future line of research.

APPENDIX

PRESCREENER ALGORITHM DETAILS

This appendix provides technical descriptions of the two prescreener algorithms employed in this paper, as well as their fusion. Note that, prior to prescreening, preprocessing is applied in which the time of the ground response is estimated at each spatial location, and then each A-scan is shifted so that the ground response occurs at the same time index across all spatial locations. Subsequently all data at, or above, the ground response is removed. These two preprocessing steps are common in GPR BTD [21], [29], [66], [67]. Finally, note also that all of the prescreener parameters, including the preprocessing, were established prior to the discrimination design process. These parameters were assumed to be fixed by the discriminator designers.

A. F2 Prescreener

The F2 prescreener is comprised of a series of smoothing and constant-false-alarm-rate (CFAR) filtering steps [21], [68]. CFAR filtering estimates how anomalous an input datum is based upon the surrounding data. In this context, CFAR filtering assumes that the (nonanomalous) observed data are generated by a nonstationary Gaussian process [68]. Based upon this assumption, there are many variants of CFAR filtering, however, the basic process involves computing a rolling (or moving) estimate of the mean and variance of the data, denoted μ_{bg} and σ_{bg}^2 , respectively. The output of the filter at each input datum p is computed by $(p - \mu_{bg})^2 / \sigma^2$.

With this basic definition of CFAR filtering, the F2 prescreener involves the following major steps.

- 1) At each time index, median filtering the GPR volume in the down-track direction, to mitigate noise.
- 2) At each time index, subtraction of the mean in the cross-track dimension, in order to remove panel-specific signal variations.
- 3) Depth binning, whereby each A-scan is divided into nonoverlapping bins (i.e., subsets of time indices). The set of A-scan values within each bin is replaced by the average of the top two values within that bin.
- 4) CFAR filtering is applied along the time axis.
- 5) The processed GPR volume is summed along the time axis, resulting in a single value at each spatial location (i.e., a 2-D spatial map of intensities).
- 6) The 2-D CFAR and Gaussian smoothing operations are applied, in that order, to the spatial map.
- 7) Nonmaximum suppression to reduce false alarms.

Alarms are obtained from the processed spatial map by applying a threshold to the resulting intensity values, and identifying connected components of pixels with intensities above the threshold. This threshold was set in order to balance

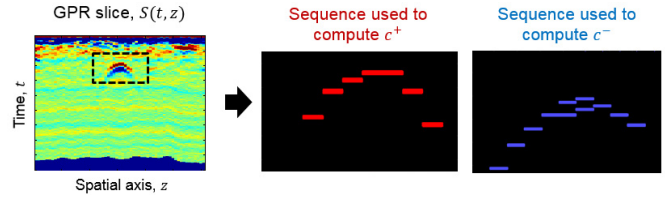


Fig. 16. (Left) Illustration of a GPR slice. (Middle) Sequences of local maxima values (in red) that were identified for computing c^+ . This sequence was identified within the black dashed box in the GPR slice. The full procedure for identifying such sequences, and computing c^+ , are described in Table III. The sequences of local minima values (in blue) that were identified for computing c^- .

the detection rate and the false alarm rate of the prescreener. This balance, and therefore the threshold, was set by the sponsoring agency based upon their operational needs.

B. CCY Prescreener

The CCY prescreener implementation begins by statistically whitening each point in the GPR volume based upon statistics computed on neighboring points at the same depth (time index). This step mitigates signal attenuation with respect to depth. Once this preprocessing is completed, a CCY measurement is computed at every spatial location in the GPR volume. This CCY measurement is itself computed from two CCY measurements: one computed on a down-track slice of GPR data, and one computed on a cross-track slice of GPR data. We have found that the presence of a CCY in both tracks of data increases the probability that a threat is present.

Given either slice of GPR data, the CCY calculation proceeds in the same manner, and produces two CCY measures, c^+ and c^- that are summed to obtain one CCY measure for the slice. A detailed description of the algorithm used to compute c^+ and c^- is presented in Table III, but we outline the algorithm here. The algorithm attempts to identify sequences of high magnitude pixels that form a concave curve. This search proceeds by considering the positive signals and the negative signals separately. For example, we consider only the positive signal by setting all of the negative pixel values to zero, and then identifying local maxima in the resulting image that exceed some value threshold. This processing is illustrated on real GPR data in Fig. 16 (middle image). c^+ is a measure of the CCY of the sequence of local maxima identified in the image. A similar process is applied to identify the negative signals to obtain c^- . The negative image maxima are also illustrated in Fig. 16.

The final estimate of CCY at a particular spatial location consists of summing the two CCY measures from the down-track slice (i.e., c^+ and c^-) with those from the cross-track slice (i.e., four total). Once a CCY measure is computed for every spatial location, a smoothing filter is applied to the resulting 2-D map of CCY values. The CCY prescreener employs nonmaximum suppression to the resulting 2-D map in order to obtain alarms. This threshold was set in order to balance the detection rate and the FAR of the prescreener. This balance, and therefore the threshold, was set by the sponsoring agency based upon their operational needs.

C. Fusion of the Prescreeners

Due to the complementary nature of the F2 (energy-focused) and CCY (shape-focused) prescreeners, it has been advantageous to prescreening performance to fuse their respective outputs. This can be accomplished by treating the alarms made by F2 and CCY as though they were generated from a single prescreener. However, it is likely that many pairs of alarms (i.e., one from F2 and one from CCY) will correspond to the same object (e.g., buried threat, subsurface rocks, and roots). This may result in wasteful processing of the same object via several different alarms, and create unnecessary vehicle stops or slowdowns if the object is perceived as a threat.

To address this problem, we merge alarms that are likely to be redundant. A simple and effective proxy for alarm redundancy is the relative proximity of the two alarms, and so alarms are merged using a proximity threshold. The exact details of the merging process are omitted here due to space considerations. In the event of a merger, however, the two alarms are usually replaced with a single alarm, and the decision statistic assigned to the new alarm is given by a weighted sum (i.e., an average) of the statistics from the individual alarms.

In general, the F2 and CCY decision statistics can differ by orders of magnitude, potentially interfering with the effectiveness of the weighted average, because the magnitude of one prescreener statistic dominates the other. To mitigate this problem, the CCY statistics are rescaled as follows: $a \cdot \text{CCY}^b + c$. In our experiments, the three parameters, a , b , and c , were determined using training data to maximize the area under the ROC curve on the training data.

ACKNOWLEDGMENT

The authors would like to thank M. Scholl and P. Howard for their support of this work. They would also like to thank M. DeLong and Duke Research Computing at Duke University, Durham, NC, USA, for their computational support for this work.

REFERENCES

- [1] D. Nabelek and K. C. Ho, "Detection of deeply buried non-metal objects by ground penetrating radar using non-negative matrix factorization," *Proc. SPIE*, vol. 9454, May 2015, Art. no. 945419.
- [2] K. L. Lee and M. M. Mokji, "Automatic target detection in GPR images using Histogram of Oriented Gradients (HOG)," in *Proc. 2nd Int. Conf. Electron. Design (ICED)*, Aug. 2014, pp. 181–186.
- [3] A. Hamdi, O. Missaoui, and H. Frigui, "An SVM classifier with HMM-based kernel for landmine detection using ground penetrating radar," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 4196–4199.
- [4] K. C. Ho, L. Carin, P. D. Gader, and J. N. Wilson, "An investigation of using the spectral characteristics from ground penetrating radar for landmine/clutter discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1177–1191, Apr. 2008.
- [5] F. Abujarad, G. Nadim, and A. Omar, "Wavelet packets for GPR detection of non-metallic anti-personnel land mines based on higher-order-statistic," in *Proc. 3rd Int. Workshop Adv. Ground Penetrating Radar (IWAGPR)*, May 2005, pp. 21–25.
- [6] H. Frigui, O. Missaoui, and P. Gader, "Landmine detection using discrete hidden Markov models with Gabor features," *Proc. SPIE*, vol. 6553, May 2007, Art. no. 65532A.
- [7] D. Reichman, J. M. Malof, and L. M. Collins, "Algorithm development for deeply buried threat detection in GPR data," *Proc. SPIE*, vol. 9823, May 2016, Art. no. 98231A.
- [8] D. P. Nabelek and K. C. Ho, "Detection of shallow buried objects using an autoregressive model on the ground penetrating radar signal," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 870911.
- [9] A. Karem, A. B. Khalifa, and H. Frigui, "A Fisher vector representation of GPR data for detecting buried objects," *Proc. SPIE*, vol. 9823, May 2016, Art. no. 98231C.
- [10] P. Kłeszk, M. Kapruziak, and B. Olech, "Boosted classifiers for antitank mine detection in C-scans from ground-penetrating radar," in *Proc. Adv. Intell. Syst. Comput.*, vol. 342, 2015, pp. 191–205.
- [11] Y. Zheng and A. Elmaghraby, "A vehicle threat detection system using correlation analysis and synthesized X-ray images," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 87090V.
- [12] V. Kovalenko, A. G. Yarovoy, and L. P. Lighthart, "A novel clutter suppression algorithm for landmine detection with GPR," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3740–3750, Oct. 2007.
- [13] K. Tbarki, S. Ben Said, and N. Ellouze, "Non-linear filtering for landmine detection on ground penetration radar data," in *Proc. 2nd Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Mar. 2016, pp. 200–205.
- [14] K. Tbarki, S. B. Said, R. Ksantini, and Z. Lachiri, "Covariance-guided landmine detection and discrimination using ground-penetrating radar data," *Int. J. Remote Sens.*, vol. 39, no. 2, pp. 289–314, 2018.
- [15] S. Lameri, F. Lombardi, P. Bestagini, M. Lualdi, and S. Tubaro, "Landmine detection from GPR data using convolutional neural networks," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug./Sep. 2017, pp. 508–512.
- [16] J. M. Malof, D. Reichman, and L. M. Collins, "Discriminative dictionary learning to learn effective features for detecting buried threats in ground penetrating radar data," *Proc. SPIE*, vol. 10182, May 2017, Art. no. 101820W.
- [17] I. Giannakis, S. Xu, P. Aubry, A. Yarovoy, and J. Sala, "Signal processing for landmine detection using ground penetrating radar," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 7442–7445.
- [18] J. N. Wilson, P. Gader, W.-H. Lee, H. Frigui, and K. C. Ho, "A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection and discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2560–2572, Aug. 2007.
- [19] E. Tebchrany, F. Sagnard, V. Baltazart, J.-P. Tarel, and X. Dérobert, "Assessment of statistical-based clutter reduction techniques on ground-coupled GPR data for the detection of buried objects in soils," in *Proc. 15th Int. Conf. Ground Penetrating Radar*, Jun./Jul. 2014, pp. 604–609.
- [20] Y. Liao, L. W. Nolte, and L. M. Collins, "Decision fusion of ground-penetrating radar and metal detector algorithms—A robust approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 398–409, Feb. 2007.
- [21] P. Gader, W.-H. Lee, and J. N. Wilson, "Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2522–2534, Nov. 2004.
- [22] P. Kłeszk, A. Godziuk, M. Kapruziak, and B. Olech, "Fast analysis of c-scans from ground penetrating radar via 3-D Haar-like features with application to landmine detection," *Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3996–4009, Jul. 2015.
- [23] P. A. Torrione, K. D. Morton, R. Sakaguchi, and L. M. Collins, "Histograms of oriented gradients for landmine detection in ground-penetrating radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1539–1550, Mar. 2014.
- [24] D. Reichman, L. M. Collins, and J. M. Malof, "Some good practices for applying convolutional neural networks to buried threat detection in Ground Penetrating Radar," in *Proc. 9th Int. Workshop Adv. Ground Penetrating Radar (IWAGPR)*, Jun. 2017, pp. 1–5.
- [25] S. E. Yuksel, J. Bolton, and P. D. Gader, "Landmine detection with multiple instance hidden Markov models," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Sep. 2012, pp. 1–6.
- [26] A. Manandhar, K. D. Morton, L. M. Collins, and P. A. Torrione, "Multiple instance hidden Markov models for GPR-based landmine detection," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 87091L.
- [27] E. Pasolli, F. Melgani, and M. Donelli, "Automatic analysis of GPR images: A pattern-recognition approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2206–2217, Jul. 2009.
- [28] D. Reichman, L. M. Collins, and J. M. Malof, "Learning improved pooling regions for the Histogram of Oriented Gradient (HOG) feature for buried threat detection in ground penetrating radar," *Proc. SPIE*, vol. 10182, May 2017, Art. no. 101820V.
- [29] R. N. Nagashree, N. Aswini, A. Dyana, and C. H. S. Rao, "Detection and classification of ground penetrating radar image using textural features," in *Proc. Int. Conf. Adv. Electron. Comput. Commun.*, Oct. 2014, pp. 1–5.

- [30] R. Mendez-Rial, U. Uschkerat, F. I. Rial, and M. A. González-Huici, "Evaluation of landmine detection performance applying two different algorithms to GPR field data," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 87091J.
- [31] J. Wood, J. Wilson, and J. Bolton, "Extracting edge histogram detector features from ground penetrating radar data without ground alignment," *Proc. SPIE*, vol. 8357, May 2012, Art. no. 83571W.
- [32] H. Frigui and P. D. Gader, "Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic K -nearest neighbor classifier," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 1, pp. 185–199, Feb. 2009.
- [33] D. Reichman, L. M. Collins, and J. M. Malof, "Improvements to the Histogram of Oriented Gradient (HOG) prescreeener for buried threat detection in ground penetrating radar data," *Proc. SPIE*, vol. 10182, May 2017, Art. no. 101820U.
- [34] A. Hamdi and H. Frigui, "Ensemble hidden Markov models with application to landmine detection," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 75, 2015.
- [35] A. Kareem, A. Fadeev, H. Frigui, and P. Gader, "Comparison of different classification algorithms for landmine detection using GPR," *Proc. SPIE*, vol. 7664, Apr. 2010, Art. no. 76642K.
- [36] J. Bolton, S. E. Yuksel, and P. Gader, "Multiple instance learning for hidden Markov models: Application to landmine detection," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 87091M.
- [37] H. Frigui and P. Gader, "Detection and discrimination of land mines based on edge histogram descriptors and fuzzy K -nearest neighbors," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2006, pp. 1494–1499.
- [38] D. Reichman, L. M. Collins, and J. M. Malof, "On choosing training and testing data for supervised algorithms in ground-penetrating radar data for buried threat detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 497–507, Jan. 2018.
- [39] C. R. Ratto, K. D. Morton, L. M. Collins, and P. A. Torrione, "A hidden Markov context model for GPR-based landmine detection incorporating stick-breaking priors," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2011, pp. 874–877.
- [40] H. Frigui, L. Zhang, P. Gader, J. N. Wilson, K. C. Ho, and A. Mendez-Vazquez, "An evaluation of several fusion algorithms for anti-tank landmine detection and discrimination," *Inf. Fusion*, vol. 13, no. 2, pp. 161–174, Apr. 2012.
- [41] J. Bolton, P. Gader, H. Frigui, and P. Torrione, "Random set framework for multiple instance learning," *Inf. Sci.*, vol. 181, no. 11, pp. 2061–2070, 2011.
- [42] P. A. Torrione, C. S. Throckmorton, and L. M. Collins, "Performance of an adaptive feature-based processor for a wideband ground penetrating radar system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 2, pp. 644–658, Apr. 2006.
- [43] J. S. Varsanik, J. W. Roberts, T. W. Chevalier, and A. D. Mulliken, "GPR preprocessing optimization with signal-to-clutter metrics," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 87091O.
- [44] M. A. González-Huici, I. Catapano, and F. Soldovieri, "A comparative study of GPR reconstruction approaches for landmine detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4869–4878, Dec. 2014.
- [45] F. Giovanneschi and M. A. González-Huici, "A preliminary analysis of a sparse reconstruction based classification method applied to GPR data," in *Proc. 8th Int. Workshop Adv. Ground Penetrating Radar (IWAGPR)*, Jul. 2015, pp. 1–4.
- [46] S. E. Yuksel, J. Bolton, and P. Gader, "Multiple-instance hidden Markov models with applications to landmine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6766–6775, Dec. 2015.
- [47] C. S. Won, D. K. Park, and S.-J. Park, "Efficient use of MPEG-7 edge histogram descriptor," *ETRI J.*, vol. 24, no. 1, pp. 23–30, 2002.
- [48] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [49] C. Chang and C. Lin, "LIBSVM," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
- [50] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. New York, NY, USA: Academic, 2008.
- [51] K. C. Ho, P. D. Gader, and J. N. Wilson, "Improving landmine detection using frequency domain features from ground penetrating radar," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2004, pp. 1617–1620.
- [52] S. Harris, K. C. Ho, and A. Zare, "On the use of log-Gabor features for subsurface object detection using ground penetrating radar," *Proc. SPIE*, vol. 9823, May 2016, Art. no. 98230K.
- [53] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
- [54] Q. Chen, Z. Song, J. Dong, Z. Huang, Y. Hua, and S. Yan, "Contextualizing object detection and classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 13–27, Jan. 2015.
- [55] E. Temlioglu, I. Erer, and D. Kumlu, "Histograms of dominant orientations for anti-personnel landmine detection using ground penetrating radar," in *Proc. 4th Int. Conf. Elect. Electron. Eng. (ICEEE)*, Apr. 2017, pp. 329–332.
- [56] D. Reichman, L. M. Collins, and J. M. Malof, (2018). "gprHOG and the popularity of Histogram of Oriented Gradients (HOG) for buried threat detection in ground-penetrating radar." [Online]. Available: <https://arxiv.org/abs/1806.01349>
- [57] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [58] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
- [59] H. Chen and A. G. Cohn, "Probabilistic robust hyperbola mixture model for interpreting ground penetrating radar data," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–8.
- [60] H. Harkat, Y. Elfakir, S. D. Bennani, G. Khaissidi, and M. Mrabti, "Ground penetrating radar hyperbola detection using scale-invariant feature transform," in *Proc. Int. Conf. Elect. Inf. Technol. (ICEIT)*, May 2016, pp. 392–397.
- [61] T. Noreen and U. S. Khan, "Using pattern recognition with HOG to automatically detect reflection hyperbolas in ground penetrating radar data," in *Proc. Int. Conf. Elect. Comput. Technol. Appl. (ICECTA)*, Nov. 2017, pp. 1–6.
- [62] D. Reichman, L. M. Collins, and J. M. Malof, "How much shape information is enough, or too much? Designing imaging descriptors for threat detection in ground penetrating radar data," *Proc. SPIE*, vol. 10628, Apr. 2018, Art. no. 106280E.
- [63] H. Frigui, L. Zhang, P. Gader, and D. Ho, "Context-dependent fusion for landmine detection with ground-penetrating radar," *Proc. SPIE*, vol. 6553, Apr. 2007, Art. no. 655321.
- [64] L. E. Besaw and P. J. Stimac, "Deep convolutional neural networks for classifying GPR B-scans," *Proc. SPIE*, vol. 9454, May 2015, Art. no. 945413.
- [65] J. M. Malof, J. Bralich, D. Reichman, and L. M. Collins, "Improving the histogram of oriented gradient feature for threat detection in ground penetrating radar by implementing it as a trainable convolutional neural network," *Proc. SPIE*, vol. 10628, Apr. 2018, Art. no. 106280D.
- [66] M. P. Masarik, J. Burns, B. T. Thelen, J. Kelly, and T. C. Havens, "GPR anomaly detection with robust principal component analysis," *Proc. SPIE*, vol. 9454, May 2015, Art. no. 945414.
- [67] J. Roberts, Y. Shkolnikov, J. Varsanik, T. Chevalier, and A. D. Mulliken, "Robust entropy-guided image segmentation for ground detection in GPR," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 87091N.
- [68] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.



Jordan M. Malof (S'15–M'18) received the B.S. degree from the University of Louisville, Louisville, KY, USA, in 2008, and the Ph.D. degree from Duke University, Durham, NC, USA, in 2015.

He is currently an Assistant Research Professor with Duke University, where he focuses on statistical modeling, machine learning, and computer vision for remote sensing applications.



Daniel Reichman received the B.S.E.E. and M.S.E.E. from The Cooper Union for the Advancement of Science and Art, New York, NY, USA, in 2013, and the Ph.D. degree in electrical and computer engineering from Duke University, Durham, NC, USA, in 2017.

He is currently a Research Scientist with the Applied Machine Learning Laboratory, Duke University. His research interests include machine learning, statistical signal processing, and computer vision applied to remote sensing applications.



Andrew Karem is currently pursuing the Ph.D. degree in computer engineering and computer science with the University of Louisville, Louisville, KY, USA.

He has been a Research Assistant with the Multimedia Research Lab, University of Louisville for the past few years. His research interests include image processing and machine learning with emphasis on multiple instance learning, and the detection of buried explosive objects using ground penetrating radar and wideband electromagnetic induction sensors.



Hichem Frigui (S'92–M'98) has been active in the research fields of machine learning; data mining; audio, image, and video processing; and multisensor information fusion. He has developed and applied algorithms from the above fields to a wide range of applications including: detection of buried explosive objects using ground penetrating radar and wideband electromagnetic induction sensors; segmentation and analysis of images and video; analysis of CT images for the diagnosis of lung cancer; and computational tools for predicting material properties. He is currently a Professor of computer engineering and computer science and the Director of the Multimedia Research Lab with the University of Louisville, Louisville, KY, USA.



K. C. Ho (S'89–M'91–SM'00–F'09) was born in Hong Kong. He received the B.Sc. degree (Hons.) in electronics and the Ph.D. degree in electronic engineering from the Chinese University of Hong Kong, Hong Kong, in 1988 and 1991, respectively.

From 1991 to 1994, he was a Research Associate with the Royal Military College of Canada, Kingston, ON, Canada. He joined the Bell-Northern Research, Montreal, QC, Canada, in 1995, as a Scientific Staff Member. From 1996 to 1997, he was a Faculty Member with the Department of Electrical Engineering, University of Saskatchewan, Saskatoon, SK, Canada. Since 1997, he has been with the University of Missouri, Columbia, MO, USA, and is currently a Professor with the Electrical Engineering and Computer Science Department. He is an Inventor of nine patents in USA and 13 patents in Canada, Europe, and Asia on geolocation and signal processing for wireless communications. His research interests include sensor array processing, source localization, subsurface object detection, wireless communications, and adaptive processing.

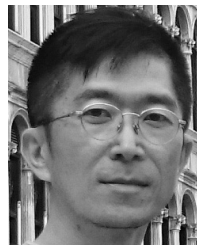
Dr. Ho received the Senior Faculty Research Award multiple times, the Junior Faculty Research Award, and the Teaching Award from the College of Engineering, University of Missouri. He was active in the development of the ITU-T Standard Recommendations from 1995 to 2012. He was the Rapporteur of ITU-T Q15/SG16: Voice Gateway Signal Processing Functions and Circuit Multiplication Equipment/Systems from 2009 to 2012, and the Associate Rapporteur of ITU-T Q16/SG16: Speech Enhancement Functions in Signal Processing Network Equipment in 2012. He was the Editor of the ITU-T Recommendations G.160: Voice Enhancement Devices from 2006 to 2012, G.168: Digital Network Echo Cancellers from 2000 to 2012 and G.799.2: Mechanism for Dynamic Coordination of Signal Processing Network Equipment from 2004 to 2009. He was a Technical Chair of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2016. He served as the Past Chair in 2015, the Chair from 2013 to 2014, and the Vice Chair from 2011 to 2012 of the Sensor Array and Multichannel Technical Committee of the IEEE Signal Processing Society. He was on the organizing committees of the IEEE SAM 2008 Workshop and the IEEE CAMSAP 2011. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2009 to 2013 and 2003 to 2006 and the IEEE SIGNAL PROCESSING LETTERS from 2004 to 2008.

Dr. Ho received the Senior Faculty Research Award multiple times, the Junior Faculty Research Award, and the Teaching Award from the College of Engineering, University of Missouri. He was active in the development of the ITU-T Standard Recommendations from 1995 to 2012. He was the Rapporteur of ITU-T Q15/SG16: Voice Gateway Signal Processing Functions and Circuit Multiplication Equipment/Systems from 2009 to 2012, and the Associate Rapporteur of ITU-T Q16/SG16: Speech Enhancement Functions in Signal Processing Network Equipment in 2012. He was the Editor of the ITU-T Recommendations G.160: Voice Enhancement Devices from 2006 to 2012, G.168: Digital Network Echo Cancellers from 2000 to 2012 and G.799.2: Mechanism for Dynamic Coordination of Signal Processing Network Equipment from 2004 to 2009. He was a Technical Chair of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2016. He served as the Past Chair in 2015, the Chair from 2013 to 2014, and the Vice Chair from 2011 to 2012 of the Sensor Array and Multichannel Technical Committee of the IEEE Signal Processing Society. He was on the organizing committees of the IEEE SAM 2008 Workshop and the IEEE CAMSAP 2011. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2009 to 2013 and 2003 to 2006 and the IEEE SIGNAL PROCESSING LETTERS from 2004 to 2008.



Joseph N. Wilson (M'05) received the B.S. degree in applied mathematics with an emphasis on computer science from Florida State University, Tallahassee, FL, USA, in 1977, and the M.S. degree in applied mathematics and computer science and the Ph.D. degree in computer science from the University of Virginia, Charlottesville, VA, USA, in 1980 and 1985, respectively.

Since 1984, he has been with University of Florida, Gainesville, FL, USA, where he is currently an Associate Professor with the Computer and Information Science and Engineering Department. He was an Associate Chair of the department from 1994 to 2001. His research interests include machine learning, image and signal processing, and information security.



Wen-Hsiung Lee received the B.S. degree in electronics engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 1990, and the master's and Ph.D. degrees from the University of Florida, Gainesville, FL, USA, in 1994 and 2001, respectively.

Since 2006, he has been a Senior Scientist with CSES (formerly NIITEK Inc.), Sterling, VA, USA.



William J. Cummings received the B.A. degree in physics from The University of Virginia, Charlottesville, VA, USA, in 1990.

He was involved in engineering/engineering leadership/management roles for more than 30 years developing and testing sensor systems for flight control/safety, navigation, drug discovery, and buried explosive hazards. He focused on industry committees to develop product development standards. Since 2005, he has been a Lead Engineer and a Program Manager for GPR development as well as a management positions with CSES (formerly NIITEK Inc.), Sterling, VA, USA.

number of engineering management positions with CSES (formerly NIITEK Inc.), Sterling, VA, USA.



Leslie M. Collins (M'96–SM'01) was born in Raleigh, NC, USA. She received the B.S.E.E. degree from the University of Kentucky, Lexington, KY, USA, in 1985, and the M.S.E.E. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1986 and 1995, respectively.

She was a Senior Engineer with the Westinghouse Research and Development Center, Pittsburgh, PA, USA, from 1986 to 1990. She joined Duke University, Durham, NC, USA, in 1995, as an Assistant Professor and was promoted to Associate Professor in 2002. Her research interests include incorporating physics-based models into statistical signal processing algorithms, and she is pursuing applications in subsurface sensing as well as enhancing speech understanding by hearing impaired individuals.

Dr. Collins is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.