

# Unsupervised Local Spatial Mixture Segmentation of Underwater Objects in Sonar Images

Avi Abu and Roeel Diamant 

**Abstract**—In this paper, we focus on the segmentation of sonar images to achieve underwater object detection and classification. Our goal is to achieve accurate segmentation of the object’s highlight and shadow regions. We target a robust solution that can manage different seabed backgrounds. Segmentation of sonar images is a challenging task. Speckle noise and intensity inhomogeneity may cause false detections and complex seabed textures, such as sand ripples and seagrass, often leading to false segmentation. In this paper, we propose our local spatial mixture (LSM) method for image segmentation of sidescan deployed sonar systems of any type. This new method estimates pixel labels in sonar images by incorporating the possible spatial correlation between neighboring pixels for improved segmentation. LSM modifies the expectation-maximization algorithm by adding an intermediate step (I-step) between the expectation (E-step) and maximization (M-step) steps. To combat intensity inhomogeneity, we employ a new initialization algorithm, one whose thresholds are set automatically to achieve and maintain robustness in various underwater environments. Using multiple evaluation indexes that measure the geometrical features of the segmented objects, we tested LSM using synthetic and real sonar images, one of which is obtained from our own sea experiment. Our results show that LSM achieves improved segmentation performance over the state-of-the-art methods of four different approaches; LSM is also robust to different seabed textures and intensity inhomogeneity. We share the sonar images from our sea experiments.

**Index Terms**—Expectation maximization (EM), gamma distribution, object detection, sand ripples, sonar image segmentation, speckle noise.

## I. INTRODUCTION

**S**ONAR images have been widely used in the fields of marine geology, marine biology, and marine archeology [1]. Sonar systems, such as the synthetic aperture sonar (SAS), multibeam sonar, and sidescan sonar, can produce high-resolution images of the seafloor, even in murky water. The importance of object detection and classification using remote sensing images can be found in a wide range of applications, such as natural geological hazards, environmental monitoring, and urban planning. An effective land-use classification method, based on a library

of pretrained object detectors, is proposed in [2]. In this paper, we offer local spatial mixture (LSM): a method that produces accurate segmentation of the object’s borders and of the image’s background. LSM is aimed specifically for images produced by a side-mounted sonar system, such as sidescan sonar, multibeam, and SAS. Image segmentation is the second step of the process of automatic detection and classification (ADAC) and its quality determines to a great extent the performance of the classifier. In this respect, LSM is a blind segmentation method that does not require *a priori* knowledge of the shape or size of the object. LSM is also robust to the types of both sonar system and seabed.

Sonar image segmentation is an important step in the ADAC chain. A typical ADAC scheme is presented in Fig. 1. The ADAC process works as follows: first, an ROI is detected from the original image. Then, segmentation is performed for the ROI. The segmentation results are used to extract the geometrical features of the detected objects. In the final step, a classification is performed based on the extracted features. In this paper, we focus on the segmentation part of the ADAC process for sonar images. We perceive segmentation as the key to effective ADAC performance. Three types of regions are identified in sonar images: highlight, shadow, and background [3]. The *highlight* originates from reflections of the acoustic wave from the object; *shadow* regions are created due to the lack of acoustic reverberation behind the object; and background originates from seafloor reverberation. An example of all three regions combined in a single sonar image is presented in Fig. 2. The background regions are identified by rectangles, and the highlight and shadow regions are identified by red lines. Our goal is to accurately identify the pixels in the image that relate either to the highlight or to the shadow regions of an object, and which are separated from the background. That is, in contrast to pure segmentation, we only care about the correct labeling of pixels in relation to the object.

Sonar image segmentation is mostly challenging due to strong reflections from the seafloor and the inhomogeneity of the water column. These phenomena can cause a false segmentation of the image background as an object highlight/shadow. Another challenge is posed by complex seabed textures, such as sand ripples, seagrass, and boulders. Boulders and seagrass may increase the areas in the image with false segmentation. Furthermore, the echoes of the object and sand ripples may blend together, causing a brutal distortion of the geometrical features of the object’s shadow. These distortions are because traditional segmentation techniques for (optical) natural images cannot be directly employed for sonar images.

Manuscript received November 12, 2017; revised March 19, 2018 and June 6, 2018; accepted August 1, 2018. This work was supported by the NATO Science for Peace and Security Programme under Grant G5293. This paper was presented in part at the 2017 IEEE OCEANS Conference, Aberdeen, Scotland, June 2017. (Corresponding author: Avi Abu.)

Associate Editor: J. Cobb.

The authors are with the Department of Marine Technology, University of Haifa, Haifa 3498838, Israel (e-mail: aabu@campus.haifa.ac.il; roeed@univ.haifa.ac.il).

Digital Object Identifier 10.1109/JOE.2018.2863961

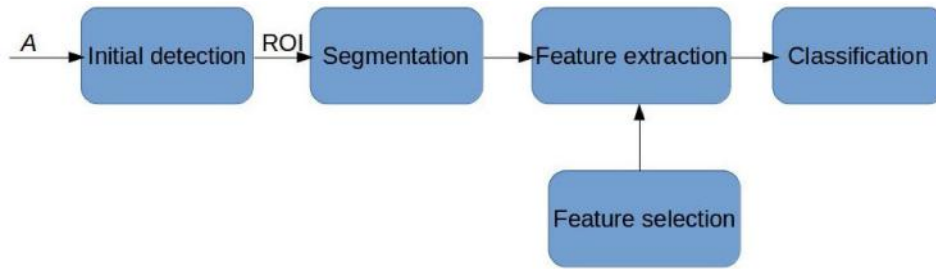


Fig. 1. Scheme of an ADAC system. Region of interest (ROI) is first detected. Segmentation is then extracted for each ROI. Based on the segmentation map, geometrical features of the detected objects within the ROI are measured. Using these features, the classifier assigns a class to each object. The contributions of this paper are focused on image segmentation.

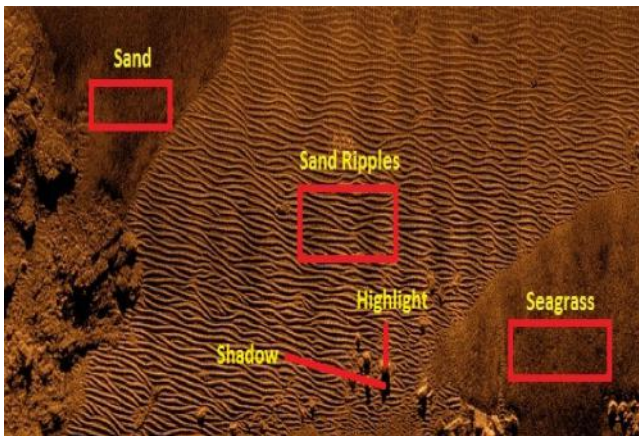


Fig. 2. Complex seabed texture in a sidescan sonar image. Examples of shadow and highlight regions are represented by red lines. Examples of background are represented by rectangles.

Supervised and unsupervised methods are the main approaches in sonar image segmentation. The former uses a database of labeled examples of the target object, whereas no prior training set is available for the latter. Given a reliably labeled database, supervised segmentation will produce better results than unsupervised segmentation. However, since sonar images tend to vary significantly, according to the angle and distance of the observed object, a reliable database is available only for a very limited set of objects. This problem is often exacerbated, due to the existence of rocks and the varying structure of the sea surface, which may produce false detections when using a limited database. In this paper, we focus on unsupervised image segmentation. We assume that we know the parameters of the sonar system (i.e., its carrier frequency and overlooking angle) as well as the depth of the water. We also assume that we have an upper bound for the size of the object and thus can calculate the expected size of the object in the sonar image. Still, we assume no training set for the object of interest. In other words, the actual shape and structure of the target in the sonar image are not known. Under this setting, we perform unsupervised sonar image segmentation.

Our LSM segmentation algorithm is targeted to be computationally inexpensive and to show robustness to intensity inhomogeneity. LSM first performs an initial segmentation of the sonar image into three clusters: highlight, shadow, and background. Then, segmentation is performed using the expectation–maximization (EM) algorithm [4]. To achieve robustness to

different backgrounds, we have solved the complex analysis for a generalized mixture distribution model. The result is a vast improvement in segmentation compared to the commonly used Gaussian mixture model. To avoid local minima points during the EM process, we issue a novel initialization procedure based on the detection of local shadow and local highlight regions.

The main contributions of this paper are summarized as follows.

- 1) A mixture model-based method to significantly improve segmentation accuracy compared to four benchmarks that represent the state-of-the-art methods: Dempster–Shafer’s theory, Markov random field (MRF), graph cut, and fuzzy logic.
- 2) A novel intermediate step that utilizes the expected spatial dependence of the image’s pixels. The result is a new soft–hard method, which evaluates the posterior.
- 3) A novel initialization procedure with an automatic selection of thresholds that handles the intensity inhomogeneity of sonar images and ultimately reduces the risk of false segmentation and local minima points.
- 4) A statistical analysis of the segmentation results and analysis of real sonar images using novel quantitative measures.

We investigate LSM performance in both numerical simulations and using real sonar images. For the former, we implemented a simulator to produce sonar images. For the latter, we use the EdgeTech database, which includes multiple sonar images from different marine environments [5]. The results show that, at a cost of a slight complexity increase, our algorithm exceeds the state-of-the-art methods of four different approaches in terms of false segmentation, object shape measures, and object geometrical features.

This paper is organized as follows. In Section II, we survey the state-of-the-art methods of image segmentation. In Section III, the system model is presented. Section IV contains the details of LSM, whereas numerical and experimental results are presented in Section V. Finally, concluding remarks are provided in Section VI. The notations used in this paper are summarized in Table I.

## II. RELATED WORK

We identify three main approaches for image segmentation. The first is the fuzzy-based segmentation, the second is the mixture-model MRF, and the third is the active contour

TABLE I  
LIST OF MAJOR NOTATIONS

Notation	Explanation
$A$	Original sonar image
$A_i$	Intensity level of $i$ th pixel
$N$	Number of pixels in $A$
$\mathcal{L}$	Number of labels
$l_i$	Label of $i$ th pixel
$I_i$	Indicator vector of $i$ th pixel
$\pi_{i,l}$	Indicator vector of $i$ th pixel
$u^i$	Horizontal axis of $i$ th pixel
$v^i$	Vertical axis of $i$ th pixel
$\phi(x)$	Gamma distribution
$(\alpha_l, \mu_l)$	Parameters of Gamma distribution (for the $l$ th label)
$\mathcal{N}_i$	Second-order neighborhood of $i$ th pixel
$O$	Group of outlier pixels
$\tau$	Incidence threshold
$\delta$	Intensity threshold
$\beta_s$	Shadow label threshold
$\beta_h$	Highlight label threshold
$\Theta$	Set of parameters involved in the mixture model
$\mathbf{y}$	The complete data set in the mixture model
$\bar{A}(u, v)$	The integral image of $A$
$\mathcal{B}_s$	Background map
$\mathcal{S}$	Shadow map
$\mathcal{H}$	Highlight map

(AC) method. Regarding the former, fuzzy  $C$ -means (FCM) is a widely used algorithm in image segmentation [6] and [7]. FCM performs segmentation via nonlinear optimization, which is based on an objective function. The MRF approach usually involves the estimation of the posterior probability to manage the spatial dependencies between neighboring pixels, such that outliers with respect to their neighborhood are excluded [8]. The AC model has been applied successfully to image segmentation [9], and is considered to be robust to noise and to generate smoothed and accurate contours. The segmentation problem can also be solved using graph theory. For example, in graph cut image segmentation, by minimizing the Gibbs energy function, a label is assigned to each node in the graph's representation of the image [10]. While fuzzy-based and AC-based segmentation

work well for certain types of images, they require manual adaptation for a different seabed structure. To obtain robustness to background types and to combat nonhomogeneity, we find mixture model-based segmentation to be a better fit. Mixture-models MRF-based segmentation is less sensitive to noise and outliers in sonar image than FCM, AC, and graph cut. This is mainly since MRF utilizes the local spatial interactions between neighboring pixels; as a result, the label of the pixel is less influenced by outliers. On the contrary, in FCM, the spatial correlation between neighboring pixels is not taken into account; in AC, the gradient method leads to low segmentation accuracy; and graph cut based segmentation is very sensitive to initialization. To handle the nonhomogeneity in sonar images and to avoid trapping to local minima, a manual adaptation of crucial parameters is needed in fuzzy-based segmentation, whereas in AC, manual initialization is required [11]. Also, the approach of parametric shape templates [12] in AC is severely degraded in sonar data, and should be adapted manually to generate satisfactory results. In contrast, employing a mixture model of distributions, such as the gamma distribution, allows flexibility in segmenting sonar data. This is because, by setting its parameters appropriately, the representation through gamma distribution captures a verity of distributions [13]. The flexibility of MRF-based segmentation, such as the MRF in [8], in handling noise in different regions in the sonar image is reflected by the prior distribution that varies for each pixel corresponding to each label and is set according to the local spatial information.

Due to its importance and numerous applications, the problem of sonar image segmentation has been widely studied. Recently, the AC method has been applied successfully for image segmentation [14] and [15]. Two main approaches exist for AC methods: *edge based* and *region based*. *Edge-based* methods utilize image gradients to identify the object's boundaries [16] and [17]. In sonar image segmentation, these are related to the boundaries of the highlight and shadows that characterize an object. *The region-based* method models the background and object regions statistically and finds an optimum, where the model best fits the image [18]. Another interesting approach involves the integration of the image's geometric properties. This is performed in the gradient vector flow (GVF) algorithm, which extends the gradient vector of an edge map, thereby suppressing noise. Wang *et al.* [19] proposed normally biased GVF to combat weak edge preserving and to achieve noise robustness.

The main limitation of the above-mentioned algorithms is their sensitivity to the initialization step. In particular, if the AC is initialized without comprising a part of the object, the algorithm will diverge. Recently, Fandos and Zoubir [20] proposed a new polygonal AC algorithm for sonar imagery that utilizes the likelihood function of the object region and background region. The initial contour is drawn from the results of the Markovian segmentation. Li *et al.* [21] proposed the region scalable fitting (RSF) model that utilizes local information in the cost function to deal with intensity inhomogeneity. This local information is based on a weighted average of pixel intensities in a Gaussian window in the object region and background region. However, since the cost function of the RSF model is nonconvex, the final segmentation results are very sensitive to the initial selection of

the segments. Javed *et al.* [22] proposed a method that improves RSF in terms of accuracy and efficiency. The model is built on local entropy and local variance as inputs to fuzzy inference engine to extract proper weights to be used in the cost function. Yang *et al.* [15] proposed a nonlocal means-based despeckling filter as a pretreatment to improve robustness to multiplicative noise for sidescan sonar images. The authors also integrated an edge-driven constraint into the RSF model to avoid convergence to local minima and to reduce complexity. Jing *et al.* [23] proposed a novel global minimization AC model (GMACM), which is a convex version of the RSF. GMACM has been applied for oil slick segmentation in synthetic aperture radar (SAR) images with satisfactory results. However, GMACM only utilizes the local mean information, which is not sufficient to accurately model an image with intensity inhomogeneity; hence, its performance is limited. Recently, Song *et al.* [24] proposed a globally statistical AC model (GSACM) that takes into account not only the local average, but also the local variance of the image's intensities. GSACM is more robust to the initial selection of object and background boundaries since the algorithm solves an unconstrained convex cost function, in contrast to the GMACM, which omits the constrained term when minimizing the cost function.

In [11], Fandos *et al.* apply a high complexity graph theory based method aimed for a synthetic aperture sonar image. The method splits the image into two groups of pixels, one assigned to the shadow region and the other to the background. The iterative conditional modes (ICM) method is used for the segmentation of the highlight region. Celik and Tjahjadi [25] propose an unsupervised multiscale segmentation method for seabed mapping using sonar imagery. A multiresolution representation of the image is obtained by applying the wavelet transform. A feature set is then computed for each pixel and  $K$ -means clustering is used. For initialization, the particle swarm optimization method is used in [26], whereas in [27], initialization is based on the fractal segmentation. Maximum likelihood (ML) was used to estimate the distribution parameters of the Markov model, and the iterative conditional estimation algorithm serves for final segmentation.

Finite mixture models have also been widely used in image segmentation to incorporate spatial correlation between pixels [28] and [29]. Zhang *et al.* [30] substituted the prior probability of the pixel label with an estimate of the MRF distribution. The diffused EM (DEM) algorithm [31] uses an anisotropic diffusion step between the E-step and the M-step of the EM algorithm. Recently, the MRF and Dempster-Shafer theory (DST) have been utilized to handle spatial correlation among neighboring pixels [32] in sonar images. In DST-based clustering, pixel labels are used as evidence. Then, according to Dempster's rule, all possible combinations of hypotheses are considered for decision-making for pixel  $i$ . The work by Nguyen and Wu [8], which represents the state-of-the-art method in MRFs-based segmentation, incorporated the spatial correlation among neighboring pixels by using a novel smoothing prior. This new model was applied in the MRF distribution to form the log-likelihood function initialized via the  $K$ -means algorithm. Yet, a limitation of this method is that a single parameter controls the smoothing prior,

which may limit the tolerance to noise. Wu and Chung [33] proposed a compound MRF model to represent the boundary of an object. Analyses conducted on simulated images, as well as on real clinical images, reveal an improvement in segmentation and accurate object boundary results. Pan *et al.* [34] proposed a new mixture model, based on the Student's  $t$ -distribution [35] and the MRF. The main advantage of the Student's  $t$ -distribution is that it is heavily tailed, and hence provides a much more robust approach than the standard Gaussian mixture model (GMM). In [36], at a high computational cost, Nikou *et al.* used local spatial interactions between neighboring pixels. The probability of pixel labels is modeled using the Dirichlet compound multinomial density function. To impose smoothness, the Gauss MRF is utilized on the Dirichlet parameters.

While the above surveyed sonar segmentation methods perform well in simple underwater environments, they fail to segment the image in complex seabed textures, such as sand ripples or seagrass. This is mostly due to the problem of intensity inhomogeneity, but also due to nonaccurate segmentation initialization. Considering this challenge, we offer a novel robust initialization algorithm and a new model for incorporating the spatial information among neighboring pixels.

### III. SYSTEM MODEL

This paper focuses on the segmentation of sonar images produced from sonar systems mounted on autonomous underwater vehicles (AUV). The goal of image segmentation is to accurately estimate the pixel labels. Let  $A$  be a two-dimensional (2-D) sonar image of size  $N$  with  $A_i, i = 1, 2, \dots, N$ , being the intensity level of the  $i$ th pixel. Let  $\mathcal{L}$  be a set of labels for the image pixels with  $|\mathcal{L}| = L$ , where  $|\cdot|$  stands for the dimension of a group. We consider three labels  $\mathcal{L} = \{\mathbb{S}, \mathbb{B}, \mathbb{H}\}$ , where  $\mathbb{S}$ ,  $\mathbb{B}$ , and  $\mathbb{H}$  stand for shadow, background, and highlight, respectively. The result of the segmentation process is an assignment  $l_i \in \mathcal{L}$  for each pixel  $i$ . Let  $\mathbf{e}_l$  denote an  $L \times 1$  vector whose  $l$ th component is one and the rest are zero. Also, let  $\mathbf{I}_i$  be an indicator vector with  $\mathbf{I}_i = (I_{i,1}, \dots, I_{i,L})^T \in \{\mathbf{e}_1, \dots, \mathbf{e}_L\}$ .  $\mathbf{I}_i$  is a discrete random variable with a probability function defined by  $\text{Prob}(\mathbf{I}_i = \mathbf{e}_l) = \pi_{i,l} \forall i, l$  whose distribution is given as

$$f(\mathbf{I}_i) = \prod_{l=1}^L \pi_{i,l}^{I_{i,l}}. \quad (1)$$

We use the gamma distribution with parameters  $\alpha_l$  and  $\mu_l$  suitable for sonar images [37] such that

$$\begin{aligned} f(A_i | I_{i,l} = 1, \theta_l) &= \frac{\left(\frac{\alpha_l}{\mu_l}\right)^{\alpha_l} (A_i)^{\alpha_l-1} \exp\left(-\frac{\alpha_l A_i}{\mu_l}\right)}{\Gamma(\alpha_l)} \\ &\equiv \phi(A_i | \theta_l) \end{aligned} \quad (2)$$

where  $\theta_l = [\alpha_l, \mu_l]^T$  and  $\Gamma(\cdot)$  is the gamma function, and we have

$$f(A_i | \mathbf{I}_i, \theta_l) = \prod_{l=1}^L \phi(A_i | \theta_l)^{I_{i,l}}. \quad (3)$$

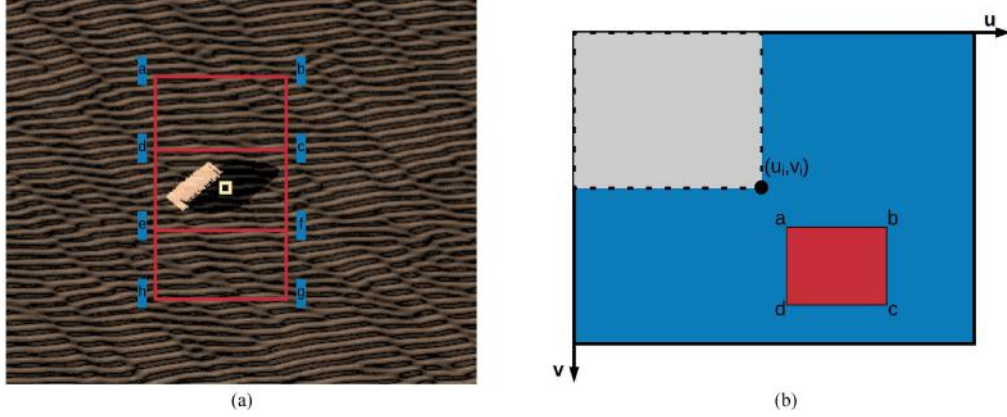


Fig. 3. Schematic diagrams of the windows template for background, shadow and highlight calculation, and the integral image. (a) Upper and lower red windows contribute to the calculation of the background  $B_s^i$ , and the yellow window contribute to the calculation of the shadow map  $S^i$  and the highlight map  $H^i$ . (b) Illustration of the integral image. The blue window refers to the original image, and the sum of all pixels in the gray window is  $\bar{A}(u^i, v^i)$ .

The set of parameters is defined by  $\Theta \equiv (\pi_1, \dots, \pi_N, \theta_1, \dots, \theta_L)$ , where  $\pi_i = (\pi_{i,1}, \dots, \pi_{i,L})$ ,  $i = 1, \dots, N$ . Using (1) and (3), we get the likelihood as follows [4]:

$$f(A_i, \mathbf{I}_i | \theta_i) = \prod_{l=1}^L [\pi_{i,l} \phi(A_i | \theta_l)]^{I_{i,l}}. \quad (4)$$

Denote the complete data set by  $\mathbf{y}^T = (\mathbf{A}^T, \mathbf{I}_1^T, \dots, \mathbf{I}_N^T)$ , where  $\mathbf{A}$  is an  $N \times 1$  vector containing all pixel intensities. Then, assuming the components of  $\mathbf{A}$  are independent identically distributed, as well as the components of  $\{\mathbf{I}_i\}$ , the conditional distribution of the complete data set is

$$f(\mathbf{y} | \Theta) = \prod_{i=1}^N \prod_{l=1}^L [\pi_{i,l} \phi(A_i | \theta_l)]^{I_{i,l}}. \quad (5)$$

Note that, while traditionally the prior does not depend on specific pixels, in (5), the priors  $\pi_{i,l}$  are pixel dependent. We use this formalization to allow incorporating intensity inhomogeneity of sonar images. Also, the likelihood function in (5) does not incorporate local correlations between neighboring pixels. Hence, applying the ML algorithm directly on (5) does not utilize all available information.

#### IV. PROPOSED LSM METHOD

In this section, we present the details of our LSM segmentation method. LSM is composed of two main stages: the first is the initialization stage and the second is the clustering stage. The clustering stage is based on a modification of the EM algorithm and is initialized by the first stage as a good starting point to avoid trapping to a local minimum. We start with the initialization of the segmentation procedure and follow with a full description of LSM.

##### A. Initialization Method

Due to the frequent speckle noise in sonar images, it is crucial to produce good initial clusters to improve the convergence rate as well as the final segmentation results. Referring to the pseudocode in Algorithm 1, our initialization process is

composed of three stages. In the first stage, the shadow segmentation process is performed. In the second stage, we identify the highlight-related samples. Finally, the remaining pixels are assigned the background label and the prior is initialized with  $\pi_i^{(1)} = (0, 1, 0)$ . To accelerate the processing time of the shadow and highlight segmentations, we process the integral image.

1) *Integral Image*: An *integral image* is a representation of the original image, which allows for a very fast computation of rectangular arrays [38]. Let  $u^i$  and  $v^i$  denote the horizontal and vertical axes of the  $i$ th pixel in image  $A$ , respectively. Let  $\mathcal{K}_{(u,v)} \triangleq \{(j) : u^j \leq u, v^j \leq v\}$  be the set of all pixels located above and to the left of the pixel at location  $(u, v)$ . The integral image of  $A$  is given by

$$\bar{A}(u, v) = \sum_{j \in \mathcal{K}_{(u,v)}} A_j. \quad (6)$$

Fig. 3(a) illustrates the process of a windowing architecture. The upper and lower red windows contribute to the calculation of the background, whereas the yellow window contributes to the calculation of the shadow and highlight. Using the integral image, the mean value of pixels inside a window is easily calculated and complexity improves. In Fig. 3(b), the blue region stands for the original image and the sum of all pixels in the gray region is  $\bar{A}(u^i, v^i)$ . Let  $\{a, b, c, d\}$  be the four corners of a window [see Fig. 3(b)], then the sum of all pixels inside this window is given by:  $\bar{A}(u^a, v^a) - \bar{A}(u^b, v^b) - \bar{A}(u^d, v^d) + \bar{A}(u^c, v^c)$ .

2) *Shadow Segmentation*: Two maps are calculated based on the integral image, namely the background map and the shadow map. The background value of pixel  $i$  in the background map is given by

$$B_s^i = \frac{1}{2\gamma_b^2} \left[ \bar{A}(u^a, v^a) - \bar{A}(u^b, v^b) - \bar{A}(u^d, v^d) + \bar{A}(u^c, v^c) \right. \\ \left. + \bar{A}(u^e, v^e) - \bar{A}(u^f, v^f) - \bar{A}(u^h, v^h) + \bar{A}(u^g, v^g) \right] \quad (7)$$

where  $\{a, b, c, d\}$  and  $\{e, f, g, h\}$  are the four corners of the upper and lower windows, respectively [see Fig. 3(a)]. The above-

mentioned value is calculated using a split-window template composed of three equally sized squared windows with a size of  $\gamma_b^2$ . The size of the window is set according to the size of the object of interest such that the object is contained within the middle window. The parameter  $\gamma_b$  is selected such that an object of interest is fully contained within, and as a result, it does not contribute to the estimation of the background. Therefore, the exact size of the object is not required, but only its bounds. Let the center of this template be in a  $(u^i, v^i)$  position. The background map is the mean pixel value while accounting for the top and bottom windows only. We avoid the middle window since the object of interest may be covered by this window. These windows are stacked vertically and not horizontally since the highlight of the object, in sonar imagery, lies from left/right (depends on the sonar system) of the shadow region.

The shadow map  $\mathcal{S}^i$  corresponds to the mean value of the pixels at a squared window located at  $(u^i, v^i)$  with a size of  $\gamma_s^2$

$$\mathcal{S}^i = \frac{1}{\gamma_s^2} \left[ \bar{A}(u^a, v^a) - \bar{A}(u^b, v^b) - \bar{A}(u^d, v^d) + \bar{A}(u^c, v^c) \right]. \quad (8)$$

Parameter  $\gamma_s$  is chosen to be small enough to keep the object's boundaries. Our experiments reveal that  $\gamma_s = 5$  gives the best results. The intensity of pixels in a shadow region is lower than the intensity of the background. Thus, shadow segmentation is extracted by identifying pixels with a local mean intensity that is lower than the mean background intensity. For shadow segmentation, we develop a new initialization method that best fits sonar images. Our method is inspired by Williams [1]. However, the work in [1] deals only with object detection, where the features of the object are of less interest, whereas to achieve the additional aim of object classification, our work also segments the object's regions. More specifically, in contrast to [1], in our case, the size of the window for calculating the shadow map should be small to maintain accurate segmentation along the object's boundaries. Moreover, we automatically select the threshold parameter according to the image's intensity distribution to allow for accurate segmentation in all of the object's regions. The label of the  $i$ th pixel is initialized as a shadow label if  $\mathcal{B}_s^i > \beta_s \mathcal{S}^i$ , where  $\beta_s$  is an automatically set threshold parameter, as we will explain further. In this case, the prior for pixel  $i$  is replaced with  $\pi_i^{(1)} = (1, 0, 0)$ .

3) *Highlight Segmentation*: In sonar images, the background's intensity level is mostly dependent on the structure of the seabed. A sonar image may contain different types of seabed structures, such as sand or boulders. In such a case, the variations in the background's intensity level are high and the task of highlight segmentation is challenging. Considering this challenge, we apply the  $k$ th-order statistics [39]. Let  $\mathcal{R}_{(u,v)}^{\gamma_s}$  and  $\mathcal{R}_{(u,v)}^{\gamma_b}$  be the set of all pixels in a squared window with sizes  $\gamma_s^2$  and  $\gamma_b^2$ , respectively, centered at  $(u, v)$  excluding the  $k$ th pixels of the largest intensities in each row of the window. The highlight segmentation is calculated using a small rectangle centered in the pixel of interest.

Formally, The highlight map  $\mathcal{H}^i$  is given by

$$\mathcal{H}^i = \sum_{j \in \mathcal{R}_{(u^i, v^i)}^{\gamma_s}} A_j \quad (9)$$

and the background map is

$$\mathcal{B}_h^{i_0} = \sum_{j \in \mathcal{R}_{(u^{i_0}, v^{i_0})}^{\gamma_b}} A_j \quad (10)$$

where  $i_0$  is an arbitrarily located pixel in the upper right-hand corner of the original image. We label pixel  $i$  as highlight if  $\mathcal{H}^i > \beta_h \mathcal{B}_h^{i_0}$ , where  $\beta_h$  is an automatically set threshold parameter, as explained further. The prior is replaced with  $\pi_i^{(1)} = (0, 0, 1)$ . Note that unlike  $\mathcal{B}_s^i$ , which is calculated for each pixel in the image,  $\mathcal{B}_h^{i_0}$  is constant to preserve the object's highlight boundaries.

4) *Automatic Setting of Threshold Parameters*: We now describe a method to automatically set the threshold parameters  $\beta_s$  and  $\beta_h$ . This is required to increase the robustness of our algorithm to different seabed environments. Using a fast approach with reasonable segmentation results, we utilize the ICM method [40] to extract an initial partition of the image into three clusters, namely highlight, shadow, and background, denoted by  $C_{\mathbb{H}}$ ,  $C_{\mathbb{S}}$ , and  $C_{\mathbb{B}}$ , respectively. Next, we calculate *intensity thresholds*  $\mathcal{F}_s$  and  $\mathcal{F}_h$ , defined as the intensities for which the number of pixels with false segmentation is minimized. The intensity threshold  $\mathcal{F}_s$  is determined by solving

$$\mathcal{F}_s = \arg \min_{\mathcal{F}} \left( \sum_{i=1}^{|C_{\mathbb{S}}|} \mathcal{I}(A_i | i \in C_{\mathbb{S}} > \mathcal{F}) + \sum_{j=1}^{|C_{\mathbb{B}}|} \mathcal{I}(A_j | j \in C_{\mathbb{B}} < \mathcal{F}) \right) \quad (11)$$

where  $\mathcal{I}$  is an indicator function that is unity if its argument is true, and zero otherwise. Similarly,  $\mathcal{F}_h$  is given by solving

$$\mathcal{F}_h = \arg \min_{\mathcal{F}} \left( \sum_{j=1}^{|C_{\mathbb{B}}|} \mathcal{I}(A_j | j \in C_{\mathbb{B}} > \mathcal{F}) + \sum_{k=1}^{|C_{\mathbb{H}}|} \mathcal{I}(A_k | k \in C_{\mathbb{H}} < \mathcal{F}) \right). \quad (12)$$

Since the thresholds  $\beta_s$  and  $\beta_h$  stand for the gap between the local shadow's mean and background and local highlight's mean and background, we set the threshold parameters  $\beta_s$  and  $\beta_h$  as the ratio between the intensity thresholds and the average of the shadow or background clusters. Formally

$$\beta_s = \frac{\mathcal{F}_s}{\frac{1}{|C_{\mathbb{S}}|} \sum_{i \in C_{\mathbb{S}}} C_{\mathbb{S}}^i} \quad (13)$$

$$\beta_h = \frac{\mathcal{F}_h}{\frac{1}{|C_{\mathbb{B}}|} \sum_{j \in C_{\mathbb{B}}} C_{\mathbb{B}}^j}. \quad (14)$$

The initialization process is described in Fig. 4, and its pseudocode is given in Algorithm 1.

## B. Image Segmentation Using Modified EM

Our LSM sonar image segmentation algorithm is similar in its structure to the DEM method [31]. Yet, while DEM showed

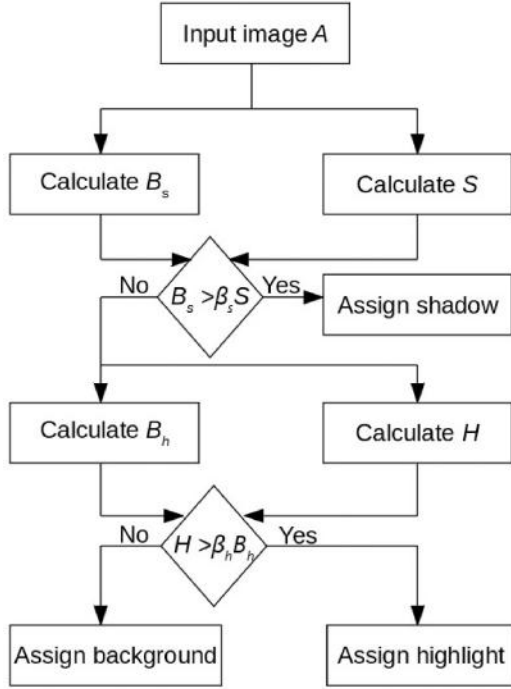


Fig. 4. Flowchart of the proposed initialization algorithm. Shadow and highlight regions are detected by local windows. The thresholds  $\beta_s$  and  $\beta_h$  are set automatically. Pixels that are labeled as neither shadow nor highlight are labeled as background.

satisfactory results for natural images, the results for sonar images severely degrade (see [37, Figs. 4 and 5]). To fit the case of sonar images, our LSM method replaces the DEM's diffusion step with an intermediate step, which takes into account the local intensity inhomogeneity in sonar images to achieve more accurate segmentation results. The key idea is to achieve robustness to different seabed structures by allowing the labeling of each pixel to be influenced by the states of its neighbors. While MRF-based methods utilize the spatial dependence among neighboring pixels via the smoothing prior according to some model, our approach is based on utilizing both the local texture and local intensity of the neighborhood.

1) *Gamma Mixture Model for Spatially Independent EM*: The basic EM consists of a conditional expectation (E-step) of the log likelihood and its maximization (M-step). The E-step is given by

$$\mathcal{C}(\Theta|\Theta^{(m)}) \equiv E\{\ln(f[\mathbf{y}|\Theta])|\Theta^{(m)}, \mathbf{A}\} \quad (15)$$

where  $\Theta^{(m)}$  are the estimates of the parameter sets  $\Theta$  at the  $m$ th iteration. Based on (5) and (15), the conditional expectation is given by

$$\begin{aligned} \mathcal{C}(\Theta|\Theta^{(m)}) &= E\left[\sum_{i=1}^N \sum_{l=1}^L I_{i,l} [\ln(\pi_{i,l}) + \ln(\phi(A_i|\theta_l))]\right] \Big| \Theta^{(m)}, \mathbf{A} \\ &= \sum_{i=1}^N \sum_{l=1}^L \underbrace{E(I_{i,l}|\Theta^{(m)})}_{v_{i,l}^{(m)}} [\ln(\pi_{i,l}) + \ln(\phi(A_i|\theta_l))]. \end{aligned} \quad (16)$$

---

#### Algorithm 1: Proposed Initialization Algorithm.

---

**Input:** The original image  $A$ ;

**Output:** The pixel labels;

**Begin**

- 1: Calculate  $\bar{A}$  using (6);
  - 2: Calculate  $\mathcal{B}_h^{i_0}$  using (10);
  - 3: Set the thresholds  $\beta_s$  and  $\beta_h$  using (13) and (14);
  - 4: **for** each pixel  $i$  **do**
  - 5:   Calculate background map  $\mathcal{B}_s^i$  using (7);
  - 6:   Calculate shadow map  $\mathcal{S}^i$  using (8);
  - 7:   **if**  $\mathcal{B}_s^i > \beta_s \mathcal{S}^i$  **then**
  - 8:     pixel  $i$  is labeled as shadow;
  - 9:   **else**
  - 10:    Calculate highlight map  $\mathcal{H}^i$  using (9);
  - 11:    **if**  $\mathcal{H}^i > \beta_h \mathcal{B}_h^{i_0}$  **then**
  - 12:     pixel  $i$  is labeled as highlight;
  - 13:    **else**
  - 14:     pixel  $i$  is labeled as background;
  - 15:    **end if**
  - 16:   **end if**
  - 17: **end for**
- End**
- 

The expression for the posterior  $v_{i,l}^{(m)}$  is given by [4]

$$v_{i,l}^{(m)} = \frac{1}{\mathcal{D}} \pi_{i,l}^{(m)} \phi(A_i|\theta_l^{(m)}) \quad (17)$$

where  $\mathcal{D}$  is a normalization factor given by  $\sum_{l=1}^L \pi_{i,l}^{(m)} \phi(A_i|\theta_l^{(m)})$ . In the M-step, the parameter set  $\Theta^{(m)}$  is updated by

$$\Theta^{(m+1)} = \arg \max_{\Theta} \mathcal{C}(\Theta|\Theta^{(m)}) \quad (18)$$

with the constraints  $\sum_{l=1}^L \pi_{i,l} = 1$  for  $i = 1, \dots, N$ . This maximization problem is solved using a Lagrange multiplier  $\eta$  and by maximizing the following quantity:

$$\Upsilon = \sum_{l=1}^L v_{i,l}^{(m)} [\ln(\pi_{i,l}) + \ln(\phi(A_i|\theta_l))] + \eta \left( \sum_{l=1}^L \pi_{i,l} - 1 \right). \quad (19)$$

The solution of  $\partial \Upsilon / \partial \pi_{i,l} = 0$  yields the following equation:

$$\frac{v_{i,l}^{(m)}}{\pi_{i,l}} + \eta = 0. \quad (20)$$

From (17), we can get

$$\sum_{l=1}^L v_{i,l}^{(m)} = 1 \quad (21)$$

and using (20) and (21), we get the update

$$\pi_{i,l}^{(m+1)} = v_{i,l}^{(m)}. \quad (22)$$

$A_{i,1}$	$A_{i,2}$	$A_{i,3}$
$A_{i,8}$	$A_i$	$A_{i,4}$
$A_{i,7}$	$A_{i,6}$	$A_{i,5}$

Fig. 5. The second-order neighborhood configuration of pixel  $i$  in  $\mathcal{N}_i^l$ . This local neighborhood contains eight pixels.

The update for the parameter  $\mu_l$  is the solution of the equation  $\partial\Upsilon/\partial\mu_l = 0$  and is given by the closed form expression

$$\begin{aligned}\mu_l^{(m+1)} &= \frac{\sum_{i=1}^N \pi_{i,l}^{(m+1)} A_i}{\sum_{i=1}^N \pi_{i,l}^{(m+1)}} \\ &= \frac{\sum_{i=1}^N v_{i,l}^{(m)} A_i}{\sum_{i=1}^N v_{i,l}^{(m)}}.\end{aligned}\quad (23)$$

The estimation of  $\alpha_l$  is set by solving  $\partial\Upsilon/\partial\alpha_l = 0$ , i.e.,

$$\ln(\alpha_l^{(m+1)}) - \psi(\alpha_l^{(m+1)}) = -\frac{\sum_{i=1}^N v_{i,l}^{(m)} \ln\left(\frac{A_i}{\mu_l^{(m)}}\right)}{\sum_{i=1}^N v_{i,l}^{(m)}}\quad (24)$$

where  $\psi(\cdot)$  is the digamma function. Since (24) has no closed solution, we adopt the following approximation [41]:

$$\ln(x) - \psi(x) \approx \frac{1}{2x} + \frac{1}{12x^2}.\quad (25)$$

Substituting (25) into the left-hand side of (24) leads to an approximate closed solution for  $\alpha_l^{(m+1)}$ , as the positive root of the equivalent quadratic equation.

2) *Intermediate Step*: In LSM, we utilize the expected spatial dependence between the image's pixels through an I-step performed in-between the E-step and the M-step. To that end, we already assign labels to pixels during the I-step such that in the M-step, we update the set parameters  $\Theta^{(m)}$  based on the assigned labels. The labeling of pixels in the I-step is performed by

$$\hat{l}_i^{(m)} = \arg \max_{l \in \mathcal{L}} v_{i,l}^{(m)}\quad (26)$$

where  $v_{i,l}^{(m)}$  is obtained during the E-step. Let the neighboring group of a pixel  $i$  be  $\mathcal{N}_i$ . Note that  $\mathcal{N}_i$  includes a set of pixels such that  $i \notin \mathcal{N}_i$  and  $\forall i' \in \mathcal{N}_i, i \in \mathcal{N}_{i'}$ . In our application, the second-order neighborhood system is chosen [20]. The neighborhood of pixel  $i$  is defined by  $\mathcal{N}_i = \{A_{i,1}, A_{i,2}, A_{i,3}, A_{i,4}, A_{i,5}, A_{i,6}, A_{i,7}, A_{i,8}\}$ . Fig. 5 shows such a neighborhood configuration. Let  $\Psi_{\mathcal{N}_i}^{\check{l}}$  be the number of pixels in  $\mathcal{N}_i$  with an assigned label equal to  $\check{l}$  and  $\tau$  and  $\delta$  be the

incidence threshold and intensity threshold, respectively. Let  $\mathcal{O}$  be the group of outlier pixels. Then, pixel  $i \in \mathcal{O}$  if the following two conditions hold:

- 1)  $\exists \check{l} \in \mathcal{L}$  s.t.  $\Psi_{\mathcal{N}_i}^{\check{l}} > \tau$ ;
- 2)  $|A_i - \text{median}(\mathcal{N}_i)| > \delta$ .

Due to the intensity inhomogeneity of sonar images, we give special treatment to pixels, which are identified as outliers. For each identified outlier, we enforce the smooth posterior to maintain the maximum value of the label that corresponds with the majority of its pixel's neighborhood. Labeling in the I-step is performed by introducing a smoothed posterior,  $\bar{v}_{i,l}^{(m+1)}$  whose value in the  $(m+1)$ th iteration is

$$\bar{v}_{i,l}^{(m+1)} = \begin{cases} \zeta_{i,l}, & \text{if } i \in \mathcal{O} \\ \frac{\Lambda_{i,l}^{(m+1)}}{\sum_{l=1}^L \Lambda_{i,l}^{(m+1)}}, & \text{otherwise} \end{cases}\quad (27)$$

$$\zeta_{i,l} = \begin{cases} 1, & \text{if } l = \hat{l}_i^{(m)} \\ 0, & \text{otherwise} \end{cases}\quad (28)$$

and the weight  $\Lambda_{i,l}^{(m+1)}$  of each  $i$ th pixel for each label  $l$  is

$$\Lambda_{i,l}^{(m+1)} = \left[ \frac{1}{\sqrt{2\pi} (\sigma_l)^2} \cdot \exp\left(-\frac{(A_i - \mu_l^{(m)})^2}{(\sigma_l)^2}\right) \right]^{1 - \frac{\Psi_{\mathcal{N}_i}^{\check{l}}}{|\mathcal{N}_i|}}.\quad (29)$$

The parameter  $\sigma_l$  is the standard deviation of the pixels belonging to label  $l$  and is calculated using the parameter sets  $\theta_l$  as follows:

$$\sigma_l = \frac{\mu_l^{(m)}}{\sqrt{\alpha_l^{(m)}}}.\quad (30)$$

Note that in (29), to exclude local outliers and to keep the boundaries of the object, higher weight is given to the majority labels of the local neighborhood as well as to the label that suits the statistics of the pixel, i.e.,  $\Lambda_{i,l} \rightarrow 1$  as  $\Psi_{\mathcal{N}_i}^{\check{l}} \rightarrow |\mathcal{N}_i|$ , which reflects the local label information. Moreover, as the statistics of  $A_i$  suits the model of the  $l$ th label,  $\Lambda_{i,l}$  gets higher value. From (29), we observed that a priority is given to a label  $l$  in  $\Lambda_{i,l}^{(m+1)}$  whose  $\Psi_{\mathcal{N}_i}^{\check{l}}$  value is high. This means that the smooth prior of the  $i$ th pixel seeks a homogeneous environment.

3) *Connecting the I-step and the M-step*: The connection between the I-step and the M-step is performed as follows. First, posterior  $v_{i,l}^{(m)}$  is substituted by  $\bar{v}_{i,l}^{(m+1)}$  from (27) in the update of the prior probability  $\pi_{i,l}$  in the M-step. Next, the parameter set is updated through a hard decision on the smooth posterior. This is done to improve the convergence rate. Formally, at the beginning of the M-step, we replace the posterior  $v_{i,l}^{(m)}$  with

$$\bar{\omega}_{i,l}^{(m+1)} = \begin{cases} 1, & \text{if } l = \hat{l}_i^{(m+1)} \\ 0, & \text{if } l \neq \hat{l}_i^{(m+1)} \end{cases}\quad (31)$$

where the assigned label for pixel  $i$  at iteration number  $(m+1)$  is given by

$$\hat{l}_i^{(m+1)} = \arg \max_{l \in \mathcal{L}} \bar{v}_{i,l}^{(m+1)}.\quad (32)$$



**Algorithm 2:** LSM Segmentation Algorithm.**Input:** The original image  $A$ ;**Output:** Pixel labels  $\{\tilde{l}_i\}$ ;**Begin**

- 1: Initialize the priors  $\{\pi_i^{(1)}\}, \forall i$ , using algorithm (1);
- 2: Set  $\{v_{i,l}^{(1)}\} \leftarrow \{\pi_i^{(1)}\}$ ;
- 3: Calculate  $\{\theta_l^{(1)}\}, \forall l \in \mathcal{L}$ ;
- 4: Set  $m = 2$ ;
- 5: **while**  $m < M_{\max}$  **do**
- 6: Calculate the posterior probabilities  $\{v_{i,l}^{(m)}\}$  using (17);
- 7: Perform a hard decision on  $\{v_{i,l}^{(m)}\}$  and get pixels label  $\{\tilde{l}_i^{(m)}\}$ ;
- 8: Calculate  $\{\bar{v}_{i,l}^{(m+1)}\}$  and  $\{\bar{\omega}_{i,l}^{(m+1)}\}$  using (27)–(32);
- 9: Substitute  $\{v_{i,l}^{(m)}\} \leftarrow \{\bar{v}_{i,l}^{(m+1)}\}$  and update prior probabilities using (22);
- 10: Calculate gamma distribution parameters using (23) and (24) with  $\{v_{i,l}^{(m)}\} \leftarrow \{\bar{\omega}_{i,l}^{(m+1)}\}$ ;
- 11: **if**  $|\mathcal{U}^{(m)} - \mathcal{U}^{(m-1)}| < \epsilon$  **then**
- 12: Stop the loop and output the pixel labels  $\{\tilde{l}_i^{(m)}\}$ ;
- 13: **else**
- 14: Set the loop counter  $m = m + 1$  and go to (6);
- 15: **end if**
- 16: **end while**
- 17: Get pixel labels  $\{\tilde{l}_i^{(m)}\}$ ;

**End**

To summarize the operation of our segmentation algorithm, we illustrate a single cycle of the segmentation process in Fig. 6. The algorithm iterates until the number of maximum iteration steps  $M_{\max}$  is reached or the results converge. Let  $\mathcal{U}^{(m)}$  be the log of the likelihood given in (5). Then, the latter is determined when  $|\mathcal{U}^{(m)} - \mathcal{U}^{(m-1)}| < \epsilon$ , where  $\epsilon$  is a user-defined parameter. The pseudocode for the process is given in Algorithm 2.

## V. EXPERIMENTAL AND COMPARATIVE STUDIES

In this section, we explore the performance of LSM, as well as those of four state-of-the-art methods that represent different segmentation approaches and that yielded the best results for the segmentation of sonar images (to achieve object detection and classification). The first benchmark is the E-DS-M [32], which uses the evidence theory in the EM's I-step; the second is the Nguyen method in [8], which employs an improved formalization for the smoothing prior in the MRF model; the third benchmark is the robust fuzzy thresholding technique [42], which utilizes soft clustering via fuzzy membership degree; and the fourth benchmark is the graph cut algorithm [10], which uses a graph representation of the image. An efficient implementation of the graph cut method for sonar image segmentation [43] has been utilized. In this implementation, the initial seeds are chosen after the ICM segmentation results. The implementation of the graph cut algorithm, for sonar imagery, was performed according to [11]. In this implementation, the highlight region was segmented according to the ICM method, and the graph cut

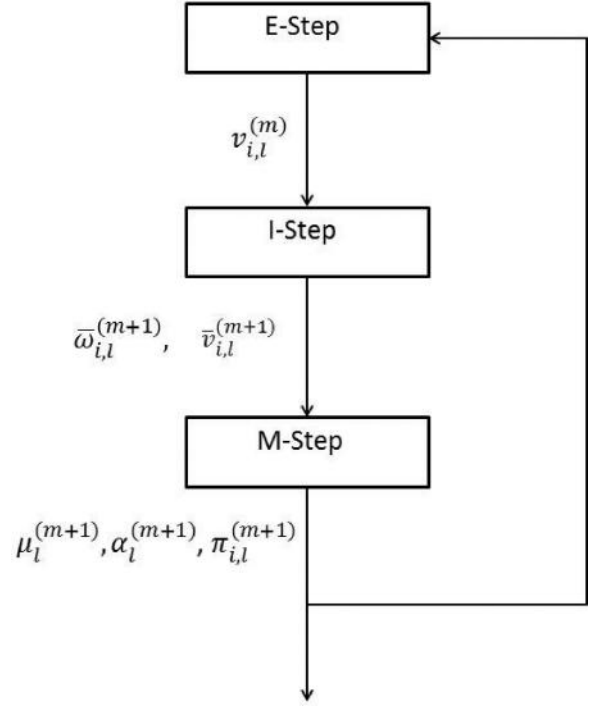


Fig. 6. Single cycle of the proposed segmentation algorithm. The intermediate step is inserted between the E-step and M-step of the EM algorithm so as to utilize local spatial information.

algorithm splits the image into two groups of pixels, one assigned to the shadow and the second to the background.

To study the effect of the seabed texture on the proposed segmentation method, three types of seabed textures are analyzed: sand, seagrass, and sand ripples. The proposed initialization algorithm is compared with the most common approaches, namely the  $K$ -means [44] algorithm and the four-region [45] method. We test performance over both simulated sonar images and real sonar images. To verify the performances of LSM as blind segmentation method, we analyze the segmentation results on different kinds of object's shapes. This includes a synthetic cylindrical object and arbitrary geometry of real underwater objects.

Unless otherwise mentioned, the following setting is considered: the compared methods are initialized with a four-region, with a window of  $5 \times 5$  pixels. This window has been demonstrated to produce satisfactory results [45]. The parameters of the proposed initialization algorithm are set with:  $\gamma_b = 50$ ,  $\gamma_s = 5$ , and  $k = 3$ , which give the best initialization results. The threshold parameters  $\tau$  and  $\delta$  are set to 4 and 1.5, respectively. The maximum number of iterations is set to 50 and the convergence threshold is set to  $\epsilon = 0.001$ . To choose the parameters of the benchmark methods, we perform a statistical analysis. For the E-DS-M, we choose  $\gamma_1 = 0.1$  and  $\gamma_2 = 1.5$  as the best fit, as obtained in the numerical analysis of [32]. For the Nguyen method,  $\beta = 12$  is used [8] and the neighborhood  $\partial_i$  is a square window with a size of  $3 \times 3$  pixels. For the fuzzy thresholding method, the best results are obtained using a median filter with a neighborhood block with a size of  $3 \times 3$  and a fuzziness index

of 2 [42]. As a good compromise [43], the parameters  $h$  and  $v$  of the graph cut implementation were set with 0.1

### A. Quantitative Measures

The performance analysis of a sonar image segmentation is more than the counting of erroneous pixel identification and should involve determining the shape and formation of the segmented objects. Yet, we have not yet found a proper definition for such measures. Considering this need, in this section, we formalize our quality measures for the segmentation of underwater images containing a submerged object. As statistical quantitative measures, we use the misclassification ratio (MCR) [46] and the probabilistic rand (PR) index [47]. MCR is defined as the number of misclassified pixels normalized by the total number of pixels and is given by

$$\text{MCR} = \frac{\sum_{i=1}^N \mathcal{I}(\bar{I}_i \neq G_i)}{N}. \quad (33)$$

The PR index measures the similarity between each pixel pair in the image and is given by

$$\text{PR} = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i < j}} \left[ \mathcal{I}(\bar{I}_i = \bar{I}_j) p_{ij} + \mathcal{I}(\bar{I}_i \neq \bar{I}_j) (1 - p_{ij}) \right] \quad (34)$$

where  $p_{ij} = \text{Prob}(G_i = G_j)$ . Clearly, segmentation improves, the higher the PR is the lower the MCR is.

To allow for a quantitative assessment of the segmented object, we determine three shape measures, namely, the position fit  $\zeta_p$ , the shape fit  $\zeta_s$ , and the size fit  $\zeta_n$  [48]. The *position fit* is given by

$$\zeta_p = \frac{1}{2} \sum_{c \in C_{\text{H}}, C_{\text{S}}} \left( 1 - \frac{\Delta_u^c + \Delta_v^c}{2} \right) \quad (35)$$

where  $\Delta_u^c$  is the difference in the horizontal direction of the center of the object (in region  $c$ ) between the segmentation results and the ground-truth map,  $\Delta_v^c$  is the difference in the vertical direction, and  $C_{\text{H}}$  and  $C_{\text{S}}$  are the highlight and shadow clusters, respectively. Explicit expressions for  $\Delta_u^c$  and  $\Delta_v^c$  are given by

$$\Delta_u^c = \frac{1}{N_u^c} |\bar{u}_{S_i}^c - \bar{u}_{G_i}^c| \quad (36)$$

$$\Delta_v^c = \frac{1}{N_v^c} |\bar{v}_{S_i}^c - \bar{v}_{G_i}^c| \quad (37)$$

where  $N_u^c$  and  $N_v^c$  are the number of rows and columns in the image,  $\bar{u}_{S_i}^c$  and  $\bar{v}_{S_i}^c$  are the average of the abscissas and ordinates of the object (in region  $c$ ), and  $\bar{u}_{G_i}^c$  and  $\bar{v}_{G_i}^c$  represent the ground-truth map averages. The size-fit quality measure is given by

$$\zeta_n = \frac{1}{2} \sum_{c \in C_{\text{H}}, C_{\text{S}}} \left( 1 - \frac{|N_f^c - N_g^c|}{N_f^c + N_g^c} \right) \quad (38)$$

where  $N_f^c$  and  $N_g^c$  are the number of pixels of the object at region  $c$  in the segmentation results map and the ground-truth

map, respectively. Shape fit is given by

$$\zeta_s = \frac{1}{2} \frac{N_{f \cap g}^c}{N_{f \cup g}^c} \quad (39)$$

where  $N_{f \cap g}^c$  is the number of pixels in the intersection between the results map and the ground-truth map in the object region  $c$  and  $N_{f \cup g}^c$  is the union between the two maps. Clearly, segmentation results improve as  $\zeta_p$ ,  $\zeta_s$ , and  $\zeta_n$  increase.

To further examine segmentation performance, we consider the geometrical features of the segmented object. Specifically, we consider the perimeter  $P$  of the object, area ratio  $R_{\text{area}}$ , and solidity  $\text{Sol}$ . The *area ratio* is defined as the ratio between the areas of the shadow and the highlight regions, and *solidity* is the ratio between the area of the object and the area of its convex hull. Based on these geometrical features, we define the following evaluation indices:

$$\kappa_{\text{perim}} = \frac{1}{2} \left( |P_f^{\text{S}} - P_g^{\text{S}}| + |P_f^{\text{H}} - P_g^{\text{H}}| \right) \quad (40)$$

$$\kappa_{\text{sol}} = \frac{1}{2} \left( |\text{Sol}_f^{\text{S}} - \text{Sol}_g^{\text{S}}| + |\text{Sol}_f^{\text{H}} - \text{Sol}_g^{\text{H}}| \right) \quad (41)$$

$$\kappa_{\text{area}} = |R_{\text{area},f} - R_{\text{area},g}| \quad (42)$$

where  $P_f^{\text{S}}$  and  $P_g^{\text{S}}$  are the segmented and ground-truth perimeters of the object's shadow, respectively,  $P_f^{\text{H}}$  and  $P_g^{\text{H}}$  are the segmented and ground-truth perimeters of the object's highlight, respectively,  $\text{Sol}_f^{\text{S}}$  and  $\text{Sol}_g^{\text{S}}$  are the segmented and ground-truth solidities of the object's shadow, respectively,  $\text{Sol}_f^{\text{H}}$  and  $\text{Sol}_g^{\text{H}}$  are the segmented and ground-truth solidities of the object's highlight, respectively, and  $R_{\text{area},f}$  and  $R_{\text{area},g}$  are the segmented and ground-truth area ratios of the object, respectively. In general, segmentation improves, the lower the values of  $\kappa_{\text{perim}}$ ,  $\kappa_{\text{area}}$ , and  $\kappa_{\text{sol}}$  are.

### B. Results on the Synthetic Images

An example of the synthetic sonar images used in our simulations is shown in Fig. 9(a). The figure contains three images of sand, seagrass, and sand ripples. Sand and seagrass seabed textures are generated according to the models in [49], whereas the sand ripples texture is generated according to [50]. The object region and background are synthesized separately. The object's intensity level is modeled by a gamma distribution with  $\mu_s = 120$  and  $\sigma_s = 10$  for the shadow region and  $\mu_h = 1$  and  $\sigma_h = 0.1$  for the highlight region. Similar to the model proposed in [37], the object and background are superimposed as follows:

$$A_i = \begin{cases} 0.8\chi_{i,h} + 0.2\chi_{i,b}, & \text{if } i \in \{C_{\text{H}}\} \\ 0.8\chi_{i,s} + 0.03\chi_{i,b}, & \text{if } i \in \{C_{\text{S}}\} \\ \chi_{i,b}, & \text{if } i \in \{C_{\text{B}}\} \end{cases} \quad (43)$$

where  $\chi_{i,h}$ ,  $\chi_{i,b}$ , and  $\chi_{i,s}$  stand for the intensity level of the  $i$ th pixel in the highlight, background, and shadow regions, respectively. The size of the synthetic images is  $300 \times 300$ . A cylinder structure is considered as the object in all synthetic images. To test performance statistically, we performed 1000 Monte Carlo simulation runs. In each run, a new background

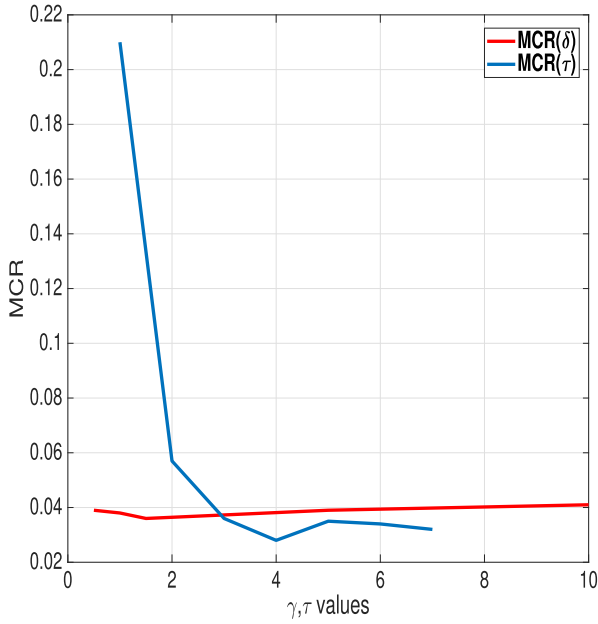


Fig. 7. Misclassification ratio (MCR) values of LSM with different  $\gamma$  and  $\delta$  values using the synthetic seagrass image in Fig. 9(a). The best results are obtained with  $\tau = 4$  and  $\delta = 1.5$ .

image was realized, each of which was corrupted by a speckle noise with an STD of 0.3.

1) *Parameters Study of  $\gamma$  and  $\delta$* : To comment on the sensitivity of our algorithm to parameter choice, in this section, the effect of the two parameters  $\gamma$  and  $\delta$  is analyzed. These two parameters control the size of the outliers group  $\mathcal{O}$ . Specifically, a small value of  $\gamma$  increases the number of outliers, whereas a large value increases the false segmentation. In contrast,  $\delta$  controls the relative intensity threshold for outlier declaration. We study these two parameters in the synthetic *seagrass* image presented in Fig. 9(a). In Fig. 7, we show the MCR (33) values calculated for different values of  $\gamma$  and  $\delta$ . To set the optimal thresholds, a 2-D search for the threshold combination of the incidence threshold and intensity threshold is required. Fortunately,  $\Psi_{N_i}^l \in \{1, \dots, 8\}$  such that the complexity of this operation is low. Hence, we calculate the  $MCR(\gamma, \tau_j)$  values  $\forall \tau_j = \{1, \dots, 8\}$ , and select those  $\gamma$  and  $\tau$  values that lead to minimum MCR. We observe that the best MCR results are obtained for  $\gamma = 4$  and  $\delta = 1.5$ .

2) *Results for the Proposed Initialization Algorithm*: The initialization results of the benchmark methods,  $K$ -means, four-region, and our LSM are presented in Fig. 8 for the *seagrass* (upper row) and *sand ripples* (lower row) examples. We notice that the object's shadow is almost fully blended with shadow pixels. Similarly, due to the high-intensity inhomogeneity that exists in the *seagrass* image, the results presented in Fig. 8(c) show false segmentation for four-region initialization. However, the results also show that the borders of the highlight and shadow of the cylindrical object are preserved. The results of Fig. 8(e) show that the proposed initialization method achieves a clean segmentation performance. For the sand ripple background, the results presented in Fig. 8(g)–(j) show a much more noisy segmentation for all four initialization methods. Note that

the results of the proposed method in Fig. 8(e) and (j) for pixels near the edges are mostly labeled as the background. This is because, due to the windowing, fewer pixels are included in  $\mathcal{B}_s^i$  near the edges of the image, and thus the local background is estimated on less pixels than windows located far from the edges. As a result, near the edges, less pixels pass the shadow and the highlight tests.

Fig. 8(c) shows that four-region is better than  $K$ -means in terms of false shadow segmentation. However, clearly the best segmentation result is obtained when using the LSM. In particular, the object's borders are more accurate and the false segmentations rate is low.

3) *Quality Analysis*: Segmentation results for benchmark methods and those of our LSM are illustrated in Fig. 9(b)–(d), respectively. The assigned pixels to the labels  $\{\mathcal{S}, \mathcal{B}, \mathcal{H}\}$  are shown in different colors: black for shadow, brown for background, and yellow for highlight. While all three methods segment the image well, segmentation performance degrades significantly the more complex the background is. For the *seagrass* image in Fig. 9(a), we observe many false segmentations, produced by the Nguyen and fuzzy thresholding methods. This includes an enlargement of the object's shadow region. In contrast, E-DS-M has a cleaner background region compared to the Nguyen method. However, the borders of the highlight and shadow regions are corrupted. This phenomena does not occur with our LSM method, which produces clean segmentation, and the object's highlight and shadow are more accurate, compared to the benchmark methods. A similar effect is observed for the *sand ripples* image in Fig. 9(a). The problem of false segmentation that appears when operating the Nguyen method is more acute, and the object's shadow region blends in with the sand ripples. The performances of E-DS-M, graph cut, and our LSM are cleaner than the Nguyen, and these methods keep the object's borders in the highlight region. However, in the E-DS-M and graph cut methods, the object's shadow borders are less accurate than the ones produced by the LSM.

4) *Quantitative Analysis*:

a) *MCR and PR Analysis*: We start by showing the performance in terms of MCR (33) and 1-PR, where PR is given in (34). The MCR and PR measures are indicators of the image segmentation's accuracy. The results are shown in Fig. 10 along with the 95% confidence intervals. All methods, except for the fuzzy thresholding, have an average MCR of less than 1% and an average PR of above 98% for the *sand* image, which is relatively homogeneous. However, for the complex *sand ripples* and *seagrass* images, which have strong intensity inhomogeneity, the performances of the compared methods severely degrade, whereas those of our LSM are not affected. LSM's small confidence interval shows its robustness to different background realizations. We therefore conclude that LSM is specifically targeted to a complex seabed environment.

b) *Analysis of Shape Measures*: Since for object classification, the shape of the detected object is of interest, in our analysis, we also consider the shape, size, and position of the segmented highlight and shadow regions. Table II shows the shape measures,  $\zeta_n$  (38),  $\zeta_s$  (39), and  $\zeta_p$  (35) for the three compared schemes. The values in the table are the mean values and

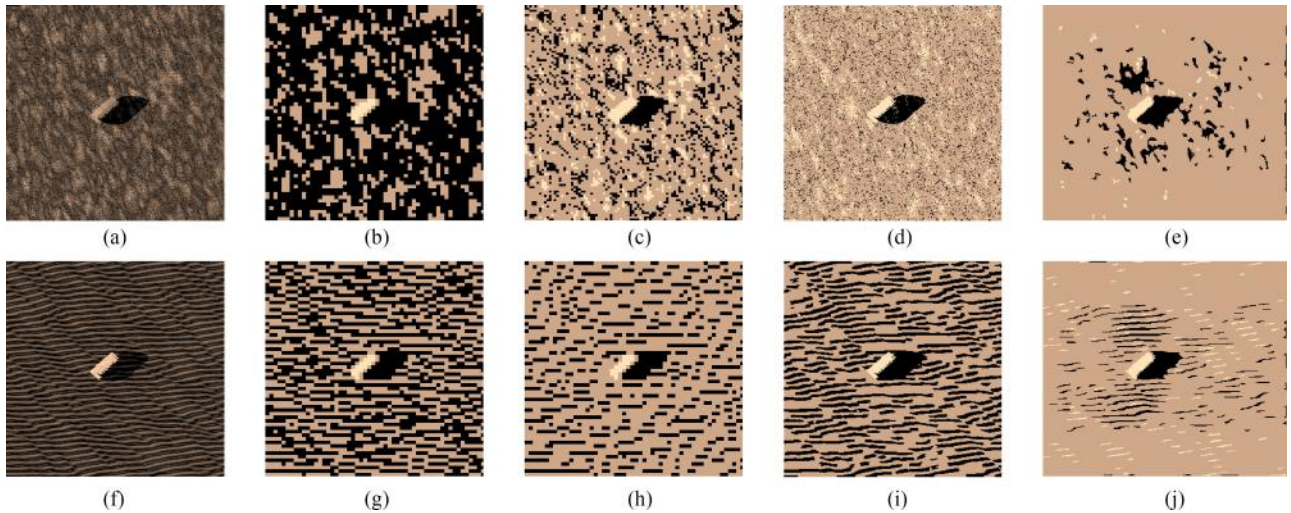


Fig. 8. Comparison of the initialization methods for the segmentation of the *seagrass* and *sand ripple* images: (a) Original *seagrass* image. (b)–(e) Provide results with *seagrass* image of *K*-means, four-region, ICM, and our LSM, respectively. (e) Original *sand ripples* image. (g)–(j) Provide results with *sand ripples* image of *K*-means, four-region, ICM, and LSM, respectively. Results show that our LSM produces the cleanest and most accurate results.

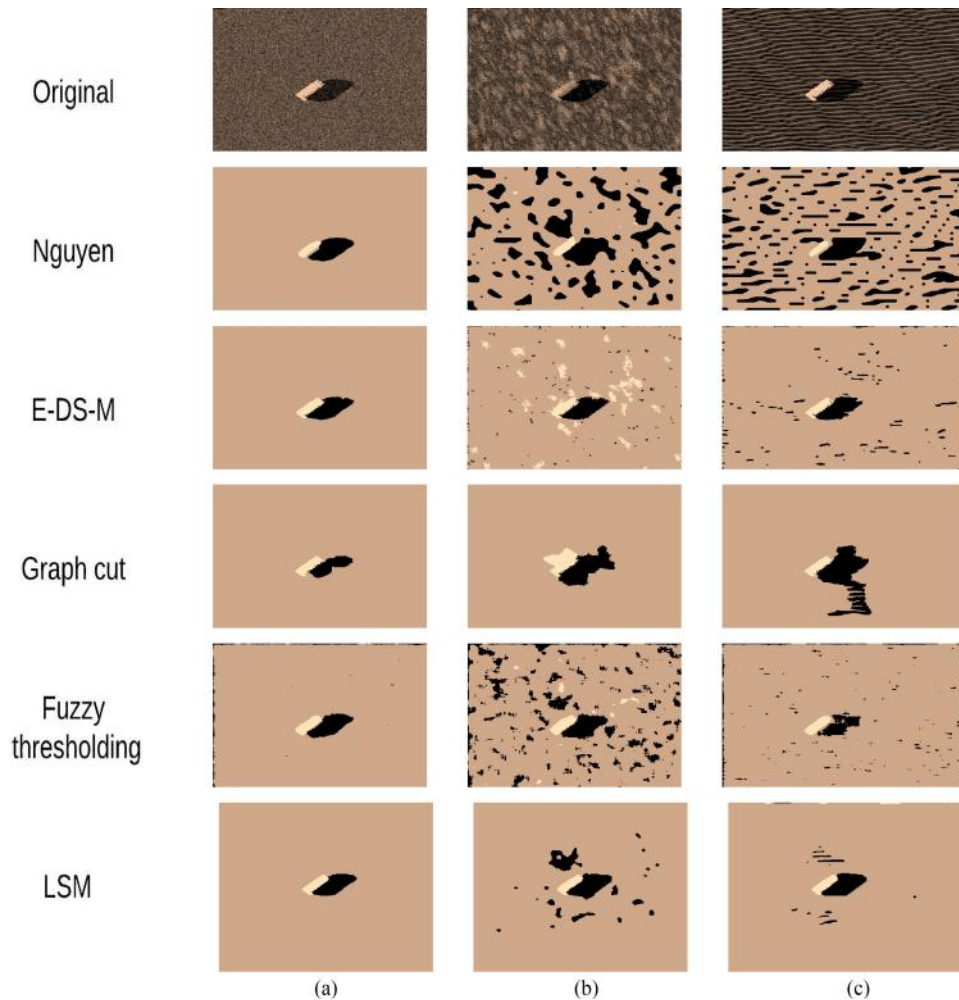


Fig. 9. Final segmentation results for the synthetic sonar images. Each column corresponds to a different sonar image. Synthetic sonar images of cylindrical objects and final segmentation results of Nguyen, E-DS-M, graph cut, fuzzy thresholding, and LSM (from top to bottom, respectively). (a)–(c) For sand, seagrass, and sand ripples background, respectively. The colors in the segmentation maps represent: black for the shadow region, yellow for the highlight region, and brown for the background region. The results show the most accurate segmentation result for our LSM.

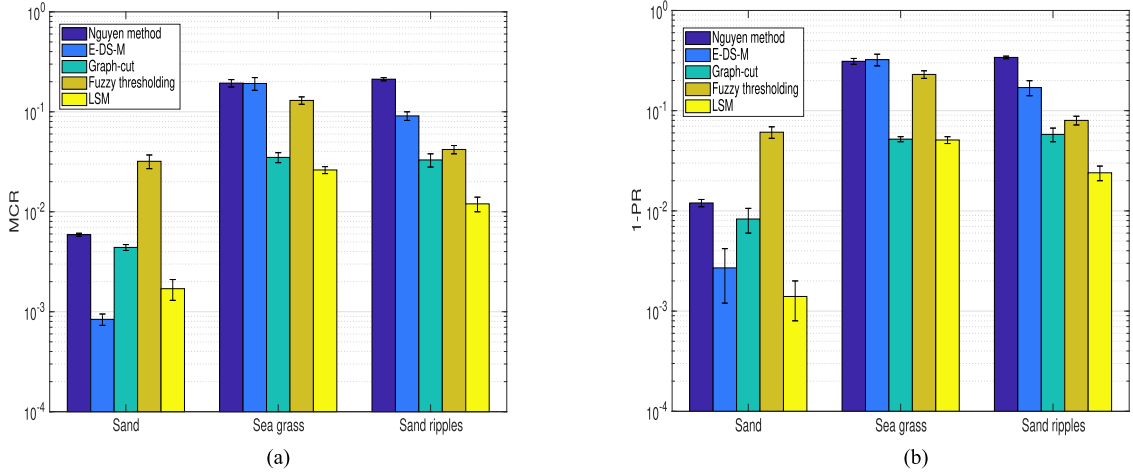


Fig. 10. PR (34) and MCR (33) values for the synthetic sonar images with speckle noise ( $STD = 0.3$ ). The 95% confidence interval is introduced on each bar in the figure. (a) Results for the MCR. (b) Results for the (1-PR). The results show that the LSM consistently generates better MCR and PR than the comparison algorithms.

TABLE II  
COMPARISON OF THE GENERAL SHAPE PROPERTIES  $\zeta_s$  (39),  $\zeta_n$  (38), AND  $\zeta_p$  (35) FOR SYNTHETIC SONAR IMAGES WITH SPECKLE NOISE ( $STD = 0.3$ )

Image Name	Sand ripples			Sea grass			Sand		
	$\zeta_n$	$\zeta_s$	$\zeta_p$	$\zeta_n$	$\zeta_s$	$\zeta_p$	$\zeta_n$	$\zeta_s$	$\zeta_p$
LSM	<b>0.994</b> [0.99,0.997]	<b>0.9</b> [0.9,0.91]	<b>0.998</b> [0.997,0.998]	<b>0.989</b> [0.98,0.99]	<b>0.888</b> [0.87,0.9]	<b>0.996</b> [0.996,0.997]	0.98 [0.97,0.98]	0.9 [0.89,0.91]	<b>0.998</b> [0.997,0.998]
Nguyen method	0.69 [0.65,0.73]	0.52 [0.51,0.54]	0.989 [0.986,0.993]	0.79 [0.57,0.82]	0.52 [0.45,0.55]	0.95 [0.78,0.99]	0.75 [0.74,0.77]	0.609 [0.606,0.617]	0.994 [0.993,0.995]
E-DS-M	0.98 [0.97,0.99]	0.9 [0.88,0.91]	0.996 [0.995,0.997]	0.84 [0.52,0.94]	0.65 [0.52,0.91]	0.96 [0.88,0.99]	<b>0.983</b> [0.981,0.987]	<b>0.95</b> [0.945,0.954]	0.996 [0.995,0.997]
Graph-cut	0.76 [0.754,0.765]	0.65 [0.62,0.68]	0.97 [0.961,0.974]	0.68 [0.74,0.77]	0.5 [0.49,0.57]	0.98 [0.971,0.985]	0.922 [0.918,0.928]	0.84 [0.835,0.847]	0.997 [0.995,0.997]
Fuzzy thresholding	0.92 [0.917,0.923]	0.77 [0.763,0.78]	0.995 [0.996,0.998]	0.96 [0.95,0.97]	0.8 [0.76,0.82]	0.994 [0.993,0.995]	0.98 [0.98,0.987]	0.85 [0.845,0.854]	0.996 [0.995,0.997]

Results show that for complex seabed texture, such as sand ripple and seagrass, LSM achieves the best results.

the 95% confidence intervals (in brackets). From Table II, we observe that the average  $\zeta_p$  for all methods is almost not affected by the background type. We also observe some inconsistencies between the performance of the five methods for different backgrounds, in terms of the other measures. For example, E-DS-M performs poorly for the *seagrass* image, whereas those of the Nguyen method reduces for the *sand ripples* image. However, the proposed method shows robust performance with an average of  $\zeta_n$ ,  $\zeta_s$ , and  $\zeta_p$  above 0.98, 0.88, and 0.99, respectively, for all images. This suggests that the proposed algorithm also outperforms the compared algorithms in terms of the shape of the segmented object. We report that similar trends are obtained for higher speckle noise.

c) *Analysis of Geometrical Features*: The object's geometrical properties are usually of vast importance in the object classification process. In this section, we analyze our segmentation method performances and the four benchmark methods, in terms of the geometrical features of the segmented highlight and shadow regions. Table III shows the segmentation

performance in terms of our geometrical quality measures,  $\kappa_{\text{perim}}$  (40),  $\kappa_{\text{area}}$  (42), and  $\kappa_{\text{sol}}$  (41). From the results, we conclude that the perimeter of the object in the compared methods is mostly corrupted for *seagrass* and *sand ripples* backgrounds. Also in this case, LSM achieves the best performance. Furthermore, the LSM's relatively small confidence intervals and the consistency of the results for the different background types show that the LSM is robust for the seabed environment.

### C. Results on Real Sonar Images

To complete our investigation, we now discuss results obtained for a set of real sonar images shown in Fig. 12(a). The three images were produced by an EdgeTech 4125 sidescan sonar and we refer to them as *body*, *fish*, and *bicycle*. The size of the images is  $230 \times 217$ ,  $250 \times 385$ , and  $350 \times 500$  pixels, respectively.

1) *Initialization Results*: We next explore the tradeoff between the miss segmentation and false segmentation of the

TABLE III  
COMPARISON OF THE GEOMETRICAL FEATURES' ERRORS  $\kappa_{\text{perim}}$  (40),  $\kappa_{\text{area}}$  (42), AND  $\kappa_{\text{sol}}$  (41) FOR  
SYNTHETIC SONAR IMAGES WITH SPECKLE NOISE (STD = 0.3)

Image Name	Sand ripples			Sea grass			Sand		
	$\kappa_{\text{perim}}$	$\kappa_{\text{area}}$	$\kappa_{\text{sol}}$	$\kappa_{\text{perim}}$	$\kappa_{\text{area}}$	$\kappa_{\text{sol}}$	$\kappa_{\text{perim}}$	$\kappa_{\text{area}}$	$\kappa_{\text{sol}}$
LSM	<b>13.98</b> [11.99,18]	<b>0.047</b> [0.007,0.089]	<b>0.04</b> [0.03,0.04]	<b>21.49</b> [16.99,24.5]	<b>0.094</b> [0.01,0.21]	<b>0.03</b> [0.015,0.05]	<b>12.579</b> [10.99,14.5]	<b>0.1</b> [0.03,0.16]	0.042 [0.0034,0.0048]
Nguyen method	74.52 [46,107.5]	7.23 [5.7,8.95]	0.12 [0.08,0.17]	69.3 [48.5,94]	2.57 [1.64,3.9]	0.11 [0.06,0.16]	43.1 [41,44.5]	5.12 [4.63,5.57]	0.14 [0.12,0.15]
E-DS-M	48.51 [35,65]	0.057 [0.006,0.13]	0.077 [0.06,0.09]	80.62 [45,125.5]	0.42 [0.09,0.65]	0.16 [0.06,0.23]	19.52 [15,24.5]	0.16 [0.13,0.19]	0.023 [0.015,0.032]
Graph-cut	276 [240,294]	4.45 [3.5,5.2]	0.2 [0.12,0.3]	80.5 [54,94.5]	0.7 [0.35,0.84]	0.11 [0.03,0.15]	34 [24,37]	0.73 [0.5,0.84]	<b>0.04</b> [0.001,0.005]
Fuzzy thresholding	78 [65,80.5]	0.7 [0.03,0.16]	0.14 [0.11,0.17]	35 [33,37]	0.1 [0.05,0.2]	0.09 [0.02,0.15]	36 [32,40.5]	0.22 [0.13,0.29]	0.05 [0.015,0.062]

Results show that LSM achieves the most accurate geometrical features of the segmented object.

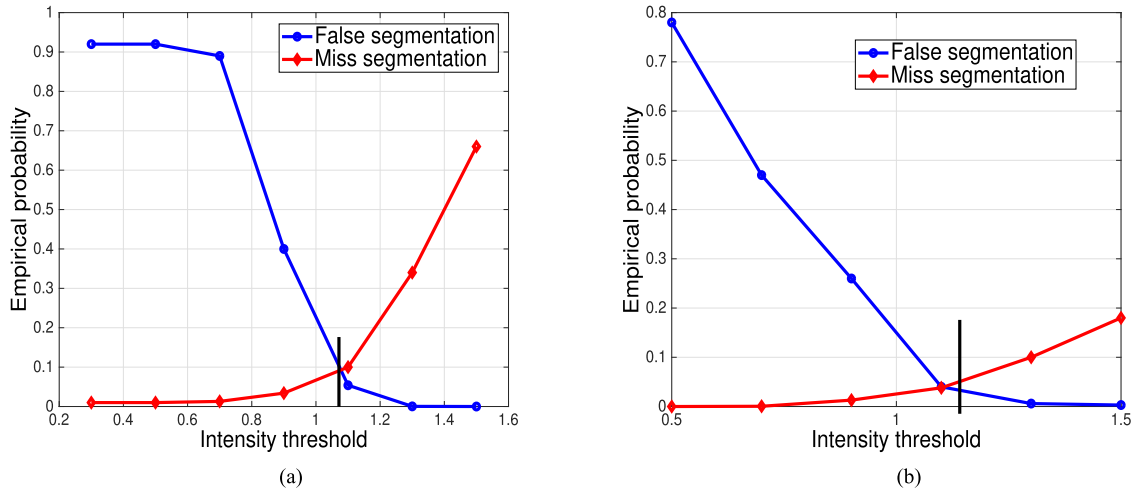


Fig. 11. False segmentation and miss segmentation values [from (44) to (47)] for the *body* image versus the intensity threshold  $\mathcal{F}$ . The results of automatic selection of thresholds are  $\mathcal{F}_h = 1.07$  and  $\mathcal{F}_s = 1.15$  and are marked with black lines. (a) Results for the highlight region. (b) Results for the shadow region. The results show that our automatic procedure for choosing the threshold achieves a favorable tradeoff.

highlights and shadows over the real sonar images at the output of the initialization process. Let  $\text{FS}_h$  and  $\text{FS}_s$  be the false segmentation indices in the highlight and shadow regions, respectively, such that

$$\text{FS}_h = \frac{1}{|\widehat{\mathcal{C}}_{\text{H}}|} \sum_{i \in \widehat{\mathcal{C}}_{\text{H}}} \mathcal{I}(\bar{l}_i \neq G_i) \quad (44)$$

$$\text{FS}_s = \frac{1}{|\widehat{\mathcal{C}}_{\text{S}}|} \sum_{i \in \widehat{\mathcal{C}}_{\text{S}}} \mathcal{I}(\bar{l}_i \neq G_i) \quad (45)$$

where  $\widehat{\mathcal{C}}_{\text{H}}$  and  $\widehat{\mathcal{C}}_{\text{S}}$  are the clusters of highlight and shadow in the segmentation results, respectively,  $\bar{l}_i$  is the label of pixel  $i$  in the segmentation results, and  $G_i$  is the label of pixel  $i$  in the ground-truth map  $G$ . Similarly, miss segmentations  $\text{MS}_h$  and  $\text{MS}_s$  are defined as the average number of pixels that belong to the highlight and shadow clusters in  $G$ , respectively, and were

not correctly assigned in the segmentation results. Formally

$$\text{MS}_h = \frac{1}{|\mathcal{C}_{g,h}|} \sum_{i \in \mathcal{C}_{g,h}} \mathcal{I}(\bar{l}_i \neq G_i) \quad (46)$$

$$\text{MS}_s = \frac{1}{|\mathcal{C}_{g,s}|} \sum_{i \in \mathcal{C}_{g,s}} \mathcal{I}(\bar{l}_i \neq G_i) \quad (47)$$

where  $\mathcal{C}_{g,h}$  and  $\mathcal{C}_{g,s}$  are the highlight and shadow clusters in the ground-truth map  $G$ , respectively.

Fig. 11(a) and (b) shows the false segmentation and miss segmentation values of the initialization results for the highlight and shadow regions of the *body* image. The curves in Fig. 11 are generated by computing (44)–(47) from the results of the proposed initialization method for each intensity threshold  $\mathcal{F}$  on the curves. The automatic process for setting the threshold levels determine the intensity thresholds  $\mathcal{F}_h$  and  $\mathcal{F}_s$  to be 1.07 and 1.15, respectively, and is represented by the vertical lines in Fig. 11(a) and (b). We observe that our process for automatic determination

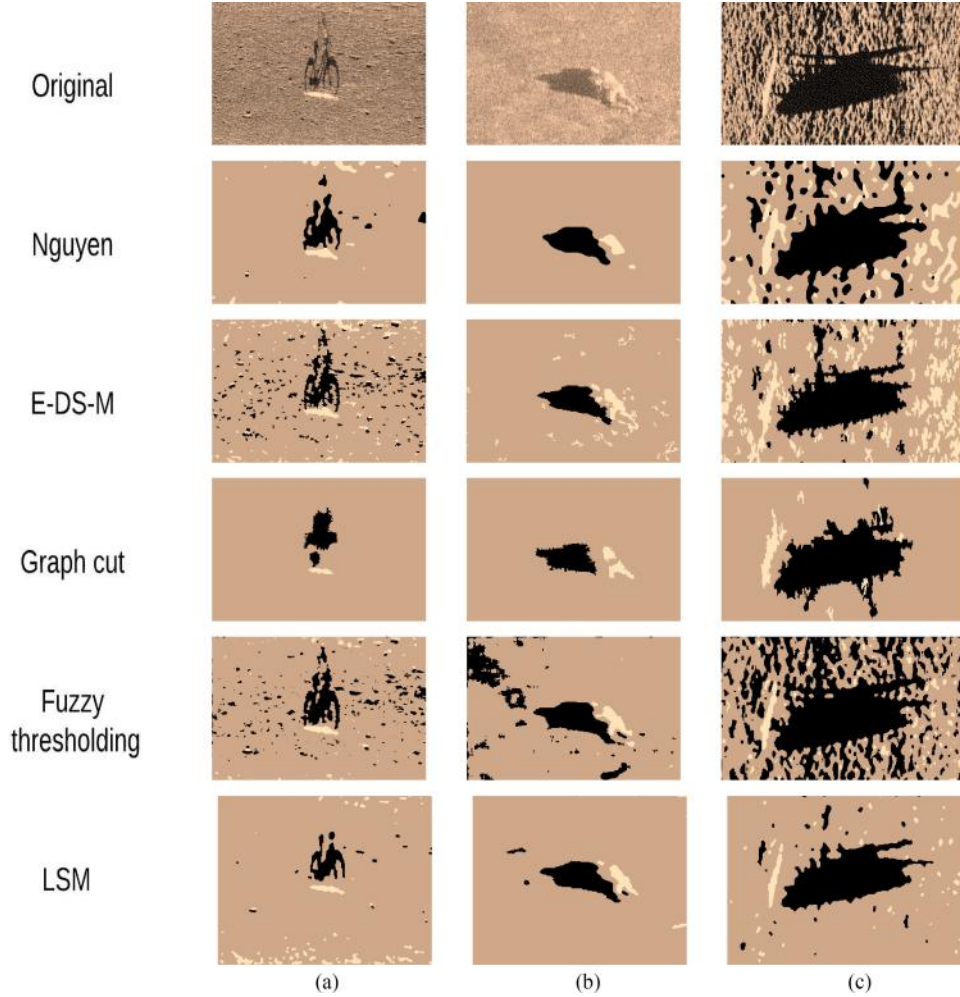


Fig. 12. Final segmentation results for real sonar images (original image courtesy of EdgeTech [5]). The color in the segmentation maps represent: black for the shadow region, yellow for the highlight region, and brown for the background region. Real sonar images and final segmentation results of the Nguyen, E-DS-M, graph cut, fuzzy thresholding, and LSM (from top to bottom, respectively). (a)–(c) For bicycle, body, and fish, respectively. The best segmentation results are obtained by the LSM.

of the thresholds determines the intensity thresholds for both the highlight and shadow regions around the crossing point of the curves, i.e., our method produces a favorable tradeoff between the false segmentation and the miss segmentation results.

2) *Segmentation Results*: The segmentation results are presented in Fig. 12. For the *bicycle* image, we observe obvious false segmentation areas when using the Nguyen, E-DS-M and fuzzy thresholding benchmarks, i.e., a large region is incorrectly labeled as highlight. For the *body* image, the shadow of the object was accurately segmented by all methods, except for the graph cut. However, using the Nguyen method, the highlight of the object is corrupted. Although it still exists, the highlight corruption is less apparent in using the E-DS-M; yet, a much cleaner and accurate segmentation is achieved by the LSM. Conditions of high intensity inhomogeneity can be clearly seen in the *fish* image. This inhomogeneity causes a serious degradation in the segmentation performance of all methods. However, compared to the benchmark methods, the LSM and graph cut methods produce cleaner results with no obvious false segmentation. The results in Fig. 12 show that the segmentation performance of

graph cut is *cleaner*, i.e., fewer background-based false negative exists. However, for ADAC’s main task of object classification, the objective of the segmentation part is not just to provide a clean segmented image, but also to keep the object’s formation and avoid distortions to its highlight and shadow regions. This is important since, later, the segmented highlight and shadow areas are used as extracted features for the classification. In that context, the results in Fig. 12 show that graph cut significantly distorts the highlight and shadow regions of the objects, whereas LSM provides an accurate object-related segmentation.

#### D. Sea Trial Results

To comment on the performance of LSM in a realistic sea environment of complex seabed and over different sonar systems, we performed three sea experiments. In these experiments, we used a 400-kHz multibeam sonar EM 2040 mounted on the RV Bat-Galim research ship and two Kraken-made SAS systems mounted on the sides of our own A18 5.5 m Eca Robotics Inc., AUV [51]. A picture of the SAS system onboard our AUV is

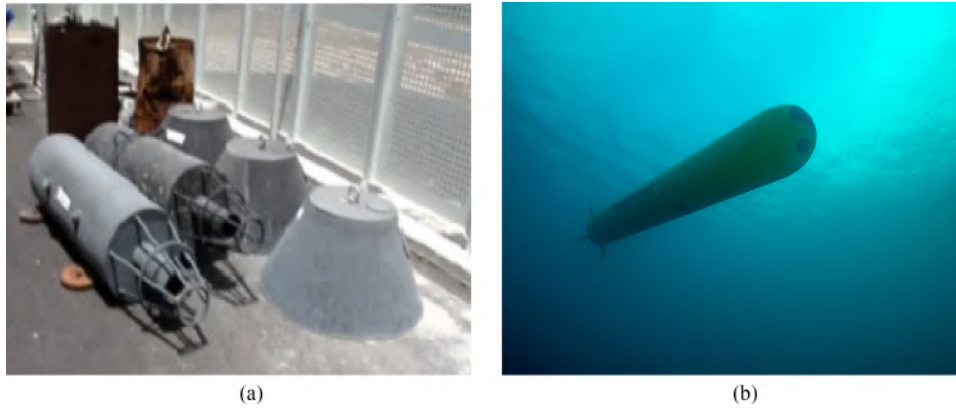


Fig. 13. Photographs of the targets considered in the sea trial study and the Eca Robotics AUV A18. (a) Two targets were used: one of a cone shape (right-hand side of the figure) and other of the tube shape (left-hand side of the figure). (b) A18 AUV and its SAS system. Picture taken during our sea experiment

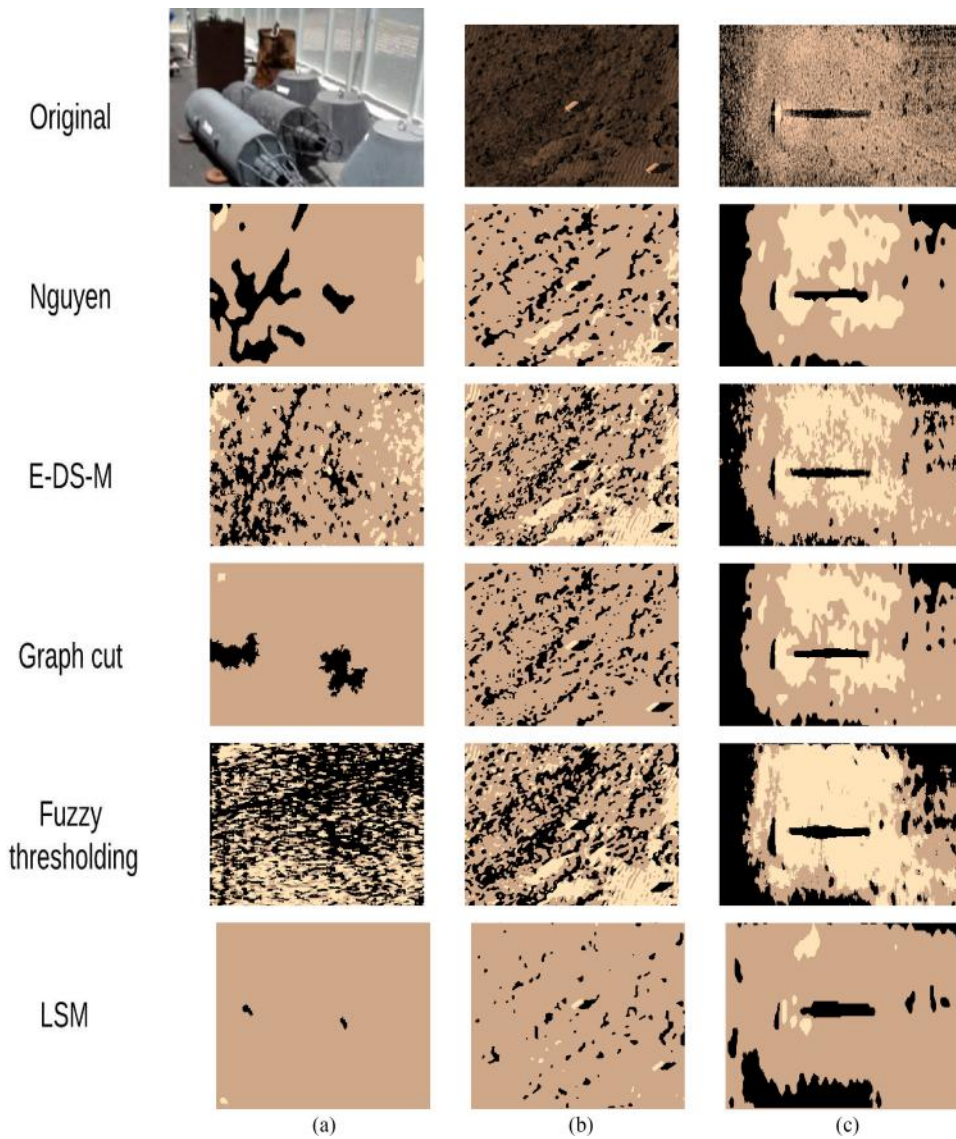


Fig. 14. Final segmentation results from our sea trial. (a) Multibeam (EM 2040) sonar image. (b) Multibeam sonar image with two synthetic cylindrical objects. (c) Two-sided synthetic aperture sonar (Kraken). Real sonar images and final segmentation results of the Nguyen, E-DS-M, graph cut, fuzzy thresholding, and LSM (from top to bottom, respectively). However, LSM and graph cut achieves the most clean results, LSM has the most accurate results.



shown in Fig. 13(b). The first trial included the multibeam system scanning an area of depth 25 m and detecting two targets of truncated metallic cones whose picture is shown in Fig. 13(a). The second experiment involved the multibeam system scanning an area of extreme complex seabed that includes rocks, sand, and reef at depth of 20 m. The third experiment included the AUV scanning with its SAS as deserted gas well as depth of 1000 m. The three produced images are of complex background and include diverse targets. To allow reproducibility, we share these original three sonar images in [52].

The resulting original multibeam sonar image from the first experiment is shown in Fig. 14(a). The image includes two targets (shown in the middle of the image), a diverse background of sand (right side), rocks covered with sediments (middle), exposed rock (left upper corner), and a crack (left-hand side). Segmentation results for LSM and the other four benchmark methods are shown below the original image. From the segmentation results, we observe that the shadow of the object is clearer than the highlight. Hence, only pixels that relate to the shadow of the objects are segmented well. Comparing the segmentation performance, we conclude that also for the sonar image generated in our experiment, LSM outperforms the four benchmark methods in terms of segmentation accuracy of shadow and background regions, as well as in preserving the shape of the segmented object.

The sonar image obtained in the second trial is shown in Fig. 14(b). The image includes a highly complex background of rocks, boulders, and sand ripples. To test segmentation performance, we included two synthetic targets shown in the middle and right lower side of Fig. 14(b). One target is placed on a rocky area, whereas the other is placed on sand ripples. We observe that all methods segmented successfully the object on the rocky area, whereas the segmentation results for the object on the sand ripple were satisfactory only for graph cut and LSM. The segmentation results in Fig. 14(b) revealing that LSM obtained the most clean background.

The sonar image obtained in the third trial is shown in Fig. 14(c). In column (c) of Fig. 14, we introduce the segmentation results of LSM and the benchmark for the SAS image. In the middle of the figure, we see the sonar reflection from the gas well. This structure includes a 4-m-high metal pole whose highlight intensity is extremely high and whose shadow is very long. The result is an image with a complex distribution for the highlight and shadow regions. However, we observe that all methods segment successfully the shadow region of the object, strong intensity inhomogeneity in the original image causes a severe degradation in the segmentation results of the highlight and background regions. In the segmentation results of fuzzy thresholding, the highlight region of the object is merged with the background. Also here, LSM achieves the most clean results while keeping the object's boundaries.

### E. Comparison of Running Time

Since the I-step is performed as part of the EM's main loop, it does not influence the complexity of the whole process and

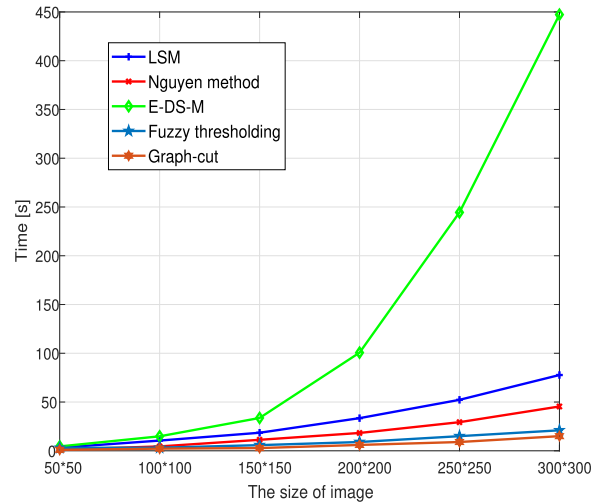


Fig. 15. Running time for the five compared segmentation methods for different sized sonar images. The maximum number of iterations is 50 and the convergence threshold is set to  $\epsilon = 0.001$ . The results show that the graph cut method is faster than the other algorithms.

the complexity of the LSM is similar to the basic unconstrained EM algorithm, namely,  $\mathcal{O}(M_{\max}NL)$ . This is similar to the complexity of the Nguyen method, whereas the complexity of the E-DS-M is  $\mathcal{O}(M_{\max}NL^2)$ .

To demonstrate the running time of the LSM and compare it with that of the benchmark methods, in Fig. 15, we compare the average running time of the five methods, as a function of the sonar image size. All experiments were performed using a MATLAB R2011a installed onboard an Intel(R) Core(TM) i7-3770 CPU @3.4 GHz, 8GB RAM processor. We observe that for an image of  $150 \times 150$  pixels and above, the E-DS-M is time consuming and cannot be used for real-time applications. We also observe that the Nguyen method is slightly faster than our LSM. The proposed algorithm is time consuming compared to the graph cut and fuzzy thresholding methods. However, the performance and accurate results of the LSM, as shown above, compensates for this drawback.

## VI. CONCLUSION

In this paper, we introduced the LSM: a new, unsupervised segmentation algorithm used for object detection and classification in sonar images. LSM aims to overcome the problem of intensity inhomogeneity, which exists in complex seabed textures, such as sand ripples or seagrass, to accurately label the pixels related to the object's highlight and shadow. Our algorithm is built around a simple, yet effective model, which incorporates the expected correlation among neighboring pixels using an intermediate step between the E-step and the M-step of the EM algorithm. To achieve robustness to different seabed structures, we also proposed an automatic process for setting the segmentation thresholds used during the algorithm's initialization process. We measured LSM performance using novel quality measures that examine not only the label's assignment result, but also the segmentation quality in terms of the segmented shape, structure,

and area. Our numerical and experimental results showed that, compared to benchmark segmentation methods, the LSM produces accurate segmentation for complex seabed structures, even in the presence of significant speckle noise. Since the main challenge in sonar imagery is still the noisy background, future research on joint image denoising and segmentation is planned.

## REFERENCES

- [1] D. P. Williams, "Fast target detection in synthetic aperture sonar imagery: A new algorithm and large-scale performance analysis," *IEEE J. Ocean. Eng.*, vol. 40, no. 1, pp. 71–92, Jan. 2015.
- [2] G. Cheng, J. Han, L. Gou, Z. Liu, S. Bu, and J. Ren, "Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4238–4249, 2015.
- [3] M. Mignotte, C. Collet, P. Perez, and P. Boutheymy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.
- [4] G. S. Sanjay and T. J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1014–1028, Jul. 1998.
- [5] 2017. [Online]. Available: <http://www.edgetech.com>
- [6] D. Pham, "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recognit. Lett.*, vol. 20, no. 1, pp. 57–68, Jan. 1999.
- [7] X. Ye, L. Wang, and T. Wang, "Sonar image segmentation on fuzzy C-mean using local texture feature," in *Proc. Int. Conf. Med. Eng.*, 2011, pp. 115–120.
- [8] T. M. Nguyen and Q. M. J. Wu, "Fast and robust spatially constrained Gaussian mixture model for image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 621–635, Apr. 2013.
- [9] R. R. Ganta, S. Zaheeruddin, N. Baddiri, and R. R. Rao, "Segmentation of oil spill images with illumination-reflectance based adaptive level set model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1394–1402, Oct. 2012.
- [10] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [11] R. Fandos, L. Sadamori, and A. M. Zoubir, "High quality segmentation of synthetic aperture sonar images using the min-cut/max-flow algorithm," in *Proc. 19th Eur. Signal Process. Conf.*, Sep. 2011, pp. 51–55.
- [12] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *J. Amer. Stat. Assoc.*, vol. 86, no. 414, pp. 376–387, 1991.
- [13] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*. Hoboken, NJ, USA: Wiley, 2011, ch. 22.
- [14] R. R. Ganta, S. Zaheeruddin, N. Baddiri, and R. R. Rao, "Segmentation of oil spill images with illumination-reflectance based adaptive level set model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1394–1402, Oct. 2012.
- [15] G. Huo, S. X. Yang, Q. Li, and Y. Zhou, "A robust and fast method for sidescan sonar image segmentation using nonlocal despeckling and active contour model," *IEEE Trans. Cybern.*, vol. 7, no. 99, pp. 1–18, Mar. 2016.
- [16] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, Feb. 1997.
- [17] M. Lianantonakis and Y. R. Petillot, "Sidescan sonar segmentation using texture descriptors and active contours," *IEEE J. Ocean. Eng.*, vol. 32, no. 3, pp. 744–752, Jan. 2007.
- [18] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.
- [19] Y. Wang, L. Liu, H. Zhang, Z. Cao, and S. Lu, "Image segmentation using active contours with normally biased GVF external force," *IEEE Signal Process. Lett.*, vol. 17, no. 10, pp. 875–878, Oct. 2010.
- [20] R. Fandos and A. M. Zoubir, "Optimal feature set for automatic detection and classification of underwater objects in SAS images," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 454–468, Jun. 2011.
- [21] C. Li, C. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.
- [22] U. Javed, M. M. Riaz, A. Ghafoor, and T. A. Cheema, "SAR image segmentation based on active contours with fuzzy logic," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 181–188, Feb. 2016.
- [23] Y. Jing, J. An, and Z. Liu, "A novel edge detection algorithm based on global minimization active contour model for oil slick infrared aerial image," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2005–2013, Jun. 2011.
- [24] H. Song, B. Huang, and K. Zhang, "A globally statistical active contour model for segmentation of oil slick in SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2402–2409, Dec. 2013.
- [25] T. Celik and T. Tjahjadi, "A novel method for sidescan sonar image segmentation," *IEEE J. Ocean. Eng.*, vol. 36, no. 2, pp. 186–194, Apr. 2011.
- [26] L. Hongpoa, S. Junb, W. Haic, T. Shuhuac, and T. Zhiguoc, "High resolution sonar image segmentation by PSO based fuzzy cluster method," in *Proc. 4th Int. Conf. Genetic Evol. Comput.*, Apr. 2010, pp. 18–21.
- [27] X. Wang, X. Ye, C. Fang, Z. Zhang, and L. Zhao, "A novel automatic segmentation algorithm for sonar imagery," in *Proc. IEEE Int. Conf. Mechatronics Autom.*, 2008, pp. 336–241.
- [28] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 494–498, Mar. 2005.
- [29] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.
- [30] J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Process.*, vol. 3, no. 4, pp. 404–420, Jul. 1994.
- [31] G. Boccignone, V. Caggiano, P. Napolitano, and M. Ferraro, "Image segmentation via multiresolution diffused expectation-maximization," in *Proc. IEEE Int. Conf. Image Process.*, 2005, vol. 1, pp. 289–292.
- [32] T. Fei and D. Kraus, "Dempster-Shafer theory supported EM approach for sonar image segmentation," *Trans. Syst., Signals Devices*, vol. 9, no. 3, pp. 1–43, 2014.
- [33] J. Wu and C. S. Chung, "A segmentation model using compound Markov random fields based on a boundary model," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 241–252, Jan. 2007.
- [34] X. Pan, H. Zhu, and Q. Xie, "Image segmentation by student's-t mixture models based on Markov random field and weighted mean template," *Int. J. Signal Process., Image Proc. Pattern Recognit.*, vol. 9, no. 2, pp. 313–322, 2016.
- [35] G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich, "Spatially varying mixtures incorporating line processes for image segmentation," *J. Math. Imag. Vis.*, vol. 36, no. 2, pp. 91–110, Feb. 2010.
- [36] C. Nikou, A. Likas, and N. P. Galatsanos, "A Bayesian framework for image segmentation with spatially varying mixtures," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2278–2289, Sep. 2010.
- [37] T. Fei and D. Kraus, "An expectation-maximization approach assisted by Dempster-Shafer theory and its application to sonar image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 1161–1164.
- [38] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
- [39] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ, USA: Wiley, 2003, ch. 8.
- [40] J. Besag, "On the statistical analysis of dirty images," *J. Roy. Stat. Soc.*, vol. B-48, pp. 259–302, 1986.
- [41] A. R. Webb, "Gamma mixture models for target recognition," *Pattern Recognit.*, vol. 33, no. 12, pp. 2045–2054, 2000.
- [42] S. A. Fernandez, A. H. Curiale, and G. V. S. Ferrero, "A local fuzzy thresholding methodology for multiregion image segmentation," *Knowl.-Based Syst.*, vol. 83, pp. 1–12, 2015.
- [43] R. Fandos, "ADAC system design and its application to mine hunting using SAS imagery," Ph.D. dissertation, Dept. Elect. Eng., Univ. Darmstadt, Darmstadt, Germany, 2012.
- [44] E. W. Forgy, "Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications," *Biometrics*, vol. 21, pp. 768–769, 1965.
- [45] R. Fandos and A. M. Zoubir, "Enhanced initialization scheme for a three region Markovian segmentation algorithm and its application to SAS images," in *Proc. 10th Eur. Conf. Underwater Acoust.*, 2010, vol. 3, pp. 1323–1331.
- [46] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [47] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 34–41.

- [48] E. V. D. Lucca, C. C. Freitas, A. C. Frery, and S. J. S. Santanna, "Comparison of SAR segmentation algorithms," in *Proc. IXth Braz. Symp. Remote Sens.*, 1998, pp. 1097–1108.
- [49] J. T. Cobb, "Sonar image modeling for texture discrimination and classification," Ph.D. dissertation, Dept. Elect. Eng., Univ. Florida, Gainesville, FL, USA, 2011.
- [50] T.-D. Miao, Q.-S. Mu, and S.-Z. Wu, "Computer simulation of aeolian sand ripples and dunes," *Phys. Lett. A*, vol. 288, no. 11, pp. 16–22, 2001.
- [51] 2017. [Online]. Available: <http://www.ecagroup.com>
- [52] 2017. [Online]. Available: [https://drive.google.com/file/d/1OaKVyBanzRqicVK05qmYlqk\\_VIMJmDKV/view?usp=sharing](https://drive.google.com/file/d/1OaKVyBanzRqicVK05qmYlqk_VIMJmDKV/view?usp=sharing)
- [53] M. Ahmed, S. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithms for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193–199, Mar. 2002.
- [54] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.
- [55] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy C-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 573–584, Feb. 2013.
- [56] X. Tian, L. Jiao, and X. Zhang, "A clustering algorithm with optimized multiscale spatial texture information: Application to SAR image segmentation," *Int. J. Remote Sens.*, vol. 34, no. 4, pp. 1111–1126, Feb. 2013.
- [57] F. Zhao, L. C. Jiao, H. Q. Liu, and X. B. Gao, "A novel fuzzy clustering algorithm with non local adaptive spatial constraint for image segmentation," *Signal Process.*, vol. 91, no. 4, pp. 988–999, Apr. 2011.
- [58] J. Ji and K. Wang, "A robust nonlocal fuzzy clustering algorithm with between-cluster separation measure for SAR image segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4929–4936, Oct. 2013.
- [59] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [60] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, Jun. 2016.
- [61] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
- [62] L. Chen, C. L. P. Chen, and M. Lu, "A multiple-kernel fuzzy C-Means algorithm for image segmentation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 5, pp. 1263–1274, Apr. 2011.
- [63] H. C. Huang, Y. Y. Chuang, and C. S. Chen, "Multiple kernel fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 120–134, Feb. 2012.
- [64] D. Xiang, T. Tang, C. Hu, and Y. Su, "A kernel clustering algorithm with fuzzy factor: Application to SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 7, pp. 1290–1294, Jul. 2014.
- [65] R. Shang *et al.*, "A spatial fuzzy clustering algorithm with kernel metric based on immune clone for SAR image segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1640–1652, Apr. 2016.
- [66] S. Xie, F. Khomh, Y. Zou, and I. Keivanloo, "An empirical study on the fault-proneness of clone migration in clone genealogies," in *Proc. IEEE Conf. Softw. Maintenance Reeng. Reverse Eng.*, Feb. 2014, pp. 94–103.



**Avi Abu** received the B.Sc. and M.Sc. degrees in electrical engineering from Tel Aviv University, Tel Aviv, Israel, in 1998 and 2003, respectively. He is currently working toward the Ph.D. degree at the Underwater Acoustic and Navigation Laboratory, Haifa University, Haifa, Israel, in 1998.

His research interests include sonar image segmentation, underwater object detection, and sonar image denoising.



**Roe Diamant** received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion, Israel Institute of Technology, Haifa, Israel, in 2002 and 2007, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada, in 2013.

From 2001 to 2009, he was a Project Manager and Systems Engineer with Rafael Advanced Defense Systems, Haifa, Israel, where he developed a commercial underwater modem with network capabilities. In 2015 and 2016, he was a visiting Professor at the University of Padova, Padova, Italy. He currently leads the Underwater Acoustic and Navigation Laboratory as a Senior Lecturer (equivalent to an Assistant Professor) at the Department of Marine Technology, University of Haifa, Haifa, Israel. His research interests include underwater acoustic communication, underwater navigation, object identification, and classification.

Dr. Diamant is currently an Associate Editor for the IEEE OCEAN ENGINEERING. He was the recipient of the Israel Excellent Worker First Place Award from the Israeli Presidential Institute in 2009, the NSERC Vanier Canada Graduate Scholarship in 2010, and two best paper awards.