# An Automatic Approach to the Detection and Extraction of Mine Features in Sidescan Sonar

Scott Reed, Yvan Petillot, and Judith Bell

***Abstract*—Mine detection and classification using high-resolution sidescan sonar is a critical technology for mine counter measures (MCM). As opposed to the majority of techniques which require large training data sets, this paper presents unsupervised models for both the detection and the shadow extraction phases of an automated classification system. The detection phase is carried out using an unsupervised Markov random field (MRF) model where the required model parameters are estimated from the original image. Using *a priori* spatial information on the physical size and geometric signature of mines in sidescan sonar, a detection-orientated MRF model is developed which directly segments the image into regions of shadow, seabottom-reverberation, and object-highlight. After detection, features are extracted so that the object can be classified. A novel co-operating statistical snake (CSS) model is presented which extracts the highlight and shadow of the object. The CSS model again utilizes available *a priori* information on the spatial relationship between the highlight and shadow, allowing accurate segmentation of the object's shadow to be achieved on a wide range of seabed types. Results are given for both models on real and synthetic images and are shown to compare favorably with other models in this field.**

***Index Terms*—*A priori* information, automated mine detection, image analysis, Markov random field (MRF) models, shadow extraction, statistical snakes.**

## I. Introduction

THE ANALYSIS of sidescan sonar images in the field of mine countermeasures (MCM) is traditionally carried out by a skilled human operator. This analysis is difficult due to the large variability in the appearance of the sidescan images as well as the high levels of noise usually present in the images. With the advances in autonomous underwater vehicle (AUV) technology, automated techniques are now required to replace the operator to carry out this analysis on-board.

Complete MCM systems are usually composed of a detection and a classification process such as the systems by Dobeck *et al.* [1], Ciany *et al.* [2], [3], and Aridgides *et al.* [4]. All three of these systems operate using the detection/classification framework although they operate using very different models. Dobeck implements a matched filter in [1] to detect mine-like objects (MLOs) after which both a $K$-nearest neighbor neural network classifier and a discriminatory filter classifier are used to classify the objects as mine or not-mine. The detection process is relatively simple and is primarily for identifying regions which definitely do not contain MLOs. The classification process then uses up to 45 features for every possible MLO to determine which are real MLOs and which are false alarms. The system in [2] utilizes an adaptive thresholding technique for the detection after which geometric features are extracted, allowing each MLO to be classified as mine or not-mine. Adaptive Clutter Filter technology is used in [4] to suppress the background clutter after which classification is carried out on an optimum set of features. These systems are similar in that the detected MLO is classified simply as mine or not-mine by considering a set of features and that all three require training using a large amount of ground truth data. The success of these models is thereafter dependent on the similarity between the training data and the test data with poor results being observed when the difference between the two is high [2]. It has also been shown in [5] that the success of trained models can be dependent on the choice of data used to train the system. The reported successes of these models have been dramatically improved by fusing the results of the individual models [2], [6], [7] together. This is based on the premise that as the individual models use different mathematical functions to carry out their procedures, fusing the results together will both confirm suspected MLOs and help remove false alarms. This idea has provided encouraging results and could be easily extended to other automated MCM systems, both supervised and unsupervised.

Instead of considering the computer aided detection/classification (CAD/CAC) problem as being completely integrated, research is often carried out on a specific aspect of the problem. Detection of possible MLOs has been attempted using fractal-based analysis [8], spatial point processes [9], and dual hypothesis theory [10] where an object is characterized as a disruption in the local texture field. However, the success of these models is heavily dependent on large training samples and simplifying modeling assumptions, raising questions to their widespread applicability. Thresholding and clustering theory has been used in [11] and [12] to segment the sidescan sonar image into regions of object-highlight, shadow, and background after which neighboring object-highlight and shadow regions were labeled as possible MLOs. This idea has been developed further in [13] where two Markov random field (MRF) models were used to segment the images using the *a priori* knowledge that object-highlight regions generally lie close to shadow regions. While the technique is not a detection model as such (it identifies possible object-highlight pixels rather than regions), it does demonstrate that MRF models can provide a suitable vehicle for modeling *a priori* information. The use of *a priori* information is convincingly demonstrated in [14] and [15] where an MRF tech-

Manuscript received April 22, 2002; revised August 23, 2002.

The authors are with the Ocean Systems Laboratory, Department of Computing and Electrical Engineering, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, U.K.

Digital Object Identifier 10.1109/JOE.2002.808199

nique was used to model some of the available information on the sonar process.

After an MLO has been detected, a classification procedure is required to determine whether the detected object is a false alarm or not. While many systems define classification as simply determining whether an object is mine or not-mine, geometric analysis can be used in the classification stage to determine the shape of the object [16]. Mines can often be described by simple objects such as cylinders, spheres, and truncated cones, therefore ensuring that, if the MLO can be classified as one of these objects, it can be identified as a mine with a high degree of confidence. A nonpositive classification as one of these objects leads to the MLO being identified as not-mine. Fawcett [17] has attempted this form of classification using simple features drawn from a mugshot of the object (this process assumed prior detection of the object). The technique is interesting yet was tested using only synthetic data where the success rate deteriorated when complex backgrounds where added to the object mugshots. The extracted highlight region of the object has also been considered in [18] and [11] for classification but is usually too variable and dependent on the specific sonar conditions to be used as a reliable classification feature. A popular feature to use is the object's shadow region which is generally more dependable and can be used to accurately classify the object if it can be extracted accurately.

Extraction of the shadow using classical edge-driven deformable models [19], [20] is generally not possible due to the high levels of noise in sidescan imagery. Models have been developed to overcome this problem using fuzzy logic [21], histogram thresholding [22], and statistical models [23]. Although these models offer good results on relatively flat seabeds, the presence of sand ripples often leads to inaccurate shadow extraction [24]. Quidu *et al.* [22] classify the object by extracting features from the shadow and comparing these to a training set. Due to the nonlinear nature of the sonar process (the same object at different ranges and orientations will produce completely different shadow regions), the features first had to be range normalized. Deformable templates have also been used in [25] and [26] to directly classify the object. Mignotte *et al.* [25] approximated the shadows produced by a cylinder and a sphere as a parallelogram and spline templates respectively, using affine transformations on these templates to find the best fit to the shadow. While good results are observed, these template models are disadvantaged for classification purposes in that they usually include the assumption that the MLO will match one of the tested templates. Also, altering the shape of the shadow template directly instead of considering the relationship between the objects parameters (size and orientation) and the resultant shadow region will affect the ability to determine the object's dimensions during the classification stage.

The solution presented here, which aims at solving the automated MCM problem, is a three-tier process as summarized in Fig. 1. The first stage detects MLOs in the sidescan data. Having identified these possible targets, the second stage extracts the shadow cast by the object to be used later in the classification stage. This classification stage will use the shadow information to provide information on the shape and dimensions of the detected object. This paper concentrates on the first two stages of the process.

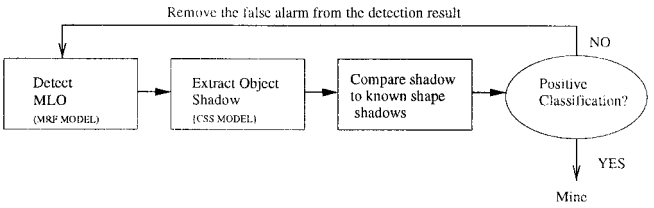


Fig. 1. Overall proposed detection and classification system. The first two parts are considered in detail in this paper.

A novel, automated detection model is presented to fulfill the first stage of the process in Fig. 1. This utilizes an MRF model to carry out a detection-oriented segmentation on the raw sidescan image. While most detection models which consider the underlying label field use a two-tier process (the image is first segmented after which the detection problem is considered), this model will directly segment the image into regions of object-highlight, seabottom reverberation, and shadow using available *a priori* spatial information on the appearance of mine signatures in sidescan sonar. Results will then be presented on both real and synthetic images.

The detection phase identifies (areas where the model has identified a mine-like signature) which need to be extracted from the image for further examination. A novel co-operating statistical snakes (CSS) model is then presented which provides an accurate and robust method for extracting both the objects highlight and shadow regions. The model segments the object-highlight and the shadow region by considering the image as being composed of three separate statistical regions. Using *a priori* information on the relationship between the object-highlight and the shadow, accurate segmentation can be achieved on seabed types where other models would fail. Results are given again on both real and synthetic images.

The paper will be laid out as follows. Section II details the sidescan process and discusses what *a priori* knowledge on objects in sidescan sonar is used within this paper. Section III will detail the unsupervised detection model. Section IV will outline the CSS shadow extraction model and highlight the link between the two separate processes while Section V will conclude the paper.

## II. Objects in Sidescan Sonar

For the purposes of this paper, it is assumed that all objects are discrete and protrude above the seabed, but are still connected to it [14]. If it is assumed that the sonar sound pulse moves without refraction, the process can be approximated by tracing rays, similar to the ray-tracing method used for simulating optical scenes [27]. This produces the geometrical situation pictured in Fig. 2. As the object is denser or has a higher reflectivity than the background, the return from the object surface (points A–B) is much stronger than the background. The sonar shadow (points B–C) is produced due to the object effectively blocking the sonar waves from reaching this region of the seabed. While this model is not correct in all cases (extreme range, floating objects), MCM data is usually taken with a sonar fish at low altitude. This ensures

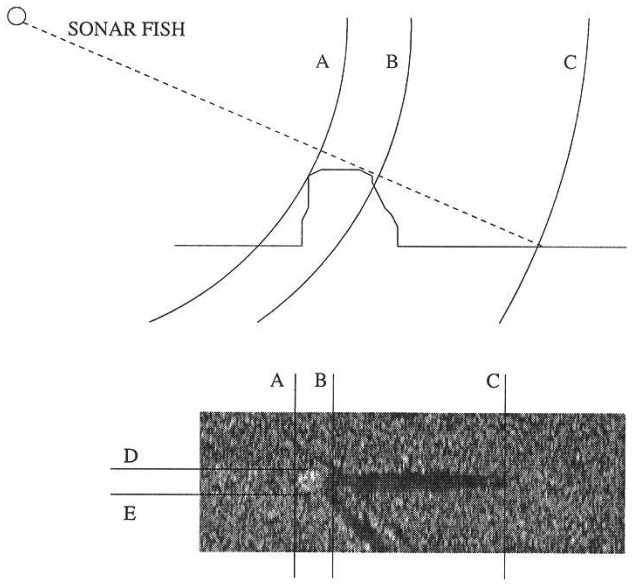


Fig. 2. The formation of object in sidescan sonar images using the ray-based approach to modeling.

that the objects produce shadows and therefore comply with the model described in Fig. 2.

Fig. 2 illustrates the geometry for one line of a sidescan image. As the full image is created by repeating this process for each pulse as the AUV moves through the water, the shadow region produced by the object can only be as wide as the object in the sonar image (points D–E in Fig. 2).

The object signature observed in Fig. 2 allows common characteristics to be modeled and used in both the detection and the CSS model. As MLOs are small, the highlight observed is also small, isolated, and compact. Due to the usual MCM procedure of using a low altitude sonar fish, this small highlight will be accompanied by a shadow region.

## III. Unsupervised Object Detection

### A. Introduction

The first stage in the automated MCM process is the detection of possible MLOs within the raw Sidescan image. While many mine detection models act directly on the noisy Sidescan image, promising results have been obtained by first trying to segment the image to recover the underlying label field (in this paper, the allowed labels are shadow, seabottom-reverberation and object-highlight) [14], [24]. An MRF model provides a reliable framework for obtaining this underlying field by incorporating pixel dependencies into the segmentation model (i.e., a pixel surrounded by shadow pixels is most likely to belong to the shadow class itself). This ability to simply and effectively model the inter-spatial dependencies between pixels has ensured that simple MRF models have been used for a wealth of applications, obtaining accurate segmentation results in the presence of strong noise [28]–[30]. However, within the context of sidescan imagery, where there is a large variation in the appearance and complexity of the images, more complicated models containing parameter estimation phases are required to ensure a confident segmentation. The MRF model used in this section extends the two-class anisotropic MRF model in [25] to develop a detection-orientated segmentation model. This model uses *a priori* knowledge on the size and appearance of mine signatures in sidescan sonar to directly segment the images into regions of object-highlight, shadow and seabottom-reverberation.

### B. MRF Theory

A general MRF model consists of two fields, the observed image $Y$ and the underlying "true" label field $X$ which we wish to recover. A pixel $s$ is assigned a label $x_s$ based on two criteria. The first is dependent on the labels of the neighboring pixels and is controlled by a local Markovian probability term. The second criteria considers the probability of label $x_s$ producing observed gray level $y_s$. This requires that each possible label field has a corresponding noise distribution from which its observed graylevels can be drawn. Therefore, the MRF model must first have the capacity to determine the parameters of the Markovian probability term as well as the parameters of the noise distributions.

We consider a more complex set of three random fields $Z = (X, Y, O)$ where we define $Y = \{Y_s, s \in S\}$ as the field of observations (this is the raw sidescan image) where each $Y_s$ takes it value from the possible gray-level values $\{0 \cdots 255\}$. Label field $X = \{X_s, s \in S\}$ is the underlying label field which we wish to recover and so $X_s$ can take the value $\{e_0 = shadow, e_1 = seabottom\text{-}reverberation, e_2 = object\text{-}highlight\}$. $O = \{O_s, s \in S\}$ is defined as the object field where each $O_s$ is drawn from $\{o_0 = object, o_1 = non\text{-}object\}$. This field can be determined directly by considering label field $X$ where $O_s = o_0(object)$ if $X_s = e_2(object\text{-}highlight)$ and $O_s = o_1(non\text{-}object)$ otherwise. Label Field $O$ therefore shows the clustering of object pixels. Based on the observed data $Y = \{Y_s, s \in S\}$, the detection process can be cast as an analysis of the conditional probability $P(X, O|Y)$, the probability of the "unobserved" true data given the observational data. Using Bayes theorem, this probability can be expressed as

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

$P_{Y|X}(y|x)$ is the likelihood term where the data is assumed to be independently conditioned on labeling process $X$. It can therefore be defined as a product of the individual pixel probabilities $P_{Y|X}(y|x) = \prod_{s \in S} P_{Y_s|X_s}(y_s|x_s)$ where $P_{Y_s|X_s}(y_s|x_s)$ is the probability of observed gray-level $y_s$ being drawn from the noise distribution used to represent label $x_s$. $P_X(x)$ is the Markovian prior distribution used to model the dependencies between pixels of the label field $X$. $P_{O|X}(o|x)$ is a prior probability which uses *a priori* information on the size and geometry of mine signatures in sidescan sonar to discourage clusterings of $O_s = o_0(object)$ which have the wrong size. Expressing the posterior distribution as $P_{X,O|Y}(x, o|y) \propto \exp\{-U(x, y, o)\}$ [31], the underlying label field can be obtained by minimizing the following posterior energy:

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

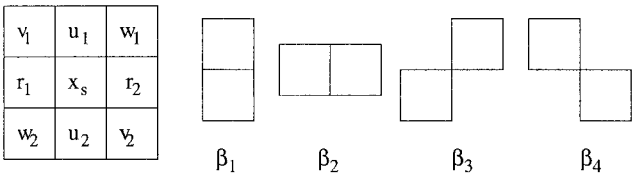


Fig. 3. Second-order neighborhood system with associated cliques and their labeling notations.

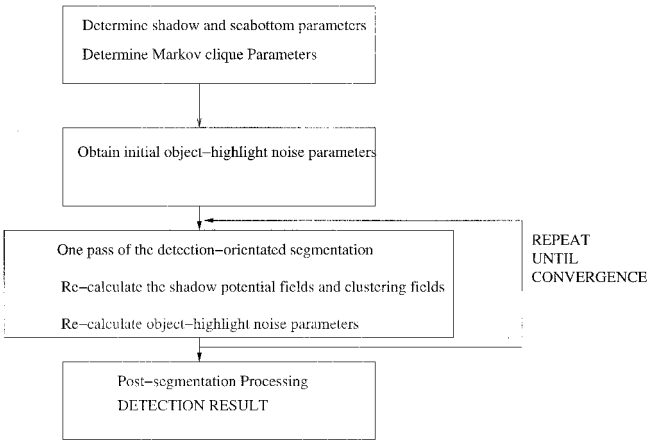


Fig. 4. Overview of the detection model.

The first term on the right-hand side $\Phi_s(x_s, y_s) = -\ln P_{Y_s|X_s}(y_s|x_s)$ is the energy term relevant to the likelihood function $P_{Y|X}(y|x)$. The second term describes the dependency of label $x_s$ on the label values of the neighboring pixels of $s$ where a second-order anisotropic model has been used. Fig. 3 shows the four allowed cliques for this neighborhood, thereby showing how $\beta_{st} = \beta_1, \beta_2, \beta_3$, or $\beta_4$ depending on the relative position of the neighboring pixel to pixel $s$.

The third term acts only on pixels with label $x_s = e_2(object\text{-}highlight)$. This uses an adaptation of a potential term derived in [13], utilizing the *a priori* information that a mine highlight should have a shadow to the right of it (port configuration). The fourth term uses more *a priori* information and favors the clustering of $O_s = o_0(object)$ pixels only if they are of the right size. This function models the belief that mine-like signatures are in general compact and separated.

An overview of the entire detection-orientated segmentation process quantified in (2) can be seen in Fig. 4. The separate components within this process will now be considered in detail.

### C. Estimation of the Markovian $\beta$ Parameters and Noise Parameters

For the estimation of the Markovian parameters $\Theta_x = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ and the noise parameters, the image is first considered to be composed of only two regions: shadow and nonshadow. The likelihood term for the shadow class $P_{Y_s|X_s}(y_s|e_0)$ is assumed to be a Gaussian with mean gray-level and variance $m_{\text{sh}}$ and $\sigma^2_{\text{sh}}$, respectively. The likelihood term for the seabottom-reverberation class $P_{Y_s|X_s}(y_s|e_1)$ is described by a shifted Rayleigh law with minimum gray-level $\min_{\text{sea}}$ and variance $\sigma^2_{\text{sea}}$, thereby requiring noise parameters $\Theta_y = \{m_{\text{sh}}, \sigma^2_{\text{sh}}, \min_{\text{sea}}, \sigma^2_{\text{sea}}\}$ to be estimated. Justification for using a Rayleigh distribution for the seabottom-reverberation class can be seen in [32]. This argues that isotropic seabed regions are described well by Rayleigh distributions while it is assumed that the luminance within shadow regions is essentially due to electronic noise and so is described by a Gaussian distribution.

As mine-like objects are known *a priori* to be small and clustered [10], the estimation of these parameters without consideration to the third class $X_s = e_2(object\text{-}highlight)$ was expected to yield accurate results. Determining estimates to $\Theta_x$ and $\Theta_y$ was done using the Iterative Conditional Estimation (ICE) model described in detail in [31], and [33] and summarized here. The ICE technique first requires initial estimates $\Theta_x^{[0]}$ and $\Theta_y^{[0]}$ to the parameters $\Theta_x$ and $\Theta_y$. The iterative technique then defines $\Theta_x^{[k+1]}$ and $\Theta_y^{[k+1]}$ to be the conditional expectations of parameter estimators $\hat{\Theta}_x$ and $\hat{\Theta}_y$, respectively, at iteration $k+1$ dependent on the data $Y = y$ and the current parameter fits $\Theta_x^{[k]}$ and $\Theta_y^{[k]}$. Appealing to the law of large numbers, these terms are related by

$$\Theta_x^{[k+1]} = \frac{1}{n}\left[\hat{\Theta}_x\left(x_{(1)}\right) + \cdots + \hat{\Theta}_x\left(x_{(n)}\right)\right] \quad (3)$$

$$\Theta_y^{[k+1]} = \frac{1}{n}\left[\hat{\Theta}_y\left(x_{(1)}, y\right) + \cdots + \hat{\Theta}_y\left(x_{(n)}, y\right)\right]. \quad (4)$$

Both $\Theta_x^{[k+1]}$ and $\Theta_y^{[k+1]}$ can therefore be calculated by drawing $x_{(i)}, i = 1 \cdots n$ realizations from the posterior distribution $P_{X|Y,\Theta}(x|y, \Theta^{[k]})$ where $n$, the number of realizations, is set to 1. The Gibbs Sampler was used to generate samples from this posterior distribution which was represented by a simplified version of the posterior energy in (2). This gave the posterior energy term $U_{\text{est}}$ described by

$$U_{\text{est}}(x, y) = \sum_{s \in S} \Phi_s(x_s, y_s) + \sum_{\langle s,t \rangle} \beta_{st}[1 - \delta(x_s, x_t)]. \quad (5)$$

The last two terms of (2) have been neglected as these deal with the $X_s = e_2(object\text{-}highlight)$ class and are therefore not used in this parameter estimation step. For the ICE technique to work, initial estimates $\Theta_x^{[0]}$ and $\Theta_y^{[0]}$ to the model parameters are required, as is a method for determining $\hat{\Theta}_x$ and $\hat{\Theta}_y$ at each iteration.

*1) Determining $\hat{\Theta}_x$ and $\hat{\Theta}_y$:* Determining the estimator of the Markov parameters, $\hat{\Theta}_x$, is done by considering label field $X$ and uses a least squares technique as developed by Derin *et al.* [34]. Defining $\eta_s$ as the second-order neighborhood of pixel $s$ as shown in Fig. 3, the Markovian probability can be written as

$$P_{X_s|X_{\eta_s}}(x_s|\eta, \Theta_x) \propto \exp\{-\Upsilon(x_s, \eta)^T \Theta_x\} \quad (6)$$

where, using the labels in Fig. 3, we have

$$\Upsilon(x_s, \eta) = [I(x_s, u_1) + I(x_s, u_2),\ I(x_s, r_1) + I(x_s, r_2),\ I(x_s, w_1) + I(x_s, w_2),\ I(x_s, v_1) + I(x_s, v_2)]^T \quad (7)$$

where $I = 1 - \delta$ and $\delta$ is the Kronecker delta function. This probability describes the dependency of label $x_s$ on the labels of the pixels in neighborhood $\eta_s$. For a given neighborhood $\eta_s$, the ratio of the probabilities of pixel $s$ being a shadow ($x_s = e_0$)

or a seabottom-reverberation ($x_s = e_1$) pixel can be calculated using (6) and taking the logarithm to give

$$\ln \frac{P_{X_s|X_{\eta_s}}(e_1|\eta)}{P_{X_s|X_{\eta_s}}(e_0|\eta)} = \ln \frac{P_{X_s, X_{\eta_s}}(e_1, \eta)}{P_{X_s, X_{\eta_s}}(e_0, \eta)} = [\Upsilon(e_0, \eta) - \Upsilon(e_1, \eta)]^T \Theta_x. \quad (8)$$

For each possible neighborhood configuration $\eta$, the second term in (8) can be approximated using simple histogramming where the number of times each configuration occurs in the label field is counted to give

$$\frac{P_{X_s, X_{\eta_s}}(e_1, \eta)}{P_{X_s, X_{\eta_s}}(e_0, \eta)} = \frac{\#\{s \in S: x_s = e_1, x_{\eta_s} = \eta\}}{\#\{s \in S: x_s = e_0, x_{\eta_s} = \eta\}}. \quad (9)$$

This creates an over-determined set of equations for the four unknowns which can be solved in a least-squares sense to provide an estimate $\hat{\Theta}_x$ for the Markovian model parameters $\Theta_x$.

Determining the estimator of the noise parameters $\hat{\Theta}_y$ is achieved by considering the complete data $(x, y)$ and is obtained using a simple maximum-likelihood method where the individual components of $\hat{\Theta}_y$ can be determined by

$$\hat{m}_{\rm sh}(x, y) = \frac{1}{N_{e_0}} \sum_{s \in S: x_s = e_0} y_s \quad (10)$$

$$\hat{\sigma}^2_{\rm sh}(x, y) = \frac{1}{N_{e_0} - 1} \sum_{s \in S: x_s = e_0} (y_s - \hat{m}_{\rm sh})^2 \quad (11)$$

$$\widehat{\min}_{\rm sea}(x, y) \simeq \min\{y_s: x_s = e_1\} - 1 \quad (12)$$

$$\hat{\sigma}^2_{\rm sea}(x, y) = \frac{1}{2N_{e_1}} \sum_{s \in S: x_s = e_1} \left(y_s - \widehat{\min}_{\rm sea}\right)^2 \quad (13)$$

where $N_t$ is the number of pixels with label $t$.

*2) Obtaining Initial Estimates $\Theta_x^{[0]}$ and $\Theta_y^{[0]}$:* Once an initial label field $X^{[initial]}$ has been determined, the initial estimates $\Theta_x^{[0]}$ and $\Theta_y^{[0]}$ can be obtained using the techniques described in the previous section. $X^{[initial]}$ is obtained by first splitting the image into $M$ nonoverlapping windows. Each window is assigned a vector $\mathbf{x_i}$, where $0 \leq i \leq M$. Each vector $\mathbf{x_i}$ is composed of two components $m_i$, the mean gray level, and $l_i$, the minimum gray level. These vectors are then clustered into either a shadow or seabottom-reverberation group using a $k$-means clustering algorithm [31]. From this clustering algorithm, the maximum-likelihood estimates of $\Theta_y^{[0]}$ can be obtained. The label field $X^{[\rm initial]}$ can then be initialized using simple maximum-likelihood considerations [essentially segmenting the image using only the first term on the right-hand side of (5)]. From this, $\Theta_x^{[0]}$ can be obtained using the least-squares method described in the previous section. Starting from initial parameter estimates $\Theta_y^{[0]}$ and $\Theta_x^{[0]}$, the ICE model can thereafter produce more accurate estimates.

### D. Obtaining and Updating Initial Object-Highlight Noise Parameters

The appearance of object-highlight regions in sidescan sonar is dependent on a large variety of factors such as the material and orientation of the object involved. It therefore cannot be described by a well-defined noise law as with the shadow and seabottom-reverberation regions. However, due to the fact that the objects protrude above the seafloor and that they have often have a higher degree of reflectivity than the seafloor, object-highlight regions are generally among the brightest parts of the sidescan image (which are typically quantized to 8 b). Teaching algorithms to model the object-highlight noise distribution based on training sets would prove problematic due to the large number of variables that dictate the appearance of the highlight regions. For instance, sonar images taken on the same heading but different altitudes would produce very different results. An AUV on a different heading altogether would likely produce an image unrecognizable as the same area of seabed. Due to these complexities, it is necessary to deduce the object-highlight noise distribution on a per-image basis, using a distribution that simply models the vague *a priori* belief that the highlight regions are the brightest parts of the image. A normalized linear equation is used with the form

$$P_{Y_s|X_s, \Theta_y}(y_s|e_2, \Theta_y) = A_{\rm ob}(m_{\rm ob} y_s + c_{\rm ob}) U(y_s - g_{\min}) U(g_{\max} - y_s) \quad (14)$$

where $U(.)$ is the Heaviside function, $m_{\rm ob}$ is the gradient of the line, $c_{\rm ob}$ is the intersect point, $g_{\min}$ and $g_{\max}$ are the minimum and maximum allowed gray levels and $A_{ob}$ ensures the function to be normalized within the allowed limits $\{g_{\min}, g_{\max}\}$. Initially we have no information on the expected range of the object-highlight pixels and so the conservative values of $m_{\rm ob} = 1.0$, $c_{\rm ob} = 0.0$, $A_{\rm ob} = 3.07 \times 10^{-5}$, $g_{\min} = 0$ are allocated. $g_{\max}$ is allocated the highest gray-level value in the image. This produces a normalized triangular function. Using these parameter estimates for the object-highlight regions along with the final parameter estimates $\hat{\Theta}_y$, the label field $X$ can be initialized for all three classes. This is demonstrated in Fig. 5 where three images containing mines are shown along with the initial labeling for field $X$ prior to segmentation.

As Fig. 5 shows, the accurate parameter estimation of $\Theta_y$ using the ICE technique has led to a good initialization for the shadow and seabottom-reverberation regions. The description of the object-highlight regions is much poorer due to the lack of *a priori* information, highlighting the need for the two prior terms in (2) to provide an accurate segmentation.

As the detection-orientated segmentation continues, the priors which affect the $x_s = e_2(object\text{-}highlight)$ pixels (these two terms are explained in the next section) will begin to remove many of the false alarms allowing the noise distribution for the object-highlight regions to be updated. This allows the numerical values used for initialization to be updated as the segmentation proceeds. $m_{\rm ob}$ and $c_{\rm ob}$ are updated by a Least-squares method similar to that used in the estimation of $\Theta_x$ where a general linear line is fitted to a histogram of the pixels labeled $X_s = e_2(object\text{-}highlight)$.

Parameters $g_{\min}$ and $g_{\max}$ are estimated from the object-highlight histogram while $A_{\rm ob}$, the normalizing constant, is calculated by

$$A_{\rm ob} = \frac{2}{m_{\rm ob}(g^2_{\max} - g^2_{\min}) + c_{\rm ob}(g_{\max} - g_{\min})}. \quad (15)$$

### E. Modeling the A Priori Information

Objects in sidescan sonar leave a recognizable signature characterized by a highlight region followed by a region of shadow.

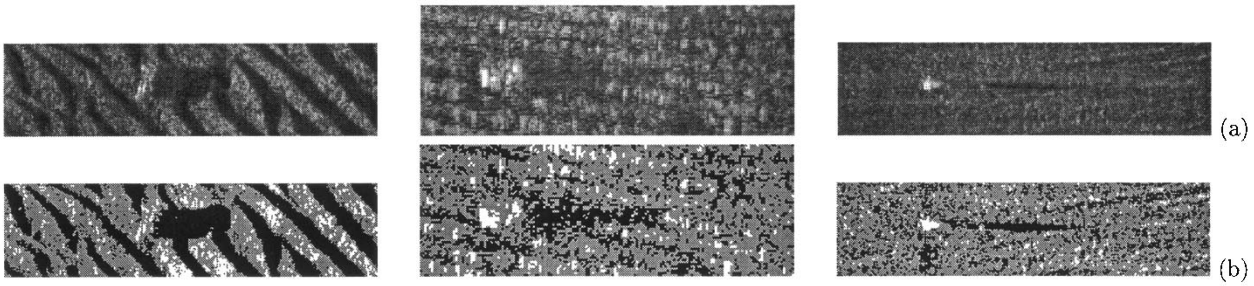


Fig. 5. (a) Three images containing mines. (b) Initial three-class labeling of field $X$ prior to the detection-orientated segmentation where black represents shadow regions, gray represents seabottom-reverberation regions and white represents object-highlight regions.

As discussed in Section II, the highlight regions of these objects also generally appear in small dense clusters surrounded by regions of seabottom-reverberation or shadow. This known *a priori* information can be modeled to increase the robustness of the detection algorithm.

*1) The Shadow Prior Energy Term:* The term $-\sum_{s\in S}\delta(x_s, e_2).\ln\Psi_X(s)$ in (2) acts only upon pixels with label $x_s = e_2(object\text{-}highlight)$ and discourages pixels not in the proximity of a shadow region from being labeled $x_s = e_2(object\text{-}highlight)$ [13]. *A priori* information on the geometry of the signature (all examples here are for port configuration) allow this criteria to become more specific in that the shadow region must lie to the right of the highlight region. We define a shadow pixel $s^o_{i,j}$ situated at row $i$ and column $j$ of the image which generates a potential field $\phi_{s^o_{i,j}}(r)$ such that

$$\phi_{s^o_{i,j}}(r) = \frac{1}{r}\exp\left(-\frac{r}{\sigma}\right) \tag{16}$$

where $r(r \neq 0)$ is the distance from pixel $s^o_{i,j}$ and $\sigma$ controls the rate of drop-off of the potential field. This is set at $\sigma = 5$ throughout to allow a smooth drop-off in the potential. Utilizing the *a priori* information that the shadow is always to the right of the highlight region in port configuration, we can express the total potential field at pixel $s_{k,l}$ as

$$\Psi_X(s_{k,l}) = \sum_{s\in S:\, x_s = e_0} \phi_{s^o_{i,j}}(d)\delta_{i=k}U(j-l) \tag{17}$$

where $d$ is the distance between pixels $s_{k,l}$ and $s_{i,j}$, $\delta$ is the Kronecker delta function, and $U(.)$ is the Heaviside function.

*2) The Clustering Prior Energy Term:* The final term in energy equation (2), $-\sum_{s\in S}\chi_s(x_s, o_s)$, considers object field $O$ to promote situations where object-highlight regions appear in small dense clusters. This *a priori* knowledge on the size of the objects being searched for is described by object parameters $\Theta_o = \{l_{\min}, l_{\max}\}$ where $l_{\min}$ and $l_{\max}$ are the minimum and maximum size of objects being searched for, respectively. This assumes that the image has been geo-referenced prior to analysis where each pixel is therefore a measure of distance rather than time. New inertial navigational sensors (INS) systems for geo-referencing can offer good estimates on both the AUV's position and velocity, ensuring that the geo-referenced image is a more accurate representation of the scene than the raw image. However, to account for possible errors, $l_{\min}$ and $l_{\max}$ are given conservative estimates, simply ensuring that clusterings of pixels which are obviously too small or large to be a mine are unlikely to remain labeled as object-highlight regions. Parameter $l_{\min}$ is used to define a mask $M$ as shown in Fig. 6 where the size of regions $I$ and $B$ will depend on the pixel resolution.

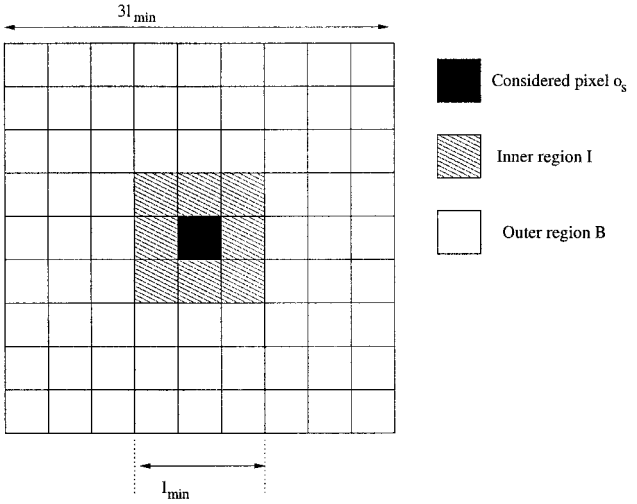


Fig. 6. Definition of object mask used in prior. The physical size of the objects being looked for is used to define the size of the inner region I.

Mask $M$ is comprised of three regions: the pixel under consideration $o_s$, an inner region $I$ with width $l_{\min}$, and an outer region $B$ with width $3l_{\min}$. The object-clustering field $\chi_s(x_s, o_s)$ at pixel $s$ can be determined by

$$\chi_s(x_s, o_s) = \frac{4}{N_I}\sum_{t\in I}\left(\delta_{o_t=o_0} - \frac{1}{2}\right)\left(\delta_{o_s=o_0} - \frac{1}{2}\right) - \frac{2}{N_B}\sum_{v\in B}\left(\delta_{o_v=o_1} - \frac{1}{2}\right) \tag{18}$$

where $N_I$ is the number of pixels in region $I$, $N_B$ is the number of pixels in region $B$, and $t$ and $v$ are integers used to sum over all the pixels in regions $I$ and $B$, respectively. This function is maximized when region $I$ is composed entirely of object pixels and region $B$ is made up of nonobject pixels, thus rewarding scenarios where the object cluster is small and isolated. The function encourages clustering of object-highlight pixels of the size of typical mines and is also useful in that it does not discriminate between the probability of a pixel belonging to the shadow or seabottom reverberation class as $o_s = o_1(non\text{-}object)$ for both $x_s = \{e_0, e_1\}$.

### F. The Segmentation Process

Achieving the global minima of (2) is a computationally huge task. The iterated conditional modes (ICM) technique [28] dramatically lightens computational demands by swiftly converging onto a local minimum. Segmentation is carried out using a raster scan where each pixel is considered in turn. Each pixel $s$ is assigned a label $x_s$ so as to always minimize energy $U(x, y, o)$ in (2).

After every sweep through the image, object field $O$ is updated from label field $X$. The shadow potential field $\Psi_X(s)$ is recalculated for each pixel $s$ and the object-highlight noise parameters are also recalculated. While these calculations should theoretically occur after every pixel label change, real-time constraints make this impractical and, in practice, good results are obtained with the updates being calculated after every sweep.

### G. Postsegmentation Processing

The detection-orientated segmentation process produces a field $X$ which is segmented into regions of shadow, seabottom-reverberation, and object-highlight. While the last two terms in (2) discourage regions of object-highlight which do not conform to the known mine signature in sidescan, false alarms can occur. To remove these, a postsegmentation process is carried out. This process will first use $\Theta_o = \{l_{\min}, l_{\max}\}$ to remove object-highlight regions which lie outside the acceptable size range. It will also remove object-highlight regions which do not lie in close proximity to a shadow region by defining a maximum allowed distance $D_{\max}$. The set limits for these techniques need not be rigidly defined and could be made case-specific. For example, if the model was looking for tethered mines, both the shadow potential $\Psi_X(s)$ and the post-segmentation distance $D_{\max}$ could be altered to detect the expected signature left by such a mine. $D_{\max}$ was set to 5 pixels in this model.

*1) The Size of the Object:* Model parameters $\Theta_o = \{l_{\min}, l_{\max}\}$ describe the minimum and maximum size of potential objects being searched for by the model. The maximum and minimum dimensions of each region were calculated by ensuring each pixel labeled as object-highlight ($x_s = e_2$) was assigned—$r_s^{\max}$ and $r_s^{\min}$. These equate to the maximum and minimum run length of object-highlight pixels through each pixel $x_s = e_2$, considering only vertical and horizontal runs of pixels. As the images were geo-referenced previously, these run-lengths could be equated to the physical dimensions of the object. To ensure that each pixel within an object-highlight region was assigned the same dimensions, a simple iterative labeling algorithm using a mode filter was carried out as illustrated in Fig. 7.

*2) The Distance From the Object Region to the Nearest Shadow Region:* Each pixel with $x_s = e_2(object\text{-}highlight)$ was given a minimum distance $d_s^{\min}$ to the nearest shadow region to the right (port configuration). The labeling algorithm described in Fig. 7 then ensured that every pixel within an object-highlight region was assigned the same distance value.

Once each pixel $x_s = e_2(object\text{-}highlight)$ had been labeled, regions which did not conform to the model could be simply removed and replaced with $x_s = e_1(seabottom\text{-}reverberation)$ pixels.

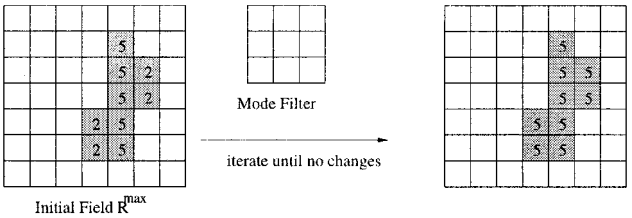


Fig. 7. Explanation of the labeling process using a mode filter to ensure that every pixel within each object is described by the same size dimensions.

### H. Results

The detection results given in this section assume that the objects present in the images have $\Theta_o = \{0.3, 1.8\}$ where these values are in meters. The detection model is first demonstrated on two synthetic images generated using the sidescan sonar simulator model developed by Bell [27]. This simulator was used to provide fully ground truthed data where exact details of the scene and objects could be controlled. The first example is a simple scenario where all the objects present appear on an isotropic seabed while the second example is more difficult due to the presence of the sand ripples. Both images have been geo-referenced so that each pixel has a resolution of 0.08 $\times$ 0.08 m. As Fig. 8 shows, the model succeeds in identifying all the objects in both images (the objects are marked white in the segmentation), offering no false alarms.

An important part of evaluating an object detection system is to test the model on images containing no objects. Fig. 9 contains two real sidescan sonar images where there is a high level of clutter but no objects. These images have been geo-referenced where each pixel has a resolution of 0.15 $\times$ 0.15 m. As can be seen, the detection model correctly identifies that there are no objects present regardless of the high amounts of clutter. Another complex image containing sand ripples and clutter but no objects is shown in Fig. 10. This image has resolution 0.08 $\times$ 0.08 m, with the detection model again correctly identifying no objects.

Fig. 11 contains two real sidescan images taken from another trial. The difference in appearance between these images and those in Figs. 9 and 10 is quite obvious, highlighting the need for a robust detection system to cope with the large variation in the appearance of sidescan sonar images. The first contains only one object which the model correctly identifies. The second contains multiple objects for which we have no ground truth data results. However, the results obtained agree well with a skilled operator's interpretation of the image. It should be noted that these two images are not geo-referenced as no navigational data were available. The images have therefore been assumed to have a pixel resolution of 0.08 $\times$ 0.08 m.

Fig. 12 shows a final real sidescan image. The image contains four mines lying on a sand ripple seabed surrounded by large amounts of clutter. The images has been georeferenced so that the pixels have a resolution of 0.08 $\times$ 0.08 m. The detection model successfully detects three of the mines along with several

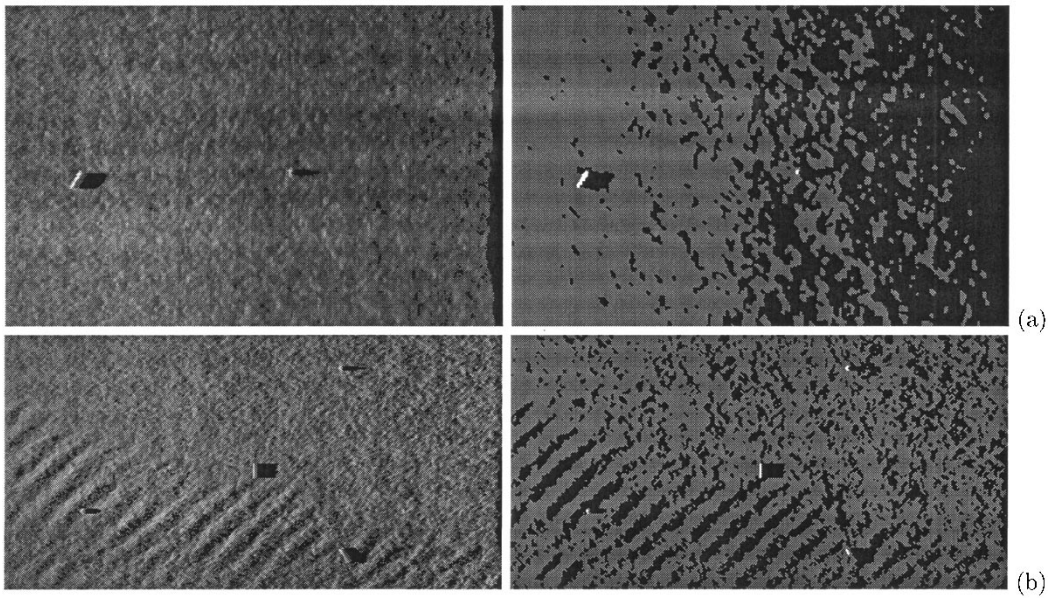


Fig. 8. (a) Synthetic image containing two mine-like objects and the detection-orientated segmentation result identifying all objects. (b) Synthetic image containing four mine-like objects on a complex seabed and the detection-orientated segmentation result correctly identifying all objects.

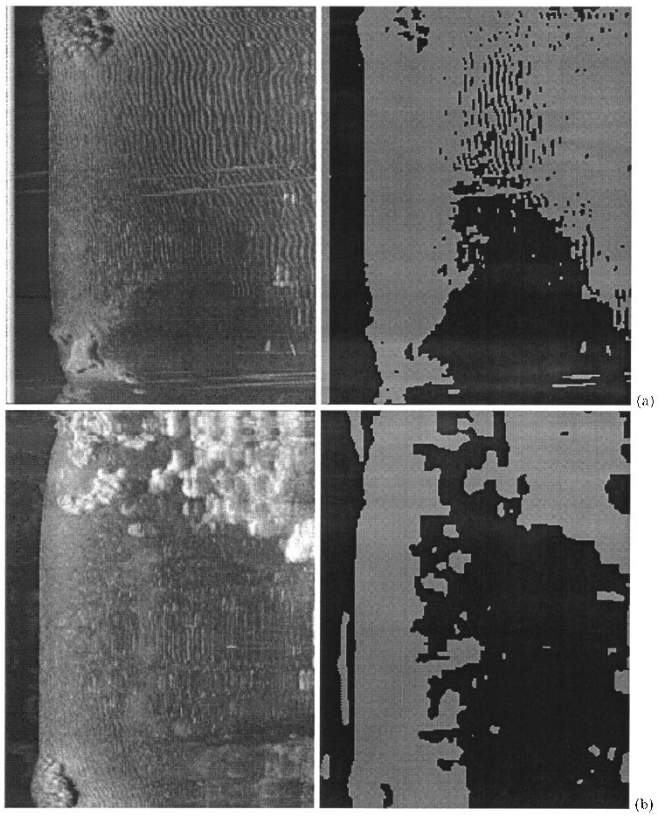


Fig. 9. (a) Real sidescan image containing no mine-like objects and the detection-orientated segmentation result which correctly detects no objects. (b) Another real sidescan image containing no mine-like objects and the detection-orientated segmentation result correctly identifying no objects.

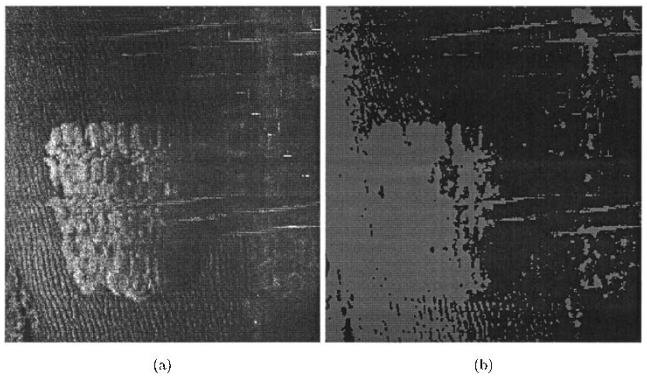


Fig. 10. (a) Real Sidescan image containing no mines but high levels of clutter and complex seabed types. (b) Detection-orientated segmentation model correctly detecting no objects in the image.

false alarms. The failure to detect the fourth mine arose from the region behind the object having a relatively high gray level therefore being labeled seabottom-reverberation. This caused the post-segmentation phase to remove the fourth mine as it had no accompanying shadow region. The false alarms detected all have sizes and signatures comparable to a mine-like object and are a result of the image containing a lot of object-like clutter. With the final three-tier classification system, it is hoped that these detections will be removed after the classification phase (see Fig. 1). Using the classification stage to remove false alarms could eventually allow the postsegmentation process of the detection model to be relaxed, thereby ensuring that the fourth mine in Fig. 12 is not removed.

### *I. Summary*

This section has introduced an automated detection algorithm which conducts a detection-orientated segmentation of the image using an MRF model along with *a priori* spatial information on the expected signature of mines in sidescan. The model has been tested on real and synthetic images, both of which contained clutter and a variety of seabed types. Results

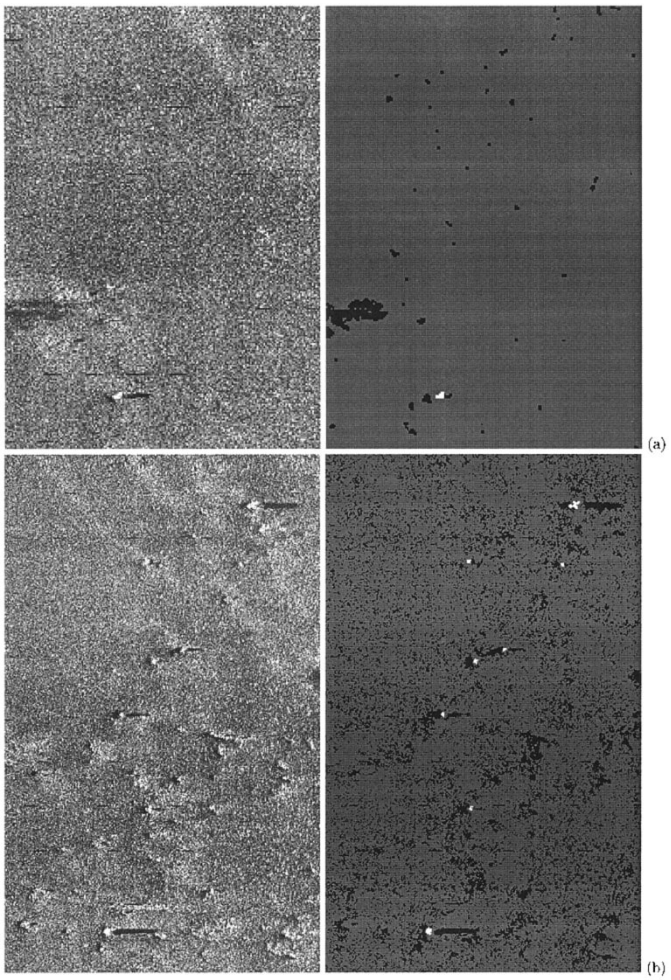


Fig. 11. (a) Real sidescan image containing one object and the detection-orientated segmentation result, correctly showing the one object. (b) Real sidescan image containing multiple objects as well as the detection-orientated segmentation result which correctly identifies all objects present.

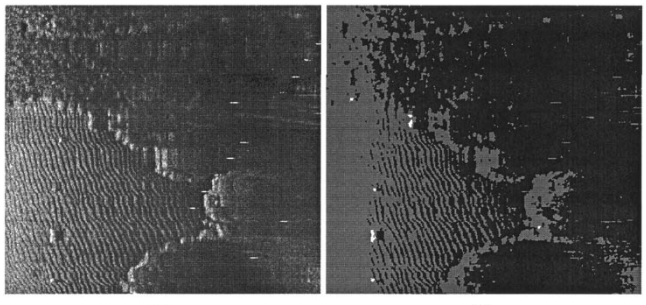


Fig. 12. (a) Real sidescan image containing four mines. (b) The detection-orientated segmentation result shows that three of the mines have been detected. Several false alarms with mine-like signatures were also detected.

were very promising with false alarms occurring only in one image where either the seabed or clutter presented a mine-like signature. This would suggest that a texture-based model would be useful to complement the spatial-based model presented here to provide a complete automated unit.

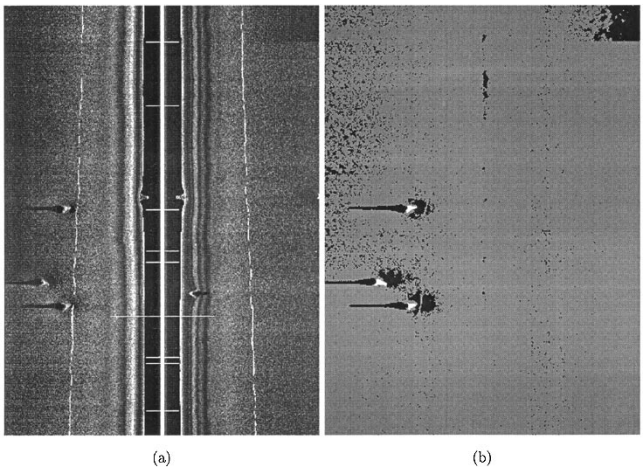


Fig. 13. (a) Sidescan image containing three mine-like objects. (b) Detection-orientated result successfully detecting the three objects. The images have been altered to be the same size for visual purposes.

## IV. Extracting the Object Features

After a mine-like object has been detected, its shadow can be extracted and used later in the classification process (Fig. 1). If the shadow of the object can be matched to the shadow from a well-known mine shape such as a cylinder or a truncated cone, the object's shape can be correctly classified and identified as a mine. For this process to be possible, it is necessary to first obtain an accurate segmentation of the shadow.

Several models have been proposed for extracting an object's shadow [25], [23], [21] which offer good results on flat seabeds but can yield poor results when complex seabeds such as sand ripples are involved. This is because the shadows due to the sand ripples are generally described by the same statistics as the object's shadow, often leading to inaccurate segmentations. The Co-operating Statistical Snake (CSS) model described here extracts both the object-highlight and the shadow. While the use of the object-highlight for classification purposes is limited, the known *a priori* information on the relationship between the highlight and shadow can be used to ensure that the shadow segmentation is accurate. The CSS model approximates the image as three homogeneous regions—object-highlight, shadow and background and so uses two statistical-snakes [35] to segment both the object-highlight and shadow. The *a priori* information between the object-highlight and shadow is used to constrain the movement of the snakes so as to achieve accurate segmentation results regardless of the seabed type involved.

If a mine-like object has been detected using the model described in Section III, some of the available information from the detection result can be used to overcome the initialization problem which is inherent in many segmentation algorithms. As the size and position in the image of the MLOs are known from the detection result, the CSS model can be accurately initialized by considering the label field. This is demonstrated in Figs. 13 and 14. Fig. 13 contains a reduced, raw sidescan image and the final detection result showing three MLOs. Fig. 14 contains the extracted label field for each of the objects as well as the initial and final segmentation results of the CSS model. The initialization of the object-highlight and shadow boxes was conducted by first restricting the two snakes to a rectangular form and using

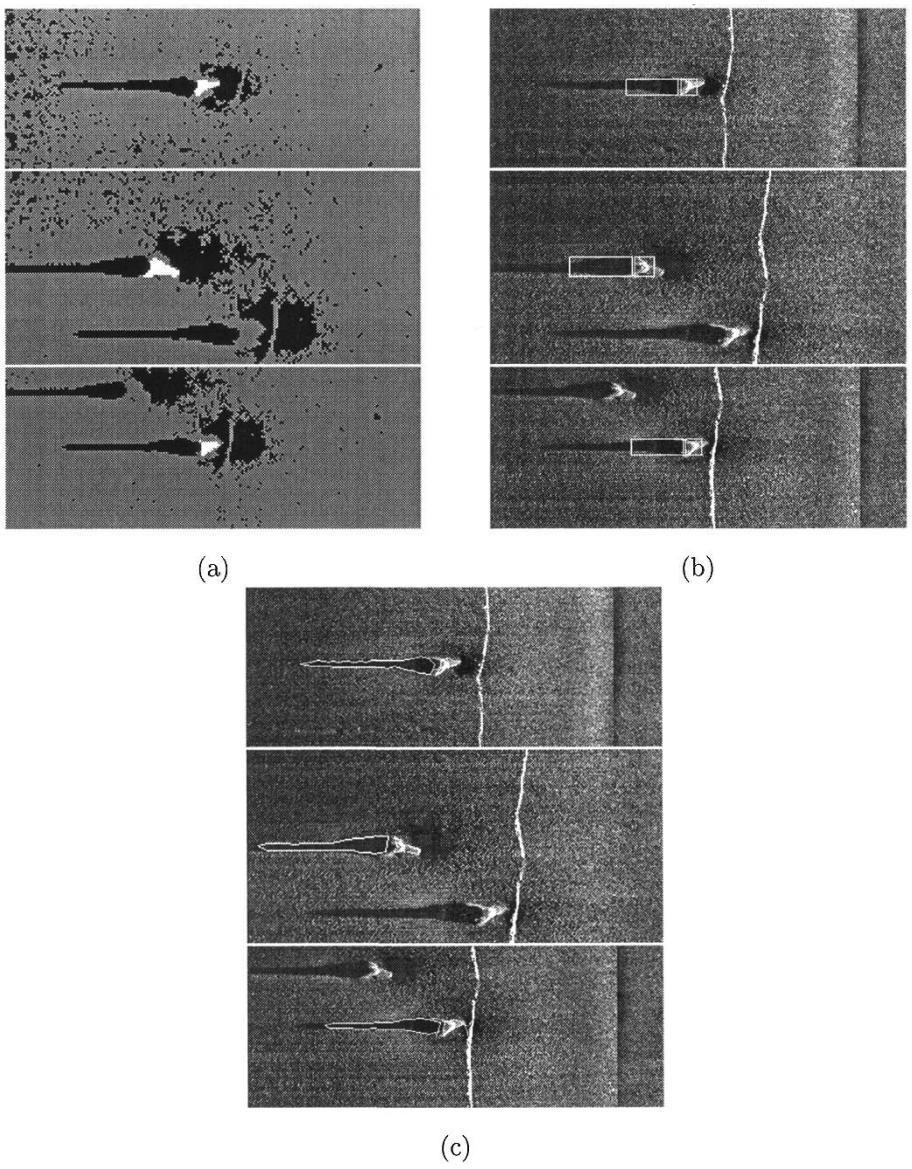


Fig. 14. (a) Label field mugshots of objects detected in Fig. 13 using the MRF detection model. (b) Initialization of the CSS model using label field. (c) Segmentation result using the CSS model.

a term that considered the homogeneity of the object-highlight pixels and shadow pixels within the object-highlight and shadow snakes, respectively, as well as the boxes' position with respect to the center of the object. This ensured confident initial conditions for the CSS snake every time.

Ideally the detection CSS models should be completely integrated, as demonstrated in Figs. 13 and 14, with the detection result providing the initialization step for the CSS snakes. However, due to the sensitivity of the data involved, mine images used for test purposes are often provided as mugshots, having already assumed that the object has been detected. This is the case for the rest of the data presented in this section of this paper.

For ease of notation, the raw data mugshots of the mines will be referred to as $\mathbf{y} = y(i, j)$, specifying a row and column position of each pixel instead of the $y_s$ notation used to refer to a specific data pixel in the detection model.

### A. The Statistical Snakes

Assume that the observed scene (the raw sidescan image) $\mathbf{y}$ is composed of three areas: object-highlight, shadow, and background, of which we wish to segment the object-highlight and shadow regions. We consider the image $\mathbf{y} = y(i, j)$ to be composed of $N_i \times N_j$ pixels where the highlight's gray levels $\mathbf{h}$, the shadow's gray levels $\mathbf{p}$, and the background pixels $\mathbf{b}$ are assumed to be uncorrelated and have $N_h$, $N_p$, and $N_b$, pixels, respectively. All three regions are described by probability density functions (pdfs) $p^{\mu_h}$, $p^{\mu_p}$, and $p^{\mu_b}$ where $\mu_h$, $\mu_p$, and $\mu_b$ are the parameters of the three pdfs.

We define a template window function $\mathbf{w} = \{w(i, j) \mid (i, j) \in [1, N_i] \times [1, N_j]\}$ which defines the shapes of the two snakes at any given time. Defining $w(i, j)$ to be equal to 2 inside the highlight, 1 inside the shadow, or 0 everywhere else, the image becomes composed of three regions

$\Omega_h = \{(i, j) \,|\, w(i, j) = 2\}$, $\Omega_p = \{(i, j) \,|\, w(i, j) = 1\}$, and $\Omega_b = \{(i, j) \,|\, w(i, j) = 0\}$. The observed image can be viewed as the sum of the three components

$$y(i,j) = h(i,j)\delta_{w(i,j)=2} + p(i,j)\delta_{w(i,j)=1} + b(i,j)\delta_{w(i,j)=0} \tag{19}$$

where $h(i, j)$, $p(i, j)$, and $b(i, j)$ are values drawn from their respective probability distributions and $\delta$ is the Kronecker delta function. Without any *a priori* knowledge, the best $\mathbf{w}$ is chosen by maximizing the likelihood

$$P[\mathbf{y} \,|\, \mathbf{w}, \mu_h, \mu_p, \mu_b] = P(\chi_h \,|\, \mu_h)P(\chi_p \,|\, \mu_p)P(\chi_b \,|\, \mu_b) \tag{20}$$

where

$$P(\chi_u \,|\, \mu_u) = \prod_{(i,j)\in\Omega_u} p^{\mu_u}[y(i, j)]. \tag{21}$$

The likelihood is expressed as a product of probabilities as the distributions are assumed to be uncorrelated and $\chi_u = y(i, j) \,|\, (i, j) \in \Omega_u (u = h, p \text{ or } b)$.

The likelihood function in (20) depends on the parameters of the probability functions as well as the template $\mathbf{w}$. The parameters $\mu_u$ where $u \in \{h, p, b\}$ are computed using a maximum-likelihood approach. Assuming that the three regions are described by exponential distributions allows the parameter estimates $\mu_u$ $u \in \{h, p, b\}$ to be injected back into (20) to obtain an expression for the likelihood that is simply dependent on the sums of gray levels [35]. However, the detection model described in Section III modeled the shadow, background, and object-highlight regions using separate noise distributions (Gaussian, Rayleigh, and triangular, respectively). Allocating an exponential distribution which can accurately model all three of these distributions is a difficult and unlikely task. For simplicity and based on the assumption that the regions are statistically quite separated, the Gaussian distribution was chosen as the most suitable exponential function to describe the three regions. While not exact, the results demonstrate that this assumption is sufficiently valid to provide accurate segmentation results. This leads to the log-likelihood function [36]

$$l_{\text{gauss}}(\mathbf{y}, \mathbf{w}) = -N_h H(\theta_h) - N_p H(\theta_p) - N_b H(\theta_b) \tag{22}$$

where $H(q) = \log(q)$ and

$$\theta_u = \frac{1}{N_u(\mathbf{w})} \sum_{(i,j)\in\Omega_u} y(i,j)^2 - \left\{ \frac{1}{N_u(\mathbf{w})} \sum_{(i,j)\in\Omega_u} y(i,j) \right\}^2 \tag{23}$$

where $N_u(\mathbf{w})$ is the number of pixels in region $u$ and $u \in \{h, p, b\}$. An iterative scheme must now be used to maximize this equation and produce the final segmentation result. To lower the computation time, the two-dimensional (2-D) summations in (23) can be converted to a one-dimensional (1-D) summation around the borders of the two snakes.

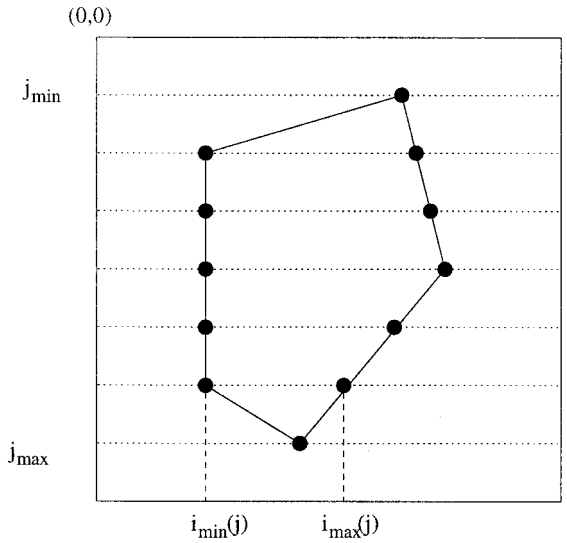


Fig. 15. Explanation of how the 2-D summation over the image gray levels can be modified to a 1-D summation around the snake.

### B. Converting the Likelihood to a 1-D Problem

The terms $\theta_h$, $\theta_p$, and $\theta_b$ need to be computed every iteration, making it important that the calculation is fast.

*1) Object-Highlight and Shadow Regions:* For the shadow and highlight regions, which are enclosed by their respective snakes and by considering Fig. 15, the first term on the right-hand side of (23) can be written as

$$\frac{1}{N_u(\mathbf{w})} \sum_{(i,j)\in\Omega_u} y(i,j)^2 = \frac{1}{N_u(\mathbf{w})} \sum_{j=j_{\min}}^{j=j_{\max}} \sum_{i=i_{\min}(j)}^{i=i_{\max}(j)} y(i,j)^2 \tag{24}$$

allowing the inside summation to be written as

$$\sum_{i=i_{\min}(j)}^{i=i_{\max}(j)} y(i, j)^2 = F[i_{\max}(j), j] - F[i_{\min}(j) - 1, j] \tag{25}$$

where

$$F[\tau, \eta] = \sum_{i=0}^{\tau} y(i, \eta)^2. \tag{26}$$

This term $F[\tau, \eta]$ can be deduced for every pixel before the segmentation process, so similarly, if we express $G[\tau, \eta] = \sum_{i=0}^{\tau} y(i, \eta)$, (23) can be rewritten as

$$\theta_u = \frac{1}{N_u(\mathbf{w})} \sum_{j=j_{\min}}^{j=j_{\max}} (F[i_{\max}(j), j] - F[i_{\min}(j) - 1, j]) - \left\{ \frac{1}{N_u(\mathbf{w})} \sum_{j=j_{\min}}^{j=j_{\max}} (G[i_{\max}(j), j] - G[i_{\min}(j) - 1, j]) \right\}^2 \tag{27}$$

for $u \in \{h, p\}$.

TABLE I
RELATIONSHIP BETWEEN $C(i, j)$, $\gamma(i, j)$ AND $\zeta(i, j)$ FOR THE TWO SNAKES, HIGHLIGHT AND SHADOW

| *snake* | $C(i,j)$ | $\gamma[(i,j)]$ | $\zeta[(i,j)]$ |
|---|---|---|---|
| highlight | 1 | F[i,j] | G[i,j] |
| highlight | -1 | F[i-1,j] | G[i-1,j] |
| highlight | 0 | 0 | 0 |
| highlight | 3 | $\frac{y^2(i,j)}{3}$ | $\frac{y(i,j)}{3}$ |
| shadow | 2 | F[i,j]/2 | G[i,j]/2 |
| shadow | -2 | F[i-1,j]/2 | G[i-1,j]/2 |
| shadow | 5 | 0 | 0 |
| shadow | 4 | $\frac{y^2(i,j)}{4}$ | $\frac{y(i,j)}{4}$ |

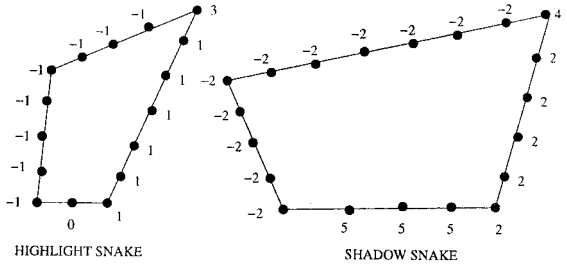


Fig. 16. Description of how the Huffman encoding system varies between the highlight and shadow snakes. This is so the model can distinguish between the separate regions.

This can be expressed as a summation around the snakes boundary $\delta_{\text{bound}}$ to give

$$\theta_u = \frac{1}{N_u(\mathbf{w})} \sum_{\delta_{bound}} C(i,j)\gamma[(i,j)] - \left\{ \frac{1}{N_u(\mathbf{w})} \sum_{\delta_{bound}} C(i,j)\zeta[(i,j)] \right\}^2 \quad (28)$$

for $u = \{h, p\}$ [35]. $C(i, j)$ is the associated Huffman Encoding for each point on the snake, $\zeta[(i, j)]$ and $\gamma[(i, j)]$ are defined in Table I, and the coding system is summarized in Fig. 16. Note that the Huffman codes have been arbitrarily chosen here so $C_{high}(i, j) \in \{-1, 0, 1, 3\}$ and $C_{\text{shad}}(i, j) \in \{-2, 2, 5, 4\}$ where the codes were computed by vector considerations as in Chesnaud *et al.* [35].

*2) The Background Region:* Once $\theta_h$ and $\theta_p$ have been calculated, it is simple to deduce $\theta_b$ by noting that

$$\sum_{(i,j)\in\Omega_b} y(i,j)^2 = \sum_{(i,j)\in\mathbf{y}} y(i,j)^2 - \sum_{(i,j)\in\Omega_h} y(i,j)^2 - \sum_{(i,j)\in\Omega_p} y(i,j)^2 \quad (29)$$

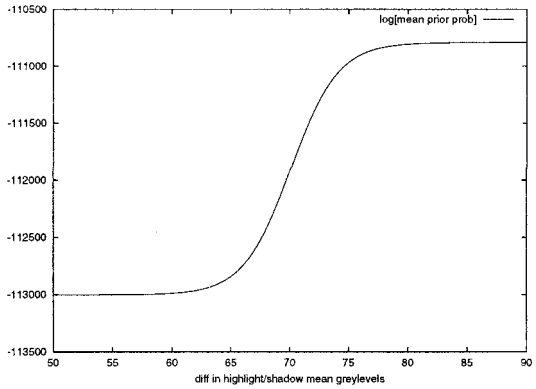


Fig. 17. The form of the log of the prior likelihood which awards large differences in the mean gray level of the regions inside the two snakes.

and similarly

$$\sum_{(i,j)\in\Omega_b} y(i,j) = \sum_{(i,j)\in\mathbf{y}} y(i,j) - \sum_{(i,j)\in\Omega_h} y(i,j) - \sum_{(i,j)\in\Omega_p} y(i,j) \quad (30)$$

to allow the calculation of $\theta_b$.

### C. Modeling the a Priori Information

As in the detection model, *a priori* information can be modeled to improve the model's capability to successfully segment the highlight and shadow. The CSS model uses two different priors to award certain snake configurations:

- a high difference in mean graylevel between the highlight and shadow snake;
- scenarios where the highlight snake and the shadow snake have similar centroid positions and similar heights.

These two priors and their effects on the shadow extraction will now be considered.

*1) The Mean Prior:* Object highlights are generally amongst the brightest regions of an image while shadow regions are amongst the darkest. A prior term of the form [10]

$$\log[P_{\text{mean}}(\mathbf{w})] = \mu \tanh\left(\frac{1}{2}\alpha(m_\Delta - \beta)\right) + c \quad (31)$$

was used to award scenarios where there was a large difference in the mean gray level of the pixels within the two snakes. $m_\Delta$ is the difference in the mean gray level, variables $\mu$ and $c$ are used to ensure that this log function lies in the same dynamic range as the log-likelihood term of the statistical snake while $\beta$ controls the tanh functions crossover location, and $\alpha$ controls the crossover rate. The general function has the form shown in Fig. 17.

As Fig. 17 demonstrates, $m_\Delta$ under a given value are all classified as equally "bad" while $m_\Delta$ over a given amount are all classified as equally "good." The transition between theses two states is controlled by $\alpha$, the crossover rate, which has been selected in this case to give a smooth change. The shape of this

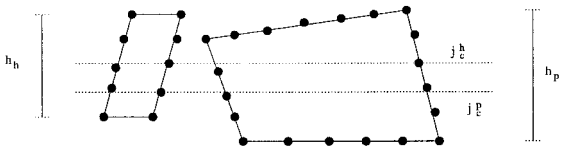

Fig. 18. *A priori* knowledge on the relationship between an objects highlight and its shadow can be used to constrain the two snakes.

prior term stops the two snakes from simply collapsing to ensure a high $m_{\Delta}$.

*2) The Position Prior:* The presence of sand ripples can often lead to an incorrect shadow extraction due to the ripple shadows corrupting the results. However, in the cases where an object highlight is present, it is possible to become more assertive as to which shadow regions are due to the object and which are not. We first define the variables $j_c^h$, $j_c^p$, $h_h$, and $h_p$ as described in Fig. 18 where $j_c^t$ is the $y$ coordinate of the center of snake $t = \{h, p\}$ and $h_t$ is the maximum height of snake $t$ such that

$$j_c^t = \frac{1}{N_t} \sum_{r \in \delta t} j_r \tag{32}$$

where $\delta t$ is the boundary of snake $t$, $N_t$ is the number of pixels on the boundary edge, and $h_t = \max\{j_r\} - \min\{j_r\}$ where $r \in \delta t$. We can now define the differences $\Delta j_c$ and $\Delta h$ where

$$\Delta j_c = j_c^h - j_c^p \tag{33}$$

$$\Delta h = h_h - h_p. \tag{34}$$

The ideal scenario is when both $\Delta j_c$ and $\Delta h$ are equal to zero and so we define the allowed spread in these variables as a Gaussian distribution with mean 0. Assuming that the distributions of $\Delta j_c$ and the $\Delta h$ are independent of each other, the combined log prior term can be written as a sum of the individual log terms to give

$$\log[P_{\text{position}}(\mathbf{w})] = C - t_1|\Delta j_c|^2 - t_2|\Delta h|^2 \tag{35}$$

where $t_1$ and $t_2$ and are constants. These determine the penalty for moving away from the ideal case where both $\Delta j_c$ and $\Delta h$ equal zero. $C$ is a constant to ensure that the prior lies in the same dynamic range as $l(\mathbf{y}, \mathbf{w})$.

*3) Determining the Prior Constants:* Both the mean and the position priors discussed previously contain constants which need to be determined before the segmentation can begin. This is carried out as a presegmentation calculation. The two snakes are initially restricted to only four points each and kept in rectangular form. A quick iterative process is carried out where the rectangular snakes' positions and dimensions (height, width, position, and distance apart) are altered randomly within an allowed range. At each position, the log-likelihood of the statistical snake is measured using (22). To simplify this process, both boxes have the same $j_c$ and height $h$ while the lengths of the highlight and shadow boxes are kept at $h$ and $3h$, respectively. This restricts the size of the parameter space to four parameters ($j_c$, $i_c^{\text{high}}$, $h$, and $d$) where $i_c^{\text{high}}$ is the $i$ coordinate of the center of the highlight box and $d$ is the distance between the two boxes. This simplistic box model for the two snakes allows a thorough search through the parameter space for estimating the dynamic range of the likelihood model as well as finding a favorable initial starting point for the two snakes.

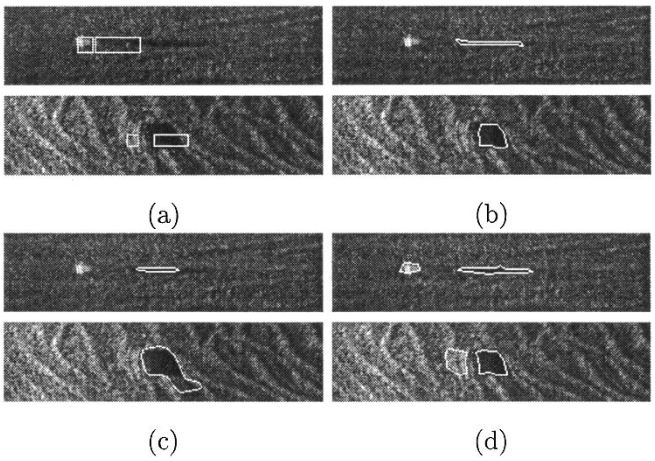


Fig. 19. (a) Two images containing mines with initial CSS snakes shown. (b) Segmentation result obtained using the SS model. (c) Segmentation result obtained using the MRF/CS model. (d) Segmentation result obtained using the CSS model.

As with all deformable models, a good initial starting point is highly desirable. If the mugshot image of the object came from the detection model detailed in Section III, an accurate initialization of the CSS model is possible using the label field, as shown in Figs. 13 and 14, where accurate size and positional data can be extracted. However, for the data shown in this section for which no *a priori* size or positional information is available, the initialization of the two snakes is carried out while the prior constants are being estimated and is determined by using the snake positions which maximize the difference in the mean gray levels of the two box-snakes $m_{\Delta}^{\max}$.

Using the rectangular snakes, the log likelihood $l(\mathbf{y}, \mathbf{w})$ from (22) was calculated iteratively. $L_{\min}$ and $L_{\max}$ were allocated the lowest and highest log likelihood found, respectively. Defining the largest difference in log likelihood $\Delta L = L_{\max} - L_{\min}$ allowed the prior constants to be defined as

$$c = L_{\min} + \frac{1}{2}\Delta L \quad \mu = \frac{1}{2}\Delta L \quad \beta = \frac{2}{3} m_{\Delta}^{\max} \tag{36}$$

$$\alpha = 0.5 \quad C = L_{\max} \quad t_1 = t_2 = \frac{\Delta L}{10}. \tag{37}$$

These values ensured that prior terms had roughly the same dynamic range as the log likelihood $l(\mathbf{y}, \mathbf{w})$. This is important for ensuring that the relative importance of the different terms can be controlled so that the log-likelihood term $l(\mathbf{y}, \mathbf{w})$ can remain the dominant segmentation term while the prior terms simply constrain the snakes' movements in a sensible manner.

### D. The Segmentation Process

The segmentation process has to maximize the posterior function

$$J(\mathbf{y}, \mathbf{w}) = \lambda_0 \log[P_{\text{reg}}(\mathbf{w})] + (1 - \lambda_0)(\lambda_1 \log[P_{\text{position}}(\mathbf{w})] + \lambda_2 l(\mathbf{y}, \mathbf{w}) + (1.0 - \lambda_1 - \lambda_2) \log[P_{\text{mean}}(\mathbf{w})]) \tag{38}$$

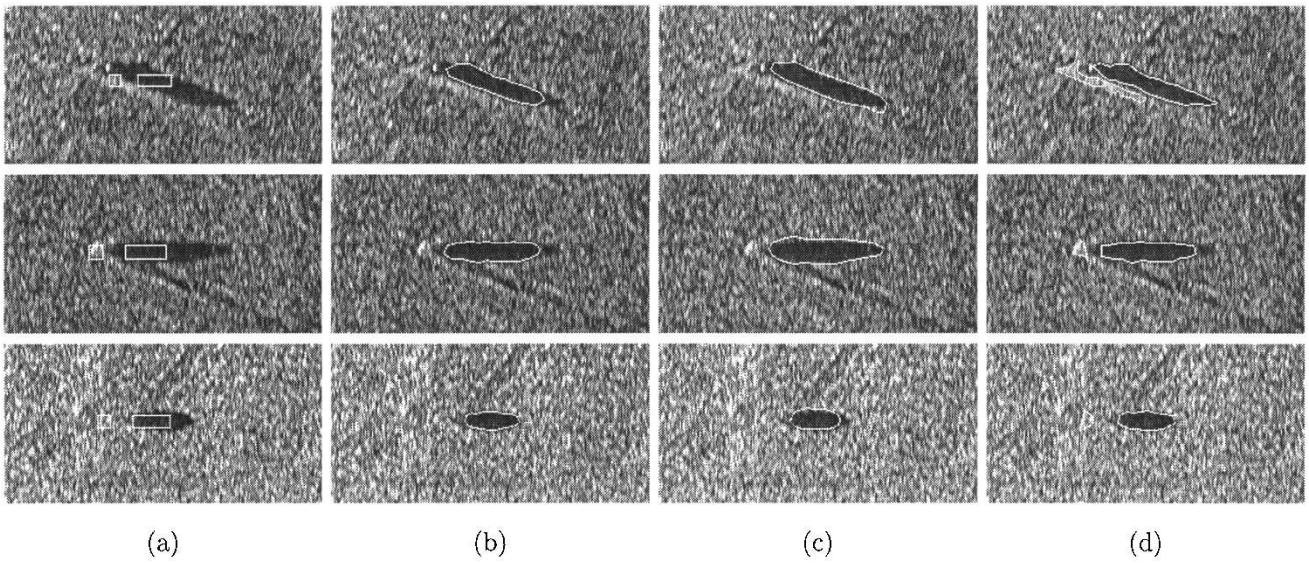


Fig. 20. (a) Three images containing mines with initial CSS snakes shown. (b) Segmentation result obtained using the SS model. (c) Segmentation result obtained using the MRF/CS model. (d) Segmentation result obtained using the CSS model.

where $\log[P_{\text{reg}}(\mathbf{w})]$ is a smoothing prior [35] and $\lambda_k \; k \in \{0, 1, 2\}$ are weights used to control the importance of each term.

A multiscale maximum approach was used to segment the highlight and shadow regions where both snakes were initialized with only four points. The iterative approach randomly selects a point after which its displacement from its old position is again determined randomly. The displacement uses two 1-D Gaussian proposal distributions such that, for the $i$-displacement, $i_{\text{new}} = i_{\text{old}} + \varsigma$ where $\varsigma$ is drawn from a Gaussian with mean $i_{\text{old}}$ and standard deviation $\sigma = 3$. After each displacement, a check must be carried out to ensure that none of the snake segments cross (the model operates under the assumption that the snakes are simply connected) after which the decision on whether to keep or reject the new configuration is made.

New points were added when convergence had been achieved with the present set of snake points (this was defined to be reached when the best fit solution had not changed for 200 iterations). $N$ new points were added between points $n$ and $m$ where $N$ was the integer solution to $d_{(n,m)}/10$, $d_{n,m}$ simply being the distance between the two points. This allows the snake progressively more flexibility as the algorithm proceeded. Accurate segmentation results were seen to be obtained after two additions of points.

Although all three terms lie within the same dynamic range, it was important that the statistical snake term $l(\mathbf{j}, \mathbf{w})$ remained the dominant term. $\lambda_1$ and $\lambda_2$ were maintained at 0.2 throughout while $\lambda_0$ was initialized at 0.0 and incremented by 0.05 every time new points were added. This ensured that the snakes maintained a smooth form as they were given more flexibility of movement.

### *E. Results*

Results are given on seven real and two synthetic sidescan images to allow the model to be tested over a large range of conditions. The performance of the CSS model is compared to the performance of two alternative models. The first is a single statistical snake (SS) model as described in [23]. The second is a classical-based snake technique as discussed in [24] where the image is first binarized using a two-class hierarchical MRF model (MRF-CS). The snake is driven by an energy term which considers both the homogeneity of shadow pixels inside the snake and the proximity of the snake to the edges of the binarized image described by an edge potential field [25]. As an aside, it should be noted that the MRF-CS model generally provides a smoother contour than the SS model due to its edge potential term. Rather than insisting that the MRF-CS snake lie directly on the shadow boundary, the edge potential term allows the snake to simply lie in the proximity of the edge and so generally acts as a smoothing agent to the model.

The initial starting point for the CSS model's snakes are also shown. As discussed before, when using raw sidescan data, the results from the detection model outlined in Section III can be used to accurately initialize the CSS model. However, as most of the data was obtained as mugshots, the CSS model was initialized using the method described in Section IV-C3 while the prior constants were being estimated. While this gave a poorer initialization point than using the detection result, the model is still successful in obtaining the correct segmentation. The other two models only segment the shadow and so were initialized using the CSS model's shadow initialization position.

Fig. 19 contains two images of objects lying on a flat seabed. The first image contains an object with clear object-highlight and shadow regions ensuring both the SS and the CSS models provided an accurate segmentation result. The MRF-CS solution detects a smaller shadow region as the MRF two-class segmentation removed part of the shadow region. The second image contains a sharp drop in graylevel with range as well as an object with very little highlight. Both the SS and the CSS model provide good segmentations (even though there is no distinctive highlight region) while the MRF-CS model gives a poor segmentation. This was due to the extreme range variation in graylevel leading to a poor MRF two-class segmentation.

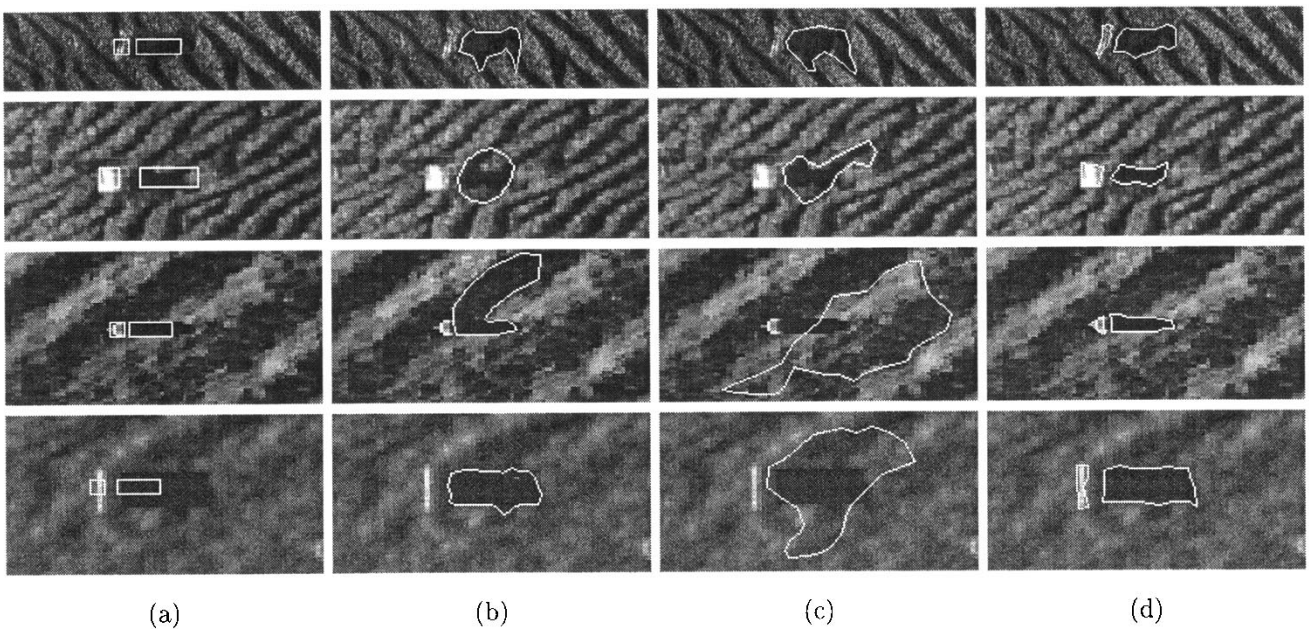


Fig. 21. (a) Four images containing mines on a sand ripple seabed with initial CSS snakes shown. (b) Segmentation result obtained using the SS model. (c) Segmentation result obtained using the MRF/CS model. (d) Segmentation result obtained using the CSS model.

Fig. 20 contains three noisy images of objects (one cylinder and two spheres) lying on a flat seabed. All three objects have either an indistinctive or no highlight region. However, the shadow regions are relatively clear and all three models provide accurate shadow segmentation results.

Fig. 21 contains two real and two synthetic images where the objects can be seen lying on sand ripple seabeds. In all four cases, both the SS and the MRF-CS models provide poor segmentation results as they cannot distinguish between the object shadow and the ripple shadows. The CSS model, constrained by its priors, achieves good segmentation results in all four cases.

### *F. Summary*

A novel CSS model has been presented for extracting the shadow of unknown objects in Sidescan imagery for future classification. Whilst the extraction of the shadow is relatively simple on a flat seabed, the presence of clutter or ripple shadows confuses the situation leading to inaccurate segmentations using standard techniques. *A priori* information on the expected signature of objects in Sidescan imagery was used to constrain the snakes' movement so that accurate segmentation results could be obtained regardless of the seabed type involved. The CSS model's extraction of the highlight region is also useful in the later classification phase, where the size and orientation information of the highlight region can be used to constrain the possible object shapes which could have produced the observed shadow region.

## V. Overall Conclusion and Future Research

This paper has presented automated models for both object detection and feature extraction in sidescan imagery. The detection model used spatial *a priori* knowledge on the size and geometry of object signatures in sidescan within the framework of an MRF model to provide accurate detection results even when large amounts of clutter were present. The model provides an interesting alternative to the current trend of trained detection models as in [1], [9], [2], and [14], making it applicable for a wide range of data without the problem of requiring suitable training data. This model was tested on both real and synthetic data offering good results in all cases.

Once an object has been detected, its shadow can be extracted for future classification. A novel CSS model was presented which extracted both the object highlight and its shadow. This technique demonstrated how the inclusion of *a priori* information could again provide more accurate results. Specifically, the problems inherent when considering complex seabed backgrounds as noted in [17] and [24] did not impact the accuracy of the results obtained using the CSS model. The CSS model was favorably compared with a statistical snake model and a MRF-based model with results presented on real and synthetic data.

Although this paper has concentrated on the detection of MLOs in sidescan imagery, suitable alteration of the priors involved would allow the described techniques to be applied to other fields such as pipeline or trawling scar detection. Future research will concentrate on using the CSS model results for classification purposes as well as developing a texture-orientated detection model thereby producing the building blocks for a complete automated classification system.

## Acknowledgment

The authors would like to thank B. Zerr at GESMA for providing many of the sidescan images for both the detection and the CSS models. They would also like to thank QinetiQ for providing data as well as the SACLANT Center, La Spezia, Italy, for allowing the inclusion of data from the BP'02 test runs.

## REFERENCES


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

[2] C. M. Ciany and W. Zurawski, "Performance of Computer Aided Detection/Computer Aided Classification and data fusion algorithms for automated detection and classification of underwater mines," in *Proc. CAD/CAC Conf.*, Halifax, NS, Canada, Nov. 2001.

[3] C. M. Ciany and J. Huang, "Computer aided detection/computer aided classification and data fusion algorithms for automated detection and classification of underwater mine," in *Proc. MTS/IEEE Oceans Conf. and Exhibition*, vol. 1, 2000, pp. 277–284.

[4] T. Aridgides, M. Frenandez, and G. Dobeck, "Adaptive 3-dimensional range–crossrange-frequency filter processing string for sea mine classification in side-scan sonar imagery," *Proc. SPIE*, vol. 3079, pp. 111–122, 1997.

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

[6] 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 *Proc. MTS/IEEE Oceans Conf. and Exhibition*, vol. 1, 2001, pp. 135–142.

[7] G. J. Dobeck, "Algorithm fusion for automated sea mine detection and classification," in *Proc. MTS/IEEE Oceans Conf. and Exhibition*, vol. 1, 2001, pp. 130–134.

[8] L. M. Linett, S. J. Clarke, C. St. J. Reid, and A. D. Tress, "Monitoring of the seabed using sidescan sonar and fractal processing," in *Proc. Conf. Underwater Acoustics Group*, 1993, vol. 15, pp. 49–64.

[9] L. M. Linett, D. R. Carmichael, S. J. Clarke, and A. D. Tress, "Texture analysis of sidescan sonar data," in *Proc. IEE Conf. Texture Analysis in Radar and Sonar* London, U.K., Nov. 1993.

[10] B. R. Calder, L. M. Linnet, and D. R. Carmichael, "Spatial stochastic models for seabed object detection," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 3079, pp. 172–182, 1997.

[11] S. G. Johnson and M. A. Deaett, "The application of automated recognition techniques to side-scan sonar imagery," *IEEE J. Oceanic Eng.*, vol. 19, pp. 138–144, Jan. 1994.

[12] S. Guillaudeux, "Some image tools for sonar image processing," in *Proc. CAD/CAC Conf.*, Halifax, NS, Canada, Nov. 2001.

[13] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Three class Markovian segmentation of high resolution sonar images," *Comput. Vis. Image Und.*, vol. 76, pp. 191–204, Dec. 1999.

[14] B. Calder, "Bayesian spatial models for SONAR image interpretation," Ph.D. dissertation, Heriot-Watt University, September 1997.

[15] B. R. Calder, L. M. Linnet, and D. R. Carmichael, "Bayesian approach to object detection in sidescan sonar," *IEE Proc.—Vis. Image Signal Proc.*, vol. 45, no. 3, 1998.

[16] 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. 6th Int. Symp. Unmanned Untethered Submersible Technology*, 1989, pp. 417–424.

[17] J. A. Fawcett, "Image-based classification of side-scan sonar detections," in *Proc. CAD/CAC Conf.*, Halifax, NS, Canada, Nov. 2001.

[18] P. H. Pidsley and M. A. Way, "Processing for CAD and image based CAC," in *Proc. CAD/CAC Conf.*, Halifax, NS, Canada, Nov. 2001.

[19] C. Xu and J. L. Prince, "Snakes, shapes and gradient vector flow," *IEEE Trans. Image Processing*, vol. 7, pp. 359–269, Mar. 1998.

[20] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, pp. 321–331, 1987.

[21] V. J. Myers, "Image segmentation using iteration and fuzzy logic," in *Proc. CAD/CAC Conf.*, Halifax, NS, Canada, Nov. 2001.

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

[23] S. Reed, J. Bell, and Y. Petillot, "Unsupervised segmentation of object shadow and highlight using statistical snakes," in *Proc. GOATS Conf.*, La Spezia, Italy, 2001.

[24] S. Reed, Y. Petillot, and J. Bell, "Unsupervised mine detection and analysis in side-scan sonar: A comparison of Markov Random Fields and statistical snakes," in *Proc. CAD/CAC Conf.*, Halifax, NS, Canada, Nov. 2001.

[25] 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. Machine Intell.*, vol. 22, pp. 129–141, Feb. 2000.

[26] I. Quidu, Ph. Malkasse, G. Burel, and P. Vilbe, "Mine classification based on raw sonar data: An approach combining Fourier descriptors, statistical models and genetic algorithms," in *Proc. OCEANS MTS/IEEE Conf. and Exhibition*, vol. 1, 2000, pp. 285–290.

[27] J. Bell, "A model for the simulation of sidescan sonar," Ph.D. dissertation, Heriot-Watt University, August 1995.

[28] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc.*, vol. B-48, no. 3, pp. 259–302, 1986.

[29] M. L. Comer and E. J. Delp, "Segmentation of textured images using a multiresolution Gaussian autoregressive model," *IEEE Trans. Image Processing*, vol. 8, pp. 408–420, Mar. 1999.

[30] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[31] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Processing*, vol. 9, pp. 1216–1231, July 2000.

[32] K. W. Stewart, D. Chu, S. Malik, S. Lerner, and H. Singh, "Quantitative seafloor characterization using a bathymetric sidescan sonar," *IEEE J. Oceanic Eng.*, vol. 19, pp. 599–610, Oct. 1994.

[33] W. Pieczynski, "Statistical image segmentation," *Mach. Graphics Vis.*, vol. 1, no. 1/2, pp. 261–268, 1992.

[34] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs Random Fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 39–55, Jan. 1987.

[35] C. Chesnaud, P. Refregier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 1145–1157, Nov. 1999.

[36] C. Chesnaud, V. Page, and P. Refregier, "Improvement in robustness of the statistically independant region snake-based segmentation method of target-shape tracking," *Opt. Lett.*, vol. 23, pp. 488–491, Apr. 1998.


**Scott Reed** received the M.Phys. degree in astrophysics (with Honors) from Edinburgh University, Edinburgh, U.K., in 1999 and the M.Sc. degree in information technology from Heriot-Watt University, Edinburgh. He is currently working toward the Ph.D. degree at Heriot-Watt University where he is researching object detection and classification models in underwater image processing.

His current research interests include Bayesian statistics, Markov random field models, pattern recognition techniques, image segmentation, and fuzzy logic.

**Yvan Petillot** received the Ph.D. degree from the Universite de Bretagne Occidentale in 1996. His dissertation work focused on real-time pattern recognition using optical processors.

He is currently a Lecturer in the school of Engineering and Physical Sciences in Heriot-Watt University, Edinburgh, U.K. Since finishing his doctoral work, he has been working on various successful research projects in the field of image processing (ESPRIT III, HICOPOS, MAST III and ARAMIS) and is currently an investigator on two European Union projects (FPV AMASON and AUTOTRACKER). His fields of interests include image analysis and understanding, sonar image processing, navigation, and sensor fusion.

**Judith Bell** received the M.Eng. degree (with Merit) in electrical and electronic engineering and the Ph.D. degree from Heriot-Watt University, Edinburgh, U.K., in 1992, and 1995, respectively. Her thesis examined the simulation of sidescan sonar images.

She is currently a Lecturer at Heriot-Watt University and is extending the modeling work to examine modeling of high-frequency active sonars and the use of such models for the verification and development of algorithms for processing sonar images.