HIGH QUALITY SEGMENTATION OF SYNTHETIC APERTURE SONAR IMAGES USING THE MIN-CUT/MAX-FLOW ALGORITHM

Raquel Fandos, Leyna Sadamori and Abdelhak M. Zoubir

Signal Processing Group, Institute of Telecommunications
Technische Universität Darmstadt, Merckstr. 25, 64283 Darmstadt, Germany
phone: +49 6151-16 70804, fax: +49 6151-16 3778, {fandos, sadamori, zoubir}@spg.tu-darmstadt.de

ABSTRACT

In the context of automatic detection and classification for mine hunting applications, a high quality segmentation of sonar images is mandatory. Assuming a Markov Random Fields representation of the images, we propose a min-cut/max-flow segmentation algorithm. We introduce an original initialization of the graph cut algorithm based on the segmentation result of an Iterative Conditional Modes (ICM) segmentation approach. A large database of synthetic aperture sonar images containing 378 spherical and cylindrical man made objects has been segmented using both the ICM algorithm and the graph cut approach. Both sets of results have been automatically classified according to a set of significant features. Results are compared.

1. INTRODUCTION

The high resolution achieved by Synthetic Aperture Sonar (SAS) images encourages the development of Automatic Detection and Automatic Classification (ADAC) systems for mine hunting applications [1, 2]. Image segmentation is the first step of most ADAC systems and its quality determines to a great extent the performance of the classifier. Three labels are typically considered: the highlights of the objects `hlt`, their shadows `shw` and the background `bkg`. In Fig. 1 the segmentation of a sonar image showing a cylindrical man made object is illustrated.

Markov Random Fields (MRF) have traditionally been utilized to model sonar images for segmentation purposes [3, 4]. They consist of two fields, $(\mathbf{y},\mathbf{x})$. The observation field $\mathbf{y} = (y_1,\ldots,y_p,\ldots,y_{|\mathcal{P}|})$ is defined by the intensity y_p of each pixel $p \in \mathcal{P}$. The label field $\mathbf{x} = (x_1,\ldots,x_p,\ldots,x_{|\mathcal{P}|})$ is the 'ground truth' that we want to recover. Each x_p corresponds to one of the possible labels $\{\texttt{hlt},\texttt{shw},\texttt{bkg}\}$. The probability density function (pdf) of $\mathbf{x}$, $P(\mathbf{x})$, can be expressed as $P(\mathbf{x}) = \prod_{p\in\mathcal{P}} P(x_p)$, where $P(x_p)$ depends only on the neighboring pixels of p (see Sec. 2). Segmenting $\mathbf{y}$ is equivalent to estimating $\mathbf{x}$. According to the Bayes theorem, the optimal $\mathbf{x}$ maximizes the *a posteriori* probability, i. e.,

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}}\{P(\mathbf{x}|\mathbf{y})\} = \arg\max_{\mathbf{x}}\{P(\mathbf{x})\cdot P(\mathbf{y}|\mathbf{x})\}. \qquad (1)$$

This is a computationally unaffordable task and, therefore, a suboptimal $\hat{\mathbf{x}}$ is required. The Iterative Conditional Modes (ICM) algorithm [5] has traditionally been employed and, for sonar images, produces reasonable results when combined with the Iterative Conditional Estimation algorithm for estimating the MRF parameters (see [1]). An initialization is also required (see [6]).

However, challenging sea bed scenarios such as sand ripples or backgrounds with parts almost as dark as the objects shadow can lead to poor segmentation results, as shown in Fig. 1, where the shadow of the cylindrical object is much broader than the cylinder itself and has, moreover, a very irregular shape.

In [7] it is demonstrated that a min-cut/max-flow algorithm can be used to estimate $\mathbf{x}$. Thus, we have utilized a computationally efficient implementation of a graph cut algorithm [8] to segment the sonar images. Using a graph representation for the image, a min-cut/max-flow algorithm splits it into two groups of pixels, one assigned to the shadow label `shw` and the second to the background label `bkg`. The ICM result for the `hlt` label is in general satisfactory, and therefore is not regarded by the graph cut approach. To the best knowledge of the authors, it is the first time that graph cut theory is applied for segmentation of sonar images.

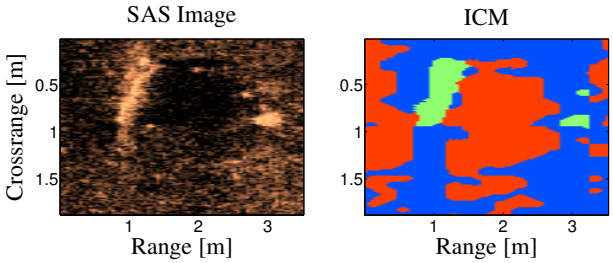

Figure 1: SAS image and ICM segmentation result. The `hlt` label is depicted in green, the `shw` label in red and the `bkg` label in blue.

In Sec. 2 the graph theory is summarized. Sec. 3 deals with the modeling of both regional and boundary properties of the sonar images within the graph theory frame. The initialization of the graph after the ICM segmentation result is described in Sec. 4. In Sec. 5, we present the segmentation results provided by both the ICM based algorithm (implemented as proposed in [1]) and the graph cut for a database of real SAS images with 378 man made objects. The performance of the two algorithms is compared using a statistical classifier. Finally, conclusions are provided in Sec. 6.

2. GRAPH THEORY

A directed weighted graph $\mathcal{G} = \{\mathcal{V},\mathcal{E}\}$ consists of a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. An edge represents a connection between two ordered nodes, that is, $\mathcal{E} = \{\{p,q\}|p,q \in \mathcal{V}\}$. A function $w\colon \mathcal{E} \to \mathbb{R}^+$ assigns a positive real valued weight to each edge, denoted by $w_{\{p,q\}}$. Since the graph is directed $w_{\{p,q\}} \neq w_{\{q,p\}}$ [9].

Grid graphs are typically employed in computer vision to represent images, since the alignment of nodes in rows and columns is a natural representation of the image pixels $p \in \mathcal{P}$. A neighborhood system, $\mathcal{N} = \{\{p,q\}|p,q \in \mathcal{P}\}$, has to be chosen to establish the edges configuration connecting the different pixel nodes. For our application, the second order neighborhood system is considered, which implies that each pixel is connected to its 8 direct neighbors.

For the purpose of image segmentation some extra nodes denoted as terminals are required. Each terminal corresponds to one of the possible pixel labels. For binary segmentation we need two terminals, the source s and the sink t. In a terminal graph, each pixel is connected not only to its 8 neighbors, but also to the terminal nodes s and t. Hence, we distinguish two kinds of edges: n-links (neighboring links) are edges between two pixel nodes and t-links (terminal links) are edges between a pixel node and a terminal. Thus, the two-terminal graph is defined as:

$$\mathcal{G}\colon \mathcal{V} = \mathcal{P}\cup\{s,t\},\ \mathcal{E} = \mathcal{N}\cup\underbrace{\{\{p,s\},\{p,t\}|p\in\mathcal{P}\}}_{t\text{-links}}. \qquad (2)$$

A two-terminal graph representing a 3×3 image is depicted in Fig. 2.

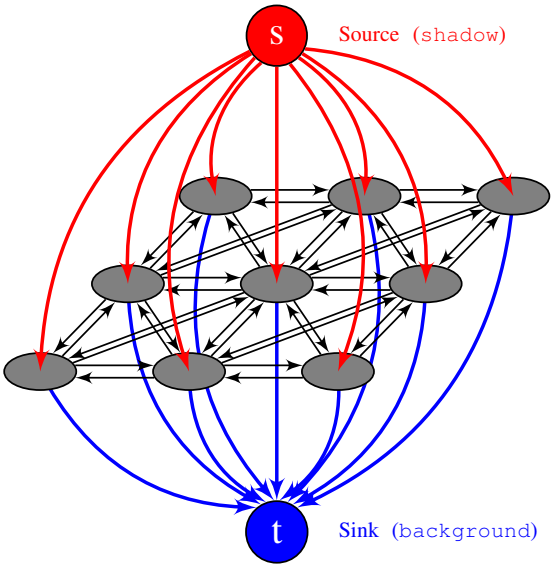

Figure 2: Graph representation of a 3×3 image with two terminals and a second order neighborhood system for the n-links.

2.1 Min-Cut / Max-Flow

An s/t cut C (hereafter referred to only by cut) on a graph separates the set of nodes $\mathcal{V}$ into two disjoint subsets $\mathcal{S} \subset \mathcal{V}$ and $\mathcal{T} \subset \mathcal{V}$, $\mathcal{S} \cap \mathcal{T} = \emptyset$, such that the source $s \in \mathcal{S}$ and the sink $t \in \mathcal{T}$. In our application, $\mathcal{S}$ corresponds to the `shw` label and $\mathcal{T}$ to the `bkg` label. A cut $C = \{\mathcal{S}, \mathcal{T}\}$ is a subset of $\mathcal{E}$ containing all edges $\{p,q\}$ where $p \in \mathcal{S}$ and $q \in \mathcal{T}$. The cost of a cut $|C|$ is defined as the sum of the weights of the edges in C, i.e., $|C| = \sum_{\{p,q\} \in C} w_{\{p,q\}}$. The minimum cut is defined as a cut on graph $\mathcal{G}$ that has minimum cost.

The min-cut/max-flow theorem states that for any directed linear graph the maximum flow value from s to t is equal to the cost of the minimum cut separating s and t [9]. In other words, finding the minimum cut of a graph is equivalent to finding its maximum flow.

In order to illustrate the concept of flow in a graph, let us interpret the directed graph as a network and the edges as pipes connecting the nodes. Each pipe has a certain capacity $c_{\{p,q\}}$ that corresponds to the weight of the edge $w_{\{p,q\}}$. Now we can push a flow $f(s,t)$ through the network leaving the source and arriving at the sink. According to the min-cut/max-flow theorem, $f_{\max} = |C|_{\min}$.

Before stating the conditions that define a finite flow f in a network, let us denote all outgoing edges from node p and all incoming edges to node p by,

$$\begin{aligned} O(p) &= \{\{p,q\} \in \mathcal{E} | q \in \mathcal{V}\} \\ I(p) &= \{\{q,p\} \in \mathcal{E} | q \in \mathcal{V}\}, \end{aligned} \tag{3}$$

respectively. The first condition is given by,

$$\sum_{q \in O(p)} f(p,q) - \sum_{q \in I(p)} f(q,p) = \begin{cases} f & \text{if } p = s \\ -f & \text{if } p = t \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

which is comparable to Kirchhoff's current law. Assuming that outcoming flows are positive and incoming flows are negative, the sum of all outgoing and incoming flows must be zero for all nodes but the source and the sink. The flow emerging from the source, f, is equal to the flow arriving at the sink. Secondly, capacities must be finite, i.e., $c_{\{p,q\}} < \infty$. Finally, the flow within an edge cannot exceed its capacity, $f(p,q) \leq c_{\{p,q\}}$.

edge	$w_{\{p,q\}}$	link
$\{p,q\}$	$B_{p,q}$	n-link
$\{p,s\}$	$\lambda \cdot R_p(\texttt{bkg})$	t-link
$\{p,t\}$	$\lambda \cdot R_p(\texttt{shw})$	t-link

Table 1: Edge weighting

2.2 Efficient Implementation

We have adopted the min-cut/max-flow algorithm proposed in [8], which is broadly used in the literature. It is based on the augmenting path concept [9]. The algorithm works on a residual graph $\mathcal{G}_f$, which is initialized as $\mathcal{G}$. In each iteration, a path along non saturated edges from s to t is searched in $\mathcal{G}_f$. The smallest capacity along the path determines the maximum flow Δf that can be pushed. The residual capacities of the edges along the augmented path are reduced by Δf, while the residual capacities of the reverse edges are increased by the same amount. The total flow from s to t is increased, $f = f + \Delta f$. The algorithm terminates when there are no more $s \rightarrow t$ possible paths.

3. EDGE WEIGHTING

Segmenting an image using graph theory is equivalent to finding the minimum cut of its associated graph. Therefore, the segmentation result is determined by the edge weights. There are two kinds of edges, the n-links and the t-links (see Sec. 2). The former link each pixel with its neighbors, while the latter link each pixel with the source and the sink. Hence, it is natural that the weights of the n-links account for the so-called boundary properties of the image (related to $P(\mathbf{x})$ in Eq. (1)) while the t-links depend on its regional properties ($P(\mathbf{y}|\mathbf{x})$ in Eq. (1)).

We can express the cost of the label field $\mathbf{x}$ as [10]:

$$E(\mathbf{x}) = \lambda \cdot R(\mathbf{x}) + (1-\lambda) \cdot B(\mathbf{x}) \tag{5}$$

where the coefficient $\lambda \in [0,1]$ specifies the relative weighting of the regional property term $R(\mathbf{x})$ with respect to the boundary property term $B(\mathbf{x})$, and

$$\begin{aligned} R(\mathbf{x}) &= \sum_{p \in \mathcal{P}} R_p(x_p) \\ B(\mathbf{x}) &= \sum_{\{p,q\} \in \mathcal{N}} B_{\{p,q\}} \cdot \delta_{x_p \neq x_q}. \end{aligned} \tag{6}$$

Note that the boundary term associated to an edge, $B_{\{p,q\}}$, contributes to $B(\mathbf{x})$ only if $x_p \neq x_q$.

Let us now describe how to assign the weights $w_{\{p,q\}}$ to the edges so that the expression in Eq. (5) corresponds to the cost of the cut defined by a certain labeling $\mathbf{x}$, that is, $E(\mathbf{x}) = |C|$. Considering the definition of a cut on a two terminal graph (see Sec. 2.1), we can make the following statements regarding a cut C:

- if $p \in \mathcal{S}$ then $\{p,t\} \in C$
- if $p \in \mathcal{T}$ then $\{p,s\} \in C$
- $\{p,q\} \in C$ iff $p \in \mathcal{S}$ and $q \in \mathcal{T}$.

For every node $p \in \mathcal{P}$, exactly one t-link is severed by the cut. If two neighboring pixels p and q are labeled differently, the edge with its origin in $\mathcal{S}$ and destination in $\mathcal{T}$ is severed by the cut. Then, an assignment of weights to the edges of graph $\mathcal{G}$ according to Table 1 ensures a minimization of $E(\mathbf{x})$ by the minimum cut on $\mathcal{G}$.

3.1 Regional Properties

The more likely a pixel p is to belong to a region, the lower the regional cost $R_p(x_p)$ of assigning the corresponding label to the pixel must be. Hence, it is reasonable to express the regional cost as a function F of the pdf, $P(y_p|x_p)$, of the two regions of the image,

$$R_p(x_p) = F(P(y_p|x_p)), \quad x_p = \{\texttt{shw}, \texttt{bkg}\}. \tag{7}$$

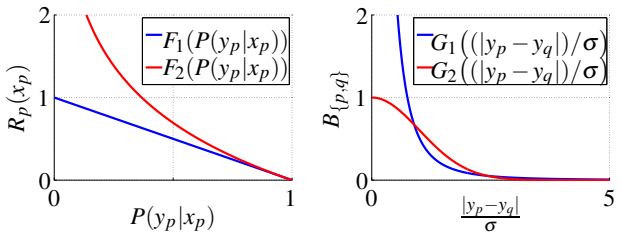

Figure 3: Regional and Boundary properties. For G_1, $m = \frac{1}{4}$ and $n = 3$ have been chosen.

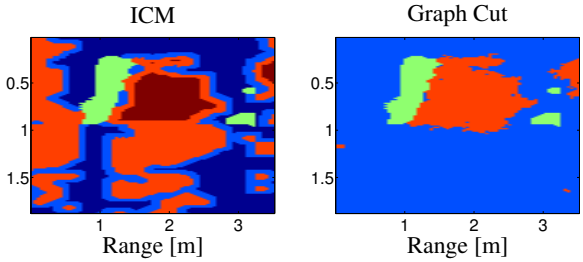

Figure 4: ICM and graph cut segmentation results with $r = 0.2$ and $\lambda = 0.1$ for the sonar image in Fig. 1. The seeds for the graph cut algorithm initialization are highlighted in the ICM image.

Since $R_p(x_p)$ should decrease with increasing $P(y_p|x_p)$, F must be a monotonically decreasing function. Two options have been considered:

$$F_1(x) = 1-x \quad (8)$$
$$F_2(x) = -\ln(x). \quad (9)$$

Fig. 3 shows F_1 and F_2 versus P. While F_1 and F_2 are similar for high values of P, they differ significantly for small P. The difference between the segmentation results that F_1 and F_2 produce is, however, negligible. This is due to the fact that edges with high weights do not get saturated and therefore do not determine the maximum flow. For the examples shown in Sec. 5, F_2 has been chosen.

In order to estimate the pdf of the two regions, $P(y_p|\texttt{shw})$ and $P(y_p|\texttt{bkg})$, the pixels seeds (see Sec. 4) are used. A Weibull distribution has been assumed for both regions [3].

3.2 Boundary Properties

The boundary properties account for the fact that neighbor pixels with similar intensities should belong to the same region. Therefore, the boundary cost $B_{\{p,q\}}$ can be defined as a function G of the magnitude of the pixel intensity difference normalized by the standard deviation σ,

$$B_{\{p,q\}} = G\left(\frac{|y_p - y_q|}{\sigma}\right). \quad (10)$$

Two function families have been studied:

$$G_1(x) = (m+x)^{-n}, \quad n,m \in \mathbb{Q}^+ \quad (11)$$
$$G_2(x) = \exp\left(-\frac{x^2}{2}\right), \quad (12)$$

where n and m need to be chosen. An example of each function family is depicted in Fig. 3. Both functions are similar when $|y_p - y_q| > \sigma$, but differ greatly otherwise. Again, the segmentation results that both functions produce are almost identical, since only the high weights differ and those do not influence the minimum cut. For the examples shown in Sec. 5, G_2 has been used.

4. INITIALIZATION: SEEDS

It is possible to fix the label of a group of pixels, the so-called seeds. The subsets $\mathcal{O} \subset \mathcal{P}$ and $\mathcal{B} \subset \mathcal{P}$, $\mathcal{O} \cap \mathcal{B} = \emptyset$, denote the sets of seeds that are initially labeled as `sdw` and `bkg`, respectively. This influences the labeling of the adjacent pixels via the boundary properties.

In our application the seeds are chosen after the ICM segmentation result. First, a rectangular structuring element is employed to morphologically erode the shadow region. Its dimensions are proportional to those of the ICM shadow, that is, if the size of the smallest rectangle that completely contains the ICM segmented shadow has a size of $M \times N$ pixels, then the structuring element has dimensions $r \cdot M \times r \cdot N$, with $r \in (0,1)$. The pixels that remain and that, moreover, lay to the right of a highlight region, are added to the $\mathcal{O}$ set of seeds. After, the `bkg` region is eroded and the remaining pixels are considered as $\mathcal{B}$ seeds. The greater r is, the less pixels are assigned to the seed sets, that is, the graph cut segmentation is less influenced by the ICM result. Fig. 4 includes the ICM and the graph cut segmentation results for the SAS image in Fig. 1. The seeds that stem from the ICM segmentation with $r = 0.2$ are highlighted.

We need to assure that the seed pixels do not change label during the max-flow search. According to [10] this is achieved by setting:

- if $p \in \mathcal{B}$, $w_{\{p,s\}} = 0$ and $w_{\{p,t\}} = L$
- if $p \in \mathcal{O}$, $w_{\{p,t\}} = L$ and $w_{\{p,s\}} = 0$

where $L = 1 + \max_{p\in\mathcal{P}} \sum_{q:\{p,q\}\in\mathcal{N}} B_{\{p,q\}}$.

5. RESULTS

5.1 Segmentation Results

We have accomplished an empirical study in order to determine suitable values for the parameters λ (see Eq. (5)) and r (see Sec. 4). In Fig. 5 the segmentation of two SAS images is shown, assuming values 0, 0.1 and 0.5 for λ. A value $r = 0.1$ has been chosen. If $\lambda = 0$ only the boundary properties are considered. With $\lambda = 0.5$ both regional and boundary properties have the same weight. While the latter configuration is sensitive to noise, the former is too strongly determined by the initialization. A good trade-off is $\lambda = 0.1$.

Fig. 7 shows the segmentation that corresponds to $r = \{0.05, 0.1, 0.15\}$. Low values of r imply that most of the pixels are used as seeds. Hence, the graph cut segmentation is too much influenced by the ICM result and does not add any significant value. On the other hand, if r is too high, too few seed pixels are considered to estimate the pdf for the regional weights (see Sec. 3.1), which might result in a poor performance. A good compromise is $r = 0.1$.

Five SAS images –about 30,000 m^2– containing 378 man made objects (128 cylinders and 250 spheres) have been used to test this algorithm. The ICM based algorithm [1] has been systematically applied. Beside the man made objects (class $\mathbb{M}$), 1795 clutter regions (class $\mathbb{C}$) have been segmented. Fig. 8 shows 5 snapshots of the SAS images, together with the ICM and the min-cut/max-flow segmentation results. A value of $\lambda = 0.1$ has been chosen, and the erosion structuring element is built with $r = 0.1$. The example in the first row shows a high quality sonar image. The ICM segmentation result is good and the graph cut segmentation is almost identical. A dark background area is segmented together with the shadow in the second example by the ICM algorithm. The initialization of the graph cut algorithm with seeds that lay to the right of the highlight region allows for distinguishing the shadow of the object highlight. The spherical objects on row four lays on an uneven background with sand ripples. The ICM algorithm segments some of the sand ripples shadows together with the shadow of the highlight object while the graph cut algorithm gets rid of them. The last example contains a snapshot of the sonar images where no man made object is present. Still, some clutter is segmented.

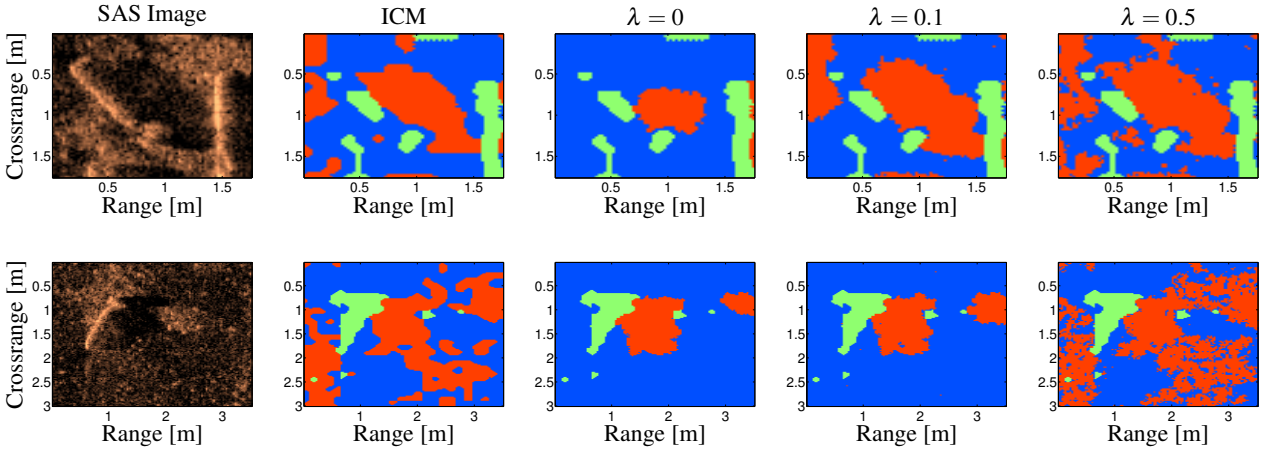

Figure 5: Study on λ. The ICM segmentation result of two SAS images is compared with the graph cut results that stems from $\lambda = \{0, 0.1, 0.5\}$ assuming $r = 0.1$. The first example presents a poor segmentation for $\lambda = 0$ (result too strongly determined by the initialization). The second example illustrates the degradation of the results due to a too high λ.

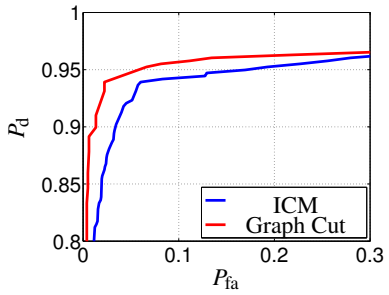

Figure 6: ROC curves

5.2 Classification Results

Based on the classification system proposed in [2], optimal feature sets matching both segmentation approaches have been found. They consist of a combination of geometrical and statistical features for both the shadow and the highlight. Fig. 6 depicts the ROC curve for both algorithms, that is, the empirical probability of detection, P_d (classification of $\mathbb{M}$ objects as $\mathbb{M}$) vs. the probability of false alarm, P_{fa} (classification of $\mathbb{C}$ objects as $\mathbb{M}$). For $P_{fa} > 0.3$, the difference in P_d is negligible. For low false alarms, however, the graph cut segmentation produces higher P_d than the ICM approach. If we fix $P_{fa} = 0.025$, which corresponds to 0.0015 false alarms per squared meter (45/1795 $\mathbb{C}$ classified as $\mathbb{M}$ objects), the ICM provides $P_d = 0.87$ (329/378 $\mathbb{M}$ detected), while $P_d = 0.94$ (355/378 $\mathbb{M}$ detected) for the min-cut/max-flow segmentation. Notice that no distinction has been done between spherical and cylindrical objects.

6. CONCLUSION

Assuming a Markovian image model, a computationally efficient implementation of a graph cut algorithm has been employed to segment a SAS image database. It is the first time that this algorithm is utilized for sonar applications. We have proposed a novel initialization of the graph based on the result of the ICM based algorithm proposed in [1]. Several regional and boundary properties and their influence on the segmentation have been investigated. A comparison between the graph cut and the ICM results has been accomplished. If the image quality is high and the background is uniform, the ICM algorithm already provides satisfactory results. In the case of uneven sea bed, however, applying the graph cut significantly improves the ICM results, increasing the probability of detection up to 7 % for a fixed false alarm rate.

Acknowledgment

We would like to thank Dr. A. Kraft and Dr. K. Siantidis from ATLAS ELEKTRONIK GmbH for providing the data and for their support and valuable comments.

REFERENCES

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

[2] R. Fandos and A. M. Zoubir, "Optimal feature set for automatic detection and classification of underwater objects in SAS images," *IEEE Journal of Selected Topics in Signal Processing*, 2011, to appear.

[3] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Three-class Markovian segmentation of high resolution sonar images," *Computer Vision and Image Understanding*, vol. 76, no. 3, pp. 191 – 204, 1999.

[4] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 1, pp. 90 – 105, 2003.

[5] J. Besag, "On the statistical analysis of dirty images," *Journal of the Royal Statistical Society*, vol. B-48, pp. 259 – 302, 1986.

[6] R. Fandos and A. M. Zoubir, "Enhanced initialization scheme for a three-region markovian segmentation algorithm and its application to SAS images," in *Proceedings of the European Conference on Underwater Acoustics*, 2010.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Markov Random Fields with efficient approximations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 1998, pp. 648 –655.

[8] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124 – 1137, 2004.

[9] L. R. Ford and D. R. Fulkerson, *Flows in Networks*, N. J. Princeton, Ed. Princeton University Press, 1962.

[10] Y. Boykov, O. Veksler, and R. Zybih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222–1239, 2001.

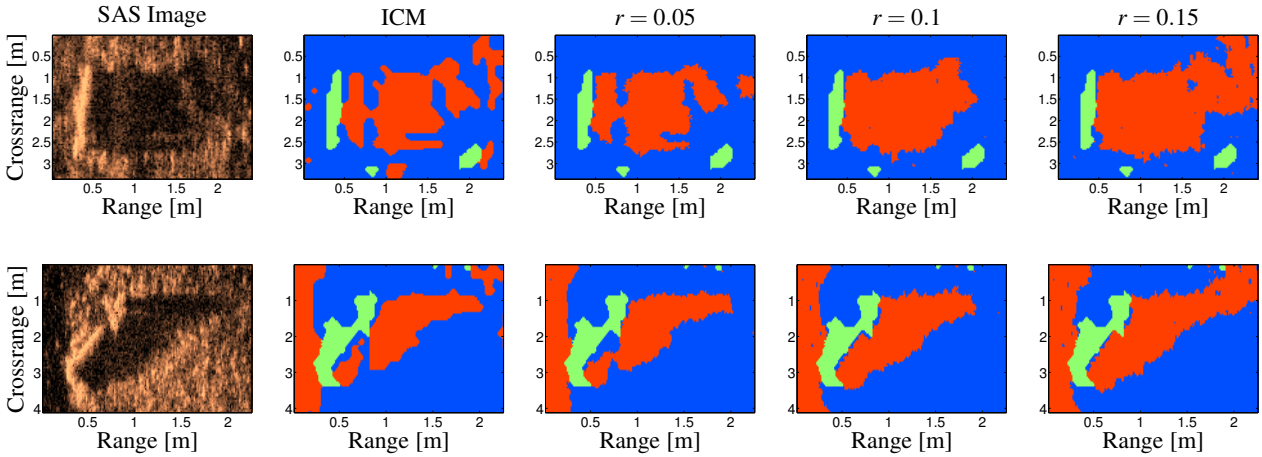

Figure 7: Study on r. The ICM segmentation result of two SAS images is compared with the graph cut results for $r = \{0.05, 0.1, 0.15\}$ and $\lambda = 0.1$. The graph cut result for the first examples is very close to the ICM solution for $r = 0.05$. For higher values of r, the result differs and is more convenient for further classification purposes. Both examples show a poor segmentation for high r values ($r = 0.15$).

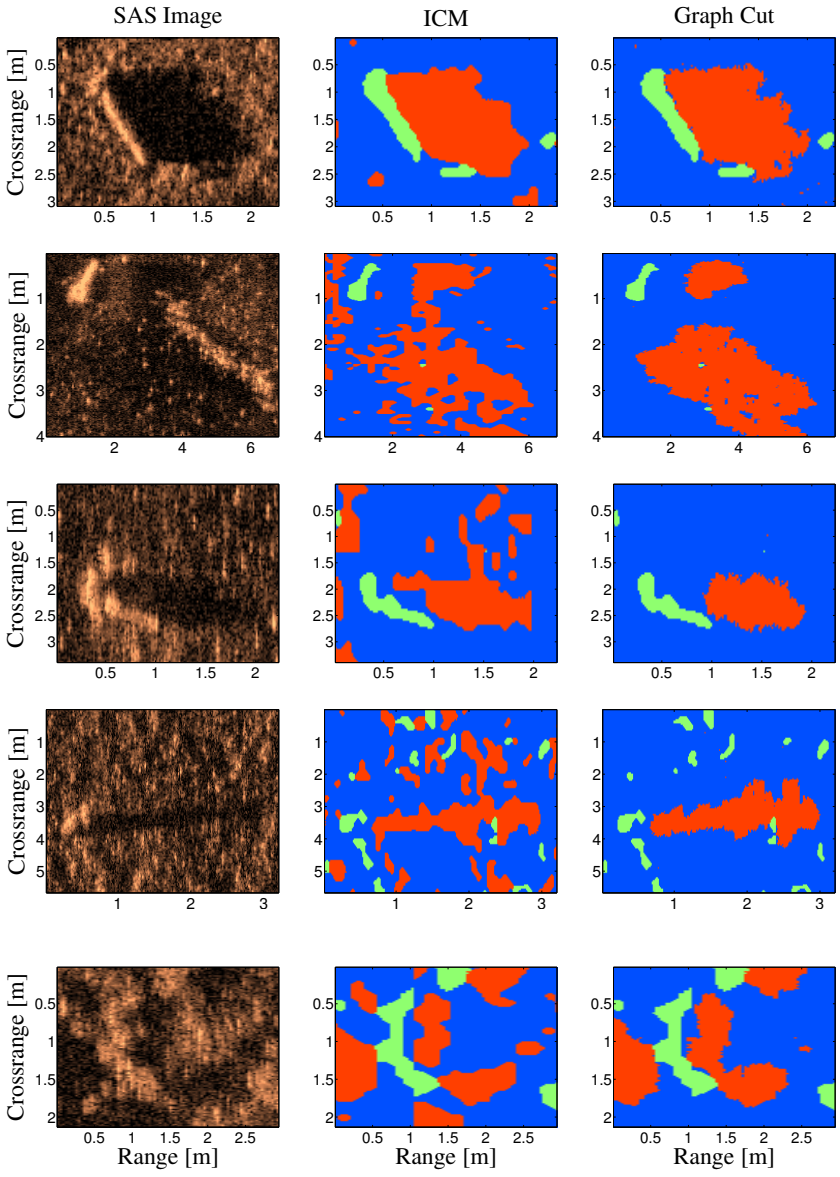

Figure 8: ICM segmentation algorithm vs. graph cut based segmentation algorithm for 5 snapshots of SAS images. The parameters $\lambda = 0.1$ and $r = 0.1$ have been chosen.