CHALMERS

Tools for classification of mine-like objects in synthetic aperture sonar images

Master of Science Thesis

JOHAN ENGSTRÖM

Department of Signals and Systems
Division of Biomedical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2007
Report No. EX095/2007

Tools for classification of mine-like objects in synthetic aperture sonar images

Johan Engström
enga@etek.chalmers.se

November 13, 2007

Abstract

Analyzing sonar images is monotonous and time-consuming but important work; a great number of mines are placed in oceans around the world, some of them are left over from the world wars. Sea mines are also a potential terrorist threat. Using synthetic aperture sonar (SAS) instead of conventional sidescan increases the spatial resolution of the sonar image and therefore also improves the use of a CAD/CAC system based on image analysis. This thesis is a preliminary study into the application of such a system to SAS images and gives a short introduction to sonar systems. It presents two segmentation methods and proposes a number of features to be used for classification purposes. The first segmentation method is based on region growing and uses local statistics (local mean and standard deviation). The second method uses Markov random fields theory. The features discussed describe the smoothness of the object surface and object shape and size. Some features show promise for classification purpose and the results for the segmentation schemes are also promising. However a larger test set is needed to draw any firm conclusions.

Foreword

Working with this thesis has been interesting and stimulating work from which I have learned a lot, not only about image analysis and sonar systems, but also about structuring a project. I am very grateful for the chance to work on this thesis, and would like to thank all at SAAB Underwater Systems (SUS) for supporting me in different ways, ranging from providing data and discussing ideas, to giving me inspiration and company over a cup of coffee. In particular I would like to thank my supervisor at SUS, Mikael Trieb. Dr Geoffrey A. Shippey at Chalmers University of Technology is also acknowledged for discussing ideas, reviewing the text and giving helpful comments. Without his help the thesis would not have been the same.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

The first sea mine was invented in 1776 by the American inventor David Bushnell. Since then, sea mines have been used and played a major role in many wars and the defence of different areas.
During the world wars, around 165000 mines were placed in the Baltic Sea and the West Sea, and many of these remain unexploded on the seabed [1]. One of the reasons why sea mines have been used so much is that it is a cheap way to cause great damage to enemy ships. For that reason, they could also present a terrorist threat.
Mine-hunting systems employ sophisticated sonars to produce an image of the seafloor. Traditionally, mine-like objects are then detected and classified by a trained human operator. Mine countermeasures are often carried out over large areas, generating large amounts of sonar data. Analysing all this data is monotonous and time-consuming work. For this reason, and also the human factor that different operators make different classifications, there is a requirement for automatic detection and classification systems. A wide research literature on the subject exists and some examples can be found in the bibliography.
This Master thesis is part of the SAPPHIRES project at SAAB Underwater Systems in Motala. This project is concerned with the development of an AUV (autonomous underwater vehicle) carrying a SAS (synthetic aperture sonar) system. The introduction of SAS techniques, derived originally from SAR (synthetic aperture radar), increases the spatial resolution of sonar images substantially, and should make detection and classification easier, whether automatically or by human operator. The improved resolution has increased interest in applying image analysis techniques to sonar images. In the future CAD/CAC (computer-aided detection and classification) based on image analysis will be a major part of sonar mine-hunting systems, in parallel with other types of signal-processing applied directly to the received signals. Perhaps the human operator will not be eliminated for a long time to come, but image analysis will simplify

the operator's job and increase accuracy.

1.2 Scope of this thesis

It is a major task to develop a mine classification system for SAS imagery, requiring time and effort well outside the scope of this thesis. Moreover a large training set, including all the common types of sea mine plus a wide variety of natural features is required. However the AUV itself is still under development, which limits the available number of experimental images. Hence this thesis is intended to be a first CAD/CAC study applied to SAS imagery, covering an overview of the problem and some software tools for use in a future system. The main focus is on the segmentation of detected objects into highlight and shadow regions, which will have a great impact on the success of the system, but some classification features are also discussed. One physical limitation is that the SAPPHIRE sonar frequency is too high for bottom penetration, so only objects lying on, or projecting above, the seabed are considered. Completely buried objects are excluded.
Chapter 2 gives a brief introduction to sonar systems. Chapter 3 discusses the different regions in a SAS image, while Chapter 4 discusses image statistics. Chapters 5 and 6 introduce two different segmentation algorithms. Chapter 7 considers several classification features. Results are presented in Chapter 8. In the last chapter, Chapter 9, are conclusions and proposals for future work discussed.

1.3 Mine-like object (MLO)

There is no exact definition of a mine-like object, (MLO). They are simply objects which look like mines to a sonar. There exist many different types of sea mines. Most countries have their own types and some of them are secret. A list of publicly known mines can be found in [2] giving also sizes. The size of a mine varies considerably, but it can be seen from this list that an object smaller than 0.3 [m] in any dimension is probably not a mine. According to the list the largest sea mine in the world measures $3.4 \times 3.4 \times 3.0$ [m], but there exist mines which are larger than this in one dimension. Sea mine can look quite different from one another, but many common mines are based on simple geometric shapes such as sphere, cylinder, truncated cone or box.

1.4 CAD/CAC system

A CAD/CAC system can be structured in different ways. In this thesis, the strategy is firstly to detect candidate objects in a large area image, following which the highlight and shadow regions are segmented for use in feature extraction. These features are then used to classify the object as a MLO or otherwise.

This thesis is concerned with segmentation and feature extraction, assuming the suspicious object has been detected and framed in a much smaller area image.

Chapter 2

Introduction to synthetic aperture sonar

2.1 Sonar

SONAR stands for sound navigation and ranging. It is used in underwater applications instead of RADAR because electromagnetic waves suffer great attenuation in water. An active sonar system consists of one or more arrays of transducers which convert electrical energy to sound and vice versa, acting as transmitters (projectors) and/or receivers (hydrophones). The transmitter directs a sound pulse in a specific frequency band at some sector of the seabed. This is called a ping. The receiver array records the echo reflected back towards the sonar. In a passive sonar, no sound is projected, and only the sound originating from targets and the environment is received at the array.
By taking the relative position of the receiver array elements into account, together with the time delay of the received signals, the contribution to the received signal from different points in the sector can be calculated.
Sonar is used in different modes, sometimes pointing downwards or obliquely downwards towards the seabed for bathymetric estimation, sometimes pointing sideways along the surface of the seabed. Synthetic aperture sonar is a version of this latter mode, and it is the one we are concerned with here [3].

2.2 Types of echo

2.2.1 Specular reflection and backscatter

At a perfectly smooth surface, all the energy is reflected in the specular (mirror-like) direction, meaning that the incident angle θ_i (Fig. 2.1) is equal to the reflected angle θ_r. No object encountered by a sonar is perfectly smooth, but it will be effectively smooth if the roughness dimension is very much less than the sonar wavelength. The sonar will only receive specular reflection when

$\theta_i = \theta_r = 0^o$, i.e. when the plane of the surface is normal to the direction of the transmitted pulse. If this direction is nearly horizontal, then the surface must be near vertical. Hence specular reflections are usually received from smooth surfaces of man-made objects standing up from the seabed, but occasionally from smooth near-vertical rock walls.
A rough surface will act both as a reflector and a scatterer. The dominant component depends on the roughness of the surface relative to wavelength. At high enough frequencies, most surfaces act as pure scatterers with fairly uniform reflected intensities in all directions (Fig. 2.2). There is no intensity peak as the incident angle changes. Even here, surfaces normal to the incident beam produce stronger echoes, because there is a greater surface area contained within each range interval.
When the surface is neither perfectly smooth, nor rough enough to act only as a scatterer, the intensity will have the characteristic illustrated in Fig. 2.3, i.e. the sound is scattered in all directions, but there is still a peak in the specular direction.

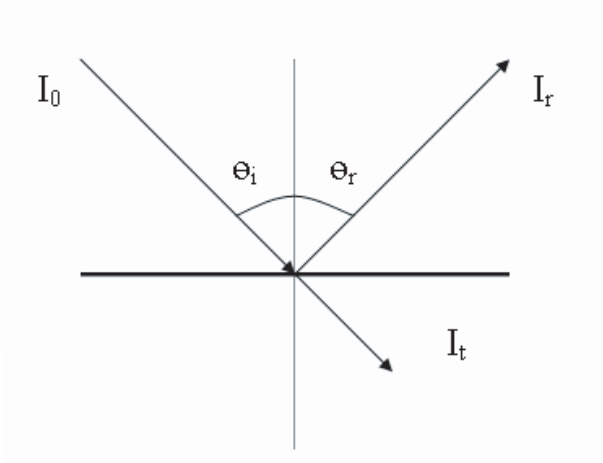


Figure 2.1: Reflection at a smooth surface

Not all the incident energy is reflected back at a surface. Some of it, I_t in Fig. 2.1 is transmitted through the surface and either absorbed by the medium or reflected at a further interface. The reflected sound intensity, I_r, is given by quantity 2.1 [3].

$$\frac{I_r}{I_0} = \frac{Z_r - Z_w}{Z_r + Z_w} \tag{2.1}$$

where Z_w is the specific acoustic impedance of the medium, in this case water, and Z_r is the specific acoustic impedance of the reflector.

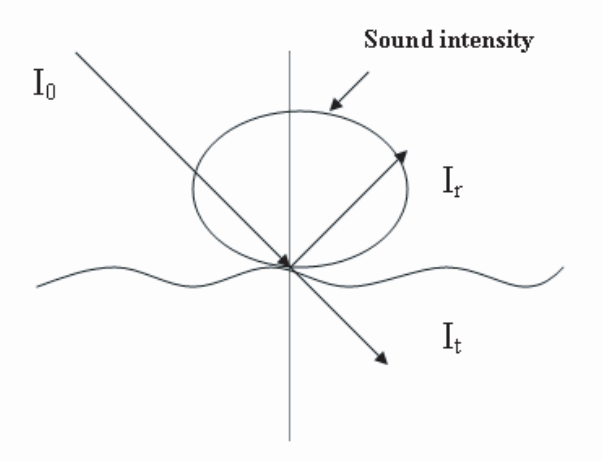


Figure 2.2: Reflection at surface rough enough to act only as a scatter

2.2.2 Seabed reverberation

Reverberation from the seabed varies significantly depending of the appearance of the seabed. The bottom will never be perfectly plane and therefore some scattering will occur and the intensity pattern will look like Fig. 2.2 or 2.3 depending on roughness. The specific acoustic impedance of the bottom may also be significant different, depending on bottom material. For example seabeds consisting of soft sediment can be almost transparent, while rock outcrops return almost all the incident energy. Hence reflected intensity varies considerably, Eq. 2.1. A rocky seabed will reflect more of the energy back toward the sonar, making the detection and discrimination of MLOs harder [4]. It is therefore more likely for a seabed containing scattered rocks to be chosen as a minefield. More details on how the appearance of the seabed affects the CAD/CAC system are discussed in Chapter 3.

2.2.3 Sea-surface reverberation

The sea-surface also reflects or scatters sound, which can affect the sonar system due to multipath propagation of sound in water. The sound travelling from the sonar to an object on the seafloor and back can take alternative paths. The first possibility is the direct-direct path, Fig. 2.4. Other possibilities are the surface-direct, Fig. 2.5, bottom-direct Fig. 2.6, direct-surface, direct-bottom, surface-bottom etc. [4]. Multipath propagation might lead to ghost targets and loss of image contrast by filling in of shadows [5].

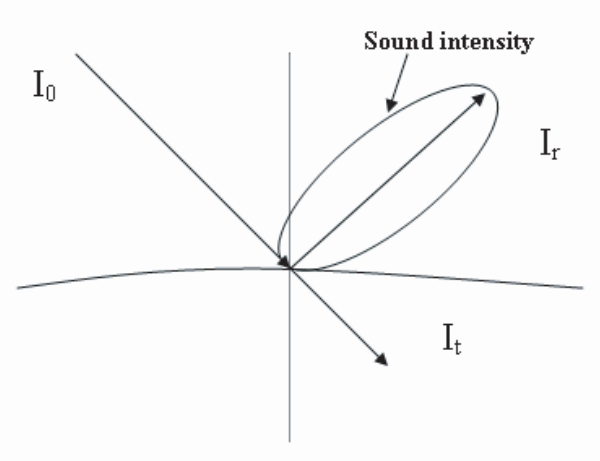


Figure 2.3: Reflection at rough surface

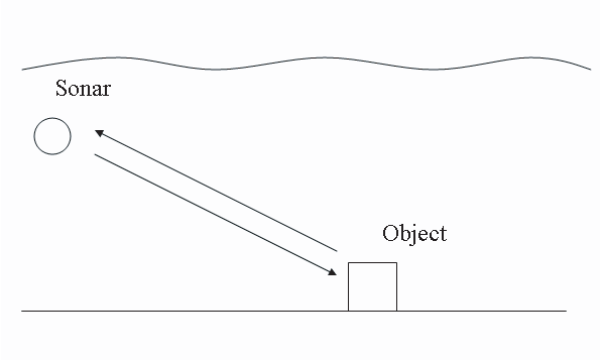


Figure 2.4: Direct - direct path

2.2.4 Object echo

The target strength, TS, is a measure indicating the relationship between incident intensity and reflected intensity measured at some reference point, usually 1 meter from object center.

$$TS = 10 \log \frac{I_r}{I_a} \tag{2.2}$$

TS depends on target shape. For example a perfectly smooth sphere will reflect the sound in different directions, giving a lower TS for a given direction, while a box perpendicular to the sound wave will reflect the energy in one direction, since $\theta_i = \theta_r = 0^o$ in Fig. 2.1, resulting in a high TS for that direction. While the TS of a sphere is independent of incident angle TS for a box is not. If the box is not perpendicular to the sound waves no energy will be

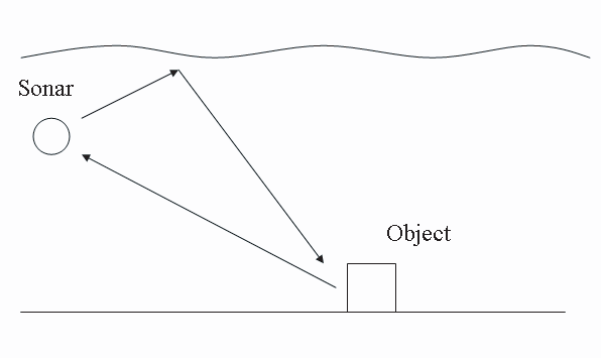


Figure 2.5: Surface - direct path

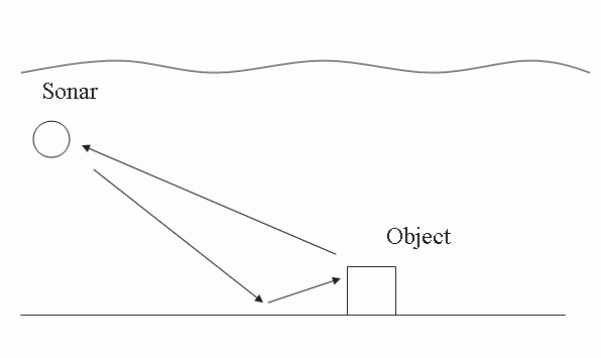


Figure 2.6: Bottom - direct path

reflected in the specular direction, meaning that TS is zero for that direction. As discussed in the previous section, the surface roughness also affects the reflected intensity. For MLO:s this might reveal a difference between man-made and natural objects. A rough object like a rock will scatter the sound more than a man-made smoother object with the same geometrical shape [3].

2.2.5 Volume reverberation

Since both density and sound speed varies in sea, this rise refraction in the volume. Marine life also causes scattering [4].

2.2.6 Transmission loss

Transmission loss from a source to an arbitrary point is defined as the ratio between the sound intensity in that point, I_1, and intensity in a reference point,

I_0, at a given distance (normally 1 meter) from the acoustic center of the source.

$$TL = 10log_{10}(\frac{I_0}{I_1}) \text{ dB} \tag{2.3}$$

Transmission loss is important to consider in detection of MLOs but also in classification because firstly absolute echo strength can be a useful classification feature, so it is important to know how to compensate correctly for attenuation versus range. Secondly, absorption losses lead to relatively higher attenuation of the high frequencies, which leads to an effective loss of bandwidth, and deteriorating range resolution with increasing range.
The main contributions to transmission loss are absorption and spreading losses, both increasing with target range. There are some other losses normally classed as transmission loss which are independent of range, for example attenuation due to bubbles in the water column. It is worth noting that transmission losses are two-way for an active sonar but not for a passive one [3].

Spreading loss

Assuming a homogenous and lossless medium, the power is the same in every point in the medium, $P_0 = P_1$. If the medium also is unbounded the spreading is spherical. Since power is given by area times intensity, transmission loss due to spreading becomes

$$\begin{aligned} P = 4\pi r^2 I \Longrightarrow \{P_0 = P_1\} \Longrightarrow 4\pi r_0^2 I_0 = 4\pi r_1^2 I_1 \Longrightarrow \\ \frac{I_0}{I_1} = \frac{r_0^2}{r_0^2} \Longrightarrow \{r_0 = 1\} \Longrightarrow \\ TL_{spreading} = 10log_{10}(r_1^2) = 20log_{10}(r_1) \text{ dB} \end{aligned} \tag{2.4}$$

Spherical spreading is a simplification. When an upper and lower surface boundary is present the spreading is rather cylindrical, and given by

$$\begin{aligned} P = 2\pi r H I \Longrightarrow \{P_0 = P_1\} \Longrightarrow 2\pi r_0 H I_0 = 2\pi r_1 H I_1 \Longrightarrow \\ \frac{I_0}{I_1} = \frac{r_0}{r_0} \Longrightarrow \{r_0 = 1\} \Longrightarrow TL_{spreading} = 10log_{10}(r_1) \text{ dB} \end{aligned} \tag{2.5}$$

Neither of these models is perfectly true. The true value probably lies somewhere between them, so sometimes $TL = 15log_{10}(r_1)\ dB$ is used. However both models shows that spreading loss increases with range [3].

Absorption

Absorption losses are due to conversion from acoustic energy to heat. The absorption loss for a plane wave with intensity I travelling a small additional distance dx, through an absorptive medium is given by Eq. 2.6.

$$dI/I = -ndx \tag{2.6}$$

Integrating between two ranges r_0 and r_1, and setting

$$\alpha = 10n \log_{10}(e) \tag{2.7}$$

gives

$$\alpha = \frac{10 \log I_0 - 10 \log I_1}{r_1 - r_0} \text{ dB/km} \tag{2.8}$$

α is the absorption coefficient and

$$TL_{absorption} = 0.01 \alpha r \tag{2.9}$$

Absorption in seawater is caused mainly by viscosity and ionic relaxation, and is frequency dependant. Different models exist for this frequency dependence, but in most models α increases with f^2 [4].

2.2.7 Ambient noise

The environment give rise to acoustic noise. The noise depends on frequency and seastate. One frequently used model for the different seastates is given by the Knudsen Spectra in Fig. 2.7 [4].

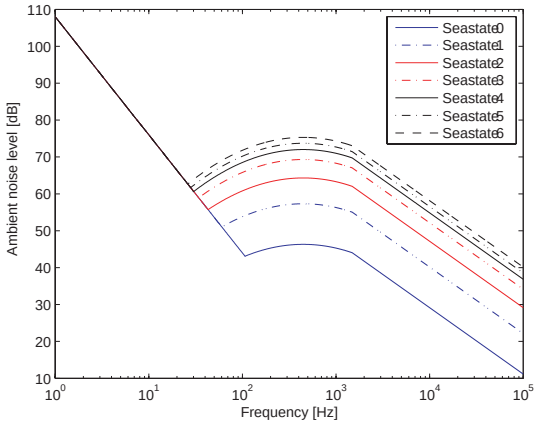


Figure 2.7: Knudsen spectra

2.3 Sidescan sonar

Sidescan sonar is the traditional instrument for underwater minehunting. A sidescan sonar is mounted on a towfish, surface vessel, or more recently on

an UUV (unmanned underwater vehicle). The insonified sector of the seabed, the footprint, is perpendicular to the track of the vehicle. The geometry of a sidescan system is shown in Fig. 2.8.

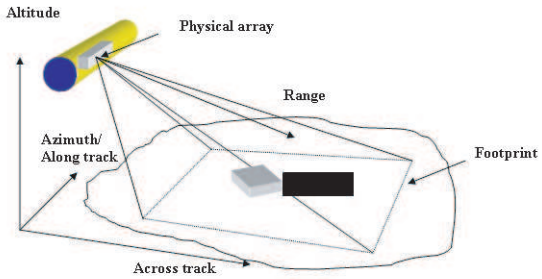


Figure 2.8: Geometry for a sidescan system

In sidescan systems the beam usually points in a fixed direction with respect to the platform. In more advanced systems it can be steered to compensate for heading variation.

With sidescan sonar (but not with SAS) there is beamforming [6], summing signals from different array elements both on transmission and reception, so pulses are conceptually transmitted from the centre of the array and received at the centre of the same array. If beam steering is used, the signals are delayed in time depending on azimuth angle. Beamforming gives a signal of energy, which can be converted to a greyscale value, versus time. A sidescan sonar image is then built up line by line.

The energy versus time signal, or energy versus range, gives an ambiguity - the hyperbola in Fig 2.9 showing positions at the same range that cannot be distinguished. As can be seen, the ambiguity width is larger at long than at short range. In order to cover the seabed continuously at close range, the distance travelled by the platform between consecutive pings must be short enough, see Fig 2.10. Hence the footprints for pings at long ranges will overlap. [3].

Due to this point-spread effect, sidescan images have the typical smoothed-out appearance visible in Fig. 2.11. The positive aspect is that speckle, random noise described in Section 3.4, is also smoothed out.

The down-range resolution of the resulting image depends on the bandwidth of the sound pulse, while the along-track (azimuth) resolution depends on the beam-width of the transmitter and receiver array, as given by Eq. 2.10. [7]

$$\delta = \frac{rc}{Df_c} = \frac{r\lambda_c}{D} \tag{2.10}$$

This equation shows that to increase azimuth resolution there are two alternatives, increasing the aperture length, D or reducing the wavelength λ_c. The

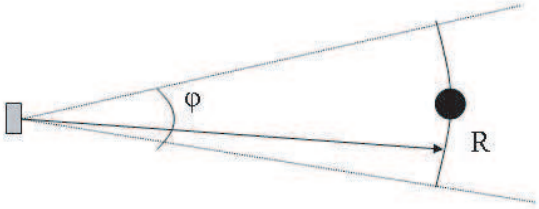


Figure 2.9: A single ping

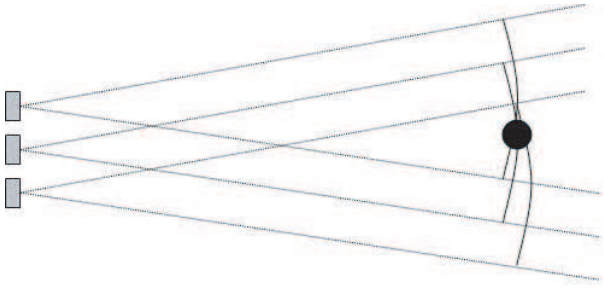


Figure 2.10: Overlapping pings

aperture length, D, is usually limited by the size of the vehicle. Wavelength is inversely proportional to frequency, so reducing wavelength implies increasing frequency and hence attenuation in the water column, reducing the maximum useful range of the system. This trade-off is important for mine-hunting systems, since good spatial resolution is needed to detect and classify objects, while long range is needed to search large areas quickly.

2.4 Synthetic aperture sonar

Synthetic aperture sonar (SAS) presents a method of achieving high spatial resolution without increasing frequency and sacrificing range. SAS is an adaptation of synthetic aperture radar (SAR) which is a mature technique used routinely in satellite monitoring of the earths surface [8]. However for sonar, it is a relatively new technique. The main principle is that a synthetic array is constructed from successive locations of the physical array - see Fig. 2.12. The length of the synthetic aperture is limited by the number of pings which insonify the same point on the seabed. Synthesis is achieved by coherent addition of their echoes, meaning that the phase of the echoes with respect to each transmitted pulse must be preserved.

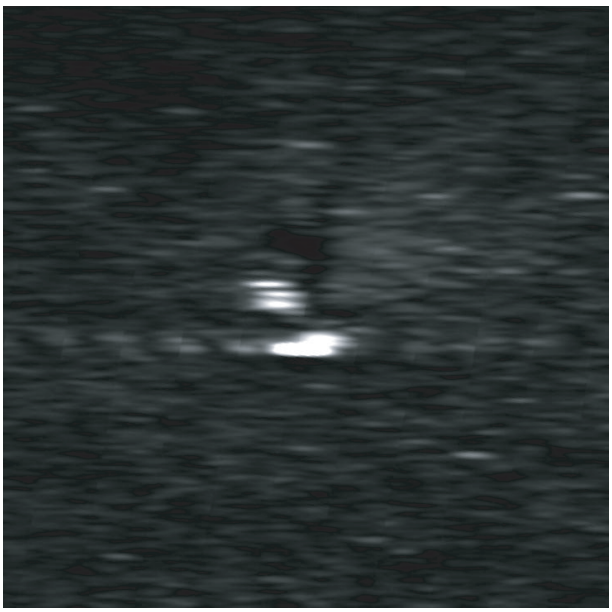


Figure 2.11: Sidescan image of an anchor

In the ideal situation where the sonar platform moves along a perfectly straight track without change of heading, this coherent addition presents little problem. However in the real situation where the track not only deviates from a straight line, but the deviation is not accurately known, this is a hard task. Motion errors arise because of water movement and the hydrodynamics of the vehicle. Image degradation occurs as soon as these motion errors exceed $\lambda/8$, where λ is the mid-wavelength. This limit equals 1.875 mm for a 100 kHz sonar [7].
These motion errors must be estimated and compensated to achieve coherent addition of successive pings. Motion errors also arise in SAR, and are compensated using the radar data itself, by methods termed "autofocus" [9]. However the problem is different with SAS because the motion error can change randomly from one platform position to the next. The correction problem is more difficult, which is one of the main reasons why it has taken a long time for the underwater community to accept the SAS technique. With SAS, methods of correcting navigation from the sonar data are termed "autopositioning" [10] or "micronavigation" [11]. The resolution gained by using SAS can be seen by comparing Fig. 2.11 to Fig. 2.13, the previous is a sidescan image of an anchor and the latter is the same anchor using synthetic aperture sonar.

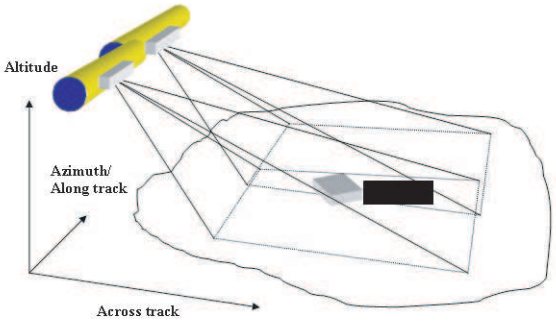


Figure 2.12: Geometry for a SAS system

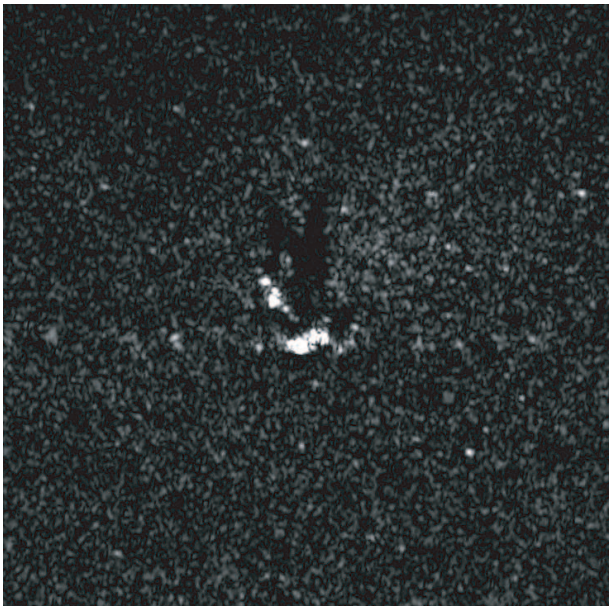


Figure 2.13: SAS image of an anchor

Chapter 3

Image regions

3.1 Introduction

One can divide a sonar image into three different regions, assuming one or more objects is present, highlights from object(s), shadow region(s) and background (i.e bottom).

3.2 Highlight

Objects protruding from the seabed return a strong echo towards the sonar. The reason for this is that there are more reflectors normal to the direction of sound propagation. Only the front surface border, the blue line in Fig. 3.1, of the object contributes to the highlight since the rest of the object is concealed behind this surface. The red surface will not affect the appearance of the highlight, unless it is insonified by a different ping. This means that object size in the across-track direction cannot be determined from the highlight.

As discussed in Section, 2.2.1 surface roughness might be a feature discriminating man-made objects, often with a smooth surface, from natural objects with somewhat rougher surfaces. In SAS images, this might be seen in the highlight region, where the texture will differ between surfaces with different roughness.

3.3 Shadow

The shadow area is the part of the footprint that is behind an object. The geometrical reason for the shadow can be seen in a 2D side view in Fig 3.2 where $d - a$ is the swath width, $\max(range) - \min(range)$ in across track direction, of the footprint. The object prevents the sound from insonifying the area behind it, represented in the figure by the line $\overrightarrow{bc}$. In the ideal situation

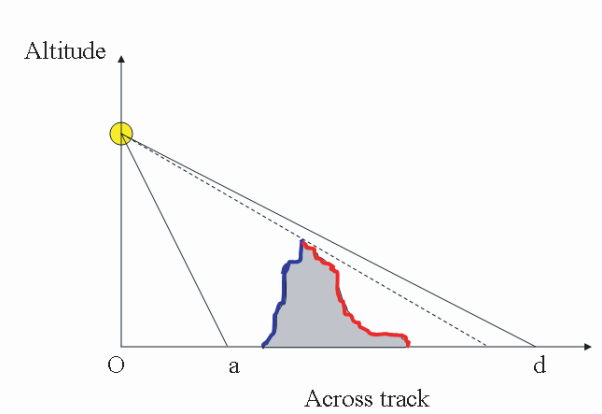


Figure 3.1: 2D side view of highlight principle

the shadow area will be totally dark. No echo can originate from this area since no sound has excited it.

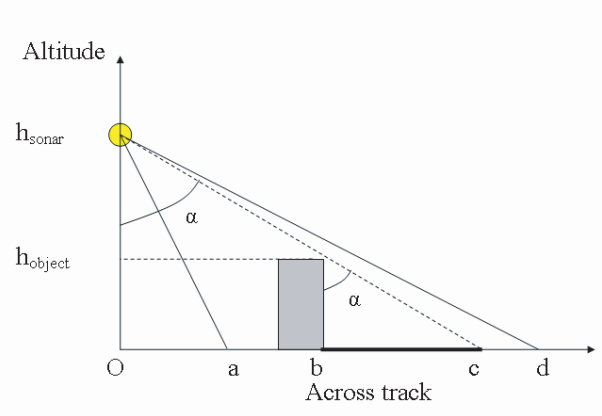


Figure 3.2: 2D side view of sonar system principle

There are different reasons why the ideal situation is not realized, giving non-zero pixels in the shadow area, for example electrical noise in the sensors and environmental noise, Section 2.2.7. Another important reason is multipath propagation, Section 2.2.3. There is an additional reason with synthetic aperture sonars due to the fact that the shadow area rotates round the object as the platform moves. The geometry changes from ping to ping and the shadow area will be slightly different, as can be seen in Fig. 3.3, eroding the edge of the shadow area.

The across-track length of the shadow, $c-b$ in Fig. 3.2 is given by equations

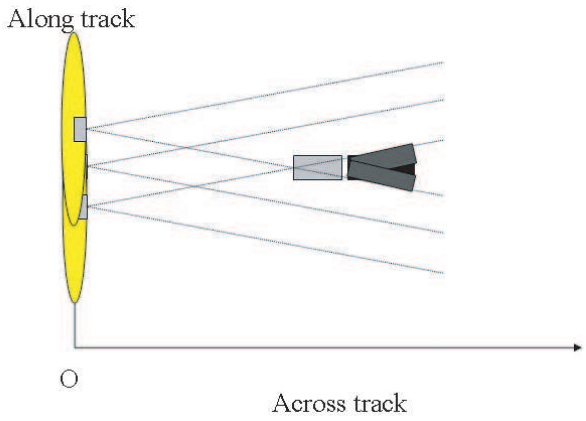


Figure 3.3: Schematic top view of SAS:s influence of shadow region

3.1 and 3.2

$$c - b = h_{object} \times \tan(\alpha) \tag{3.1}$$

$$\tan(\alpha) = \frac{c}{h_{sonar}} \tag{3.2}$$

Equation 3.1 and 3.2 show that, at a fixed range, shadow length in the across-track direction increases with decrease in sonar altitude. For an object that is not rotationally symmetric around the altitude axis, the shape of the shadow also changes with insonification angle.

3.4 Background

The background is defined as the region which is neither classified as highlight or shadow, i.e. the sea-bottom. As discussed in Section 2.2.2 the seabed varies considerably in terms of absorption and scattering depending on material and type. Rocks and other particles smaller than the sound wavelength, randomly distributed on the seabed, scatter the sound in different directions. Constructive and destructive interference of sound waves scattered from these particles give the background a granular look with relative bright pixels mixed with darker ones (termed speckle) [12].
The theoretical distribution of the absolute echo strength for random scatterers returning normally-distributed complex echoes is a Rayleigh distribution, and is the simplest model used for background echoes [13]. Rayleigh is a special case of the Weibull distribution(explained in Section 6.2.6). Fig.3.4 shows the histogram of the background region and the best fit of a Rayleigh and a general Weibull distribution for Test Image 1, Fig. A.1.
One thing that might complicate segmentation into highlight and shadow regions is when bottom structures, for example sand dunes, cast shadows. The difference

between shadows from sand-dunes and shadows from MLO:s is that shadows from MLO:s tend to be extended in the across-track dimension while sand dune shadows are extended along track.

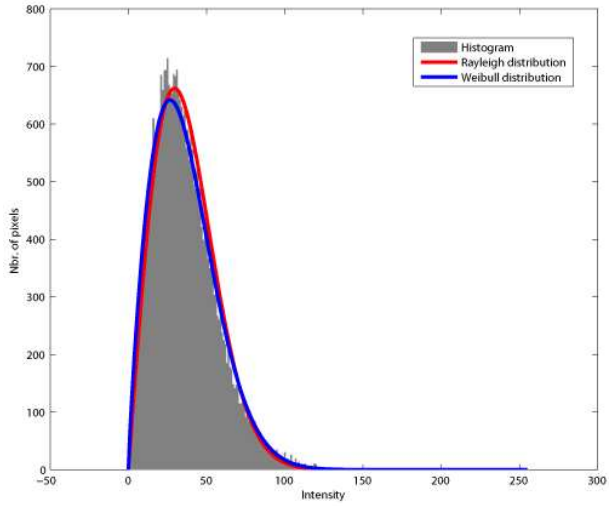


Figure 3.4: Histogram of background pixels for Test Image 1, Fig. A.1, together with Rayleigh and Weibull estimation

Chapter 4

Image statistics

4.1 Introduction

Local statistics are widely used in CAD/CAC approaches to sonar images, common measurements being local mean and local standard deviation, for example [14] and [15] for segmentation purposes. Other approaches use higher order statistics for example in [16] for detection and in [17] for classification. In this thesis local mean, local standard deviation and normalized local standard deviation are used for segmentation, Chapter 5, and classification purposes, Chapter 7.

4.2 Calculation

For a set of N random variables x_n variance, σ^2 is given by Eq. 4.1.

$$\sigma^2 = \frac{(\sum(x_n - m)^2)}{N} \tag{4.1}$$

where m is the local mean,

$$m = \frac{\sum x_n}{N} \tag{4.2}$$

Local standard deviation, σ, is given by Eq. 4.1 and 4.2,

$$\sigma^2 = \frac{(\sum(x_n - m)^2)}{N} = \frac{(\sum(x_n^2 - 2x_n m + m^2)}{N} \Longrightarrow \{Eq.4.2\}$$
$$\Longrightarrow N\sigma^2 = \sum x_n^2 - \frac{\sum x_n^2}{N} = \{m^2 = \frac{\sum x_n^2}{N^2}\} = \sum x_n^2 - Nm^2 \Longrightarrow$$
$$\sigma = \sqrt{\frac{\sum x_n^2 - Nm^2}{N}} \tag{4.3}$$

Standard deviation normalized with local mean, Eq. 4.4 is also used here.

$$NormStd = \frac{\sigma}{m} \tag{4.4}$$

4.3 Window size

The size of the window in which the local statistic is calculated is an important factor. The reason is that if too large a window is used, there will be no position of the window where all pixels are highlights or shadow. Therefore the background will affect the local statistics for all pixels. This is demonstrated by Fig. 4.1 and Fig. 4.2 where local statistics for a white object (size $5x5$) on a random background are calculated with a window smaller, (size $3x3$), than the object, Fig. 4.1, and a window larger, (size $7x7$), than the object, Fig. 4.2. The mean values are not significantly different, the main difference being that the peak-value is higher when the window smaller than the object. For standard deviation and normalized standard deviation there are more significant differences. Since the regions in a real SAS image are never all white or all black, the effect of different window size will not be as large as in this simple example, but it is still important to choose an appropriate window size for the region of interest.

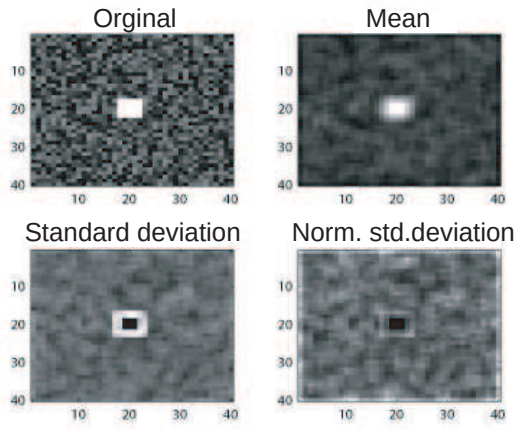


Figure 4.1: Local statistics for a white object on a noisy background, window size $3x3$

4.4 Local statistics of SAS-images

It is discussed in [14] and in the attached references that highlight pixels have high local mean but standard deviation with varying characteristics, while shadow pixels have low local mean and standard deviation. Local statistics for background pixels depend on the complexity of the seabed, in terms of structures

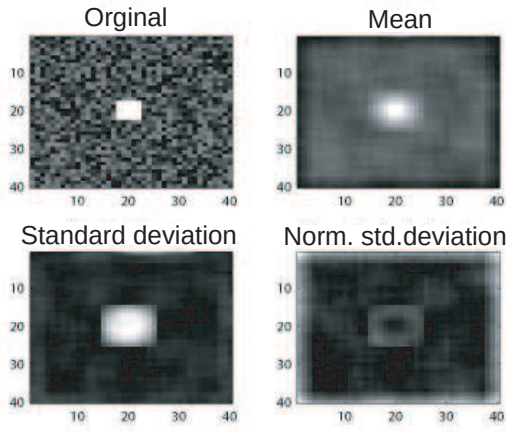


Figure 4.2: Local statistics for a white object on a black background, window size $7x7$

visible in the sonar image. Fig. 4.3 shows the standard deviation versus mean plane, using a $5x5$ window, for Test Image 3 Fig. A.3 after manual segmentation. Here can some of the statistical properties discussed above be seen, most of the highlight pixels, red dots in the figure, having high local mean and high local standard deviation while most pixels in the shadow region, blue dots in the figure, have low local mean and standard deviation.

The normalized standard deviation, Eq. 4.4, shows the relationship between local mean and local standard deviation. As can be seen in Fig. 4.4 this measure might be used as an edge detector. The figure shows normalized standard deviation in a $5x5$ neighbourhood for all pixels in Test image 3, Fig. A.3. This measure is also discussed as a classification feature in Chapter 7.

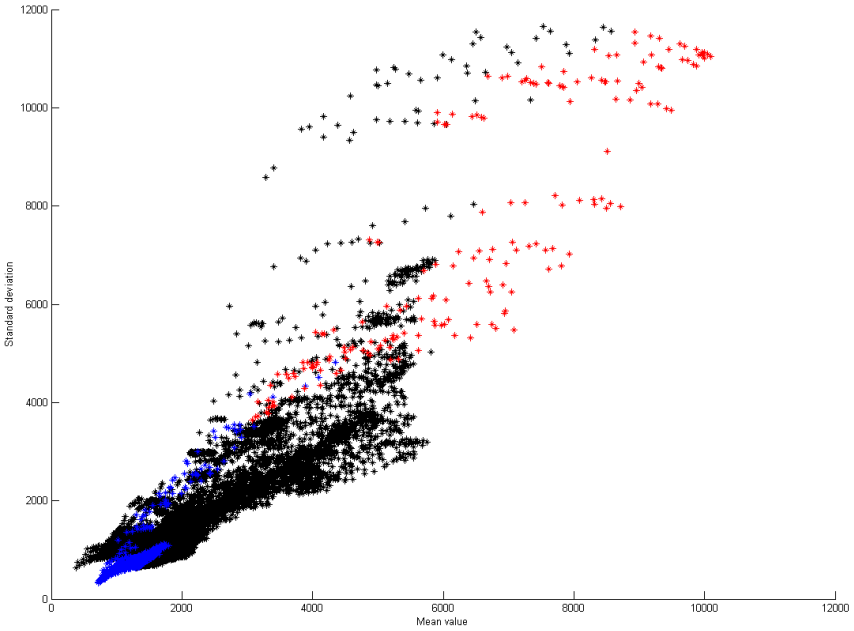


Figure 4.3: Standard deviation versus mean plane for Test Image 3 Fig. A.3. Red points are for highlight pixels, blue for shadow and black for background.

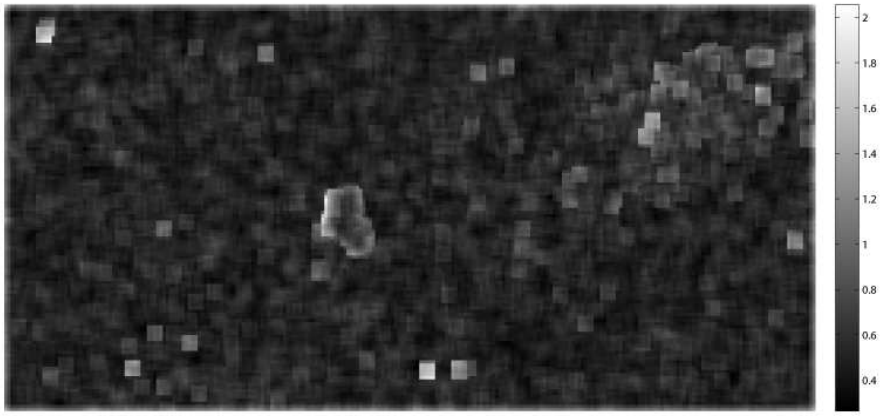


Figure 4.4: Normalized standard deviation for Test Image 3 Fig. A.3

Chapter 5

Segmentation using a region growing approach

5.1 Introduction

For an image with a flat background it might be possible to segment the interesting regions only by thresholding, but for a more complex image a more sophisticated method is needed. The approach considered here is based on the local statistics properties discussed in Chapter 4 and region growing, a basic image segmentation algorithm [18]. A short theoretical introduction is given below. For a deeper understanding an image analysis book, for example [18], is recommend. Hereafter follows a detailed description of the different steps in the algorithm. Segmentation results for all Test Images, using the same parameter configuration, can be found in Appendix A and discussions and comparison between manual segmentation and computer segmentation can be found in Chapter 8.

5.2 Theory

5.2.1 Definitions

In digital image systems a square lattice of sites is used and therefore some problems with distance might appear. A diagonal movement in the grid, a Euclidean distance equal to $\sqrt{2}$, is often considered as the same distance as a movement up and down or to either side. When this is the case the D_8- or chessboard, definition 1, distance is used.

Definition 1 $D_8[(i,j),(h,k)] = max(|i-h|,|j-k|)$

Two pixels X(a,b) and X(m,n) are neighbours if their distance apart is equal to 1. 8-neighbours are defined in definition 2. Definition 3 defines a region.

Definition 2 *Two pixels $P_0 = X(a,b)$ and $P_1 = X(m,n)$ are 8-neighbours if $D_8[P_0, P_1] = 1, P_0 \neq P_1$*

Definition 3 *A set of pixels where there exists an 8-neighbour path between every pair of pixels is called a region.*

In region-growing algorithms, two more assumptions about regions are made. There exists a binary homogeneity evaluation function H on the S regions R_i

$$H(R_i) = true,\ i = 1, 2, \ldots, S \tag{5.1}$$

$$H(R_i \cup R_j) = false,\ i \neq j\ R_i\ adjacent\ to\ R_j \tag{5.2}$$

Eq. 5.1 tells that a region is homogeneous, according to some measure H. 5.2 state that region segmentation is maximal. If two adjacent regions are merged the resulting region is no longer homogeneous [18].

5.2.2 Region growing

Region growing is initiated with the image segmented into small regions, possibly consisting of single pixels, satisfying Eq. 5.1. Adjacent regions are merged in turn if a given criterion is satisfied. When no more regions can be merged, the segmentation is maximal based on the homogeneity evaluation function and merging criterion [18].

5.3 Segmentation of sonar images

5.3.1 General idea

The idea in this segmentation approach is to use the local statistics discussed in Chapter 4. The mean versus standard deviation plane (Fig. 4.3) shows that some highlight pixels can be found by simple thresholding, while the remaining highlight pixels and the shadow pixels might be hard to segment using thresholding alone. To solve this problem, an approach is used where first the highlight is segmented, and then knowledge of the shadow position, see Section 3.3, is used together with local statistics. This approach might give some false regions which can be removed by using *a priori* knowledge:

- A true highlight region from an interesting object lies within a certain size bracket
- A true shadow region from an interesting object has approximately the same along- track dimensions as the corresponding highlight
- A true shadow region has across-track dimensions larger than a certain value
- A true highlight region is followed by a true shadow region.

5.3.2 Highlight

Initiation

Examination of the standard deviation versus mean plane for all test images shows that pixels with the highest local mean always lie within the highlight. Hence highlight seed pixels are found by choosing a threshold for local mean, for example the red line in Fig. 5.1.

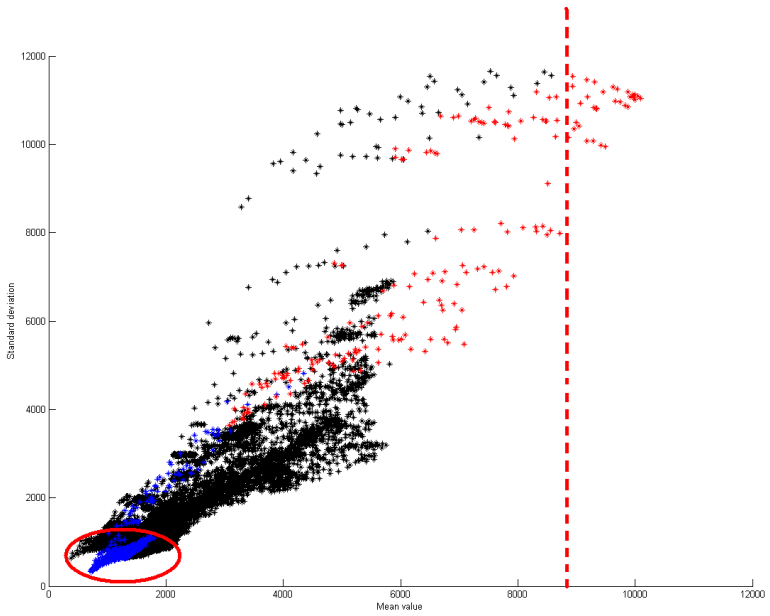


Figure 5.1: Standard deviation versus mean plane for Test Image 3 Fig A.3. The red line marks the threshold for finding seed pixels. The red ellipse marks pixels with typical shadow characteristics, i.e. low mean-value and standard deviation

The threshold, T_1 is set by a linear combination, Eq. 5.3, of the mean value for all pixels in the original image, X, size MxN, and the maximum value in the local mean image, X_{MEAN}. With a structured background, for example where small rocks are scattered on the seabed, background pixels might still be falsely initiated as highlight pixels. This risk is addressed later in processing.

$$T_1 = (1 - A_1) \times \max(X_{MEAN}) + A_1 \times (\frac{\sum_{m \in M} \sum_{n \in N} (x(m,n))}{MN}) \qquad (5.3)$$

Fig. 5.2 shows segmentation step by step for Test Image 4, Fig. A.4. Initiation can be seen in Fig. 5.2(a).

Region growing

In the next step, neighbouring pixels are incorporated in the region if their greyscale value is higher than T_2, a new lower threshold, given by a linear combination, Eq. 5.4, of the average of already classified highlight pixels and the average of all pixel values in the image.

$$T_2 = (1 - A_2) \times (\frac{\sum_{m \in M} \sum_{n \in N} (x(m,n) \subset HIGHLIGHT)}{N_{HIGHLIGHT}}) + A_2 \times (\frac{\sum_{m \in M} \sum_{n \in N} (x(m,n))}{MN}) \quad (5.4)$$

where $N_{HIGHLIGHT}$ is the number of highlight pixels in the image. The result for Test Image 4, Fig. A.4, can be seen in Fig. 5.2(b).

Final highlight regions

When no more pixels can be incorporated, holes and pixels not already classified as highlight but surrounded only by highlight pixels are filled. Regions failing the first *a priori* criterion are then removed. The final segmentation for Test Image 4, Fig. A.4, is shown in Fig.5.2(c).

5.3.3 Shadow

Initiation

Seed pixels for the shadow regions, one for each highlight region, are found by defining a window, SW, in the area behind the highlight in which the seed pixels are found. These are pixels where the local mean and local standard deviation are below the thresholds T_3,T_4 (Eq.5.5 and Eq. 5.6) respectively. The size of SW is set by the along-track size of the highlight region and the minimum expected width of the shadow, across-track. The shadow-segmentation steps can be seen in Fig. 5.2(d) - 5.2(g)

$$T_3 = (1-A_3) \times \min(X_{MEAN}(SW)) + A_3 \times (\frac{\sum_{m \in M_{SW}} \sum_{n \in N_{SW}} (X_{MEAN}(m,n))}{M_{SW} N_{SW}}) \quad (5.5)$$

$$T_4 = (1-A_4) \times \min(X_{SD}(SW)) + A_4 \times (\frac{\sum_{m \in M_{SW}} \sum_{n \in N_{SW}} (X_{SD}(m,n))}{M_{SW} N_{SW}}) \quad (5.6)$$

Region growing

For the shadow-growing step, two more thresholds T_5 and T_6 are defined, again using linear combination (Eq. 5.7 and 5.8). These relax the requirements for a pixel neighbouring a shadow pixel to be classified as shadow. Growing of the

shadow is accepted if a neighbouring pixel either has local standard deviation below T_5 or local mean below T_6. Growth outside the previous defined window is now permitted. However to fulfil the second *a priori* criterion, growth beyond the along-track dimension of the highlight is forbidden.

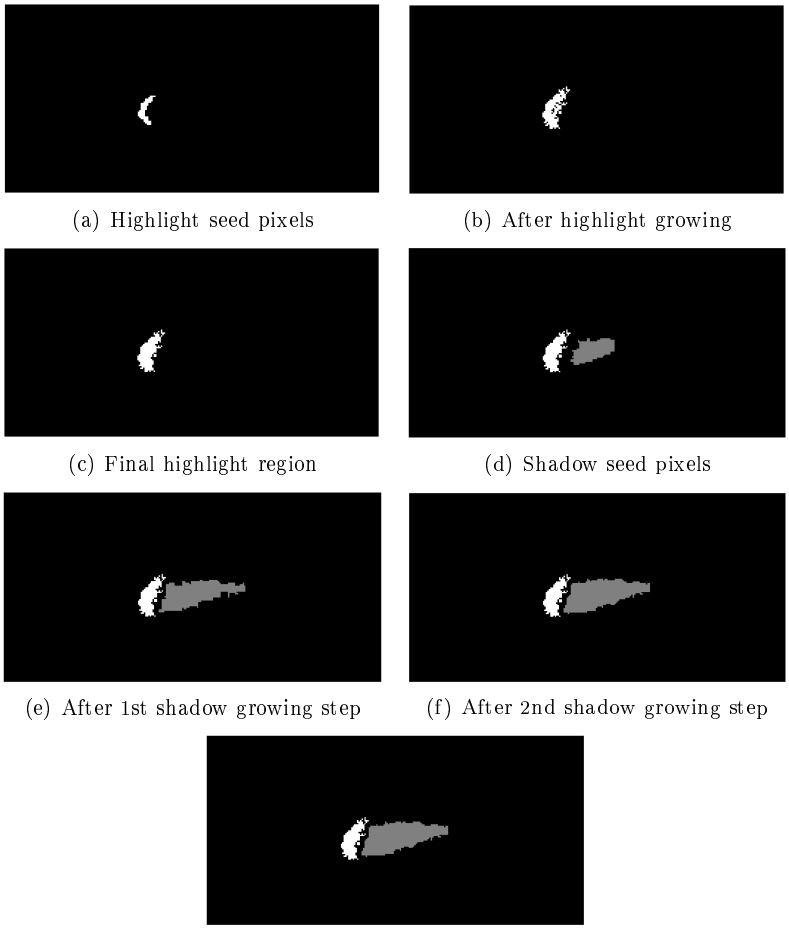
$$T_5 = (1 - A_5) \times \frac{\sum_{m \in M} \sum n \in NX_{SD}(m,n) \subset SHADOW}{N_{SHADOW}} + A_5 \times (\frac{\sum_{m \in M} \sum n \in NX_{SD}(m,n)}{MN}) \quad (5.7)$$

$$T_6 = (1 - A_6) \times \frac{\sum_{m \in M} \sum n \in NX_{MEAN}(m,n) \subset SHADOW}{N_{SHADOW}} + A_6 \times (\frac{\sum_{m \in M} \sum n \in NX_{MEAN}(m,n)}{MN}) \quad (5.8)$$

Where N_{SHADOW} is the number of shadow pixels in the image.

Final segmentation

After shadow growth, holes are filled in the same way as for the highlight regions. Finally, regions that don't fulfil all *a priori* criteria are removed.



(a) Highlight seed pixels

(b) After highlight growing

(c) Final highlight region

(d) Shadow seed pixels

(e) After 1st shadow growing step

(f) After 2nd shadow growing step

(g) Final segmentation

Figure 5.2: Segmentation of Test Image 4, Fig. A.4

Chapter 6

Segmentation using a Markov random fields approach

6.1 Introduction

The second segmentation approach is an adaptation of the scheme proposed in [19] and [20] for segmentation of conventional sidescan images. In the references, MRF (Markov random fields) theory is used to achieve an MAP (maximum *a posterior*) segmentation of images heavily corrupted by noise. The algorithm is divided into two steps, firstly the separation of shadow from not-shadow regions in the image, followed by separation of the not-shadow regions into highlight and background. All required parameters are estimated from the image itself, which makes it a promising approach for segmenting SAS images. Unfortunately parameter estimation is computationally heavy and time consuming. However the algorithm can be implemented on parallel processing hardware [21] which will mitigate that problem in the longer term. Because of the long execution times, only the first step - shadow segmentation - has been implemented so far.

The underlying strategy is as follows. The observed image is regarded as a MRF corruption of the unknown segmented binary image. Then a sampling algorithm is used to obtain a large number of realisations of the binary image, allowing the most likely one to be selected.

6.2 Theory

6.2.1 Neighbourhood

The spatial relationship between a set of pixels, S, is defined via a neighbourhood system,η

Definition 4 $\eta = \{\eta_i \,|\, \forall i \in S\}$

where η_i are the pixels neighbouring pixel i (see Definition 2, Chapter 5). A neighbourhood system satisfies:

- $i \notin \eta_i$
- $i \in \eta_{i'} \iff i' \in \eta_i$

The first relationship states that a pixel is not a neighbour to itself. The second states that the neighbourhood relationship is commutative. A clique is a subset of S containing only neighbouring pixels. A pair-site clique is defined as

Definition 5 $C_2 = \{\{i, i'\} \,|\, i' \in \eta_i,\, i \in S\}$

As shown in Fig. 6.1 there are four types of pair-site cliques determined by their orientation.

6.2.2 Markov random fields

A set of random variables $X = (X_s, s \in S)$ defined on the set, S, and taking values in L is defined to be a MRF if and only if

Definition 6

$$P(X) > 0 \; \forall x \in X \tag{6.1}$$

and

$$P(x_i | x_{S-\{i\}}) = P(x_i | x_{\eta_i}) \tag{6.2}$$

X_{S-i} denotes all members of S except i The second property, called Markovianity states that the probability of an event x_i for pixel i is determined only by the pixel's immediate neighbourhood.

6.2.3 Gibbs random fields

Definition 7 *A set of random variables X is a Gibbs random field (GRF) if and only if its configuration obeys a Gibbs distribution*

Definition 8 *A Gibbs distribution takes the form:*

$$P(X) = \frac{e^{-\frac{U(X)}{T}}}{\sum_{x \in X} e^{-\frac{U(X)}{T}}} \tag{6.3}$$

$$U(x) = \sum_{c \in C} V_c(x) \tag{6.4}$$

where T is a constant normally set to 1. V_c is the clique potential expressing the contrast between the elements. U(x) then expresses the contrast between pixel x and its neighbourhood. The expression used here will be given in Eq. 6.8

The Hammersley-Clifford Theorem [22] states that X is a MRF on S with respect to η if and only if X is a GRF on S with respect to η.

6.2.4 MAP segmentation

Considering the observed image as a noisy observation Y of the underlying random field, X, the *a posterior* probability P(X|Y) is given by Bayes Law [22], Eq. 6.5.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{6.5}$$

In our case, the underlying random field is the binary segmented image. Observation Y is available, so P(Y) is certain and can be ignored. The conditional probability P(Y|X) is given by Eq 6.6 [19].

$$P(Y|X) = \prod_{s \in S} P(y_s|x_s) \tag{6.6}$$

where $P(y_s|x_s)$ is found by fitting a Weibull distribution (Section 3.4 and 6.2.6) to the greyscale probability density function of Y for the two possible values, $x_s = 0$ and $x_s = 1$ in X. By modelling X as a MRF and using the Hammersly-Clifford Theorem, the probability $P(x_i)$ is shown [22] to be:

$$P(x_s|x_{S-\{s\}}) = P(x_s|\eta_s) = \frac{e^{-\sum_{c \in \eta_s} V_c(x)}}{\sum_L e^{-\sum_{c \in \eta_s} V_c(x)}} \tag{6.7}$$

The local energy, $\sum_{c \in \eta_s} V_c(x)$, the sum of clique potentials, is defined as [23]:

$$\sum_{c \in \eta_s} V_c(x) = \Theta(x_s, \eta)\phi_x^T \tag{6.8}$$

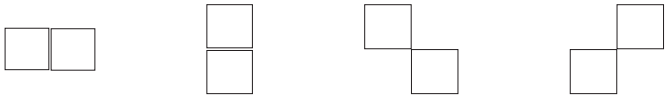
where $\phi_x = \beta_1, \beta_2, \beta_3, \beta_4$ are *a priori* parameters connected to the different cliques in Fig. 6.1. $\Theta(x_s, \eta)$ is defined for the neighbourhood shown in Fig. 6.2 as

$$\begin{aligned} \Theta(x_s, \eta) = [&I(x_s, u_1) + I(x_s, u_3),\ I(x_s, u_2) + I(x_s, u_4), \\ &I(x_s, v_1) + I(x_s, v_3),\ I(x_s, v_2) + I(x_s, v_4)] \end{aligned} \tag{6.9}$$

where

$$I(z_1, z_2) = 0\ if\ z1 = z2,\ 1\ otherwise \tag{6.10}$$

This mean that I = 0 for neighbouring pixels if they lie within the same segment, but I = 1 if they lie on opposite sides of a segmentation boundary.



(a) β_1 (b) β_2 (c) β_3 (d) β_4

Figure 6.1: Cliques

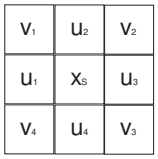


Figure 6.2: Neighbourhod

6.2.5 K-mean classification

In a K-mean classifier, every class is initiated as a seed point c_i in the N-dimensional space defined by the N features used. Every variable x_i defines a point p_i with coordinates given by the feature values, $p_i = (feature_1, feature_2, , \ldots, feature_N)$. Every variable is assigned to the class defined by the nearest point c_i , i.e. the point with the lowest Euclidian difference to p_i. When all variables have been classified, new seed points for the classes are given by the mean values of all points assigned to the class. This procedure continues until the seed points remain unchanged [23].

6.2.6 Weibull distribution

A regular Weibull distribution, W(α,C) is a 2-parameter distribution determined by a scale parameter, α, and a shape parameter, C. In [19] a third parameter, min, is introduced to shift the distribution:

$$W(y; min, C, \alpha) = \frac{C}{\alpha} \frac{(y - min)}{\alpha}^{C-1} e^{\frac{-(y-\min)^C}{\alpha^C}} \tag{6.11}$$

Fig. 6.3 show the effects of the different parameters. In Fig. 6.3(a) $\alpha = 40$, C = 2, and min is varied. In Fig. 6.3(b), min = 0, C = 2, and α is varied. In Fig. 6.3(c), min = 0, $\alpha = 40$, and C is varied. Fig. 6.3 also shows the flexibility of a Weibull distribution, as pointed out in [19].

When C=1, W follows an exponential law, while when C=2, W is a Rayleigh distribution. This is the theoretical distribution of the absolute echo strength for random scatterers returning normally-distributed complex echoes, and is the simplest model used for background echoes [13].

6.3 Segmentation of SAS images

6.3.1 Estimation of parameters

The clique parameters ϕ_x and Weibull distribution parameters $\phi_y = (\min, C, \alpha)$ are estimated by the ICE algorithm [19], whose structure is shown in Fig. 6.5. The algorithm is divided into three stages:

1. Initiation stage to find approximate parameters for the given image
2. Generation of sample images with these parameters, using a Gibbs sampler (Section 6.3.1)
3. New parameter estimation

The algorithm is iterative and terminates when the variance of the last few iterations falls below a certain value. This section will describe the particular implementation of the algorithm used in this research to segment SAS images.

Initiation

The initiation stage aims to obtain good initial values for the clique and distribution parameters. First, a rough segmentation is carried out, classifying every pixel by a K-mean procedure using the two features: local mean and local standard deviation (Section 4.2) plus a third feature: local minimum greyscale value. The window size used here is 5×5. Experiments in this thesis show that using three classes in the k-mean procedure, and then merging two of them to form the non-shadow class, gives better results then using only two. The classes are initiated by using *a priori* knowledge that shadow will have low feature values while non-shadow will have higher. Parameter estimation is carried out in the same way as in the rest of the algorithm described below. The maximum likelihood segmentation is achieved by for each pixel in the observed image, y, choosing the segmentation which maximises the conditional probability given by initiation values of ϕ_y, $P(y_s|\phi_y^0)$.

Gibbs sampler

The purpose of the Gibbs Sampler in the ICE algorithm is to generate binary images which obey the *a posterior* probability distribution $P(X|Y)$ given by current parameters. A binary image of size $M \times N$ has 2^{MN} possible configurations. Calculating the probability for every single one of these is impractical, so the Gibbs Sampler uses Eq. 6.5 - 6.8 to calculate the probability of an event pixel by pixel. One pixel of the image is changed in each step, and the probability becomes:

$$P(x_s|y_s) = P(y_s|x_s)\frac{e^{-\sum_{c\in\eta_s} V_c(x)}}{\sum_L e^{-\sum_{c\in\eta_s} V_c(x)}} \tag{6.12}$$

This operation is carried out for a number of raster scans of the whole image. According to [22] 50 scans suffice for most images.

Estimation of distribution parameters, ϕ_y

The Weibull distribution parameters (Eq. 6.11), $\phi_y = (\min, C, \alpha)$, are estimated by a maximum likelihood approach. The log-likelihood, $\ln P_{y|\phi_y}(y|\phi_y)$, for M independent samples is given by

$$\ln P_{y|\phi_y}(y|\phi_y) = \prod_{i=1}^{M} \frac{C}{\alpha} \frac{(y-\min)}{\alpha}^{C-1} \exp^{\frac{-(y-\min)^C}{\alpha^C}} \tag{6.13}$$

In [19] $\hat{\phi_y}$ is found by zeroing the partial derivatives of $\ln(P_{y|\phi_y}(y|\phi_y)$ with respect to each parameter. The parameter, min, is estimated as

$$\hat{\min_{ML}} = y_{min} - 1 \tag{6.14}$$

where y_{min} is the minimum value of y. After $\hat{\min}_{ML}$ is found, the distributions are shifted by setting $\tilde{y_i} = y_i - \hat{\min}_{ML}$ Then α is estimated from

$$\hat{\alpha_{ML}} = (\frac{1}{M} \sum (\tilde{y_i}^{C_{ML}})^{\frac{1}{C_{ML}}} \tag{6.15}$$

However Eq. 6.15 involves the parameter, C, still unknown. Hence 6.15 is combined with the result of zeroing the partial derivative of Eq. 6.13 with respect to C to obtain

$$\frac{\sum_{i=1}^{M} (\tilde{y_i}^{\hat{C_{ML}}} \ln \tilde{y_i})}{\sum_{i=1}^{M} \tilde{y_i}^{\hat{C_{ML}}}} - \frac{1}{M} \sum_{i=1}^{M} \ln \tilde{y_i} = \frac{1}{\hat{C_{ML}}} \tag{6.16}$$

Eq. 6.16 is solved iteratively.

Estimation of Markov parameters , ϕ_x

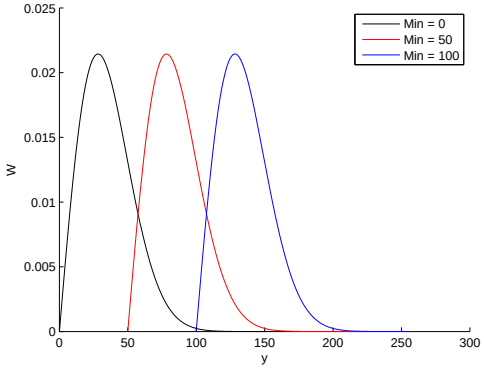
The MRF clique parameters, $\phi_x = \beta_1, \beta_2, \beta_3, \beta_4$, are estimated following a method detailed in [23]. This method has shown good test results in the references. Combining Eq. 6.7 and 6.8, and calculating the conditional probabilities for $x_s = 0$ and $x_s = 1$ gives the equation:

$$[\Theta(x_s = 0, \eta_s) - \Theta(x_s = 1, \eta_s)]^T \phi_x = ln(\frac{P(x_s = 0)|\eta_s}{P(x_s = 1)|\eta_s}) \tag{6.17}$$

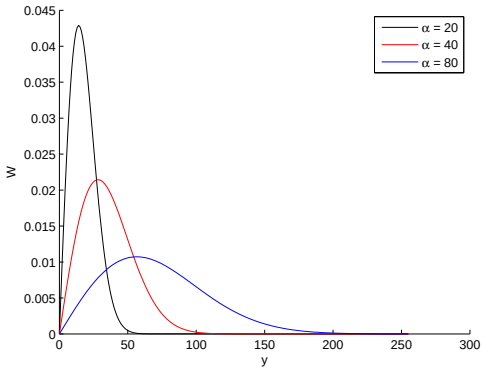
The ratio on the right hand side of this equation is calculated by counting the number of pixels in the image taking the values 0 and 1, and with a given neighbourhood, η_s. Since there are 8 neighbouring pixels taking two alternative values, there are 256 possible neighbourhood configurations. Hence the system of equation is overdetermined, and can be solved using the least square error method to obtain a solution for $\hat{\phi_s}$. $\Theta(x_s = 0, \eta_s)$ and $\Theta(x_s = 1, \eta_s)$ do not change, so need only be calculated once.

6.3.2 MAP segmentation

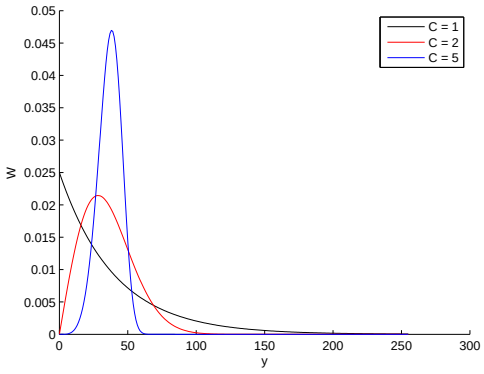
When the required parameters have been estimated, the MAP segmentation is achieved using one of the maximization methods discussed in [13]. Here we have used the ICM, Iterated Conditional Modes, which seems one of the most straightforward methods. This method uses a greedy approach of maximizing the local probability given by Eq. 6.12 pixel by pixel, and iterating until it converges. More about this method, and other maximization methods can be found in [22].



(a) Varying parameter min



(b) Varying parameter α



(c) Varying parameter C

Figure 6.3: Weibull distribution

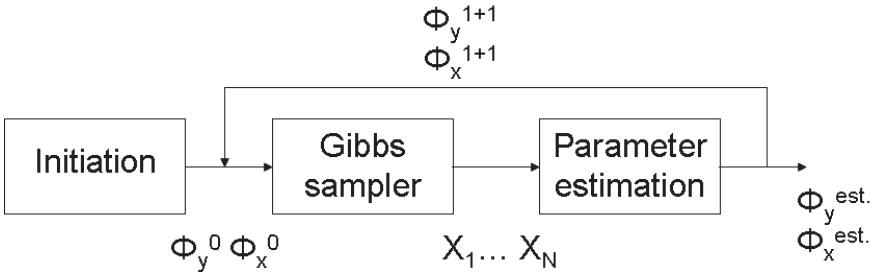


Figure 6.4: Schematic view of ICE algorithm

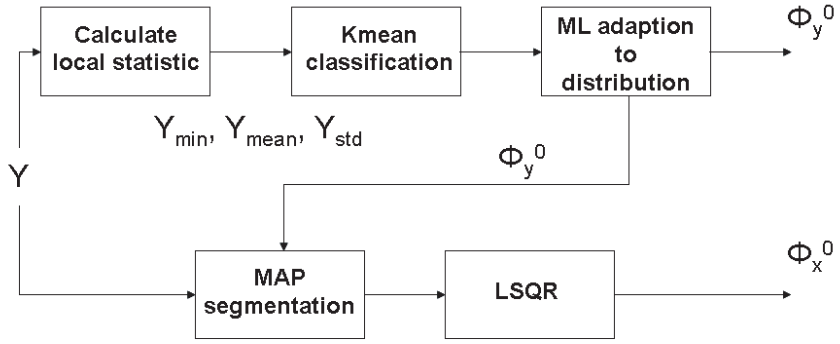


Figure 6.5: Schematic view of initiation part of ICE algorithm

Chapter 7

Classification features

7.1 Introduction

Discriminating man-made objects from natural ones is a hard and important task. To be useful the discrimination must be done with high confidence. Usually a number of different features are needed to achieve sufficient confidence. Alternative possible classification features are proposed here. These features are evaluated in Chapter 8.

7.2 Lines

One characteristic that might discriminate man-made from natural objects is the presence of straight edges, regularity and corners. In some situations, boxes, cylinders and other shapes with sharp edges will cast a shadow with straight edges. If the object has a regular shape, the shadow will have parallel lines. Sharp corners also imply a man-made object. For the shadow, only the region border is of interest since any non-black pixels are due to noise. In highlight regions, there might exist lines due to the object surface, so it could be interesting to look for edges in the greyscale image. However none of the test images has this characteristic, so this aspect is not considered here.

7.2.1 Hough transform

The standard Hough transform is designed to detect straight lines in images. Representing a line L(x,y) with a vector s where s is perpendicular to L and given by

$$s = x\cos\theta + y\sin\theta \tag{7.1}$$

In the s,θ-space, the line L is represented by a single point. Line detection by Hough Transform is based on this fact [18]. For every pixel in the region border, every possible line in all direction is transformed to the s,θ-space. Since every line is represented by a single point in s,θ-space, this point is incremented in the

transform space- the accumulation matrix H. Theoretically every edge-pixel has an infinite number of lines intersecting it. In practice, the resolution determines how many line directions are tested for each pixel. Finding the lines is done by thresholding the accumulation matrix and taking the inverse transform of the peaks.

7.2.2 Straight line detection

The Hough transform is robust and insensitive to noise since it can detect lines even when there is missing data, or an edge pixel in the wrong location due to segmentation errors. To be sure of finding well-defined lines it is important to choose an appropriate threshold, T. The proposition here is to determine this threshold by:

$$T = \gamma \times n, \quad n = number\ of\ border\ lines. \tag{7.2}$$

Where γ is a constant, $\gamma = 0.25$ was used for the results presented in next chapter. The results of line detection for Test Image 2 are shown in Fig. 7.1

(a) Highlight (b) Shadow

Figure 7.1: Result of line detection for Test Image 2

7.2.3 Parallel lines detection

Using the Hough transform, it is straightforward to determine whether detected lines are parallel by looking at the θ value.

$$\theta_j = \theta_i, i \neq j \Longrightarrow Line\ i\ and\ j\ are\ parallel \tag{7.3}$$

Due to noise, the equality should be replaced by approximate equality, $\pm 5^o$ was used in the tests.

7.2.4 Right-angle corner detection

Similarly, detection of right-angled corners is straightforward, by again looking at θ value for different lines.

$$|\theta_j - \theta_i| = 90^o \Longrightarrow\ Line\ i\ and\ j\ form\ a\ right-angle\ corner \tag{7.4}$$

The same inequality interval as above is used.

7.3 Size

Object size is of great interest in classification and identification systems. Too large or too small an object is unlikely to be a mine and can be ignored, or at least given less priority in the mine countermeasure part of the system. Since the exact size of many mines is known, an accurate size can give a good indication of what type of mine it might be.

7.3.1 Height

The height of an object, h, is given by Eq. 7.5, see Section 3.3 for visualization (Fig. 3.2) and derivation (Eq.3.1 and Eq. 3.2).

$$h = \frac{ShadowLength \times Altitude}{ShadowRange} \text{ [m]} \tag{7.5}$$

where ShadowLength is the length of the shadow, Altitude is the height of the sonar platform above the seabed at target range. ShadowRange is the maximum swath distance between the sonar and the shadow. Because of the erosion of shadows, see Section 3.3, the height can only be an approximation to the true height.

7.3.2 Width

For an object parallel to the track of the vessel, the width of the object in the along-track dimension is given by Eq. 7.6

$$w = l\rho \text{ [m]} \tag{7.6}$$

where l is the along-track length of the highlight in pixels and ρ is the image resolution in $[\frac{m}{pixels}]$. If the object is not parallel to the track, l will not be equal to the length of the highlight region. To estimate the width of the object, by finding the largest elongation at any rotation, the shape is transformed using the Hotelling/Kauren-Loeve transform [24] before calculating the width.

7.4 Shape

Many mines have shapes reminiscent of basic geometrical shapes like spheres, cylinders, boxes and truncated cones. At some insonification angles, objects shaped like boxes and cylinders will cast rectangular shadows. A sphere lying on the bottom will cast an elliptical shadow. For truncated cones the shadow will look like a truncated triangle. For objects that are not rotationally symmetric, for example boxes and cylinders lying on their side, the shadow shape depends on the direction of insonification. A shadow shape like the above and/or a shape that is mirror-symmetric around some axis strongly indicates a man-made object.

7.4.1 Rectangularity

The rectangularity, R, of a region is given by comparing its area with a rectangle of the same length and width [18].

$$R = \frac{A}{ab} \tag{7.7}$$

where A is the area of the region, a is the maximum span of the region in any direction and b is the span at right-angles to that direction. To make the measure invariant under rotation the object is transformed using the Hotelling/Kauren-Loeve Transform [24] before calculation.

7.5 Smooth surface

As discussed in Section 3.2, the highlight texture might be a feature that discriminates man-made from natural objects. There exist many different measures which can be used to reveal texture differences; the ones used here were peak intensity value, mean intensity and normalized standard deviation for all highlight pixels

Chapter 8

Results

8.1 Segmentation using a region growing approach

Fig. 8.1 shows the results of the region growing approach for Test Images 4 and 7, compared with the manual segmentation used to test the classification features.

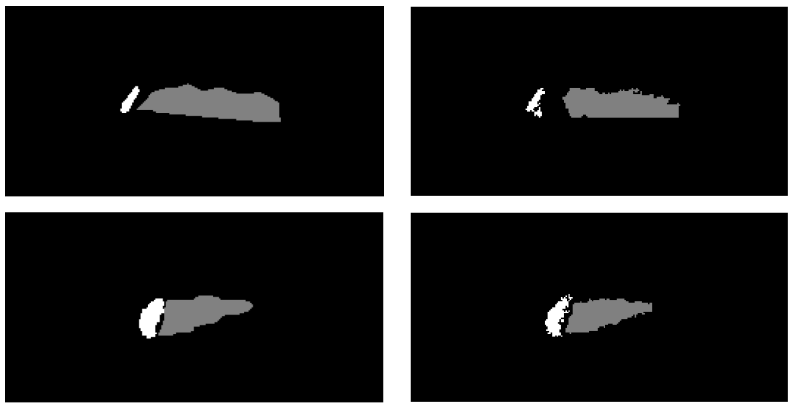


Figure 8.1: Comparison between manual segmentation and computer segmentation for Test Images 4 and 7. The left-hand images are manually segmented

It is worth noting the difference in the shadows for Test Image 7. In the computer segmented version, the shadow has a straight line that is not present in the manual segmented image. This artificial boundary line is due to the second *a priori* criterion in the segmentation method. In Appendix A, the result for all test images are shown together with the original images. The same parameter set was used for all images, with good results except for Test Image 2, where different parameters are needed.

8.2 Segmentation using a Markov random field approach

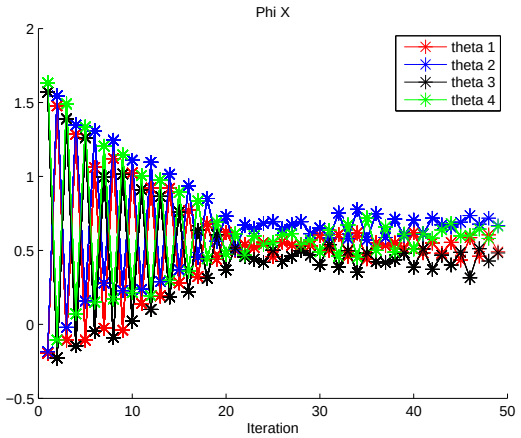
Fig. 8.2 shows the result of the MRF segmentation approach for Test Image 6. The result after initiation with the ICE algorithm is shown in Fig. 8.2(a), Fig. 8.2(b) shows the result after 25 iterations and Fig. 8.2(c) shows the final result. The result of the parameter estimation step can be seen in Fig. 8.3, where 8.3(a) shows the convergence of the Markov parameters and 8.3(b) shows the estimated probability density functions.

(a) After initation

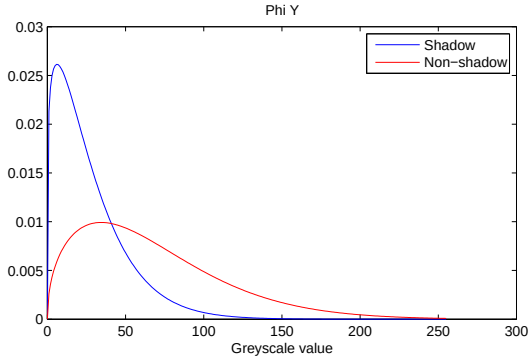
(b) After 25 iterations

(c) Final result

Figure 8.2: Result for MRF segmentation for Test Image 6



(a) Estimation of Markov parameters



(b) Estimated p.d.f. for shadow and non-shadow

Figure 8.3: Result for parameter estimation

The results are promising, but not as good as the results for the conventional sidescan images presented in the references.

8.3 Classification features

In order to evaluate the classification features, the Test Images were manually segmented. Table 8.3 lists the features describing the smoothness of object surface. Intensities are given with a nominal scale factor. Some compensation for attenuation with range is already applied in the image reconstruction algorithm, so no further correction is applied.

As can be seen from the table, mean intensity looks the most useful as a classification feature, while standard deviation normalized by mean is promising and might be used together with other features as an indicator for man-made objects. These results indicate that peak intensity does not seem to be a helpful classification feature.

Table 8.3 shows the result for the shape features discussed in Chapter 7. Test Image 1 has two highlight regions belonging to the same object, so two values are given for the highlight-shape features. The object height given is the maximum estimated height. The mines in Test Images 6 and 7 are box-shaped with approximate size $1.0 \times 0.7 \times 0.7$ [m].
As discussed in Chapter 7, object size is a useful feature, when it can be estimated accurately. As can be see for Test Images 6 and 7 object size is not estimated sufficiently accurately to identify the mine type by itself, but might be useful together with other features. Regarding the other shape features, rectangularity shows promise as an indicator for man-made objects. The line-based features(longest line, parallel lines and corners) could also be useful, particularly parallel lines. However there is a question whether they are sufficiently robust. In [25] the same idea was used, but with a different choice of parameters (threshold and resolution). Those results showed that the line-based features were not useful at all.

Test Image	Object	Peak intensity	Mean intensity	$\frac{Std.Deviation}{Mean}$	Range [m]
1	Anchor	6.22	1.69	0.72	38
2	Torpedo mine	5.73	1.68	0.58	64
3	Rock	7.08	1.11	0.93	53
4	Rock	5.65	1.01	0.80	39
5	Rock	7.97	0.84	1.33	40
6	Mine	4.76	1.57	0.82	50
7	Mine	7.59	1.98	0.65	56

Table 8.1: Features describing smoothness of the object surface. The results are for the test images shown in Appendix A after manual segmentation

Im.	Object	Corners lines		Parallel lines		Rectangularity		Longest line [m]		Highlight size [m]	Obj. height [m]
		H.	S.	H.	S.	H.	S.	H.	S.		
1	Anchor	0	0	1	0	0.6493	0.5155	0.81	1.69	1.10	0.5
		0		0		0.6664		1.00		1.01	
2	Torpedo mine	0	0	1	1	0.7897	0.7166	1.44	6.10	2.96	1.1
3	Rock	0	0	0	0	0.7218	0.7243	1.00	2.06	1.30	0.5
4	Rock	0	0	0	0	0.7133	0.6841	1.06	2.44	1.84	0.8
5	Rock	0	0	1	0	0.7771	0.6420	0.75	1.31	1.13	0.2
6	Mine	0	0	0	0	0.8407	0.8395	1.06	4.63	0.92	0.9
7	Mine	0	0	0	1	0.8222	0.7819	0.88	4.94	1.23	0.8

Table 8.2: Features describing object shape. The results are for manually segmented test images. Since Test Image 1 has two highlight regions, two values are given. The mines in Test Images 6 and 7 are box-shaped with approximate size $1.0 \times 0.7 \times 0.7[m]$

Chapter 9

Conclusions and future work

9.1 General

As a first study of a CAD/CAC system for the SAPPHIRES vehicle, the results are encouraging, but much more work is needed to achieve a functional CAD/CAC system. Future work needs to be done in all three approaches discussed in this thesis. An important step is to build up a larger test set of images for assessing the classification features and finding a more robust way of choosing parameters. A sufficiently large test set will allow the application of alternative classification methods, for example neural nets [18]. This step would be premature with the small amount of available data.

When the vehicle is fully developed, including autopositioning [10] and interferometry [26], the image resolution can be expected to be improved, and the methods discussed here might be even more useful.

9.2 Segmentation using a region growing approach

The results of the region growing approach indicate that the general method might be useful. One big concern is the large number of parameters needed to set the different thresholds. For future development, it would be desirable to estimate these parameters from the image itself, as in the MRF algorithm. Another possibility would be to use additional statistical measures, for example higher order statistics.

9.3 Segmentation using a Markov random fields approach

The results of the MRF segmentation here are not as good as for the conventional sidescan images presented in the references but good enough to justify further investigation. In many of the cited references, statistical snakes [20] were applied

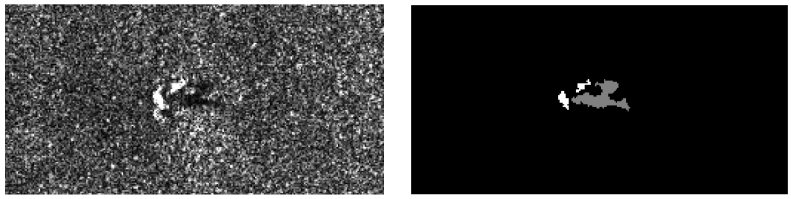
after the segmentation step to improve segmentation and to remove shadows due to seabed structures. Another possibility might be to try some alternative maximization method instead of ICM, as discussed in [13].

9.4 Classification features

It is hard to draw any firm conclusions of the value of the classification features considered from such a small test set. Some of the features discussed in the thesis seem to be useful. For example the features describing object surface might be used to discriminate man-made from natural objects and the shape features can be used as indicators for man-made objects. Other interesting features to evaluate in the future are symmetry as a continuous feature measure discussed in [27], and higher order statistics discussed in [17].

Appendix A

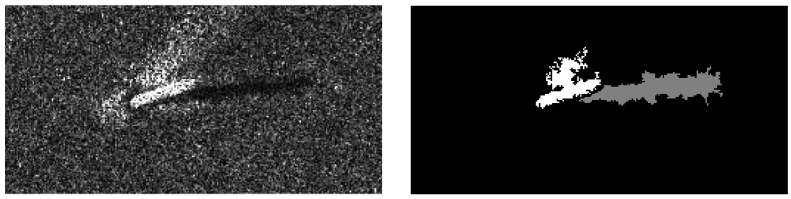
Test images



(a) Original image

(b) Computer segmented image,using same parameter set for all Test Images

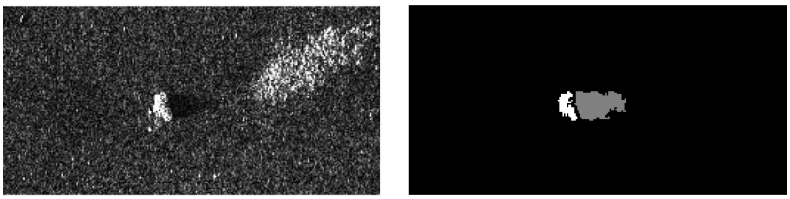
Figure A.1: Test image 1. Anchor at range $\approx$ 38 m, altitude $\approx$ 9.6 m



(a) Original image

(b) Computer segmented image,using same parameter set for all Test Images

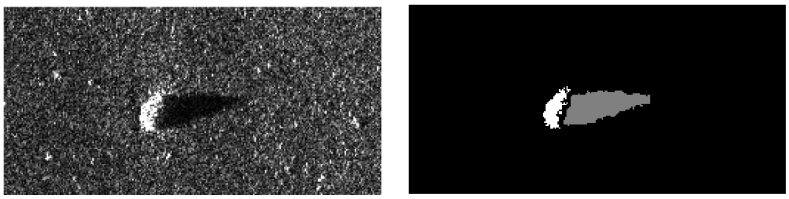
Figure A.2: Test image 2. Torpedo mine at range $\approx$ 64 m, altitude $\approx$ 16.9 m



(a) Original image

(b) Computer segmented image,using same parameter set for all Test Images

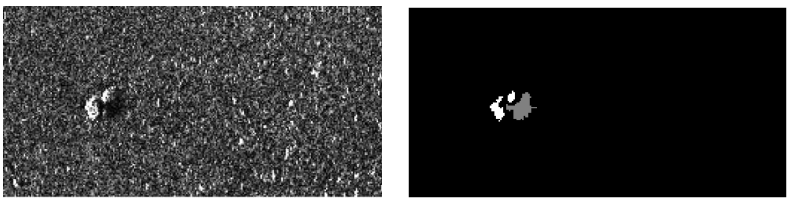
Figure A.3: Test image 3. Rock at range $\approx$ 53 m, altitude $\approx$ 10.9 m



(a) Original image

(b) Computer segmented image,using same parameter set for all Test Images

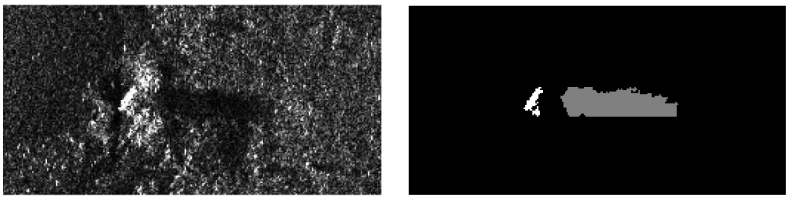
Figure A.4: Test image 4. Rock at range $\approx$ 39 m, altitude $\approx$ 9.0 m



(a) Original image

(b) Computer segmented image,using same parameter set for all Test Images

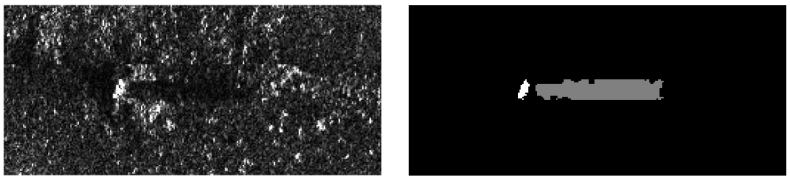
Figure A.5: Test image 5. Rock at range $\approx$ 40 m, altitude $\approx$ 8.4 m



(a) Original image

(b) Computer segmented image,using same parameter set for all Test Images

Figure A.6: Test image 6. Mine at range $\approx$ 50 m, altitude $\approx$ 8.5 m



(a) Original image

(b) Computer segmented image,using same parameter set for all Test Images

Figure A.7: Test image 7. Mine at range $\approx$ 56 m, altitude $\approx$ 8.5 m

Bibliography

[1] Östlund Patrick, http://www.marinen.mil.se/index.php?lang=S&c=news&id=36510, *Försvarmakten*. 10th of April 2007.

[2] "Oceanography and Mine Warfare". *Ocean Studies Board, National Research Council*. ISBN: 978-0-309-06798-0. 2000. http://books.nap.edu/openbook.php?isbn=0309067987

[3] Ashley, D. Waite. "SONAR for Practising Engineers". *John Wiley & Sons Ltd*. ISBN: 0-471-49750-9. 2002

[4] Urick, Robert J. "Principles of underwater sound". *McGraw-Hill, Inc.* ISBN: 0-07-066087-5. 1983.

[5] Bellettini A., Pinto M., Wang L. "Effect of multipath on synthetic aperture sonar". *WCU 2003, Paris, september 7-10, 2003*

[6] Nielsen, Richard O. "Sonar signal processing". *Artech House, Inc.* ISBN: 0-89006-453-9. 1991.

[7] Banks Simon. "Studies in High Resolution Synthetic Aperture Sonar", *PHD thesis,UCL , September 2002*

[8] Berens Patrick. "Introduction to Synthetic Aperture Radar (SAR)", *Research Institute for High-Frequency Physics and Radar Techniques (FHR), Research Establishment for Applied Science (FGAN), Wachtberg Germany*

[9] Koo V. C., Lim T. S., Chuah H.T. "A Comparison of Autofocus Algorithms for SAR Imagery". *Progress In Electromagnetics Research Symposium 2005, Hangzhou, China, August 22-26*

[10] Jönsson Mattias, Parastates Elias, Shippey Geoffrey, Pihl Jörgen, Karlsson Peter, Dalberg Eva, Nillsson Bernt. "Signal Processing Methods for Active Synthetic Aperture Sonar". *FOI- Swedish defence research agency* ISSN:1650-1942,FOI-R-0528–SE June 2002

[11] Bellettini Andrea, Pinto Marc A. "Theoretical Accuracy of Synthetic Aperture Sonar Micronavigation Using a Displaced Phase-Center Antenna". *IEEE Journal of Oceanic Engineering, Vol. 27, No. 4, October 2002*

[12] Steven A. Fortune "Phase Error Estimation for Synthetic Aperture Imagery" *PHD thesis University of Canterbury, Christchurch, New Zealand. June 2005*

[13] Mignotte M., Collet C., Pérez P., Bouthemy P. "Sonar Image Segmentation Using an Unsupervised Hierarchical MRF Model" *IEEE Transactions on Image Processing, Vol 9., No 7., July 2000*

[14] Maussang F., Chanussot J., Hétet A. "Automated Segmenation of SAS Images using the Mean - Standard Deviation Plane for the Detection of Underwater Mines". *Proceedings of MTS/IEEE Oceans'03 conference, San Diego, California, USA September 2003, pp. 2155-2160*

[15] Maussang F., Chanussot J., Hory C., Hétet A. "Synthetic Aperture Sonar Imagery: Towards a Classification of Underwater Mines in the Mean - Standard Deviation Plane". *Proceedings of Physics in Signal and Image Processing (PSIP'03), pp 137-140,Grenoble, France, January 2003*

[16] Maussang F., Chanussot J., Hétet A. "Higher order statistics for the detection of underwater mines in SAS imagery". *Proceedings of the Seventh European Conference on Underwater Acoustics, ECUA 2004, Deltf, The Neterlands, 5-8 July 2004*

[17] Maussang F., Rombaut M., Chanussot J., Hétet A. "Fusion of Local Statistical Detectors in SAS Imagery Classification".*Proceedings of Oceans'05 - Europe, Vol.1, pp 299-304, Brest,France, 20-23 June 2005*

[18] Sonka M., Hlavac V., Boyle R. "Image Processing, Analysis, and Machine Vision". *Brooks/Cole Publishing Company.* ISBN: 0-534-95393-X. 1998.

[19] Mignotte M., Collet C., Pérez P., Bouthemy P. "Three-Class Markovian Segmentation of High-Resolution Sonar Images" *Computer Vision and Image Understanding Vol.76, No. 3, December, pp 191-204, 1999.*

[20] Reed S., Petillot, Bell J. "An Automatic Approach to the Detection and Extraction of Mine Features in Sidescan Sonar" *IEEE Journal of oceanic engineering Vol. 28, No. 1, January 2003.*

[21] Geman S., Geman D. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images" *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6,pp.721-741,Nov 1984*

[22] Li S. Z. "Markov Random Field Modeling in Image Analysis" *Springer-Verlag.* ISBN: 4-431-70309-8. 2001.

[23] Mignotte M., Collet C., Pérez P., Bouthemy P. "Unsupervised segmentation applied on sonar images" *Proceedings International Workshop EMMCVPR'97: Energy Minimisation Methods in Computer Vision and Pattern Recognition, volume LNCS 1223, pages 491-506, Venice, Italy, May 1997. Springer-verlag.*

[24] Gonzalez R. C., Woods R. E. "Digital Image Processing" *Prentice-Hall Inc.* ISBN: 0-201-18075-8. 2002.

[25] Engström J., Trieb M., Shippey G.A. "Tools for classification of mine-like objects in synthetic aperture sonar images". *Proceedings of Institute Of Acoustics International Conference on Detection and Classification of Underwater Targets, Vol. 29, Pt.6, September 2007*

[26] Bonifant W.W. "Interferometric synthetic aperture sonar processing", Master's thesis, Georgia Institute of Technology, July 1999

[27] Zabrodsky H., Peleg S., Avnir D.. "Symmetry as a Continuous Feature". *IEEE Transactions on pattern analysis and machine intelligence, vol 17, no 12, December 1995*