Model-based approach to the detection and classification of mines in sidescan sonar

Scott Reed, Yvan Petillot, and Judith Bell

This paper presents a model-based approach to mine detection and classification by use of sidescan sonar. Advances in autonomous underwater vehicle technology have increased the interest in automatic target recognition systems in an effort to automate a process that is currently carried out by a human operator. Current automated systems generally require training and thus produce poor results when the test data set is different from the training set. This has led to research into unsupervised systems, which are able to cope with the large variability in conditions and terrains seen in sidescan imagery. The system presented in this paper first detects possible minelike objects using a Markov random field model, which operates well on noisy images, such as sidescan, and allows *a priori* information to be included through the use of priors. The highlight and shadow regions of the object are then extracted with a cooperating statistical snake, which assumes these regions are statistically separate from the background. Finally, a classification decision is made using Dempster–Shafer theory, where the extracted features are compared with synthetic realizations generated with a sidescan sonar simulator model. Results for the entire process are shown on real sidescan sonar data. Similarities between the sidescan sonar and synthetic aperture radar (SAR) imaging processes ensure that the approach outlined here could be made applied to SAR image analysis.

OCIS codes: 100.0100, 330.1880.

1. Introduction

Rapid developments in autonomous underwater vehicle technology have altered the direction of mine-counter measures research toward more automated techniques.[1–3] These techniques generally require training, and their success can be dependent on the similarity between the training and test data sets. The approach detailed here uses a three-phase process. The first phase employs model based on a Markov Random Field (MRF) model to directly segment the raw image into regions of object highlight, shadow, and background. Unlike many previous detection models,[4,5] this one requires no training. The MRF structure of the model also allows known information to be modelled and included through the use of priors.[6]

The second phase uses a cooperating statistical snake (CSS) model[7] to consider each of the detected minelike objects (MLOs). This model was originally developed to ensure the accurate segmentation of object shadow regions onto complex seabeds, such as sand ripples, when other models failed.[8] The model segments both the highlight and shadow regions of the object by assuming the regions to be statistically separate, thereby enforcing a dependency between the two snakes and constraining their movement. The CSS model is also effective in identifying false alarms.

The third and final phase of the procedure entails classification of the MLO by use of a sonar simulator model. By using the known range and height of the MLO, one can iteratively produce synthetic presentations of possible objects (restricted to cylinders, spheres, and truncated cones here). A Dempster–Shafer (DS) approach[9–11] is used to assign a belief to each of the possible classes, taking into account the degree of match (using the Hausdorff distance[12]) and the plausibility of the synthetic object's parameters. This novel approach extends the traditional mine or not-mine classification to provide useful shape and size information. The DS framework also permits multiview analysis. This is important for sidescan surveys, which generally use a lawn-mower ap-

The authors are with the Ocean Systems Laboratory, School of Engineering and Physical Sciences, Heriot-Watt University, Riccarton Campus, Edinburgh, EH14-4AS, United Kingdom. S. Reed's e-mail address is S.Reed@hw.ac.uk.

Received 11 April 2003; revised manuscript received 29 July 2003; accepted 27 August 2003.

0003-6935/04/020237-10$15.00/0

proach, often covering the same area of seabed multiple times.

In Sections 2 and 3 we review the detection and CSS models previously presented by Reed *et al.*[7] In Section 4 we present the results for these first two phases. The classification theory and results from both the monoview and multiview analyses are discussed in Section 5. In Section 6 we present our conclusions.

2. Detection Model

A. MRF Theory

General MRF models are composed of two fields: the observed image Y and the underlying true label field X. A pixel s is assigned a label x_s based on two probability measurements. The first of these considers the probability that label x_s will produce observation y_s. The second considers the labels of the neighboring pixels. This interspatial dependency among pixels has ensured the successful use of MRF models in a variety of difficult segmentation problems.[13–15]

We consider a set of three random fields $Z = (X, Y, O)$. Field $Y = \{Y_s, s \in S\}$ is the field of gray-level observations and thus takes its values from the gray-level range $\{0 \ldots 255\}$. Label field $X = \{X_s, s \in S\}$ is the underlying label field that we wish to recover with the segmentation. Label X_s can take values $\{e_0 = \text{shadow}, e_1 = \text{seabottom reverberation}, e_2 = \text{object highlight}\}$. Field $O = \{O_s, s \in S\}$ is defined as the object field; O_s is drawn from $\{o_0 = \text{object}, o_1 = \text{nonobject}\}$ and is determined directly by consideration of label field X. Label field O, therefore, shows the clustering of object pixels. The probability of the unobservable true data given the observed field Y can be expressed by use of Bayes theorem as follows:

$$P_{X,O|Y}(x, o|y) \propto P_X(x) P_{O|X}(o|x) P_{Y|X}(y|x). \quad (1)$$

By expressing the posterior distribution as $P_{X,O|Y}(x, o|y) \propto \exp\{-U(x, y, o)\}$,[16] the desired underlying label field can be obtained by minimizing posterior energy:

$$U(x, y, o) = \sum_{s \in S} \Phi_s(x_s, y_s) + \sum_{\langle s,t \rangle} \beta_{st}[1 - \delta(x_s, x_t)] - \sum_{s \in S} \delta(x_s, e_2) \ln \Psi_X(s) - \sum_{s \in S} \chi_s(x_s, o_s). \quad (2)$$

The first and second terms on the right-hand side correspond to the likelihood and Markovian terms used in general MRF models. The third and the fourth terms incorporate some of the knowledge about the appearance of objects in sidescan imagery. The third term acts only on pixels with label $x_s = e_2$(object highlight). A directional potential field generated by pixels labeled $x_s = e_0$(shadow) discourages pixels far from a shadow region from being labeled $x_s = e_2$(object highlight). This uses an adapted potential term[6] and models the *a priori* information that a mine highlight region usually has a corresponding shadow region (see Fig. 1). The fourth term considers field O and favours clustering of $x_s = e_2$(object highlight) in compact, separated clusters of minelike dimension. (See Reed *et al.*[7] for a complete explanation of the detection model.)

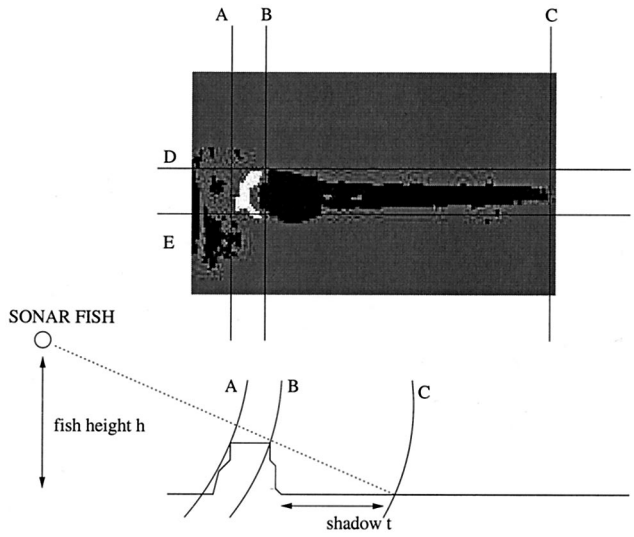

Fig. 1. As in radar, the wave (here it is sound) is blocked by objects and a shadow is generated. Given the relative position of the sonar fish, an estimate of the object's height and size can therefore be obtained. This information can then be used to remove obvious false alarms.

B. Postprocessing Phase

The detection-orientated segmentation detailed in Subsection 2.A highlights possible MLOs. Some of these will be obvious false alarms and can be removed. First, object highlight regions that are obviously too large or small can be removed (see area enclosed within ABDE in Fig. 1). The height of the MLO can also be computed and used by consideration of the length of the accompanying shadow region t, along with navigational data such as the sonar fish height h. This method is useful in the removal of false alarms produced by complex sea floors.

3. Cooperating Statistical Snake Model

A. CSS Model Theory

The CSS model[7] was developed in response to the failure of conventional techniques[8,17] to extract the MLOs shadow region on complex seafloors. The CSS model extracts both the object highlight and shadow regions. We consider the detected MLO's mugshot image $\mathbf{y} = y(i, j)$ with its corresponding template image $\mathbf{w} = w(i, j)$, the latter of which defines the shape of the two snakes at any given time. Indices i and j represent pixel positions in the image. It is assumed that the image is composed of object-highlight, object-shadow, and background regions, described by $\Omega_h = \{(i, j)|\omega(i, j) = 2\}$, $\Omega_p = \{(i, j)|\omega(i, j) = 1\}$ and $\Omega_b = \{(i, j)|\omega(i, j) = 0\}$, respectively. All three regions are described by probability density functions (pdf's) p^{μ_h}, p^{μ_p}, and p^{μ_b}, where μ_h, μ_p, and

μ_b are the parameters of the three pdf's. If the probability densities are assumed to be Gaussian,[18] then the segmentation $\mathbf{w}$ can be obtained by maximizing

$$l_{\text{gauss}}(\mathbf{y}, \mathbf{w}) = -N_h \log(\theta_h) - N_p \log(\theta_p) - N_b \log(\theta_b), \tag{3}$$

where

$$\theta_u = \frac{1}{N_u(\mathbf{w})} \sum_{(i,j)\in\Omega_u} y(i,j)^2 - \left\{ \frac{1}{N_u(\mathbf{w})} \sum_{(i,j)\in\Omega_u} y(i,j) \right\}^2 \tag{4}$$

and $N_u(\mathbf{w})$ is the number of pixels in region u and $u \in \{h, p, b\}$. The log-likelihood term in Eq. (3) is simply a function of gray-level sums and thus can be rewritten as a summation around the perimeter of the two snakes.[7,18] This ensures that the iterative process required for finding the most likely solution is computationally fast.

B. Incorporating *A Priori* Information

The *a priori* knowledge that the object-highlight region is generally much brighter than the object-shadow region can be modeled using prior term

$$\log[P_{\text{mean}}(\mathbf{w})] = \mu \tanh\left(\frac{1}{2} \alpha(m_\Delta - \beta) \right) + c, \tag{5}$$

where m_Δ is the difference in mean gray level among the pixels within each of the two snakes. The other parameters are constants that control the dynamic range and crossover rate of the function. The flat-topped form of the tanh function prevents the snakes from simply collapsing, thereby ensuring a high m_Δ value. As seen in Fig. 1, the object-highlight region and corresponding object-shadow region must also have roughly the same along-track size (enclosed by D and E). We define $\Delta_{\max} = \max\{i_h\} - \max\{i_p\}$ and $\Delta_{\min} = \min\{i_h\} - \min\{i_p\}$, where i_h and i_p are the y coordinates on the perimeter of the object-highlight and object-shadow snakes, respectively. The height prior term can then be defined as

$$\log[P_{\text{position}}(\mathbf{w})] = C - t_1 U(|\Delta_{\max}| - \zeta)|\Delta_{\max}|^2 - t_2 U(|\Delta_{\min}| - \zeta)|\Delta_{\min}|^2, \tag{6}$$

where t_1 and t_2 are constants that penalize large differences in $\Delta_{\max}$ and $\Delta_{\min}$. The constant C ensures that the prior operates in the correct dynamic range. U is the Heaviside function, which allows the snakes some flexibility of movement, where constant ζ is set arbitrarily to a small, nonzero number.

The final posterior energy to be maximized during the segmentation process is

$$J(\mathbf{y}, \mathbf{w}) = \lambda_0 \log[P_{\text{reg}}(\mathbf{w})] + (1 - \lambda_0) \times \{\lambda_1 \log[P_{\text{position}}(\mathbf{w})] + \lambda_2 l(\mathbf{y}, \mathbf{w}) + (1.0 - \lambda_1 - \lambda_2) \log[P_{\text{mean}}(\mathbf{w})]\}, \tag{7}$$

where $\log[P_{\text{reg}}(\mathbf{w})]$ is a smoothing prior[18] and λ_k $k \in \{0, 1, 2\}$ are weights used to control the importance of each term. Here λ_1 and λ_2 are set at 0.2 and 0.6, respectively, and λ_0 is incremented gradually throughout the process.[7]

C. Initializing the CSS Model

The detection result from Section 2 ensures a good initialization for the CSS model. Both snakes are initialized as rectangles with only four points. The maximum and minimum rows and columns for each object were used to define the object-highlight rectangle. The object-shadow snake was initialized by considering the homogeneity of the e_0(shadow) pixels within it, whereas the height of the object-shadow rectangle was set to that of the object-highlight rectangle.

D. Removing False Alarms with the CSS Model

Complex backgrounds can provide situations in which the MRF-based detection model falsely identifies a MLO that the postprocessing phase does not remove. In these situations the CSS model can often identify the false alarm. The CSS model operates on the assumption that there are three distinct statistical regions (object highlight, object shadow, and background). When an object is present, the prior $\log[P_{\text{mean}}(\mathbf{w})]$ prevents the object-highlight snakes from expanding. False alarms that do not have these three distinct distributions often result in an uncontrolled expansion of the snakes. If the snakes expand beyond minelike dimensions, the MLO can be identified as a false alarm and can be removed.

4. Detection and Shadow Extraction Results

The combination of the detection and CSS models was tested on more than 200 sidescan images. These images were obtained from the BP'02 (battle preparation) trials conducted at the NATO Saclant Underwater Research Centre in Italy. This large database of images ensured that the model was tested on a large variety of terrains under dramatically different conditions. Of the 200 images, 70 objects were marked by human operators as possible MLOs, many of which were the same object seen from different views. The model succeeded in detecting 56 of these isolated objects, resulting in an 80% detection rate. Many of the objects that were not detected were removed because of the presence of a watermark in many of the images, a phenomenon that is due to the sea-surface return when the sonar fish lies close to the surface. When this watermark corrupts the MLO's shadow, the object-height calculation is affected, which results in the object's removal. Manual removal of the watermark resulted in a detection rate of 91%. Future research will explore ways to automate this process. The detection-CSS model detected 55 false alarms, usually due to the presence of complex seabed types. This resulted in an average of 0.275 false alarms or images. The detection-CSS model is demonstrated in Fig. 2, which illustrates a complex example in which some of the objects are lying on sand ripples. The process is

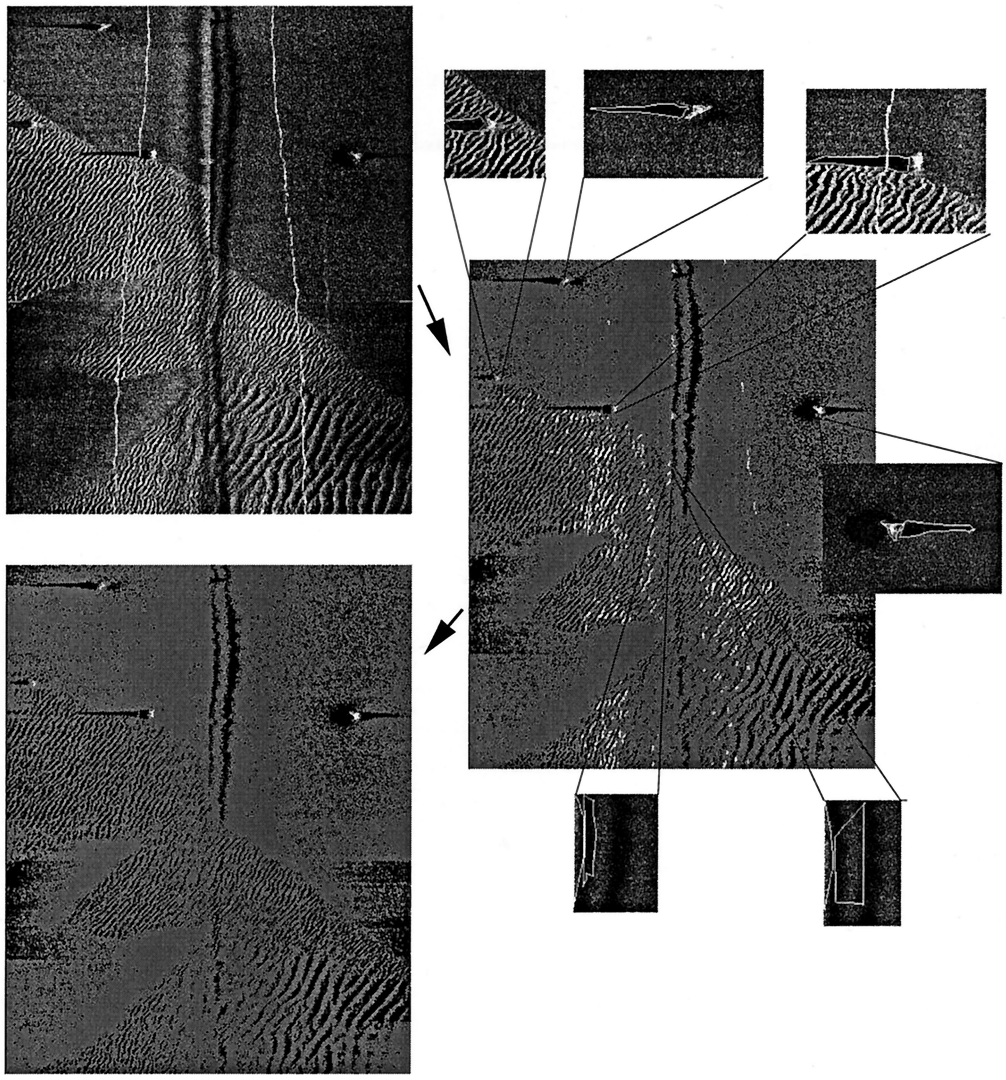

Fig. 2. Detection CSS model result for an image containing objects hidden within the sand ripple seafloor. The first image is the raw sidescan image. The second image contains the MRF detection result before the post-processing or CSS stages. The third image shows the detection result obtained. Accurate shadow segmentation results for all objects have been obtained due to the constrained movement of the CSS model.

explained by use of three images; arrows indicate the image order. The first image contains the raw sonar image. The second image contains the initial MRF detection result. The third image contains the completed detection result after the postprocessing and CSS processes have been completed. All images considered were 1024 $\times$ 1000 pixels in size. Figure 2 shows that the correct detection results are obtained and that the object shadows are accurately extracted even when they are corrupted by shadows from the ripples. This was possible because of the constraining behavior of the CSS model. All the false alarms are removed during the postprocessing and CSS stages. It should be noted that only a few of the false alarms detected by the CSS model have been shown. The computation time for this result was 163 s on a Pentium 4 1.3-GHz personal computer. A simpler sonar image would require substantially less analysis time.

5. Classification of Minelike Objects

Man-made objects generally produce more regularly shaped shadows than do natural objects. This characteristic can be used to classify an unknown MLO. Previous research has focused on template-matching models,[17,19] which attempt to fit shadow templates of possible man-made objects to the MLO's shadow. Although these approaches have yielded good results, they generally do not take into account the underlying sonar shadow formation process, that is, the plausibility of the tested templates is generally not considered during the testing process. The method detailed here, however, uses a sonar simulator model to produce synthetic shadow representations from considered classes (limited here to cylinders, spheres, and truncated cones). Sonar simulator models have been used in the past to train feature-based supervised classification models.[20] Here the simulator

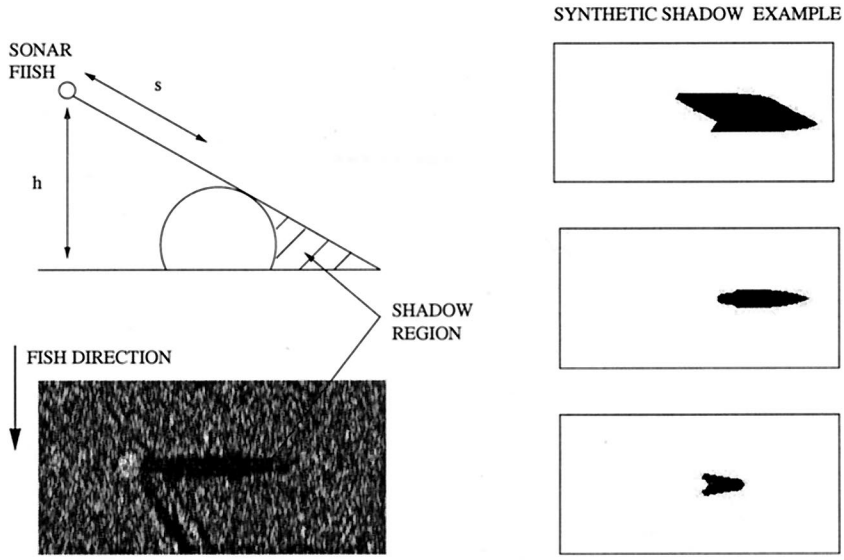

Fig. 3. Description of the shadow formation process. Individual pings are added together to form an overall sonar image (bottom left-hand side). Synthetic shadow representations from the three considered classes are also shown on the right-hand side: cylinder (top), sphere (middle), and truncated cone (bottom). These can be compared with the real sonar shadow to find a match.

model is used within the classification process. The best match from each class is considered by a DS model. Although most models produce a hard, classification result, the DS approach allocates to each class a belief that can be updated as more information (e.g., another view) is made available.

The classification decision uses predominantly shadow information. The highlight region can be very dependent on the MLO's composition, where complex elastic scattering effects can significantly alter the highlight's appearance. These effects are very difficult to incorporate accurately into a simple simulator model. As a result, only the elongation of the highlight region is considered during the classification phase. This increases the separability between the classes, as discussed later in Subsection 5.C. The shadow region is related predominantly to the shape of the object so that information regarding the object's composition is not required. In addition, because the classification process is iterative, simulation of the shadow alone significantly decreases the computational effort.

A. Generating Synthetic Shadows

The shadow-forming process in sidescan imagery is summarized in Fig. 3, which shows the cross section of a spherical object at slant range s and depth h from the sonar fish. As shown, the sonar cannot reach the region of seafloor behind the object. As the sonar fish moves along, successive pings are placed together until a complete sonar image is formed. Figure 3 also shows examples of synthetically generated shadow regions from the cylinder, sphere, and truncated cone classes. Object shadows from each of the classes can be generated by consideration of a simple ray-tracing model. This model is an approximation from an existing sonar simulator,[21] which assumes isovelocity conditions and a simple point source receiver, where the source and receiver are colocated. Compensated for effects, such as the beam pattern, is assumed by the model. This reduction in model complexity is appropriate when only the generated shadow is of interest rather the full backscattered signal, which would require a more complicated simulator. The cylinder, truncated cone, and sphere classes are assumed to be completely described by parameters $\Theta_{\text{cyl}} = \{r_{\text{cyl}}, l_{\text{cyl}}, d_{\text{cyl}}, \phi_{\text{cyl}}\}$, $\Theta_{\text{sph}} = \{r_{\text{sph}}, d_{\text{sph}}\}$ and $\Theta_{\text{cone}} = \{r_{\text{cone}}, d_{\text{cone}}\}$, respectively, where l is the length, r is the radius and d is the depth of each object. The angle of the cylinder with respect to the along-track direction is represented by ϕ_{cyl}. With these parameters synthetic shadow representations from these object classes can be generated under the same sonar conditions (sonar fish height, range, resolution) that the MLO was detected.

B. Comparing Shadow Regions

The Hausdorff distance[12] is a technique that measures the resemblance between two shapes. If $A = \{a_1, \ , a_p\}$ and $B = \{b_1, \ , b_q\}$ are defined as the points on the perimeter of the real and the synthetic shadow regions, respectively, then the Hausdorff distance is defined as

$$H(A, B) = \max[h(A, B), h(B, A)] \tag{8}$$

where

$$h(a, b) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{9}$$

and $\| \cdot \|$ is some underlying norm on the points of A and B.

Function $h(A, B)$ is a directed Hausdorff distance and is computed by first calculating the distance between each point in A to its nearest neighbour in B.

The maximum of this set of values is $h(A, B)$. Therefore if $h(A, B) = d$, each point in A must be within distance d from some point in B. A similar process is carried out to compute the second directed Hausdorff distance $h(B, A)$. The Hausdorff distance $H(A, B)$ is designated the maximum value of the two directed distances $h(A, B)$ and $h(B, A)$. $H(A, B)$ is therefore a measure of mismatch between A and B.

Parameters Θ_{cyl}, Θ_{sph}, and Θ_{cone} were iteratively changed so that the best match—and the smallest Hausdorff distance $H(A, B)$—could be found for each class. Initial parameter estimates for the iterative search were provided by applying moment analysis on the MLO's extracted highlight region obtained by the CSS model. It should be noted that this initialization step was possible only because the CSS model extracted both the MLO's highlight and shadow regions. Large margins were set on each of the initial parameter estimates to define a discretized parameter space for each class that must be searched to find the best fit to the MLO's shadow. The parameter space for both Θ_{sph} and Θ_{cone} was two dimensional, allowing an exhaustive search to be used. The parameter space for Θ_{cyl} was four dimensional and required Monte Carlo Markov chain techniques[8] to ensure a good maximum estimate. Estimates for the best Hausdorff distance for each class took approximately 60 s to compute on a Pentium 4 1.3-GHz personal computer.

C. Obtaining Class Membership Functions

Assuming that the best Hausdorff solution H_j for each class $j \in \{\text{cyl, sph, cone}\}$ was obtained with object parameters Θ_j^b and that the highlight region extracted from the MLO had elongation α, an overall class membership function can be defined by

$$\omega_j^{\text{final}}(H_j, \Theta_j^b, \alpha) = \omega_j^{\text{haus}}(H_j)\omega_j^{\text{par}}(\Theta_j^b)\omega_j^{\text{elong}}(\alpha). \quad (10)$$

Function $\omega_j^{\text{haus}}(H_j)$ considers the best Hausdorff distance value for each class. The shape of this function was determined for each class by training. The data used for training and testing were taken from two different data sets, each under very different sonar conditions. The training and testing data sets were completely disjointed, with no overlap. On the objects of known class j, a perfect sonar simulator would be expected to produce synthetic shadow representations with $H_j \approx 0$. However, the ray-tracing model used here produced roughly Gaussian distributions around nonzero Hausdorff values, leading to the function

$$\omega_j^{\text{haus}}(H_j) = 1 \quad \text{if } H_j \le \bar{m}_j$$
$$= \exp\left[-\frac{(H_j - \bar{m}_j)^2}{2\sigma_j^2}\right] \quad \text{if } H_j > \bar{m}_j,$$

where $\bar{m}_j$ and σ_j^2 are the mean and variance values of the class Gaussian distributions, respectively. *A priori* information on the shape and expected dimensions of minelike objects was introduced through the use of $\omega_j^{\text{par}}(\Theta_j^b)$. This function used simple trapezium fuzzy functions to allocate high values ($\approx$1) to parameters believed to be minelike and low values ($\approx$0) to the others. The trapezium functions were made suitably broad to ensure that objects with parameters relatively close to minelike dimensions could still achieve a high membership function.

Function $\omega_j^{\text{elong}}(\alpha)$ considered the elongation of the highlight region. MLO's with high elongation are more likely to belong to the cylinder class, whereas MLOs with low elongation probably belong to either the sphere or truncated-cone class. These functions were again simple trapezium functions determined from the training data, where $\omega_{\text{sph}}^{\text{elong}}(\alpha) = \omega_{\text{cone}}^{\text{elong}}(\alpha)$.

The overall membership function $\omega_j^{\text{final}}(H_j, \Theta_j^b, \alpha)$ lay in the [0, 1] range. These membership functions were used within a DS model to provide a classification decision. The use of fuzzy functions within a DS framework was chosen over other classification models for a variety of reasons. First, the limited amount of data available for training made methods such as the K-NN (nearest neighbor) classifier or a neural-network-based approach difficult. Second, because the number of parameters varied with class, a fuzzy classifier model seemed simpler to implement than a clustering model. The simplicity of the sonar simulator model also required that functions $\omega_j^{\text{par}}(\Theta_j^b)$ be robust enough to cope with inaccuracies inherent to the simple ray-tracing assumptions used in the simulator. An improved sonar simulator model would perhaps permit a more rigorous clustering approach to the classification but at the expense of slowing the iterative process.

D. Dempster–Shafer Model

Dempster–Shafer theory,[10] frequently used as an alternative to Bayesian theory[22] and fuzzy logic[23] for data fusion, allows the representation of imprecision and uncertainty through the definition of two functions: plausibility (Pls) and belief (Bel). These are derived from a mass function m, which is analogous to the well-known probability density function. Mass functions are defined on the power set of the space of discernment D. For classification purposes, D may be the set of possible classes. Specific to DS theory, D may also contain union of classes.[11] Denoting 2^D as the power set of D, we can define mass function $m(A)$ for every element A of 2^D such that

$$m(\emptyset) = 0, \qquad \sum_{A \subset 2^D} m(A) = 1. \quad (11)$$

For the mine classification model presented here, the allowed classes were $A = \{\text{clutter, cyl, sph, cone, sph} \cup \text{cone}\}$. Class $A = \text{sph} \cup \text{cone}$ was used to model confusion between the sphere and cone classes. The mass functions were generated from the class membership functions $\omega_j^{\text{final}}(H_j, \Theta_j^b, \alpha) j \in \{\text{cyl, sph, cone}\}$. Given the set of mass functions, the belief (Bel) and

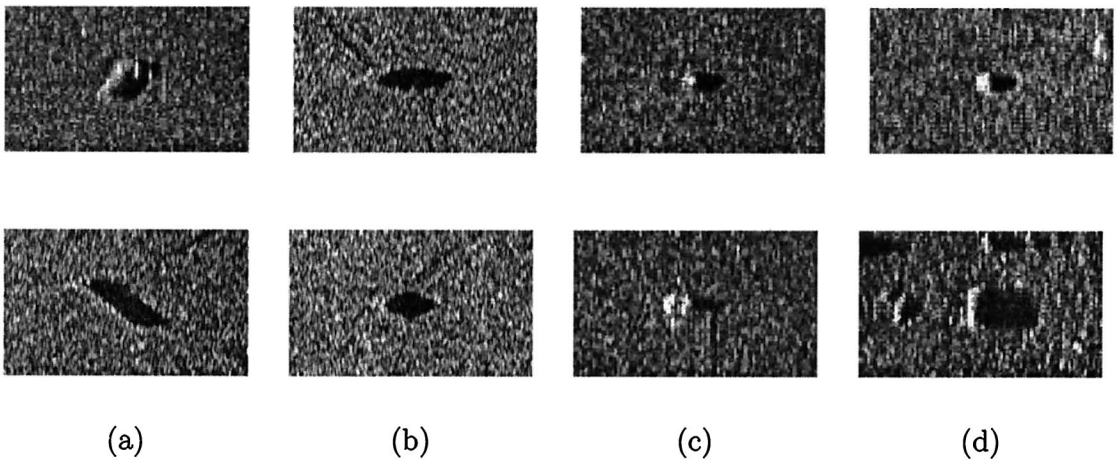

Fig. 4. Examples of (a) cylinder, (b) sphere, (c) truncated cone, and (d) clutter objects used to test the classification model.

plausibility (Pls) of each possible outcome can be determined by

$$\mathrm{Bel}(\emptyset) = 0,$$
$$\mathrm{Bel}(A) = \sum_{B \subset A} m(B), \qquad \forall A \subset D, A \neq \emptyset, \quad (12)$$

$$\mathrm{Pls}(\emptyset) = 0,$$
$$\mathrm{Pls}(A) = \sum_{B \cap A \neq \emptyset} m(B), \qquad \forall A \subset D, A \neq \emptyset. \quad (13)$$

A decision can be made by considering either the plausibility or belief function. In this paper, the MLO was allocated the singleton class (i.e., the MLO cannot be classified as sph ∪ cone) with the maximum belief.

E. Multiview Classification

When the same object has been viewed from multiple aspects, DS theory allows the monoview mass functions to be combined. Considering mass function m_k from source k ($k = 1, \ldots, n$), this rule is expressed by

$$(m_1 \oplus m_2 \oplus \ldots \oplus m_n)(A) = \frac{\sum_{B_1 \cap \ldots \cap B_n = A} m_1(B_1) m_2(B_2) \ldots m_n(B_n)}{1 - \sum_{B_1 \cap \ldots \cap B_n = \emptyset} m_1(B_1) m_2(B_2) \ldots m_n(B_n)} \quad (14)$$

for all nonempty subsets A of D. The summation on the bottom line of Eq. (14) is often referred to as the conflict and is ≤ 1. A value of 1 means that the evidence from the sources are completely conflicting and thus cannot be fused. Once the fused mass functions have been determined, the belief (Bel) functions can be determined as in Eq. (12), and a multiview classification result can be obtained.

F. Results

The classification model was first tested on monoimage cases. This assumes that each MLO was detected only once. The test data were provided by DRDC–Atlantic in Canada and Groupe de Etudes Sous-Marine l'Atlantique (GESMA) in France. The DRDC-Atlantic data were collected with a Klein 5500 multibeam 550-kHz high-resolution sidescan sonar. The images considered were 90 × 60 pixels in size and had been processed to ensure that both the across and along-track resolution were 0.10 m. The data provided from GESMA were obtained with a DF1000 dual-frequency sidescan sonar and provided images with a resolution of 0.03 m × 0.03 m. Examples from the four considered classes (cylinders, spheres, truncated cones, and clutter) can be seen in Fig. 4. As shown, the clutter objects are often visually very similar to the man-made objects.

The classification model was tested on 50 objects: 10 each from the cylinder, truncated cone, and cylinder classes and 20 clutter objects. The testing data were completely disjointed from the training data. The results are shown in Fig. 5. The solid curve represents a standard mine–not-mine classification process. The model correctly identified more than 90% of the mines and correctly classified approximately 50% of the clutter objects. The dashed curve denotes a more specific classification in which the classification is deemed correct only if the mine's class was also successfully identified. Under these constraints, more than 80% of the mines were still correctly classified when the same clutter classification rate was maintained. Obtaining the correct shape classification (as well as the parameter information) makes it possible to identify the mine type and thus affects how the specific threat is handled. The difference in the two classification performances is due to cases in which the sonar conditions are such that the shadow regions from all classes are very small and similar, leading to confusion between the classes. Better separability between classes should be achieved with a more complex sonar simulator.

The multiview classification model is demonstrated on two different objects. The first is a cylindrical object seen from four different views. The second object is a truncated cone, which has also been detected in four different passes. The different views can be seen in Fig. 6. The observed difference

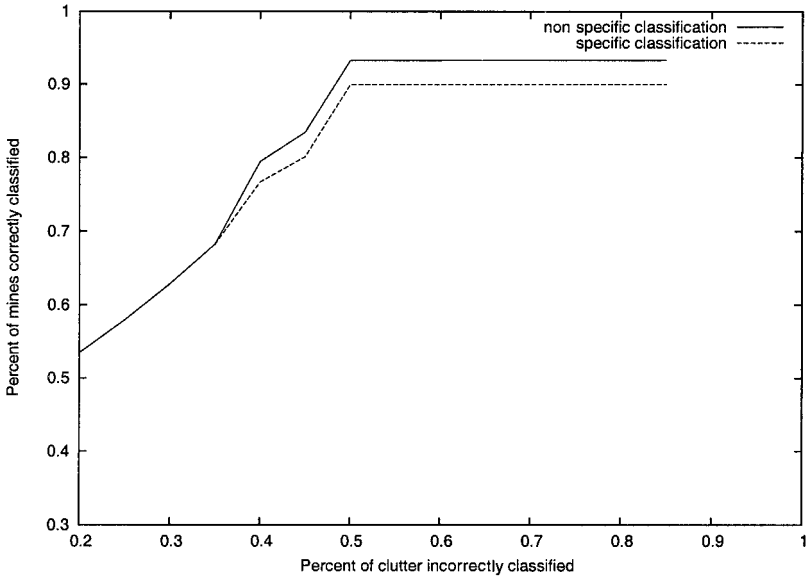

Fig. 5. Percentage of correctly classified mine objects plotted against the percentage incorrectly classified clutter objects.

in appearance of each object under different sonar conditions highlights the difficulties that a feature-based classification model would have in classifying each of the objects as the same shape. However, the sonar simulator model described here uses this information to model the underlying shadow forming process, allowing a correct classification to be obtained. Tables 1 and 2 show the monoimage and multiimage classification results for the cylinder and truncated cone images, respectively. Tables 1 and 2 show how both the objects are correctly classified with a strong belief. The cylindrical example is straightforward, with three of the four images providing the correct monoview classification result. Object 4 in Fig. 6 has a weak belief in the cylinder class owing to large estimates for $\Theta_{\text{cyl}}{}^{b}$. However, the fused result produces a very high belief for the cylinder class. The truncated cone example is more difficult, with three out of the four images providing incorrect monoview classification results. However, when images are fused, the consistently high truncated cone belief is used to correctly classify the object. Although only two examples are shown here, they are representative of other results obtained. The results demon-

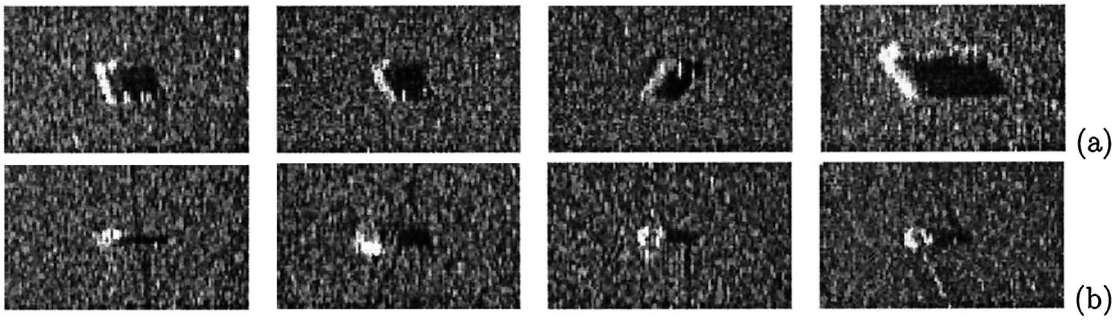

Fig. 6. (a) Four different views of the same cylinder. (b) Four different views of the same truncated cone. These views are taken from different directions, fish heights, and slant ranges.

Table 1. Belief Functions for Different Classes for the Individual Images and the Overall Fused Result for a Cylindrical Object[a]

Monoimage Belief					Fused Belief				
Object	Cyl	Sph	Cone	Clutter	Objects Fused	Cyl	Sph	Cone	Clutter
1	0.789	0.084	0.084	0.295	1	0.789	0.084	0.084	0.295
2	0.917	0.083	0.083	0.167	1, 2	0.945	0.009	0.009	0.064
3	0.917	0.083	0.083	0.167	1, 2, 3	0.989	0.001	0.001	0.012
4	0.324	0.151	0.151	0.826	1, 2, 3, 4	0.969	0.001	0.001	0.030

[a]Fused belief functions change as new images are added to the multiview classification.

Table 2. Belief Functions for Different Classes for the Individual Images and the Overall Fused Result for a Truncated Cone Object[a]

Monoimage Belief					Fused Belief				
Object	Cyl	Sph	Cone	Clutter	Objects Fused	Cyl	Sph	Cone	Clutter
5	0.145	0.313	0.371	0.605	5	0.145	0.313	0.371	0.605
6	0.194	0.194	0.561	0.632	5, 6	0.047	0.102	0.350	0.642
7	0.045	0.500	0.652	0.091	5, 6, 7	0.007	0.156	0.700	0.179
8	0.131	0.456	0.362	0.441	5, 6, 7, 8	0.002	0.178	0.634	0.199

[a]Fused belief functions change as new image are added to the multiview classification.

strate the advantages of considering multiview analysis for classification.

6. Conclusion

This paper has presented a model-based mine detection and classification system for use in sidescan imagery. The detection model used a MRF model to detect possible MLOs. This model allowed the inclusion of *a priori* information through the use of priors. Unlike many detection models currently in use, the process is completely automated and requires no training. The highlight and shadow regions of the MLOs are then extracted with a CSS model for classification. The model obtained accurate segmentation results—even on complex seabeds—by restraining the movements of the snake by use of *a priori* knowledge on the relationship between the highlight and shadow regions. The CSS model also allowed many false alarms to be removed from the initial detection result. A novel model-based classification system was then presented. This model classified the object by modeling the underlying physical shadow-forming process. This system extended the normal mine–not-mine classification to provide shape and size information on the object. The classification decision was provided by a DS framework, which allowed monoimage and multiimage analyses. This feature is especially desirable in sidescan surveys, where the same object is often viewed multiple times. Results were presented for real sidescan data. This work would also be directly applicable to detection and classification models in other media such as SAR imagery.

The authors thank the Mine and Torpedo Defence group at DRDC-Atlantic (Canada), the NATO Saclant Undersea Research Centre (Italy) and GESMA (France) for providing the sidescan data used to present our results.

References

1. G. J. Dobeck, J. C. Hyland, and L. Smedley, "Automated detection and classification of sea mines in sonar imagery," in *Detection and Remediation Technologies for Mines and Minelike Targets II*, A. C. Dubey and R. L. Barnard, eds. Proc. SPIE **3079,** 90–110 (1997).
2. C. M. Ciany and J. Huang, "Computer aided detection/computer aided classification and data fusion algorithms for automated detection and classification of underwater mines," in *Proceedings of the MTS/IEEE Oceans Conference and Exhibition*, (Institute of Electrical and Electronics Engineers, New York, 2000), Vol. 1, pp. 277–284.
3. T. Aridgides, M. Ferdandez, and G. Dobeck, "Fusion of adaptive algorithms for the classification of sea mines using high resolution side scan sonar in very shallow water," in *Proceedings of MTS/IEEE Oceans Conference and Exhibition*, (Institute of Electrical and Electronics Engineers, New York, 2001), Vol. 1, pp. 135–142.
4. L. M. Linnett, S. J. Clarke, C. St. J. Reid, and A. D. Tress, "Monitoring of the seabed using sidescan sonar and fractal processing," in *Conference of Underwater Acoustics Group* (Institute of Acoustics, St. Albans, Hertfordshire, UK, 1993) Vol. 15, pp. 49–64.
5. B. R. Calder, L. M. Linnett, and D. R. Carmichael, "Spatial stochastic models for seabed object detection," in *Detection and Remediation Technologies for Mines and Minelike Targets II*, A. C. Dubey and R. L. Barnard, eds., Proc. SPIE **3079,** 172–182 (1997).
6. M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Three class Markovian segmentation of high resolution sonar images," Comput. Vision Image Understand **76,** 191–204 (1999).
7. S. Reed, Y. Petillot, and J. Bell, "An automated approach to the detection and extraction of mine features in sidescan sonar," IEEE J. Ocean Eng. **28,** 90–105 (2003).
8. S. Reed, J. Bell, and Y. Petillot, "Unsupervised segmentation of object shadow and highlight using statistical snakes," in *Autonomous Underwater Vehicle and Ocean Modelling Networks: GOATS 2000 Conference Proceedings CP-46*, (NATO Saclant Undersea Research Centre, La Spezia, Italy, 2001) pp. 221–236.
9. B. Zerr, E. Bovio, and B. Stage, "Automatic mine classification approach based on AUV manoeuvrability and the COTS side scan sonar," in *Autonomous Underwater Vehicle and Ocean Modelling Networks: GOATS 2000 Conference Proceedings CP-46*, (NATO Saclant Undersea Research Centre, La Spezia, Italy, 2001), pp. 315–322.
10. S. LeHegart-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," IEEE Trans. Geosci. Remote Sens. **35,** 1018–1031 (1997).
11. I. Bloch, "Some aspects of Dempster–Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account," Pattern Recog. Lett. **17,** 905–919 (1996).
12. D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," IEEE Trans. Pattern Anal. Mach. Intell. **15,** 850–863 (1993).
13. J. Besag, "on the statistical analysis of dirty pictures," J. R. Stat. Soc. Ser. B. Methodol. **48,** 259–302 (1986).
14. M. L. Comer and E. J. Delp, "Segmentation of textured images using a multiresolution Gaussian autoregressive model," IEEE Trans. Image Process. **8,** 408–420 (1999).
15. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6,** 721–741 (1984).
16. M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Sonar

image segmentation using an unsupervised hierarchical MRF model," IEEE Trans. Image Process. **9,** 1216–1231 (2000).
17. E. Dura, J. Bell, and D. Lane, "Superellipse fitting for the classification of mine-like shapes in side-scan sonar images," *Proceedings of MTS/IEEE Oceans Conference and Exhibition*, (Institute of Electrical and Electronics Engineers, New York, 2002), Vol. 1, pp. 23–28.
18. C. Chesnaud, P. Refregier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models," IEEE Trans. Pattern Anal. Mach. Intell. **21,** 1145–1157 (1999).
19. M. Mignotte, C. Collet, P. Perez, 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. **22,** 129–141 (2000).
20. P. Galerne, K. Yao, and G. Burel, "Object classification using neural networks in sonar imagery," in *New Image Processing Techniques and Applications: Algorithms, Methods, and Components II*, P. Refregier, R.-J. Ahlers, eds., Proc. SPIE **3101,** 306–314 (1997).
21. J. Bell, "A Model for the Simulation of Sidescan Sonar," Ph.D. dissertation (Heriot-Watt University, Edinburgh, Scotland, 1995).
22. A. H. S. Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: fusion of landsat TM and SAR images," IEEE Trans. Geosci. and Remote Sens. **32,** 768–778 (1994).
23. I. Bloch, "Information combination operators for data fusion: a comparative review with classification," IEEE Trans. Syst. Man Cybern. **26,** 52–67 (1996).