Image Processing Techniques for Sector Scan Sonar

by

Lukas Anton Hendriks

Thesis presented at the University of Stellenbosch in partial fulfilment of the requirements for the degree of

Master of Science in Engineering

Department of Electrical Engineering
University of Stellenbosch
Private Bag X1, 7602 Matieland, South Africa

Study leader: Mr J Treurnicht

October 2009

Declaration

By Submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicity otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 23 November 2009

Abstract

Sonars are used extensively for underwater sensing and recent advances in forward-looking imaging sonar have made this type of sonar an appropriate choice for use on Autonomous Underwater Vehicles. The images received from these sonar do however, tend to be noisy and when used in shallow water contain strong bottom reflections that obscure returns from actual targets.

The focus of this work was the investigation and development of post-processing techniques to enable the successful use of the sonar images for automated navigation.

The use of standard image processing techniques for noise reduction and background estimation, were evaluated on sonar images with varying amounts of noise, as well as on a set of images taken from an AUV in a harbour.

The use of multiple background removal and noise reduction techniques on a single image was also investigated. To this end a performance measure was developed, based on the dynamic range found in the image and the uniformity of returned targets. This provided a means to quantitatively compare sets of post-processing techniques and identify the "optimal" processing.

The resultant images showed great improvement in the visibility of target areas and the proposed techniques can significantly improve the chances of correct target extraction.

Uittreksel

Sonars word algemeen gebruik as onderwater sensors. Onlangse ontwikkelings in vooruit-kykende sonars, maak hierdie tipe sonar 'n goeie keuse vir die gebruik op 'n Outomatiese Onderwater Voertuig. Die beelde wat ontvang word vanaf hierdie sonar neig om egter raserig te wees, en wanneer dit in vlak water gebruik word toon dit sterk bodemrefleksies, wat die weerkaatsings van regte teikens verduister.

Die fokus van die werk was die ondersoek en ontwikkeling van naverwerkings tegnieke, wat die sonar beelde bruikbaar maak vir outomatiese navigasie.

Die gebruik van standaard beeldverwerkingstegnieke vir ruis-onderdrukking en agtergrond beraming, is geëvalueer aan die hand van sonar beelde met verskillende hoeveelhede ruis, asook aan die hand van 'n stel beelde wat in 'n hawe geneem is.

Verdere ondersoek is ingestel na die gebruik van meer as een agtergrond beramings en ruis onderdrukking tegniek op 'n enkele beeld. Hierdie het gelei tot die ontwikkeling van 'n maatstaf vir werkverrigting van toegepaste tegnieke. Hierdie maatstaf gee 'n kwantitatiewe waardering van die verbetering op die oorspronklike beeld, en is gebaseer op die verbetering in dinamiese bereik in die beeld en die uniformiteit van die teiken se weerkaatsing. Hierdie maatstaf is gebruik vir die vergelyking van verskeie tegnieke, en identifisering van die "optimale" verwerking.

Die verwerkte beelde het 'n groot verbetering getoon in die sigbaarheid van teikens, en die voorgestelde tegnieke kan 'n betekenisvolle bedrae

lewer tot die suksesvolle identifisering van obstruksies.

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations who have contributed to making this work possible:

- IMT for funding the project
- Mr J Treurnicht for his help and guidance
- My family and friends for their support

Dedications

Hierdie tesis word opgedra aan my ouers,
vir hul ondersteuning, geduld, moed inpraat en liefde.

Contents

List of Figures

List of Tables

Nomenclature

Small Letters

c	Speed of sound in water

Capital Letters

ΔD	Mean dynamic range
ΔM	Maximum negative deviation
$\mathcal{P}$	Performance measure
T	Threshold
T_p	Pulse length

Acronyms

AUV	Autonomous Underwater Vehicle
FFT	Fast Fourier Transform
IMT	Institute for Marine Technology

Chapter 1

Introduction

1.1 Background and Motivation

This project is part of a greater research initiative into the automation of mine hunting for the purpose of harbour safety. The research is focussed on the development of the necessary subsystems to enable an *Autonomous Underwater Vehicle* (AUV) to successfully navigate and search for suspicious objects in a harbour environment, without the need for direct human control. To enable complete automation, there are a few critical hurdles that need to be overcome, one of which is the need for obstacle avoidance.

Whilst surveying a given area, the AUV may have multiple possible paths it can take from one point to the next. It is however necessary to ensure that these paths are safe, and not obstructed in some way. This requires the use of the appropriate sensors and the development of the necessary algorithms to interpret the data from these sensors. As such, this project was proposed as an initial investigation into the sensors commonly used for collision avoidance, and the signal processing algorithms required to enable their automated use.

This project was funded and run in conjunction with the Institute of

Maritime Technology (IMT), who have been very forthcoming in sharing their knowledge and infrastructure whenever it was needed. To help steer the project in the right direction, they provided a commercially available sector scan sonar, which is representative of the sensors commonly used for collision avoidance.

The sonar output is a set of intensity values, that is a representation of the amount of reflected energy from a specific direction at increasing range units. This data is used to create a bitmap image of the surveyed area. These images are a rather crude representation of the surroundings, but they do provide an adequate means for an operator to identify obstacles.

The focus of the project is the identification of important parameters of this specific sonar and the post-processing required to enable the use of the output images for automated navigation.

The images have low resolution and as such it is very difficult to differentiate between the features found in objects and those found in the background. The only real way of differentiating between the two is the variation in local intensities. This is something that is easily done by a human observer, but can require considerable further processing to enable an automated detection. This is illustrated in fig. 1.1, which shows

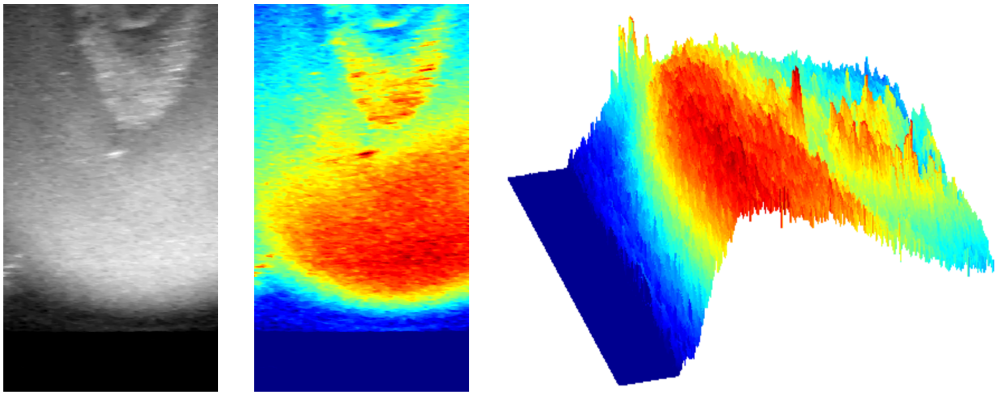


Figure 1.1: An example of the same sonar image displayed in three different ways.

three different representations of the same sonar intensity map.

The first image shows the intensity values as a grey-scale image, and the reflection from a target is visible as a "V" shape in the top half of the image. When compared to the second image, where the intensities are mapped to colours, both the object and background suddenly appear less uniform, with the variations in both becoming more apparent. On further inspection of the three-dimensional image, the variation in both the background and the object become even more pronounced. Presenting the data in varying ways circumvents the pattern recognition inherent to human observation and gives slightly more insight into the difficulties found with these images.

From this initial inspection it is already clear that the non-uniformity of both the background and the object will make it very difficult to identify the edge of an object as a mere jump in intensity. Furthermore, since the object and background contain similar intensity values, an object cannot immediately be identified based solely on its intensity range. Improving the uniformity of objects and accentuating the object features in order to make them easily discernible from the background, will be of great benefit. Alternatively, suppressing the background should have a similar result.

From this, the first objective is the investigation of the imaging process and identification of parameters that influence the resulting images. Investigation of these parameters could lead to some improvement of the image quality, or at least help identify areas that could benefit from further processing.

Second, is the identification of appropriate image processing techniques. Because of the lack of detail found in theses images, the focus is on the use of basic image processing tools to accentuate the object features. Especially those used for *noise reduction*, *thresholding* and *background removal*, in both the spatial and frequency domain, are relevant.

1.2 Thesis Organisation

Chapter two provides the relevant background on digital image processing needed for the discussions in later chapters. The chapter starts by discussing the basic principles of digital images, and then goes on to discuss the implementation and purpose of various techniques.

Chapter three is an introduction to sonar technology. Specifically on the sector scan sonar that was used in the research. The chapter starts with an overview of the basic principles of sonar and the differences between various types of sonar. This is followed by a discussion on the characteristics specific to the sector scan sonar and how this relates to the output images. The chapter concludes with a review of previous work done using the images from a sector scan sonar.

In *chapter four* a detailed analysis is undertaken of the influence of certain sonar properties on the resulting images. The chapter also introduces two real world data sets and identifies how they may benefit from further processing.

Chapter five provides a detailed discussion on the proposed image processing techniques and how they influence the images. This evaluation focusses on noise reduction and improvement of uniformity in the images, as well as identifying techniques that can be used to successfully remove or suppress bottom reflections.

In *chapter six* the proposed techniques are further scrutinised and evaluated according to a proposed quantitative performance measure. The way that the different techniques interact and how they can be used together are also evaluated.

Chapter 2

Image Processing Background

As a precursor to a discussion on sonar, this chapter provides an introduction to some of the basic principles of digital image processing. Many of the techniques discussed here are encountered in the literature on previous work done with sonar, where a working knowledge of image processing is assumed. As such, the principles and techniques presented in this chapter are discussed in general in order to provide a background to later work.

2.1 Digital Image Fundamentals

In general, the imaging process can be described as *the combination of an "illumination" source and the reflection or absorption of energy from that source by the elements of the "scene"* [3]. Although the specifics of an imaging process may vary, such as the reflection of visible light from a three-dimensional scene as used in photography, or the reflections of electromagnetic waves in radar, this basic principle holds true. Regardless of how an image is acquired, the objective is to generate digital images from the sensed data [3].

In most cases the output from a sensor is a continuous voltage or cur-

rent waveform related to the physical phenomenon being sensed, which then needs to be converted to digital form. In order to create a digital representation of the image space, it is necessary to convert both the spatial and amplitude information contained in this waveform to discrete values. To this end, *sampling* refers to the digitisation in the spatial domain and *quantisation* is the digitisation of amplitude values. A digital image can formally be described as *the sampled and quantised representation of a continuous image field*.

In order to further explain this process, let $g(x, y)$ denote a continuous, infinite-extent, ideal image field representing, for example, the reflected intensity values from an arbitrary object. In a perfect image sampling system, samples of the ideal image could be obtained by multiplying the ideal image with a spatial sampling function

$$s(x, y) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \delta(x - k\Delta x, y - l\Delta y) \tag{2.1.1}$$

composed of an infinite array of Dirac delta functions arranged in a grid of spacing $(\Delta x, \Delta y)$. The sampled image can then be represented as

$$f(k, l) = g(k\Delta x, l\Delta y)s(x, y) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} g(k\Delta x, l\Delta y)\delta(x - k\Delta x, y - l\Delta y), \tag{2.1.2}$$

where it is observed that $g(x, y)$ may be brought into the summation and evaluated only at the sample points $(k\Delta x, l\Delta y)$ [4]. Each of these continuous intensity values, $f(k, l)$, is then quantised into a discrete value.

This digitisation process leaves us with a matrix of discrete values. As a convention, this resultant matrix has M rows and N columns, with the value at each coordinate corresponding to a single pixel value. The complete $M \times N$ digital image can be written in the following matrix

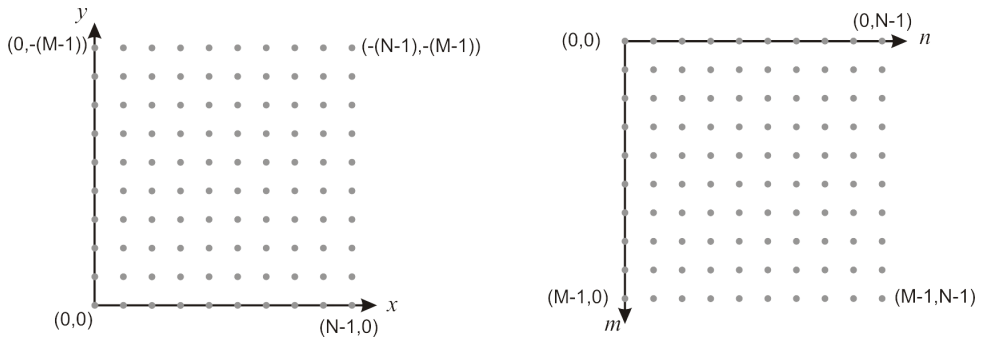


Figure 2.1: Comparison of Cartesian and image space

form:

$$f(m,n) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix} \quad (2.1.3)$$

It should be noted that $f(m,n)$ is used to define the discrete image space, and $f(x,y)$ continuous space. Unfortunately this is not a standard convention throughout the image processing literature and $f(x,y)$ is used for discrete space in some cases [3] [4]. To avoid confusion with Cartesian space, and to implicitly show the difference between continuous and discrete space, m and n will be used for the two-dimensional axis. It should also be noted that the image space is a flip of the Cartesian definition, as in Fig. 2.1. Since the Cartesian convention is a common understanding of two-dimensional space, this difference can cause some confusion, especially where the distances between pixels need to be calculated.

2.2 Noise Reduction

The presence of noise is an undeniable reality of any imaging procedure and can be added by a variety of sources, such as the medium through which the images is taken or the electronics of the sensing equipment. Regardless of how the noise enters the system it is always an unwanted byproduct. Since this is a common problem found in image processing, a great number of techniques have been developed for the reduction of noise. In most cases it is however difficult to quantify or assign a number to what constitutes an improvement. The procedures are therefore mainly judged through inspection and requires tweaking of parameters to obtain the best results.

2.2.1 Image Averaging

A simple way to reduce the noise of an image is to take the average of a number of images of the same scene. To illustrate this process, let $g_i(x, y)$ be the ith image taken of the same original scene $f(x, y)$ corrupted by additive noise $\eta_i(x, y)$; that is,

$$g_i(x, y) = f(x, y) + \eta_i(x, y). \tag{2.2.1}$$

Assuming that the noise at all coordinates (x, y) is uncorrelated and the scene, $f(x, y)$, does not change, averaging of $g_i(x, y)$ over a sequence of P images gives the relation

$$f(x, y) = \frac{1}{P}\sum_{i=0}^{P-1} g_i(x, y) - \frac{1}{P}\sum_{i=0}^{P-1} \eta_i(x, y) \tag{2.2.2}$$

The noise term will tend towards its ensemble average $E\{\eta(x, y)\}$ for large values of P. Assuming zero-mean Gaussian noise, the ensemble average is *zero* at all (x, y) [4]. From this the original image can be estimated

as

$$\hat{f}(x,y) = \frac{1}{M} \sum_{i=0}^{M-1} g_i(x,y) \tag{2.2.3}$$

which in the ideal case means that the noise has been completely removed.

Although the theory suggests a large number of images are needed to completely remove the noise, in practise however, a marked reduction can already be noticed after averaging only two or three images, making this a very practical solution.

2.2.2 Spatial Filtering

Filtering in the spatial domain refers to operations performed with the values of neighbouring pixels and the corresponding values of a matrix with the same size as the neighbourhood. This matrix is commonly called a *mask*, *kernel* or *window* [3]. It is also, more often than not, referred to as a *filter*, which is originally associated with a function in the frequency domain. As such, it is commonly called a *spatial filter*, in order to make a clear distinction.

A general description of a linear spatial filter, can be given by

$$g(m,n) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(m+s, n+t) \tag{2.2.4}$$

where f is an image of size $M \times N$, w is the filter mask of size $M_w \times N_w$ and $a = (M_w - 1)/2$ and $b = (N_w - 1)/2$. The process consists of moving the filtering mask from pixel to pixel and calculating the *response* to the filter at each pixel (x, y), as illustrated in Fig. 2.2.

Nonlinear spatial filters are also implemented by sliding a mask past the image. In most cases, however, the operation is based on the specific values of the underlying neighbourhood and do not explicitly use the

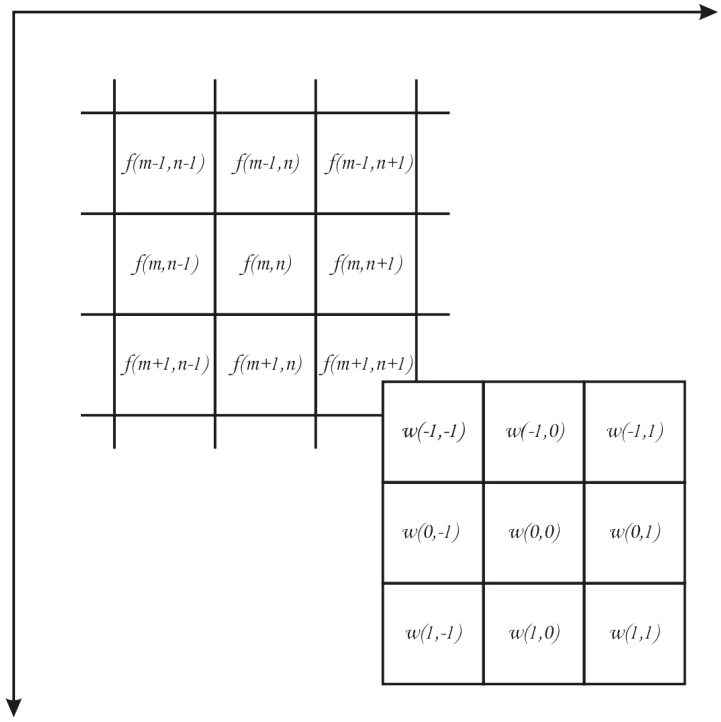


Figure 2.2: The mechanics of spatial filtering. The figure shows a 3×3 mask and the underlying image section

coefficients in the sum-of-products manner of Eq. 2.2.4 [3].

As a practical note, care should be taken when spatial filters are evaluated at the edges of an image. For example, when a 3×3 mask is applied to an image as in Fig. 2.2, and evaluated at $(0, 0)$, the window would be partly outside the image space. One approach is to rectify this problem through *padding* the outer edges of the image. As in signal processing, this padding can be realised by adding zeros, but this results in a somewhat darkened outer edge. A better solution is to mirror the values of the image around the edges. This adds no new information to the image and does not create distorted edges. This is similar to ignoring the values where the window and the image do not overlap, but it is much simpler to implement.

2.2.2.1 Averaging Filters

An averaging filter is a linear spatial filter, as described in the previous section. The output of the filter is the weighted average of the pixels under the mask. It is also commonly referred to as a smoothing or low-pass filter, referring to the blurring or smudging effect of the filter. It is for this reason that this type of filter is commonly used to reduce certain types of noise, which manifests as sudden changes in intensity. This does however have a detrimental effect on any sharp edges that the image may contain, which commonly sets a limitation on the size of the kernel. Apart from noise, the filter can also be used to get rid of false contours or image artifacts.

With reference to Eq. 2.2.4 a general expression for a weighted average filter can be given as

$$g(m,n) = \frac{\sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(m+s, n+t)}{\sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t)} \tag{2.2.5}$$

Since the denominator is the sum of the filter coefficients, it need only be calculated once [3].

2.2.2.2 Median Filter

The median filter is the most common implementation of order-statistic filters, which rely on the order or ranking of the pixels in the encompassed image area. The median is calculated by sorting the values from largest to smallest, or vice-versa, and picking the middle value. As such it is a nonlinear filter, and cannot be described by simple summation.

Median filters are very effective at reducing impulse or salt-and-pepper noise that cause intensity spikes, without the blurring caused by linear

smoothing filters. As such this filter is very popular and in most cases produce better results than simpler linear filters. It does however tend to have greater computational complexity than linear spatial filters and as such require careful implementation when efficiency is of concern.

2.2.3 Filtering in the Frequency Domain

In most applications of signal processing, including image processing, the Fourier transform and the frequency domain is a fundamental and invaluable tool. It was therefore thought prudent to highlight some of its properties.

The two-dimensional discrete Fourier transform (DFT) of a function, of size $M \times N$, is given by the equation

$$F(u,v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) e^{-j2\pi(um/M+vn/N)} \tag{2.2.6}$$

This expression must be evaluated at all $u = \{0, 1, 2, \ldots, M-1\}$ and $v = \{0, 1, 2, \ldots, N-1\}$. Similarly the inverse transform is given as

$$f(m,n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(um/M+vn/N)} \tag{2.2.7}$$

The two-dimensional DFT retains all the properties of the one-dimensional version, such as its periodic and symmetrical nature. For a full discussion on the Fourier transform and its properties see [4], [3] or [5].

One of the most important properties of the Fourier transform and an important relationship between the frequency and spatial domain is the *convolution theorem*. The discrete convolution of two functions $f(m,n)$ and $h(m,n)$ is denoted by $f(m,n) * h(m,n)$ and is defined by the expres-

sion

$$f(m,n) * h(m,n) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f(k,l)h(m-k,n-l) \tag{2.2.8}$$

It should be noted that this is very similar in form to Eq. 2.2.4 and results in the same window shifting operation. As such, linear spatial filtering is merely an implementation of the convolution operator [4] [3].

Letting $F(u,v)$ and $H(u,v)$ denote the Fourier transforms of $f(m,n)$ and $h(m,n)$ it can be shown that $f(m,n) * h(m,n)$ and $F(u,v)H(u,v)$ form a Fourier transform pair [3] [5]. This result can formally be stated as

$$f(m,n) * h(m,n) \Leftrightarrow F(u,v)H(u,v) \tag{2.2.9}$$

or

$$F(u,v)H(u,v) \Leftrightarrow f(m,n) * h(m,n) \tag{2.2.10}$$

where the double arrow is used to indicate the *forward* and *inverse* Fourier transform.

From this theorem it should be apparent that linear filtering in the spatial and frequency domains are closely related, or rather, a filter in the spatial domain always has an analogue in the frequency domain, and vice-versa. An important result of the convolution theorem is that local effects in one domain constitutes global effects in the other, since convolution, which is dependent on local information, results in a global multiplication.

The power of the Fourier transform is that it can be used to identify and remove global effects, which tend to be very difficult when only using local information. As a result, the Fourier transform can be used as an effective way of combatting periodic noise, which show up as peaks in the frequency domain. By then suppressing or removing these peaks

the noise can be near completely eliminated without disturbing the rest of the image data. Furthermore, noise effects tend to be confined to the higher frequencies, and depending on the noise and image information, a low pass filter can be used to isolate the image information and remove the noise.

2.3 Image Segmentation

Segmentation is the decomposition of an image into regions or objects and is an important part of image analysis, especially where it is part of an automated system. The human mind is a excellent pattern recognition system, especially to visual input, and is easily able to discern different objects in images. Automation of this process is however one of the most difficult tasks in image processing.

2.3.1 Thresholding

A common way to discern objects from background information is the difference in their intensity or amplitude range. A *threshold* value is then set to separate the background and object information.

Fig. 2.3 gives two examples of objects contained in background information. The respective histograms are given to show the distribution of pixel values throughout the image. In both images the objects are easily discernible through inspection, but the histograms can be misleading. In both cases the histogram shows an apparent division or valley between two peaks; one in both the darker and lighter part of the spectrum. It should however be apparent that only Fig. 2.3(a) can be segmented by setting a simple threshold in the middle of the two peaks. In Fig. 2.3(b) this type of thresholding would add parts of the background to the object and vice-versa, as shown in the thresholded images.

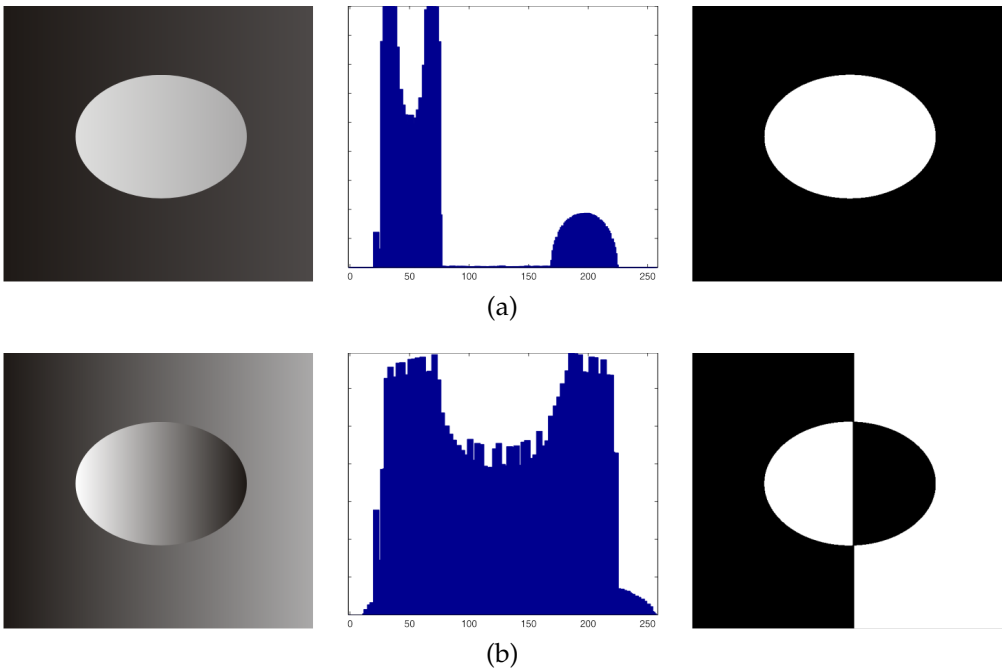


Figure 2.3: Two examples of an object on a background, along with their respective histograms and thresholded representations.

2.3.1.1 Basic Thresholding

The simplest form of thresholding is to divide the image histogram using a single global threshold, T. Where the histogram is easily separable, this type of global threshold is more than adequate.

The easiest way to determine the threshold is through inspection of the histogram. Although this is rarely the optimal threshold, it can be an adequate solution, especially where the input images vary very little.

In [3], an automatic selection of the threshold is suggested using the following procedure:

1. Select an initial estimate for T. If no information is available, the average image value might be a good initial estimate.

2. Segment the image using T. This will produce two groups of pixels: G_1 consisting of all pixels with values $> T$ and G_2 with values $\leq T$.

3. Compute the average grey level μ_1 and μ_2 for each region.

4. Compute a new threshold value:

$$T = \frac{1}{2}(\mu_1 + \mu_2) \tag{2.3.1}$$

5. Repeat steps 2 through 4 until difference in T in successive iterations is less than a predefined parameter T_0.

In [6] a similar procedure is discussed, with the only real difference that μ_2 in Eq. 2.3.1 is replaced with T^{t-1}, the threshold from the previous iteration.

In many cases such as Fig. 2.3(b) a global threshold does not segment the background and objects adequately. An approach for handling this situation is to divide the image into subimages and individually finding a appropriate threshold. This process is known as *adaptive thresholding*, since the threshold used for each pixels is dependent on the local statistics [3]. The window or subimage size is the main concern when using adaptive thresholding in this way. For more accurate local threshold estimation, a window can be moved from pixel to pixel, similar to spatial filtering, and the threshold calculated for each instance.

2.3.1.2 Optimal Thresholding

A threshold can be considered optimal when it is estimated in such a way that it produces a minimum average segmentation error. When using this type of method, it is assumed that the image contains two intensity regions or modes. Each mode can be considered as a random quantity and their histogram is used as an estimate of their probability density function (PDF). Fig. 2.4 shows two PDFs, where z is intensity and $p(z)$ is the PDF of the whole image. $p(z)$ can be described as

$$p(z) = P_1 p_1(z) + P_2 p_2(z) \tag{2.3.2}$$

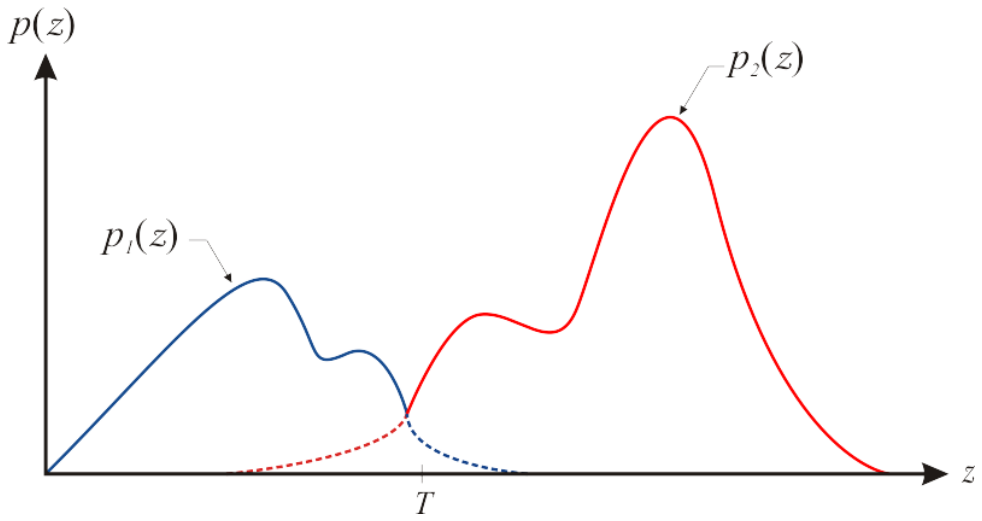


Figure 2.4: Probability density functions of two regions in an image.

where P_1 and P_2 are the probability of occurrence of each mode. The threshold T is used to separate the two modes. Using this threshold and the probabilities of each mode occurring, the overall probability of an error occurring is given by

$$E(T) = P_1 \int_T^{\infty} p_1(z)\, dz + P_2 \int_{-\infty}^{T} p_2(z)\, dz \tag{2.3.3}$$

where each integral represents the probability that a value will occur that will be classified wrongly [3].

Although the methods might differ greatly all optimal techniques are in essence trying to find a threshold value T that minimises Eq. 2.3.3.

2.3.2 Region Growing

Region growing is a conceptually very simple method of segmenting an image which easily groups together regions of pixels with similar characteristics. The procedure is started by selecting a seed point, from where a region of similar pixels are *grown* by adding connected pixels [4] [3] [7].

0	0	0	0	5	6	7	8
0	0	0	0	6	7	9	10
5	7	0	0	0	8	10	10
6	6	0	0	0	9	9	8

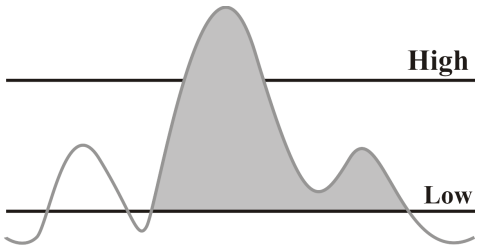


Figure 2.5: Example of how region growing and thresholding can be used to complement each other

The first complication is the selection of seed points. If no information about possible regions exists, the procedure is started at an arbitrary location and stops when all the pixels have been evaluated. Secondly constraints need to be placed on the growing, or to what constitutes *similar* pixels. These constraints may become somewhat complex, but are usually determined for each application.

An example of how this process can be used is given by Fig. 2.5. The grid represents an image space and the values contained in it. For this example, lets assume that an object is identified by containing values equal to or greater than nine. However, it is also known that objects can contain values as low as five. The image contains two regions of values above the lowest threshold, but only one that contains the necessary peak values. In order to fully describe the object whilst excluding the area without the necessary peak, the value of nine is used as an initial threshold. This identifies the values with the darker background shading. As is apparent, this is only identifies a small part of the object and it would be preferable to identify the object completely. These initially identified pixels can now be used as the seed points for a region-growing algorithm. By evaluating the pixel values adjacent to the seed points, new object pixels can be identified and their neighbours evaluated. The result is the complete shaded area, which identifies the object completely, whilst the all other pixels are excluded.

Chapter 3

Principles and Applications of Sector Scan Sonar

A prerequisite to the autonomous navigation of an underwater vehicle is the ability of the vessel to detect objects obstructing its planned path. It is therefore necessary for the vessel to have some sort of sensor that has the ability to detect objects at a adequate range and resolution. Both light and electromagnetic waves, commonly used under normal atmospheric conditions, undergo strong attenuation underwater and as such have very limited range. As a result, the ability of sound to propagate great distances under water has been extensively researched and the resulting sonar technology has been developed as a means to navigate and identify obstructions under water.

3.1 Introduction to Imaging Sonar

An initial distinction can be made between two classes of sonar, namely *passive* and *active* sonar. Passive or listening sonar consists of a receiver system that monitors all incoming sound. This system is used to identify targets from the specific noise they produce, such as the mixture of sound

caused by propellers and the hull vibrations caused by motors. This type of sonar is limited by the fact that it cannot directly measure the range to a target and commonly requires a large array of transducers. Passive sonar is therefore limited to use in military submarine hunting.

The second class of sonar, active sonar, consists of both a transmitter and receiver. The transmitter produces an acoustic pulse or *ping* and the receiver system listens for the resulting reflections or echoes. The distance to an object is then calculated based on the speed of sound in the medium and the time delay between the ping and the echo. Although all active sonars are built on this basic principle, the variety of applications have resulted in a great variation in the types of active sonar available. This ranges from the most basic single-transducer *echo sounder*, used to determine water depth, to large-array high-resolution imaging sonar.

In both classes of sonar the conversion from acoustic to electrical energy, and in the case of active sonar transmission, from electrical to acoustic, is done by devices called *transducers*. These are similar to antennas found in electromagnetics and as such also have a specific directivity, as indicated by their beam pattern. Depending on the application, a single transducer or an array of transducers may be used for both transmission and reception.

The sub-class of active sonar of most interest to this application is that of *imaging sonar*, and can divided into two basic types: *side-scan sonar* and *forward-looking sonar*. Side scan sonar provide very high resolution images of the ocean floor, used for accurate mapping of the seabed and shipwreck and mine hunting, amongst others [8]. Fig. 3.1 shows the basic operation of a side scan sonar along with an example of the resulting image from survey of a shipwreck. The sonar is mounted on either side of a vessel or towed behind it at a constant depth. It then produces a narrow acoustic beam to each side of a track line and the resulting reflections or *backscatter* is recorded. As the vehicle moves forward the sonar continuously scans the ocean floor and this continuous data input is used to produce an image.

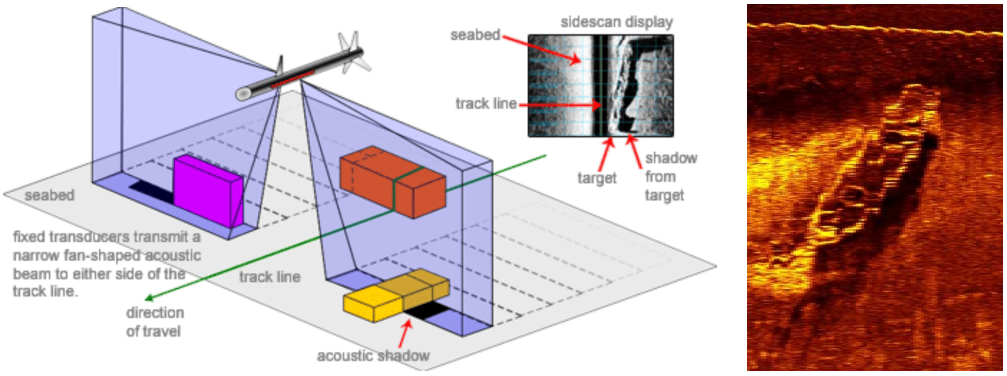


Figure 3.1: Operation of Side Scan Sonar and the resulting image of a shipwreck [1], 2008

Forward looking sonar are the type of sonar most commonly used for collision avoidance, but also finds application in mine detection and surveillance. This type of sonar commonly uses a single transducer that is mechanically scanned in the horizontal plane, to complete a sweep of a so-called sector. The sonar produces a single ping at each angle and waits for the return before stepping to the following angle, continuing until the entire sector is scanned. The returns from each ping is then used to create the image. This type of forward looking sonar is also commonly called *sector scan sonar* in order to differentiate it from the modern *multi-beam forward looking sonar*. This new development is an extension of sector scan sonar and by replacing the single mechanically scanned transducer with an array of transducers results in greatly improved update rate and resolution. The price difference between these two types of forward looking sonar can be quite significant, but depending on the application the increase in scan speed as well as the increased resolution can warrant the extra expense.

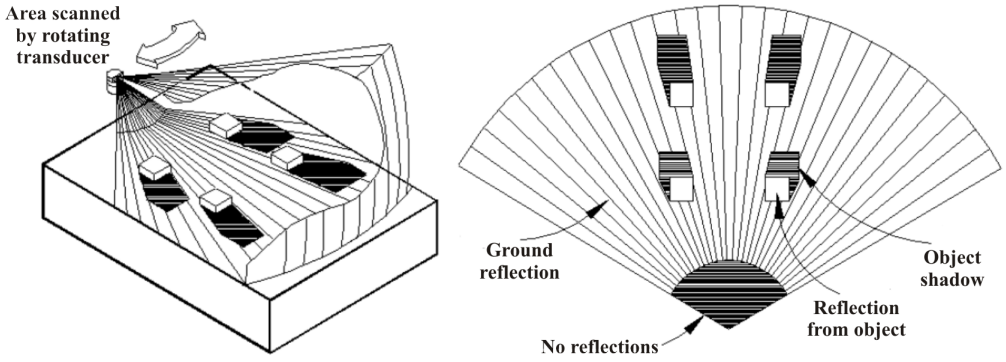


Figure 3.2: Diagram showing scanning procedure of sector scan sonar and a idealisation of the expected return [2]

3.2 Sector Scan Sonar Principles

As explained in the previous section, sector scan sonar falls under the class of active sonar. It has gained popularity for use as navigational sonar for Autonomous Underwater Vehicles (AUVs), because of its small size and relatively low cost.

This type of sonar is characterised by a fan shaped beam that is rotated mechanically to create a spatial map of its surrounding area. Fig. 3.2 gives a diagram of the scanning process and how this relates to the output images. The beam usually has a width of 1°–3° in the horizontal plane and around 30° in the vertical plane. A very important characteristic of this type of beam pattern is that it has no vertical resolution, i.e. the sonar has no way of discerning the depth or height of an object. It also means that if two objects appear directly above each other, the sonar return will show only one object.

Fig. 3.3 shows an example of a image taken with a sector scan sonar. The image was taken in a harbour at a floor depth of 14 m at a scanning distance of 50 m. The resulting image shows the returns from an anchored boat along with the harbour wall. Also of note in the image is the strong bottom reflections resulting from great difference between the water depth and the chosen scanning range.

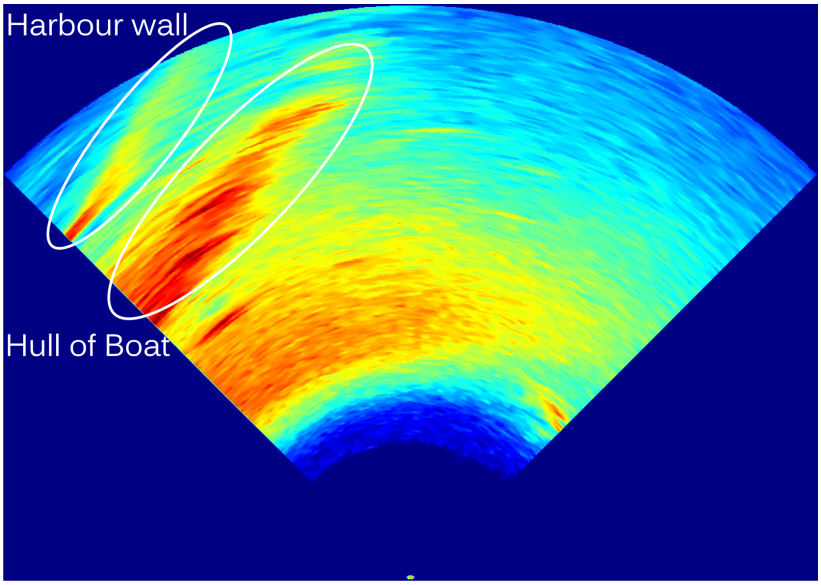


Figure 3.3: Image taken with a sector scan sonar in a harbour. The harbour wall and hull of a boat are visible, as indicated.

As with all active sonar the sector scan sonar emits an acoustic pulse to produce echoes from nearby objects. The pulse is created by exciting the sonar transducer for a short amount of time at either a single frequency or over a frequency range. The amount of acoustic energy that the sonar emits, and therefore the maximum range it can survey, is controlled by the *pulse amplitude* and *pulse duration*. The amplitude of the pulse is constrained by the transducer's physical attributes as well as *cavitation*, a phenomenon referring to the creation of a vacuum in the liquid medium, and resulting effects, when excessive pressures are applied to it [9].

3.2.1 Range Resolution

In general, *resolution* is defined as *the minimum distance separating two objects where they remain distinguishable* [10]. This a very important attribute when evaluating the performance of a sonar system.

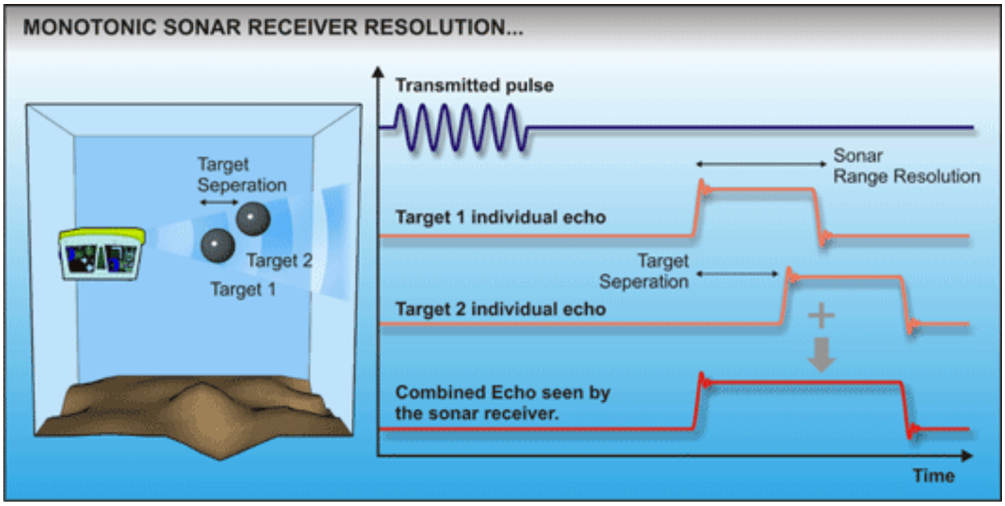


Figure 3.4: Diagram explaining the range resolution constraint of a monotonic sonar system [1], 2008

The choice of frequency has a large influence on the attainable resolution and, as will be shown, an increase in frequency translates into an increase in resolution. Unfortunately, absorption in water is proportional to the square of the frequency and is a major contributor to the attenuation of sound in water, which reduces the range of propagation. Therefore, there is always a tradeoff between resolution and range [1, 11].

When a single frequency (monotonic) system is used, the *range resolution* is given by

$$\textit{Range resolution} = (c \times T_p)/2 \tag{3.2.1}$$

where c is the velocity of sound in water, typically around 1500 m/s, and T_p is the pulse duration. Fig. 3.4 shows the return from two separate objects using a single frequency pulse (the return pulses are the output of an *envelope detecting* receiver). From the diagram it should be apparent that it is not possible to distinguish two separate objects that are separated by less than the range resolution as stated in eq. 3.2.1 [10, 12]. It is therefore possible to increase the resolution of a monotonic sonar by shortening the pulse length. This however, limits the amount of acoustic energy that

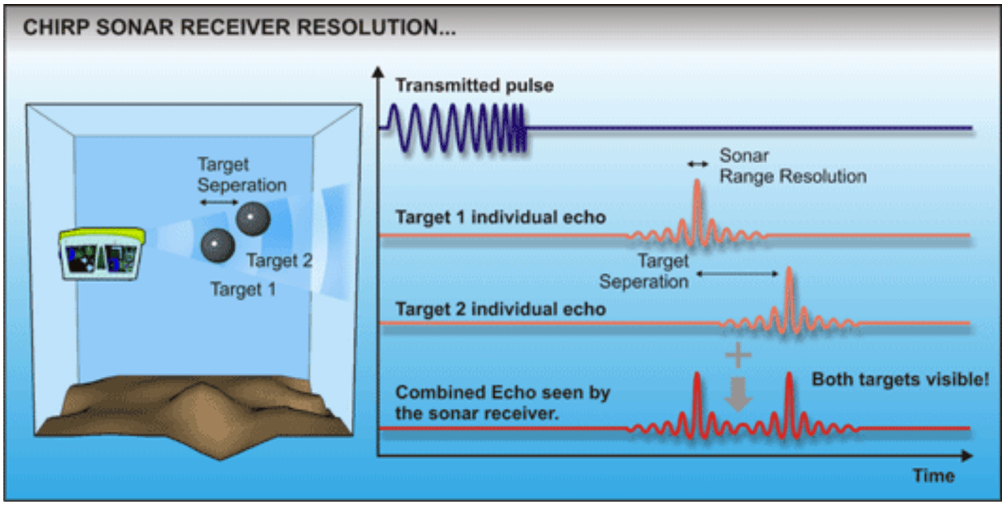


Figure 3.5: Diagram explaining the range resolution benefits of a CHIRP sonar [1], 2008

can be emitted and hence the maximum range that can be surveyed.

A common way of overcoming this constraint is to use a frequency sweep or *chirp* signal instead of a single frequency. With this technique the frequency increases from the start to the end of the output pulse. Using this type of output signal the attainable range resolution is given by

$$\textit{Range Resolution} = \frac{c}{2\,\Delta f} \tag{3.2.2}$$

where c is the velocity of sound in water and Δf is range of the frequency sweep or *bandwidth* [10] [13]. Eq. 3.2.2 also holds true for the monotonic case, as it is shown in [13] that the effective bandwidth of a monotonic signal is nearly equal to the inverse of the pulse length ($\beta = \frac{3}{2T_p}$). Using this property of CHIRP signals, the resolution can be improved by simply increasing the bandwidth of the frequency sweep. This excludes the need to shorten the pulse duration and with it the maximum survey range. Fig. 3.5 shows the two object returns of a CHIRP signal after matched filtering. The two objects can clearly be distinguished as most of the energy is now contained in the main lobe of the resultant signal.

The available bandwidth of a sonar is related to the physical attributes of the transducer and is therefore limited [11].

3.2.2 Angle Resolution

The angle or lateral resolution is given by the sonar's main vertical beam width. For common sector scan sonar this is in the order of 1°–3°. The beam width is mainly influenced by the physical dimensions of the acoustic transducer as well as the frequency used. As rule, a higher frequency and smaller transducer produces a narrower beam [10] [11]. Also of note is that since beamwidth stays constant with range, the effective resolution decreases as the range increases, as seen in fig. 3.2.

It should be clear that the attainable resolution is, for the most part, a compromise between the frequency used and the range that needs to be surveyed. There are of course further constraints such as the physical properties of available transducers, size and cost. Ultimately, the choice of system specifications are related to the needs of the application.

3.2.3 Target Strength

The final constraint on sonar images results from the sonar specifications as well as the physical properties of the objects being surveyed. The acoustic image of an object is a representation of its *target strength*. In many cases the acoustic image will resemble the optical one, but usually contain much less detail. In most cases rough objects reflect sound well and in all directions, making good sonar targets. Smooth angular objects tend to give strong reflections in the head-on or perpendicular direction, but can become near invisible when viewed from the wrong angle. Objects that form corners tend to have a strong reflection regardless of the direction it is being illuminated from and although it may be a physically small object it can appear large in the sonar image [2, 9]. The target strength is also related to the frequency being used, since the ra-

tio of target size to wavelength needs to be much greater than one for a target to be properly resolved [9]. The "roughness" as mentioned, is also with respect to the wavelength. As such, the smaller the wavelength the "rougher" an object becomes, increasing its target strength.

3.2.4 Image Composition

The composition of images from raw sonar data is not covered in any detail in the literature on detection from sector scan images. For the most part the authors use images previously acquired or taken with the manufacturer's supplied software, i.e. they do not mention interfacing directly with hardware and merely present the images under consideration. In most cases it is also somewhat outside the scope of the papers to delve into the mechanics of how the images were produced. As will be shown in later chapters, it can however be very beneficial to work with raw data and it is therefore important to gain insight into how the images are created.

As noted, sector scan images have a fan-shaped beam that sweeps across the horizontal plane in set angle steps. As illustrated in Fig. 3.2 this divides the three-dimensional space into equally-sized wedges, each wedge constituting a single ping. The volume illuminated by each ping is divided into a set of *range bins*. Each range bin is given a value pertaining to the amount of acoustic energy reflected from that volume. In practice this is done by dividing the received signal into sections of time T, and assigning a single value to each of these time sections [14]. Since time and distance are related through the velocity of sound, these time sections constitute a set of range bins.

When the sonar is set up, it is given values denoting the maximum range to be surveyed, as well as the number of *range bins* to be used (these values are subject to hardware constraints and are limited to a certain range of values by the manufacturer). As a result, when the size of a range bin is greater than the achievable range resolution, the number of

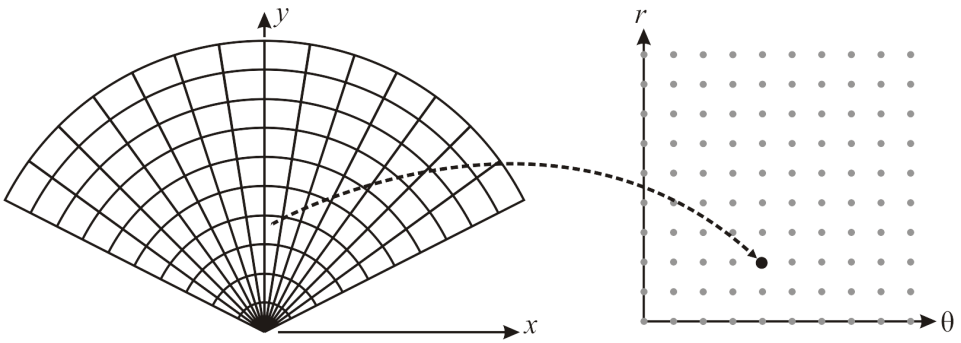


Figure 3.6: Diagram explaining conversion from Cartesian to polar space.

range bins will denote the effective range resolution, where

$$\textit{Effective Range Resolution} = \frac{\textit{Total Range}}{\textit{Number of Range Bins}} \tag{3.2.3}$$

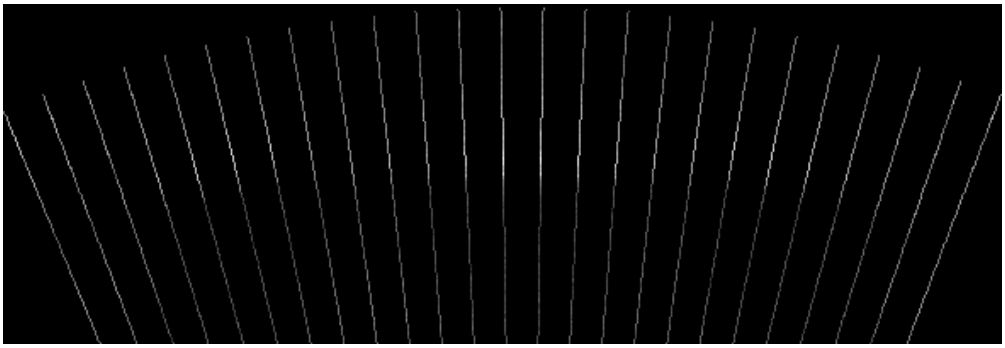
The data received from the sonar for each ping is therefore an array of values associated with a single angle. In effect, the sector scan sonar converts the continuous three dimensional Cartesian space it surveys, into a discrete polar coordinate plane, as partially illustrated in fig. 3.6.

In order to create the sonar image of the surveyed space, the polar data needs to be transformed back to Cartesian space. There is of course a simple relationship between the Cartesian and polar coordinate systems, given by

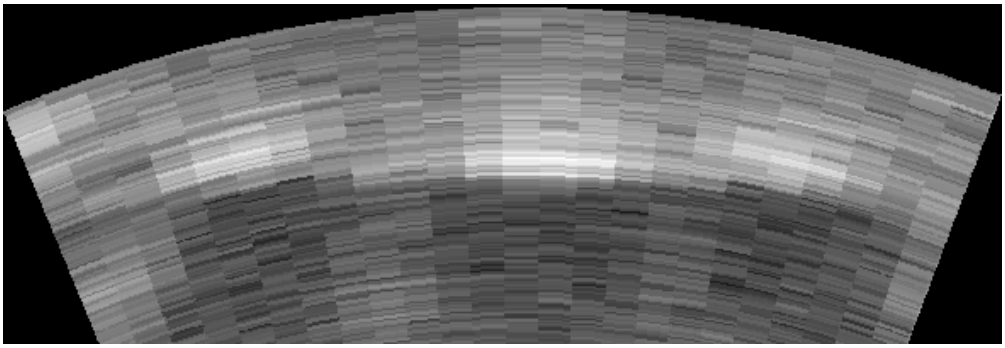
$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta. \tag{3.2.4}$$

This conversion is easily done, and except for the minor rounding errors when working in discrete space, gives a faithful reproduction of the measured data.

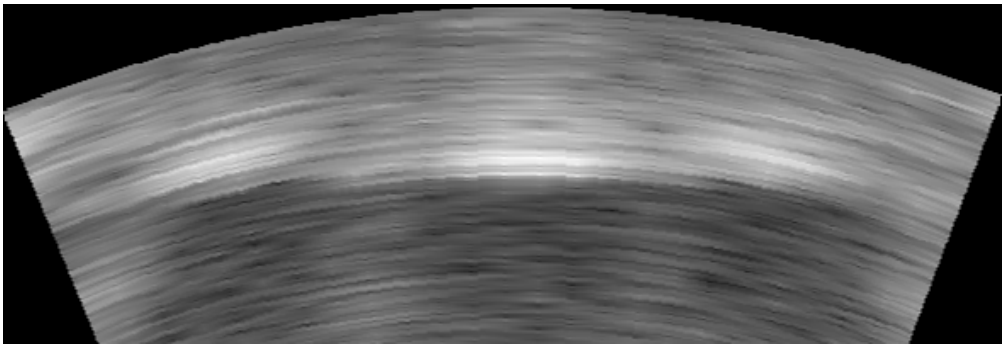
A common result found after transformation from polar to Cartesian coordinates is a sparsely covered image area, as in fig. 3.7(a). This occurs when the angle between consecutive pings is not small enough. To put this into perspective, if a single ping contains 800 bins, the angle between



(a)



(b)



(c)

Figure 3.7: (a): Sparsely transformed Cartesian image. (b)(c): Result of different types of interpolation

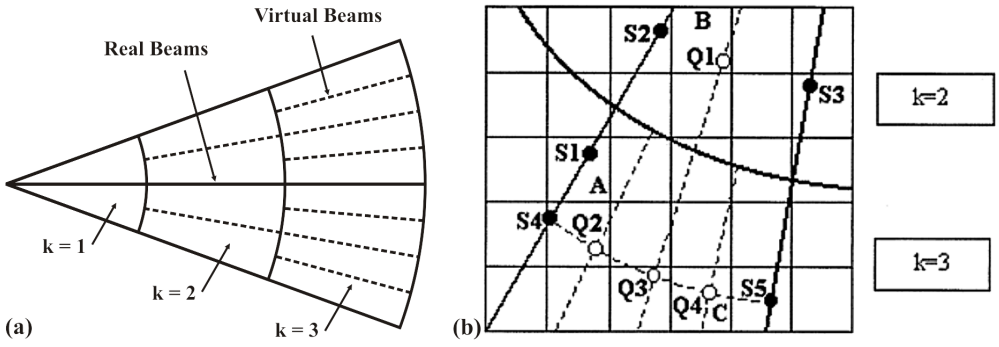


Figure 3.8: Diagrams describing the virtual-beams technique

two scans needs to be 0.143° to fill the two adjacent pixels the furthest from the origin. When this is compared to a common beam width of 3°, sampling at such an angle will, in most cases, hold no benefit. To fill the gaps in the image the unknown values are *interpolated* from the available data. Two possible methods for interpolating the data of sector scan sonar are described in [15].

The first technique, called *cell-filling*, consists of giving every pixel falling inside the area of a single range bin, the same value. This technique involves computing the four corners of the resolution cell and assigning all the contained pixels the centre value, which is the value returned from the sonar. If a pixel falls in two overlapping cells it is given the average value. Fig. 3.7(b) shows the resulting image from this interpolation technique. Although this technique does have theoretical merit, the result is somewhat crude, as illustrated by the vertical banding that is present in the image. It can also become computationally intensive, since the four corners will have to calculated for each range bin.

The second technique proposed by [15], called the *virtual-beams technique*, is illustrated in Fig. 3.8(a) and described as follows:

> If N beams are regularly distributed over a sector of ψ radians and the maximum acceptable distance from any pixel to a

beam trace is d, all the pixels up to the radius

$$r_k = \frac{2^k d(N-1)}{\psi} \tag{3.2.5}$$

will be at a distance less than $2^k d$ from the closest beam trace. As shown in Fig. 3.8, if one divides the sector into regions limited by radii $[r_k, r_{k+1}]$, $k = 1, 2, \ldots$, and if one creates $2^{k-1} - 1$ equispaced virtual-beam segments between every two adjacent real beam traces in every region k (see fig. 3.8(a)), the distance from any pixel to a real or virtual-beam trace will be less than d. A linear interpolation of the values of real samples in adjacent beam traces may be performed to compute the values of virtual samples. The interpolation coefficients used for the virtual segment i in region k, $i = 1, \ldots, 2^{k+1}$, are $c_{ik} = i/2^{k-1}$ and $1 - c_{ik}$.

Moreover, only the virtual samples that directly contribute to a pixel assignment are computed. For instance, in fig. 3.8(b),the value of the real sample S_1 is directly assigned to pixel A, whereas S_2 and S_3 (in region $k = 2$) are used to obtain the value of the virtual sample $Q_1 = (S_2 + S_3)/2$ that is assigned to pixel B. Analogously, S_4 and S_5 (in region $k = 3$) are used to obtain the value of $Q_4 = (S_4 + 3S_5)/4$ assigned to pixel C; instead, the value of Q_2 is not computed, as its closest pixel can be assigned by using the real value S_4 [15].

The virtual-beams technique, as described above, is really just an implementation of linear interpolation. The empty pixels are given the average value of the two closest valued pixels, weighted by the distance to each. The resulting image is shown in fig. 3.7(c). This technique gives much better defined objects and since it uses averaging it has the added benefit of noise rejection.

Further attention will be given to image composition in a later chap-

ter. The principles described here are however very important as background for much of the work to follow.

3.3 Review of Related Work

A limited amount of literature exists on the processing of sector scan sonar images. A great deal of work has been done on data extraction from side scan images, but due to the difference between the images produced by the two types of sonar the focus of the work differs greatly and is not really applicable. Initial work on sector scan images processed only a single image at a time [16], whereas later work focused on using image sequences to identify and track objects [17] [18].

Throughout the literature the procedure for processing sector scan images can be divided into three consecutive operations: *enhancement*, *segmentation* and either *classification* or *tracking* of objects. The classification and tracking of objects is outside the scope of this thesis. It does however create the context within which the previous work on enhancement and segmentation was done. It is therefore covered as briefly as possible.

Lane and Chantler [16] report on classification of objects in sector scan images using qualitative feature matching. The process entails calculating a set of *feature measures* from the features of each detected object and then comparing these values to a database of pre-classified objects. They identify two attributes crucial to successful automatic classification.

First, when objects are viewed from different perspectives and using different sonar devices, they commonly have a different appearance. Hence, the description features used need to be invariant of these changes. Second, the set of features should be suitably discriminatory, to ensure different objects can be properly distinguished.

They divide the process into two separate stages, *perception* and *classi-*

fication. The perception stage is equivalent to the enhancement and segmentation procedures, as previously mentioned. The processing starts by reducing speckle noise with a small (3×3 or 5×5) median filter. The background levels in the raw image are estimated using a large, 30×30, median filter. This estimation works under the assumption that the objects contained in the image are smaller than this filter and should therefore be removed, leaving just the background. This background image provides a per pixel segmentation value for the denoised image. This produces two binary images respectively representing candidate objects (values above the estimated threshold) and shadow observations (values below the estimated threshold). Erosion and dilation are applied to remove any spurious pixels. The resulting image apparently provides sufficient object observations and as such, each object is identified and separated.

A series of grey level and feature measures are calculated and produces a description vector for each observation. Using unsupervised pattern recognition techniques, a set of quantitative features providing the best separation in the clustering of objects, was identified from a larger set of measures. *Size, brightness, variance and elongation* were identified as the features providing the best separation. To provide invariance to changing appearance of objects they convert the quantitative features, as calculated previously, to so-called qualitative values (big, midbright, invariant) to create a further level of abstraction. The mapping to the qualitative values can be configured dynamically according the feature values contained in a specific image, which contributes to their invariance.

Using a set of pre-labelled images a library of *exemplars* containing the qualitative features of the contained objects is created. Using the Euclidian distances between an observed object's attributes and those of an exemplar, a quantitative value can be given denoting the likelihood of the object belonging to that class of object. The procedure did provide somewhat satisfactory results. For the most part it was successful at classifying objects that showed only slight variation between images,

but struggled with highly variable objects.

Chantler and Lane et al. [6] provide algorithms for the tracking of objects in sector scan images using the image processing technique of optical flow. This paper is an extension of the work done in [17] and gives a more thorough description of the methods used.

Firstly they identify a set of principal difficulties with tracking objects in sector scan sonar images. Similar to [16] they note the changing of objects from one image to the next; the fact that single objects may split into many and vice versa; objects quickly changing direction causing incorrect motion prediction and objects also disappearing as they move out of the vertical sonar beam as initial difficulties. Images also contain multiple static and moving objects which may merge and split between images. Finally, inadequacies in segmentation can lead to false objects caused by noise being tracked.

They begin their procedure with initial noise reduction, by applying a 5×5 median filter. Since a set of consecutive images are used, the temporal data contained in this set is used to separate static from moving objects. At each pixel point, a one-dimensional (1-D) fast Fourier transform (FFT) is calculated along the temporal axis, i.e. using the corresponding pixel's value from all images in the sequence. The static and moving objects are then separated using the low and high frequency components of this data, respectively. An inverse FFT is then applied to the low-pass and band-pass filtered data, which results in a set of data for both the static and moving objects. Next, in order to extract only the significant objects, the images containing moving objects are thresholded, using the following procedure:

1. Assuming no knowledge about the location of observations, consider the minimum grey level μ in an image, $E(x, y)$, as a pre-estimate of the segmentation threshold. Pixels whose grey level are larger than μ can initially be regarded as observation pixels. Let η^0 be the mean value of those pixels. An initial segmentation threshold can

then be defined as

$$T^0 = \frac{1}{2}(\mu + \eta^0) \tag{3.3.1}$$

2. At step t, compute η^t as the mean observation grey level, where segmentation into background and observation at step t is defined by the threshold value T^{t-1} determined in the previous step. All the pixels with a value of less than μ, $E(x_0, y_0) < \mu$, are replaced by μ. Other pixel's values are not changed.

3. Let

$$T^t = \frac{1}{2}(\eta^t + \mu) \tag{3.3.2}$$

T^t now provides an updated threshold to segment background and observation.

4. If $T^t = T^{t-1}$, halt; otherwise return to 2.

Although this procedure is automatically stopped, between five and ten iterations was found to be sufficient.

The next stage is to estimate the 2-D motion of these dynamic observations relative to the sonar. They employ optical flow and use the changing brightness field between images as a direct measure of relative motion between objects and observer. Using *Lucas' method* the velocity of each pixel is estimated. The velocity for each object is taken as the average velocity of the pixels it contains. Using this average velocity its position in the following image can be predicted.

To perform the tracking, they employed a strategy that constructs online, frame by frame, a tree of possible tracking solutions among the observations in all frames. At each image frame as new nodes are created, confidences linking nodes in the tree can be revised, thus maintaining the current best estimate of likely observation tracks.

The procedures they implemented produced satisfactory results. The position prediction had a maximum error of 5% using the optical flow motion estimates. The tree-based tracking system also showed adequate

results when dealing with the changing appearance of objects as well as the merge and split of static and moving objects.

Pettilot and Lane et al. [19], describe a framework for segmentation, tracking of objects and motion estimation in sonar images. The framework is used to design an obstacle avoidance and path planning system for underwater vehicles and can be divided into five modules. The following section provides an overview of each, as taken from the text.

1. *Segmentation*: The purpose of this module is to identify the regions of interest containing obstacles. As the vehicle is moving close to the seabed, high backscatter seabed returns are expected. This backscatter must be estimated and removed when possible in the segmentation process.

2. *Feature Extraction*: Once the image has been segmented, potential obstacles and their features (position, moments, area) are computed. These features will be used later to discard false alarms and track the obstacles and the vehicle.

3. *Tracking*: This module provides a dynamic model of the obstacles. Moreover, considering the amount of data to be processed, the tracking drives the segmentation and reduces the computational cost.

4. *Workspace Representation*: From obstacles and features extracted from the current image, we can build a symbolic representation of the vehicle's surrounding. Combining this representation with previous instances of the vehicle's environment, a dynamic workspace is built and constantly updated. It forms the basis for the path planning algorithm. In this workspace, each object is represented by its current position, shape and estimated velocity. The objects shape is assumed to be elliptic and real objects are represented as ellipses. This workspace can be seen as a local map of symbolic objects with their associated estimated static (shape, position) and dynamic (velocity) properties.

5. *Path Planning*: A nonlinear programming technique based on a *constructive solid geometry* (CSG) representation of the obstacles is used for path planning. Each obstacle in the workspace is represented as a constraint that has to be met in the search space (that is, the path must not cross the obstacle) while minimising the Euclidean distance to the goal.

The segmentation is divided into two layers. The first layer provides a basic and fast system for identifying new objects. The procedure starts with a denoising filter. They decided to use a 7×7 Gaussian filter, as it provided results comparable to that of the commonly used median filter, but at a reduced computational cost. To complete the first layer segmentation, a threshold needs to be calculated. For this calculation an adaptive thresholding technique based on the image histogram was used. Under the assumption that objects are small when compared to the image size, the histogram provides a good estimate of the noise probability density function. The values contained in the predicted locations of previously identified objects are removed from the histogram calculation. Using a fixed false alarm rate a global threshold is derived from the histogram. The assumption is that this will remove all clutter and noise that do not have an intensity comparable to that of assumed objects. This technique is used on a subsampled image to quickly identify new areas of interest for the more elaborate second layer segmentation.

The second layer segmentation is only performed on the areas of interest identified in the first segmentation layer as well as the predicted areas of previously identified objects being tracked. A double threshold is set using the previously computed image histogram. A high threshold is set to remove as much clutter as possible and identify the areas of highest intensity. These areas represent the maximum reflection of each object, but are usually not a good representation of the full size of the object. Therefore, a lower threshold is also set and all the pixels that are above this threshold and connected to the high intensity areas are incorporated into the total identified area. This double threshold technique

has the benefit of removing mid-level peaks caused by noise, whilst still identifying the objects more completely.

The objects are tracked by Kalman filters, which shows robustness in noisy environments, such as sector scan images. It also returns the uncertainty of estimated parameters, which can be included in the path planning module as a further safety measure. The state vector is composed of the position in x and y coordinates and the area of the object, along with their first and second order derivatives.

The tracking algorithm was tested by comparing the estimated motion of the AUV from the sonar images and the telemetry received from inertial navigation sensors. The inertial sensors cannot compensate for possible currents and has a much slower refresh rate than the estimation algorithm, which might lead to differences in the measured and estimated motion. Despite this, the estimated heading differed by less than a percentage point from the inertial sensor data. The workspace representation and path planning was implemented, as mention previously, and provided smooth paths in the presence of both static and moving objects.

These three papers provide a thorough representation of the current work done on the use of sector scan images. The first two also form the base for work done by other researchers [20] [18], which focus on different ways of tracking objects and do not offer further expansion of enhancement or segmentation techniques.