euronoise

Target Detection of Man Made Objects in Side Scan Sonar Images - Segmentation based False Alarm Reduction -

Max Neumann[a], Christian Knauer[a], Bodo Nolte[b], Dieter Brecht[b], Wolfgang Jans[b] and Alfons Ebert[c]

[a]Freie Universität Berlin, Takustr. 9, 14195 Berlin, Germany
[b]Forschungsanstalt der Bundeswehr für Wasserschall und Geophysik, Klausdorfer Weg 2-24, 24148 Kiel, Germany
[c]FGAN-FOM, Gutleuthausstraße 1, 76275 Ettlingen, Germany
bodonolte@bwb.org

This paper describes a fast and robust algorithm significantly reducing the number of false detections produced by simple screening algorithms for side scan sonar (SSS) images. The presented algorithm consists of two processing steps. First, an iterative segmentation process is carried out in order to separate the image into shadow and background. This segmentation is based on an energy function that combines the local neighborhood segment information and the amplitude of a pixel. By minimizing this function, a clear shadow, the most significant target characteristic, can be extracted. Second, a robust classification approach is applied based on the *Region of Interest* (ROI) and the shadow contour, utilizing the area of the shadow, the first two statistical moments of the pixel amplitudes and the existence of parallel lines (Hough transformation).
The algorithms were tested using a data set with approx. 2400 ROIs, containing about 200 targets and 270 target like stones or sand ripple structures. This data set was gathered during five different measurement campaigns in the Baltic Sea and the Mediterranean Sea using three different SSS systems (*Benthos C3D*, *Klein2000* and *Marine Sonics*). These data were collected by *FWG* and *WTD71* as well as by *Atlas Elektronik* with the *SeaOtter MK1* AUV.

1 Introduction

Modern side scan sonars provide high resolution images of the seafloor. This allows the detection of ground mines and other hazardous objects in harbors, coastal areas and commercially / strategically important waterways. Image analysis including detection and classification is usually done (or at least supported) by computer based algorithms. For many reasons (safety, efficiency, covertness), most concepts for state-of-the-art and future systems are based on autonomous underwater vehicles (AUVs) as platform. In order to support autonomy, algorithms need to be developed which ensure automatic (unsupervised), robust and efficient detection, localization and classification of significant objects from the sonar images.

One of the most reliable features allowing a discrimination/classification of proud targets from sonar images is the typical object shadow cooccurring with the highlight structure. In this paper we present a simple two-step approach exploiting the shadow structure and shape. The input for the algorithm is a Region of Interest (ROI) which is provided by a prior detection algorithm. This pre-detection algorithm (not presented here) can be simple and fast, and a high false alarm rate is accepted to avoid target overlooking.

In the presented procedure, the false alarm rate is reduced in two processing steps. First the ROI is segmented by an iterative algorithm using fuzzy functions into *highlight*, *background* and *shadow*. In the next step, the extracted shadow is used for a classification into the classes *potentially man made* or *natural object*. All natural objects are sorted out and the man made objects can be used for further analysis.

2 Segmentation

For detection of man made objects, the shadow is the most interesting part of the image. Many studies are dealing with an automatic extraction of the object's shadow. Simple approaches use global thresholds for segmentation [3]. These methods generate good results on gravel ground. For bumpy seabed containing silt or rocks other methods with local thresholds are more promising [4]. The main disadvantage of all threshold based variations is the poor robustness against speckle and noise. The difference between a dark noise pixel and a dark shadow pixel can not be found only in the pixel's color. Since SSS images are typically very noisy, some segmentation approaches exploit the type of noise (its specific statistical properties) with Markov Random Field concepts. These methods use the different kinds of noise in a local neighborhood of a point $p \in \mathscr{P}$ where $\mathscr{P}$ is the set of all pixels. By minimizing a noise related function, the shadow region can be extracted [8],[7],[5] or the sea bottom can be segmented [1].

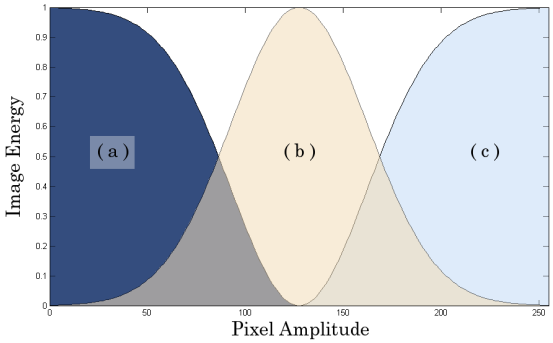

Figure 1: Image energy for the three segment classes *shadow* (a), *background* (b) and *highlight* (c)

Another approach that includes the local neighborhood concept combines two fuzzy functions [6]. One function evaluating the pixel brightness $I_p \in [0,255]$ and one evaluating the connectivity. This kind of segmentation is used in the present paper.

With the pixel segment association s:

$$\begin{aligned} s &: p \rightarrow \{-1,0,1\} \\ s_p &= \begin{cases} -1 & \text{if } p \in \text{shadow} \\ 0 & \text{if } p \in \text{background} \\ +1 & \text{if } p \in \text{highlight} \end{cases} \end{aligned}$$

and a *Gaussian Bell* curve:

$$\mathscr{N}_{\overline{I_p},\sigma}(x) = e^{-\frac{(x-\overline{I_p})^2}{2\sigma^2}}$$

with mean $\overline{I_p}$ and standard deviation σ, an image energy evaluating the pixel brightness can be defined as:

$$\begin{aligned} E_{img} &: (p,s_p) \rightarrow [0,1] \qquad (1) \\ E_{img}(p,s_p) &= \begin{cases} |s_p| - (2 \cdot s_p - 1) \cdot \mathscr{N}_{\overline{I_p},\sigma}(I_p) \\ \quad \text{if } s_p \cdot (\overline{I_p} - I_p) \leq 0 \\ 0 \quad \text{otherwise} \end{cases} \end{aligned}$$

This energy function can easily be illustrated by figure 1. In case p is a background pixel with $s(p) = 0$, the image energy E_{img} will be calculated by the Gaussian Bell curve with the center at the mean pixel amplitude $\overline{I_p}$ of the image. A

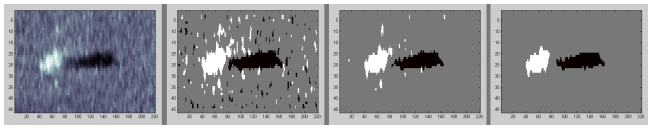

Figure 2: ROI ($7m \times 5m$) with an object (left) and iterations of segmentation after $t = 0, 2, 4$ steps. White = highlight, gray = background, black = shadow

background pixel reaches maximum energy values for pixel amplitudes around the mean amplitude of the image.

In the case that p is part of the shadow area with $s(p) = -1$ and the pixel amplitude is beneath the mean $\overline{I_p}$, the energy is calculated as the difference between one and the Gaussian curve from before. If $I_p > \overline{I_p}$, the image energy is set to zero because a shadow has to be darker then the average amplitude. Nearly the same is done for getting the image energy of a highlight pixel, but with opposite sign. To sum up, high energy is associated to dark shadow pixels, bright highlight pixels and background pixels if they have midrange amplitudes.

The global maximum of E_{img} as a function of s can be used to derive a first segmentation for every pixel p. The same segmentation can be found by a simple threshold based segmentation with thresholds at $\mathcal{N}_{\bar{x},\sigma}(I_p) = \frac{1}{2}$.

Due to noise and speckle, there are still many wrong and deformed segments. These errors can not eliminated by just comparing the pixel amplitudes. Therefore the neighborhood energy is introduced. This energy utilizes the simple feature of real object shadows that, compared to a dark noise pixel, the local neighborhood of a shadow pixel consists of many other shadow pixels.

With the set of local neighborhood pixels κ the neighborhood energy can be defined as:

$$\begin{aligned} E_{hood} &: p \to [0,1] \\ E_{hood}(p) &= \mathcal{N}_{|\kappa(p)|,\frac{|\kappa(p)|}{4}}(\alpha(p)) \end{aligned} \tag{2}$$

where $\alpha(p)$ counts the neighborhood pixels in the same segment class:

$$\begin{aligned} \alpha &: p \to [0,|\kappa|] \\ \alpha(p) &= |\{p_N \in \kappa(p) | s(p) = s(p_N)\}| \end{aligned}$$

So the more neighbors of p are in the same segment class as p, the higher the neighborhood energy becomes. In the present work the 3×3 neighborhood is used.

To combine the energies from equation (1) and (2) the weighted sum is chosen with the parameters $k_1, k_2 \in [0,1]$ and $k_1 + k_2 = 1$:

$$E_{total} = k_1 \cdot E_{img} + k_2 \cdot E_{hood}$$

A pixelwise maximization of E_{total} as a function of $s(p)$ will result in an improved segmentation into *highlight*, *shadow* and *background*. For every pixel p the segment $s(p)$ is chosen in the way that $E_{total}(p,s)$ is maximal.

It is obvious that nearly all segment flipping actions will happen only to pixels at segment borders. A pixel only surrounded by pixels of the same segment class will have a maximal neighborhood energy. For runtime reasons therefore only the energy of the segment border pixels is calculated and their class is flipped if necessary. The algorithm stops after

Algorithm 1 Iterative three class segmentation

```
given    :  p ∈ 𝒫 and I(p)
quested  :  segmentation s ∈ {−1,0,1}
              -1  shadow
               0  background
               1  highlight
```

```
s      ←  initial threshold based
          segmentation
i      ←  0
while s_old ≠ s ∧ i < 20
   s_old      ←  s
   i          ←  i+1
   κ(p)       ←  neighbors of p
   𝒫_border   ←  {p ∈ 𝒫 | ∃q ∈ κ(p) :
                    s(p) ≠ s(q)}
  for each p ∈ 𝒫_border
        s(p)  ←  s' ∈ {−1,0,1}, so that
                 E_total(p, s(p) = s') is maximal
  end for
end while
```

a predefined depth of iteration or by reaching a stable state where no pixels are flipping their segment anymore.

The different iteration steps can be seen in figure 2. The ROI in the image on the left side has a size of $7m \times 5m$ and contains a cubiform anchor. From left to right there are the three segmentations after $0, 2, 4$ iterations. The first segmentation at $t = 0$ is equivalent to a simple threshold based partitioning of the image, like aforementioned. After four iterations all segments based on noise and speckles are eliminated and only the object highlight and shadow remain.

This algorithm is very fast in practice. In general the segmentation is stable after two to seven iterations. The only disadvantage is the poor segmentation result when finding the object highlights. The highlight structure typically consists of individual bright speckles which are not necessarily connected. The segmentation performance is consequently lower as against the sharp shadow structure. For that reason only the shadows are taken for the following classification step.

3 Classification

The classification for false alarm reduction is based on the major shadow segment from the segmentation. Three fuzzy functions $f_1, f_2, f_3 \in [0,1]$ are extracted and combined to an object-likeness-function. The three functions describe the shadow area, the distribution of the pixel amplitudes and the existence of parallel lines, where higher function values stand for higher object likeness of the corresponding feature.

The first function f_1 can be extracted directly from the shadow area A:

$$f_1 = \mathcal{N}_{\bar{A},\sigma}(A) \tag{3}$$

where $\bar{A}$ is a typical area, known a priori and σ is a high standard deviation for scaling deviant shadow areas.

The second function f_2 evaluates the distribution of the pixel's gray values. An object in SSS has a highlight and a shadow. In the histogram of the amplitudes this fact is noticeable by many low values and some high values but less

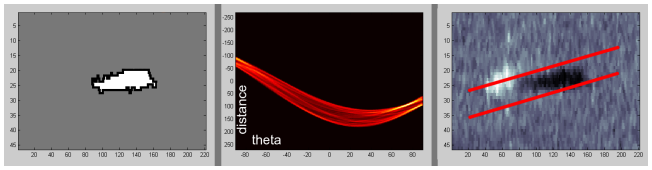

Figure 3: From left to right: extracted shadow (white) with contour (black) in a $7m \times 5m$ ROI, parameter space of the Hough transformation with two parallel lines at $\theta = 70^o$ and the same parallel lines transformed back into the ROI with the object

values in the middle of the amplitudes range. So, compared to normal background pixels with no object, the standard deviation of the pixel amplitudes is much higher. To get all object pixels with highlight and shadow, the found shadow is enlarged to every side by $\frac{1}{4}m$ and by $1m$ to the left side, where the highlight should be. With the enlarged pixel set $U \subseteq P$:

$$f_2 = sc\left(\frac{\sigma(\{I(p)|p \in U\})}{\sigma(\{I(p)|p \in P \backslash U\})}\right) \quad (4)$$

where sc is a function scaling the quotient to $[0,1]$.

In general, most man made objects have simple geometries like cylinders or cubes, so the existence of parallel lines indicates a not natural object on the seafloor. The last fuzzy function f_3 identifies these parallel lines in the shadow contour through the *Hough Transformation* [2].

As shown in figure 3, the Hough transformation extracts every line through many edge points. A line is represented by the parameter d as the distance of the line to the center and the angle θ to the x-axis. For each point of the shadow's contour, all possible lines through this point are calculated and written into the parameter space (histogram). One point in the parameter space is associated with one line. Clusters of points in the parameter space are lines through many points on the shadow contour, as seen in figure 3 in the middle. The two parallel lines g_1 and g_2 with the most points on them and the same angle θ are chosen for the last feature:

$$f_3 = \frac{|\{p_{xy} \in \text{shadow's contour}|y = g_1(x)\}|}{|\{p_{xy} \in \text{shadow's contour}|y = g_2(x)\}|} \quad (5)$$

W.l.o.g., g_1 contains less border pixel than g_2, so f_3 is maximal for a shadow with parallel lines of the same length.

The object likeness *score* can now be calculated with the fuzzy functions from equations 3,4,5 and the scaling factors $k_1, k_2, k_3 \in [0,1]$ with $k_1 + k_2 = 1$:

$$score = (k_3 + (1 - k_3) \cdot f_3) \cdot (k_1 f_1 + k_2 f_2)$$

4 Results

The segmentation and classification was tested automatically without individual parameter adjustment on 2436 ROIs from five different measurement campaigns. Four datasets where captured at the Baltic Sea with three different sonars, one was captured at the Mediterranean Sea. The ROIs contain 202 verified man made objects, 209 big stones and 2025 false targets like ripples, small stones, speckles or empty seabed. All ROIs from the test case where selected semiautomatically and classified manually for "ground truth". The big stones

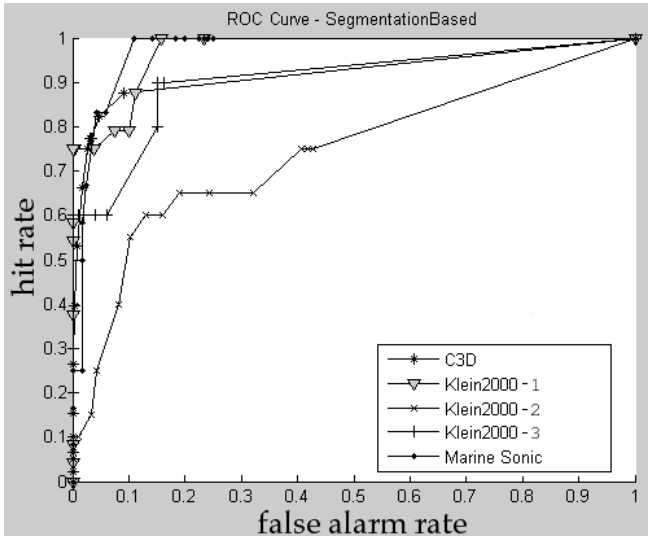

Figure 4: ROC curve for segmentation based false alarm reduction. The outlier, *Klein2000 - 2*, is caused by many ripple marks captured during the campaign

are not counted for the hit rate and not for the false alarms. The ROC curve can be seen in figure 4. Most of the false ROIs where separated out successfully. Only the *Klein2000 - 2* dataset is almost resistant to false alarm reduction. This is a dataset with many ripple marks, which are structures with long and smooth shadows and highlights with many parallel lines. Sometimes the ripple shadows are classified as man made and some objects in the dataset are classified as natural because the object shadow is melted together to the ripple shadows.

Nevertheless, most of the false targets in the other datasets were sorted out correctly. Furthermore the algorithm is very robust. No parameters had to be changed for the different datasets which were captured with different sonars at different places with different frequencies.

5 Conclusion

In this paper a new approach for false alarm reduction is presented. The experimental validation with different data sets has shown that the algorithm works robust on SSS images containing no major ripple marks, even if they are captured by different sonars.

The algorithm is based on a fast iterative segmentation approach which combines pixel amplitudes and the local neighborhood information. The segmentation and the classification work independent from each other, so each part can easily be modified or replaced.

Further improvements can also be expected by different combinations of the fuzzy functions. There are many types of possible associations for the presented fuzzy functions which have not been analyzed in the present work and which can further improve the robustness and the hit rate of the algorithm.

References

[1] J. Gazagnaire, P. Beaujean, and J.Stack. Combining model-based and in situ performance prediction to eval-

uate detection & classification performance. In *Proceeding of the Institute of Acoustics*, volume 29, pages 11–18, 2007.

[2] P.V.C. Hough. Machine analysis of bubble chamber pictures. In *International Conference on High Energy Accelerators and Instrumentation*, pages 554–556, Geneva; Schweiz, 1959. CERN.

[3] Wolfgang Jans. Aufklärung und Bildverarbeitung. In *Symposium Sensorik*, Mannheim, October 2002.

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, September 2004.

[5] Max Mignotte, Christophe Collet, Patrick Pérez, and Patrick Bouthemy. Three-class markovian segmentation of high resolution sonar images. *Computer Vision and Image Understanding: CVIU*, 76(3):191–204, 1999.

[6] V. L. Myers. Sonar image segmentation using iteration and fuzzy logic. In *Proceedings on CAD/CAC Conference*, Halifax, Nova Scotia, Canada, 2001.

[7] Dzung L. Pham and Jerry L. Prince. A generalized em algorithm for robust segmentation of magnetic resonance images. In *Conference on Information Sciences and Systems*, number 33, March 1999.

[8] Patrick Pérez. Markov random fields and images. *CWI Quarterly*, 11(4):413–437, 1998.