

*SACLANT UNDERSEA  
RESEARCH CENTRE  
REPORT*



REDUCTION OF LOW FREQUENCY  
ACTIVE SONAR CLUTTER  
THROUGH IMAGE PROCESSING

*S. Dugelay, D.A. Abraham*

December 1997

The SACLANT Undersea Research Centre provides the Supreme Allied Commander Atlantic (SACLANT) with scientific and technical assistance under the terms of its NATO charter, which entered into force on 1 February 1963. Without prejudice to this main task – and under the policy direction of SACLANT – the Centre also renders scientific and technical assistance to the individual NATO nations.

---

This document is approved for public release.  
Distribution is unlimited

---

SACLANT Undersea Research Centre  
Viale San Bartolomeo 400  
19138 San Bartolomeo (SP), Italy

tel: +39-187-540.111  
fax: +39-187-524.600

e-mail: [library@saclantc.nato.int](mailto:library@saclantc.nato.int)

**NORTH ATLANTIC TREATY ORGANIZATION**

Reduction of low frequency  
active sonar clutter  
through image processing

S. Dugelay and D.A. Abraham

---

The content of this document pertains to  
work performed under Project 041-3 of  
the SACLANTCEN Programme of Work.  
The document has been approved for  
release by The Director, SACLANTCEN.



Jan L. Spoelstra  
Director

intentionally blank page

**Reduction of low frequency active  
sonar clutter through image processing**

S. Dugelay and D.A. Abraham

**Executive Summary:**

One of the problems encountered in using low frequency active sonar for the detection, classification and localization of low Doppler submarine targets in shallow water is the large number of false clutter detections from sea bottom features. These detections can overload automatic tracking and classification algorithms. This report investigates the potential for image processing algorithms exploiting information from multiple beams to reduce the number of clutter detections by associating them over bearing and range and then discounting those too large to be a submarine. Application to real data has shown reductions greater than 80% from the number of target detections on all beams using a standard automatic detection algorithm to the number of submarine-like objects in the range-bearing image. Work in clutter reduction for low frequency active sonar is on-going at SACLANTCEN.

intentionally blank page

**Reduction of low frequency active  
sonar clutter through image processing**

S. Dugelay and D.A. Abraham

**Abstract:** Large numbers of false clutter detections arise in the use of low frequency active sonar for the detection of low Doppler submarine targets in shallow water. Traditional detection algorithms operate individually on each beam output searching for targets at all ranges. Detection algorithms such as the Page test [1] are designed for target echoes that are extended in range owing to the multipath propagation of shallow water channels and reflection off of the target. However, the target echo and bottom features may extend over multiple beams either physically or by bleeding through the sidelobes of the beamformer. This indicates that detections need to be associated across bearing or the detector must be designed to account for targets and clutter that are spread over multiple beams.

This report considers the detection/association issue from an image processing perspective, applying a Markov random field (MRF) model to the image formed from the range-bearing sonar data. The Markov random field model exploits a priori information such as the distribution of the reverberation, minimal information about the distribution of the target echo and the relationship between each range-bearing cell and its neighbouring cells. Maximum a posteriori (MAP) estimates of the labeling of each range-bearing cell (i.e., target or reverberation) are obtained rapidly through an iterative algorithm with an initialization provided by the Page test detector output. The objects in the resulting range-bearing image are then tested according to their size, ruling out any too large to be submarines. Application to real data has shown a reduction of greater than 80% from the number of Page test detections over all the beams to the number of submarine-like objects in the MRF-MAP image.

**Keywords:** clutter reduction ◦ detection ◦ segmentation ◦ classification ◦ Markov random fields ◦ image processing

## Contents

1	Introduction . . . . .	1
2	Brief description of the sonar and data acquisition . . . . .	3
2.1	The sonar and preliminary data processing . . . . .	3
2.2	Image construction . . . . .	4
2.3	Image description . . . . .	5
3	A Markov random field solution . . . . .	8
3.1	The theory of Markov random fields . . . . .	8
3.2	Applying theory to experimental data . . . . .	14
3.3	Results and conclusions . . . . .	21
4	Conclusions and prospects . . . . .	27
	References . . . . .	30
	Annex A - A review of image processing techniques in general . . . . .	32
A.1	Feature extraction . . . . .	33
A.2	Image segmentation . . . . .	38
A.3	Classification . . . . .	39

# 1

## Introduction

---

One of the problems encountered in the detection and classification of submarines in shallow water using low frequency active sonar is the large number of false clutter detections from sea bottom features. Too large a number of clutter detections from each ping will overload subsequent signal processing such as tracking and classification algorithms. In this study, we investigate the potential for image processing algorithms to reduce the number of clutter detections. Image processing has, as an advantage over single beam detection algorithms, the ability to associate detections occurring on multiple adjacent beams.

Direct application of standard image processing techniques to active sonar data is not necessarily trivial as the sonar image obtained from beamforming towed array data is in an irregular polar grid of range-beam cells as opposed to a rectangular grid. The goal of this report is to determine if image processing is indeed a useful tool in submarine detection and discrimination. Certainly, for many other underwater applications, image processing is now a crucial step in interpretation and for geological applications, several techniques have been applied using data from various sensors:

- Blondel *et al.* [2] used textural information contained in TOBI (Towed Ocean Bottom Instrument 30kHz) images. This approach is not based on any physical properties of the phenomena or geometry of the sonar. Artifacts cannot therefore be taken into account.
- Linett *et al.* [3] proposed a segmentation method based mainly on textural analysis either by fractal methods or spatial analysis for high frequency sonar data.
- Jiang *et al.* [4] developed a method of hierarchical Markov models for side-scan sonar images extracting textural features.
- IFREMER developed a Markovian model combined with angular reverberation for low-frequency (13kHz) multibeam echosounder images [5].

We can also cite the following authors for their work in minehunting using image processing methods:

- Pipelier [6] proposed a method for a sidescan minehunting sonar based on Markov random fields. This study demonstrated the interest of using Markovian models for image segmentations although the model employed was kept very simple and required more development.
- Thourel [7] also utilized Markov fields combined with a multiresolution or multimodel approach for minehunting applications. While background intensity data are considered to follow a chi-squared distribution law, mine shadows are assumed Gaussian distributed.
- Stage and Zerr [8] took advantage of the high resolution present in minehunting sonar images in order to compute local statistical properties around a pixel. This introduces an average filtering process slightly degrading the image but which in their case was not detrimental. The method developed detects objects by checking if they belong to the background or not; i.e. if their local statistical distribution reasonably follows the distribution of the background (a chi-squared distribution in intensity) or not.

It is to be noted that most recent applications have been constructed using Markov random fields (MRF): the main advantage of Markov random field theory is that it is a simple mathematical model that allows different levels of interpretation. High frequency applications (in geology) have a tendency to base their method on textural analysis or averaging schemes. In the present case, textural information is unknown or unavailable and poor resolution renders averaging dangerous; only Markov random fields seem to have a probabilistic theoretical background solid enough to incorporate a priori knowledge and to produce reliable results. Unfortunately, a literature search revealed very little in applying image processing to the acoustic detection of submarines.

We will determine if it is possible to discriminate targets from strong bottom features, while working in a noisy environment. Short descriptions of the sonar system geometry and data acquisition are necessary if an adequate method is to be developed; this will be described in Section 2. Images are of course greatly dependent on the acquisition system itself, and the construction process; we briefly discuss this process in Section 2.2. Section 3 describes the application of Markov random fields to the active sonar clutter reduction problem including the modifications required and results demonstrating the viability of the approach on real data. Finally, our conclusions will summarize the problem and proposed solution, giving way to numerous future prospects in the final section. A brief review of the most important image processing techniques is presented in Annex A, where their constraints, disadvantages, and applicability to the current problem are discussed.

Brief description of the sonar  
and data acquisition

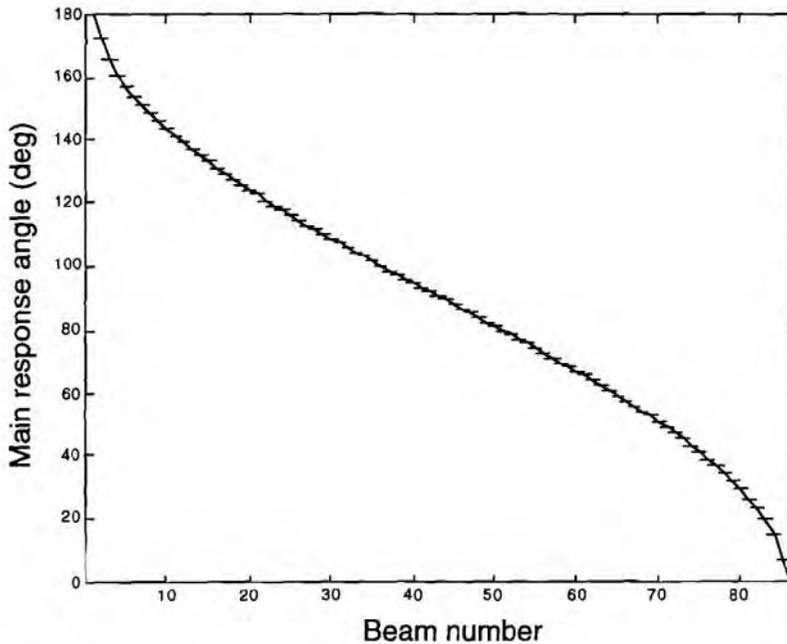
---

*2.1 The sonar and preliminary data processing*

The data analyzed in this report were collected using the TVDS source and the Centre's low frequency towed line array. The transmission waveform was a 2.29 second hyperbolic frequency modulated (HFM) waveform sweeping from 460 to 565 Hz. The data received from 128 sensors spaced one meter apart were beamformed using Hanning shading on the array to beams spaced so that the beampatterns overlapped 3dB down from their main response axis at 700 Hz, resulting in 86 beams spanning from forward to aft endfire.

The data from each beam were matched filtered, basebanded so that the center of the waveform band (512.5 Hz) shifted to zero Hertz, and decimated to a sampling frequency of 125 Hz. The resulting data are nearly statistically independent from sample to sample in range and provide a range sampling interval of 6 meters assuming 750 m/s for the two-way speed of sound. The data were then normalized using a trimmed-mean order statistic normalizer [9] with leading and lagging windows each consisting of 100 samples and removal of the upper and lower quarter of the combined auxiliary data.

As the beamformer was designed with 3dB overlap in the beampatterns at 700 Hz, we expect a larger degree of overlap at lower frequencies. Evaluation of the beampatterns indicated that a point target arriving on the main response axis of a beam is attenuated by approximately 6 dB on adjacent beams and 35 dB two beams away. Thus, we expect that submarines with high target strength will be detected on up to three adjacent beams even at far ranges or angles away from broadside where the range-bearing cell centers may be separated by distances greater than the physical extent of the submarine. The correspondence between beam number and main response axis is depicted in Fig. 1 where zero degrees indicates forward endfire.

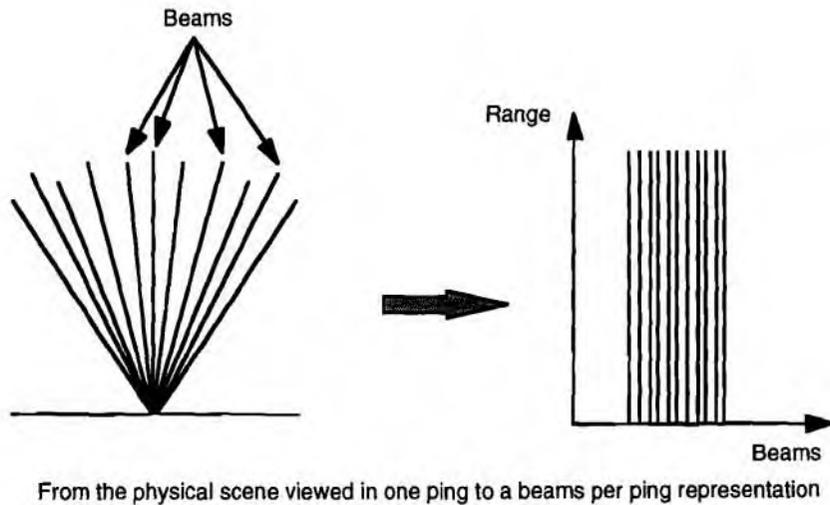


**Figure 1** *Beamformer main response angles as a function of beam number.*

## 2.2 Image construction

The image construction process is greatly dependent on the constraints imposed by the beamforming. Our major priority is constructing an image where samples have been transformed as little as possible. This means that we would like for example to avoid geometric transformations such as projections from a polar coordinate system into a cartesian coordinate system and interpolation. We also would like to avoid having to average samples falling into the same pixel. Such transformations can be incorporated into an image processing technique, but in this tentative study, we want to keep matters fairly simple and straightforward.

A representation of physical reality would be a “fan-view” image. This supposes geometric transformation of the data and interpolation. A priori knowledge of pixel distributions may then be lost. Such techniques introduce inter-pixel correlation which we want to avoid. We also want to keep the resolution of the time series (6m). Therefore, the easiest and most convenient representation is simply a bearing vs. range representation for each ping as shown in Fig. 2, where the varying distance between adjacent range-bearing cells is accounted for in the image processing algorithm. This representation is useful for detecting objects in a single ping and conveys a global view of the area in one take.



**Figure 2** *Bearing vs. range representation*

A similar representation can be produced for a closer look at a particular beam when multiple pings are available: pings vs. range for one beam, as shown in Fig. 3. This kind of image is useful in tracking a feature. Static features if detected by the same beam will appear moving due to the motion of the transmit and receive antennae. Since we have knowledge of the ship's speed and direction, it is possible to determine whether a target is indeed moving or static.

Finally, a combination of both representations produces a temporal bearing vs. range image as seen in Fig. 4. This is particularly useful if one is detecting objects and estimating relative motion at the same time.

### 2.3 *Image description*

In this section, we will enumerate the most important image characteristics which make the low frequency active sonar problem unique in itself, and therefore demand specific modelling.

From the description of the construction process above, it is clear that an important characteristic of the image is that it does not translate the real physical view of the scene but conveys a distorted view. Indeed, pixels which are neighbours in the image, are in fact situated meters apart in reality. Also, the separating distance depends on range and bearing. It will be necessary to keep this in mind when applying image processing techniques.

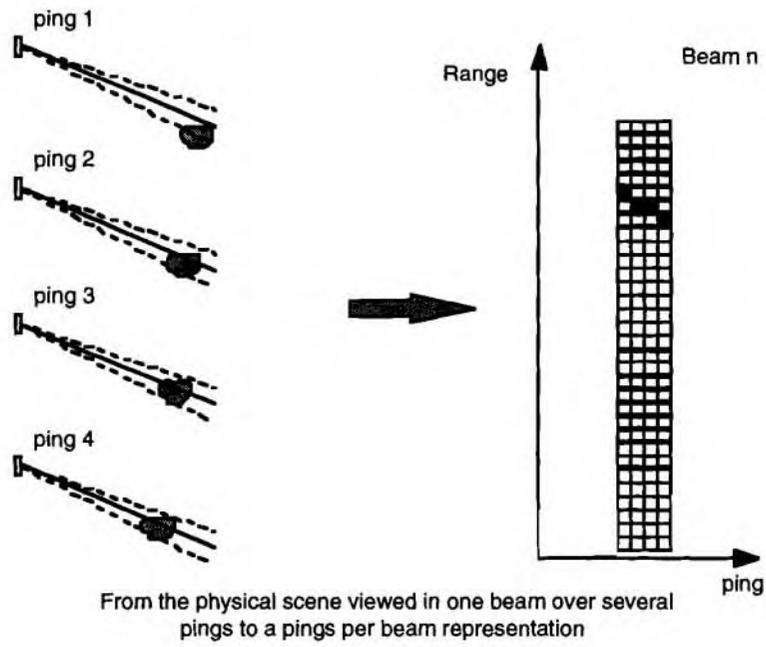


Figure 3 *Pings vs. range representation*

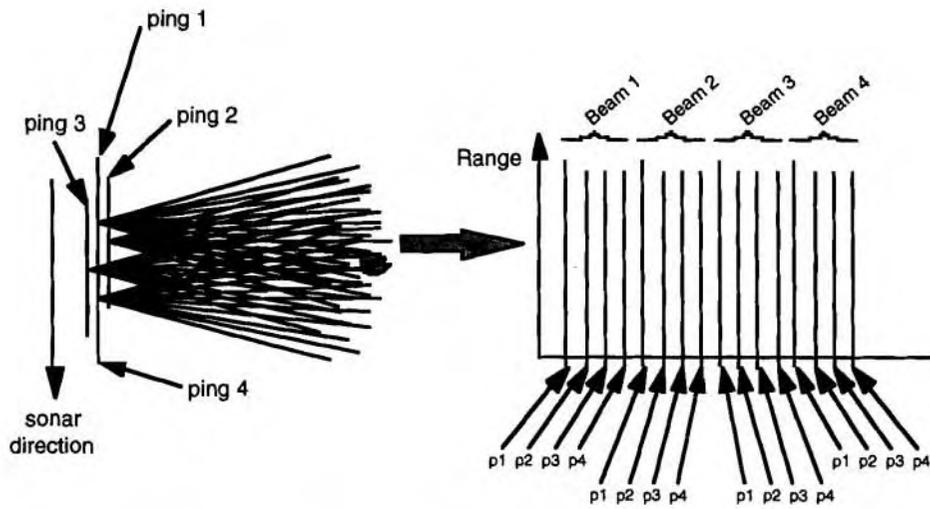


Figure 4 *Temporal bearing vs. range representation*

Secondly, we may observe a spreading effect of a feature dependent on range as is demonstrated in Fig. 5. At close range, a bottom feature may encompass several beams because of its size. At further ranges, the distance between beams increases, and the feature will not be seen in as many beams. This spreading effect can be estimated and it should be particularly useful when classifying targets based on their shape.

Finally, no information is readily available on target pixel intensities or probability distributions. We will assume that background pixels are independent and that intensities reasonably follow a chi-squared distribution with 2 degrees of freedom (i.e., Rayleigh distributed reverberation amplitude). The hypothesis of such a law is not rejected by the statistical Kolmogorov-Smirnov test [10] on the real data processed for this report.

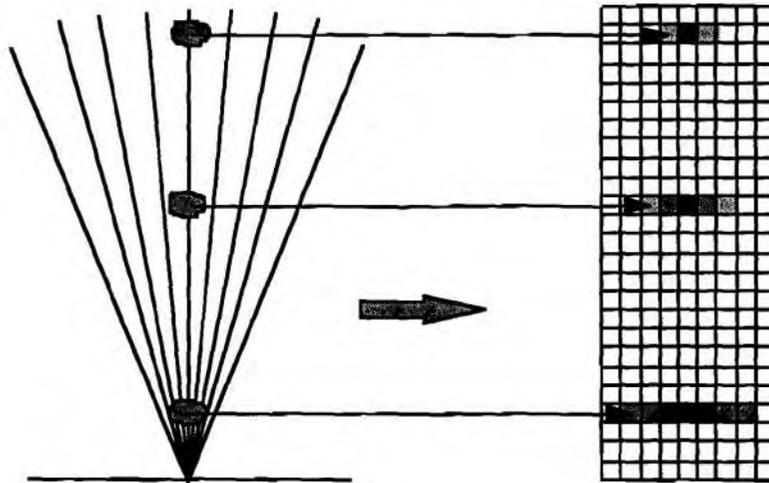


Figure 5 *Spreading effect*

# 3

## A Markov random field solution

---

Markov random field theory is a branch of probability theory for analyzing the spatial or contextual dependencies of physical phenomena [11, 12, 13]. Its most interesting characteristic is the ability to define a local model when in reality it is global. In particular, this allows dictating that a pixel will only depend on its neighbours while adequately modeling the global relationships of the pixels within the image. In this section, we will introduce notations and definitions relevant to image processing [4, 5, 6, 7, 11, 12, 13, 14, 15]; then, we will expose the application of MRF theory to low frequency active sonar data.

### 3.1 The theory of Markov random fields

#### 3.1.1 Sites and labels

The goal of segmentation is to affect a label to each pixel of the image. The labelling problem is then specified in terms of sites and a set of labels. Let  $S = \{s = \{i, j\} | 1 \leq i \leq m, 1 \leq j \leq n\}$  be a set of sites. Its elements correspond to the locations at which the image is sampled (in the present case, we have a regularly spaced lattice). This set can be conveniently re-indexed by a single number  $k$ , where  $k$  takes on values in  $\{1, 2, \dots, N\}$  with  $N = m \times n$ .

A label represents an event that may occur at a site. Let  $L$  be a set of labels. In the discrete case, a label assumes a discrete value in a set of  $M$  labels:

$$L = \{l_1, l_2, \dots, l_M\} \quad (1)$$

The labelling problem is then to assign a label from the set  $L$  to each of the sites in  $S$ . As an example, in the case of edge detection in an image, the problem is to assign a label  $x_i$  from the set  $L = \{edge, non - edge\}$  to site  $i \in S$  where elements in  $S$  index the image pixel. The set

$$x = \{x_1, x_2, \dots, x_N\} \quad (2)$$

is called a labelling of the sites in  $S$  in terms of the labels  $L$ ,  $x_i \in L$ . In the terminology of random fields, this is also called a configuration. A configuration, or

labelling, can correspond to an image, an edge map, or an interpretation of image features in terms of object features, and so on. When all the sites have the same label set, the set of all possible labellings, called the configuration space, is the following cartesian product:

$$\Omega = \underbrace{L \times L \cdots \times L}_{N \text{ times}} = L^N \quad (3)$$

### 3.1.2 Neighbourhood system and cliques

The inter-relationship between pixels is described through a neighbourhood system associated with the set of sites  $S$ . The neighbourhood system  $V$  is defined as:

$$V = \{v_i | i \in S\} \quad (4)$$

where  $v_i$  is the set of sites neighbouring  $i$ . The neighbouring relationship has the following properties:

1. a site is not a neighbour to itself:  $i \notin v_i$
2. the neighbouring relationship is mutual:  $i \in v_j \Leftrightarrow j \in v_i$

The pair  $[S, V]$  constitutes a graph where  $S$  contains the nodes and  $V$  determines the links between the nodes according to a neighbourhood relationship. A clique  $c$  for  $[S, V]$  is defined as a subset of  $S$ :

1.  $c$  is a single site  $c = \{i\}$  or
2. two elements of  $c$  are neighbours according to the neighbourhood system  $V$

A clique is in fact either a single site, or a set of sites in a neighbourhood which are linked to each other via the neighbourhood system. These sets will be useful, for example, if we want to distinguish links between vertical neighbours, horizontal neighbours or diagonal neighbours.

The order of a clique corresponds to the number of elements contained in the clique:

$$\begin{aligned} C_1 &= \{i | i \in S\} \\ C_2 &= \{\{i, j\} | j \in v_i, i \in S\} \\ C_3 &= \{\{i, j, k\} | i, j, k \in S \text{ are neighbours to one another}\} \end{aligned}$$

In Fig. 6, we present two neighbourhood systems and the associated cliques on a lattice of regular sites. The type of a clique on a regular lattice is determined by its

size, shape and orientation. As the order of the neighbourhood system increases the number of cliques grows rapidly and so do the involved computational expenses. In many cases, whatever the neighbourhood system, only second order cliques will be considered.

It is also possible to define cliques on an irregular lattice where the neighbourhood system may be defined as a disc. In Fig. 7 is an example of irregularly spaced sites, a neighbourhood for one site (marked by the dashed circle), and the corresponding cliques. Note that the set  $\{m, i, f\}$  does not form a clique because  $f$  and  $m$  are not neighbours.

### 3.1.3 Markov random fields: definition

Let  $X = \{X_1, X_2, \dots, X_N\}$  be a family of random variables defined for the set  $S$ , where each variable  $X_i$  takes a value  $x_i$  in  $L$ . The family of variables  $X$  is called a random field and we define  $x = \{x_1, x_2, \dots, x_N\} \in \Omega$  as a configuration of  $X$ , corresponding to a realisation of the field. For a discrete set of labels  $L$ ,  $P(X_i = x_i)$  is the probability of the variable  $X_i$  taking the value  $x_i$ , and the joint probability is given by  $P(X = x) = P(X_1 = x_1, \dots, X_N = x_N)$ .

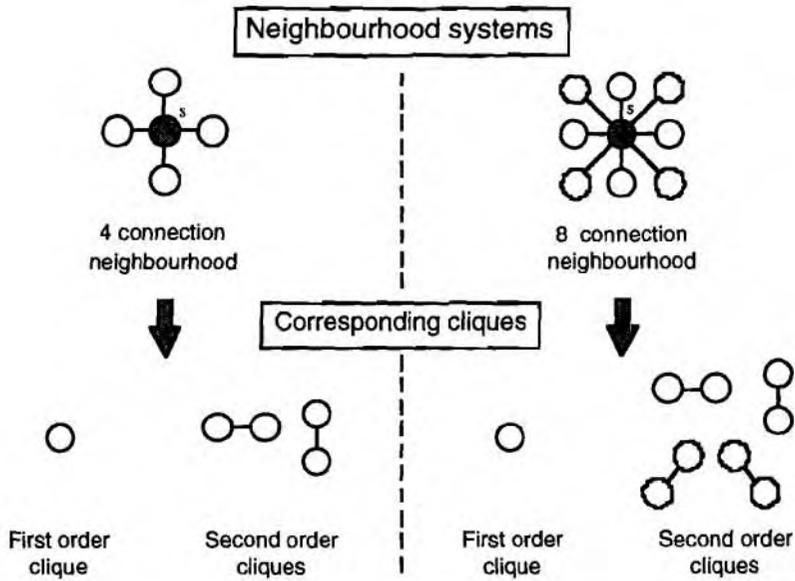
Definition:

*$X$  is said to be a Markov random field (MRF) on  $S$  with respect to the neighbourhood system  $V$  if and only if:*

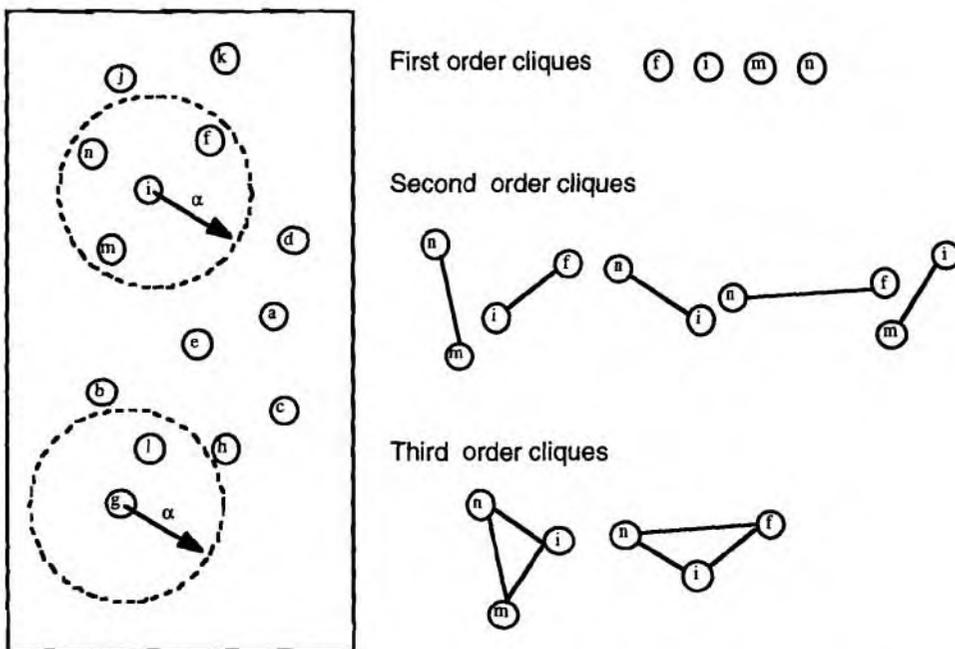
1.  $P(X = x) > 0, \forall x \in \Omega$
2.  $P(x_i | x_{S-\{i\}}) = P(x_i | x_{v_i})$

*where  $x_{S-\{i\}}$  denotes the set of labels at the sites  $S - \{i\}$ , and  $x_{v_i} = \{x_j | j \in v_i\}$  are the labels of the neighbourhoods of  $i$ .*

The first condition imposes positivity and can usually be satisfied in practice. The second condition (Markovianity) describes the local interaction between pixels: only neighbouring labels have direct interactions on each other. It is always possible to select a sufficiently large  $V_i$  so that the Markovian property holds; the largest of neighbourhoods consisting of all other sites, any field  $X$  is an MRF with respect to such a neighbourhood system. An MRF is said to be homogeneous if  $P(x_i | x_{v_i})$  does not depend on the position of the site  $i$  in  $S$ ; it is said to be isotropic if  $P(x_i | x_{v_i})$  does not depend on clique orientation.



**Figure 6** *Neighbourhoods and cliques on a regular lattice*



**Figure 7** *Neighbourhoods and cliques on an irregular lattice*

### 3.1.4 Gibbs random fields

A set of random variables  $W$  is said to be a Gibbs random field (GRF) on  $S$  with respect to a neighbourhood system  $V$  if and only if its configurations obey a Gibbs distribution in the following form:

$$P(w) = \frac{1}{Z} e^{-U(w)/T} \quad (5)$$

where  $Z = \sum_{w \in \Omega} e^{-\frac{1}{T}U(w)}$  is a normalizing constant called the partition function,  $T$  is a constant called the temperature which will be assumed to be 1 unless stated otherwise, and  $U(w)$  is the energy function. The energy is a sum of clique potentials over all possible cliques  $C$ :

$$U(w) = \sum_{c \in C} V_c(w) \quad (6)$$

The temperature  $T$  controls the sharpness of the distribution. When the temperature is high, all configurations tend to be equally distributed. Near the zero temperature, the distribution concentrates around the global energy minima.

### 3.1.5 Hammersley-Clifford theorem (energy minimisation)

The Hammersley-Clifford theorem establishes the equivalence between the local property of a Markov random field and the global property of the Gibbs random field. The theorem states:

*Let  $V$  be the neighbourhood system associated with the set of sites  $S$ .  
Then,  $W$  is a MRF with respect to  $V$  if and only if  $W$  is a GRF.*

### 3.1.6 Baye's framework (MAP)

In most images, neighbouring pixels generally display similar intensities; when texture is present, relations or particular orientations are discernible. These regularities describe correlations or likelihoods. According to Baye's theory, when both the a priori distribution and likelihood function of a pattern are known, the best that can be estimated from these sources of knowledge is the Baye's labelling.

$P(X = x)$  is the a priori probability,  $P(Y = y|X = x)$  the probability of  $Y$  conditional on  $X$ . Baye's formula then gives us the a posteriori probability:

$$P(X = x|Y = y) = \frac{P(Y = y|X = x) P(X = x)}{P(Y = y)} \quad (7)$$

In applying MRFs to active sonar data,  $X$  are the labels assigned to each range-bearing cell and  $Y$  are the corresponding normalized intensity data.

A cost function  $C(x, \hat{x})$  is defined as a measure of the discrepancy between the field of labels and its estimation, and thereby characterizes the risk associated with this estimation. The optimal estimator is obtained by minimizing the risk  $R(\hat{x})$ :

$$R(\hat{x}) = \sum_{x \in \Omega} C(x, \hat{x}) P(X = x | Y = y) \quad (8)$$

The cost function  $C(x, \hat{x})$  is chosen according to preference; two popular choices are the quadratic cost function:

$$C(x, \hat{x}) = \|\hat{x} - x\|^2 \quad (9)$$

where  $\|a - b\|$  is the distance between  $a$  and  $b$ , and the (0 - 1) or  $\delta$  cost function

$$C(x, \hat{x}) = 1 - \delta(x, \hat{x}) \quad (10)$$

where  $\delta(a, b) = 1$  if  $a = b$  and 0 otherwise. The maximum a posteriori estimator is the estimator that penalizes all incorrect configurations in the same way using the (0 - 1) cost function. Then the MAP estimate is found by:

$$\begin{aligned} \hat{x}_{opt} &= \arg \min_{\hat{x} \in \Omega} R(\hat{x}) \\ &= \arg \max_{\hat{x} \in \Omega} P(X = \hat{x} | Y = y) \\ &= \arg \max_{x \in \Omega} P(Y = y | X = x) P(X = x) \end{aligned} \quad (11)$$

### 3.1.7 Optimization algorithm

The number of possible configurations makes it impossible to calculate the probability for each configuration and taking the optimal solution. Therefore, optimization algorithms have been developed. These algorithms can be divided into two categories: deterministic or stochastic. In this study, we have constrained ourselves to a deterministic algorithm: ICM (Iterated Conditional Mode [14]), where the temperature  $T$  is equal to 1. The main advantage of this algorithm is its speed to converge to a solution. The optimal solution is found if initialization is correct. This is in fact the biggest drawback of this approach since the algorithm needs an initialization fairly close to the global solution in order to converge to it. Nevertheless, in our case, a good initialization is generally feasible. The general outline of the algorithm is the following:

- Draw an initial configuration  $x^{(0)}$  as close as possible to the optimal configuration,

- Scan the set of sites  $S$ . Each site  $s$ , at iteration  $n$  is characterized by its state  $x_s^{(n)}$  and the configurations  $x_{v_s}^{(n)}$  in its neighbourhood  $v_s$ . Its new state  $x_s^{(n+1)}$  corresponds to that minimizing the conditional energy  $U(l|x_r^{(n)}, r \in v_s)$ ,
- Continue scanning the image for a determined number of iterations, or until a stop test has been fulfilled.

### 3.2 *Applying theory to experimental data*

#### 3.2.1 *The problem and image characteristics*

Our objective in the proposed processing of the range-bearing sonar image is the removal of clutter objects too large to be a submarine. This is accomplished by a segmentation process which detects objects within the image that are distinct from the background, followed by a pre-classification that separates these objects as either target-like or clutter. The latter step is not fully a classification as it can only rule out objects according to a size criterion, and is thus called ‘pre-classification’. To accomplish these tasks we first describe the pertinent assumptions and characteristics of the image:

- it is reasonable to assume a chi-squared distributed background with two degrees of freedom (i.e., Rayleigh distributed reverberation amplitude),
- due to the representation, the image is inhomogeneous and anisotropic,
- submarines are considered as small objects, hence shape and size are a useful discrimination feature,
- without removal of the displacement caused by motion of the tow ship from ping to ping, bottom features appear to be moving in the pings-per-beam image, though in a known manner.

In the next section, we will explain how this information has been introduced into the model in order to achieve the goal.

#### 3.2.2 *Modelling the energy terms*

Let us remember that the goal of the segmentation is to maximise the maximum a posteriori (MAP) probability. Given Baye’s formula, maximising the MAP probability requires maximisation of the product of the a priori probability  $P(X = x)$  and

the conditional probability  $P(Y = y|X = x)$ .  $P(Y = y)$  is a constant which does not enter the maximisation process.

We are therefore looking to model two probabilities:

- the conditional probability, which will be identical for both the single ping image and the consecutive pings for a single beam representation, and
- the a priori probability, which will enable us to incorporate a priori knowledge on interactions between pixels, and therefore depends on the representation mode.

These probabilities will both be modelled using a Gibbs distribution. The maximisation of the product of the two probabilities is then equivalent to the minimisation of the sum of the two energy functions  $U(Y = y|X = x)$  (the conditional energy) and  $U(X = x)$  (the a priori energy).

#### (A) The conditional probability

Only information on the background is available beforehand if we want to implement an unsupervised segmentation process. We therefore need to construct a conditional probability that will test if a pixel is indeed in the background or not. The quickest and most efficient way of obtaining this information is thresholding. The idea is to say that if the intensity of a pixel is sufficiently close to the mean of a chi-squared distribution with two degrees of freedom, then it is most likely a part of the background.

The conditional probability is then constructed in the following manner. First note that the probability that a variable  $Y$  following a chi-squared distribution be comprised between 0 and  $y$  is given by:

$$p(y) = 1 - \exp(-y/2) \quad (12)$$

What we need to know is the threshold value  $y_t$  when the probability of  $Y$  taking a value larger than  $y_t$  is "small". If we define "small" as  $\alpha$  then  $y_t$  can be calculated from the following equation:

$$P(Y > y_t) = 1 - p(y_t) = \exp(-y_t/2) = \alpha \quad (13)$$

The conditional probability is then defined as follows:

- If  $y < y_t$  then

$$P(Y_s = y|X = 0) = p \quad (14)$$

$$P(Y_s = y|X = 1) = 1 - p \quad (15)$$

- If  $y > y_t$  then

$$P(Y_s = y|X = 0) = 1 - q \quad (16)$$

$$P(Y_s = y|X = 1) = q \quad (17)$$

where  $X$  is the field of labels,  $X = 0$  corresponds to the background label, and  $X = 1$  to the object label.

The most delicate step will be defining  $p$  and  $q$ . We could define  $p$  equal to  $\alpha$  if we knew that histograms of the background pixel values and the object pixel values were separate. The problem arises when we allow histograms to overlap, and moreover when we do not have access to the histogram of the object pixel values.

To set some ideas, we have estimated histograms of the background amplitude, and that of the objects after a Page test detector [1] has been applied to the signal in each beam, and we have plotted them in the same graph (Fig. 8). Clearly, although estimation may be somewhat uncertain when it comes to the histogram of the detected objects, pixels with intensity lower than the threshold may in fact belong to an object. The value  $\alpha$  defined for a chi-squared distribution may then not be adequate for the conditional probability we want to model.

We have found, empirically, that values for  $p$  and  $q$  of 0.6 and 0.8 seem to adequately model our conditional probability. That is, if the intensity is lower than the threshold, then the pixel is most likely to be a background pixel but there is still a good chance of it being in an object. Whereas, if the intensity is higher than the threshold, then surely the pixel cannot be from the background and the probability that it is indeed coming from an object is high.

From eq. (5) and eqs. (14–17), we then have, ignoring the constant terms, the following conditional energies for each label:

- If  $y < y_t$  then

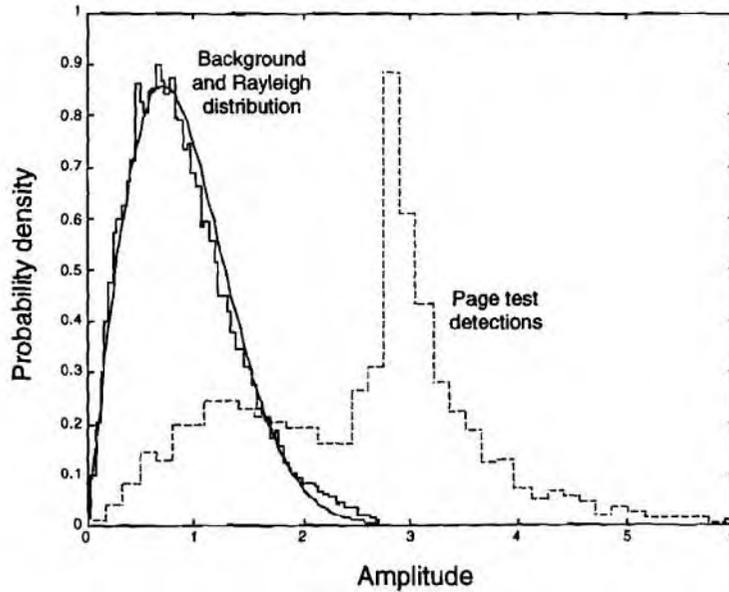
$$U(y|0) = -\ln(p) \quad (18)$$

$$U(y|1) = -\ln(1 - p) \quad (19)$$

- If  $y > y_t$  then

$$U(y|0) = -\ln(1 - q) \quad (20)$$

$$U(y|1) = -\ln(q) \quad (21)$$

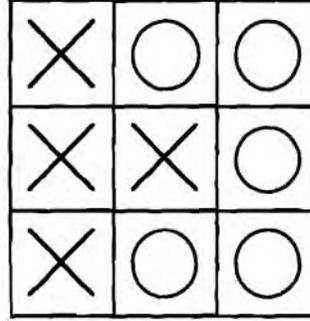


**Figure 8** Histograms of the pixel amplitudes from background and objects detected by the Page test.

**(B) The a priori probability**

The role of the a priori energy is to model the possible interactions in a clique. In this study, as in most studies, cliques are restricted to a maximum order of two; so, we are left to deal with interactions between two pixels or single sites. This energy may also be used to correct for any geometrical artifact resulting from the image construction process. This means that even though the image may not be a physical representation of the scene, real physical information can still be incorporated into the model. The a priori probability is generally modeled as a counting function called the Ising model [15]. This means that the process will count the number of different labels in the pixel neighbourhood and this energy term will act as a homogenizing function. In Fig. 9 is a schematic example of a pixel and its neighbourhood, and the role of homogenizing function. The pixel the process is trying to classify is the center one,  $X_s$ , which is labelled with a cross. It has 3 “cross” neighbours and 5 “circle” neighbours. The pixel will tend to be labelled as a circle, unless its intensity is too far away from the “circle” intensity; then, it will stay as it is. Neighbouring pixels ( $X_t, t \in v_s$ ) may also be weighted depending on their position for example, or correlation. The energy term for each class then counts the number of neighbouring pixels labeled with that class and weighted by their influence:

$$U(X_s = l) = -\alpha \ln(c_s) - \beta \sum_{t \in v_s} c_t \delta(x_t, l) \tag{22}$$



**Figure 9** *Using neighbouring labels*

where  $l$  equals either 0 for background or 1 for an object,  $\alpha$  and  $\beta$  are respectively homogenizing parameters for the first order and second order cliques,  $c_s$  and  $c_t$  are weighting coefficients depending on the clique and,  $\delta(a, b) = 1$  if  $a = b$ , and 0 otherwise. In this study,  $\alpha$  and  $\beta$  are equal and increased every iteration from 0.5 in steps of 0.1 until convergence. This allows acceleration of regularisation without wiping away small features too quickly, slowly moving from a conditional probability dominant model to an a priori dominant model [14, 5].

We have decomposed the energy of eq. (22) into two terms: the first order clique and the second. First, let us deal with the second order clique term. Here, we will re-introduce the distances which have been dropped in the image construction process; that is, the influence of a neighbouring pixel will be weighted by a relative inverse distance in the real physical scene as depicted in Fig. 10. As the distance between beam centers increases with range, we will in reality introduce higher relative weighting between pixels in the same beam as the weighting between pixels in different beams decreases with range. Note that this concept then implies an inhomogeneous Markov random field [5]. Two types of cliques have been distinguished: cliques for intrabeam pixel links (i.e., vertical second order cliques) and cliques for interbeam pixel links (i.e., horizontal and diagonal cliques). The weighting coefficients are then given by the following equations:

- Intrabeam links

$$c_t = 1 \tag{23}$$

- Interbeam links

$$c_t = \sqrt{d(R_0, \Delta\theta_0)/d(R_s, \theta_{(s,t)})} \tag{24}$$

where  $R_0$  is the starting range of the data (in this case  $R_0 = 6000m$ ),  $\Delta\theta_0$  is the minimum interbeam distance (i.e., at broadside),  $R_s$  is the range of the pixel  $s$ ,  $\theta_{(s,t)}$

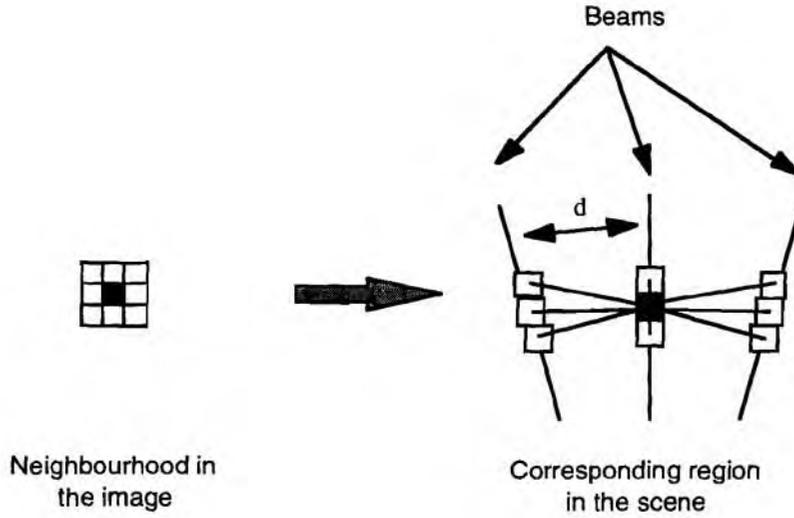


Figure 10 *Weighting the influence of neighbouring pixels*

is the interbeam angle between the beam  $B_s$  and the beam  $B_t$ , and  $d(R, \theta)$  is the interbeam distance at range  $R$  for a beam separation  $\theta$ .

As for the first order clique, we have modeled this term with the following considerations. We do not want small objects to disappear too easily, since targets can be seen in as few as three pixels. This means small objects (surrounded by background) must have a chance of surviving. On the other hand, a small cluster of background pixels, surrounded by object pixels should disappear quite easily with the homogenizing term. As illustrated in Fig. 11, this can be seen as a Markov chain from iteration to iteration [11, 16]:

- at iteration  $n$  the pixel has been affected a label and is considered to be in state  $x_s^{(n)}$  either background (0), or target (1),
- at iteration  $n + 1$  we can define the following transition probabilities:

$$P(x_s^{(n+1)} = 0 | x_s^{(n)} = 0) = p_{00} \quad (25)$$

$$P(x_s^{(n+1)} = 1 | x_s^{(n)} = 0) = p_{10} = 1 - p_{00} \quad (26)$$

$$P(x_s^{(n+1)} = 0 | x_s^{(n)} = 1) = p_{01} = 1 - p_{11} \quad (27)$$

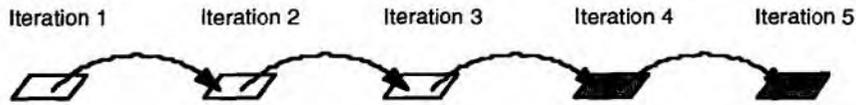
$$P(x_s^{(n+1)} = 1 | x_s^{(n)} = 1) = p_{11} \quad (28)$$

We have decided to consider that transition probabilities are determined once and do not vary as the segmentation process iterates. We have chosen  $p_{00}$  equal to 0.5

and  $p_{11}$  equal to 0.9. These values are purely heuristic; further study would be needed to determine their influence on the final result. All that can be said is that  $p_{11}$  determines, in a way, the smallest detectable object. The weighting coefficient for first order cliques is then written as:

$$c_s = [p_{00}\delta_{(x_s=0)} + p_{01}\delta_{(x_s=1)}] \delta_{(l=0)} + [p_{10}\delta_{(x_s=0)} + p_{11}\delta_{(x_s=1)}] \delta_{(l=1)} \quad (29)$$

where  $l$  is the argument of eq. (22) and  $x_s$  is the previous value of the state of  $s$ .



States of one pixel from iteration to iteration may be modeled as a Markov chain

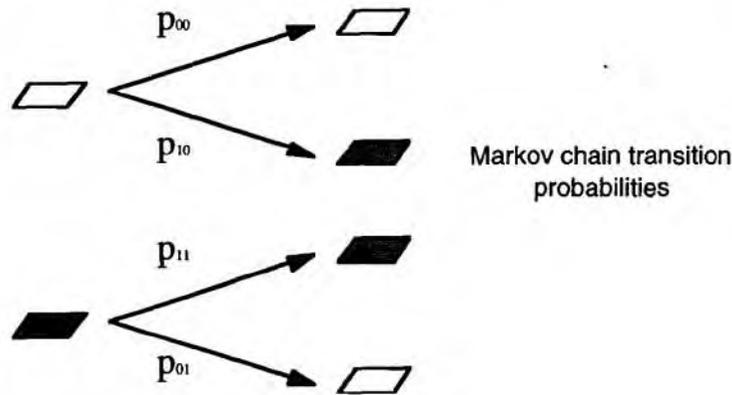


Figure 11 Markov chain process for each site

(C) Pre-classification

The only pertinent information on the objects that we can reasonably exploit is the size; either by saying it is too wide, too high, or just too big. This is simply a matter of scanning the segmented image, counting the width, height and area of each object, and declaring if it is of a reasonable size to be a submarine or not. Submarines can be as small as the Croatian modified UNA class vessel [17] measuring 22.5m long and 3.4m wide, or in a similar range the Russian Losos class submarine, the type of vessel responsible for violating Swedish waters, measuring 29.2m long by 3.9m wide. In the best case, these submarines will appear in the image as two or three pixels. On the other hand, the largest submarine, the Russian Typhoon reaches an impressive 171.5m long and 13m wide. Depending on the target's aspect angle it can, for example, at broadside be seen in as many as three beams and two time samples; the area then covered is at least six pixels. At forward or aft endfire, it could

also be detected in a single beam, and at least 28 time samples<sup>1</sup>. Pre-classification must therefore consider these type of cases; if it is based solely on area, it may then misclassify thin objects. The scheme is then to say that “round big” objects are most likely not submarines; “round big” objects being more than 3 pixels wide in beams and 3 pixels long in range. We have also checked that a point target cannot be seen in more than three beams at long ranges: so, targets which are spread over more than three beams are automatically rejected. However, even at long ranges, a target may appear in three beams owing to the spacing of the beams for the data processed — the beams were spaced so that their beampatterns overlap 3 dB down from their main response axes (MRAs) at 700 Hz. If a point target is exactly on the MRA of a beam, it will appear 6 dB down at the output of the matched filter on the adjacent beams and over 30 dB down two beams away.

### 3.3 Results and conclusions

We have tested the method on a set of ten normalised data files; note that no submarines are actually contained in these files. Each file corresponds to a ping of data acquired between 6km and 36km at 6m intervals (5000 samples). We will present a close-up of an interesting zone for one segmentation but main results will be displayed in a table.

For all segmentation results, a 3 beam by 5 range cell neighbourhood window was used, and the Page test was applied to each beam signal in order to obtain a correct initialisation for the segmentation algorithm. In Table 1, we have summarized several stages in the detection and classification process:

- for each beam, we applied a Page test detection algorithm to the beam signal and counted the number of targets detected (first column: Page test); we then had “segmented” signals for each beam
- we then constructed the image using the “segmented” signals and counted the number of detected objects (second column: Image),
- using the previous image as a segmentation initialisation, we applied the segmentation process and counted the remaining number of objects after homogenization (third column: Segmentation),
- finally, using size criteria, we discarded some objects considered too big and counted the last remaining potential targets (last column: pre-classification).

---

<sup>1</sup>Note that these estimates are particular to the sonar system in use and do not account for the spreading of the target echo in time that results from propagation through a shallow water environment.

**Table 1** *Number of potential targets detected after different stages.*

File	Page test	Image	Segmentation	Pre-classification
p001n	1751	1115	422	343
p002n	1797	1193	424	340
p003n	1672	1129	402	327
p004n	1744	1140	391	316
p005n	1689	1129	372	309
p006n	1749	1124	378	302
p007n	1709	1154	385	308
p008n	1793	1212	418	333
p008n	1798	1184	445	361
p010n	1731	1183	411	327

From this table, we see that exploiting spatial information by just laying down individual beam signals as an image already reduces the number of detections by approximately 30 to 35%. The segmentation process cleans up the initialisation and regroups detections further reducing the number of detections by 65%. Finally, the classification step, which is based only on size hypotheses, reduces the number of potential targets by approximately 22%. From the Page test to the last classification step, the number of potential targets has been indeed reduced from an average of 1750 detections to an average of 310, leaving only 18% of the potential targets.

It should be noted that a low threshold was used within the Page test and therefore many potential targets are detected. The threshold  $y_t$  employed to define the conditional probabilities (eqs. (14) to (17)) was determined by the boundary  $\alpha$  and was "compatible" with the Page test; that is, a simple thresholding of the image using  $y_t$  gave similar results. If the Page test threshold were to be changed,  $y_t$  should be modified accordingly; this was beyond the scope of this study and still needs to be investigated.

Nevertheless, with the Page test initialisation, convergence of the segmentation process was quick as can be seen in Fig. 12. The convergence rate was estimated by counting the percentage of pixels changing label from one iteration to another (flips). When this percentage was small enough, the process was considered stable and we assumed that the global minimum had been attained. Convergence occurs for most of the files after only 6 iterations; this can be explained by the fact that the initialisation of the segmentation process is very close to the optimal solution.

In the following figures, each stage of the clutter reduction process is illustrated by viewing an extract of one ping over all the beams but only 400 range samples

(2.4km). Figure 13 contains the normalized data over this region and is quite difficult to analyse visually, though some strong features are distinguishable. Figure 14 displays the Page test detection results which produce a reasonable segmentation; however, features tend to be split and the segmented image contains many detections which are in reality noise. The MAP estimation of the labels obtained from the Page test initialisation is shown in Fig. 15. Note that the image has been purged of small features believed to be noise and that detections are grouped, forming connected regions. Removal of bottom features through size discrimination (pre-classification) produces the final image shown in Fig. 16 where big features have been discarded (grey) and potential targets are depicted in black. Since pre-classification was established only by means of size, many small bottom features remain. This indicates that supplementary information such as target strength, probability distribution, target time series analysis, or association with known clutter (i.e., from a geographical database or previous pings) may allow further discrimination and a greater reduction in the number of clutter detections.

Segmentation results have enabled us to estimate histograms of characteristic features of bottom objects, such as width, height, and area. Unfortunately, most detected features are small and the histograms do not reveal any interesting information on the large bottom objects. However, study of such histograms for target size would be extremely interesting and useful.

The algorithm described in Section 3.2 was designed to preserve small target like objects; however, this ability has not been explicitly demonstrated and should be considered in future research.

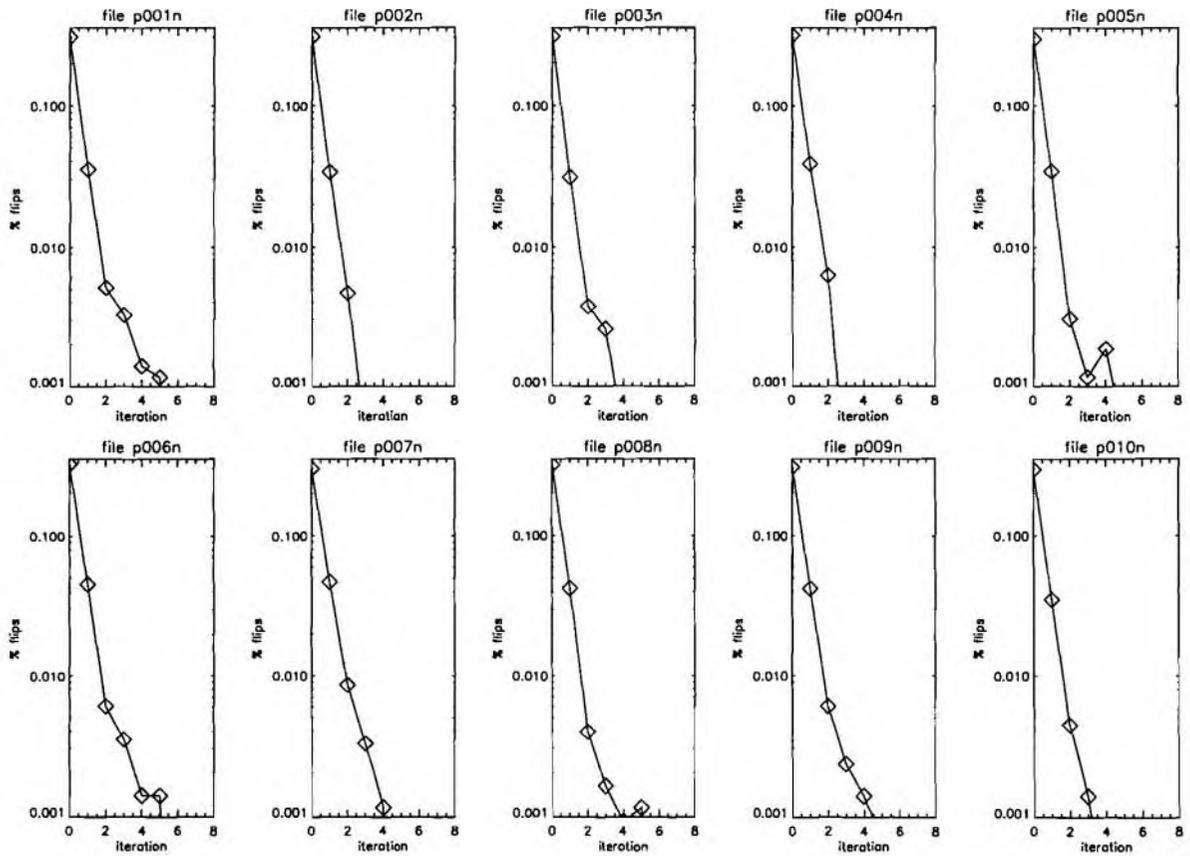


Figure 12 *Convergence rate of the MAP estimation of the image pixel labels*

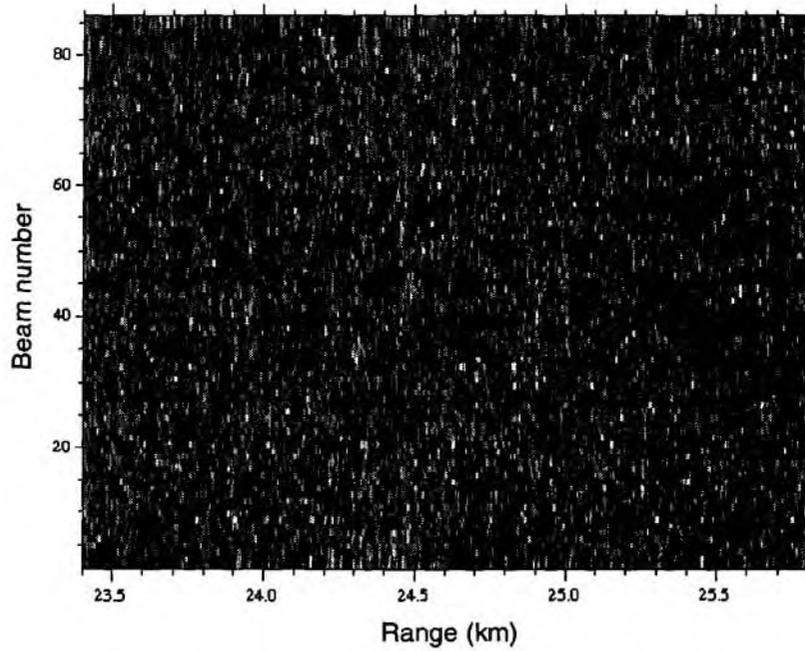


Figure 13 *Original image of the normalized sonar data*

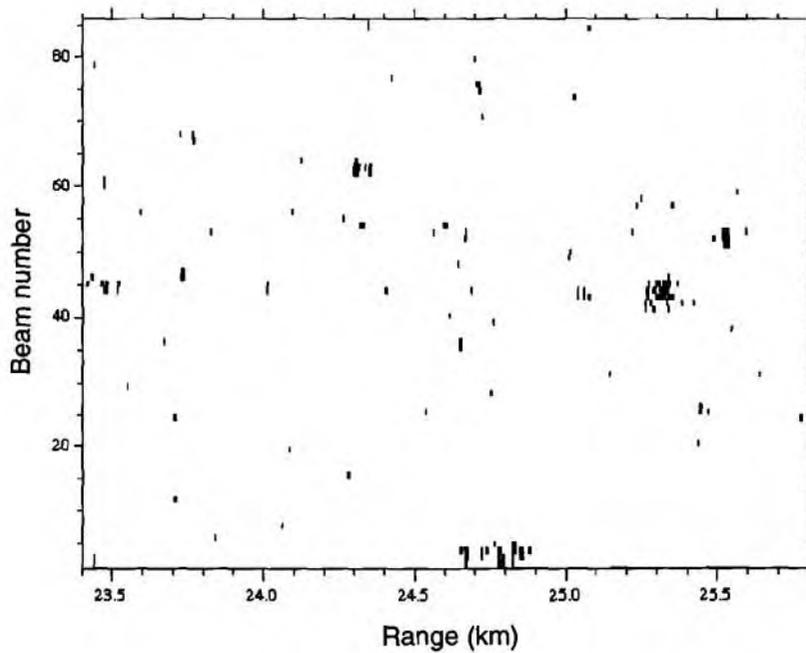


Figure 14 *Page test detections from each beam*

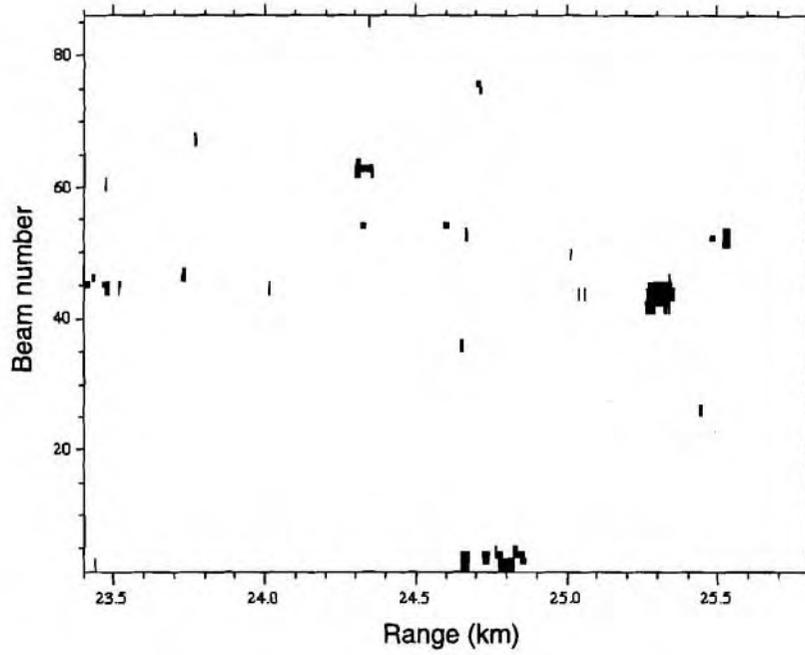


Figure 15 Segmentation of the image via MAP estimation of the image pixel labels

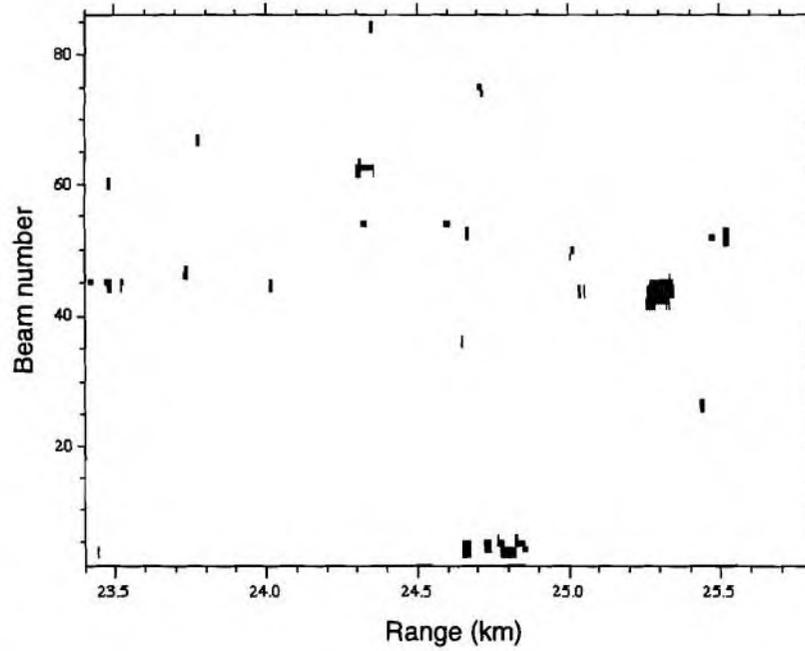


Figure 16 Segmentation and pre-classification of the image into bottom features (grey) and potential targets (black)

---

## Conclusions and prospects

---

The goal of this study was to evaluate the performance of an image processing approach to classify potential submarine detections. Clearly, this technique is particularly interesting since it will take into account spatial information discarded with current signal processing approaches. We decided to employ Markov random fields which is a theory based on particularly strong probabilistic foundations and is well suited in modelling low frequency sonar images. We attempted to introduce all knowledge of the problem may it be geometrical, physical or purely intuitive into the solution. With all this, we can obtain reliable segmentation results. The segmentation algorithm produces clean segmentations from a Page test detector initialisation in very few iterations. Convergence is indeed quick due to the good quality of the initialisation step, and a good adequation of the model to our images. This procedure also considerably reduces the number of potential targets detected.

Classification of the detected objects relies solely on size criteria; further studies should be undertaken in comparing actual images of submarines to those of bottom features. It is to be noted that no comparative study could be applied for this work since no such information was available. Nevertheless, with certain size hypotheses, results show that a good number of potential detected targets (20%) can be quite easily distinguished as bottom features and can therefore be discarded. Small bottom features will of course remain as potential targets. Additional discrimination parameters are then necessary to avoid confusion; such parameters could possibly be linked to the intensity of the object or the decay from beam to beam. The classification step can also be incorporated into the actual segmentation algorithm, either using size criteria or maybe intensity if proven to be pertinent. The performance of the classification step is highly dependent on the construction of the beams and the range resolution. A higher resolution image obtained through greater bandwidth transmit waveforms and larger aperture arrays or high resolution beamforming may improve the classification performance as better size estimation would be possible. Additionally, as the data considered in this report were beamformed such that the suppression on adjacent sidelobes was only 6 dB, reduction of the bleeding of strong targets and bottom features into adjacent beams may help in rejecting bottom features that physically extend over only two or three beams.

The greatest drawback of the MRF theory is its inappropriateness to real-time applications; the algorithm employed here was coded under IDL and performance

(time-wise) cannot be considered significantly. However, optimal coding in C and C++ [18] greatly reduce computational time; it is also to be noted that because of the particular structure of MRFs and the optimisation algorithm, it is possible to parallelise the algorithm, thus reducing again computational time.

It would also be quite interesting to test stochastic optimisation algorithms, in particular simulated annealing. This method is certainly more expensive computationally and time-wise; however, it can free itself completely from an imprecise initialisation and it can be proven theoretically that the process always converges to the optimal solution and not to a local minimum.

In the future, many different approaches can be explored, either based on MRFs or other imaging techniques:

- Detection of motion in the multiple pings in one beam image: this is in fact almost readily available. Only a closer study of diagonal cliques is necessary to ensure reliable results.
- Detection of features and motion in the temporal range bearing representation mode is probably the most promising and challenging aspect of image processing: we will cite Bouman and Liu [19] who worked with much success on segmentation of moving objects (mainly cars) in videos; their approach is based on a hierarchical Markov random field model. This approach generates segmentations of each ping image into different components: detection of objects and detection of movement from one ping to another.
- Target tracking would consist of focusing on one particular object: this could be achieved by segmentation correlations from ping to ping, motion detection and prediction. This would be especially interesting when a target is “jumping” from one beam to another.
- Correlation of segmentations from ping to ping could give information on the movement of the towed array.
- Hierarchical random fields have shown faster segmentations and more rapid homogenisation than other methods. Thus, when computational time is critical and the noise level is relatively high they may be effective.

Techniques that may be applicable, though are not considered feasible for near-term solutions include:

- Wavelets are now being used more and more in image processing. They are well known for their capabilities in image compression. Little is known about their application in image segmentation, but certainly they could be very useful

when the noise level is very high: applying wavelets means decomposing the image into frequency bands, it then being possible to reconstruct the image with less noise. However, it may be difficult to incorporate a priori information into wavelet-based algorithms.

- Neural networks could also present an alternative solution; though it would then be imperative to have more information on targets.
- Fractals have recently exploded in the image processing domain and are still an enormous new area to explore. For example, fractal dimensions may be useful in connecting sea-surface reverberation to sea-surface roughness [20] using a scatterometer. In the present case, it is not clear whether resolution would be fine enough to extract pertinent fractal parameters.

We have presented here only a few possibilities, but many more could be found while exploring different routes. However, it is to be reminded that most image segmentation algorithms always give segmentation results; the quality and reliability of the results depend on the precision with which one models the observed data. While some algorithms can integrate constraints, MRFs can incorporate much more a priori information (be it on pixel interactions, image construction or data characteristics) than any other segmentation algorithm, thus keeping a close and essential bond between the image and the physical phenomenon.

## References

---

- [1] D. A. Abraham and P. K. Willett, "Active signal detection in shallow water using the Page test," SR-252, SACLANT Undersea Research Centre, November 1996.
- [2] P. Blondel, J.C. Sempere, and V. Robjoug, "Textural analysis and structure-tracking for geological mapping: applications to sonar images from Endeavour Segmenta, Juan de Fuca ridge," *Proceedings IEEE Oceans '93*, vol. 3, pp. 208-213, 1993.
- [3] L.M. Linett, S.J. Clarke, and D.R. Carmichael, "The analysis of sidescan sonar images for seabed types and objects," *Proceedings Second European Conference on Underwater Acoustics*, vol. 2, pp. 733-738, 1994.
- [4] M. Jiang, W.K. Stewart, and M. Marra, "Segmentation of seafloor sidescan imagery using Markov random fields and neural networks," *Proceedings IEEE Oceans '93*, vol. 3, pp. 456-461, 1993.
- [5] S. Dugelay, *Caracterisation des fonds marins a partir de donnees sondeur EM12*, Ph.D. thesis, Universite Paris XI Orsay, 1997.
- [6] F. Pipelier, "Rapport de stage: Segmentation d'image sonar modelisee par des champs markoviens," Tech. Rep., Thomson-Sintra, France, 1993.
- [7] P. Thourel, *Segmentation d'images sonar par modelisation markovienne hierarchique et analyse multiresolution*, Ph.D. thesis, Universite Rennes, 1996.
- [8] B. Stage and B. Zerr, "Detection of objects on the seabottom using backscattering statistics dependant on the observation point," *IEEE Journal of Oceanic Engineering*, vol. 22, no. 1, pp. 40-46, January 1997.
- [9] P. P. Gandhi and S. A. Kassam, "Analysis of CFAR processors in nonhomogeneous background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 427-445, July 1988.
- [10] M. Fisz, *Probability Theory and Mathematical Statistics*, John Wiley & Sons, Inc., N.Y., third edition, 1963.
- [11] S.Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, Tokyo, fourth edition, 1995.
- [12] J.-P. Cocquerez and S. Philipp, *Analyse d'Images: Filtrage et Segmentation*, Masson, S.A., Paris, first edition, 1995.

- [13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, November 1984.
- [14] J.E. Besag, "On the statistical analysis of dirty pictures," *J.R. Statistic. Soc. B*, , no. 36, pp. 259–302, 1986.
- [15] M. Sigelle and R. Ronfard, "Modeles de Potts et relaxation d'images de labels par champs de Markov," *Traitement du signal*, vol. 9, no. 6, pp. 449–458, 1989.
- [16] B. Zerr, "Private communication," 1997.
- [17] Captain Richard Sharpe RN, Ed., *Jane's Fighting Ships*, Jane's Information Group Limited, Coulsdon, Surrey, U.K., one hundredth edition, 1997.
- [18] J.-M. Augustin, "Private communication," 1996.
- [19] P. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE transaction on pattern analysis and machine intelligence*, vol. 13, no. 2, pp. 99–113, 1991.
- [20] B. Chapron, "Private communication," 1995.
- [21] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc., Englewood Cliffs, NJ, first edition, 1989.
- [22] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, Inc., Canada, first edition, 1973.
- [23] A. Rosenfeld and A.C. Kak, *Digital Picture Processing*, Academic Press, Inc., Oval Road, London, second edition, 1982.

## Annex A

### A review of image processing techniques in general

---

In this section, we will review some classical feature detection methods and interesting features for classification. In each case, we will point out the advantages and drawbacks, and why the methods may or may not be suitable to the low frequency active sonar problem.

Image analysis basically involves the study of feature extraction, segmentation, and classification techniques [21]. In Fig. 17 we have listed the main methods used in each step as described by Jain [21]; more general information on image processing can be found in [11, 22, 12, 23]. It should be noted that feature extraction is a step that is performed for both segmentation and classification, often requiring different features for each task.

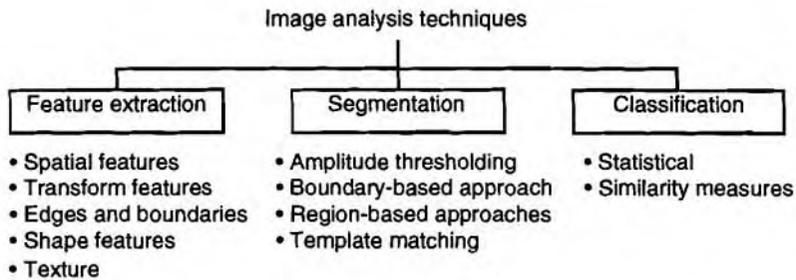


Figure 17 *Imaging techniques*

### A.1 Feature extraction

In this section, we will enumerate features describing a set of connecting pixels in an image. These features may be used in a segmentation process [12], but also

- to describe a region (shape parameters),
- to measure characteristics of an object (size, radiometry, texture),
- in post-processing, to merge regions based on similarity criteria or size,
- to classify regions,
- in an interpretation stage, to recognize an object or part of an object,

#### A.1.1 Spatial features

The most significant spatial features are [21]:

- Amplitude features: transformations can be applied to pixels in a sliding window such as averaging or median filtering. It is then possible to compute amplitude features such as averages, variances, or still inverse contrast ratio.
- Histogram features: define  $h_X$  as the histogram of the grey levels. Then, common features of  $h_X$  are its moments, absolute moments, central moments, and entropy.

In this particular case, it is recommended to avoid averaging. Indeed, due to the somewhat poor resolution of the image (6m in range), small features will tend to disappear in an averaging process, and therefore not be detected. Histograms are interesting when one can compare histograms originating from different features. In our case, only the histogram of the background is readily available before segmentation.

#### A.1.2 Transform features

Image transforms provide the frequency domain information in the data [21]. Transform features are extracted by zonal filtering (slit or aperture) the image in the selected transform space. The usual transformation is of course the Fourier transform, although other transforms such as Haar and Hadamard (wavelets) are also potentially useful. These transforms are in general interesting when texture is present.

### A.1.3 Edge features

Edge points can be defined as locations between pixels where grey levels abruptly change [12]. These points are very useful in segmentation and identification of objects in a scene. Transformations such as gradient operators (Roberts, Sobel, Kirsch), compass operators (gradient operators in a selected number of directions), Laplace operators, or stochastic gradients can be applied to the image making edge detection easier.

In the case of underwater acoustic imaging, these kinds of operators do not work well: grey levels rarely abruptly change, and the noise level in underwater applications is generally high (edge detectors are efficient when the noise level is low).

### A.1.4 Shape features

Boundaries are connected edges that characterize the shape, size and orientation of an object. Boundaries may be represented in different ways such as chain codes, fitted line segments, B-Spline representations, or control points [12, 21, 23].

Shape can also be determined by the region the object occupies. Let us define the following binary array

$$u(m, n) = \begin{cases} 1 & \text{if } (m, n) \in R \\ 0 & \text{otherwise} \end{cases} \quad (30)$$

as a simple representation of the region  $R$ .

A run-length code consists of the start address of each string of 1s, followed by the length of that string in a given direction. For each direction, we can associate a matrix of run-lengths  $P_\theta = \{p_\theta(i, j)\}$ ; the element  $p_\theta(i, j)$  of the matrix represents the number of run-lengths  $j$  pixels long (in the direction  $\theta$ ) composed of pixels with a grey level  $i$ . In a region, we can then extract different features from this dimension  $L \times n_\theta$  matrix where  $L$  is the number of grey levels in the image and  $n_\theta$  is the length of the longest code in the  $\theta$  direction in the region:

- number of run-length codes

$$SLP = \sum_{i=0}^{L-1} \sum_{j=1}^{n_\theta} p_\theta(i, j) \quad (31)$$

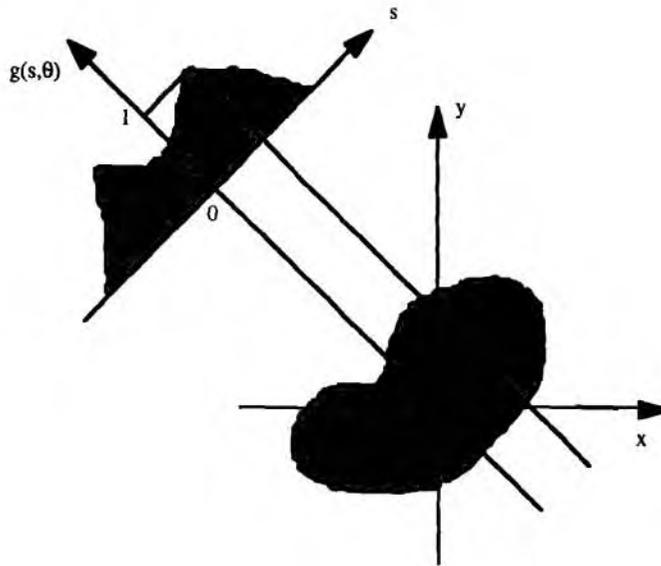
- proportion of short run-length codes

$$RF_1 = \frac{1}{SLP} \sum_{i=0}^{L-1} \sum_{j=1}^{n_\theta} \frac{p_\theta(i, j)}{j^2} \quad (32)$$

- proportion of long run-length codes

$$RF_2 = \frac{1}{SLP} \sum_{i=0}^{L-1} \sum_{j=1}^{n_\theta} j^2 p_\theta(i, j) \quad (33)$$

A projection  $g(s, \theta)$  is simply the sum of the run-lengths of 1's along a straight line oriented at angle  $\theta$  at a distance  $s$  as seen in Fig. 18. This in fact produces a histogram of the number of pixels that project into a bin at a distance  $s$  along a line of orientation  $\theta$ . From this histogram, features such as the first moments of  $g(s, 0)$  and  $g(s, \pi/2)$  will give the center of mass coordinates of the region  $R$ . We may also



**Figure 18** *Projection imaging geometry*

want to measure certain geometric attributes of a connected region  $R$  where pixels in  $R$  are described by their coordinates  $(x_i, y_i)$  such as:

- perimeter  $P(R)$ ; may be computed as the sum of the distances between consecutive pixels on a boundary (sum of 1s and  $\sqrt{2}$ ). The result is slightly different if it is estimated from inside the boundary, or outside of the region,

- area  $S(R) = K$  or number of pixels in the region,
- radii:  $R_{min}$  and  $R_{max}$  are respectively the minimum and maximum distances from the center of mass to the boundary,
- corners: these are locations on the boundary where the curvature becomes unbounded,
- bending energy: this is still another attribute associated with the curvature [21],
- roundness:  $\gamma = 4\pi S(R)/P^2(R)$ , for a disc,  $\gamma$  is minimal and is equal to 1,
- symmetry: commonly, there are two types of symmetry of shapes, rotational and mirror. Distances from the center of mass to different points on the boundary can be used to analyse symmetry of shapes.

Or, still we may compute moment-based features:

- center of mass:  $x_m = \frac{1}{K} \sum_{i=1}^K x_i$  and  $y_m = \frac{1}{K} \sum_{i=1}^K y_i$
- orientation: first eigenvector of the inertia matrix  $\begin{pmatrix} a & c \\ c & b \end{pmatrix}$  with

$$a = \frac{1}{K} \sum_{i=1}^K (x_i - x_m)^2 = \frac{1}{K} \sum_{i=1}^K x_i^2 - x_m^2 \quad (34)$$

$$b = \frac{1}{K} \sum_{i=1}^K (y_i - y_m)^2 = \frac{1}{K} \sum_{i=1}^K y_i^2 - y_m^2 \quad (35)$$

$$c = \frac{1}{K} \sum_{i=1}^K (x_i - x_m)(y_i - y_m) = \frac{1}{K} \sum_{i=1}^K x_i y_i - x_m y_m \quad (36)$$

Principal orientation of the object is defined by its angle  $\alpha$  with the  $x$  axis:

$$\tan 2\alpha = \frac{2c}{a - b} \quad (37)$$

- bounding rectangle: rectangle whose sides are parallel to the eigenvectors of the inertia matrix,

- best fit ellipse: the second order moments of the best fit ellipse are equated to those of the object. Let  $a$  and  $b$  denote the semimajor and semiminor axes of the best fit ellipse. The least and greatest moments of inertia for an ellipse are:

$$I_{min} = \frac{\pi}{4}ab^3, \quad I_{max} = \frac{\pi}{4}a^3b \quad (38)$$

For orientation  $\theta$ , the above moments can be calculated as

$$I'_{min} = \sum_{(x_i, y_i) \in R} [(x_i - x_m)\cos\theta - (y_i - y_m)\sin\theta]^2 \quad (39)$$

$$I'_{max} = \sum_{(x_i, y_i) \in R} [(x_i - x_m)\sin\theta - (y_i - y_m)\cos\theta]^2 \quad (40)$$

Resolving  $I'_{min} = I_{min}$  and  $I'_{max} = I_{max}$  gives us

$$a = \left(\frac{4}{\pi}\right)^{1/4} \left(\frac{(I'_{min})^3}{I'_{min}}\right)^{1/8}, \quad b = \left(\frac{4}{\pi}\right)^{1/4} \left(\frac{(I'_{max})^3}{I'_{max}}\right)^{1/8} \quad (41)$$

- eccentricity:  $\epsilon = \frac{(b-a)^2 + 4c}{S(R)}$

These features will be indispensable when classifying: for example, the area or bounding rectangle of the detected object in the single ping image will enable us to rule out objects which are too big to be submarines. We may also want to check if objects in the pings-per-beam representation are considered stationary or not; in this case, orientation may be a good descriptor.

#### A.1.5 Texture

Texture describes the structural pattern of a surface such as wood, sand, grass, or cloth. The term texture is generally used to describe a repetition of a basic texture element containing several pixels [12]. Textural features can be computed to describe the texture such as:

- autocorrelation function
- entropy
- log contrast
- Haralick texture parameters

In practise, texture (in the imaging sense), is very hard to define and may vary from person to person depending on their perception; it is then not surprising to find several definitions of contrast or homogeneity for example that tend to give us a mathematical definition of the same visual phenomenon. However, it is definite that in the case of the low frequency sonar images obtained for the current application, no texture is present. Texture can be correlated to frequency and resolution: the higher the frequency, generally the higher the bandwidth (and thus resolution), the smaller the insonified area, and in the case of seafloor sounding, the less the penetration. The signal is then of "finer" quality and texture will then describe fine interface details such as ripples or dunes.

## *A.2 Image segmentation*

Image segmentation consists of decomposing a scene into its components. The image segmentation techniques listed in Fig. 17 are briefly described in the following subsections. The reader is referred to [12, 21] for further detail.

### *A.2.1 Amplitude thresholding or window slicing*

Thresholding is particularly useful when pixel amplitudes are sufficient to characterize the object [21]. The most delicate step is choosing the threshold. Commonly, several approaches are used:

1. Select the threshold from a histogram of the image.
2. Select the threshold so that a predetermined number of samples are below the threshold.
3. Adaptively choose the threshold from local neighbourhood histograms.
4. If a probabilistic model is known for different classes, select the threshold to minimize the probability of error.

In the present application, thresholding is an excellent and quick solution in a first approach for segmentation. Indeed, the amplitude of the pixel is the most characteristic feature in discriminating between background and object. However, if such a technique were to be applied alone, the obtained segmented image would be very noisy. It is therefore interesting to apply this method combined with a homogenizing process.

#### A.2.2 *Boundary-based approaches*

Segmentation can be achieved by using boundary information, either by contour following, connectivity, edge linking, graph searching, curve fitting, Hough transform, or Markov random fields. This is however rather difficult when objects are touching, or overlapping, or if a break occurs in the boundary due to noise or artifacts [21].

#### A.2.3 *Region-based approaches and clustering*

The main idea in this case is to identify regions that have similar features. Hence, clustering techniques (pixel grouping) are well adapted and can be applied for image segmentation [12, 21].

One example is the region growing algorithm that merges similar regions together unless they are sufficiently different. The difficulty lies in defining "sufficiently different". We may propose several heuristic criteria:

1. Merge two regions together if their similarity distance is less than a threshold.
2. Merge two regions together if there are no strong edge points between them.
3. Merge two regions together if the number of weak boundary points over the length of the common boundary is greater than a threshold. We define weak boundary points as locations where pixels on either side have a magnitude difference less than a threshold.

#### A.2.4 *Template matching*

Objects can be detected by template matching [21]. This means that we know specifically the size and shape of the objects which are to be detected. In our case, no specific size or shape may be defined a priori. This is then an impossible approach.

#### A.3 *Classification*

Classification is a phase which in reality is very close to feature extraction. We will use object descriptors to give them a "name".

### A.3.1 Statistical

It is possible to classify an object based on statistical descriptors such as pdf of the pixel intensities comprised in the object, or statistics such as mean, variance or textural parameters. This supposes we have the same kind of a priori information on the different classes which are present in the image. This information is either learned in a training step before segmentation, or we use theoretical distributions. By applying statistical tests (such as Kolmogorov-Smirnov) it is then possible to classify the object as in a class or not.

### A.3.2 Similarity measures

Similarity measures will choose the most similar group to classify the object: either based on the mean intensity, shape features, texture, etc. For any two feature vectors  $x_i$  and  $x_j$  some of the commonly used similarity measures are:

$$\text{Dot product: } \langle x_i, x_j \rangle = x_i^T x_j = \|x_i\| \|x_j\| \cos(x_i, x_j)$$

$$\text{Similarity rule: } S(x_i, x_j) = \frac{\langle x_i, x_j \rangle}{\langle x_i, x_i \rangle + \langle x_j, x_j \rangle - \langle x_i, x_j \rangle}$$

$$\text{Weighted Euclidean distance: } d(x_i, x_j) = \sum_k [x_i(k) - x_j(k)]^2 w_k$$

$$\text{Normalized correlation: } \rho(x_i, x_j) = \langle x_i, x_j \rangle / \sqrt{\langle x_i, x_i \rangle \langle x_j, x_j \rangle}$$

Report no. changed (Mar 2006): SR-272-UU

Initial Distribution for SR 272

*Ministries of Defence*

DND Canada	10
CHOD Denmark	8
DGA France	8
MOD Germany	15
HNDGS Greece	12
MARISTAT Italy	9
MOD (Navy) Netherlands	12
NDRE Norway	10
MOD Portugal	5
MDN Spain	2
TDKK and DNHO Turkey	5
MOD UK	20
ONR USA	42

*Scientific Committee of National Representatives*

SCNR Belgium	1
SCNR Canada	1
SCNR Denmark	1
SCNR Germany	1
SCNR Greece	1
SCNR Italy	1
SCNR Netherlands	2
SCNR Norway	1
SCNR Portugal	1
SCNR Spain	1
SCNR Turkey	1
SCNR UK	1
SCNR USA	2
French Delegate	1
SECGEN Rep. SCNR	1
NAMILCOM Rep. SCNR	1

*NATO Commands and Agencies*

NAMILCOM	2
SACLANT	3
CINCEASTLANT/	
COMNAVNORTHWEST	1
CINCIBERLANT	1
CINCWESTLANT	1
COMASWSTRIKFOR	1
COMMAIREASTLANT	1
COMSTRIKFLTANT	1
COMSUBACLANT	1
SACLANTREPEUR	1
SACEUR	2
CINCNORTHWEST	1
CINC SOUTH	1
COMEDCENT	1
COMMARAIMED	1
COMNAVSOUTH	1
COMSTRIKFOR SOUTH	1
COMSUBMED	1
NC3A	1
PAT	1

*National Liaison Officers*

NLO Canada	1
NLO Denmark	1
NLO Germany	1
NLO Italy	1
NLO Netherlands	1
NLO Spain	1
NLO UK	1
NLO USA	1

**Sub-total** 208

SACLANTCEN 30

**Total** 238