Unified Design of a Feature-Based ADAC System for Mine Hunting Using Synthetic Aperture Sonar

Raquel Fandos, Abdelhak M. Zoubir, *Fellow, IEEE*, and Konstantinos Siantidis

***Abstract*— A system for automatic detection and classification (ADAC) of underwater objects for mine hunting applications is proposed. The system consists of three steps: segmentation, feature extraction, and classification. This paper focuses on two design issues: the selection of the optimal classifier and the selection of the optimal feature subset. Often, the comparison of classification systems is based on a pre-selected feature set. However, a different subset might yield a different ranking. We apply a resampling algorithm that assesses the classifier performance without constraints to any specific feature subset. Once a classifier is chosen, a feature selection algorithm estimates the optimal feature subset. We propose a novel extension of the sequential forward selection (SFS) and the sequential forward floating selection (SFFS) methods, which mitigates their main limitations, i.e., the nesting problem. Instead of keeping the best alternative at each iteration, a set of *D* options is stored. The performance of the so-called *D*-SFS and *D*-SFFS is tested on simulated and real data, significantly outperforming the standard algorithms. The proposed methods are also used for designing an ADAC system for mine hunting based on two extensive databases of synthetic aperture sonar images.**

***Index Terms*— Classification algorithms, feature extraction, image classification, image processing, image recognition, object detection, object segmentation, sea floor, sonar detection, synthetic aperture sonar.**

I. Introduction

THERE exists an increasing interest in automatic target recognition for remote sensing applications [1]–[6]. In particular, in the context of underwater mine hunting, the first works on computer-aided detection and computer-aided classification (CAD/CAC) utilized sidescan sonar images [7]–[11]. The degradation of the resolution of the images, especially for long range areas, limits their performance. The development of synthetic aperture sonar (SAS) technologies over the last years and the high resolution imagery it provides [12], [13], have increased the attention toward automatic detection and classification (ADAC) for mine hunting.

Manuscript received December 8, 2011; revised July 2, 2012 and December 16, 2012; accepted March 22, 2013. Date of publication July 24, 2013; date of current version February 27, 2014. This work was supported by the collaboration with ATLAS ELEKTRONIK GmbH. This work was performed at the company site in Bremen, Germany.

R. Fandos and A. M. Zoubir are with the Department of Telecommunications, Signal Processing Group, Technische Universität Darmstadt, Darmstadt 64283, Germany (e-mail: rfandos@spg.tu-darmstadt.de; zoubir@spg.tu-darmstadt.de).

K. Siantidis is with ATLAS ELEKTRONIK GmbH, Bremen W-2800, Germany (e-mail: konstantinos.siantidis@atlas-elektronik.com).

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.2013.2260863

Military applications are often linked to such systems and hence, few studies with real data have been published. They are typically based on either template fitting [14]–[19] or feature description of the objects. In this paper, we consider the latter. Feature-based ADAC systems for mine hunting consist of three steps.

1) *Segmentation:* The image is partitioned into three regions: the highlight of the objects, their shadows, and the background. Besides man-made objects, physical features of the terrain, such as rocks and sand ripples may also be segmented.
2) *Feature extraction:* Each object (highlight and associated shadow) is characterized by a set of features.
3) *Classification:* Each object is classified according to the comparison of its feature vector and those of a training set. Ideally, the physical features of the terrain are classified as clutter and the man-made objects as mines. If not, they constitute false alarms and missed detections, respectively.

Sidescan sonar ADAC systems based on feature schemes have been employed in several studies [7], [9]–[11], [20]–[23]. Most of them (see Section III) use a variety of geometrical, statistical, and textural features. A common approach to maximize the performance is the fusion of several simpler systems [10], [11], [24], [25]. Alternatively, the fusion of multiple views of the same object can be exploited [26]–[28].

In [29], we described a single view, feature-based ADAC system for SAS, and we compared the performance of several types of descriptors (e.g., normalized central moments and principal components). In this paper, we propose a unified ADAC system that is a continuation of the work presented in [29]. A more sophisticated segmentation algorithm is introduced and the feature set is extended, in an attempt to minimize the influence that poor segmentation has on the feature values. More importantly, the focus of this paper is on a novel framework for both the selection of the optimal classifier and the selection of the optimal subset of features. Indeed, they are interdependent. When classification systems are compared, typically the same feature subset is used for all of them. However, a different feature subset might produce a different ranking. In [30], a resampling algorithm [31] for quality assessment of classifier performance, which is unconstrained to any specific feature subset, is proposed.X We utilize it to choose the optimal classification system. Moreover, the algorithm provides confidence intervals for the optimal

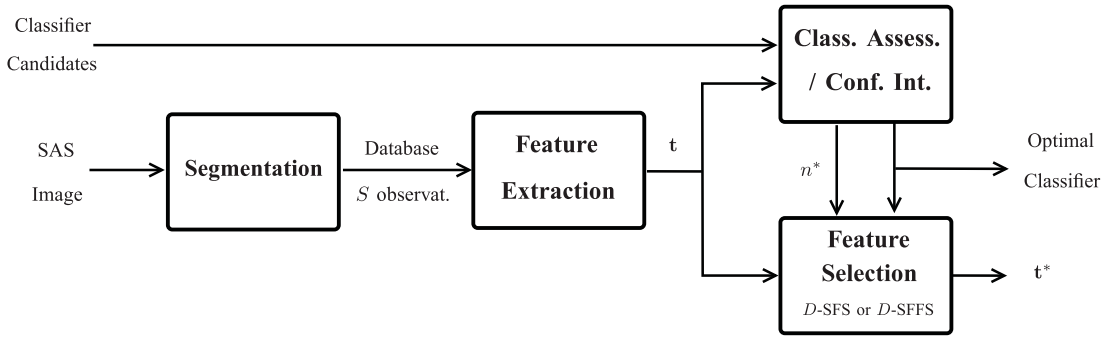


Fig. 1. Scheme of the ADAC system design procedure for mine hunting proposed in this paper. The output of the system, $\mathbf{t}^*$ and the optimal classifier, will be employed by the working ADAC system.

dimension of the feature vector. Then, a feature selection algorithm is required to estimate the actual optimal feature subset subject to that classifier. Knowing *a priori* the expected dimensionality of the feature subset, reduces the search space. Given a classification problem where S observations are available and a feature vector $\mathbf{t}$ of dimension N is provided for each observation s, $1 \leq s \leq S$, it is of importance to select a subset of n^* features, $\mathbf{t}^* = \{t_1^*, t_2^*, \ldots, t_{n^*}^*\}$, with $n^* < N$. This procedure reduces not only the cost of recognition by reducing the number of features that need to be computed but also, due to the curse of dimensionality [32], it generally provides a better classification accuracy. Numerous feature selection algorithms exist in [33]. In this paper, we propose an extension of two well-established ones, the sequential forward selection (SFS) and the sequential forward floating selection (SFFS). While the former sequentially adds features to the optimal subset, the latter allows as well for removal. The algorithms suffer from limitations that undermine their performance, e.g., the nesting of feature subsets (the best n-feature subset does not necessarily contain the best subset of $n-1$ elements). The extension that we propose alleviates them by storing the best D options (with $D > 1$) instead of choosing a single feature at each step. Thus, the extended algorithms are called D-SFS and D-SFFS, respectively. Their performance has been tested on 120 simulated data sets and on six standard databases of the UCI machine learning repository [34].

Fig. 1 shows a flowchart of the ADAC system design procedure we propose. First, the SAS image is segmented. Each object constitutes an observation s of the database, $1 \leq s \leq S$, which is characterized by a feature vector $\mathbf{t}$. After selecting a set of classifier candidates (e.g., Mahalanobis' classifier and k-Nearest Neighbor), their performance is assessed according to [30] and the best one is chosen. Moreover, confidence intervals for the optimal number of features n^*, subject to the selected classifier, are provided. Finally, a feature selection algorithm (either the D-SFS or the D-SFFS method) estimates the optimal feature subset $\mathbf{t}^*$. For the working ADAC system, consisting of the segmentation, feature extraction and classification steps, only the features of the optimal subset $\mathbf{t}^*$ need to be extracted, and only the selected classifier candidate is utilized.

Two databases of real data have been employed to test the proposed algorithms.[1] The first one consists of 437 man-made objects (308 spheres and 129 cylinders) and 3604 clutter objects. The second database has been collected by the NATO Undersea Research Center (NURC), in the framework of a NATO project for mine countermeasures [35]. It comprises 45 wedge-shaped objects, 67 cylinders, and 73 truncated cones that were measured during surveys using a synthetic aperture sonar mounted on the MUSCLE autonomous underwater vehicle. In order to study the ability of the system to avoid false alarms, the 3604 clutter elements from the first data set have been adopted in the second data set as well. Note that the imbalance between the clutter class and the mine classes is very strong. This issue is regarded in the definition of a suitable figure of merit for the system (see Section IV-A).

The main contributions of this paper are summarized as follows. Instead of combining several nonoptimal simple classification systems, each of them based on a single type of descriptors (algorithm fusion), a single optimized classification system combining multiple feature types is proposed. The resampling method in [30] is applied to compare a set of classifier candidates and the best alternative is chosen. Besides, it provides the optimal number of features. Unlike most existing mine hunting methods, which include all extracted features in the feature set, we apply the D-SFS and D-SFFS algorithms to search for the optimal feature subset. These novel algorithms show a significant performance improvement with respect to the standard SFS and SFFS.

This paper is organized as follows. Sections II and III describe briefly the segmentation and feature extraction steps, respectively. Section IV is devoted to the quality measure employed to assess classifier performance. Section V deals with feature selection. The D-SFS and D-SFFS algorithms are described and tested. Section VI provides results for the two SAS data sets. Conclusions are drawn in Section VII and the novel features are described in the Appendix.

II. Segmentation

While template fitting approaches often do not require segmenting the image [18], feature extraction methods nor-

[1]This work is part of a collaboration with ATLAS ELEKTRONIK GmbH. The experiments presented in this paper were performed at the company site in Bremen, Germany.

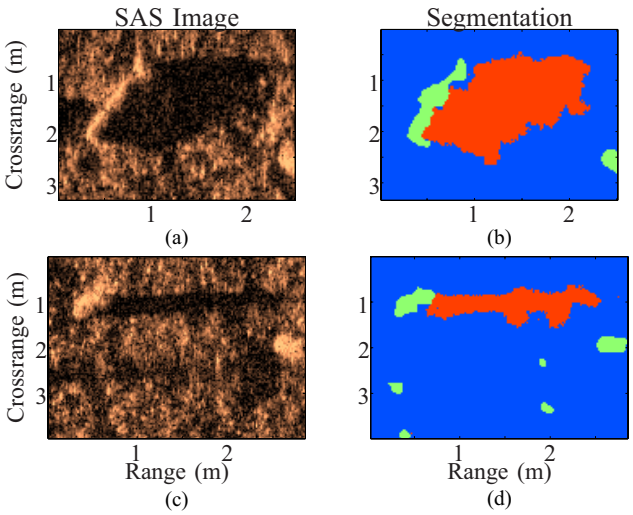


Fig. 2. Snapshots of SAS images and segmentation results with three pixel classes: hlt (green), sdw (red), and bkg (blue). The sonar system moves along the crossrange direction, following a path to the left of the scanned seabed. Some areas of the seabed remain hidden behind the objects and are not illuminated by the ultrasound energy. They originate shadows, i.e., the dark areas of the image to the right of each object. The object highlights correspond to the lighter areas of the image. Note that the shadows of the objects are, indeed, more prominent than the object highlights. Both range and crossrange are measured in meters. (a) SAS image of a cylindrical mine. (b) Segmentation result corresponding to the object in Fig. 2(a). (c) SAS image of a spherical mine. (d) Segmentation result corresponding to the object in Fig. 2(c).

mally do (although not always, see [36]). In this paper, the SAS images are segmented in order to simplify the subsequent analysis of the objects.

Three regions are generally regarded by sonar image segmentation algorithms: the highlight of the object in the scene hlt, its shadow sdw, and the background bkg. Fig. 2 includes two snapshots showing a cylindrical and a spherical man-made object and their segmentation. The resolution of the images is 2.5×2.5 cm per pixel.

The most popular segmentation algorithm for sonar applications is based on a Markovian representation of the image [37], [38]. Other algorithms, such as the min-cut/max-flow algorithm or active contours produce successful results as well [39], [40].

The segmentation algorithm used in this paper (for details see [39], [41]) is a cascade of two segmentation approaches: under a Markovian image representation, first the iterative conditional modes (ICM) algorithm is applied. Based on its segmentation of the sdw region, the min-cut/max-flow algorithm [42] is initialized, producing an improved segmentation for the objects shadows.

III. Feature Extraction

In this section, the different features that are extracted for each segmented object are described. The selection of a meaningful feature subset, that provides the best system performance, will be accomplished by the feature selection methods (see Section V).

Traditionally, the description of sonar objects has focused on the shadow shape [43], including, e.g., Fourier descriptors [20], [44] and normalized central moments [45]. The statistical properties of the image have also been considered [46], [47].

For this paper, the set of features presented in [29] has been extended. A combination of statistical and geometrical features for both the highlight and shadow regions is considered. In total, the following descriptors form the feature set $\mathbf{t}$.

1) Principal components of the segmented shadow [48], ζ_i, $1 \leq i \leq 50$.
2) Normalized central moments [49], $\mu_{i,j}$, $0 \leq i+j \leq 10$.
3) Invariant moments [49], γ_i, $1 \leq i \leq 7$.
4) 2-D-Fourier coefficients, $\phi_{i,j}$, $0 \leq i \leq 7$, $0 \leq j \leq 7$.
5) Statistical features.
6) Geometrical features of the shadow.
7) Geometrical features of the highlight.
8) Descriptors of the shadow–highlight relation.

Let us describe in more detail the four last groups. A Weibull distribution, $\mathcal{W}(\beta, \xi)$, models properly the sonar images intensity distribution [50], where β and ξ are the so-called scale and shape parameters, respectively. The parameters that maximize the likelihood function for the shadow pixels, β_{sdw} and ξ_{sdw}, the highlight pixels, β_{hlt} and ξ_{hlt}, and the background pixels, β_{bkg} and ξ_{bkg}, are estimated. The difference between them for the different regions, $\{\Delta\beta_{\text{sdw,bkg}}, \Delta\beta_{\text{hlt,bkg}}, \Delta\beta_{\text{sdw,hlt}}, \Delta\xi_{\text{sdw,bkg}}, \Delta\xi_{\text{hlt,bkg}}, \Delta\xi_{\text{sdw,hlt}}\}$ is employed as features, where $\Delta\beta_{\text{sdw,bkg}} = \beta_{\text{sdw}} - \beta_{\text{bkg}}$, and analogously for the others.

As geometrical features of the shadow, its area γ_{sdw}, perimeter ρ, compactness χ, ratio of principal axes r_{sdw}, and solidity Γ have been employed. Besides, six novel descriptors are proposed in the Appendix: orientation o, width in range direction η_u, width in crossrange direction η_v, the ratio between η_u and η_v (denoted by r_η), and the coefficients Ψ and Φ. They minimize the influence of poor segmentation scenarios, which might be due to either low image quality or challenging seabeds, such as sand ripples.

The highlight is characterized by its area γ_{hlt}, length in crossrange direction Π_{hlt}, and ratio of principal axes r_{hlt}. The shadow–highlight relation is described by the ratio between their crossrange lengths $r_{\text{sdw,hlt}}$, the difference in orientation angle Δ_o, and the mean distance between the left edge of the shadow and the right edge of the highlight, $d_{\text{sdw,hlt}}$. For details on the implementation, see [29].

The dimension of $\mathbf{t}$ is $N = 204$.

IV. Classification

Numerous classification systems exist in the literature (see [51], [52] for a review), e.g., Mahalanobis' classifier, k-Nearest Neighbor (k-NN), neural networks, support vector machines, etc. There is no overall optimal classifier, and the superiority of one over another is application-dependent.

Sometimes, the selection of a classifier is determined by external constraints, such as limited computational power. However, if the system designer is free to choose a classification system, he or she requires a quantitative measure in order to compare the performance of several classifier candidates.

TABLE I
CONFUSION MATRIX OF A THREE-CLASS SYSTEM. THE MISSED DETECTED MINES ARE INDICATED IN RED AND THE FALSE ALARMS IN BLUE

		Predicted		
		Class 1	Class 2	Class 3
Actual	class 1	$a_{1\|1}$	$a_{2\|1}$	$a_{3\|1}$
	class 2	$a_{1\|2}$	$a_{2\|2}$	$a_{2\|2}$
	class 3	$a_{1\|3}$	$a_{2\|3}$	$a_{3\|3}$

A. Figure of Merit

The natural figure of merit f of a classifier is the overall misclassification rate, P_e. Although P_e is bounded for some specific distribution, the error rate must be estimated from the available data if the distribution is unknown and the number of observations is finite. Given a database of test observations and a class label $c \in \{1, \ldots, C\}$ associated with each observation, where C is the number of classes, the misclassification rate is calculated as

$$P_e = \sum_{c_1} P(c_1) \cdot \sum_{c_2 \neq c_1} P(c_2|c_1), \quad c_1, c_2 \in \{1, \ldots, C\} \quad (1)$$

where $P(c_1)$ is the prior probability of class c_1 and $P(c_2|c_1)$ is the probability of deciding for class c_2 when the actual class is c_1.

Let us now consider our particular problem. The SAS databases consist of 3604 clutter objects (class 1) and either two (database SAS1) or three (database SAS2) other classes with a number of observations between 45 and 308. Hence, the problem is a multiclass one and imbalanced. Given the dominance of class 1, minimizing $f := P_e$ focuses on reducing $P(c|1)$ with $c \neq 1$, that is, the false alarm rate. Indeed, we are far more interested in reducing $P(1|c)$, i.e., the rate of mines classified as clutter or missed detected mines. The Neyman–Pearson test, that fixes a certain false alarm rate and calculates the corresponding missed detection rate, is more appropriate. However, the nature of some classifiers, such as support vector machines or k-nearest neighbors (employed in Section VI) does not allow for fixing the false alarm rate, and therefore the Neyman–Pearson test becomes unusable. In the sequel, we propose a new figure of merit that allows for a fair comparison of these classifiers. In short, it consists of computing a 1-D figure of merit based on the confusion matrix of the system.

The confusion matrix of a system with $C = 3$ classes is represented in Table I. Element $a_{c_2|c_1}$ accounts for the number of elements with actual class c_1 and predicted class c_2. The expression of the proposed figure of merit f reads

$$f_\lambda = \lambda \sum_{c_2 \neq 1} a_{c_2|1} + (1-\lambda) \sum_{c_1 \neq 1} \sum_{c_2 \neq c_1} a_{c_2|c_1} \quad (2)$$

for $c_1, c_2 \in \{1, \ldots, C\}$ and $0 \leq \lambda \leq 1$. There are two main differences between (1) and (2). First, the former expresses all quantities as probabilities while the latter employs absolute numbers of observations. Second, the former weighs the classification error of each class with its prior probability while the latter introduces a weighting factor λ that determines the relative importance of misclassifying clutter with respect to misclassifying a mine, independently of the prior distributions of the classes.

Note that both $f := P_e$ and $f := f_\lambda$ give the same importance to classifying a mine class observation $c_1 \neq 1$ as a different mine class, and to classifying it as clutter (missed detection). That is, in (1), the weight of $P(c_2|c_1)$ is the same for $c_2 \neq 1$ (mine) and $c_2 = 1$ (clutter), and analogously in (2). A new weighting factor μ allows for distinguishing both scenarios

$$f_{\lambda,\mu} = \lambda \underbrace{\sum_{c_2 \neq 1} a_{c_2|1}}_{\text{false alarms}} + (1-\lambda) \left[\mu \underbrace{\sum_{c_1 \neq 1} \sum_{c_2 \neq 1, c_1} a_{c_2|c_1}}_{\text{mines wrong class}} + (1-\mu) \underbrace{\sum_{c_1 \neq 1} a_{1|c_1}}_{\text{missed mines}} \right] \quad (3)$$

for $c_1, c_2 \in \{1, \ldots, C\}$ and $0 \leq \lambda, \mu \leq 1$. For $\mu = 0$, we do not penalize the assignment of a mine to a wrong class, as long as it is a mine class (i.e., it is not clutter).

B. Assessment of Classifier Performance and Optimal Number of Features

Traditionally, the comparison of two classifiers is based on an f value calculated from a single feature set (e.g., [53]–[55]). However, a second feature set might yield a different ranking. A solution to this issue is proposed in [30]. In the following, we provide a summary of the algorithm, but details are given in [30]. Instead of obtaining a single value for f, it is considered as a random variable conditional on the dimension n of the feature subset, with $n < N$. A set of possible feature vector sizes, $n_i \in \{n_1, n_2, \ldots, n_{\max}\}$, $n_{\max} \leq N$, is defined, and the probability distribution of f subject to n_i, $\hat{p}_{f|n_i}$, is estimated by applying resampling techniques. We choose the number of samples B and repeat the following steps for each n_i, noting that the feature vector $\mathbf{t}$ is considered as random conditional on the data.

1) *Step 1:* Select randomly from $\mathbf{t} = \{t_1, t_2, \ldots, t_N\}$ a set of n_i features to obtain a sample $\mathbf{t}'_b = \{t'_1, t'_2, \ldots, t'_{n_i}\}$.
2) *Step 2:* Calculate the figure of merit $f'_{b,n_i} = f(\mathbf{t}'_b)$.
3) *Step 3:* Repeat steps 1 and 2 B times to obtain a vector $\{f'_{1,n_i}, \ldots, f'_{B,n_i}\}$, from which the distribution $\hat{p}_{f|n_i}$ can be empirically estimated.

The better the performance of the classifier is, the lower are the values reached for f. Hence, if we depict $\hat{p}_{f|n_i}$ in the $f - n$ plane, the energy of $\hat{p}_{f|n_i}$ is concentrated in low f values for good classifiers. Taking this into account, the following quantitative assessment of classifier performance is proposed:

$$Q = \frac{1}{n_{\max} - n_1} \sum_{n_i = n_1}^{n_{\max}} \Delta n_i \cdot \int_0^1 w(f) \cdot \hat{p}_{f|n_i} \, \mathrm{d}f \quad (4)$$

where $\Delta n_i = n_{i+1} - n_i$. The function $w(f)$ must be monotonically decreasing, e.g., $w(f) = f^{-\alpha}$ with $\alpha > 0$, so that the contribution of $\hat{p}_{f|n_i}$ to Q is greater for lower f. The value $\alpha = 1$ produces satisfactory results.

Furthermore, $\hat{p}_{f|n_i}$ can be used to estimate the distribution of the number of features subject to f. According to Bayes' theorem

$$\hat{p}_{n|f} = \frac{\hat{p}_n \cdot \hat{p}_{f|n_i}}{\hat{p}_f}, \quad n_i \in \{n_1, \ldots, n_{\max}\} \tag{5}$$

where $\hat{p}_n$ and $\hat{p}_f$ are considered uniformly distributed and therefore, $\hat{p}_{n|f} = A \cdot \hat{p}_{f|n_i}$, $n_i = n_1, \ldots, n_{\max}$, and A is a normalizing constant. The region of the $f-n$ plane where n^* and its corresponding f^* are most probably located is limited by the 95% confidence interval of $\hat{p}_{n|f}$, $q_{n|f}^{0.025} \le n \le q_{n|f}^{0.975}$, which depends on the value of f. Feature selection algorithms are likely to provide an optimal feature subset that performs better than any randomly generated subset. Therefore, we can limit the search of the optimal feature subset to $f \le f_b^{\min}$, where $f_b^{\min}$ is the smallest figure of merit estimate

$$f_b^{\min} = \min\left\{f'_{b,n_i}, b \in \{1, \ldots, B\}, n_i \in \{n_1, \ldots, n_{\max}\}\right\}. \tag{6}$$

When $\hat{p}_{f|n_i}$ is to be used for classification quality assessment, its derivation from the figure of merit estimates $\{f'_{1,n_i}, \ldots, f'_{B,n_i}\}$ may be performed by histogram techniques, nonparametric density estimation, or by using a parametric model. However, when $\hat{p}_{n|f}$ is derived from $\hat{p}_{f|n_i}$ [see (5)] for prediction of the optimal number of features, $\hat{p}_{f|n_i}$ must be inferred by means of a parametric model that extrapolates $\hat{p}_{n|f}$ for $f \le f_b^{\min}$ (see [30] for details). In this paper, a Gaussian distribution has been adopted.

V. Feature Selection

The problem of feature selection is defined as follows [33]: given a set of N candidate features, $\mathbf{t} = \{t_1, t_2, \ldots, t_N\}$, select an n^*-element subset, $\mathbf{t}^* = \{t_1^*, t_2^*, \ldots, t_{n^*}^*\}$ with $n^* \le N$, that minimizes a given figure of merit f using a specific classifier.

Only an exhaustive search over the feature space guarantees the optimal feature set [56]. The size of the feature candidate database, N, makes this task prohibitive in most cases. Many methods exist in the literature (see [33], [57] for a review) that estimate a suboptimal set, $\mathbf{t}^*$. The obtained $\mathbf{t}^*$ for a given method depends on the data, the evaluation criterion (figure of merit), and the classifier. Although some controversy has arisen in the last years [58], [59], it is generally accepted that the sequential floating forward selection (SFFS) algorithm [60] provides the best performance.

A. Standard Methods: SFS & SFFS

Let us first briefly revisit the sequential forward selection (SFS) [61], from which SFFS is derived. Initialized with the best single feature, $t_1^* := \arg\min_{t_j \in \mathbf{t}} f(t_j)$, $1 \le j \le N$, the SFS algorithm adds one feature at a time that in combination with the already selected ones optimizes f: $\mathbf{t}_n^* := \{\mathbf{t}_{n-1}^*, t_n^*\}$, with $t_n^* = \arg\min_{t_j \in \mathbf{t} \setminus \mathbf{t}_{n-1}^*} f(\{\mathbf{t}_{n-1}^*, t_j\})$. The algorithm stops when n reaches a preselected value n_{M}, with $n_{\mathrm{M}} \le N$. The output of the SFS algorithm is not $\mathbf{t}^*$, but a set $\mathbf{t}_n^*$, $1 \le n \le n_{\mathrm{M}}$. The optimal subset $\mathbf{t}^*$ is obtained as

$$\mathbf{t}^* := \arg\min_{\mathbf{t}_n^*}\{f(\mathbf{t}_n^*)\}, \quad 1 \le n \le n_{\mathrm{M}}. \tag{7}$$

Its corresponding figure of merit, $f(\mathbf{t}^*)$, is denoted by f^*.

Note that the SFS algorithm requires a high amount of computation at each iteration and therefore, it is more appropriate to constraint the search space, limited by n_{M}. This can be achieved by considering the approach at the end of Section IV-B, where the optimal number of features is predicted to be in a preferably small range of values for n. The estimation of the 97.5% quantile of n given f, $q_{n|f}^{0.975}$, allows for constraining n_{M}. The value of $q_{n|f}^{0.975}$ decreases as f also decreases. Therefore, n_{M} is initialized to $n_{\mathrm{M}} := \max\{q_{n|f}^{0.975}\}$ and can be updated after each iteration n

$$\text{if } f_n < \min\{f_1, \ldots, f_{n-1}\}, \quad n_{\mathrm{M}} := q_{n|f_n}^{0.975} \tag{8}$$

where $f_n = f(\mathbf{t}_n^*)$. Typically, $q_{n|f}^{0.975}$ drops as f decreases and therefore, the search space is significantly reduced by this procedure.

Analogously to the SFS, the sequential backward selection (SBS) [62] is initialized with the whole candidate database $\mathbf{t}$ and removes one feature at a time.

The main drawback of both the SFS and SBS algorithms lies in the so-called nesting problem, that is, the optimal 3-subset is not necessarily contained in the optimal 4-subset, and so on [33]. The SFFS algorithm tries to overcome this issue. After initializing the feature set, $\mathbf{t}_0^* := \emptyset$, the SFFS repeats the following steps until the size of $\mathbf{t}_n^*$ reaches $n = n_{\mathrm{M}}$.

1) *Step 1 (SFS):* Find $t_n^* = \arg\min_{t_j \in \mathbf{t} \setminus \mathbf{t}_{n-1}^*} f(\{\mathbf{t}_{n-1}^*, t_j\})$ and update $\mathbf{t}_n^* := \{\mathbf{t}_{n-1}^*, t_n^*\}$, $n := n + 1$.
2) *Step 2 (Conditional SBS):*
   Find $t_n^{**} = \arg\min_{t_j \in \mathbf{t}_n^*} f(\mathbf{t}_n^* \setminus t_j)$
   If $f(\mathbf{t}_n^* \setminus t_n^{**}) \le f(\mathbf{t}_{n-1}^*)$ then
   a) update $\mathbf{t}_{n-1}^* := \mathbf{t}_n^* \setminus t_n^{**}$
   b) $n := n - 1$ and go to Step 2
   else go to Step 1.

The value of n_{M} is updated after each iteration as indicated by (8). The output of the algorithm is again a set $\mathbf{t}_n^*$, $1 \le n \le n_{\mathrm{M}}$. The optimal subset $\mathbf{t}^*$ and corresponding f^* are estimated as indicated by (7).

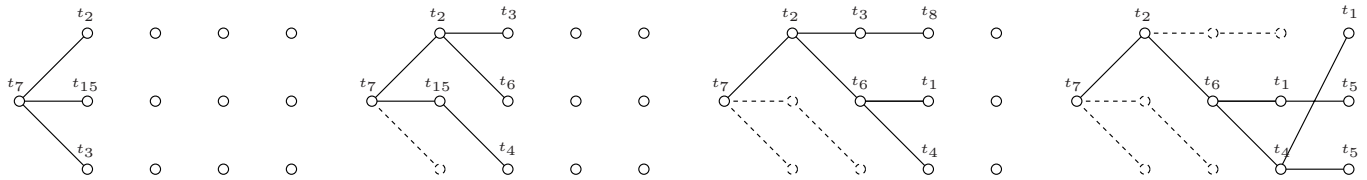
B. D-SFS

In this paper, we propose a novel modification of the SFS algorithm in order to alleviate the nesting issue. Instead of choosing the best possible feature at a time, we keep the best D candidates, with $D > 1$, in a matrix $\mathbf{T}(d, n)$, whose structure can be represented as a tree of D branches (Fig. 3 and Table II). While n, $1 \le n \le n_{\mathrm{M}}$, refers to the features index in $\mathbf{T}$, d, $1 \le d \le D$, indicates the branch. Note that the optimal 5-subset of the example, $\{t_7, t_2, t_6, t_4, t_1\}$, does not contain the best 4-subset, $\{t_7, t_2, t_3, t_8\}$.

The algorithm is formulated as follows. The matrix $\mathbf{T}$ is initialized for all branches with the best single feature: $\mathbf{T}(d, 1) := t_1^*$ $\forall d$, where $t_1^* = \arg\min_{t_j \in \mathbf{t}} f(t_j)$, $1 \le j \le N$. Then, n is increased iteratively until it reaches n_{M}.

TABLE II
EVOLUTION OF THE MATRIX $\mathbf{T}(d, n)$ FOR THE EXAMPLE IN FIG. 3. AT EACH STAGE, THE FIRST ROW CORRESPONDS TO THE LOWEST MERIT, I.E., THE OPTIMAL SUBSET FOR THAT DIMENSIONALITY

		2 Features		3 Features			4 Features				5 Features				
Branch	1	t_7	t_2	t_7	t_2	t_3	t_7	t_2	t_3	t_8	t_7	t_2	t_6	t_4	t_1
	2	t_7	t_{15}	t_7	t_2	t_6	t_7	t_2	t_6	t_1	t_7	t_2	t_6	t_1	t_5
	3	t_7	t_3	t_7	t_{15}	t_4	t_7	t_2	t_6	t_4	t_7	t_2	t_6	t_4	t_5



(a) $n = 2$ features (b) $n = 3$ features (c) $n = 4$ features (d) $n = 5$ features

Fig. 3. D-SFS working principle, with $D = 3$. Note that the same feature can be added more than once at the same iteration n as long as the corresponding branches do not coincide already at previous iterations. The evolution of the corresponding matrix $\mathbf{T}(d, n)$ is included in Table II.

1) *Step 1:* Compute the figure of merit that stems from adding each possible t_j feature to each branch d: $f_{n,d,j} = f(\{\mathbf{T}(d, \mathbf{n}), t_j\})$, where $\mathbf{n} = \{1, \ldots, n\}$ $\forall t_j \in \mathbf{t} \setminus \mathbf{T}(d, \mathbf{n})$, $\forall d$.
2) *Step 2:* Sort $f_{n,d,j}$ for all d and all j, in an ascending fashion and store the result in $\hat{f}_{n,d,j}$. Create the vectors $\hat{\mathbf{d}}$ and $\hat{\mathbf{t}}_j$ with the values of d and t_j corresponding to $\hat{f}_{n,d,j}$.
3) *Step 3:* Update the tree.
   a) Select the branches that correspond to the best merit $\mathbf{T}(d, \mathbf{n}) := \mathbf{T}(\hat{\mathbf{d}}(d), \mathbf{n})$, $1 \leq d \leq D$.
   b) Add the best feature to each branch: $\mathbf{T}(d, n+1) := \hat{\mathbf{t}}_j(d)$, $1 \leq d \leq D$.
   c) $n := n + 1$.

Different branches correspond to different values of f and therefore to different quantiles $q_{n|f}^{0.975}$. The value of n_{M} is updated after each iteration [see (8)] according to the f value of the best alternative, that is, the first row of $\mathbf{T}$, $f(\mathbf{T}(1, \mathbf{n}))$.

Once the iterative process is finished, the best alternative is selected

$$\mathbf{t}_n^* = \mathbf{T}(1, \mathbf{n}), \quad 1 \leq n \leq n_{\mathrm{M}}. \tag{9}$$

The optimal feature subset and corresponding f^* are then found according to (7).

The computational complexity of the D-SFS algorithm increases linearly with D.

C. D-SFFS

The performance of the SFFS algorithm can also be improved by tracking several branches, i.e., alternatives. Due to the SBS step in the SFFS algorithm, some branches might consist of less features than others at a given moment. Hence, we employ a different n_d for each branch d. We define a vector $\mathbf{a}$ that lists the so-called active branches. A branch d is active if it has the least amount of features, i.e., if $n_d = \min\{n_1, \ldots, n_D\}$. The state of the branches is updated only after the SBS step.

We initialize $\mathbf{T}$ as for D-SFS: $\mathbf{T}(d, 1) := t_1^*$ $\forall d$, where $t_1^* = \arg\min_{t_j \in \mathbf{t}} f(t_j)$, $1 \leq j \leq N$. At the beginning, all branches are active: $\mathbf{a} = \{1, \ldots, D\}$ and $n_d = 1$ $\forall d$. Then, we iterate until $\min\{n_1, \ldots, n_D\} = n_{\mathrm{M}}$.

1) *Step 1:* D-SFS considering only the active branches.
2) *Step 2:* Conditional SBS (Step 2 of SFFS) for each active branch.
3) *Step 3:* Update $\mathbf{a}$: $d \in \mathbf{a}$ iff $n_d \leq \min\{n_1, \ldots, n_D\}$.

Note that the tree representation used for the D-SFS is no longer possible since, for instance, the first feature might be deleted for one of the branches but not for the others. The matrix representation is preferred (see Table III for an example).

As for the D-SFS, n_{M} is updated according to the best option (first branch) at each iteration. Also, the best alternative is selected after the iterative process as indicated by (9).

The computational complexity of the D-SFFS algorithm increases linearly with D.

D. D-SFS & D-SFFS Performance

Both the D-SFS and the D-SFFS algorithms have been tested on 120 2-class synthetic data sets and on six real databases from the UCI machine learning repository [34]: the Parkinsons, Breast Cancer Wisconsin (Diagnostic), Breast Cancer Wisconsin (Prognostic), Ionosphere, Musk (version 1), and Multiple Features data sets. The simulated data sets are intentionally diverse in order to consider a variety of classification scenarios. Hence, half of the simulated data sets contain 40 features and the other half 80. The number of observations is 100 for the former and 150 for the latter. The features follow either Gaussian or bimodal Gaussian distributions of different mean values for the different classes. Both correlated and uncorrelated features are considered.

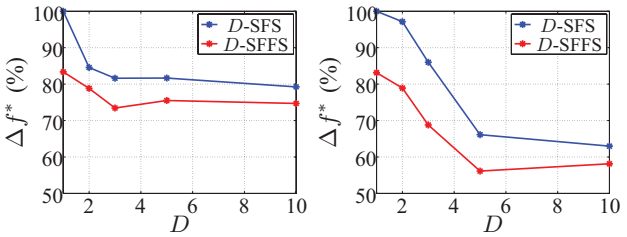
Fig. 4 illustrates the average relative improvement of the optimal figure of merit, Δf^*, that the D-SFS and D-SFFS algorithms provide with respect to the SFS f^* as a function of D. The curves have been calculated by averaging the results

TABLE III

EXAMPLE OF EVOLUTION OF THE MATRIX $\mathbf{T}(d, n)$ FOR THE D-SFFS ALGORITHM WITH $D = 3$. 3-SFS IT. $\mathbf{n}$: $\mathbf{a} = \{1, 2, 3\}$. SBS IT. $\mathbf{n}$: BRANCHES 2 AND 3 LOOSE FEATURES, $\mathbf{a} = \{2\}$. 3-SFS IT. $\mathbf{n+1}$: ONLY ON BRANCH 2. SBS IT. $\mathbf{n+1}$: ONLY BRANCH 2 IS ACTIVE, AND IT LOSES ONE FEATURE, $\mathbf{a} = \{2\}$. 3-SFS IT. $\mathbf{n+2}$: ONLY ON BRANCH 2. SBS IT. $\mathbf{n+2}$: ONLY BRANCH 2 IS ACTIVE, AND IT DOES NOT LOSE ANY FEATURE, $\mathbf{a} = \{2, 3\}$. 3-SFS IT. $\mathbf{n+3}$: ON BRANCHES 2 AND 3. SBS IT. $\mathbf{n+3}$: ALL BRANCHES ARE ACTIVE BUT NONE LOSE FEATURES, $\mathbf{a} = \{1, 2, 3\}$. 3-SFS IT. $\mathbf{n+4}$: ON ALL BRANCHES

		3-SFS it. n				**SBS it. n**				**3-SFS it. n + 1**				**SBS it. n + 1**				**3-SFS it. n + 2**			
Branch	1	t_7	t_2	t_3	t_8	t_7	t_2	t_3	t_8	t_7	t_2	t_3	t_8	t_7	t_2	t_3	t_8	t_7	t_2	t_3	t_8
	2	t_7	t_2	t_6	t_1	t_7	t_2			t_7	t_2	t_8		t_2	t_8			t_2	t_8	t_1	
	3	t_7	t_2	t_6	t_4	t_2	t_6	t_4		t_2	t_6	t_4		t_2	t_6	t_4		t_2	t_6	t_4	

		SBS it. n + 2				**3-SFS it. n + 3**				**SBS it. n + 3**				**3-SFS it. n + 4**				
Branch	1	t_7	t_2	t_3	t_8	t_7	t_2	t_3	t_8	t_7	t_2	t_3	t_8	t_2	t_8	t_1	t_6	t_5
	2	t_2	t_8	t_1		t_2	t_8	t_1	t_7	t_2	t_8	t_1	t_7	t_7	t_2	t_3	t_8	t_1
	3	t_2	t_6	t_4		t_2	t_8	t_1	t_6	t_2	t_8	t_1	t_6	t_2	t_8	t_1	t_6	t_4



(a) UCI data sets (b) Simulated data

Fig. 4. Relative decrease of f^* achieved by the D-SFS and D-SFFS algorithms with respect to the standard SFS (represented at $D = 1$). The curves have been calculated by averaging the performance of the different data sets.

from the different databases. The misclassification rate is used as figure of merit f [see (1)] and the parameter D takes the values $\{2, 3, 5, 10\}$. The results in Fig. 4 have been obtained with Mahalanobis' classifier. Linear discriminant analysis, k-nearest neighbors, and support vector machines have also been tested. They show similar performance.

For the real data sets [Fig. 4(a)], the D-SFS f^* is almost 20% lower than the SFS f^* for $D \geq 3$. Analogously, the D-SFFS reduces the value of f^* about 10% with respect to the SFFS algorithm. Increasing the value of D beyond three provides no significant improvement for either the D-SFS or the D-SFFS algorithm. Note that the optimal SFFS figure of merit outperforms the SFS f^* by more than 15%. For $D \geq 3$, however, the D-SFS algorithm outperforms the SFFS algorithm. Furthermore, the 3-SFS algorithm is computationally more efficient than the standard SFFS: while the SFFS algorithm is normally considered to be five to ten times slower than the SFS algorithm, the 3-SFS is only three times as costly.

Similar trends are observed for the synthetic examples in Fig. 4(b), with the main difference that the algorithms saturate now for $D \geq 5$ instead of 3. Otherwise, the average performance of the simulated examples is indeed superior to the real data sets results. For $D \geq 5$, the value of f^* is 25% lower for the D-SFS algorithm than for the standard SFS, and the D-SFFS reduces the figure of merit more than 20% with respect to the SFFS approach. Again, the 3-SFS average performance is almost as good as the SFFS results.

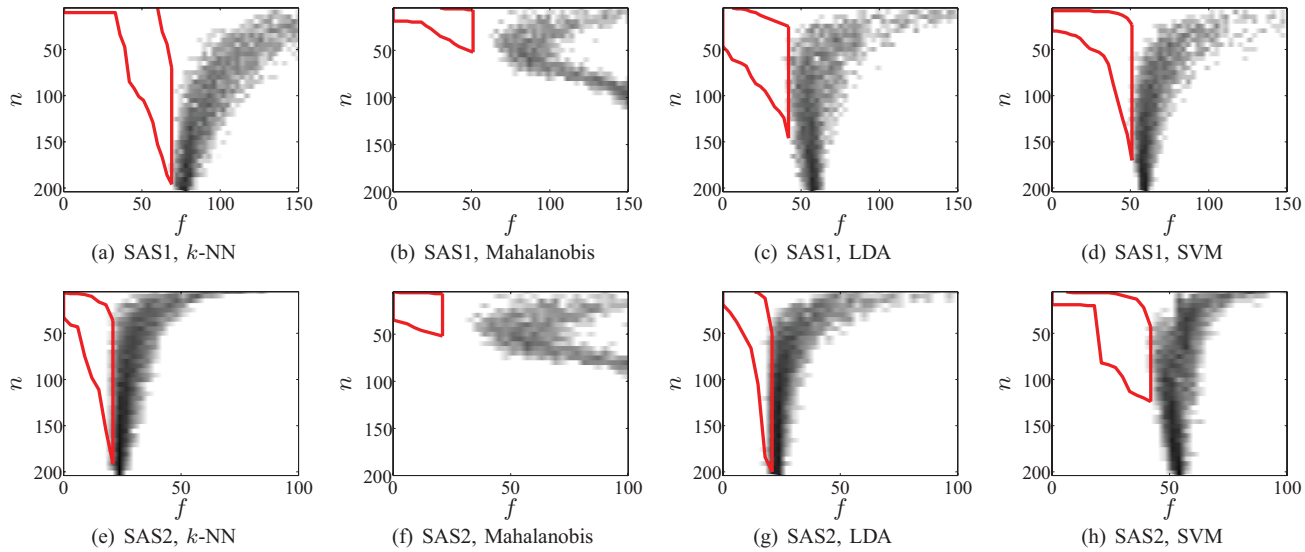
The main limitation of the D-SFS and D-SFFS methods is the fact that, although increasing D augments the probability of finding a better feature subset, it does not guarantee it. Thus, for some data sets, the value of f^* for a certain D value D_1 might be higher than for D_2, with $D_1 > D_2$.

VI. RESULTS

Two SAS databases have been used to test the algorithms above.

1) *SAS1:* It has been obtained by segmenting systematically 57 000 m^2 of SAS images, with a resolution of 2.5 × 2.5 cm. A total of 3604 clutter objects (Class 1), 308 spherical (Class 2), and 129 cylindrical (Class 3) man-made objects were segmented. Previous to the segmentation, visual inspection and classification by several experts were used to generate the ground truth.
2) *SAS2:* It has been built from a collection of snapshots of SAS images. The ground truth is known *a priori*, since the objects were placed on the seabed for this purpose. The snapshots are small and centered on the objects and, therefore, almost no clutter was obtained. In order to consider the ability of the system to avoid false alarms though, the 3604 clutter elements (Class 1) from the SAS1 database have been adopted as well. The statistical properties of both sets of images are comparable (the pixel intensity has been normalized for both databases and they follow a Weibull distribution of similar parameters), which supports the validity of this approach. Three kinds of objects are present: 45 wedge-shaped objects (Class 2), 67 cylinders (Class 3), and 73 truncated cones (Class 4).

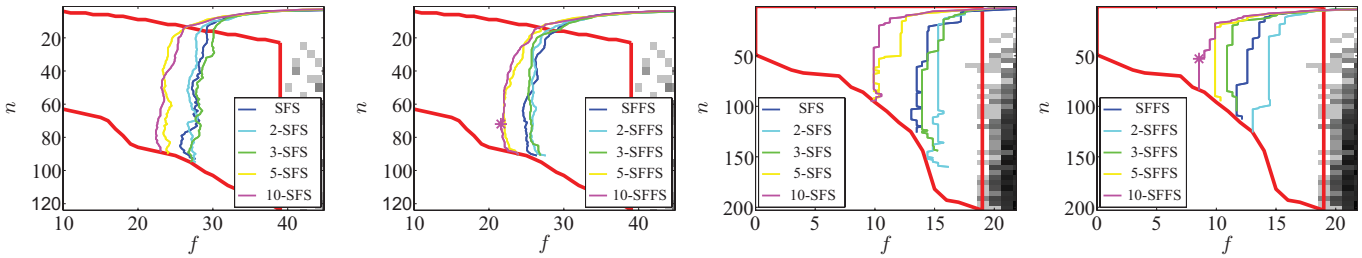
If the same database is utilized for both designing the system and testing it, an optimistic estimation of P_e is obtained [51]. In order to avoid this, a five-fold cross-validation approach has been adopted. The S available observations are randomly divided into five groups. For every feature subset tested by the feature selection methods, the classifier is designed according to four of them and tested on the fifth one. This procedure is repeated for the five groups of observations, obtaining a single value for the figure of merit for each feature subset under test. This allows for exploiting the available data (all available features are employed for the computation of the figure



(a) SAS1, k-NN (b) SAS1, Mahalanobis (c) SAS1, LDA (d) SAS1, SVM

(e) SAS2, k-NN (f) SAS2, Mahalanobis (g) SAS2, LDA (h) SAS2, SVM

Fig. 5. Quality assessment for classifier performance. The curves represent $10 \cdot \log(\hat{p}_{f|n_i})$ for $n_i = 1, \ldots, N$, calculated by histogram techniques from the figure of merit estimates. Both the database and the classification approach are indicated for each figure. The scale is common to all figures and spans between -25 dB (white) and -5 dB (black). A red line delimits the most probable region for the pair $\{f^*, n^*\}$, corresponding to $f \leq f_b^{\min}$ and $q_{n|f}^{0.025} \leq n \leq q_{n|f}^{0.975}$.



(a) SAS1, LDA, D-SFS (b) SAS1, LDA, D-SFFS (c) SAS2, LDA, D-SFS (d) SAS2, LDA, D-SFFS

Fig. 6. Feature selection results. The database, the classifier, and the feature selection algorithm are indicated for each figure. The distribution of f conditional on n, $10 \cdot \log(\hat{p}_{f|n_i})$ for $n_i = 1, \ldots, N$ is depicted in a logarithmic scale. The most probable region for $\{f^*, n^*\}$ is delimited by a red line. Note the change of both axis limits and scale with respect to Fig. 5. Each figure shows the evolution of f as n increases for either the D-SFS or the D-SFFS algorithm for a certain value of D. As a reference, the standard SFS and SFFS algorithms have been used. The best result for each database is indicated by a star.

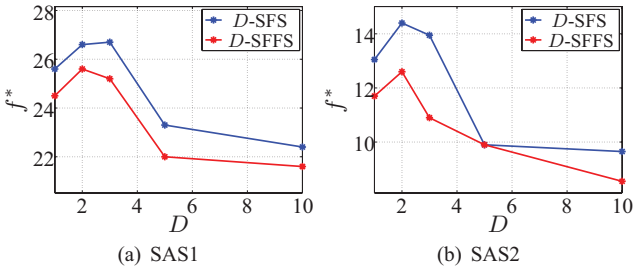
of merit) while keeping the computational cost reasonably low.

A. Classifier Performance Assessment and Confidence Intervals of the Optimal Number of Features

The classifier performance assessment presented in Section IV has been employed to compare the following classifier candidates for both data sets: a k-NN classifier with $k = 5$ [52], Mahalanobis' classifier [63], the linear discriminant analysis classifier (LDA) [52], and support vector machines (SVM) with a radial basis kernel [64].

As figure of merit, the expression proposed in (3), $f = f_{\lambda,\mu}$, has been adopted. The focus is on the minimization of the missed detected mines while keeping a reasonably low false alarm rate. Therefore, the relative weighting of the false alarms has been fixed to $\lambda = 0.1$ for both data sets. While $\mu = 0$ for the first database (no penalty is associated with assigning a wrong mine class to a mine object), $\mu = 0.5$ for the SAS2 database (the same importance is given to missed detecting a mine as to classifying it as a mine of a wrong type).

Fig. 5 shows the estimated $\hat{p}_{f|n_i}$, $n_i \in \{1, \ldots, N\}$, for both data sets and all four classifier candidates. A logarithmic



(a) SAS1 (b) SAS2

Fig. 7. D-SFS and D-SFFS performance. The results for $D = 1$ correspond to the standard SFS and SFFS algorithms.

scale has been employed, which spans between -25 dB (white) and -5 dB (black) for all examples. Each horizontal line corresponds to $\hat{p}_{f|n_i}$ for a certain n_i, and has been obtained by histogram techniques from the figure of merit estimates $\{f'_{1,n_i}, \ldots, f'_{B,n_i}\}$. Choosing a specific value of f, $\hat{p}_{n|f}$ is observed along each vertical line (only lacking the normalization constant A). The more to the left the distribution energy is, the better the classifier performs. The $f - n$ region delimited by the quantile $f \leq f_b^{\min}$ and the confidence interval $q_{n|f}^{0.025} \leq n \leq q_{n|f}^{0.975}$ is indicated by a red line. The value of $\hat{p}_{f|n_i}$ is smaller than -25 dB within this region for all

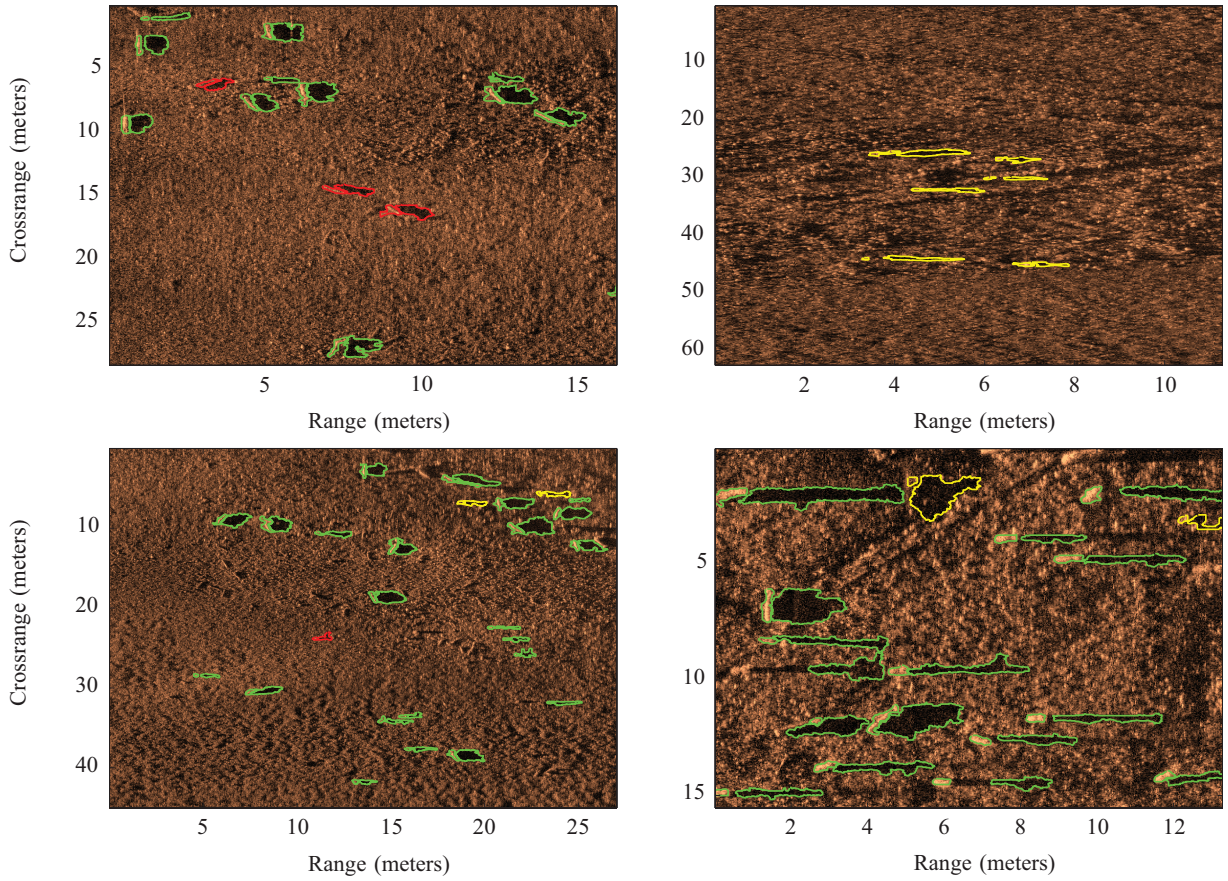


Fig. 8. Illustration of the ADAC system output: directly on the SAS images, the detected objects are indicated. The false alarms are marked in yellow, the detected mines in green, and the missed detected mines in red.

TABLE IV
PERFORMANCE ASSESSMENT Q, WITH $\alpha = 1$, OF THE k-NN CLASSIFIER WITH $k = 5$, MAHALANOBIS' CLASSIFIER, LDA, AND SVM APPROACH. FOR EACH DATA SET, THE BEST RESULT HAS BEEN HIGHLIGHTED

	SAS1	SAS2
k-NN	0.010	0.035
Mahal.	0.007	0.024
LDA	0.016	0.037
SVM	0.014	0.019

examples, and therefore appears white. A different scale is used as the feature selection results are presented (Fig. 6 in Section VI-B), which is more appropriate for studying $\hat{p}_{f|n_i}$ within this region.

The curse of dimensionality [65] is clearly visible for Mahalanobis' classifier for both data sets [see Fig. 5(b) and (f)], i.e., the figure of merit improves as n increases until a certain point and it degrades subsequently. Although the mean of $\hat{p}_{f|n_i}$, $n_i \in \{1, \ldots, N\}$, might decrease with n [e.g., Fig. 5(a)], which suggests that lower f values will be reached at higher n, the variance of $\hat{p}_{f|n_i}$ decreases as well with n. Therefore, the energy concentrated in the left tail of $\hat{p}_{f|n_i}$, where the optimal f^* will eventually be found, is smaller for values of n close to N. For this reason, the confidence intervals indicate that n^* will most probably be placed in the lower n regions.

The values of the quality assessment Q with $\alpha = 1$ [see (4)] for all classifiers and both data sets are included in Table IV. The results are in agreement with the curves in Fig. 5, that is, higher Q values correspond to distributions whose energy is concentrated in smaller values of f. Mahalanobis' classifier shows the worse performance for both examples. On the other hand, the LDA is the best candidate for both databases, although its performance is not significantly better than the SVM for the SAS1 database and the k-NN for the SAS2 example. The LDA classifier has been adopted in the following for both data sets.

B. Feature Selection Algorithms Performance

The D-SFS and D-SFFS algorithms have been employed to estimate the optimal feature subset for both databases. In order to study the influence of the parameter D, the values $D = \{2, 3, 5, 10\}$ have been considered. For comparison, the standard SFS and SFFS are used. Results are included in Fig. 6. Each curve corresponds to a different value of D, and represents the value of f for the best n-feature subset, $\mathbf{t}_n^*$, $1 \leq n \leq n_{\mathrm{M}}$. For illustration, $\hat{p}_{f|n_i}$ has also been included and the most probable region for $\{f^*, n^*\}$ is again delimited by a red line. The scale spans now between -100 dB and 0 dB for all examples. In order to focus on the area of interest, a significant change on the axes with respect to Fig. 5 has also been accomplished.

TABLE V
SAS1 CONFUSION MATRIX. IT CORRESPONDS TO $\{f^*, n^*\} = \{21.6, 72\}$, REACHED BY THE TEN-SFFS ALGORITHM. THE MISSED DETECTED MINES ARE INDICATED IN RED AND THE FALSE ALARMS IN BLUE

	Class 1	Class 2	Class 3
class 1	3478	17	109
class 2	4	240	64
class 3	6	9	114

TABLE VI
SAS2 CONFUSION MATRIX. IT CORRESPONDS TO $\{f^*, n^*\} = \{8.5, 53\}$, REACHED BY THE TEN-SFFS ALGORITHM. THE MISSED DETECTED MINES ARE HIGHLIGHTED IN RED AND THE FALSE ALARMS IN BLUE

	Class 1	Class 2	Class 3	Class 4
Class 1	3604	0	0	0
Class 2	0	38	2	5
Class 3	0	2	63	2
Class 4	0	8	0	65

The maximum dimensionality n_{M} has been limited according to (8). For example, for the five-SFFS algorithm applied to the SAS2 data set [see Fig. 6(d)], $q_{n|f}^{0.975} = 104$ for the smallest achieved value of the figure of merit, $f^* = 11$. Hence, the maximum number of iterations has been constrained to $n_{\mathrm{M}} = 104$.

Fig. 7 summarizes the results for the two data sets. The performance of both the D-SFS and D-SFFS algorithms degrades with respect to the standard SFS and SFFS for $D = \{2, 3\}$. Despite the fact that increasing D leads, in most cases, to an improved performance, it is also possible that the algorithm chooses a combination of features that, although more convenient at a given step, results in a poorer final performance. For $D \geq 5$ though, a significant improvement is observed. The best results are obtained by the ten-SFFS for both examples. The ten-SFFS outperforms the standard SFFS by 12% for the SAS1 data set and by 27% for the SAS2 example. These results are only slightly better than those provided by the five-SFFS, since the performance saturates for $D \gtrsim 5$ in both examples. D-SFFS outperforms D-SFS for a given D, but the five-SFS result is better than the standard SFFS one and it is, moreover, computationally more efficient.

The confusion matrices corresponding to the best f^* for the SAS1 and SAS2 data sets are included in Tables V and VI, respectively. The ten-SFFS algorithm provides the optimal $f^* = 21.6$ at $n^* = 72$ for SAS1. Ten mines, i.e., 2.3% of man-made objects, are missed and 126 false alarms occur, which corresponds to 0.0022 false alarms per square meter. Almost 17% of the mines are assigned to a wrong mine class (cylinders classified as spheres or vice versa), however, this is not penalized for this data set [$\mu = 0$ in (3)]. The SAS2 optimal $f^* = 8.5$ is obtained by the ten-SFFS algorithm at $n^* = 53$ (Table VI). No false alarm occurs and no mine is missed. For the SAS2 database, however, $\mu = 0.5$ and therefore, the selection of the correct mine class is considered. A total of 19 out of 185, that is, about 10% of the man-made objects, are assigned a wrong mine class. The optimal feature sets $\mathbf{t}^*$ for both databases are detailed in Table VII.

It could be thought that the good results obtained for the SAS2 data set are due to strong differences in the statistical properties of the two databases, since the clutter objects utilized by the SAS2 database are borrowed from the SAS1 data set. However, the statistical properties of both data sets have been investigated and no significant difference has been observed. Furthermore, note that only two statistical features are among the 53-element feature subset selected for the SAS2 database. Rather than to differences in the statistical properties of the databases, the good results obtained for the SAS2 database with respect to the SAS1 data set are due to a higher image quality and also to the lack of sand ripples in the surrounding seabed. When segmented, the mine objects result into rather geometrical shapes compared with the more randomly shaped clutter objects.

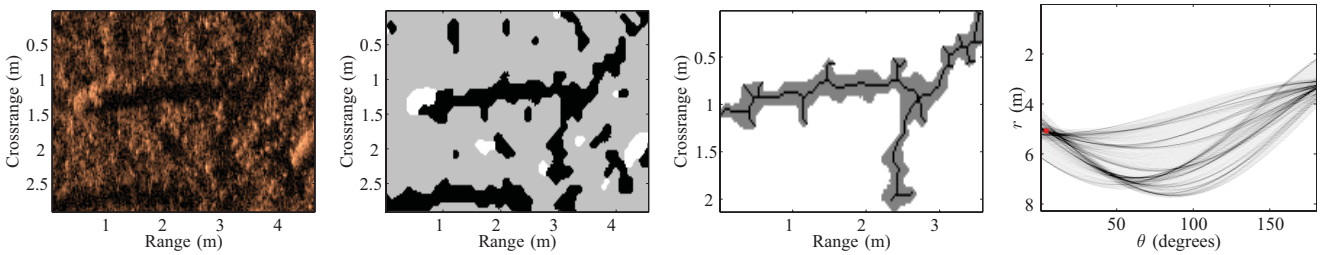
Finally, Fig. 8 includes four snapshots of the SAS1 database where classification results are shown. We indicate the detected mines (both spheres and cylinders) in green, the false alarms in yellow, and in red the missed detected mines. For the sake of clarity, we do not show the clutter that is classified as such. As expected, the false alarms are due to irregular parts of the seabed whose segmented shape is similar to either spheres of cylinders. Two main reasons account for the missed detected mines: either the segmentation is extremely poor, or the intensity of the shadow region is very light with respect to the other mines in the database. In this case, the statistical features cause the object to be classified as clutter.

VII. CONCLUSION

We presented a unified framework for the design of feature-based ADAC systems, and we applied it to the problem of mine hunting for two different databases of SAS images. Previous works on feature-based mine hunting fuse several simple nonoptimal systems in order to improve the performance. By contrast, the approach proposed in this paper pursues the design of a single classification system, which was optimized in two fundamental aspects: the choice of the classification system and the selection of the optimal feature subset. First, the segmentation process was shortly described. A collection of features was designed enforcing the insensibility of the features values to poor segmentation scenarios. Then, a resampling algorithm that provides a quantitative measure of classifier performance independently of any specific feature subset was presented. For the selection of features, we proposed a novel extension of two algorithms, the D-SFS and the D-SFFS, which address the limitations of the standard SFS and SFFS algorithms. Their complexity increases linearly with D. They were tested on 120 synthetic databases and on six standard data sets of the UCI repository. The D-SFS algorithm improved the figure of merit by about 20% with respect to the standard SFS algorithm for both synthetic and real data. In average, the D-SFFS algorithm outperformed the SFFS algorithm by 10% for the real data sets and by 25% for the simulated examples. Furthermore, the average

TABLE VII
SELECTED FEATURES FOR BOTH THE SAS1 AND SAS2 DATABASES. THEY PRODUCE THE CONFUSION MATRIX IN TABLES V AND VI

Feature Type	SAS1	SAS2
Principal Components	ζ_4, ζ_{12}, ζ_{16}, ζ_{17}, ζ_{18}, ζ_{20}, ζ_{23}, ζ_{24}, ζ_{25}, ζ_{27}, ζ_{29}, ζ_{31}, ζ_{32}, ζ_{35}, ζ_{39}, ζ_{41}, ζ_{46}, ζ_{48}, ζ_{50}	ζ_3, ζ_8, ζ_{13}, ζ_{18}, ζ_{19}, ζ_{23}, ζ_{26}, ζ_{32}, ζ_{33}, ζ_{35}, ζ_{36}, ζ_{37}, ζ_{38}, ζ_{39}, ζ_{45}, ζ_{46}
Normalized Central Moments	$\mu_{2,0}$, $\mu_{3,0}$, $\mu_{0,5}$, $\mu_{3,2}$, $\mu_{2,3}$, $\mu_{7,0}$, $\mu_{5,2}$, $\mu_{3,4}$, $\mu_{4,3}$, $\mu_{7,1}$, $\mu_{9,0}$, $\mu_{1,8}$, $\mu_{8,1}$, $\mu_{2,7}$, $\mu_{7,2}$, $\mu_{3,6}$, $\mu_{6,3}$, $\mu_{5,4}$, $\mu_{1,9}$, $\mu_{9,1}$, $\mu_{2,8}$, $\mu_{8,2}$, $\mu_{3,7}$	$\mu_{2,0}$, $\mu_{1,2}$, $\mu_{3,0}$, $\mu_{1,3}$, $\mu_{5,0}$, $\mu_{4,1}$, $\mu_{2,3}$, $\mu_{4,2}$, $\mu_{3,3}$, $\mu_{7,0}$, $\mu_{2,5}$, $\mu_{8,0}$, $\mu_{7,1}$, $\mu_{9,0}$
Invariant M.	γ_5, γ_6, γ_7	γ_2
2-D-Fourier Coefficients	$\phi_{0,5}$, $\phi_{1,0}$, $\phi_{1,4}$, $\phi_{1,5}$, $\phi_{2,0}$, $\phi_{2,2}$, $\phi_{2,4}$, $\phi_{2,6}$, $\phi_{3,0}$, $\phi_{3,3}$, $\phi_{4,4}$, $\phi_{4,5}$, $\phi_{5,0}$, $\phi_{5,2}$, $\phi_{5,5}$, $\phi_{5,6}$, $\phi_{6,1}$, $\phi_{6,5}$	$\phi_{0,6}$, $\phi_{1,0}$, $\phi_{1,1}$, $\phi_{1,3}$, $\phi_{2,1}$, $\phi_{2,2}$, $\phi_{2,4}$, $\phi_{3,3}$, $\phi_{3,5}$, $\phi_{4,6}$, $\phi_{5,0}$, $\phi_{5,3}$, $\phi_{6,4}$
Statistical	ξ_{hlt}, $\Delta\xi_{\text{sdw,bkg}}$, $\Delta\beta_{\text{sdw,hlt}}$, $\Delta\beta_{\text{hlt,bkg}}$	ξ_{bkg}, $\Delta\beta_{\text{sdw,bkg}}$
Geo. sdw	χ, o, Φ	Γ, r_η, Ψ, Φ
Geo. hlt		r_{hlt}
sdw-hlt	$r_{\text{sdw,hlt}}$, $d_{\text{sdw,hlt}}$	$d_{\text{sdw,hlt}}$, Δ_o



(a) SAS image snapshot (b) ICM segmentation result (c) Binary representation of the segmented shadow, **I**, and its skeleton (d) Hough transform of the skeleton

Fig. 9. Measure of the orientation o of the shadow of a spherical man-made object based on the Hough transform of the shadow topological skeleton. The red dot in the Hough transform diagram [Fig. 9(d)] indicates the position of the maximum, i.e., $o = 4$ degrees. Note that a local maximum appears at $\theta = 48$. This corresponds to the orientation of the sand ripple shadow segmented together with the sphere shadow. As long as the sphere shadow is longer than the sand ripple, the absolute maximum of Σ will correspond to the orientation of the sphere shadow and therefore, it will be correctly estimated.

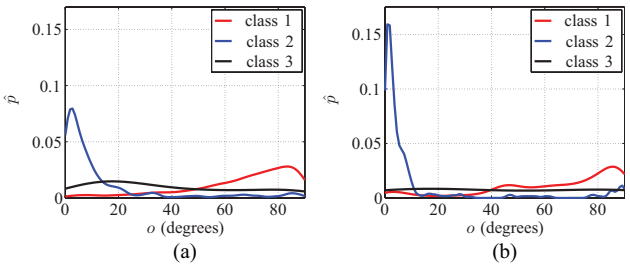


Fig. 10. Estimated distribution of the shadow orientation o (a) when computed directly from the segmented shadow and (b) when measured from the shadow skeleton. The clutter objects correspond to Class 1 (red), while the spherical and cylindrical man-made objects are assigned to Classes 2 (blue) and 3 (black), respectively.

3-SFS algorithm is as good as the SFFS algorithm in terms of figure of merit, and is, furthermore, computationally more efficient.

This unified ADAC design framework was applied to two extensive databases of SAS images with different mine classes. Four classification systems, k-NN, LDA, SVM, and Mahalanobis' classifier were compared using the resampling method. Instead of the misclassification rate, a novel figure of merit was proposed. It considered both the strong imbalance of the classes and the multiclass character of the problem. The LDA provided the best results for both databases. The D-SFS and D-SFFS algorithms were applied so as to obtain the optimal feature set. The ten-SFFS provides the best performance for the SAS1 data set, a missed detection rate of 2.3% with 0.0022 false alarms per square meter. Also for the SAS2 database, the D-SFS and D-SFFS algorithms outperformed the standard SFS and SFFS, respectively. No false alarms occur for the SAS2 data set and, moreover, all mines were detected. However, 10% of them were assigned a wrong mine class.

APPENDIX

In the following, a set of features for characterization of the shadow region is presented. It tries to enhance the accuracy of the feature measurements in situations of poor segmentation.

The orientation of a compact region, o, is normally measured as the angle that the major axis of the region forms with the abscissa of the image. For the application at hand, however, and in order to increase the robustness of the orientation estimation, the following approach has been adopted. Given the binary representation of the segmented shadow, **I**, its topological skeleton Λ [49] is extracted and the Hough transform [66] of Λ, Σ, is calculated. Loosely speaking, the skeleton of a shape corresponds to a thin version of this shape that is equidistant to its contour. Each point of Λ with Cartesian coordinates $\{u, v\}$ contributes to Σ at a certain r

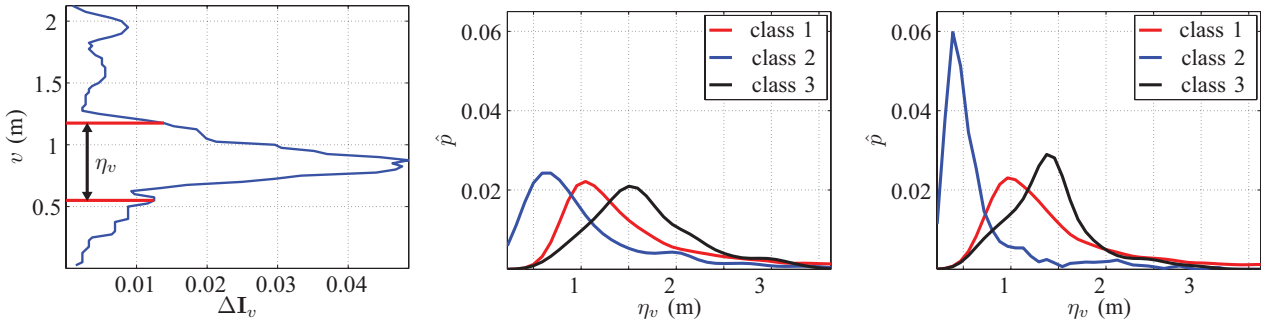


Fig. 11. Crossrange width, η_v. (a) Measure of ηv for the segmented shadow in Fig. 9. (b) pdf of ηv for the different classes in SAS1 if ηv is estimated directly from the segmented shadow. (c) pdf of ηv for the different classes in SAS1 if ηv is based on η_{I_v}

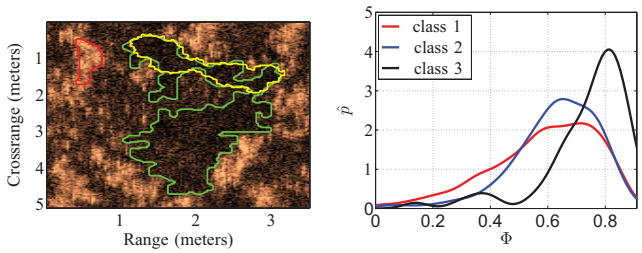


Fig. 12. Segmentation results provided by the ICM (green) and min-cut/max-flow (yellow) algorithms for a clutter object are shown in the left figure. For shadow regions that, like this one, are not well differentiated from the background, the algorithms diverge. The right figure shows the distribution of Φ for the SAS1 database.

(distance to the origin of coordinates) and θ (angle with respect to the abscissa), according to $r = u \cdot \cos\theta + v \cdot \sin\theta$. The range direction corresponds to u and v is parallel to the crossrange direction. We assign

$$o = \arg\max_{\theta}\{\Sigma(r,\theta)\} \tag{10}$$

which estimates the orientation of Λ, and therewith of the shadow region. An example is shown in Fig. 9. The shadow of the sand ripples is segmented together with the shadow of the spherical mine. This is, indeed, a common scenario that represents one of the greatest challenges for ADAC of underwater objects [67]. The proposed skeleton-based approach allows for a correct measure of the shadow orientation. Fig. 10 shows the distribution of the orientation as estimated by the angle of the shadow major axis (left) and as estimated by the proposed skeleton-based method (right). The orientation of clutter and cylindrical objects does not significantly vary. However, the distribution of the orientation for spherical objects shows a higher concentration around 0 degrees for the skeleton-based method, which is in agreement with the fact that shadows of spheres are parallel to the range direction.

The width of the shadow region in both range and crossrange directions, η_u and η_v, respectively, has also been employed. Note that, if directly measured from the segmented shadow, the value of η_v is very sensitive to poor segmentation scenarios. For example, the crossrange width of the segmented shadow $\mathbf{I}$ in Fig. 9(c) is $\eta_v > 2$ m. However, the "real" width is not greater than 0.6 m. The following approach is proposed in order to improve the measurement. The function

$$\Delta \mathbf{I}_v = \frac{1}{\bar{\Delta}} \sum_u \mathbf{I}_{uv} \tag{11}$$

is defined, where $\bar{\Delta} = \sum_u \sum_v \mathbf{I}_{uv}$ is a normalization constant. Basically, $\Delta \mathbf{I}_v$ consists on the cumulative projection of $\mathbf{I}$ on the crossrange axis. We define

$$v_1 = \min\{v\} \quad | \quad \Delta \mathbf{I}_v > \frac{1}{4} \cdot \max\{\Delta \mathbf{I}_v\} \tag{12}$$

$$v_2 = \max\{v\} \quad | \quad \Delta \mathbf{I}_v > \frac{1}{4} \cdot \max\{\Delta \mathbf{I}_v\} \tag{13}$$

and finally, the more accurate estimation of the crossrange width is calculated as $\eta_v = v_2 - v_1$. For the example in Fig. 9(c), $\eta_v = 0.62$. Fig. 11(a) illustrates $\Delta \mathbf{I}_v$, v_1 and v_2. For man-made objects that have been correctly segmented, the proposed technique is practically equivalent to a direct measurement of η_v. The distribution of η_v for the different classes in the SAS1 data set, for both a direct measurement and a measurement based on the cumulative function $\Delta \mathbf{I}_v$, is depicted in Figs. 11(b) and (c), respectively. It can be observed that the crossrange width for the sphere class (Class 2) is significantly narrower when the proposed measure is employed.

The ratio between both widths as a feature, $r_\eta = \eta_v / \eta_u$ and the amount $\Psi = \max\{\Delta \mathbf{I}_v\}/\eta_v$, have proofed as valuable descriptors as well.

Finally, let us introduce a feature that measures the reliability of the segmentation. Taking advantage of the fact that the min-cut/max-flow initialization is based on the ICM segmentation result (see Section II), a similarity measure between both segmentation results constitutes a valuable feature. When the shadow of an object is prominent with respect to the background, the min-cut/max-flow and ICM segmentation results are very similar. On the contrary, if the difference among regions is diffuse, the algorithms tend to provide significantly distinct results. This is prone to happen when the irregularities of the seabed prompt the appearance of clutter. The ratio Φ between the area of the region where both segmentation results intersect and the area of the min-cut/max-flow segmented region constitutes a good measure of the segmentation reliability. Fig. 12(a) shows the ICM segmentation of a clutter shadow and, superimposed, the min-cut/max-flow result. They significantly differ. In Fig. 12(b), the distribution of Φ for the SAS1 data set is depicted. Cylindrical

objects are segmented similarly by both algorithms and hence $\Phi \lesssim 1$ for Class 3 (cylinders). As expected, the value of Φ is lower for Class 1 objects (clutter).

ACKNOWLEDGMENT

The authors would like to thank the company for making available the data and for their support.

REFERENCES

[1] E. Giusti and A. Martorella, and M. Capria, "Polarimetrically-persistent-Scatterer-based automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4588–4599, Nov. 2011.

[2] S. Das, T. T. Mirnalinee, and K. Varghese, "Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3906–3931, Oct. 2011.

[3] R. Paladini, M. Martorella, and F. Berizzi, "Classification of man-made targets via invariant coherency-matrix eigenvector decomposition of polarimetric SAR/ISAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 3022–3034, Aug. 2011.

[4] Z. Jianxiong, S. Zhiguang, C. Xiao, and F. Qiang, "Automatic target recognition of SAR images based on global scattering center model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3713–3729, Oct. 2011.

[5] G. Gao, "An improved scheme for target discrimination in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 277–294, Jan. 2011.

[6] M. Martorella, E. Giusti, A. Capria, F. Berizzi, and B. Bates, "Automatic target recognition by means of polarimetric ISAR images and neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3786–3794, Nov. 2009.

[7] M. F. Doherty, J. G. Landowski, P. F. Maynard, G. T. Uber, D. W. Fries, and F. H. Maltz, "Side scan sonar object classification algorithms," in *Proc. Int. Symp. Unmanned Untethered Submersible Technol.*, 1989, pp. 417–424.

[8] S. G. Johnson and A. Deaett, "The application of automated recognition techniques to side-scan sonar imagery," *IEEE J. Ocean. Eng.*, vol. 19, no. 1, pp. 138–144, Jan. 1994.

[9] G. J. Dobeck, J. C. Hyland, and L. Smedley, "Automated detection and classification of sea mines in sonar imagery," *Proc. SPIE*, vol. 3079, pp. 90–110, Apr. 1997.

[10] C. M. Ciany and W. Zurawski, "Performance of fusion algorithms for Computer Aided Detection and classification of bottom mines in the shallow water environment," in *Proc. MTS/IEEE OCEANS Conf.*, vol. 4. Oct. 2002, pp. 2164–2167.

[11] T. Aridgides, M. F. Fernandez, and G. J. Dobeck, "Side-scan sonar imagery fusion for sea mine detection and classification in very shallow water," *Proc. SPIE*, vol. 4394, pp. 1123–1134, Apr. 2001.

[12] M. P. Hayes, "Synthetic aperture sonar: A review of current status," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 207–223, Jul. 2009.

[13] R. E. Hansen, H. J. Callow, T. O. Sabo, and S. A. V. Synnes, "Challenges in seafloor imaging and mapping with synthetic aperture sonar," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3677–3687, Oct. 2011.

[14] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Hybrid genetic optimization and statistical model based approach for the classification of shadow shapes in sonar imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 2, pp. 129–141, Feb. 2000.

[15] E. Dura, J. M. Bell, and D. M. Lane, "Superellipse fitting for the classification of mine-like shapes in side-scan sonar images," in *Proc. MTS/IEEE OCEANS Conf.*, vol. 1. Oct. 2002, pp. 23–28.

[16] S. Reed, Y. Petillot, and J. Bell, "Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information," *IEE Proc. Radar, Sonar Navigat.*, vol. 151, no. 1, pp. 48–56, Feb. 2004.

[17] J. Groen, E. Coiras, and D. Williams, "Detection rate statistics in synthetic aperture sonar images," in *Proc. Underwater Acoust. Meas. Conf.*, 2009, pp. 367–374.

[18] E. Coiras and J. Groen, "3D target shape from SAS images based on a deformable mesh," in *Proc. Underwater Acoust. Meas. Conf.*, 2009, pp. 303–310.

[19] H. Midelfart, J. Groen, and O. Midtgaard, "Template matching methods for object classification in synthetic aperture sonar images," in *Proc. Underwater Acoust. Meas. Conf.*, 2009, pp. 1–6.

[20] R. Balasubramanian and M. Stevenson, "Pattern recognition for underwater mine detection," in *Proc. CAC/CAD Conf.*, 2001.

[21] J. A. Fawcett, A. Crawford, D. Hopkin, M. Couillard, V. Myers, and B. Zerr, "Computer-aided detection and classification of side scan sonar images from the citadel trial," in *Proc. Inst. Acoust.*, vol. 29. 2007, pp. 3–10.

[22] S. W. Perry and L. Guan, "Pulse-length-tolerant features and detectors for sector-scan sonar imagery," *IEEE J. Ocean. Eng.*, vol. 29, no. 1, pp. 138–156, Jan. 2004.

[23] K. Siantidis and U. Hölscher-Höbing, "A system for automatic detection and classification for a mine countermeasure AUV," in *Proc. Underwater Acoust. Meas. Conf.*, 2009, pp. 1–8.

[24] G. J. Dobeck, "Algorithm fusion for automated sea mine detection and classification," in *Proc. OCEANS Conf.*, vol. 1. 2001, pp. 130–134.

[25] P.-Y. Mignotte, E. Coiras, H. Rohou, Y. Petillot, J. Bell, and K. Lebart, "Adaptive fusion framework based on augmented reality training," *IET Radar, Sonar Navigat.*, vol. 2, no. 2, pp. 146–154, Apr. 2008.

[26] B. Zerr, J. Fawcett, and D. Hopkin, "Adaptive algorithm for sea mine classification," in *Proc. Underwater Acoust. Meas. Conf.*, 2009, pp. 319–326.

[27] D. Williams and J. Groen, "Multi-view target classification in synthetic aperture sonar imagery," in *Proc. Underwater Acoust. Meas. Conf.*, 2009, pp. 699–704.

[28] J. Fawcett, V. Myers, D. Hopkin, A. Crawford, M. Couillard, and B. Zerr, "Multiaspect classification of sidescan sonar images: Four different approaches to fusing single-aspect information," *IEEE J. Ocean. Eng.*, vol. 35, no. 4, pp. 863–876, Oct. 2010.

[29] R. Fandos and A. M. Zoubir, "Optimal feature set for automatic detection and classification of underwater objects in SAS images," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 454–468, Mar. 2011.

[30] R. Fandos, C. Debes, and A. M. Zoubir, "Resampling methods for quality assessment of classifier performance and optimal number of features," *Signal Process.*, 2013, to be published.

[31] A. M. Zoubir and R. Iskander, "Bootstrap methods and applications," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 10–19, Apr. 2007.

[32] R. E. Bellman, *Adaptive Control Processes*. Princeton, NJ, USA: Princeton Univ. Press, 1961.

[33] A. K. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.

[34] A. Frank and A. Asuncion. (2010). *UCI Machine Learning Repository* [Online]. Available: http://archive.ics.uci.edu/ml/

[35] M. Pinto and A. Bellettini, "Shallow water synthetic aperture sonar: An enabling technology for NATO MCM forces," in *Proc. Undersea Defence Technol. Conf.*, 2007.

[36] J. Fawcett, "Image-based classification of sidescan sonar detections," in *Proc. CAC/CAD Conf.*, 2001.

[37] C. Collet, P. Thourel, P. Pérez, and P. Bouthemy, "Hierarchical MRF modeling for sonar picture segmentation," in *Proc. Int. Conf. Image Process.*, vol. 3. 1996, pp. 979–982.

[38] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.

[39] R. Fandos, L. Sadamori, and A. M. Zoubir, "High quality segmentation of synthetic aperture sonar images using the min-cut/max-flow algorithm," in *Proc. Eur. Signal Process. Conf.*, vol. 1. 2011, pp. 51–55.

[40] B. Lehmann, K. Siantidis, and D. Kraus, "Active contours for object characterization in sonar images," in *Proc. Underwater Defence Technol.*, 2011.

[41] R. Fandos and A. M. Zoubir, "Enhanced initialization scheme for a three-region Markovian segmentation algorithm and its application to SAS images," in *Proc. Eur. Conf. Underwater Acoust.*, vol. 3. 2010, pp. 1323–1331.

[42] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[43] F. Langner, C. Knauer, W. Jans, and A. Ebert, "Side scan sonar image resolution and automatic object detection, classification and identification," in *Proc. OCEANS Conf.*, 2009, pp. 1–8.

[44] D. Boulinguez and A. Quinquis, "Classification of underwater objects using Fourier descriptors," in *Proc. Int. Conf. Image Process. Appl.*, 1999, pp. 240–244.

[45] I. Quidu, J. P. Malkasse, G. Burel, and P. Vilbe, "Mine classification using a hybrid set of descriptors," in *Proc. OCEANS Conf.*, vol. 1. 2000, pp. 291–297.

[46] I. Tena Ruiz, D. Lane, and M. Chantler, "A comparison of inter-frame feature measures for robust object classification in sector scan sonar image sequences," *IEEE J. Ocean. Eng.*, vol. 24, no. 4, pp. 458–469, Nov. 1999.
[47] F. Maussang, J. Chanussot, A. Hétet, and M. Amate, "Higher-order statistics for the detection of small objects in a noisy background application on sonar imaging," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, pp. 25–41, 2007.
[48] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Phil. Mag.*, vol. 2, no. 6, pp. 559–572, 1901.
[49] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Boston, MA, USA: Addison-Wesley, 2001.
[50] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Three-class Markovian segmentation of high resolution sonar images," *Comput. Vis. Image Understand.*, vol. 76, no. 3, pp. 191–204, 1999.
[51] A. K. Jain, R. P. W. Duin, and M. Jianchang, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.
[52] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag, 2006.
[53] Q.-L. Tran, K.-A. Toh, D. Srinivasan, K.-L. Wong, and S. Q.-C. Low, "An empirical comparison of nine pattern classifiers," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 35, no. 5, pp. 1079–1091, Oct. 2005.
[54] M. P. Sampat, A. C. Patel, Y. Wang, S. Gupta, C.-W. Kan, A. C. Bovik, and M. K. Markey, "Indexes for three-class classification performance assessment—An empirical comparison," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 3, pp. 300–312, Mar. 2009.
[55] W. A. Yousef, R. F. Wagner, and M. H. Loew, "Comparison of nonparametric methods for assessing classifier performance in terms of ROC parameters," in *Proc. Appl. Imag. Pattern Recognit. Workshop*, 2004, pp. 190–195.
[56] T. M. Cover and J. M. Van Campenhout, "On the possible orderings in the measurement selection problem," *IEEE Trans. Syst., Man Cybern.*, vol. 7, no. 9, pp. 657–661, Sep. 1977.
[57] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.
[58] J. Reunanen, I. Guyon, and A. Elisseeff, "Overfitting in making comparisons between variable selection methods," *J. Mach. Learn. Res.*, vol. 3, pp. 1371–1382, Oct. 2003.
[59] P. Somol, J. Novovicova, and P. Pudil, "Are better feature selection methods actually better?—Discussion, reasoning and examples," in *Proc. HEALTHINF*, vol. 1. 2008, pp. 246–253.
[60] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.
[61] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Trans. Comput.*, vol. 20, no. 9, pp. 1100–1103, Sep. 1971.
[62] T. Marill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE Trans. Inf. Theory*, vol. 9, no. 1, pp. 11–17, Jan. 1963.
[63] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Hoboken, NJ, USA: Wiley, 2003.
[64] C. C. Chang and C. J. Lin. (2001). *LIBSVM: A Library for Support Vector Machines* [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[65] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
[66] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1989.
[67] D. P. Williams and E. Coiras, "On sand ripple detection in synthetic aperture sonar imagery," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Jan. 2010, pp. 1074–1077.

Raquel Fandos received the M.Eng. degree from the University of Zaragoza, Zaragoza, Spain, in 2004. She is currently pursuing the Ph.D. degree with the Technische Universität of Darmstadt, Darmstadt, Germany.

She visited KTH University, Stockholm, Sweden, for one year. In 2005 and 2008, she was with CERN, Geneva, Switzerland, as an RF Engineer for the CLIC project. Her current research interests include automatic detection and automatic classification, specially of underwater objects for mine hunting applications.

Abdelhak M. Zoubir (F'86) received the Dr.-Ing. Degree from Ruhr-Universität Bochum, Bochum, Germany, in 1992.

He was with Queensland University of Technology, Brisbane, Australia, from 1992 to 1998, where he was an Associate Professor. In 1999, he joined Curtin University of Technology, Perth, Australia, as a Professor of telecommunications and was an Interim Head of the School of Electrical and Computer Engineering from 2001 to 2003. In 2003, he moved to Technische Universität Darmstadt, Darmstadt, Germany, as a Professor of signal processing and the Head of the Signal Processing Group. He has published over 300 journal and conference papers. His current research interests include statistical methods for signal processing with emphasis on bootstrap techniques, robust detection and estimation and array processing applied to telecommunications, radar, sonar, automotive monitoring and safety, and biomedicine.

Dr. Zoubir is a Fellow of the IEEE Distinguished Lecturer (Class from 2010 to 2011). He was the Technical Chair of the 11th IEEE Workshop on Statistical Signal Processing in 2001, General Co-Chair of the 3rd IEEE International Symposium on Signal Processing and Information Technology in 2003 and of the 5th IEEE Workshop on Sensor Array and Multi-channel Signal Processing in 2008. He is the General Co-Chair of the 14th International Workshop on Signal Processing Advances in Wireless Communications in 2013 to be held in Darmstadt, Germany, the General Co-Chair of the 21st European Signal Processing Conference in 2013 to be held in Marrakech, Morocco, and the Technical Co-Chair of ICASSP-14 to be held in Florence, Italy. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2005, a Senior Editorial Board member of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING from 2009 to 2011 and he currently serves on the Editorial Boards of the *EURASIP Journals Signal Processing* and the *Journal on Advances in Signal Processing*. He currently serves a three-year term as an Editor-in-Chief of the IEEE SIGNAL PROCESSING MAGAZINE from 2012 to 2014. He was a Past Chair in 2012, a Chair from 2010 to 2011, a Vice-Chair from 2008 to 2009, and a member of the IEEE SPS Technical Committee Signal Processing Theory and Methods from 2002 to 2007. He was a member of the IEEE SPS Technical Committee Sensor Array and Multi-channel Signal Processing from 2007 to 2012. He serves on the Board of Directors of the European Association of Signal Processing.

Konstantinos Siantidis received the Diploma and Ph.D. degrees in theoretical nuclear and theoretical solid state physics from the Westfälische Wilhelms University of Münster, Münster, Germany, in 1996 and 2002, respectively.

He joint ATLAS ELEKTRONIK GmbH, Bremen, Germany, in 2003, where he is currently with the Applied Research Department with focus on automatic object recognition in sonar images. From 2004 to 2006, he was an Adjunct Professor with the University of Applied Science, Bremen. His current research interests include statistical signal and image processing, data fusion, advanced autonomy, and the applications of such methods to automatic data analysis in autonomous underwater vehicles.