Adaptive Local Fusion With Fuzzy Integrals

Ahmed Chamseddine Ben Abdallah, Hichem Frigui, *Member, IEEE*, and Paul Gader, *Fellow, IEEE*

***Abstract*—We propose a novel method for fusing different classifiers outputs. Our approach, called context extraction for local fusion with fuzzy integrals (CELF-FI), is a local approach that adapts a fuzzy integrals fusion method to different regions of the feature space. It is based on a novel objective function that combines context identification and multialgorithm fusion criteria into a joint objective function. This objective function consists of two terms: The first is designed to produce compact clusters, called contexts, and the second is designed to produce Sugeno measures for fuzzy integral fusion for each context. The terms are optimized simultaneously via alternating optimization. To test a new sample, first, its features (extracted by each algorithm) are used to assign it to each context with a fuzzy membership degree. Second, the sample confidence values (assigned by each algorithm) are fused within each context using the learned context fusion parameters. Then, the context-dependent partial confidence values are weighted by the membership degrees and averaged over all contexts to produce a final confidence value. We illustrate the performance of CELF using synthetic data, and we apply it to the problem of landmine detection using ground penetrating radar and wideband electromagnetic induction. Our extensive experiments have indicated that the proposed fusion approach outperforms all individual classifiers, the global fuzzy integral fusion method, and the basic local fusion with linear aggregation.**

***Index Terms*—Classification, classifier fusion, clustering, fuzzy integrals, local fusion.**

I. Introduction

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 problems [1]–[6] 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 advantages of the strengths of the individual classifiers, overcome their weaknesses, and achieve a higher accuracy than any individual's.

Over the past few years, a variety of schemes have been proposed for combining multiple classifiers. The most representative approaches include majority vote [7], Borda count [8], average [9], weighted average [10], Bayesian [5], and probabilistic [11]. Most of the aforementioned approaches assume that the classifier decisions are independent. For instance, the Bayesian approach requires this independence assumption in order to compute the joint probabilities. However, in practice, the outputs of multiple classifiers are usually highly correlated. Therefore, in addition to assigning fusion weights to the individual classifiers, 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 [12]–[15] and Dempster–Shafer theory [16] have this desirable property. One interesting structure using fuzzy integral is hierarchical fuzzy integral, which is also called multistep fuzzy integral [17], [18]. This structure allows the computation of the fuzzy integral from combinations of other fuzzy integrals (over smaller referential sets).

Methods for combining multiple classifiers can be classified into two main categories: global methods and local methods. Global methods assign a degree of worthiness, which is averaged over the entire training data, to each classifier. Local methods, on the other hand, 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 [19], according to which classifiers agree with each other [20], or by features of the input space [21], [22]. Then, the best classifier for each region is identified and is designated as the expert for this region [23]. Conversely, the partitioning can be defined such that each classifier is an expert in one region [24]. 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 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 [25], [26]. 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.

In [27], a local fusion approach, called context-dependent fusion (CDF), has been proposed. CDF has two main steps: First, it uses a standard clustering algorithm to partition the training signatures into groups of signatures, or contexts. Then,

Manuscript received August 13, 2010; revised March 4, 2011 and October 5, 2011; accepted December 19, 2011. Date of publication February 6, 2011; date of current version October 2, 2012. This work was supported in part by the U.S. Army Research Office under Grant W911NF-08-0255 and Grant W911NF- 07-1-0347 and by the National Science Foundation under Award CBET-0730802 and Award CBET-0730484.

A. C. Ben Abdallah is with Bloomberg LP, New York, NY 10022 USA (e-mail: ahmed.chams@gmail.com).

H. Frigui is with the University of Louisville, Louisville, KY 40208 USA (e-mail: h.frigui@louisville.edu).

P. Gader is with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: pgader@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/TFUZZ.2012.2187062

the fusion parameters are adapted to each context. CDF treats the partitioning of the feature space and the selection of local expert classifiers as two independent processes performed sequentially. However, these two tasks are not independent, and their optimization should be combined. To overcome this limitation, in [28], we proposed a generic framework for CDF, called context extraction for local fusion (CELF), which jointly optimizes the partitioning of the feature space and the fusion of the classifiers. CELF uses a simple linear aggregation 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 have this desirable property and can make the fusion of the algorithms' decisions for each context more effective.

In this paper, we generalize CELF to support fuzzy integral fusion for each context. The proposed approach, called context extraction for local fusion with fuzzy integrals (CELF-FI), partitions the feature space into different contexts and learns a Choquet integral fuser within each context. To test a new sample, first, its features (extracted by each algorithm) are used to assign it to each context with a fuzzy membership degree. Second, the sample confidence values (assigned by each algorithm) are fused within each context using the learned context fusion parameters. Then, the context-dependent partial confidence values are weighted by the membership degrees and averaged over all contexts to produce a final confidence value.

The rest of this paper is organized as follows. Fuzzy integral and CELF are described in Section II. Our fusion approach, i.e., CELF-FI, is presented in Section III. Section IV presents the application of CELF-FI to landmine detection using multisensor data. Finally, we provide the conclusions in Section V.

II. Related Work

A. Context Extraction for Local Fusion

In [28], we proposed a local fusion approach, called the CELF algorithm. CELF partitions the feature space into contexts and learns the optimal linear aggregation of the different algorithms within each context. It achieves these tasks simultaneously by optimizing a joint objective function.

Given N training observations with desired output $\mathcal{T} = \{t_j | j = 1, \ldots, N\}$ that were processed by K algorithms. Each algorithm k extracts its own feature set, i.e., $\mathcal{X}_k = \{\mathbf{x}_{kj} | j = 1, \ldots, N\}$, and generates confidence values, i.e., $\mathcal{Y}_k = \{y_{kj} | j = 1, \ldots, N\}$. The K feature sets are then concatenated to generate one global descriptor:

$$\mathcal{X} = \bigcup_{k=1}^{K} \mathcal{X}_k = \{\mathbf{x}_j = [\mathbf{x}_{1j}, \ldots, \mathbf{x}_{Kj}] | j = 1, \ldots, N\}. \quad (1)$$

CELF minimizes

$$J = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \left(\sum_{k=1}^{K} \omega_{ik} y_{kj} - t_j\right)^2 \quad (2)$$

subject to

$$\sum_{i=1}^{C} u_{ij} = 1 \ \forall j, \quad u_{ij} \in [0,1] \ \forall i, j, \quad \text{and} \quad \sum_{k=1}^{K} \omega_{ik} = 1 \ \forall i. \quad (3)$$

The first term in (2) is the sum of intracluster distances in the concatenated feature space and is the objective function of the fuzzy C-means (FCM) algorithm [29]. It seeks to partition the N samples into C clusters and represents each cluster by a center $\mathbf{c}_i$. Each sample $\mathbf{x}_j$ will be assigned to each cluster i with a membership degree u_{ij}. This term is minimized when the data are partitioned into C compact clusters with minimum sum of within-cluster distances. The second term in (2) attempts to learn cluster-dependent aggregation weights of the K algorithm outputs. In this term, w_{ik} is the aggregation weight assigned to algorithm k within cluster i, and $(\sum_{k=1}^{K} \omega_{ik} y_{kj} - t_j)^2$ is the error term resulting from aggregating the partial confidences of the sample j (i.e., $\mathbf{y}_j = [y_{1j}, \ldots, y_{Kj}]$) with the local weights of cluster i. This error term is weighted by the membership degree of $\mathbf{x}_j$ in cluster i. This term is minimized when the aggregated partial output values match the desired output. When both terms are combined and α is chosen properly, the algorithm seeks to partition the data into compact and homogeneous clusters while learning optimal aggregation weights for each algorithm within each cluster.

CELF uses a simple linear aggregation 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 that are based on the fuzzy integral [12], [30] have this desirable property.

B. Fuzzy Measure and Fuzzy Integral

The fuzzy integral has been investigated extensively for information fusion [12]–[14]. This integral defines a family of generally nonlinear aggregation operators on some function of the algorithm confidence values. The aggregation operator is defined by the fuzzy integral with respect to a nonadditive fuzzy measure. As used here, fuzzy measures are real-valued functions defined on sets of algorithms. Thus, the fuzzy integral is a mathematical construct that can be used to optimize the aggregation operator for a specific fusion application.

Definition 1 (Fuzzy Measure): Let $\mathcal{A} = \{a_1, \ldots, a_K\}$ be a finite set. A fuzzy measure g is a real-valued function defined on the power set of $\mathcal{A}$, $\mathbb{P}(\mathcal{A})$, with range [0,1], satisfying the following properties:

1) $g(\varnothing) = 0$ and $g(\mathcal{A}) = 1$.
2) Given $A, B \in \mathcal{A}$, if $A \subseteq B$, then $g(A) \leq g(B)$.

For the purpose of fusion, the set $\mathcal{A}$ is considered to contain the names of different information sources (algorithms), and for a subset $A \subseteq \mathcal{A}$, $g(A)$ is considered to be the degree of *worthiness* of this subset of information. Many fuzzy measures were introduced in the literature [31]–[35]. In this paper, we limit our study to the Sugeno measures that are a special class of fuzzy measures [36].

Definition 2 (Sugeno Measure): A fuzzy measure g is called a Sugeno measure if it satisfies the following additional property: for all $A, B \subseteq \mathcal{A}$ with $A \bigcap B = \varnothing$, there exists $\lambda > -1$ such that

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B). \tag{4}$$

It can be shown that a set function satisfying the conditions in Definition 2 is a fuzzy measure. In particular, (4) implicitly imposes the monotonicity constraints on the Sugeno measures. The value of λ can be determined for a finite set $\mathcal{A}$ using (4) and the facts that $\mathcal{A} = \cup_{i=1}^{K}\{a_k\}$ and $g(\mathcal{A}) = 1$, which leads to solving the following equation for λ:

$$1 + \lambda = \prod_{k=1}^{K}(1 + \lambda g(\{a_k\})), \text{ and } \lambda > -1. \tag{5}$$

Equation (5) is a polynomial in λ of degree $K - 1$ and can be easily solved numerically [31], [32].

The discrete Choquet integral [37] has proven to be a useful tool to fuse evidence supplied by different information sources.

Definition 3 (Choquet Integral): Let $e : \mathcal{A} \rightarrow [0, 1]$. Let $\{a_{\sigma(1)}, \ldots, a_{\sigma(K)}\}$ denote the reordering of the set $\mathcal{A}$ such that $e(a_{\sigma(1)}) \leq \ldots \leq e(a_{\sigma(K)})$, and let A_k be a collection of subsets defined by $A_k = \{a_{\sigma(k)}, \ldots, a_{\sigma(K)}\}$. The discrete Choquet integral of e with respect to g on $\mathcal{A}$ is defined as

$$C_g(e) = \sum_{k=1}^{K}\left[e(a_{\sigma(k)}) - e(a_{\sigma(k-1)})\right] \cdot g(A_k) \tag{6}$$

or

$$C_g(e) = \sum_{k=1}^{K}[g(A_k) - g(A_{k+1})] \cdot e(a_{\sigma(k)}) \tag{7}$$

where $e(a_{\sigma(0)}) = 0$, and $A_{k+1} \equiv \varnothing$.

The function e is a particular instance of the partial support (evidence) supplied by each information source in determining the confidence in an underlying hypothesis. The integral fuses this objective support with the degree of *worthiness* of the various subsets of the information sources.

The analysis of the coefficients of the fuzzy measure can be performed by the calculation of the Shapley values [38].

Definition 4 (Shapley Value): The Shapley value of g is an K-dimensional vector $\phi_g = [\phi_g(a_1), \ldots, \phi_g(a_K)]$, which is defined by

$$\phi_g(a_k) = \sum_{A \subseteq \mathcal{A}\setminus\{a_k\}} \gamma_{\mathcal{A}}(A)(g(A \cup \{a_k\}) - g(A)) \tag{8}$$

with

$$\gamma_{\mathcal{A}}(A) = \frac{(|\mathcal{A}| - |A| - 1)! \times |A|!}{|\mathcal{A}|!} \tag{9}$$

where $|A|$ indicates the cardinality of A.

The Shapley value, i.e., $\phi_g(a_k)$, with respect to a fuzzy measure g, represents the global importance of each source a_k with respect to any subset A not containing a_k. It is confined to the interval [0,1]. A value close to zero indicates that the kth algorithm is not relevant for the given data, while a value close to 1 indicates that the given algorithm is highly relevant for the given data. It can be proven that $\sum_{k=1}^{K}\phi_g(a_k) = 1$.

Another way to analyze the values of the fuzzy measure is to compute the interaction index $I_g(a_k, a_l)$ [39], [40] between pairs of information sources.

Definition 5 (Interaction Index): The mean interaction index between two sources k and l with respect to g is defined by

$$I_g(a_k, a_l) = \sum_{A \subseteq \mathcal{A}\setminus\{a_k, a_l\}} \xi_{\mathcal{A}}(A)(g(A \cup \{a_k, a_l\}) - g(A \cup \{a_k\}) - g(A \cup \{a_l\}) + g(A)) \tag{10}$$

with

$$\xi_{\mathcal{A}}(A) = \frac{(|\mathcal{A}| - |A| - 2)! \times |A|!}{(|\mathcal{A}| - 1)!}. \tag{11}$$

A positive value of the interaction index ($I_g(a_k, a_l) > 0$) induces a *conjunctive* behavior in aggregation. That is, algorithms k and l have to be both satisfied in order to have a good global score. On the other hand, a negative value of the interaction index ($I_g(a_k, a_l) < 0$) induces a *disjunctive* behavior in aggregation. That is, it suffices to satisfy one of the two algorithms, i.e., k or l, to have a good global score. A null value of the interaction index ($I_g(a_k, a_l) = 0$) induces no interaction. In this case, a linear aggregation is sufficient to have a good global score.

III. Local Fusion With Fuzzy Integrals

In the following, we propose generalizing CELF by replacing the linear fusion component with the fuzzy integral. As in CELF, we assume that we have N training observations with desired outputs $\mathcal{T} = \{t_j | j = 1, \ldots, N\}$ that were processed by K algorithms. Each algorithm k extracts its own feature set, i.e., $\mathcal{X}_k = \{\mathbf{x}_{kj} | j = 1, \ldots, N\}$, and generates confidence values, i.e., $\mathcal{Y}_k = \{y_{kj} | j = 1, \ldots, N\}$. The K feature sets are then concatenated to generate one global descriptor as in (1).

Fig. 1 displays the architecture of the proposed CELF-FI scheme. This figure highlights the two main components of the training phase, namely, *context extraction* and *decision fusion*. The context extraction step uses both the features extracted by various algorithms (indicated by solid lines in the figure) and their confidences (indicated by dotted lines) to partition the training input samples into different contexts. The decision fusion step uses the confidence values assigned by the individual algorithms (indicated by dotted lines in the figure) to adapt a fuzzy integral for each context. To test a new alarm, each algorithm extracts its set of features and assigns a confidence value. Then, as shown in the right part of Fig. 1, the features are used to assign the test sample to the closest context. The fuzzy measure of this context is then used to fuse the individual confidence values.

CELF-FI partitions the feature space and learns the fuzzy measures simultaneously by optimizing the following objective

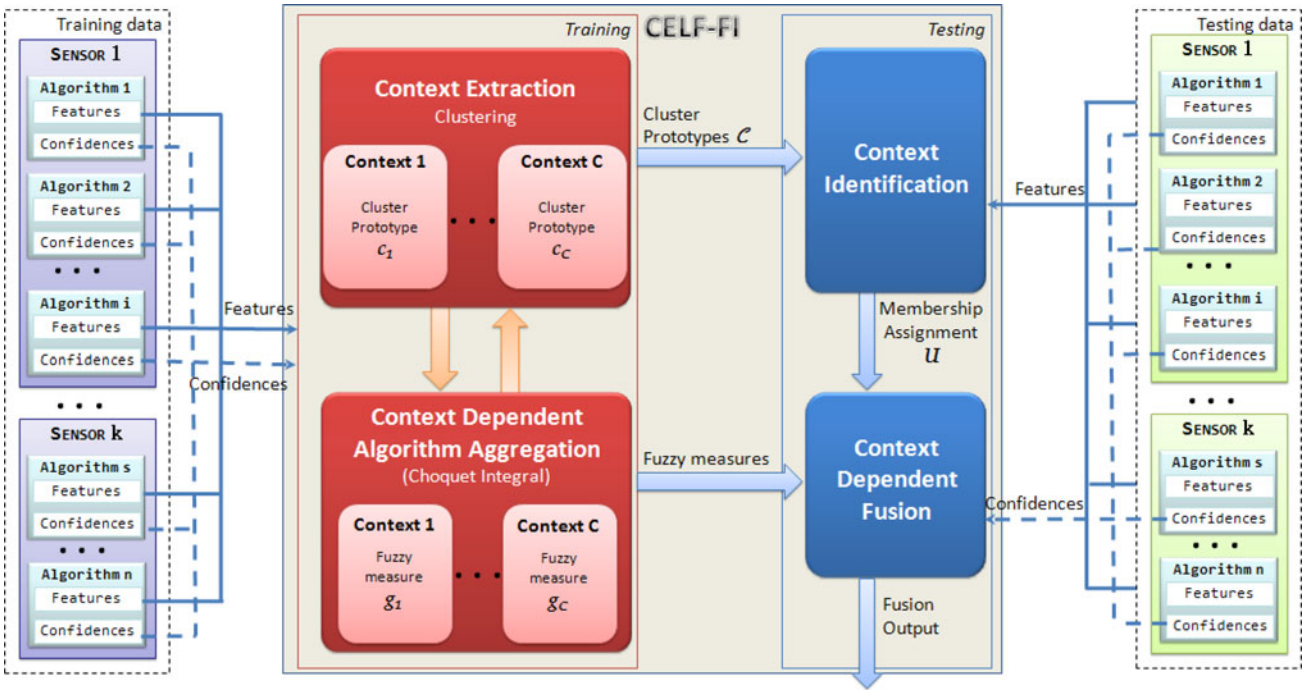

Fig. 1. Architecture of CELF-FI.

function:

$$J_{\mathrm{FI}} = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^{m}\|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^{m}\left(C_{g_i}(\hat{\mathbf{y}}_j) - t_j\right)^2 \tag{12}$$

subject to

$$\sum_{i=1}^{C} u_{ij} = 1 \quad \forall j, \quad \text{and} \quad u_{ij} \in [0,1] \quad \forall i,j. \tag{13}$$

In (12), $\hat{\mathbf{y}}_j = [\hat{y}_{1j}, \ldots, \hat{y}_{Kj}]$ is the set of confidence values assigned by the K algorithms to sample j sorted in ascending order, and g_i is the Sugeno measure associated with cluster i. For each g_i, we associate a coefficient λ_i that satisfies (4) and (5). The objective function in (12) can be rewritten as

$$J_{\mathrm{FI}} = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^{m}\|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^{m} \left(\sum_{k=1}^{K}\left[\hat{y}_{kj} - \hat{y}_{(k-1)j}\right] \times g_i(A_k) - t_j\right)^2 \tag{14}$$

where $A_k = \{k, \ldots, K\}$, and $\hat{y}_{0j} = 0$.

The first term in (12) is the unsupervised learning component. It is the sum of intracluster distances and is the objective function used in the FCM algorithm [29]. It seeks to partition the N samples into C clusters and represents each cluster by a center $\mathbf{c}_i$. Each sample j will be assigned to each cluster i with a membership degree u_{ij}. In this term, $m \in (1, \infty)$ is a constant called *fuzzifier* and is used to control the degree of fuzziness [29]. The second term in (12) is the supervised learning component. It attempts to adapt a fuzzy integral fusion to each cluster. In this term, g_i is the fuzzy measure within cluster i. This term is minimized when the fusion output values match the desired output. When both terms are combined and α is chosen properly, the algorithm seeks to partition the data into compact and homogeneous clusters, while learning optimal aggregation weights for each algorithm within each cluster.

To optimize J_{FI} with respect to $\mathbf{U} = [u_{ij}]$. We incorporate the constraints with the aid of Lagrange multipliers. We obtain

$$L^u = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^{m}\|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^{m}\left(C_{g_i}(\hat{\mathbf{y}}_j) - t_j\right)^2 + \sum_{j=1}^{N} \lambda_j \left(\sum_{i=1}^{C} u_{ij} - 1\right) \tag{15}$$

where $\Lambda = [\lambda_1, \ldots, \lambda_N]^t$ is a vector of Lagrange multipliers corresponding to the N constraints on $\mathbf{U}$ in (3). Since the memberships of the different observations are independent of each other, the aforementioned optimization problem can be reduced to N simpler independent problems. For each pattern $j = 1, 2, \ldots, N$, we formulate the augmented functional

$$L_j^u = \sum_{i=1}^{C} u_{ij}^{m}\|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{i=1}^{C} u_{ij}^{m}\left(C_{g_i}(\hat{\mathbf{y}}_j) - t_j\right)^2 + \lambda_j \left(\sum_{i=1}^{C} u_{ij} - 1\right). \tag{16}$$

Computing the derivative of L_j^u with respect to u_{ij} and setting it to 0, we obtain

$$u_{ij} = \frac{1}{\sum_{l=1}^{C}(D_{ij}/D_{lj})^{\frac{1}{m-1}}} \tag{17}$$

where

$$D_{ij} = \|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \left(C_{g_i}(\hat{\mathbf{y}}_j) - t_j\right)^2 \tag{18}$$

can be viewed as the total cost when considering point $\mathbf{x}_j$ in cluster i. This cost depends on 1) the distance between the considered point and the cluster's centroid $\mathbf{c}_i$ [first term in (18)]; and 2) the deviation of the combined algorithms' decision from the desired output [second term in (18)]. In other words, in (17), a sample j will be assigned high membership degree in cluster i if 1) sample j is close to cluster i in the feature space, i.e., small $\|\mathbf{x}_j - \mathbf{c}_i\|^2$, and 2) the confidence values that are assigned to sample j could be combined with the same Sugeno measure to match the desired output as the other samples in cluster i, i.e., small $(C_{g_i}(\hat{\mathbf{y}}_j) - t_j)^2$.

The performance of CELF-FI depends on the chosen value of α. In fact, when α is too small, the multialgorithm fusion criteria [second term in (12)] get neglected during the clustering process. As a result, the optimal solution may not be reached. On the other hand, when α is too big, the clusters may not share many common features (not similar in feature space), and our concept of context becomes not well defined. Moreover, during testing, the context identification step becomes almost random.

To minimize J_{FI} with respect to the centers $\mathbf{c}_{ik}$, we fix $\mathbf{U} = [u_{ij}]$, $\mathbf{G} = [g_{ik}]$, set the gradient to zero, and obtain

$$\mathbf{c}_i = \frac{\sum_{j=1}^{N} u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^{N} u_{ij}^m}. \tag{19}$$

Differentiation of (12) with respect to the Sugeno measure g_i does not have a closed-form solution. Thus, we use a gradient descent approach and update it in every iteration. As a convention, the measure of a singleton set $\{l\}$ is called a density and is denoted by $g_{il} = g_i(\{l\})$. Given a learning rate η, the density g_{il} is updated using

$$g_{il} = g_{il} - \eta \frac{\partial J_{\mathrm{FI}}}{\partial g_{il}} \tag{20}$$

where

$$\frac{\partial J_{\mathrm{FI}}}{\partial g_{il}} = 2\alpha \sum_{j=1}^{N} u_{ij}^m \left(\sum_{k=1}^{K} \left[\hat{y}_{kj} - \hat{y}_{(k-1)j}\right] \cdot g_i(A_k) - tj \right) \left(\sum_{k=1}^{K} \left[\hat{y}_{kj} - \hat{y}_{(k-1)j}\right] \cdot \frac{\partial g_i(A_k)}{\partial g_{il}} \right). \tag{21}$$

Note that, in (21), given a sample j and cluster i, the weight adjustment depends on the membership of the sample to the correspondent cluster. In fact, if the sample j is typical of cluster i, its membership u_{ij} is close to 1. In this case, g_i is adjusted to minimize the error J_{FI}. On the other hand, if the sample j is not typical of cluster i, its membership u_{ij} is close to 0. Therefore, g_i is not adjusted, even if the fuser misclassifies this sample.

To calculate the partial derivative $\partial g_i(A_k)/\partial g_{il}$, we use the same approach used in [41] and obtain the following cases.

Algorithm 1 CELF with Fuzzy Integrals (CELF-FI)

Inputs: $\mathcal{X}$: the features of the training data samples.
$\mathcal{Y}$: the confidences given by the different classifiers.
$\mathcal{T}$: the labels of the data samples.
C: the number of clusters.
m: the fuzzifier, $m \in (1, +\infty)$.
α, the weight of the second term in the objective function.
η, the learning rate.
Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.
$\mathcal{C}$: the cluster centers.
$\mathbf{G} = [g_{ik}]$: Sugeno measures.

1: Initialize $\mathbf{U}$ and $\mathbf{G}$.
2: **repeat**
3: Update $\mathcal{C}$ using (19).
4: Update $\mathbf{U}$ using (17).
5: Update $\mathbf{G}$ using (20) for few iterations.
6: **until** parameters do not change significantly
7: **return** $\mathcal{C}$, $\mathbf{U}$, and $\mathbf{G}$

Case 1: $k \neq K$ and $k = l$

$$\frac{\partial g_i(A_k)}{\partial g_{il}} = 1 + \lambda_i g_i(A_{k+1}) + g_{ik} g_i(A_{k+1}) \frac{\partial \lambda_i}{\partial g_{il}} + (1 + \lambda_i g_{ik}) \frac{\partial g_i(A_{k+1})}{\partial g_{il}}. \tag{22}$$

Case 2: $k \neq K$ and $k \neq l$

$$\frac{\partial g_i(A_k)}{\partial g_{il}} = g_{ik} g_i(A_{k+1}) \frac{\partial \lambda_i}{\partial g_{il}} + (1 + \lambda_i g_{ik}) \frac{\partial g_i(A_{k+1})}{\partial g_{il}}. \tag{23}$$

Case 3: $k = K$ and $k = l$

$$\frac{\partial g_i(A_k)}{\partial g_{il}} = 1. \tag{24}$$

Case 4: $k = K$ and $k \neq l$

$$\frac{\partial g_i(A_k)}{\partial g_{il}} = 0. \tag{25}$$

In (22) and (23)

$$\frac{\partial \lambda_i}{\partial g_{il}} = \begin{cases} \dfrac{\lambda_i^2 + \lambda_i}{(1 + \lambda_i g_{il}) \left(1 - (\lambda_i + 1) \sum_{k=1}^{K} \frac{g_{ik}}{1 + \lambda_i g_{ik}}\right)}, & \text{if } \lambda \neq 0 \\ K, & \text{if } \lambda = 0. \end{cases} \tag{26}$$

The resulting algorithm is summarized in Algorithm 1.

Testing phase

We should note here that using (17) to compute the membership of a given sample j to cluster i, its desired output t_j needs to be given. This information is available and is necessary to build the clusters in the training phase. However, during testing, for an unlabeled test sample, it is not possible to assign it to a cluster using (17). Instead, we identify the nearest training sample and use its label. To summarize, given an unlabeled test

sample j, we apply the following steps to generate the final fusion decision.

1) Process the sample by the different algorithms to generate a set of features, i.e., $\mathbf{x}_j$, and decision values, i.e., $\mathbf{y}_j = [y_{1j}, \ldots, y_{Kj}]$.
2) Identify the nearest training sample and use its label to assign a temporary label to the test sample.
3) Assign a membership degree to sample j in each cluster i, u_{ij}, using (18).
4) Combine the output of the different classifiers within each cluster using

$$o_{ij} = C_{g_i}(\hat{\mathbf{y}}_j). \tag{27}$$

5) Generate the final fusion decision confidence using

$$o_j = \sum_{i=1}^{C} u_{ij} o_{ij}. \tag{28}$$

A. Finding the Optimal Number of Clusters

CELF-FI requires the specification of the number of clusters. However, in most applications, this information may not be known *a priori*. To address this issue, we propose extending (12) to integrate a regularization term as in the competitive agglomeration clustering [42]. The resulting algorithm starts by partitioning the data into a large number of small clusters. As the algorithm progresses, adjacent clusters compete for data points, and clusters that lose the competition gradually become depleted and vanish. Thus, as the iterations proceed, we obtain a sequence of partitions with a progressively diminishing number of clusters. The final partition is taken to have the "optimal" number of clusters.

The proposed algorithm minimizes

$$J_C = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^2 \|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^2 \left(C_{g_i}(\hat{\mathbf{y}}_j) - t_j\right)^2 - \gamma \sum_{i=1}^{C}\left(\sum_{j=1}^{N} u_{ij}\right)^2 \tag{29}$$

subject to the constraints in (13). We should note here that in (29), γ is a constant, and the number of clusters C is not fixed.

The objective function in (29) has three components. The first two components, which are similar to those in the CELF-FI objective function (with $m = 2$), combine clustering and multialgorithm fusion. The global minimum of this component is achieved when the number of clusters C is equal to the number of samples N, i.e., each cluster contains a single data point. The last component in (29) is the sum of squares of the cardinalities of the clusters that allows us to control the number of clusters. The global minimum of this term (including the negative sign) is achieved when all points are lumped in one cluster, and all other clusters are empty. When both components are combined and γ is chosen properly, the final partition will minimize the sum of intracluster distances, while partitioning the dataset into the smallest number of clusters possible. The clusters that are depleted as the algorithm proceeds will be discarded, as explained later.

To optimize J_C with respect to $\mathbf{U} = [u_{ij}]$, we incorporate the constraints with the aid of Lagrange multipliers. We obtain

$$L = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^2 \|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^2 \left(C_{g_i}(\hat{\mathbf{y}}_j) - t_j\right)^2 - \gamma \sum_{i=1}^{C}\left(\sum_{j=1}^{N} u_{ij}\right)^2 - \sum_{j=1}^{N} \lambda_j \Big(\sum_{i=1}^{C} u_{ij} - 1\Big) \tag{30}$$

where $\Lambda = [\lambda_1, \ldots, \lambda_N]^t$ is a vector of Lagrange multipliers corresponding to the N constraints on $\mathbf{U}$ in (13). Computing the derivative of L with respect to u_{ij} and setting it to 0, we obtain

$$u_{ij} = u_{ij}^{\text{CELF-FI}} + u_{ij}^{\text{Bias}} \tag{31}$$

where

$$u_{ij}^{\text{CELF-FI}} = \left[\sum_{l=1}^{C} (D_{ij}/D_{lj})\right]^{-1} \tag{32}$$

and

$$u_{ij}^{\text{Bias}} = \frac{\gamma}{D_{ij}}(N_i - \widetilde{N}_j). \tag{33}$$

In (32) and (33), D_{ij} is the same as defined in (18):

$$N_i = \sum_{j=1}^{N} u_{ij} \tag{34}$$

is the fuzzy cardinality of cluster i, and

$$\widetilde{N}_j = \left[\sum_{l=1}^{C} \frac{N_l}{D_{lj}}\right]\left[\sum_{l=1}^{C} \frac{1}{D_{lj}}\right] \tag{35}$$

is simply a weighted average of the cluster cardinalities, where the weight of each cluster reflects its proximity to the feature point $\mathbf{x}_j$ in question.

The first component of u_{ij} in (31), i.e., $u_{ij}^{\text{CELF-FI}}$, is the membership term in the CELF-FI algorithm [see (17)]. The second component, i.e., u_{ij}^{Bias}, is a signed bias term that depends on the difference between the cardinality of the cluster of interest and the weighted average of cardinalities with respect to feature point $\mathbf{x_j}$. For clusters with cardinality higher than average, the bias term is positive, thus appreciating the membership value. On the other hand, for low cardinality clusters, the bias term is negative, thus depreciating the membership value. Moreover, this bias term is also inversely proportional to the distance of feature point $\mathbf{x_j}$ to the cluster of interest $\mathbf{c_i}$, which serves as an amplification factor. This leads to a gradual reduction of the cardinality of spurious clusters. When the cardinality of a cluster drops below a threshold, we discard the cluster and update the number of clusters. Since the initial partition has an overspecified number of clusters, each cluster is approximated by many small clusters in the beginning. As the algorithm proceeds, adjacent clusters compete between each other. As a result, only a

Algorithm 2 CELF-FI with C optimization

Inputs: $\mathcal{X}$: the features of the data samples.
$\mathcal{Y}$: the confidences given by the different classifiers.
$\mathcal{T}$: the labels of the data samples.
C_{max}: the maximum number of clusters.
m: the fuzzifier, $m \in (1, +\infty)$.
α: the weight of the second term in the objective function.
γ: the weight of the third term in the objective function.
ϵ: a given threshold.
η, the learning rate.
Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.
$\mathcal{C}$: the cluster centers.
$\mathbf{G} = [g_{ik}]$: Sugeno measures.

```
Fix the maximum number of clusters C = C_max;
Initialize U and G.
Compute the initial cardinalities N_i for 1 ≤ i ≤ C using (34);
repeat
    Update the partition matrix U using (31);
    Compute the cardinalities N_i for 1 ≤ i ≤ C using (34);
    if N_i < ε then
        discard cluster i;
    end if
    Update the number of clusters C;
    Update C using (19).
    Update U using (31).
    Update G using (20) for few iterations.
until parameters do not change significantly
return C, U, and G
```

few clusters will survive, while others will shrink and eventually become extinct.

To optimize J_C with respect to $\mathcal{C}$ and $\mathbf{G}$, we use (19) and (20), respectively, as in CELF-FI. The resulting algorithm is summarized in Algorithm 2.

B. Context Extraction for Local Fusion With Fuzzy Integrals for Multiclass Data

CELF-FI was designed and developed to support two-class data. In the following, we propose generalizing the algorithm to cover data with multiple classes (CELF-M). Given S classes, we assume that we have N training observations with desired outputs $\mathcal{T} = \{\mathbf{t}_j = [t_{j1}, \ldots, t_{jS}] | j = 1, \ldots, N\}$. t_{js} is equal to 1 if the sample j is from the sth class and 0 otherwise. These samples are processed by K algorithms. Each algorithm k extracts its own feature set, i.e., $\mathcal{X}_k = \{\mathbf{x}_{kj} | j = 1, \ldots, N\}$, and generates confidence values, i.e., $\mathcal{Y}_k = \{\mathbf{y}_{kj} = [y_{k1j}, \ldots, y_{kSj}] | j = 1, \ldots, N\}$, where y_{ksj} is the confidence assigned by classifier k to input j to be in the sth class. The K feature sets are then concatenated to generate one global descriptor, as in (1).

CELF-M minimizes

$$J_M = \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \times \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \sum_{s=1}^{S} \left(C_{g_i}(\hat{\mathbf{y}}_{sj}) - t_{js}\right)^2$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \times \sum_{s=1}^{S}\left(\sum_{k=1}^{K}\left[\hat{y}_{ksj} - \hat{y}_{(k-1)sj}\right] \cdot g_i(A_k) - t_{js}\right)^2 \quad (36)$$

subject to the constraints in (13). In (36), $\hat{\mathbf{y}}_{sj} = [\hat{y}_{1sj}, \ldots, \hat{y}_{Ksj}]$ is the set of confidence values assigned by the K algorithms to sample j to be in the sth class sorted in ascending order.

Optimizing J_M with respect to $\mathbf{U} = [u_{ij}]$ leads to

$$u_{ij} = \frac{1}{\sum_{l=1}^{C} (D_{ij}/D_{lj})^{\frac{1}{m-1}}} \quad (37)$$

where

$$D_{ij} = \|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha \sum_{s=1}^{S} \left(C_{g_i}(\hat{\mathbf{y}}_{sj}) - t_{js}\right)^2 . \quad (38)$$

Optimizing J_M with respect to the Sugeno measure g_i leads to

$$g_{il} = g_{il} - \eta \frac{\partial J_M}{\partial g_{il}} \quad (39)$$

where

$$\frac{\partial J_M}{\partial g_{il}} = 2\alpha \sum_{j=1}^{N} u_{ij}^m \sum_{s=1}^{S}\left(\sum_{k=1}^{K}[\hat{y}_{ksj} - \hat{y}_{(k-1)sj}] \cdot g_i(A_k) - t_{js}\right) \times \left(\sum_{k=1}^{K}\left[\hat{y}_{ksj} - \hat{y}_{(k-1)sj}\right] \cdot \frac{\partial g_i(A_k)}{\partial g_{il}}\right) . \quad (40)$$

Optimizing J_M with respect to the cluster centers leads to the same update equation as CELF-FI. The resulting algorithm, i.e., CELF-M, is similar to CELF-FI, except that $\mathbf{U}$ and $\mathbf{G}$ are updated using (37) and (39), respectively.

IV. Experimental Results

A. Analysis of Context Extraction for Local Fusion With Fuzzy Integrals Using Synthetic Data

To illustrate the behavior of the proposed fusion approach, we first use it to partition and fuse a simple synthetic data. This dataset is designed to illustrate the need for local fusion. It consists of 2000 samples that belong to two classes: 1000 samples from `class 0` (*negative*) and 1000 samples from `class 1` (*positive*). Suppose that each sample has been processed by three different algorithms. Each algorithm k extracts one feature (x_k) and assigns one output value (y_k). Fig. 2 displays this data in the 2-D feature space (x_1, x_2) where samples from `class 0` are represented by red dots and samples from `class 1` are represented by green dots. As can be seen, the data form two distinct clusters in the feature space, and each cluster has samples from both classes.

In Fig. 3, we display the classification results of the three individual classifiers. As can be seen, none of the three classifiers classify these data perfectly as all three figures include many misclassified samples. In fact, the accuracy of classifier 1 is 63.2%, of classifier 2 is 74.9%, and of classifier 3 is 61.7%.

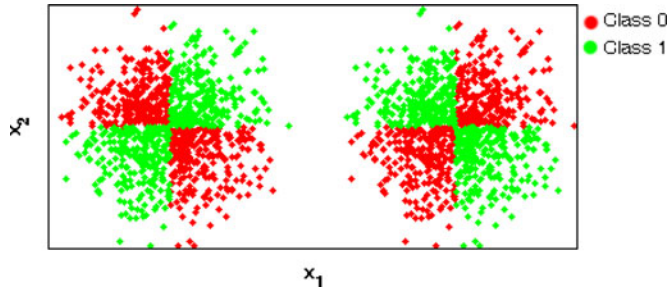

Fig. 2. Synthetic data in the 2-D feature space. Class 0 samples are shown as red dots, and class 1 samples are shown as green dots.

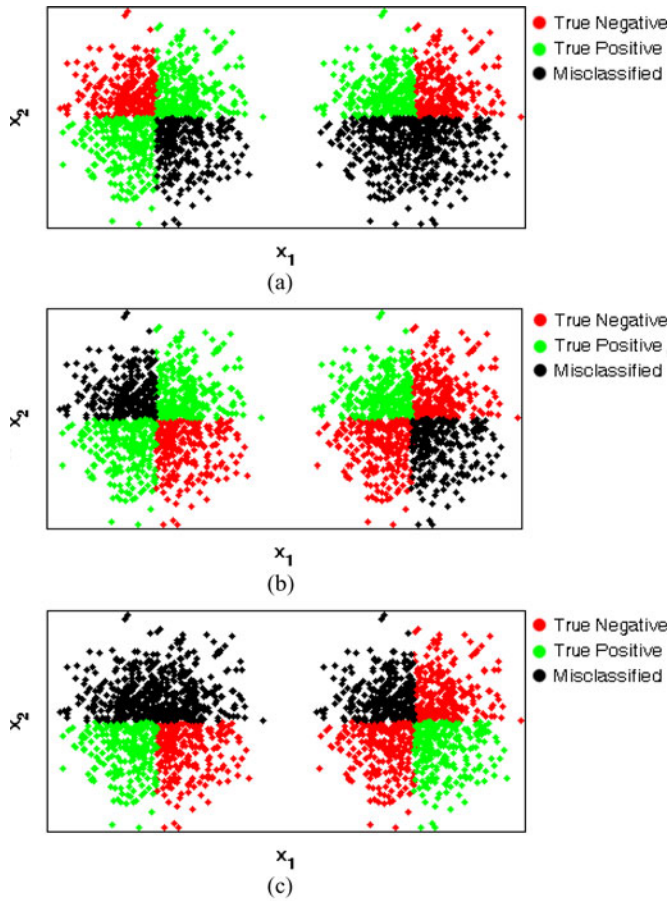

Fig. 3. Classification result of (a) the first classifier, (b) the second classifier, and (c) the third classifier.

More importantly, the performance of each classifier varies in different regions of the feature space. For instance, in the left cluster, classifier 1 has an accuracy of 75.4%, and classifier 3 has an accuracy of 49%. On the other hand, for the right cluster, the accuracy of classifier 1 is 51%, and of classifier 3 is 74.4%. This simple example illustrates the need for local fusion to take advantages of the strengths of the classifiers in different regions of the feature space.

We compare the results of CELF-FI with global fuzzy integral fusion. Fig. 4(a) displays the cumulative histogram of the confidences assigned by the global fusion algorithm. As can be seen, the fusion cannot achieve perfect classification as the distribution of the two classes overlaps. In fact, for a threshold of 0.5, the accuracy of the global fusion is 74.9% which is not any better than the best individual classifier. These results, shown in 4(b), are similar to those obtained by classifier 2 only.

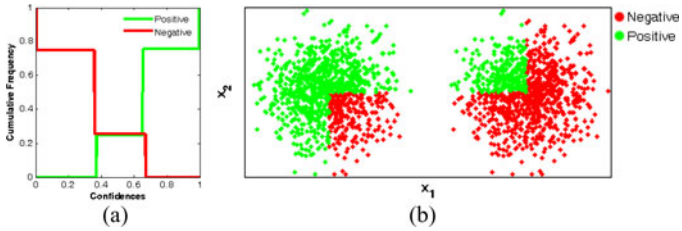

Fig. 4. Fusion results using a global fuzzy integral approach. (a) Cumulative histograms of the confidences assigned by the global fuzzy integral fusion. (b) Assigned label when the threshold is fixed to 0.5.

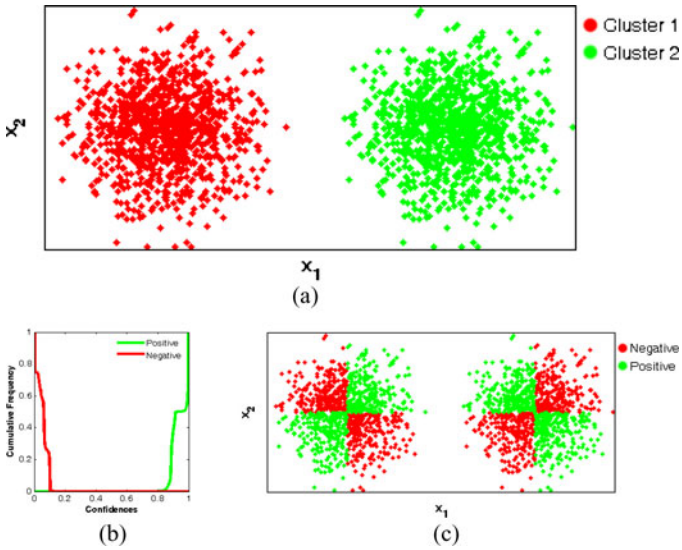

Fig. 5. Local fusion results using CELF-FI. (a) Clustered samples in the feature space. A different color is used for each cluster. (b) Cumulative histogram of the confidences assigned by the fusion algorithm. (c) Fusion results (using 0.5 as threshold).

Fig. 5(a) illustrates the clustering result using CELF-FI when the number of clusters $C = 2$, $m = 2$, $\alpha = 10$, and the learning rate η set to 0.1 (same as the one used in the global fuzzy integral). As can be seen, CELF-FI identifies the two clusters. Fig. 5(b) displays the histograms of the confidences generated by CELF-FI where the two distributions are separable and any threshold in the $[0.3, 0.7]$ range would result in an accuracy of 100%. The classification results, using a 0.5 threshold, are shown in Fig. 5(c).

In order to gain further insight into the behavior of the local and the global approaches, in Fig. 6, we display the Shapley index of each classifier, and in Fig. 7, we display the interaction indexes between each pair of classifiers assigned by the global fuzzy integral and CELF-FI (within each cluster).

As shown in Fig. 6, the global fusion assigned roughly the same Shapley score to each algorithm. This means that all the classifiers contribute to the fusion with approximately the same weight. This is expected since the three classifiers have comparable overall performance. The proposed CELF-FI, on the other

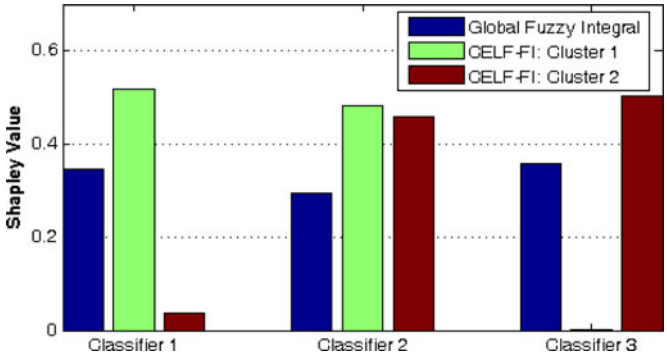

Fig. 6. Shapley values of the different classifiers assigned by the global fusion and the local fusion (within each cluster).

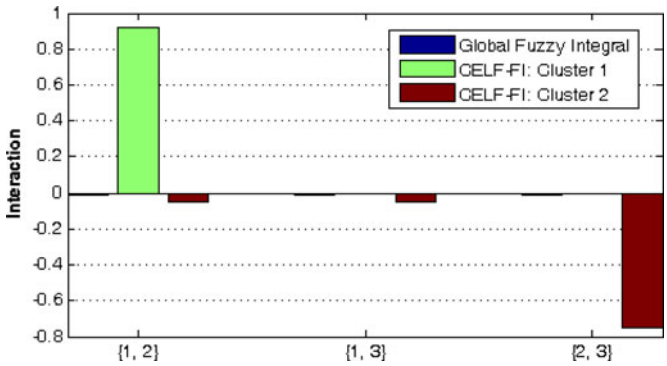

Fig. 7. Interactions indexes of the different pair of classifiers assigned by the global fusion and the local fusion (within each cluster).

hand, assigns cluster-dependent values to the Shapley indexes. In particular, for cluster 1, a high index was assigned to the first two classifiers and a low index was assigned to the third classifier. However, for cluster 2, CELF-FI learned a high index for the last two classifiers, and a low index for the first one. In fact, in order to obtain better fusion results, CELF-FI discarded the third classifier in cluster 1 and discarded the first one in cluster 2. Referring to Fig. 3, we can see that the third classifier has the worst accuracy (49%) in cluster 1, and the first classifier has the worst accuracy (51%) in cluster 2. This justifies the learned cluster-dependent fusion parameters.

Fig. 7 shows that global fusion with the fuzzy integral assigns null interaction indexes to all pairs of classifiers. Thus, a linear aggregation was used to fuse the three classifiers. On the other hand, the local CELF-FI assigns a positive interaction index to (classifier 1, classifier 2) within cluster 1 and a negative interaction index to (classifier 2, classifier 3) within cluster 2. In fact, as shown in Fig. 3, within cluster 1, classifiers 1 and 2 have to be both satisfied in order to detect samples from Class 1. However, within cluster 2, it is sufficient to satisfy classifier 2 or 3.

B. Analysis of CELF-M Using Synthetic Data

To illustrate the behavior of CELF-M, we first use it to partition and fuse a simple synthetic data. This dataset consists of 2000 samples that belong to three classes. Suppose that each sample has been processed by three different algorithms. Each

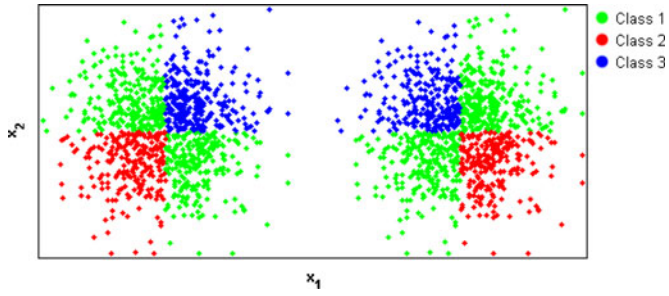

Fig. 8. Synthetic data in the 2-D feature space.

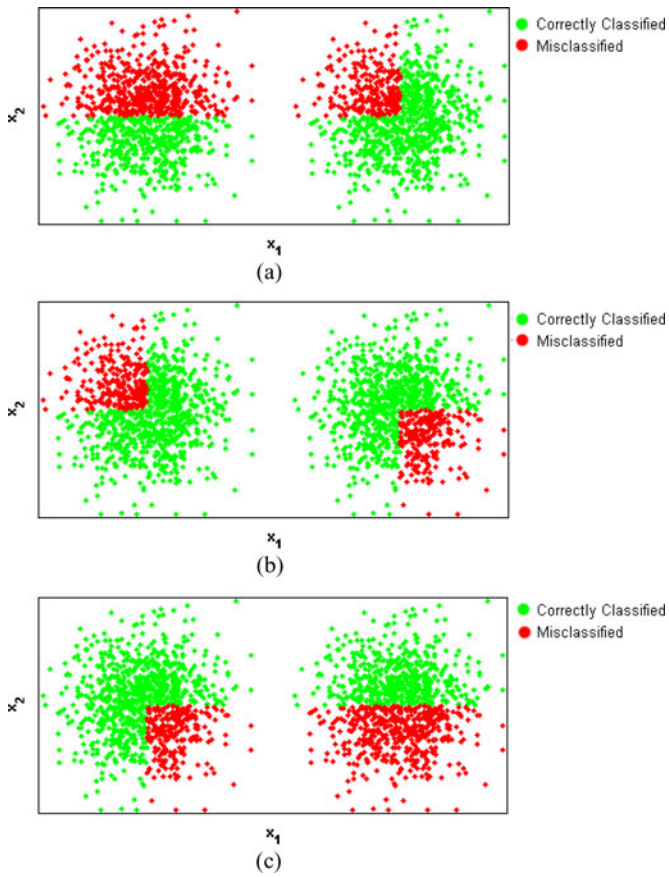

Fig. 9. Classification result of (a) classifier 1, (b) classifier 2, and (c) classifier 3.

algorithm k extracts one feature (x_k) and assigns three output value ($[y_{k1}, y_{k2}, y_{k3}]$). Fig. 8 displays these data in the 2-D feature space (x_1, x_2), where samples from `class 1` are represented by green dots, samples from `class 2` are represented by red dots, and samples from `class 3` are represented by blue dots.

In Fig. 9, we display the classification results of the three classifiers. The accuracy of classifier 1 is 62.5%, of classifier 2 is 75%, and of classifier 3 is 62.5%.

Fig. 10 displays the classification results with global fusion using the same fuzzy integral approach. The accuracy of this fusion is 75% which is no better than the best individual classifier. In fact, the results are very similar to those obtained by classifier 2 only.

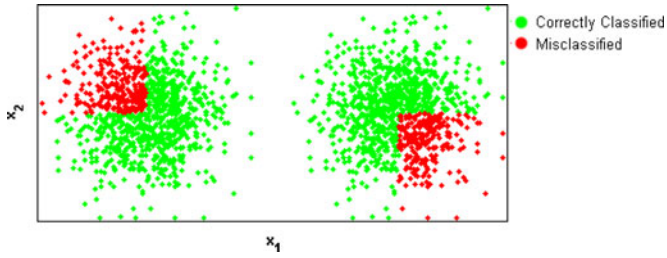

Fig. 10. Fusion results using a global fuzzy integral approach.

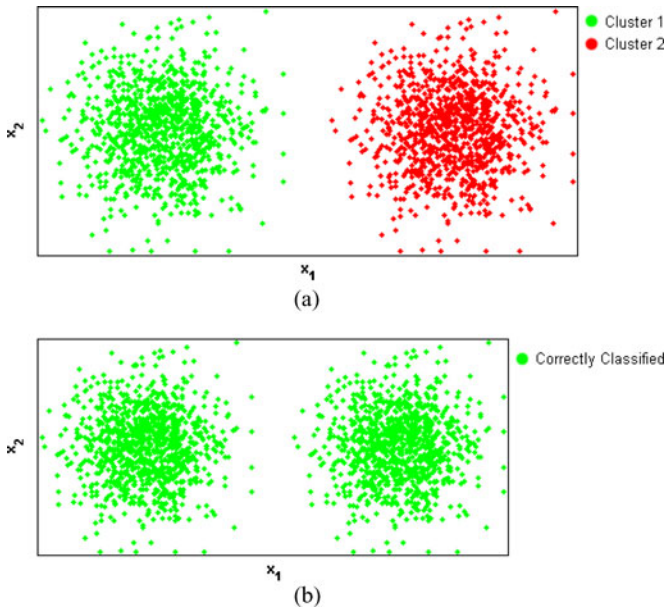

Fig. 11. Local fusion results using CELF-M. (a) Samples clustered in the feature space. A different color is used for each cluster. (b) Fusion results.

Fig. 11(a) illustrates the clustering result using CELF-M with the number of clusters $C = 2$, the fuzzifier $m = 2$, $\alpha = 10$, and the learning rate $\eta = 0.1$ (same as the one used in the global fuzzy integral). The classification results are shown in Fig. 11(b). As can be seen, CELF-M has an accuracy of 100%.

C. Application of CELF-M to Image Annotation

In this section, we report the results of applying CELF-M to image annotation. We use a subset of 1000 color images from the COREL image collection. This subset includes ten categories. Each category contains 100 images. To generate the test dataset, 25 images from each category were randomly selected. The remaining images were used for training. The training data were in part divided into training and cross-validation sets. The cross-validation set is used to select the optimal value of α. For each training set, we repeat the training process for several values of α. Then, we pick the value that results in the best accuracy on the validation set.

The goal of our experiment is to illustrate that the proposed CDF is a framework that can improve the classification accuracy by partitioning the feature space into regions and identifying local expert algorithms for each region. Thus, we did not attempt to optimize the feature extraction nor the classifier design components. We simply use a set of generic MPEG-7 descriptors [43] [color structure descriptor (CSD), scalable color descriptor (SCD), homogeneous texture descriptor (HTD), and edge histogram descriptor (EHD)]. We also use few other commonly used features [wavelet texture descriptor (WTD) and thesaurus text descriptor (TTD)]. For each set of features, we use a simple k-nearest neighbor (k-NN) classifier. The features are selected to balance the color, texture, structure, and textual properties of an image. Other descriptors and classifiers can be easily integrated into our approach.

TABLE I
DISTRIBUTION OF THE IMAGES AMONG THE TEN CONTEXTS

Cluster	1	2	3	4	5	6	7	8	9	10
Aviation	1	0	1	30	0	11	5	1	1	25
Butterflies	0	0	52	3	1	7	5	5	2	0
Cougars	3	0	0	0	48	7	0	0	0	17
Elephants	0	0	0	0	32	1	0	0	40	2
Flowers	1	0	4	6	8	14	2	2	10	28
Lions	0	0	0	0	0	17	56	0	0	2
Skiing	0	53	0	0	0	1	0	21	0	0
Sunrises	0	3	0	0	0	4	0	66	2	0
Buses	67	1	0	0	4	0	0	3	0	0
Colorado	0	0	14	1	0	47	9	2	2	0

TABLE II
ACCURACY OF THE INDIVIDUAL CLASSIFIERS AND THE FUSION ALGORITHMS

Category	WTD	CSD	EHD	HTD	SCD	TTD	FI	CELF-M
Aviation	0.72	0.76	0.81	0.61	0.81	0.68	0.88	0.96
Butterflies	0.76	0.83	0.81	0.72	0.81	0.56	0.94	0.98
Cougars	0.83	0.81	0.75	0.71	0.83	0.51	0.91	0.99
Elephants	0.65	0.81	0.85	0.55	0.73	0.73	0.85	0.94
Flowers	0.29	0.64	0.36	0.23	0.57	0.32	0.76	0.88
Lions	0.68	0.84	0.53	0.59	0.75	0.61	0.88	0.92
Skiing	0.85	0.91	0.91	0.79	0.85	0.73	0.92	0.98
Sunrises	0.84	0.93	0.64	0.80	0.95	0.72	0.96	1.00
Buses	0.93	0.95	0.83	0.61	0.96	0.64	1.00	1.00
Colorado	0.79	0.72	0.44	0.57	0.76	0.57	0.89	0.96
Overall Accuracy	0.73	0.82	0.69	0.62	0.80	0.61	0.90	0.96

We use CELF-M to fuse the six classifiers (k-NN based on WTD, CSD, EHD, HTD, SCD, and TTD with $k = 20$). We compare our results with a global fusion that uses the same fuzzy integral (i.e., we set $C = 1$). For CELF-M, we let $C = 10$ as we assume that this is sufficient to cover the variations in the given data.

Table I displays the content of the ten contexts identified by CELF-M. As can be seen, most contexts include images of "similar" categories, and thus, each one may be considered as a homogeneous context. For instance, some contexts (e.g., 1 and 2) are dominated by images from only one category. Others are dominated by images from two categories (e.g., 5 and 9). Few others include a mixture of different categories (e.g., 6 and 10).

The performance of the different classifiers and fusion algorithms is reported in Table II.

First, we note that even though the CSD-based classifier has the best overall accuracy (82%), it does not have the best performance for some categories. Second, we note that the fusion algorithms outperform all individual classifiers. In fact, these classifiers operate on different properties of the images, namely, color, texture, structure, and textual properties. This diversity allows the fusion to take advantages of the strengths of the individual classifiers and achieve a higher accuracy. Third, CELF-M

outperforms the global fuzzy integral fusion. In fact, the latter fusion tends to neglect the classifiers with low overall accuracy, even though these classifiers may perform well for some categories.

D. Application of Context Extraction for Local Fusion With Fuzzy Integrals to Landmine Detection Using Multisensor Data

Detection and removal of landmines is a serious problem affecting civilians and soldiers worldwide that has attracted the attention of several researchers in recent years [44]–[46]. The research problem for data analysis is to determine how reliably landmines can be detected and distinguished from other subterranean objects using sensor data. Difficulties arise from the large variety of landmine types, differing soil conditions, temperature and weather conditions, and varying terrain.

A variety of sensors have been proposed or are under investigation for landmine detection [47]–[50]. The key challenge to mine detection technology lies in achieving a high rate of mine detection while maintaining low level of false alarms (FAs). 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 landmine detection [51], [52]. This technology is typically used to indicate the presence or the absence of metal targets below ground. Unfortunately, many landmines are made of plastic and contain little or no metal. GPR, on the other hand, offers the promise of detecting landmines with little or no metal content. In fact, several approaches for detecting landmines and discriminating mines from clutter using GPR have been investigated [53]–[55]. Unfortunately, acceptable results have been elusive. Although systems often achieve high detection rates, it is difficult to achieve the required low FA rates. Moreover, algorithm performance can vary significantly. Therefore, fusion methods that take advantages 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.

1) Data Collection: The data used to illustrate and validate the proposed fusion method were collected using NIITEK Inc., robotic mine detection system. This system includes a GPR and a WEMI sensor and is shown in Fig. 12. It was used to acquire large sets of colocated GPR and WEMI data from two geographically distinct test sites (i.e., site A and site B). 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: *antitank metal*, *antitank with low metal content*, *antipersonnel metal*, and *antipersonnel with low metal content*. The targets were buried up to 5 in deep. Multiple datasets at different times were collected at each site resulting in a large and diverse collection of mine and clutter signatures.

In this data collection, clutter arises from two different processes. One type of clutter is emplaced and surveyed. Objects used for this clutter can be classifier into two categories: *high-metal clutter* (HMC) and *nonmetal clutter* (NMC). HMC is emplaced and surveyed in an effort to test the robustness of the detection algorithms and, in particular, those using the WEMI sensor. NMC is emplaced and surveyed in an effort to test the robustness of the GPR-based detection algorithms. The other type of clutter, which is referred to as blank, is caused by disturbing the soil.

Fig. 12. NIITEK autonomous mine detection system.

TABLE III
STATISTICS OF THE DATA COLLECTION USED IN OUR EXPERIMENT

Type		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	
FA	HMC	224	68	292	564
	NMC	72	68	140	
	Blank	52	80	132	
Total		436	439	875	875

TABLE IV
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
1”	12	104	116	70	46	116
2”	36	48	84	78	44	122
3”	28	34	62	88	18	106
4”	12	0	12	60	20	80
5”	0	10	10	0	8	8
Total	88	223	311	348	216	564

The data collection used in our experiment includes a total of 311 mine signatures and 564 clutter or FA signatures. The statistics of this collection is shown in Table III. The depth distribution for all objects is shown in Table IV.

2) Discrimination Algorithms: The raw data from each sensor are processed by various discrimination algorithms to make a decision as whether there is a mine in a certain location. Each algorithm operates on one of the sensor data to ultimately produce a set of class confidence values. It includes mainly two 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 advantages of the different sensors and detectors, we select a diverse set of algorithms that process the raw sensor data in different domains (time and frequency), extract different features, and use different classification methods. These algorithms are outlined as follows.

1) *Hidden Markov Model (HMM) Detector:* The HMM-based algorithm [53] operates on the GPR sensor data. It treats the down-track dimension as the time variable and produces a confidence that a mine is present at various positions, 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. The HMM algorithm has a background and a mine model. Each model has three states and produces a probability value. 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, i.e., states are ordered and the transition probabilities for moving to a lower numbered state are zero. See [53] for more details about this algorithm.
2) *Edge Histogram Detector (EHD) Detector:* The EHD [46] operates on the GPR sensor data. It uses translation invariant features that are based on edge histograms and a possibilistic k-NN rule for confidence assignment. 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. See [46] for more details about this algorithm.
3) *Spectral (SPECT) Detector:* The SPECT algorithm aims to capture the characteristics of a target in the frequency domain of the GPR data. It extracts the alarm spectral correlation feature (SCF) and computes a confidence value based on similarity to prototypes that characterize mine objects [56]. 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 ground level, and removes the data above and near the ground surface. 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. See [56] for more details about this algorithm.
4) *Model Fitting Detector:* The MFIT algorithm operates on the WEMI sensor data. These data were taken at 21 frequencies that were approximately logarithmically spaced from 330 Hz to 90.03 KHz. The MFIT algorithm is a feature-based classifier. First, this algorithm extracts a set of features based on fitting a model with three parameters [57] (amplitude, response time, and magnitude of asymptotes at low and high frequencies), the fitting error, and a set of three spread features. Then, a feature selection algorithm is used to reduce the seven features to four. Finally, a neural network classifier is trained using these four features to discriminate between mine and clutter objects.

For the purpose of our experiment, The training data consist of a set of mine and clutter alarms. Each alarm is processed by the EHD, HMM, SPECT, and MFIT detectors. The features extracted and the confidences generated by these discrimination algorithms were used to partition the feature space and extract the contexts. Then, within each context, the optimal set of weights is used to aggregate the confidence values of the four discrimination algorithms.

3) Motivations for Context-Dependent Fusion: Our extensive experiments have lead us to the conclusion that the performance of most detection systems can be significantly affected by various factors, and 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. These factors can have significant effects on the received sensor data and are generally unknown to an autonomous algorithm because of 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.

To illustrate the above point, in Fig. 13, we show the receiver operating characteristic (ROC) curves of the four discrimination algorithms on subsets of the data collection outlined in Section IV-D1. The different ROCs display the performance of the algorithms when different types of mines are scored. For instance, in Fig. 13(a), only antitank (AT) mines are considered. In this case, the HMM and EHD detectors have the best performance. This is because AT mines are large enough to have good GPR signatures and many of them have low metal (LM) content. This explains the relatively lower performance of the MFIT algorithm. However, for antipersonal (AP) mines, the MFIT detector has the best performance at high probability of detection (PD), as shown in Fig. 13(b). In this case, several AP mines have weak GPR signatures and cannot be easily detected by any of the GPR algorithms. Fig. 13(c) and (d) displays the ROC curves when only high metal (HM) mines or LM mines are considered. Besides the mine type, the relative performance of the different algorithms depends on other factors, such as burial depth, soil properties, and weather conditions.

The aforementioned examples suggest using different algorithms and/or features to accommodate for the different conditions. However, this task may not be as simple as it sounds

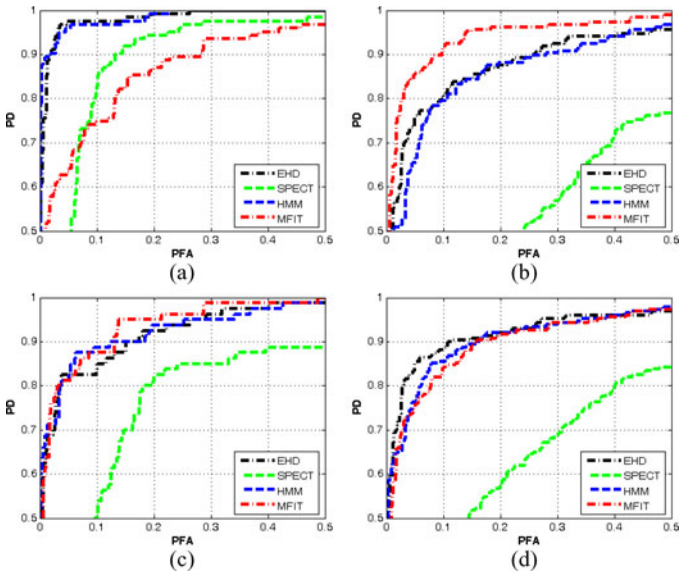

Fig. 13. Performance of the individual detectors for different types of mines when (a) only AT mines are considered, (b) only AP mines are considered, (c) only HM mines are considered, (d) only LM mines are considered. (a) AT versus clutter. (b) AP versus clutter. (c) HM versus clutter. (d) LM versus clutter.

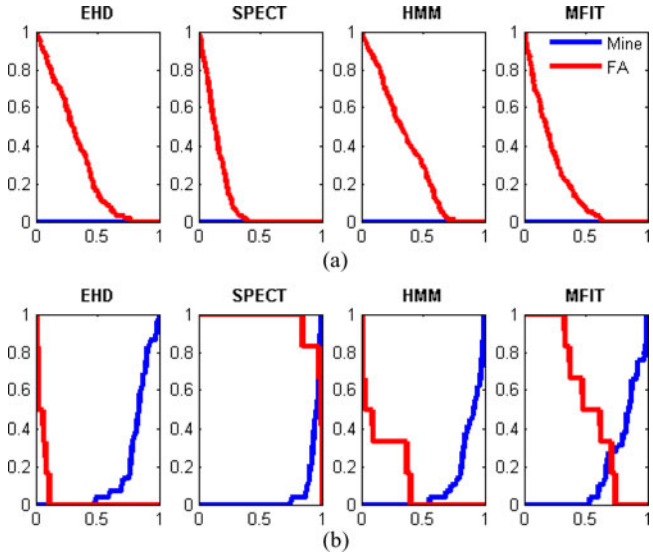

Fig. 14. Cumulative histogram of the assigned confidences by the different detectors within (a) context 1 and (b) context 9.

TABLE V
DISTRIBUTION OF THE ALARMS AMONG THE 13 CLUSTERS FOR ONE CROSS-VALIDATION SET

Cluster	Mines				FA		
	AT	AP	HM	LM	Blank	HMC	NMC
1	0	0	0	0	**74**	13	19
2	0	**16**	1	**15**	0	**78**	0
3	5	**43**	0	**48**	0	**30**	3
4	3	**43**	13	33	0	**33**	0
5	1	**19**	**20**	0	0	**22**	0
6	0	0	0	0	**51**	7	**49**
7	0	**27**	0	**27**	0	**49**	1
8	**22**	4	0	**26**	1	1	**19**
9	19	10	**29**	0	0	6	0
10	1	**12**	**12**	1	0	**38**	0
11	**23**	0	1	**22**	2	2	10
12	**37**	3	4	**36**	0	3	10
13	12	8	0	**20**	4	2	**29**

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 no trivial task and needs to be learned in an unsupervised way.

4) Results: We apply the proposed fusion method to the described dataset using sixfold cross validation. For each fold, we divide the training data into a training set and a cross-validation set. As in the previous application, the cross-validation set is used to identify the optimal value of α. For this application, the number of clusters was initially set to 20. The competitive agglomeration step within CELF-FI reduced this number to 10. Table V displays the content of the ten identified clusters in one cross-validation set. As can be seen, most clusters include alarms of similar types and, thus, may be considered as homogeneous contexts. For instance, some clusters are dominated by HM mines (e.g., cluster 9). Others are dominated by AT mines (e.g., cluster 12) or AP mines (e.g., cluster 4). In addition, some clusters include mainly mines, others include mainly clutter alarms, and few include a mixture of both. Alarms that are grouped into the same context share common GPR and/or WEMI features.

Table VI shows the assigned Shapley value of each detector within each context for one cross-validation set. As can be seen, for a given detector, CELF-FI assigns Shapley values that vary from one context to another. This is because the performance of the individual detectors is context dependent. For instance, within context 1, which contains only clutter, SCF and MFIT have better performance than EHD and HMM [see Fig. 14(a)]. Thus, CELF-FI assigns high Shapley values to these two algorithms (0.70 and 0.30, respectively). On the other hand, HMM and EHD have better performance in context 9 [see Fig. 14(b)]. Therefore, CELF-FI assigns high Shapley values to these two algorithms (0.49 and 0.42 respectively).

Table VII shows the assigned interaction indexes of each pair of detectors within each context for one cross-validation set. CELF-FI assigns indexes that vary from one context to another. For instance, within context 1, where SCF and MFIT have the best performances [see Fig. 14(a)], CELF-FI detects a conjunctive behavior in aggregation between these two detectors and assigns a positive interaction index for this pair (0.48). On the other hand, within context 9, where HMM and EHD have the best performances [see Fig. 14(b)], CELF-FI detects a disjunctive behavior in aggregation between these two detectors and assigns a negative interaction index for this pair (-0.37). An interesting case is the case of contexts 11 and 13. Within these two contexts, all interaction indexes are almost null. Thus, a linear aggregation is sufficient to have a good global score.

The overall performance of the proposed fusion method using sixfold cross validation is shown in Fig. 15. For comparison purposes, we fuse the four algorithms (HMM, EHD, SPECT, and MFIT) using the original CELF [28] and a global fuzzy

TABLE VI
SHAPLEY VALUE OF THE DIFFERENT DETECTORS WITHIN EACH CLUSTER FOR ONE CROSS-VALIDATION SET

Context	1	2	3	4	5	6	7	8	9	10	11	12	13
EHD	0.00	0.27	**0.39**	**0.44**	0.45	**0.55**	0.38	0.44	0.42	**0.66**	**0.44**	0.18	0.05
SCF	**0.70**	0.05	0.15	0.00	0.00	0.02	0.00	0.00	0.00	0.04	0.00	0.00	0.00
HMM	0.00	**0.50**	0.12	0.29	**0.49**	0.00	0.17	0.10	**0.49**	0.30	0.14	0.40	0.18
MFIT	0.30	0.18	0.34	0.27	0.06	0.43	**0.44**	**0.46**	0.09	0.00	0.42	**0.42**	**0.77**

TABLE VII
INTERACTION INDEXES OF EACH PAIR OF DETECTORS WITHIN EACH CLUSTER FOR ONE CROSS-VALIDATION SET

Context	1	2	3	4	5	6	7	8	9	10	11	12	13
{**EHD, SCF**}	0.00	0.03	-0.09	0.00	-0.00	0.02	0.00	-0.00	-0.00	-0.02	0.00	-0.00	0.00
{**EHD, HMM**}	0.00	**0.42**	-0.07	0.10	**-0.26**	0.00	0.17	-0.05	**-0.37**	-0.14	0.02	-0.18	0.00
{**EHD, MFIT**}	0.00	0.14	**-0.21**	0.09	-0.03	**0.64**	**0.48**	**-0.22**	-0.06	-0.00	0.06	-0.19	0.01
{**SCF, HMM**}	0.00	0.06	-0.03	0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.01	0.00	-0.00	0.00
{**SCF, MFIT**}	**0.48**	0.02	-0.08	0.00	-0.00	0.02	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00
{**HMM, MFIT**}	0.00	0.27	-0.06	0.06	-0.04	0.00	0.20	-0.05	-0.07	-0.00	0.02	**-0.47**	0.05

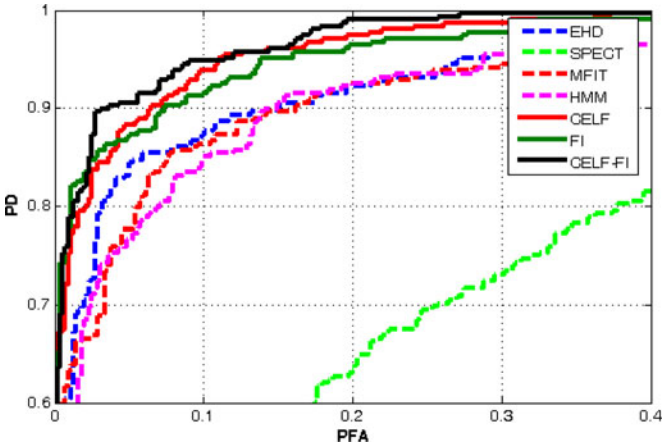

Fig. 15. PD versus PFA of CELF, fuzzy integrals fusion, CELF-FI, and the individual detectors on the entire collection using sixfold cross validation.

integral [12]–[14], [58]. Here, the CELF and CELF-FI algorithms were used to partition the feature space into 13 clusters. Fig. 15 also displays the ROC curves of the individual algorithms. First, we note that all fusion algorithms (CELF, Fuzzy Integrals, and CELF-FI) outperform all individual detectors. In fact, these detectors operate on different sensor data and use different preprocessing, feature extraction, and classification methods. This diversity allows the fusion to take advantage of the strengths of the individual detectors, overcome their weakness, and achieve a higher accuracy. Second, the proposed CELF-FI approach outperforms the original CELF approach and the global fuzzy integrals fusion. Thus, we can conclude that the nonlinear fuser used in CELF-FI performs better than the linear aggregation used in CELF and that the CDF used in CELF-FI performs better than the global fuzzy integral fusion. For instance, at a 90% PD, CELF-FI reduces the probability of false alarm (PFA) by 106% when compared with the original CELF, by 141% when compared with the original FI, and by 359% when compared with the best individual detector (EHD).

V. CONCLUSION

We have proposed a multisensor multialgorithm fusion method called CELF-FI. This approach is local and adapts the fuzzy integral fusion method to different regions of the feature space. The different regions, or contexts, are identified by partitioning the feature space formed by aggregating the features of all algorithms. A context is defined as a group of samples that share similar features (i.e., cluster) and that can be fused using the same parameters.

CELF-FI is a fuzzy approach and assigns each sample to each context with a membership degree. The final fusion is taken as the average of the local fusion outputs over all contexts weighted by the membership degrees. We have shown that the proposed adaptive fusion outperforms the global approach that uses the same fuzzy integral. We have also shown that CELF-FI outperforms the original CELF algorithm, i.e., the nonlinear aggregation used in CELF-FI performs better than the linear one used in CELF.

To identify the optimal number of contexts, we proposed an extension that starts with an overspecified number of clusters and uses competitive agglomeration to reduce the number of clusters. We have also proposed an extension of CELF-FI for data that involve more than two classes.

The proposed algorithms were applied to synthetic data to illustrate their behavior and highlight their advantages when compared with global fusion. We have also applied them to real-world applications.

In the proposed work, we used the Choquet integral with Sugeno measure to fuse the output of the individual classifiers. Future work will include integration of other types of fuzzy integrals and/or fuzzy measures [31]–[35]. We are also investigating developing a possibilistic version of the proposed approach to make it more robust.

ACKNOWLEDGMENT

The authors would like to thank R. Harmon, R. Weaver, P. Howard, and T. Donzelli for their support of this work. They would also like to thank D. Ho from the University of Missouri and F. Clodfelter and others from NIITEK, Inc., for their technical discussion, discrimination algorithms, and data collections. They would also like to thank the reviewers for their detailed and valuable comments that have improved the quality of the manuscript significantly.

REFERENCES

[1] L. Kuncheva, "'Fuzzy' versus 'nonfuzzy' in combining classifiers designed by boosting," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 6, pp. 729–741, Dec. 2003.

[2] W. Pedrycz, J. Bezdek, R. Hathaway, and G. Rogers, "Two nonparametric models for fusing heterogeneous fuzzy data," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 3, pp. 411–425, Aug. 1998.

[3] H. Seraji and N. Serrano, "A multisensor decision fusion system for terrain safety assessment," *IEEE Trans. Robot.*, vol. 25, no. 1, pp. 99–108, Feb. 2009.

[4] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Pother, "Probabilistic fusion of stereo with color and contrast for bilayer segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1480–1492, Sep. 2006.

[5] T.-Y. Kim and H. Ko, "Bayesian fusion of confidence measures for speech recognition," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 871–874, Dec. 2005.

[6] G. Pirlo and D. Impedovo, "Fuzzy-zoning-based classification for handwritten characters," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 780–785, Aug. 2011.

[7] 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.

[8] O. Melnik, Y. Vardi, and C.-H. Zhang, "Mixed group ranks: Preference and confidence in classifier combination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 973–981, Aug. 2004.

[9] P. W. Munro and B. Parmanto, "Combining neural network regression estimates with regularized linear weights," in *Advances in Neural Information Processing Systems 9*, M. Mozer, M. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 592–598.

[10] D. Wu and J. Mendel, "Aggregation using the linguistic weighted average and interval type-2 fuzzy sets," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 6, pp. 1145–1161, Dec. 2007.

[11] 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.

[12] H. Tahani and J. M. Keller, "Information fusion in computer vision using the fuzzy integral," *IEEE Trans. Syst., Man Cybern.*, vol. 20, no. 3, pp. 733–741, May/Jun. 1990.

[13] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," in *Proc. IEEE Int. Conf. Fuzzy Syst., (Int. Joint Conf. 4th IEEE Int. Conf. Fuzzy Syst., 2nd Int. Fuzzy Eng. Symp.)*, 1995, vol. 1, pp. 145–150.

[14] S. Auephanwiriyakul, J. M. Keller, and P. D. Gader, "Generalized choquet fuzzy integral fusion," *Inf. Fusion*, vol. 3, no. 1, pp. 69–85, 2002.

[15] G. Beliakov, S. James, and L. Gang, "Learning choquet-integral-based metrics for semisupervised clustering," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 3, pp. 562–574, Jun. 2011.

[16] 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, 1998.

[17] M. Sugeno, K. Fujimoto, and T. Murofushi, "A hierarchical decomposition of choquet integral model," *Int. J. Uncertainty, Fuzz. Knowl.-Based Syst.*, vol. 3, no. 1, pp. 1–15, 1995.

[18] T. Murofushi, M. Sugeno, and K. Fujimoto, "Separated hierarchical decomposition of choquet integral," *Int. J. Uncertainty, Fuzz. Knowl.-Based Syst.*, vol. 5, pp. 563–585, 1997.

[19] E. Mandler and J. Schurmann, "Combining the classification results of independent classifiers based on the dempster-shafer theory of evidence," *Pattern Recognit. Artif. Intell.*, pp. 381–393, 1988.

[20] 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.

[21] 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.

[22] Y. Xiaowei, Z. Guangquan, L. Jie, and M. Jun, "A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 105–115, Feb. 2011.

[23] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," *Pattern Recognit. Lett*, vol. 20, pp. 429–444, 1999.

[24] L. A. Rastrigin and R. H. Erensterin, *Method of Collective Recognition*. Moscow, Russia: Energoizdat, 1981 (in Russian).

[25] L. Kuncheva, "Change-glasses approach in pattern recognition," *Pattern Recogit. Lett.*, vol. 14, pp. 619–623, 1993.

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

[27] H. Frigui, L. Zhang, P. D. Gader, and D. Ho, "Context-dependent fusion for landmine detection with ground penetrating radar," presented at the SPIE Conf. Detect. Remediat. Technol. Mines Minelike Targets, Orlando, FL, 2007.

[28] H. Frigui, P. Gader, and A. C. Ben Abdallah, "A generic framework for context-dependent fusion with application to landmine detection," presented at the SPIE Defense Security Symp., Orlando, FL, 2008.

[29] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

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

[31] K. Leszczynski, P. Penczek, and W. Grochulski, "Sugeno's fuzzy measure and fuzzy clustering," *Fuzzy Sets Syst.*, vol. 15, no. 2, pp. 147–158, Mar. 1985.

[32] G. Klir and Z. Wang, *Fuzzy Measure Theory*. New York: Plenum, 1992.

[33] M. Grabisch, J. Marichal, R. Mesiar, and E. Pap, *Aggregation Functions*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[34] D. Denneberg, *Non-Additive Measure and Integral*. Dordrecht, The Netherlands: Kluwer, 1994.

[35] E. Pap, *Null-Additive Set Functions*. Norwell, MA: Kluwer, 1995.

[36] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. dissertation, Tokyo Inst. Technol., Tokyo, Japan, 1974.

[37] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets Syst.*, vol. 29, no. 2, pp. 201–227, 1989.

[38] T. Murofushi, "A technique for reading fuzzy measures (i): The shapley value with respect to a fuzzy measure," in *Proc. 2nd Fuzzy Workshop*, Nagaoka, Japan, Oct. 1992, pp. 39–48 (in Japanese).

[39] T. Murofushi and S. Soneda, "Techniques for reading fuzzy measures (iii): Interaction index," in *Proc. 9th Fuzzy Syst. Symp.*, Sapporo, Japan, May 1993, pp. 693–696 (in Japanese).

[40] M. Grabisch, "A graphical interpretation of the choquet integral," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 627–631, Oct. 2000.

[41] A. Mendez-Vazquez, P. Gader, J. 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.

[42] H. Frigui and R. Krishnapuram, "A robust competitive clustering algorithm with applications in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 450–465, May 1999.

[43] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG 7: Multimedia Content Description Language*. New York: Wiley, 2002.

[44] 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. Detect. Remediat. Technol. Mines Minelike Targets*, 2002, pp. 728–735.

[45] 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. Detect. Remediat. Technol. Mines Minelike Targets IV*, Orlando, FL, 1999, pp. 1075–1084.

[46] 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.

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

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

[49] T. T. Nguyen, D. N. Hao, P. Lopez, F. Cremer, and H. Sahli, "Thermal infrared identification of buried landmines," *Proc. SPIE, Int. Soc. Opt. Eng.*, 2005, vol. 5794, no. 1, pp. 198–208.

[50] M. Tokhi, G. Virk, and M. Hossain, "Detection of landmines using nuclear quadrupole resonance (NQR): Signal processing to aid classification," in *Climbing and Walking Robots*. Berlin, Germany: Springer-Verlag, 2006, no. 13, pp. 833–840.

[51] W. Scott, "Broadband array of electromagnetic induction sensors for detecting buried landmines," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, vol. 2, pp. II-375–II-378.
[52] P. Gao, L. Collins, N. Geng, L. Carin, D. Keiswetter, and I. Won, "Classification of landmine-like metal targets using wideband electromagnetic induction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1999, vol. 4, pp. 2327–2330.
[53] 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. 12, pp. 1867–1885, 2005.
[54] C.-C. Yang and N. Bose, "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.
[55] 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.
[56] K. Ho, L. Carin, P. Gader, and J. 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.
[57] 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. 1294–1298, Jun. 2001.
[58] G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1955.

Ahmed Chamseddine Ben Abdallah received the B.S. degree in polytechnic engineering in 2004 and the M.S. degree in mathematical engineering in 2005 from Tunisia Polytechnic School, La Marsa, Tunisia, and the Ph.D. degree in computer science and engineering from University of Louisville, Louisville, KY, in 2010.

He is currently with Bloomberg LP, New York, working on financial software development.

Hichem Frigui (M'11) 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 a 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 more than 135 journal and refereed conference articles.

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

Paul Gader (F'11) received the Ph.D. degree in mathematics from the University of Florida, Gainesville, in 1986.

He has worked as a Senior Research Scientist with Honeywell's Systems and Research Center, as a Research Engineer and Manager with the Environmental Research Institute of Michigan, and as a Faculty Member with the University of Wisconsin, Oshkosh; the University of Missouri, Columbia; and the University of Florida. He is currently a Professor with the Department of Computer and Information Science and Engineering, University of Florida. He has 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. He has approximately 260 technical publications in the areas of image and signal processing, applied mathematics, and pattern recognition, including approximately 77 refereed journal articles. His research interests include hyperspectral and Light Detection and Ranging image and signal analysis, landmine detection, handwriting recognition, machine learning/pattern recognition, fuzzy and random sets, and Choquet integration.