Context-Dependent Multisensor Fusion and Its Application to Land Mine Detection

Hichem Frigui, *Member, IEEE*, Lijun Zhang, *Student Member, IEEE*, and Paul D. Gader, *Senior Member, IEEE*

***Abstract*—We present a novel method for fusing the results of multiple land mine detection algorithms which use different sensors, features, and different classification methods. The proposed multisensor/multialgorithm fusion method, which is called context-dependent fusion (CDF), is motivated by the fact that the relative performance of different sensors and algorithms can vary significantly depending on the mine type, geographical site, soil and weather conditions, and burial depth. CDF is a local approach that adapts the fusion method to different regions of the feature space. The training part of CDF has two components: context extraction and algorithm fusion. In context extraction, the features used by the different algorithms are combined and used to partition the feature space into groups of similar signatures, or contexts. The algorithm fusion component assigns a degree of worthiness to each detector in each context based on its relative performance within the context. To test a new alarm using CDF, each detection algorithm extracts its set of features and assigns a confidence value. Then, the features are used to identify the best context, and the degrees of worthiness of this context are used to fuse the individual confidence values. Results on large and diverse ground-penetrating radar and wideband electromagnetic data collections show that the proposed method can identify meaningful and coherent clusters and that different expert algorithms can be identified for the different contexts. Typically, the contexts correspond to groups of alarm signatures that share a subset of common features. Our extensive experiments have also indicated that CDF outperforms all individual detectors and the global fusion that uses the same method to assign aggregation weights.**

***Index Terms*—Electromagnetic induction (EMI), ground-penetrating radar (GPR), land mine detection, multialgorithm fusion, multisensor fusion.**

Nomenclature

AP	Antipersonnel mines.
APLM	Antipersonnel mines with low metal content.
APM	Antipersonnel metal mines (with high metal content).
AT	Antitank mines.
ATLM	Antitank mines with low metal content.
ATM	Antitank metal mines (with high metal content).
CDF	Context-dependent fusion.
cMPP	Conditional mapped posterior probability.
EDS	Energy density spectrum.
EHD	Discrimination algorithm based on edge histogram descriptor for GPR sensor data.
EMI	Electromagnetic induction.
FA	False alarm.
GPR	Ground-penetrating radar.
HM	Mines with high metal content.
HMC	Clutter with high metal content.
HMM	Discrimination algorithm based on hidden Markov models for GPR sensor data.
LM	Mines with low metal content.
MFIT	Discrimination algorithm based on model fitting for WEMI sensor data.
MPP	Mapped posterior probability.
NMC	Clutter with no metal content.
PD	Probability of detection.
PFA	Probability of false alarm.
ROC	Receiver operating characteristics.
SCAD	Simultaneous clustering and attribute discrimination.
SPECT	Discrimination algorithm based on spectral correlation feature for GPR sensor data.
TUF	Testing/training unified framework.
UTM	Universal Transverse Mercator.
WEMI	Wideband electromagnetic induction.

Manuscript received May 18, 2009; revised October 30, 2009. Date of publication February 25, 2010; date of current version May 19, 2010. This work was supported in part by U.S. Army Research Office Grants W911NF-08-0255 and W911NF-07-1-0347 and in part by National Science Foundation Awards CBET-0730802 and CBET-0730484. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, the Office of Naval Research, the Army Research Laboratory, or the U.S. Government.

H. Frigui is with the Computer Engineering and Computer Science Department, University of Louisville, Louisville, KY 40292 USA (e-mail: h.frigui@louisville.edu).

L. Zhang is with the Center for Biomedical Imaging Statistics, Emory University, Atlanta, GA USA (e-mail: lijun.zhang.cn@gmail.com).

P. D. Gader is with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: pgader@cise.ufl.edu).

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

Digital Object Identifier 10.1109/TGRS.2009.2039936

I. Introduction

DETECTION and removal of land mines is a serious problem affecting civilians and soldiers worldwide. It is estimated that the world is littered with over 100 million land mines in over 80 countries and that 26 000 people, mostly innocent civilians, are killed or maimed by a land mine each year [1], [2]. Detection and removal of land mines is therefore a significant problem and has attracted several researchers in recent years [3]–[7]. The research problem for data analysis is to determine how reliably land mines can be detected and distinguished from other subterranean objects using sensor data. Difficulties arise from the large variety of land mine types, differing soil conditions, temperature and weather conditions, and varying terrain, to name a few.

A variety of sensors have been proposed or are under investigation for land mine detection [8]–[11]. The key challenge to

mine detection technology lies in achieving a high rate of mine detection while maintaining a low level of false alarms. The performance of a mine detection system is therefore commonly measured by a receiver operating characteristic (ROC) curve that jointly specifies the rate of mine detection and the level of false alarm.

Electromagnetic induction (EMI) and ground-penetrating radar (GPR) are two useful sensors that complement each other. EMI works by establishing a time-varying magnetic field (primary field) over a conducting target and measuring the induced magnetic field (secondary field). Several algorithms that use wideband EMI (WEMI) have been proposed for land mine detection [12]–[15]. This technology is typically used to indicate the presence or the absence of metal targets below the ground. Unfortunately, many land mines are made of plastic and contain little or no metal. GPR, on the other hand, offers the promise of detecting land mines with little or no metal content. In fact, several approaches for detecting land mines and discriminating mines from clutter using GPR have been investigated [16]–[24]. Unfortunately, acceptable results have been elusive [6], [8], [25]. Although systems often achieve high detection rates, it is difficult to achieve the required low false alarm rates (FARs). Moreover, algorithm performance can vary significantly. Therefore, fusion methods that take advantage of the strengths of different sensors and algorithms, overcome their weaknesses, adapt to the rapidly changing environmental conditions, and achieve a higher accuracy than any individual algorithm are needed.

Multisensor and multialgorithm fusion are critical components in land mine detection. A distinguishing characteristic of buried land mine detection is that sensors that detect buried objects must interact with the soil and any potential covering of the soil (such as a road surface). The physical properties of the soil can vary significantly within small areas. For example, soil can be layered with a thin layer of top soil covering a layer of clay or asphalt covering gravel covering soil and can be a heterogeneous mixture of several soil types (e.g., sand and clay). Soil can have significantly varying density within a small region [26]. Roots of vegetation can hold water. Recent rain or snow can lead to variable moisture content within the soil. Various minerals can significantly affect the transmission of electromagnetic waves through the soil. To further complicate matters, the type of material used to encase the explosive object can interact with different soils in different ways. For example, plastic casings can have very similar electrical properties as some soils under some conditions. Wood casings can absorb moisture. All these factors can have significant effects on the received sensor data and are generally unknown to an autonomous algorithm due to their wide variability over a small range. The implication for autonomous detection is that different sensors and different types of algorithms are useful for different conditions. These different algorithms must use different signal conditioning, or preprocessing, and feature extraction.

In this paper, we propose an adaptive approach to information fusion and apply it to detect buried land mines using data collected by the NIITEK robotic mine detection system. This system includes a GPR and a WEMI sensor and is shown in Fig. 1. Each sensor collects data as the system moves.

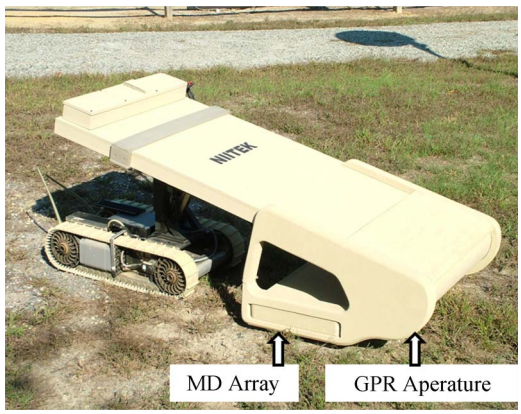


Fig. 1. NIITEK autonomous mine detection system.

The raw data from each sensor are then processed by various discrimination algorithms to make a decision as whether there is a mine in a certain location. We will also refer to these discrimination algorithms as software detectors, or simply algorithms or detectors. Throughout this paper, we will use the terms algorithm and detector interchangeably. Each detector operates on one of the sensor data to ultimately produce a set of class confidence values. It includes two main components: feature extraction and classification. The feature extraction component reduces the raw sensor data to a lower dimensional set of salient descriptors that represent the data. The classification component uses various models and methods to assign a confidence value that a mine is present at a point. To maximize the benefits of fusion and to take advantage of the different sensors and detectors, we select a diverse set of detectors that process the raw sensor data in different domains (time or frequency), extract different features, and use different classification methods. The performance of the different algorithms can vary due to factors other than the local minima of objective functions. Performance can differ due to each stage in the process and can be difficult to analyze. Our proposed fusion approach focuses on identifying the strengths and weaknesses of these algorithms in different regions of the feature space.

The rest of this paper is organized as follows. Section II-A gives a literature overview of classifier fusion. In Section III, we describe the WEMI and GPR data, the preprocessing steps, and the four distinct land mine detection algorithms that will be used to illustrate the proposed context-dependent fusion (CDF) approach which is described in Section IV. The experimental results are presented in Section V, and concluding remarks are given in Section VI.

II. Related Work

A. Combination of Multiple Classifiers

For complex detection and classification problems involving data with large intraclass variations and noisy inputs, perfect solutions are difficult to achieve, and no single source of information can provide a satisfactory solution. As a result, combination of multiple classifiers (or multiple experts) is playing an increasing role in solving these complex pattern recognition problems and has proven to be a viable alternative to using a single classifier. Classifier combination is mostly a

heuristic approach and is based on the idea that classifiers with different methodologies or different features can have complementary information. Thus, if these classifiers cooperate, group decisions should be able to take advantage of the strengths of the individual classifiers, overcome their weaknesses, and achieve a higher accuracy than any individual's.

Methods for combining multiple classifiers can be classified into two main categories: classifier selection and classifier fusion. Classifier selection methods assume that the classifiers are complementary and that their expertise varies according to the different areas of the feature space. For a given test sample, these methods attempt to predict which classifiers are more likely to be correct. Some of these methods consider the output of only one classifier to make the final decision [27]. Others combine the output of multiple "local expert" classifiers [28]. Classifier fusion methods assume that the classifiers are competitive and are equally experienced over the entire feature space. For a given test sample, the individual classifiers are applied in parallel, and their outputs are combined in some manner to take a group decision.

Over the past few years, a variety of schemes have been proposed for combining multiple classifiers. The most representative approaches include majority vote [29], Borda count [30], weighted average [31], Bayesian [32], probabilistic [33], polling methods [34], logistic regression [30], and combination by neural networks [35], fuzzy integral [36], Dempster–Shafer [37], and hierarchical mixture of experts [38].

Another way to categorize classifier combination methods is based on the way that they select or assign weights to the individual classifiers. Some methods are global and assign a degree of worthiness, that is averaged over the entire training data, to each classifier. Other methods are local and adapt the classifiers' worthiness to different data subspaces. Intuitively, the use of data-dependent weights, when learned properly, provides higher classification accuracy. This approach requires partitioning the input samples into regions during the training phase. The partition can be defined from the space of individual classifier decisions, according to which classifiers agree with each other [30], or by features of the input space [39]. Then, the best classifier for each region is identified and is designated as the expert for this region [40]. Conversely, the partitioning can be defined such that each classifier is an expert in one region [27]. This approach may be more efficient; however, its implementation is not trivial. In the classification phase, the region of an unknown sample is identified, and the output of the classifier that is responsible for this region is used to make the final decision. Data partition and classifier selection could also be made dynamic during the testing phase [41], [42]. In this case, the accuracy of each classifier (with respect to the training samples) is estimated in local regions of the feature space in the vicinity of the test sample. The most accurate classifier is selected to classify the test sample.

Another approach for building multiple classifiers is based on bagging and boosting [43]. Each classifier is trained using a different subset of the training set. The different subsets are obtained from the original using sampling. The final output is obtained by voting. Bagging specifically refers to the process of generating training subsets by sampling with replacement multiple times. A classifier is trained on each subset. All classifiers are used to classify a test sample. The outputs are combined via voting. Boosting generally refers to a more sequential process of building multiple classifiers on a training set. The general idea is that an initial classifier is trained on the training set. Points for which the initial classifier performs poorly are weighted more strongly in training a different classifier. The process is repeated multiple times in order to try and build a multiclassifier system consisting of classifiers that perform well on subsets of the training set. Boosting can cause problems by overfitting classifiers on subsets of the training data [44].

B. Fusion Methods for Land Mine Detection

Several of the fusion methods outlined in the previous section have been proposed and applied to fuse the results of land mine detection algorithms developed for WEMI (or EMI) and/or GPR sensors ([45]–[49] and references therein). Most of these methods fall under either the feature-level fusion category or the decision-level fusion category. Feature-level fusion combines various features. These features may come from different sensors or from the same raw data. In the latter case, various features can be generated by different feature extraction methods. Decision-level fusion combines decisions coming from different detection algorithms. These methods have proved to be more effective than the single detectors. However, because they rely on a global fusion strategy, they cannot adapt to different environments. For instance, if one of the algorithms has poor performance on 95% of the data but outperforms all other algorithms for the remaining 5%, then global fusion methods may not be able to take advantage of the strength of this algorithm.

Recently, a few local and adaptive fusion methods have been proposed and applied to the problem of land mine detection [50]–[52]. These methods partition the feature space into regions and adapt the fusion strategy to each region. In [50] and [51], the context extraction for local fusion approach is based on optimizing an objective function that combines context identification and multialgorithm fusion criteria into a single objective function. Since this approach involves complex optimization, only linear weighting schemes can be used for fusion. In [52], we proposed a CDF approach that performs context extraction and algorithm fusion as two independent tasks. This separation allows the integration of various fusion methods. In [52], the CDF approach was introduced with simple fusion weights and was used to fuse the results of multiple algorithms developed for a GPR sensor to detect antitank (AT) mines.

In this paper, we generalize the CDF approach and describe it in more details. In particular, we generalize it to fuse algorithms developed for multiple sensors (GPR and WEMI) to detect both AT and antipersonnel (AP) mines. We also provide a more robust approach to estimate the degree of worthiness of each algorithm in each context.

III. Data Collection and Processing

The data used to illustrate and validate the proposed fusion method were collected using the NIITEK robotic mine detection system. This system was used to acquire large collections

of colocated GPR and WEMI data from geographically distinct test sites. The NIITEK GPR sensor has been investigated for several years, and various detection algorithms have been developed and proved to be promising. In this paper, we describe three of these algorithms and use them to illustrate the proposed fusion method. On the other hand, WEMI is a relatively newer sensor, and only few algorithms, that process the data in a similar fashion, have been developed for this sensor. For this reason, only one WEMI detection algorithm is outlined and used in our experimental validation.

A. WEMI Sensor

1) Data Collection and Representation: The WEMI sensor was developed by Scott [12]. The sensor measures the response of an object at 21 logarithmically spaced frequencies over the range 330 Hz–90 kHz. The goal is to obtain characteristic spectral shapes that can help discriminate objects of interest from false alarms.

The response of the system can be modeled as

$$S(w) = A\left[I(w) + iQ(w)\right] \tag{1}$$

where w is the frequency, A is the magnitude, and the functions $I(w)$ and $Q(w)$ are the normalized real, also referred to as the in-phase, and imaginary, also referred to as the quadrature, parts of the complex response of the system. They are described in [12]. In (1), the term $I(w) + iQ(w)$ describes the shape of the response as a function of frequency. In this paper, an input data point is composed of 21 complex responses at the following measured frequencies (in hertz): 330, 390, 510, 690, 930, 1230, 1650, 2190, 2910, 3930, 5190, 6930, 9210, 12 210, 16 230, 21 630, 28 770, 38 250, 50 850, 67 650, and 90 030.

2) Discrimination Algorithm: Before feature extraction, the I and Q values are normalized between zero and one. This eliminates variation in magnitude due to several factors—such as the depth of the buried object to be detected as well as the metal mass and content—that do not affect the shape of the response curve. The magnitude can always be measured separately. After normalization, the response models proposed by Miller *et al.* [53] are used to fit the curve. The three-parameter model is given by

$$I(w) + iQ(w) = q\left(s + \frac{(iw\tau)^{1/2} - 2}{(iw\tau)^{1/2} + 1}\right) \tag{2}$$

where q, s, and τ are the three parameters describing the shape of the response curve [54]. To fit the curve, we used a built-in Matlab function *lsqcurvefit* that fits the functional form in (2) to the data. This function uses the algorithms described in [55] and [56] to find the values of the parameters (in this case, q, s, and τ) that minimize the squared error between the measured data $I(w) + iQ(w)$ and the model, given by the right-hand side of (2). The parameters resulting from this curve fit plus the error in the fit provide four features. Fig. 2 shows the response curves and their curve fits of metallic and nonmetallic objects. We note that other researchers, such as Fails *et al.* [13] and Gader *et al.* [14], have also used these model parameters as features.

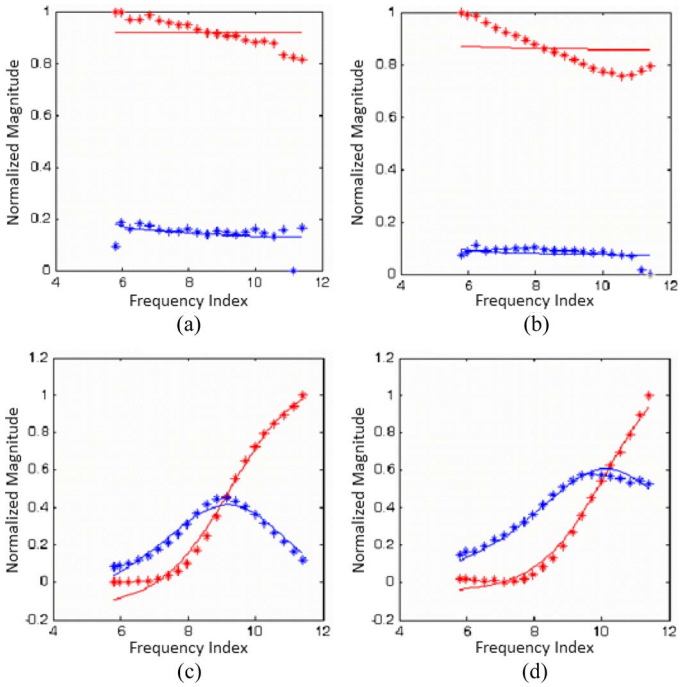


Fig. 2. (Sequences of dots) Response curves and (smooth curves) their curve fits from (a) blank, (b) a nonmetallic clutter item, (c) a metallic clutter item, and (d) an LM mine. The red color represents $I(w)$, and the blue color represents $Q(w)$.

In addition to the four features provided by the model, three spread features [57] are used. These are defined by the following equations in which I and Q represent the in-phase (real) and quadrature (imaginary) values at each frequency and N is the number of frequencies

$$\begin{aligned} Q_{\text{sum}} &= \sum_{i=1}^{N} Q_i \\ Q_{\text{spread}} &= \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} |Q_i - Q_j| \\ T_{\text{spread}} &= \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} |Q_i - Q_j| + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} |I_i - I_j|. \end{aligned} \tag{3}$$

Together, these make up the seven features used to describe a WEMI signal. Feature selection, using a greedy hill climbing with the well-known divergence measure, was performed. Four features were selected: τ, the fitting error, Q_{spread}, and T_{spread}. A multilayer perceptron (MLP) classifier was built from these features. The MLP parameters were identified through sixfold cross-validation. We will refer to this algorithm as the WEMI model fitting, or a WEMI-MFIT detector.

B. GPR Sensor

1) Data Collection and Representation: The GPR sensor [58] collects 24 channels of data. Adjacent channels are spaced approximately 6 cm apart in the cross-track direction, and sequences (or scans) are taken at approximately 1 cm down-track intervals. The system uses a V-dipole antenna that generates a wideband pulse ranging from 200 MHz to 7 GHz.

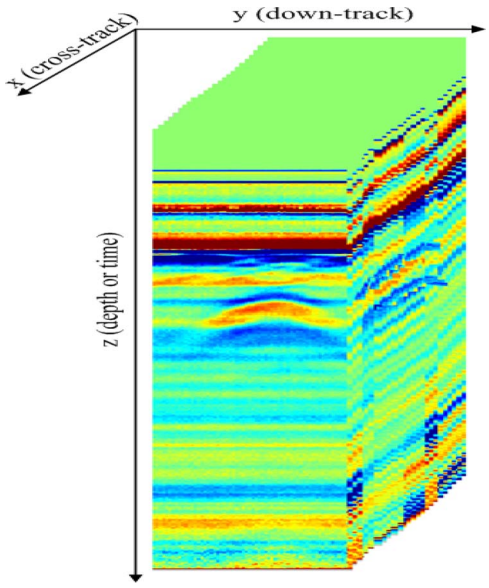


Fig. 3. Collection of few GPR scans for an AT mine with HM content buried at 2 in.

Each A-scan, i.e., the measured waveform that is collected in one channel at one down-track position, contains 416 time samples at which the GPR signal return is recorded. Each sample corresponds to roughly 8 ps. We often refer to the time index as depth although, since the radar wave is traveling through different media, this index does not represent a uniform sampling of depth. Thus, we model an entire collection of input data as a 3-D matrix of sample values $S(z,x,y)$, $z = 1,\ldots,416; x = 1,\ldots,24; y = 1,\ldots,N_S$, where N_S is the total number of collected scans and the indexes z, x, and y represent the depth, the cross-track position, and the down-track positions, respectively. A collection of scans, forming a volume of data, is shown in Fig. 3.

Fig. 4 shows several B-scans (sequences of A-scans) both down track (formed from a time sequence of A-scans from a single sensor channel) and cross track (formed from each channel's response in a single sample) at the position indicated by a line in the down track. The objects scanned are the following: 1) a high-metal (HM)-content AT mine; 2) an HM-content AP mine; and 3) a wood block. The reflections between depths 50 and 125 in these figures are the artifact of preprocessing and data alignment. The strong reflections between cross-track scans 15 and 20 are due to electromagnetic interference (or EMI). The preprocessing artifacts and the EMI can add considerable amounts of noise to the signatures and make the detection problem more difficult.

Preprocessing is an important step to enhance the GPR mine signatures for detection. In the proposed system, first, we identify the location of the ground bounce as the signal's peak and align the multiple signals with respect to their peaks. This alignment is necessary because the vehicle-mounted system cannot maintain the radar antenna at a fixed distance above the ground. Since the system is looking for buried objects, the early time samples of each signal, up to few samples beyond the ground bounce, are discarded so that only data corresponding to regions below the ground surface are processed. The remaining signal samples are divided into N depth bins, and each bin would be processed independently. The reason for this segmentation is to compensate for the high contrast between the responses from deeply buried and shallow anomalies. Discarding samples above the ground may, in some cases, cause shallowly buried mines (less than 1 in) to be missed. To alleviate this potential drawback, we have developed a different algorithm for detecting surface mines that could be included in the fusion. However, for the data collection under consideration, we did not miss any shallow mines for this reason, and for simplicity, we did not include the surface mine detector in the fusion. Another alternative that is currently under consideration is to use other background subtraction techniques without discarding any samples.

2) Discrimination Algorithms: Various algorithms have been applied to the problem of discrimination between land mines and false alarms using GPR data. Generally, automated land mine discrimination algorithms can be broken down into three phases: preprocessing, feature extraction, and confidence assignment. Preprocessing algorithms perform tasks such as normalization of the data, corrections for variations in height and speed, removal of stationary effects due to the system response, etc. Methods that have been used to perform this task include wavelets and Kalman filters [59], [60], subspace methods and matching to polynomials [61], and subtracting optimally shifted and scaled reference vectors [62]. Feature extraction algorithms reduce the preprocessed raw data to form a lower dimensional salient set of measures that represent the data. Principal component transform is a common tool that has been used to achieve this task [63]. Other feature analysis approaches include wavelets [59], image processing methods of derivative feature extraction [16], curve analysis using Hough and Radon transforms [5], and model-based methods [17]. Confidence assignment algorithms can use methods such as Bayesian [5], hidden Markov models (HMMs) [16], [17], fuzzy logic [64], rules and order statistics [65], neural networks, or nearest neighbor classifiers [66] to assign a confidence that a mine is present at a point.

In this paper, we consider fusion of one detector operating on data collected by the WEMI sensor (WEMI-MFIT, outlined in Section III-A2) and three detectors, with distinct character, operating on data collected by the GPR sensor. These algorithms have performed well in extensive field testing and are being considered for real-time implementation in handheld and vehicle-mounted GPR systems. These algorithms are highlighted in the following.

1) *HMM Detector:* The HMM-based algorithm [16], [17] treats the down-track dimension as the time variable and produces a confidence that a mine is present at various positions (x,y) on the surface being traversed. In particular, a sequence of observation vectors is produced for each point. These observation vectors encode the degree to which edges occur in the diagonal and antidiagonal directions. In particular, for every point (x_s,y_s), the strengths for the positive/negative diagonal/antidiagonal edges are computed. Then, the observation vector at a point (x_s,y_s) consists of a set of four features that encode the maximum edge magnitude over multiple depth values around (x_s,y_s). The HMM algorithm has a background and a

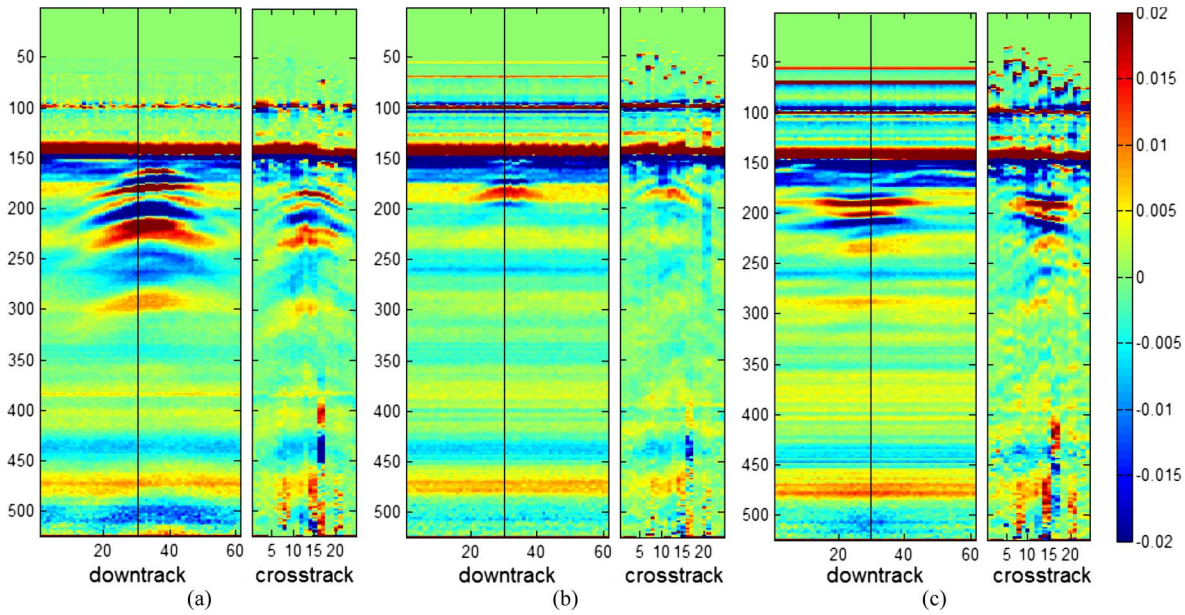


Fig. 4. NIITEK radar down- and cross-track (at a position indicated by a line in the down track) B-scan pairs for (a) an AT mine, (b) an AP mine, and (c) an NMC alarm.

mine model. Each model has three states and produces a probability value by backtracking through model states using the Viterbi algorithm [67]. The probability value produced by the mine (background) model can be thought of as an estimate of the probability of the observation sequence given that there is a mine (background) present. The mine model is a left-to-right model in that states are ordered and the transition probabilities for moving to a lower numbered state are zero. We refer the reader to [16] and [17] for more details about this algorithm.

2) *EHD Detector:* The edge histogram descriptor (EHD) uses translation-invariant features, that are based on the EHD of the 3-D GPR signatures, and a possibilistic k-nearest neighbors rule for confidence assignment [66]. The EHD captures the signature's texture and is an adaptation of the MPEG-7 EHD feature [68]. Specifically, each 3-D signature is divided into subsignatures, and the local edge distribution for each subsignature is represented by a histogram. To generate the histogram, local edges are categorized into five types: vertical, horizontal, diagonal (45° rising), antidiagonal (45° falling), and nonedges. A set of alarms with known ground truth is used to train the decision-making process. These labeled alarms are clustered to identify a small number of representatives that capture signature variations due to differing soil conditions, mine types, weather conditions, and so forth. We refer the reader to [66] for more details about this algorithm.

3) *SPECT Detector:* The spectral detector (SPECT) aims at capturing the characteristics of a target in the frequency domain. It extracts the alarm spectral correlation feature and formulates a confidence value based on similarity to prototypes that characterize mine objects [69]. The spectral features are derived from the energy density spectrum (EDS) of an alarm. The estimation of EDS involves three main steps: preprocessing, whitening, and averaging. Preprocessing estimates the ground level, aligns the data from each scan with respect to the ground level, and removes the data above and near the ground surface. This step is needed to avoid an EDS that is dominated by the response of the ground bounce. The whitening step performs equalization on the spectrum from the background so that the estimated EDS reflects the actual spectral characteristics of an alarm. Averaging reduces the variance in the EDS. We refer the reader to [69] for more details about this algorithm.

IV. CDF

A. Motivations

The proposed approach, which is called CDF, is motivated by the observation that the performance of most detection systems can be significantly affected by the mine type and the rapidly changing environmental conditions and that there is no single sensor or algorithm that can consistently outperform all others. In fact, the relative performance of different detectors can vary significantly depending on the mine type, geographical site, soil and weather conditions, and burial depth.

To illustrate the aforementioned point, in Fig. 5, we show the ROC's results of the four discrimination algorithms (EHD, HMM, SPECT, and MFIT) on subsets of data collected by the NIITEK robotic mine detection system. This data collection will be described in Section V. The different ROCs display the performance of the algorithms when different types of mines are scored. For instance, in Fig. 5(a), only AT mines are considered. In this case, the HMM and EHD detectors have the best performance, and MFIT has the worst. This is because AT mines are large enough to have good GPR signatures and many of them have low metal (LM) content. However, for

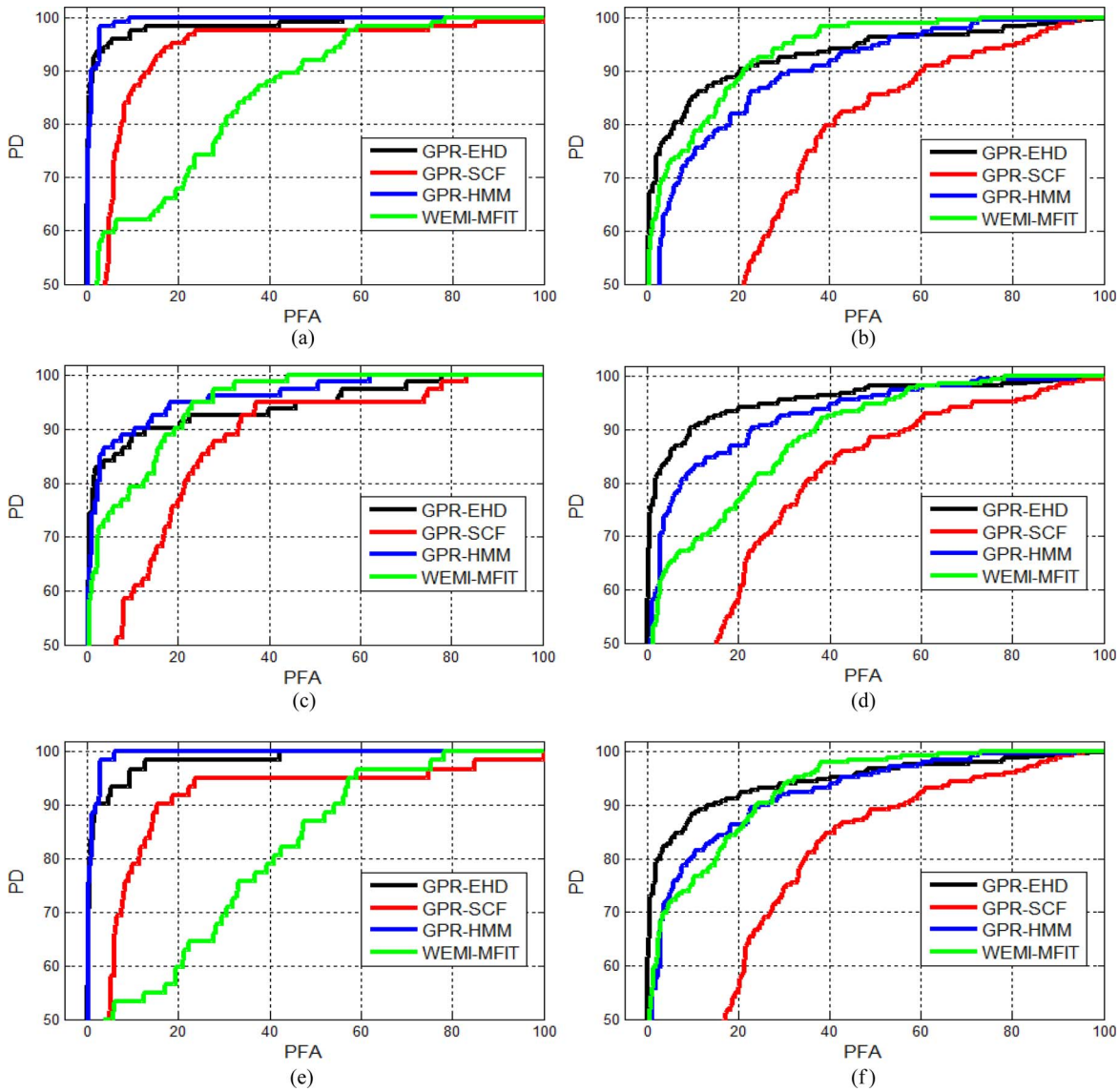


Fig. 5. PD versus PFA of four different detectors for different types of mines buried at different depths. (a) AT mines versus all clutter. (b) AP mines versus all clutter. (c) HM mines versus all clutter. (d) LM mines versus all clutter. (e) Deep mines versus all clutter. (f) Shallow mines versus all clutter.

AP mines, the MFIT detector has the best performance at a high probability of detection (PD), as shown in Fig. 5(b). In this case, several AP mines have weak GPR signatures and cannot be easily detected by any of the GPR algorithms. The relative performance of the different algorithms depends on other factors besides mine type. For instance, as shown in Fig. 5(e) and (f), the MFIT detector has a poor performance for deeply buried mines but a relatively better performance for shallow mines. The poor performance of the GPR detectors in the latter case may be due to the difficulty in decoupling the mine signatures and the background signal around the ground bounce area.

The relative performance of the different detectors is shown further in Fig. 6, where we display a scatter plot of the confidence values generated by the EHD and MFIT detectors for all alarms in the data collection. As it can be seen, the relative performance of the different algorithms can vary significantly. For instance, region R1 highlights a group of mines with LM content that are easily detected by the EHD (confidence close to one) and not by the MFIT (confidence less than 0.25). On the other hand, region R2 highlights a different set of mine alarms (both HM and LM) that are easily detected by the MFIT and

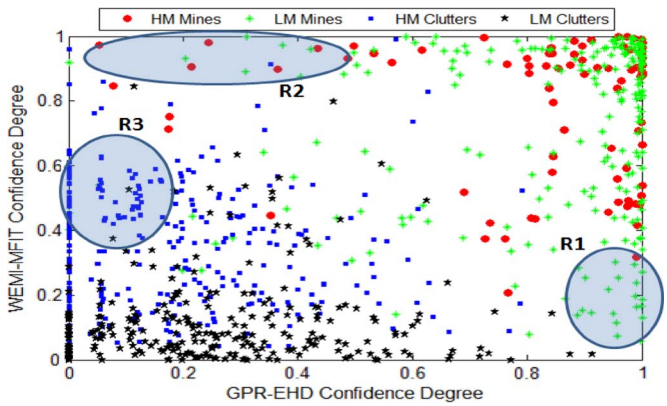


Fig. 6. Comparison of the confidence degrees assigned by the EHD and the MFIT for several mine and clutter signatures.

not the EHD. Most of these mines are buried at a depth less than 2 in, and their signatures are intertwined with the ground bounce. Region R3 displays a group of metal clutter alarms that will be detected by the MFIT and rejected by the EHD.

The aforementioned examples suggest using different discrimination algorithms and/or features, and possibly different background subtraction algorithm, to accommodate for the different mine types, burial depths, and other conditions. However, this task may not be as simple as it sounds since it is not possible to characterize the performance of each algorithm on all possible variations. Moreover, it may not be possible to know the characteristics of the test site. Thus, the selection of the optimal subset of algorithms is not a trivial task and needs to be learned in an unsupervised way.

B. CDF Framework

Motivated by the previous examples, we propose a CDF framework that can take advantage of the strengths of few algorithms in different regions of the feature space without being affected by the weaknesses of the other algorithms. CDF is a local approach that adapts the fusion method to different regions of the feature space. The partition of the input space into different regions is not explicit and does not require prior knowledge of the data type (e.g., geographical site, soil type, mine type, etc.). Instead, the training data are partitioned into clusters in a completely unsupervised manner based on the features extracted by the different detectors.

We will refer to these clusters as contexts. We do not expect the contexts to separate the mines from the clutter signatures. This is a much more challenging task and is actually the goal of the multisensor fusion algorithm. For instance, it is likely that the clustering algorithm groups some weak mines (e.g., deeply buried or AP mines) with clutter. However, it is reasonable to assume that contexts will tend to include signatures that share common attributes. For example, one context may include alarms (mines and/or clutter) with LM content, and a second context may group alarms (mines and/or clutter) collected from the same arid region. The basic motivation behind the proposed fusion is to quantify the performance of the different algorithms and develop a fusion strategy within each context.

The training component of CDF assumes that a set of labeled training alarms $\mathcal{X}$ is available. Let $\mathcal{M}$ denote the set of training mine signatures, and let $\mathcal{C}$ denote the set of training clutter signatures. That is, $\mathcal{X} = \mathcal{M} \bigcup \mathcal{C}$. We assume that each alarm has been processed by K distinct detection algorithms: $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$. Each algorithm $\mathcal{D}_i$ extracts a set of features F^i from an alarm $\mathbf{x_j}$ and assigns a confidence value y_{ij}^k to each class k. Standard fusion methods would aggregate the algorithm confidence values y_s without considering the features extracted by these algorithms. For instance, Bayesian [32] and other probabilistic methods [33] attempt to find the optimal partition of the confidence space. Other methods sort the confidence values and then combine them [30]. The proposed CDF uses both the features extracted by the algorithms and their confidence values.

In the proposed land mine application, we have $K = 4$, where $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$, and $\mathcal{D}_4$ correspond to the EHD, HMM, SPECT, and MFIT algorithms, respectively. These algorithms were selected due to their individual performance and to the minimal correlation among them. In fact, each algorithm uses different preprocessing, feature extraction, and classification method as mentioned earlier. We also have a two-class problem. Let Ω_1 denote the mine class label and Ω_2 denote the clutter class label. In this case, $y_{ij}^2 = 1 - y_{ij}^1$, and we simply drop the superscript and use y_{ij} to denote the confidence in the class of mines.

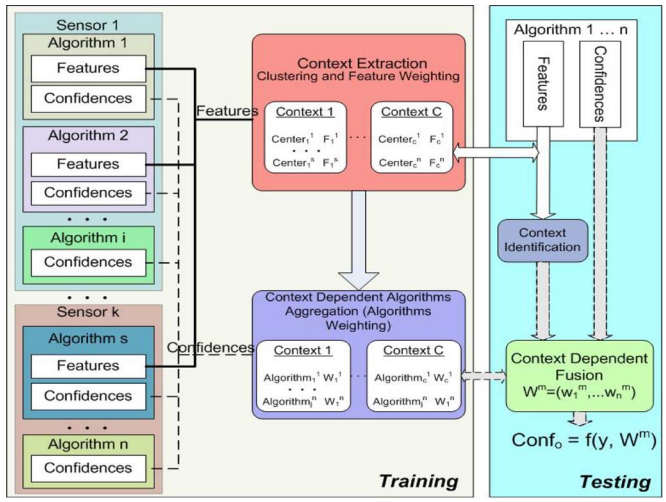


Fig. 7. Architecture of the proposed CDF.

Fig. 7 shows the overall architecture of the training and testing phases of the proposed CDF scheme. This figure highlights the two main components of the training phase, namely, context extraction and algorithm fusion. In context extraction, the features extracted by the different algorithms (from different sensors) are combined, and a clustering algorithm is used to partition the training signatures into groups of similar signatures, or contexts, and learn the relevant features within each context. Here, we are assuming that signatures that have similar response to different algorithms (e.g., signatures highlighted in the same region in Fig. 6) share some common features and would be assigned to the same cluster. The algorithm fusion component assigns a degree of worthiness to each detector in each context based on its relative performance within the context. These components are described in more details in the following sections.

1) Training Step 1: Context Extraction: The context extraction step is completely unsupervised and does not use the alarms' labels. In this step, the feature sets extracted by the different algorithms are pooled together to form a multidimensional feature space. Alarms that share common attributes (type, geographical sites, etc.) are expected to cluster in subspaces of this aggregate feature space. Thus, to identify groups of alarms with homogeneous properties, we need to partition this space into regions, or contexts.

Any clustering algorithm could be used to achieve this task. However, due to the large number and diversity of the feature sets, a clustering algorithm that can partition the data while learning the relevant features for the different clusters would provide a more reliable partition. For instance, using a standard clustering algorithm, the partition would be influenced by the

EHD features (40-D) more than the MFIT features (3-D). For this reason, we use the simultaneous clustering and attribute discrimination (SCAD) [70]. SCAD treats the features as subsets and seeks the optimal relevance weight for each subset within each cluster.

In the following, we assume that each individual detection algorithm $\mathcal{D}_s$ has its own set of features F^s with $|F^s| = k^s$. The aggregate feature space $FS = \bigcup_{s=1}^{K} F^s$ is a p-dimensional feature space with $p = \sum_{s=1}^{K} k^s$. Let $\mathcal{X} = \{\mathbf{x}_j \in \Re^p | j = 1, \ldots, N\}$ be a set of N training alarms in the p-dimensional feature space. Let $\mathbf{C} = (c_1, c_2, \ldots, c_C)$ represent the centroids of the C clusters. Let u_{ij} represent the membership of $\mathbf{x}_j$ in cluster i. The $C \times N$ fuzzy C partition $\mathbf{U} = [u_{ij}]$ satisfies [71]

$$\begin{cases} u_{ij} \in [0,1] & \forall\, i \\ 0 < \sum_{j=1}^{N} u_{ij} < N & \forall\, i, j \\ \sum_{i=1}^{C} u_{ij} = 1 & \forall\, j. \end{cases} \tag{4}$$

Let $d_{ij}^s = \|x_j^s - c_i^s\|$ be the partial Euclidean distance between $\mathbf{x}_j$ and the center of cluster i taking into account only the sth feature subset. Let $\mathbf{V} = [v_{is}]$ be the matrix of relevance weights for F^s with respect to cluster i for $s = 1, \ldots, K$. The total distance D_{ij} between $\mathbf{x}_j$ and cluster i is then computed by aggregating the partial distances and their weights, i.e.,

$$D_{ij}^2 = \sum_{s=1}^{K} v_{is} \left(d_{ij}^s\right)^2. \tag{5}$$

Using this notation, the SCAD algorithm [70] minimizes

$$J(\mathbf{C}, \mathbf{U}, \mathbf{V}; \mathcal{X}) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^m \sum_{s=1}^{K} v_{is} \left(d_{ij}^s\right)^2 + \sum_{i=1}^{C} \delta_i \sum_{s=1}^{K} v_{is}^2 \tag{6}$$

subject to the constraint in (4) and

$$\begin{aligned} & v_{is} \in [0,1] \qquad \forall\, i, s \\ & \sum_{s=1}^{K} v_{is} = 1 \qquad \forall\, i. \end{aligned} \tag{7}$$

The objective function in (6) has two components. The first one, which is similar to the fuzzy c-means objective function [71], is the sum of feature-weighted Euclidean distances to the centers, additionally weighted by constrained memberships. This component allows us to obtain compact clusters. From a feature-relevance point of view, this term is minimized when, in each cluster, only one feature is completely relevant while all other features are irrelevant. The second component in (6) is the sum of the squared feature weights. The global minimum of this component is achieved when all the features are equally weighted. When both components are combined and the coefficients δ_i are chosen properly, the final partition will minimize the sum of intracluster weighted distances, where the weights are optimized for each cluster.

The choice of δ_i in (6) is important to the performance of SCAD since it reflects the importance of the second term relative to the first term. If δ_i is too small, then the first term dominates, and only one feature in cluster i will be maximally relevant and assigned a weight of one, while the remaining features get assigned zero weights. On the other hand, if δ_i is too large, then the second term will dominate, and all features in cluster i will be assigned equal weights. Hence, the values of δ_i should be chosen such that both terms are of the same order of magnitude. This can be accomplished by updating δ_i in iteration t using

$$\delta_i^{(t)} = \frac{\sum_{j=1}^{N} \left(u_{ij}^{(t-1)}\right)^m \sum_{k=1}^{n} v_{ik}^{(t-1)} \left(d_{ijk}^{(t-1)}\right)^2}{\sum_{k=1}^{n} \left(v_{ik}^{(t-1)}\right)^2}. \tag{8}$$

In (8), the superscript $(t-1)$ is used to denote the value of the variable in iteration $(t-1)$.

Using a gradient descent optimization and a set of Lagrange multipliers for the constraints, the optimization of J with respect to $\mathbf{V}$ yields

$$v_{is} = \frac{1}{K} + \frac{1}{2\delta_i} \sum_{j=1}^{N} (u_{ij})^m \left[D_{ij}^2/K - \left(d_{ij}^s\right)^2\right]. \tag{9}$$

The first term in (9) $(1/K)$ is the default value if all K feature subsets (extracted by the K detectors) are treated equally, and no discrimination is performed. The second term is a bias that can be either positive or negative. It is positive for compact feature sets, where the partial distance is, on the average, less than the total distance (normalized by the number of features). If a feature set is compact, compared to the other features, for most of the points that belong to a given cluster (high u_{ij}), then it is very relevant for that cluster. Similarly, minimization of J with respect to $\mathbf{U}$ subject to the constraints in (4) yields

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(D_{ij}^2/D_{kj}^2\right)^{\frac{1}{m-1}}}. \tag{10}$$

Minimization of J with respect to the prototype parameters depends on the choice of d_{ij}^s. Since the partial distances are treated independent of each other (i.e., disjoint feature subsets) and since the second term in (6) does not depend on the prototype parameters explicitly, the objective function in (6) can be decomposed into K independent problems

$$J_s = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^m v_{is} \left(d_{ij}^s\right)^2, \qquad \text{for } s = 1, \ldots, K. \tag{11}$$

Each J_s would be optimized with respect to a different set of prototype parameters. For instance, if d_{ij}^s is a Euclidean

distance, minimization of J_s would yield the following update equation for the centers of subset s:

$$\mathbf{c}_i^s = \frac{\sum_{j=1}^{N} u_{ij}^m \mathbf{x}_j^s}{\sum_{j=1}^{N} u_{ij}^m}. \tag{12}$$

SCAD is an iterative algorithm that starts with an initial partition and alternates between the update equations of u_{ij}, v_{is}, and c_i^s. The SCAD algorithm is summarized as follows:

Simultaneous Clustering and Attribute Discrimination
Fix the number of clusters C;
Fix m, $m \in (1, \infty)$;
Initialize the fuzzy partition matrix $\mathbf{U}$ randomly;
Initialize the relevance weights to $1/K$;
Repeat
  Compute $(d_{ij}^s)^2$ for $1 \le i \le C$, $1 \le j \le N$,
  and $1 \le s \le K$;
  Update the relevance weights v_{is} using (9);
  Compute D_{ij}^2 using (5);
  Update the partition matrix U using (10);
  Update the centers using (12);
Until (centers to not change significantly between consecutive iterations)

We should point out here that, like many other clustering applications, there is no guarantee that SCAD will converge to the optimal partition. In fact, one can argue that, for this application, there is no such thing as "optimal" grouping. Our expectation from the clustering process is that alarms that exhibit a similarity with respect to some features would be grouped. Moreover, since we are using a fuzzy clustering algorithm, where alarms get assigned to multiple clusters with various membership degrees, the results of the subsequent fusion steps are not very sensitive to the clustering step.

2) Training Step 2: Algorithm Fusion: Any of the fusion methods mentioned in Section II-A could be integrated within our CDF. Training data from each identified context would be used to learn the optimal fusion parameters and identify *local experts* for that region of the feature space. In this paper, we report the results using linear discriminant functions, which have been mathematically analyzed for fusion [72].

The detection algorithms generate confidence values with different dynamic ranges. Thus, to avoid potential biases, we map the algorithms' output to the [0, 1] interval before fusing them. From probability theory, we know that, when a random variable z takes values in the set of real numbers, the probability distribution $f_Z(.)$ can be completely described by the cumulative distribution function (or called the probability distribution function) $F_Z(.)$, whose value at each real z is the probability that the random variable is smaller than or equal to z, i.e.,

$$z \mapsto F_Z(z) = P(Z \le z). \tag{13}$$

Using this cumulative function, the decision space $\mathbf{y}$ can be transformed into a mapped posterior probability (MPP). Here, we define the conditional MPP (cMPP) $p(\Omega_k|y_{ij})$ for each sample x_j and algorithm i in class Ω_k using the cumulative mapping of y_{ij} (confidence value assigned by algorithm i to sample j in the class of mines) for all $x_j \in \Omega_k$. Then, the performance of each algorithm can be estimated based on their cMPP value on the training data.

Within each context c, let $\boldsymbol{w}^c = \{w_1^c, w_2^c, \ldots, w_K^c\}$ be a vector of real numbers constrained to

$$\sum_{i=1}^{K} w_i^c = 1 \qquad 0 \le w_i^c \le 1 \tag{14}$$

where K is the number of algorithms to be fused.

The classifiers' weights are determined based on the relative separation of the distribution of the mine (Ω_1) and clutter (Ω_2) confidence values. Intuitively, algorithms with larger separation are considered more "expert" since they can discriminate between the class boundaries and thus will be assigned larger weights. We define the degree of separation between the mines and clutter signatures using algorithm i in context c as

$$Sep_i^c = P(\Omega_1|c) \sum_{x_j \in \mathcal{M}} \left(p(\Omega_1|y_{ij}) - p(\Omega_2|y_{ij})\right) \mu_c(x_j) + P(\Omega_2|c) \sum_{x_j \in \mathcal{C}} \left(p(\Omega_2|y_{ij}) - p(\Omega_1|y_{ij})\right) \mu_c(x_j) \tag{15}$$

where $\mu_c(x_j)$ is the fuzzy membership degree assigned by SCAD to each alarm signature x_j in cluster c. A high value (close to one) indicates that $\mathbf{x_j}$ is a typical point of cluster c. In (15), $P(\Omega_1|c)$ and $P(\Omega_2|c)$ are the prior probabilities in the mine and clutter classes given that an alarm is assigned to context c. These probabilities are estimated using

$$P(\Omega_1|c) = \frac{\sum_{x_j \in \mathcal{M}} \mu_c(x_j)}{\sum_{j=1}^{N} \mu_c(x_j)} \tag{16}$$

$$P(\Omega_2|c) = \frac{\sum_{x_j \in \mathcal{C}} \mu_c(x_j)}{\sum_{j=1}^{N} \mu_c(x_j)}. \tag{17}$$

The first term in (15) is maximized when typical mine signatures of context c [i.e., alarms with high $\mu_c(x_j)$] have high probability of being a mine and low probability of being a clutter. Similarly, the second term is maximized when typical clutter signatures have high probability of being a clutter and low probability of being a mine.

The class separations are then converted into degrees of worthiness w_i^c, such that the constraint in (14) is satisfied, using

$$w_i^c = \frac{Sep_i^c}{\sum_{i=1}^{N} Sep_i^c}. \tag{18}$$

3) Testing Step: To test a new signature (x_t) using CDF, each detector $(\mathcal{D}_i)$ would extract its set of features F^i and assigns a confidence value y_{it}. The K sets of descriptors are then used to identify the closest context. This is achieved by comparing the features of the test sample to the centroids of the clusters representing the different contexts. The partial distances, produced by the features of each algorithm, are combined using the cluster-dependent feature-relevance weights

learned in the context extraction phase. Once context c has been identified as the closest context to the sample being tested, the confidence values of the individual algorithms would be aggregated using the optimal degrees of worthiness w_i^c associated with context c, i.e., the overall confidence value assigned to the test sample x_t is computed using

$$Conf(x_t) = \sum_{i=1}^{K} w_i^c \times y_{it}. \tag{19}$$

4) Computational Complexity: The proposed CDF approach is generic and does not require a specific set of features or classifiers. Thus, its overall computational complexity cannot be determined. However, we can compare it to the alternative solution with similar settings. In this paper, we have argued that the land mine detection problem is a challenging task and that a single classifier with one set of features is not adequate, and fusion of multiple classifiers from different sensors is needed for this task. Thus, the computational complexity of CDF could be compared to the complexity of a standard global fusion method. In the next section, we illustrate the performance of CDF using the four discrimination algorithms outlined in Section III-B2. Since these classifiers and their features could be substituted by any other sets and since the same features and classifiers would be needed by any other fusion method, we do not consider their computational complexity.

The CDF has two additional steps over standard global fusion. The first one consists of partitioning the training data into clusters or contexts. This is an offline step that needs to be done only once and thus does not affect the computational complexity in the testing mode. The second step consists of identifying the closest context to a test sample. This step involves simple distance computation to identify the nearest cluster prototype. It requires $\mathcal{O}(\mathcal{C} \times p)$ computations, where $\mathcal{C}$ is the number of contexts and p is the dimensionality of the composite feature vector representing the alarm. Since the expected number of contexts $\mathcal{C}$ is typically small (less than 20) and the number of algorithms to be fused is around four, this additional computation is negligible when compared to the computation needed for feature extraction and classification.

V. Experimental Results

A. Data Statistics

The proposed CDF and the four discrimination algorithms were developed and tested with data collected using the NIITEK robotic mine detection system vehicle. The data were collected in May 2007 from two geographically distinct test sites (site A and site B) in the eastern U.S. with natural soil. The two sites are partitioned into grids with known mine locations. In all, there are 28 distinct mine types that can be classified into four categories: AT metal (ATM), AT with LM content (ATLM), AP metal (APM), and AP with LM content (APLM). The targets were buried up to 5 in deep. Multiple data collections were performed at each site, resulting in a large and diverse collection of mine and clutter signatures.

In our data collection, false alarms arise as a result of sensor signals that present a minelike character. Such signals are generally said to be a result of clutter. Clutter arises from two different processes. One type of clutter is emplaced and surveyed. Objects used for this clutter can be classified into two categories: HM clutter (HMC) and nonmetal clutter (NMC). HMC such as steel scraps, bolts, and soft drink cans was emplaced and surveyed in an effort to test the robustness of the detection algorithms and, in particular, the MFIT algorithm. NMC such as concrete blocks and wood blocks was emplaced and surveyed in an effort to test the robustness of the GPR-based detection algorithms. The other type of clutter, referred to as blank, is caused by disturbing the soil.

Overall, the data collection includes a total of 311 mine signatures and 564 nonmine, or clutter, signatures. The statistics of these collections are shown in Table I. The depth distribution for all objects is shown in Table II.

TABLE I
Statistics of the Data Collection

Type	Content	Site A	Site B	Total/Category	Total/type
AP	HM	16	40	56	187
	LM	38	93	131	
AT	HM	6	20	26	124
	LM	28	70	98	
Clutter	HMC	224	68	292	564
	NMC	72	68	140	
	BLANK	52	80	132	
Total		436	439	875	875

TABLE II
Burial Depth of All Objects in the Data Collection

	Mine			Clutter		
Depth	Site A	Site B	Total	Site A	Site B	Total
Surface	0	27	27	52	80	132
(0 1"]	12	104	116	70	46	116
(1 2"]	36	48	84	78	44	122
(2 3"]	28	34	62	88	18	106
(3 4"]	12	0	12	60	20	80
(4 5"]	0	10	10	0	8	8
Total	88	223	311	348	216	564

B. Evaluation Methods

To provide an objective and consistent evaluation of the different algorithms, we use a testing/training unified framework (TUF) system. This is a graphical user interface that supports creation of supervised learning algorithms that perform discrimination between targets and nontargets in data collected at a variety of different regions (mine lanes) in a variety of different sites. The framework employs algorithms (EHD, HMM, SPECT, and MFIT) implemented in Matlab using a control flow that processes raw data files into alarms with associated Universal Transverse Mercator coordinates and confidence values. The alarms are then processed by extracting signatures. These signatures are passed to a discrimination algorithm, which extracts a set of features and produces a confidence for each alarm. The features extracted by the

discrimination algorithms and their outputs are then passed to the fusion algorithm.

All discrimination and fusion algorithms were trained and tested using sixfold cross-validation. To assess the statistical significance of the performance improvements of the proposed fusion, for each cross-validation, we repeat the experiment 20 times. Each time, we initialize the clustering process with a different random initialization. This variation can cause SCAD to partition the data in a different way and thus lead to different contexts and different degrees of worthiness to each algorithm. The results of each run are scored using the Mine Detection Assessment and Scoring system developed by the Institute for Defense Analyses [73]. The scoring is performed in terms of PD versus probability of false alarm (PFA). Confidence values are thresholded at different levels to produce an ROC curve. For a given threshold, a mine is detected if there is an alarm within 0.25 m from the edge of the mine with confidence value above the threshold. Given a threshold, PD is defined to be the number of mines detected divided by the number of mines. For each detection algorithm and for the global fusion, we report the results by a single ROC curve. For the proposed CDF, we report the results by one ROC that averages the results of the 20 runs, and we also show the standard deviations at multiple PFA locations.

Land mine detection algorithm evaluation is application dependent. For instance, in humanitarian demining, the best algorithm may be the one at which 100% detection is achieved with the lowest FAR, no matter what other properties the ROC may display. For other time-critical demining applications, where some level of missed mines is not considered as great a cost, the best ROC may be the one at which the PD is highest at a given constant FAR. Our algorithm development efforts fall in the latter category and have been geared toward developing algorithms that are suitable for an autonomous mine detection system. In any such system, false alarms will delay the progress of the system.

C. Analysis of the Training Phase

1) Context Extraction: For each cross-validation, the training data consist of a set of colocated GPR and WEMI alarms. Each alarm is processed by the four discrimination algorithms (EHD, HMM, SPECT, and MFIT) outlined in Section III-B2. The features extracted from these alarms are then fed to SCAD to partition the aggregated feature space into C clusters. The choice of the number of clusters is not critical for this application. This number should be large enough so that most clusters contain only similar alarms. However, it should not be too large to avoid using too many small clusters that do not include enough samples to learn the optimal algorithm fusion weights. In this paper, we let $C = 10$.

Table III displays the content of the ten identified clusters for one run of one cross-validation. As it can be seen, most clusters include alarms of similar types and thus may be considered as a homogeneous context. For instance, some clusters are dominated by HM mines and HMC. Others are dominated by AT mines or AP mines. Furthermore, some clusters include mainly mine or clutter alarms. Others include a mixture of both.

TABLE III
DISTRIBUTION OF THE ALARMS AMONG THE TEN CLUSTERS FOR ONE CROSS-VALIDATION SET

Cluster	ATHM	ATLM	APHM	APLM	HMC	NMC	Blank	Total
1	0	18	0	0	0	20	0	38
2	16	0	11	0	6	0	0	33
3	0	0	12	1	28	0	0	41
4	0	0	5	15	34	0	0	54
5	0	0	0	16	39	0	0	55
6	0	0	0	0	9	31	93	133
7	4	4	24	20	54	0	0	106
8	0	3	0	45	31	1	1	81
9	2	27	0	0	0	1	0	30
10	0	42	0	22	24	51	19	158
Total	22	94	52	119	225	104	113	729

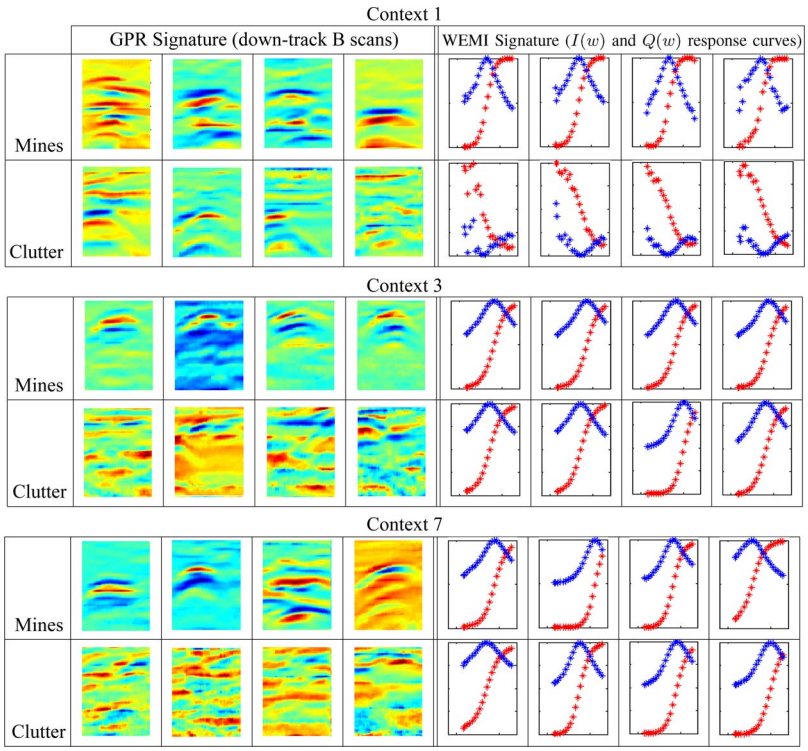
Alarms that are grouped into the same context share common GPR and/or WEMI features.

Table IV displays few representative mine and clutter alarms from three contexts. For instance, context 1 includes only AT mines with LM content and NMC (refer to Table III). Alarms from both classes have strong GPR signatures, and the GPR sensor by itself may not be sufficient to discriminate between mines and clutter within this context. The WEMI sensor, on the other hand, can easily discriminate between these samples. For context 3, which includes mainly AP mines with HM content and HMC, the GPR sensor is more reliable. For this context, the WEMI sensor cannot discriminate between mines and clutter objects since both have HM content.

The different contexts do not always correspond to alarms of the same type. If this is the case, the ground truth could be used to partition the training data into contexts. Context 7 is an example of one cluster that includes mine signatures from all four types. One alarm from each type is displayed in Table IV. In this case, other factors such as burial depth and soil properties can affect the signatures. For instance, for the GPR sensor, some shallowly buried AP mines can have signatures as strong as the deeply buried AT mines.

2) Learning Detector Degrees of Worthiness: After the context extraction step, the performance of each detector is evaluated within each context based on the degree of separation between the mine and clutter alarms assigned to it [using (15)]. Then, a degree of worthiness is assigned to each detector in each context using (18). These weights are shown in Fig. 8(b). For comparison purposes, we also assign a global weight to each detector using the entire training collection, i.e., we treat all data as one cluster and use (18). These weights are shown in Fig. 8(a). As it can be seen, overall, the EHD has the best performance, followed by the HMM and then the MFIT. However, the performance of the different algorithms can vary significantly from one context to another. For instance, the MFIT detector has the least weight in context 3. This is because this context includes mainly mines and clutter with HM content, and the MFIT cannot discriminate between these objects easily. On the other hand, context 8 includes mainly AP mines with LM content and clutter with HM content. From Fig. 5(b), we know that the EHD and MFIT outperform the other detectors

TABLE IV
SAMPLES OF REPRESENTATIVE MINE AND CLUTTER ALARMS FROM THREE DIFFERENT CONTEXTS



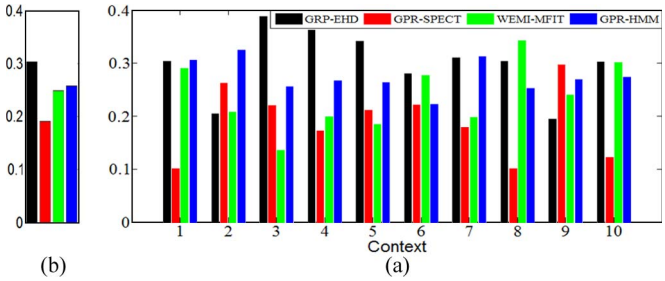


Fig. 8. Global and local degrees of worthiness of the four detectors for one cross-validation set. (a) Global weights. (b) Context-dependent weights.

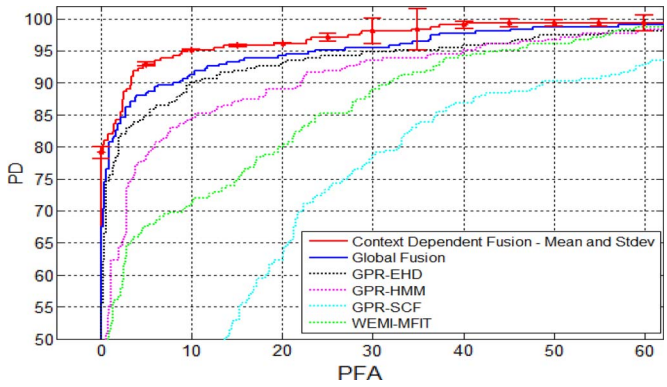


Fig. 9. Performance of the individual detectors and the global and local fusion on the entire collection using sixfold cross-validation. For the CDF, we show the mean of the ROCs of 20 runs and the standard deviation bars at selected PFA locations. For visualization purposes, the standard deviations are magnified by a factor of 30.

for AP mines. In particular, the MFIT does a better job at detecting AP mines, but the EHD is better at rejecting the HMC. Thus, a combination of these two algorithms can provide a higher PD at a lower FAR. Context 9 is another interesting one, where the SPECT detector was assigned the largest weight. This is despite the fact that, globally, the performance of this detector is not even close to the other detectors.

The aforementioned cluster-based fusion weights are intuitive and expected to be helpful as outlined in our motivation example in Section IV-A. However, here, we want to emphasize that these weights are learned from the training data without user supervision.

D. Analysis of the Testing Phase

The performance (on the testing data) of the four individual detectors using cross-validation within the TUF system is shown in Fig. 9. As it can be seen, the EHD has the best overall ROC, followed by the HMM, the MFIT, and then the SPECT. This is consistent with the performance on the training data and the global degrees of worthiness assigned to these algorithms shown in Fig. 8(a). The ROCs resulting from the global fusion and the proposed CDF are also shown in Fig. 9. For the CDF, we show the mean of the ROCs of 20 runs and the standard deviation bars at selected PFA locations. For visualization purposes, the standard deviations are magnified by a factor of 30. First, we note that, even with a simple global fusion, we obtain results that outperform all individual detectors. This is because these detectors operate on different sensor data and use different preprocessing, feature extraction, and classification algorithms. This diversity allows the fusion to take advantage of the strengths of the individual detectors,

overcome their weaknesses, and achieve a higher accuracy. Second, the proposed CDF outperforms all individual detectors and the global fusion significantly. For instance, for a 90% PD, the CDF method reduces the FAR by 63% when compared to the global fusion and by 70% when compared to the best individual detector. Similarly, at a 95% PD, the CDF method reduces the FAR by 57% when compared to the global fusion and by 69% when compared to the best individual detector. Third, we note that the standard deviations resulting from 30 random runs are very small. In fact, we had to magnify these values by a factor of 30 for the bars to be visible on the graph. This indicates that the CDF approach is not sensitive to the clustering results.

VI. Conclusion

In this paper, we have proposed a multisensor multialgorithm fusion method for land mine detection. This approach is local and adapts the fusion method to different regions of the feature space. The different regions, or contexts, are identified by partitioning the feature space in an unsupervised manner. The different contexts correspond to groups of alarm signatures that share common attributes such as mine type, geographical site, and soil and weather conditions. A degree of worthiness is assigned to each algorithm in each context based on its relative performance within the context. As a result, different expert algorithms will dominate in the classification of a new test alarm depending on the assigned context. We have shown that the proposed method can identify meaningful and coherent clusters, where different expert algorithms can be identified. Our initial experiments have also indicated that the CDF outperforms all individual detectors and the global fusion method.

Currently, we do not take into account the soil properties explicitly. We assume that, for different types of soil, the edges (extracted by EHD or HMM) and the spectral features may be different. Thus, these features could result in clusters with alarms extracted from locations with similar soil properties. Ideally, we should extract additional features (related to the statistics of an A-scan) that can capture the soil properties and not necessarily discriminate between mines and clutter. We are currently investigating such features. The proposed CDF is not restricted to specific detection algorithms. We only require the algorithms to be different to maximize the benefits of fusion. Thus, it may be beneficial to integrate other discrimination algorithms that use different raw data preprocessing and background subtraction.

Clustering is another process that needs further investigation. Currently, we use the SCAD algorithm with simple Euclidean distance for all feature sets. It may be the case that other clustering algorithms, which use different distance measures and are more robust to noise and outliers, may be more effective.

In the current implementation, a simple linear aggregation is used to assign fusion weights to the individual classifiers. This may not be the optimal way to combine the algorithms within each context. In fact, it is desirable to assign weights to subsets of classifiers to take into account the interaction between them. Fusion methods based on the fuzzy integral [74] and Dempster–Shafer theory [75], [76] have this desirable property. Similarly, one could use discriminative training methods (e.g., minimum classification error [77]) that involve optimizing the combination of the algorithms' decisions to learn a set of weights that minimize the classification error within each context. We are currently investigating these approaches.

Acknowledgment

The authors would like to thank R. Harmon, R. Weaver, P. Howard, and T. Donzelli for their support of this work; E. Rosen and L. Ayers of IDA who provided much useful software and insight; and D. Ho from the University of Missouri, L. Carin, L. Collins, and P. Torrione of Duke University, and F. Clodfelter and others from NIITEK, Inc., for their technical discussion, insights, cooperation, discrimination algorithms, and data collections.

References

[1] Landmines, Mine Action News From the United Nations, vol. 3.2, Fourth Quarter 1998.

[2] "Hidden killers: The global landmine crisis," United States Department of State Report, publication no. 10575, Sep. 1998.

[3] J. N. Wilson, P. D. Gader, W. 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.

[4] J. A. MacDonald, *Alternatives for Landmine Detection*. Santa Monica, CA: RAND Corp., 2003.

[5] S. L. Tantum, Y. Wei, V. S. Munshi, and L. M. Collins, "A comparison of algorithms for landmine detection and discrimination using ground penetrating radar," in *Proc. SPIE Conf. Detection Remediation Technol. Mines Minelike Targets*, 2002, pp. 728–735.

[6] P. D. Gader, H. Frigui, B. Nelson, G. Vaillette, and J. M. Keller, "New results in fuzzy set based detection of landmines with GPR," in *Proc. SPIE Detection Remediation Technol. Mines Minelike Targets IV*, Orlando, FL, 1999, pp. 1075–1084.

[7] P. Torrione and L. Collins, "Application of Markov random fields to landmine detection in ground penetrating radar data," in *Proc. SPIE Conf. Detection Remediation Technol. Mines Minelike Targets*, 2008, pp. 695 31B-1–695 31B-12.

[8] T. R. Witten, "Present state of the art in ground-penetrating radars for mine detection," in *Proc. SPIE Conf. Detection Remediation Technol. Mines Minelike Targets III*, Orlando, FL, 1998, pp. 576–586.

[9] I. J. Won, D. A. Keiswetter, and T. H. Bell, "Electromagnetic induction spectroscopy for clearing landmines," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 703–709, Apr. 2001.

[10] T. T. Nguyen, D. N. Hao, P. Lopez, F. Cremer, and H. Sahli, "Thermal infrared identification of buried landmines," in *Proc. SPIE Detection Remediation Technol. Mines, Minelike Targets X*, 2005, vol. 5794, pp. 198–208.

[11] S. D. Somasundaram, K. Althoefer, J. A. S. Smith, and L. D. Seneviratne, "Detection of landmines using nuclear quadrupole resonance (NQR): Signal processing to aid classification," in *Climbing and Walking Robots*, G. S. Virk, M. O. Tokhi, and M. Hossain, Eds. Berlin, Germany: Springer-Verlag, 2006, pp. 833–840.

[12] W. R. Scott, "Broadband array of electromagnetic induction sensors for detecting buried landmines," in *Proc. IGARSS*, Boston, MA, Jul. 2008, vol. 2, pp. 375–378.

[13] E. Fails, P. Torrione, W. Scott, and L. Collins, "Performance of a four parameter model for landmine signatures in frequency domain wideband electromagnetic induction detection systems," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets XII*, Apr. 2007, vol. 6553, pp. 1–8.

[14] P. D. Gader, J. N. Wilson, D. Ho, S. E. Yuksel, G. Ramachandran, and G. Heo, "Hierarchical methods for landmine detection with wideband electromagnetic induction and ground penetrating radar multi-sensor systems," in *Proc. IGARSS*, Boston, MA, Jul. 2008, pp. II-177–II-180.

[15] P. Gao, L. Collins, P. M. Garber, N. Geng, and L. Carin, "Classification of landmine-like metal targets using wideband electromagnetic induction," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1352–1361, May 2000.

[16] P. D. Gader, M. Mystkowski, and Y. Zhao, "Landmine detection with ground penetrating radar using hidden Markov models," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1231–1244, Jun. 2001.

[17] H. Frigui, K. C. Ho, and P. Gader, "Real-time land mine detection with ground penetrating radar using discriminative and adaptive hidden Markov models," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 12, pp. 1867–1885, Jan. 2005.

[18] P. Torrione and L. M. Collins, "Texture features for antitank landmine detection using ground penetrating radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2374–2382, Jul. 2007.

[19] C. Yang, "Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 743–753, May 2005.

[20] T. G. Savelyev, L. van Kempen, H. Sahli, J. Sachs, and M. Sato, "Investigation of time–frequency features for GPR landmine discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 118–129, Jan. 2007.

[21] V. Kovalenko, A. G. Yarovoy, and L. P. Ligthart, "A novel clutter suppression algorithm for landmine detection with GPR," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3740–3751, Oct. 2007.

[22] H. Frigui and P. D. Gader, "Detection and discrimination of land mines based on edge histogram descriptors and fuzzy k-nearest neighbors," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Vancouver, BC, Canada, Jul. 2006, pp. 1494–1499.

[23] D. Potin, E. Duflos, and P. Vanheeghe, "Landmines ground-penetrating radar signal enhancement by digital filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2393–2406, Sep. 2006.

[24] T. Jin and Z. Zhou, "Feature extraction and discriminator design for landmine detection on double-hump signature in ultrawideband SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3783–3791, Nov. 2008.

[25] H. T. Kaskett and J. T. Broach, "Automatic mine detection algorithm using ground penetrating radar signatures," in *Proc. SPIE Conf. Detection Remediation Technol. Mines Minelike Targets*, 1999, pp. 942–952.

[26] E. Rosen, "Investigation into the sources of persistent ground-penetrating radar false alarms: Data collection, excavation, and analysis," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets VIII*, Orlando, FL, Apr. 2003, pp. 185–190.

[27] L. A. Rastrigin and R. H. Erensterin, *Method of Collective Recognition*. Moscow, Russia: Energoizdat, 1981.

[28] R. A. Jacobs, "Methods for combining experts probability assessments," *Neural Comput.*, vol. 7, no. 5, pp. 867–888, Sep. 1995.

[29] C. Ji and S. Ma, "Combined weak classifiers," in *Advances in Neural Information Processing Systems 9*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 494–500.

[30] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, Jan. 1994.

[31] C. J. Merz and M. J. Pazzani, "Combining neural network regression estimates with regularized linear weights," in *Advances in Neural Information Processing Systems 9*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 564–570.

[32] L. Lam and C. Suen, "Optimal combination of pattern classifiers," *Pattern Recognit. Lett.*, vol. 16, no. 9, pp. 945–954, Sep. 1995.

[33] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[34] F. Kimura and M. Shridar, "Handwritten numerical recognition based on multiple algorithms," *Pattern Recognit.*, vol. 24, no. 10, pp. 969–983, 1991.

[35] M. Ceccarelli and A. Petrosino, "Multi-feature adaptive classifiers for SAR image segmentation," *Neurocomputing*, vol. 14, no. 4, pp. 345–363, Mar. 1977.

[36] P. D. Gader, M. A. Mohamed, and J. M. Keller, "Fusion of handwritten word classifiers," *Pattern Recognit. Lett.*, vol. 17, no. 6, pp. 577–584, May 1996.

[37] S. Le Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover," *Pattern Recognit.*, vol. 31, no. 11, pp. 1811–1823, Nov. 1998.

[38] M. I. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Comput.*, vol. 6, no. 2, pp. 181–214, Mar. 1994.

[39] L. I. Kuncheva, "Switching between selection and fusion in combining classifiers: An experiment," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, no. 2, pp. 146–156, Apr. 2002.

[40] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," *Pattern Recognit. Lett.*, vol. 20, no. 4, pp. 429–444, Apr. 1999.

[41] L. I. Kuncheva, "'Change-glasses' approach in pattern recognition," *Pattern Recognit. Lett.*, vol. 14, no. 8, pp. 619–623, Aug. 1993.

[42] K. Woods, W. P. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.

[43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[44] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Mach. Learn.*, vol. 37, no. 3, pp. 297–336, Dec. 1999.

[45] A. H. Gunatilaka and B. A. Baertlein, "Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 577–589, Jun. 2001.

[46] S. Perrin, E. Duflos, P. Vanheeghe, and A. Bibaut, "Multisensor fusion in the frame of evidence theory for landmines detection," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 4, pp. 485–498, Nov. 2004.

[47] 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.

[48] N. Milisavljevic and I. Bloch, "Possibilistic versus belief function fusion for antipersonnel mine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1488–1498, May 2008.

[49] R. J. Stanley, P. D. Gader, and K. C. Hoc, "Feature and decision level sensor fusion of electromagnetic induction and ground penetrating radar sensors for landmine detection with hand-held units," *Inf. Fusion*, vol. 3, no. 3, pp. 215–223, Sep. 2002.

[50] H. Frigui, P. Gader, and A. C. Benabdallah, "A generic framework for context-dependent fusion with application to landmine detection," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets*, 2008, pp. 695 31F.1–695 31F.10.

[51] H. Frigui, A. C. Benabdallah, and P. Gader, "Context-dependent fusion for landmine detection with multi-sensor systems," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets*, 2009, vol. 7303, pp. 730 328-1–730 328-10.

[52] H. Frigui, L. Zhang, P. Gader, and D. Ho, "Context-dependent fusion for landmine detection with ground-penetrating radar," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets*, 2007, p. 655 321.

[53] J. T. Miller, T. H. Bell, J. Soukup, and D. Keiswetter, "Simple phenomenological models for wideband frequency-domain electromagnetic induction," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1294–298, Jun. 2001.

[54] G. Ramachandran, P. Gader, and J. Wilson, "Gradient angle model algorithm on wideband EMI data for landmine detection," *IEEE Geosci. Remote Sens. Lett.*, 2010, to be published.

[55] T. F. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, May 1996.

[56] T. F. Coleman and Y. Li, "On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds," *Math. Program.*, vol. 67, no. 2, pp. 189–224, 1994.

[57] H. Huang and I. J. Won, "Automated anomaly picking from broadband electromagnetic data in an unexploded ordnance (UXO) survey," *Geophysics*, vol. 68, no. 6, pp. 1870–1876, 2003.

[58] K. J. Hintz, "SNR improvements in NIITEK ground-penetrating radar," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets IX*, Orlando, FL, Apr. 2004, pp. 399–408.

[59] D. Carevic, "Clutter reduction and target detection in ground-penetrating radar data using wavelets," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets IV*, A. C. Dubey, J. F. Harvey, J. T. Broach, and R. E. Dugan, Eds., 1999, vol. 3710, pp. 973–978.

[60] D. Carevic, "A Kalman filter-based approach to target detection and target–background separation in ground-penetrating radar data," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets IV*, Orlando, FL, 1999, pp. 1284–1288.

[61] A. Gunatilaka and B. A. Baertlein, "Subspace decomposition technique to improve GPR imaging of antipersonnel mines," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets V*, Orlando, FL, 2000, pp. 1008–1018.

[62] H. Brunzell, "Detection of shallowly buried objects using impulse radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 875–886, Mar. 1999.

[63] S. Yu, R. K. Mehra, and T. R. Witten, "Automatic mine detection based on ground penetrating radar," in *Proc. SPIE Conf. Detection Remediation Technol. Mines, Minelike Targets IV*, Orlando, FL, 1999, pp. 961–972.
[64] P. D. Gader, B. Nelson, H. Frigui, G. Vaillette, and J. Keller, "Fuzzy logic detection of landmines with ground penetrating radar," *Signal Process.*, vol. 80, no. 6, pp. 1069–1084, Jun. 2000.
[65] P. D. 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.
[66] H. Frigui and P. Gader, "Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic k-nearest neighbor classifier," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 1, pp. 185–199, Feb. 2009.
[67] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
[68] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG 7: Multimedia Content Description Language*. New York: Wiley, 2002.
[69] 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 Geosci. Remote Sens. Lett.*, vol. 46, no. 4, pp. 1177–1191, Apr. 2008.
[70] H. Frigui and S. Salem, "Fuzzy clustering and subset feature weighting," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2003, pp. 857–862.
[71] J. Bezdek, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Norwell, MA: Kluwer, 1999.
[72] K. Tumer and J. Ghosh, "Linear and order statistics combiners for reliable pattern classification," Ph.D. dissertation, Univ. Texas, Austin, TX, 1996.
[73] L. Ayers and E. Rosen, "MIDAS: Mine Detection Assessment and Scoring user's manual v1.1," Inst. Defense Anal., Alexandria, VA, 2004.
[74] A. Mendez-Vazquez, P. D. Gader, J. M. Keller, and K. Chamberlin, "Minimum classification error training for Choquet integrals with applications to landmine detection," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 225–238, Feb. 2008.
[75] A. P. Dempster, "A generalization of Bayesian inference," *J. R. Stat. Soc.*, vol. 30, no. 2, pp. 205–247, 1968.
[76] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[77] B. H. Juang and S. Katagiri, "Discriminative learning for minimum error classification," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 3043–3054, Dec. 1992.

Hichem Frigui (S'92–M'98) received the Ph.D. degree in computer engineering and computer science from the University of Missouri, Columbia, in 1997.

From 1998 to 2004, he was an Assistant Professor with The University of Memphis, Memphis, TN. He is currently an Associate Professor and the Director of the Multimedia Research Lab, University of Louisville, Louisville, KY. He has been active in the research fields of fuzzy pattern recognition, data mining, and image processing with applications to content-based multimedia retrieval and land mine detection. He has participated in the development, testing, and real-time implementation of several land mine detection systems. He has published over 125 journal and refereed conference articles.

Dr. Frigui has received the National Science Foundation Career Award for outstanding young scientists.

Lijun Zhang (S'07) received the M.Sc. degree from the Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai, China, in 2005 and the Ph.D. degree in computer engineering and computer science from the University of Louisville, Louisville, KY, in 2009.

He is currently a Researcher with the Center for Biomedical Imaging Statistics, Emory University, Atlanta, GA. His research interests include image processing, pattern recognition, data mining, machine learning, statistics modeling, and information fusion.

Paul D. Gader (M'87–SM'99) received the Ph.D. degree in mathematics from the University of Florida, Gainesville, in 1986.

He was a Senior Research Scientist with the Systems and Research Center, Honeywell. He was a Research Engineer and the Manager of the Environmental Research Institute of Michigan. He was a Faculty Member with the University of Wisconsin, Madison, with the University of Missouri, Columbia, and with the University of Florida, where he is currently a Professor of computer and information science and engineering. He led teams involved in real-time handwritten address recognition systems for the U.S. Postal Service and teams that devised and tested several real-time algorithms in the field for mine detection. His research interests include land mine detection, handwriting recognition, machine learning and pattern recognition, image and signal analysis, fuzzy sets, random sets, and Choquet integration. He has over 340 technical publications in the areas of image and signal processing, applied mathematics, and pattern recognition, including over 73 refereed journal articles.