Automatic Target Recognition in Sonar Imagery Using a Cascade of Boosted Classifiers

Jamil Sawas

Heriot-Watt University

School of Engineering and Physical Sciences

Thesis submitted for the degree of

Doctor of Philosophy

May 2015

Abstract

This thesis is concerned with the problem of automating the interpretation of data representing the underwater environment retrieved from sensors. This is an important task which potentially allows underwater robots to become completely autonomous, keeping humans out of harm's way and reducing the operational time and cost of many underwater applications. Typical applications include unexploded ordnance clearance, ship/plane wreck hunting (e.g. Malaysia Airlines flight MH370), and oilfield inspection (e.g. Deepwater Horizon disaster).

Two attributes of the processing are crucial if automated interpretation is to be successful. First, computational efficiency is required to allow real-time analysis to be performed on-board robots with limited resources. Second, detection accuracy comparable to human experts is required in order to replace them. Approaches in the open literature do not appear capable of achieving these requirements and this therefore has become the objective of this thesis.

This thesis proposes a novel approach capable of recognizing targets in sonar data extremely rapidly with a low number of false alarms. The approach was originally developed for face detection in video, and it is applied to sonar data here for the first time. Aside from the application, the main contribution of this thesis, therefore, is in the way this approach is extended to reduce its training time and improve its detection accuracy.

Results obtained on large sets of real sonar data on a variety of challenging terrains are presented to show the discriminative power of the proposed approach. In real field trials, the proposed approach was capable of processing sonar data real-time on-board underwater robots. In direct comparison with human experts, the proposed approach offers 40% reduction in the number of false alarms.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof Yvan Petillot, for inspiring me to pursue this very interesting and challenging topic and for his support and guidance throughout this project. His encouragement to collaborate with industry and academic partners was crucial in gaining access to field trials and providing me with an exciting and holistic learning experience during this PhD.

I would also like to thank Dr Yan Pailhas for providing the sonar simulator, See-Byte for providing the sonar augmented reality software, DSTL (Defence Science and Technology Laboratory) for providing the sonar data, and Bristol University for providing the human operator results. All of these valuable resources were key to the success of this project.

My deep appreciation goes to Dr James Nelson, Dr Nicolas Valeyrie, Mr Martin Jones, and Dr Keith Brown for their prompt and helpful comments on the manuscript both in terms of language and content.

I would also like to extend my gratitude to all friends and colleagues from the Ocean Systems Laboratory for making the work environment fun and interesting.

I cannot end without thanking my family for their absolute confidence in me.

Contents

List of Abbreviations

AdaBoost	Adaptive Boosting
AR	Augmented Reality
ATR	Automatic Target Recognition
AUV	Autonomous Underwater Vehicle
HOG	Histogram of Oriented Gradients
MCM	Mine Countermeasures
ROC	Receiver Operator Characteristics
SAS	Synthetic Aperture Sonar
SVM	Support Vector Machine

Chapter 1

Introduction

The inability of the Autonomous Underwater Vehicle (AUV) to automatically analyse the data produced by its own sensors prevents it from becoming a truly autonomous unit. Therefore, AUV responsibilities are currently limited to autonomous data collection. The development of automatic sensor analysis algorithms is the central field of the work carried out in this thesis. This thesis presents an Automatic Target Recognition (ATR) approach for use with sonar imagery. The proposed approach is capable of real-time analysis of data on standard AUV embedded processors while achieving high detection accuracy.

This chapter introduces the subject of ATR in sonar imagery and shows how critical and difficult this task is. The aims of this thesis are then listed followed by a summary of the work conducted to achieve these aims. What is original about the work is finally presented. We conclude with an outline of the structure of the thesis.

1.1 Motivation

ATR technology is an essential element for the present and future subsea systems which are becoming increasingly autonomous. This is a challenging task with many applications that have attracted attention in recent years.

Naval unexploded ordnance clearance is one of the important applications. Many mines and bombs were lost or abandoned in seas after World War I and II. They represent a constant danger to humans, the environment, and shipping to name a few.

Another key application is subsea inspection and intervention to prevent and handle disasters such as the Deepwater Horizon oilfield disaster in the Gulf of Mexico and the Fukushima Daiichi nuclear power plant disaster in Japan.

Autonomous systems equipped with ATR are also being explored to search for missing ships and planes in the ocean such as the Malaysia Airlines flight MH370 and the Air France flight 447.

Existing solutions to the problems mentioned above among others are time consuming, resource-intensive, and still rely on expert divers. Therefore, there is increasing need for AUVs equipped with efficient ATR systems to keep personnel out of harm's way and vastly reduce the time and cost associated with such tasks.

Due to the problems of limited visibility associated with video imaging in the water environment, sonar is the alternative approach commonly used for "seeing" underwater. While substantial research efforts have been expended on the automatic interpretation of video data, little attention has been given to the automatic interpretation of sonar data. ATR in sonar is a difficult task due to several reasons including:

- The high level of noise in sonar data.
- The high variations in sonar data under different sensor characteristics (e.g. frequency) and sensor technologies (e.g. sidescan sonar, Synthetic Aperture Sonar (SAS), forward looking sonar).
- The lack of publicly available sonar datasets. Published methods are therefore very hard to compare.
- Data collection operations are expensive and time-consuming. They also require human intervention and may risk lives of personnel.
- The high variations of object signature in sonar imagery. Various factors affect the view of the object in sonar imagery including the position and the orientation of the object relative to the sensor. For example, an object of a certain shape and orientation with respect to the sonar incident wave may not produce backscatter and consequently no highlight. The topology of the seafloor in the neighbourhood of the object is another important factor which impacts the view of the object. For example, the view of the object gets distorted when located on sand ripples.
- The published work in this area often has some limitations due to its strategic relevance for military applications such as mine hunting. Publications are often opaque, concealing information which makes algorithms very hard to reproduce and compare with similar work.

1.2 Aims

The general aim of this thesis is to produce a real-time and accurate approach to the problem of ATR by exploiting the information available from sonar sensors. This will enhance the capabilities of the AUV and enable it to become a truly autonomous platform. This general aim can be split into smaller and more specific objectives:

- Develop a computationally efficient ATR approach. This is an essential characteristic required to enable real-time analysis of data on-board AUVs with limited processing power. This requirement has recently assumed an increasing importance due to the increasing data sizes (high resolution, high frame rate, high range, etc.) that modern sonar systems produce (e.g. SAS, Dual Frequency Identification Sonar (DIDSON)).

- Find an ATR approach which results in a low false alarm rate while keeping a high detection rate. The ability of an underwater vehicle to behave autonomously is heavily dependent on the performance of the ATR. The false alarms of the ATR in an autonomous system might imply changes in the mission plan. A high number of false alarms may therefore result in lengthy missions. In order to replace human experts, the ATR should at least offer an equivalent performance.

1.3 Methodology

The work conducted in this research project can be summarised in the following steps:

1. The principles of sonar were studied to permit a good understanding of the data used in this thesis.

2. The literature of ATR in sonar was reviewed to identify gaps, extract the lessons learned, and use these to build on an effective approach.

3. Some existing ATR approaches from the computer vision community, traditionally used for optical images, were reviewed looking for solutions to achieve the aims of this thesis and fill some of the gaps in the literature of sonar ATR.

4. An approach from the computer vision community was chosen for further investigation. This approach uses Haar features together with a cascade of adaptively boosted decision tree classifiers [1]. Initial results were promising which proved this approach suitable for sonar data.

5. The proposed approach was studied extensively to fully understand and find the optimal way of running such a complex supervised algorithm; whose training phase is computationally very expensive (days to weeks) while its running phase is extremely efficient (milliseconds).

6. Some limitations of the proposed approach were identified and some extensions were proposed to alleviate some of these limitations. As a result, the extended approach is not only more accurate but also faster to train (minutes versus days).

7. The proposed approach was thoroughly evaluated on the synthetic and real data from a variety of challenging terrains. This also included a direct comparison with human experts. Furthermore, the approach was successfully integrated within two AUVs and evaluated in real field trials.

We have selected the cascade of boosted classifiers approach [1] from the computer vision literature and decided to try it for the first time within sonar based ATR due to its unique characteristics which we assumed likely to achieve the aims of this research. These characteristics include:

1. The proposed approach is capable of processing data extremely rapidly. This is expected to meet the real-time requirement of this thesis. The exceptional computational efficiency of this approach is mainly attributed to the cascade classifier which focuses the processing on the regions of interest in the image. The integral image representation is another factor which speeds up the processing by computing Haar features very efficiently. Further reduction in the computational complexity is achieved by reducing the feature space using Adaptive Boosting (AdaBoost).

2. The proposed approach is capable of achieving high detection accuracy while keeping the number of false alarms reasonably low. This is expected to meet the aim of low false alarm rate in this thesis.

3. The proposed approach is capable of learning from large amounts of background data; this enables the exploitation of all previously collected sonar data.

4. The proposed approach encodes objects based on the grey level differences between different regions; this is expected to extract good information about objects in sonar imagery (the highlight-shadow pairs).

5. The proposed approach is a suitable approach for handling highly imbalanced data. There are typically a lot more non-target samples than target samples in the data. The proposed approach alleviates the class imbalance issue by using the cascade classifier.

6. The proposed approach does not rely on assumptions such as whether the object of interest is tethered or lies on the seafloor. It learns directly and automatically from the data.

7. The proposed approach does not rely on navigational data or any sonar specific processes. This makes it independent from any errors involved in collecting such data or performing such processes.

8. The proposed approach is a holistic approach which is capable of performing all the traditional tasks of existing ATR systems at the same time: detection (regions of interest), classification (object/non-object), and identification (object type).

1.4 Outcomes

The key findings of the analysis conducted in this research are:

1. The proposed approach, the cascade of boosted classifiers together with Haar features, is a very effective ATR approach for sonar data. This is justified theoretically and experimentally in this thesis.

2. In comparison with some existing approaches, the proposed approach appears to offer the highest processing speed. It can process a large SAS image of 7000x2000 pixels in less than a second (using 3 GHz Intel Xeon processor), while existing approaches may take several seconds to several minutes. In real field trials, the proposed approach was capable of running real-time on-board AUV, on a standard embedded processor (PC104 - Intel 400 MHz) concurrently with all other software modules in the AUV.

3. In comparison with some existing approaches, the proposed approach appears to reduce the number of false alarms while achieving comparable detection accuracy. In direct comparison with human expert operators, the proposed approach offers around 40% reduction in the number of false alarms.

1.5 Contributions

This thesis makes the following key contributions:

1. This thesis is the first to propose the use of the cascade of boosted classifiers in the sonar domain.

2. This thesis proposes a novel method to measure the confidence in the output of the cascade classifier. The confidence values of the cascade stages are combined to compute the confidence value of the cascade.

3. This thesis proposes several optimisations to speed up the training phase of the proposed approach which is extremely time-consuming. An intelligent sampling procedure is introduced which does not re-examine samples rejected in previous stages when training a new stage. Limiting the early stages of the cascade to coarse features and gradually introducing finer features is another idea proposed to reduce the training time.

4. This thesis proposes several extensions to improve the detection performance of the proposed approach. The exploitation of information from previous and subsequent stages in making a decision at any stage of the cascade classifier is presented. The ability of the proposed approach to adapt to the context is discussed.

5. This research is the first to run extensive experiments on large collection of real heterogeneous sonar data. The collection includes hundreds of real man-made objects and covers over 20 km^2 of challenging seabed terrains. Data from various sonar technologies (sidescan, SAS, and forward looking) are included in the collection.

6. Overall this research proposes a novel ATR approach in sonar imagery which runs real-time on-board AUVs and outperforms human experts.

Some of these contributions have been documented in several publications during this PhD tenure. Some of the knowledge and the expertise gained during this research have also been transferred to industrial partners. Furthermore, the work presented here has been evaluated and exploited in research projects in collaboration with academic and industrial partners. All of these scientific activities including the publications are listed in Appendix A.

1.6 Structure of the Thesis

This chapter introduced the subject of ATR in sonar and showed how important and critical this subject is. The aims of this thesis were then listed followed by a brief description of the work conducted to achieve these aims. What is original about the work was finally presented. The remaining chapters of the thesis are organised as follows:

- **Chapter 2** gives a brief background about the subject of ATR in sonar data. The basic principles of sonar imagery are explained to permit understanding of the data used throughout the thesis. Consideration is then given to previous work in the area of ATR in sonar.

- **Chapter 3** introduces the cascade of boosted classifiers from the perspective of sonar data and provides theoretical justification for its suitability to this domain. The main thrust of the thesis is then presented which includes several extensions to this approach including: a new confidence measure, training speed optimisations, and detection performance improvements. The impact of some external

effects on the detection performance is finally discussed, including the introduction of new data and the post-processing using the Histogram of Oriented Gradients (HOG) features.

- **Chapter 4** describes the experimental results obtained in this thesis. Results on synthetic and Augmented Reality (AR) data are presented in addition to real data. A direct comparison with human experts is then presented. The results of in-water trials are then presented to demonstrate a fully autonomous solution for subsea survey, target detection, and intervention operations.

- **Chapter 5** concludes the work conducted in this thesis and summarises the key outcomes. It also discusses the limitations of the work and suggests some directions for future research in this area.

Chapter 2

Background

2.1 Introduction

By way of further introduction, in this chapter, we describe the modus operandi of Automatic Target Recognition (ATR) in sonar imagery. Principles of sonar imagery are first introduced to permit a good understanding of the data used throughout the thesis. It will be shown how complex sonar images can be relative to both human experts and ATR approaches. Consideration is then given to previous work in the area of ATR in sonar. The typical design of existing ATR approaches is first described. We then discuss the benefits and limitations of various existing techniques.

2.2 Principles of Sonar Imagery

Due to the limited propagation of light in the water environment, sound is the alternative signal commonly used for sensing underwater. Sound is also attenuated underwater, but not to the same extent as light. Although sound does not produce high quality images in comparison with light, it does provide a remarkable alternative for underwater mapping and imaging.

Sonar (SOund Navigation And Ranging) is the term which refers to the techniques which use sound waves to explore the underwater environment. This term was coined as the equivalent of Radar (RAdio Detection And Ranging) which uses radio waves in air instead. By using sound rather than light to form images on suitable sensors, sonar systems make it possible to observe the underwater environment at greater distances and where the optical visibility is poor. Sonar systems can either be passive which listen to the sounds emitted by any object, or active which emit sound and listen to the echoes. The sonar systems we deal with in this thesis are all active.

Sonar devices are an important element of underwater systems for commercial and military applications. They can be used for different purposes such as navigation,

seafloor mapping, pipeline or cable route survey, and object detection (e.g. shipwrecks, mines, and downed aircraft). Object detection, commonly referred to as Automatic Target Recognition (ATR), is the application addressed in this thesis.

Sonar devices typically produce digital images as a visual representation of what they insonify. Although sonar images may look like optical images of the seabed, they are not as easy to interpret. It is fundamental to understand the underlying image formation process of the particular sonar type to be able to interpret the images correctly. There are several types of sonar systems used for different purposes. In this thesis we mainly deal with sidescan sonar and its extension Synthetic Aperture Sonar (SAS). In one of our experiments we also process data from a forward looking sonar system. Therefore, these technologies are briefly introduced in this section, where we focus mainly on the image interpretation and the object signature. For more details, the interested reader is referred to [2] inter alia.

2.2.1 Sidescan Sonar

Sidescan sonar is the most commonly used sonar imaging device. In sidescan sonar, a narrow sound beam is transmitted at regular time intervals while the vehicle traverses through the water. Each beam covers a narrow strip to the side of the vehicle and perpendicular to the direction of travel as illustrated in Figure. 2.1. The sound signals propagate though the water and bounce off anything in their path (e.g. an object, or the seabed). The returning signals (known as the backscatter) are recorded.

Image Formation

The image is constructed from the backscatter, where each row corresponds to one strip of the seabed. Therefore, the image, broadly speaking, represents a two-dimensional scan of the insonified region of the seabed. The horizontal dimension corresponds to the time of the returning signal and the vertical dimension corresponds to the time of scanning each strip. The pixel intensity corresponds to the strength of the returning signal. Therefore, a sonar image is basically a visual representation of the acoustic backscatter.

Typical sidescan sonar systems scan the seafloor simultaneously on both sides of the vehicle (e.g. tow-fish, survey vessel, Autonomous Underwater Vehicle (AUV), Remotely Operated Underwater Vehicle (ROV)). An example of a sidescan sonar image is shown in Figure. 2.1 with an illustration of its content relative to an approximate schematic diagram of sidescan sonar. The image covers approximately 60x30 metres of the seafloor. The resolution is about 5.8 cm along range and 12 cm along track. The very dark region at the centre of the image corresponds to the water column (minimal reverberation). This region varies in width in this image due to variations in the altitude

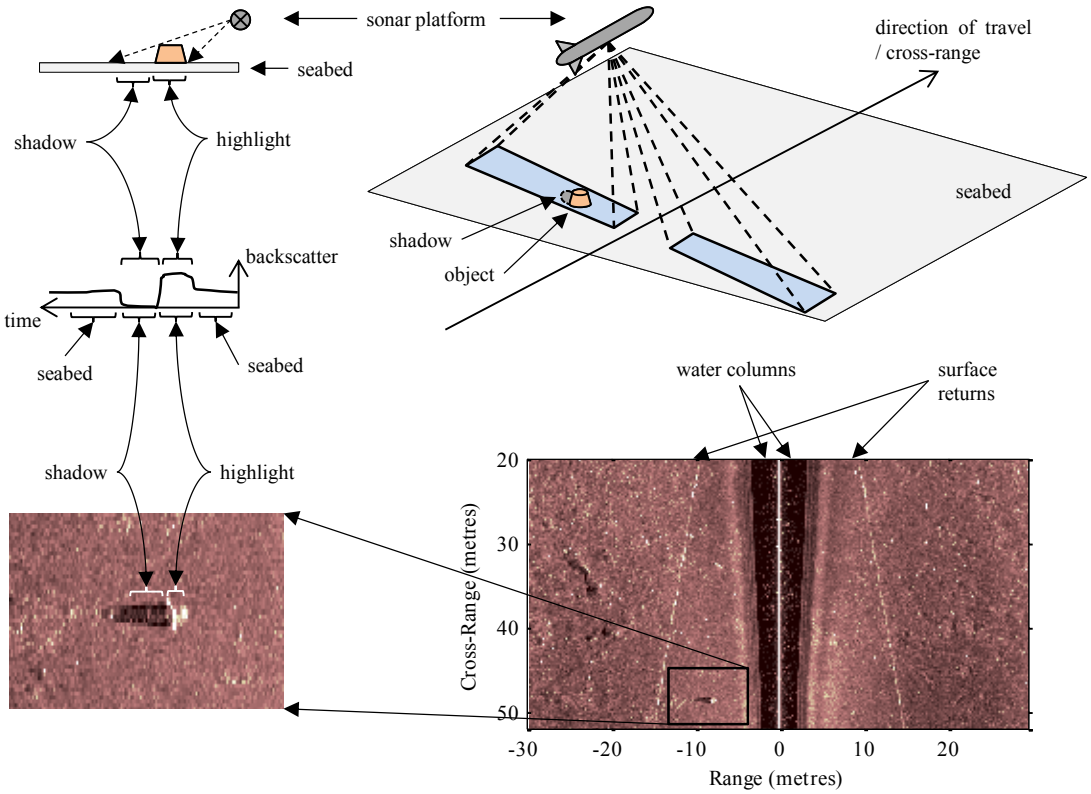

Figure. 2.1 Illustration of sidescan sonar principles. Schematic diagram of the acoustic "footprint" of sidescan sonar system is shown on top right. An example of a real sidescan sonar image is shown at the bottom. The formation of object signature (highlight/shadow pair) is illustrated on the left for a single ping (one-dimensional viewing angle).

of the vehicle. The image shows mainly sediments, some clutter, and one man-made object. The two bright lines along the sonar track represent the reflections from the sea surface. They usually occur in shallow water environment and are commonly referred to as the surface return. In addition, there are other characteristics of sidescan sonar imagery which are too many to discuss here for which the reader is referred to [3] inter alia.

Object Construction

In a sidescan sonar image, an object often appears as a bright region, usually referred to as the highlight, next to a dark region, usually referred to as the shadow. The highlight is a result of the high acoustic reflectivity of the object in comparison with the seabed, while the shadow is a result of the acoustic waves being blocked from reaching the region behind the object. Figure. 2.1 shows a man-made object which can be identified by the highlight/shadow pair. It is a dummy underwater mine of truncated cone shape (see Section 4.7 for more details).

As with any sonar system, sidescan sonar can only show objects which reflect sound back to the sonar. This is influenced by many factors including the object ma-

terial, the object aspect angle, the seabed texture, and the seabed topography. Due to the typical high grazing angles of sidescan sonar devices, objects are prone to cast shadows. Shadows are regions of low or no acoustic reflections. They are typically dark regions behind objects protruded from the seabed as illustrated in Figure. 2.1. Sonar platforms typically traverse at low altitude to ensure that objects produce good shadows.

Shadows are often very useful in identifying objects. They encode information about the size and the shape of the corresponding objects. The length of the shadow is related to the height of the object, its range, and the altitude of the sonar vessel as can be seen in Figure. 2.1. On the other hand, the shadow can only be as wide as its corresponding object. Therefore, sonar images of the same underwater object can look very different depending on the particular sonar conditions.

2.2.2 Synthetic Aperture Sonar

Sidescan sonar systems have some limitations, most notably in the trade-off between range and resolution. Low frequency systems are capable of scanning high ranges, but they produce low resolution imagery. High frequency systems are cable of producing high resolution imagery, but they are limited to low ranges. Synthetic Aperture Sonar (SAS) is an enhanced version of sidescan sonar which is capable of producing high resolution imagery up to high ranges. While each ping (the colloquial term for a sound pulse) is processed independently by sidescan sonar, SAS combines multiple pings to synthesize a large aperture (antenna). A large aperture is required to form sharp beams and consequently high resolution imagery. The term SAS also originates from its radar predecessor SAR (Synthetic Aperture Radar).

The simplified model used in Figure. 2.1 for standard sidescan sonar can also model SAS approximately. The main exception is an additional sophisticated post-processing step of the raw sonar data to coherently integrate successive returns. Therefore, the interpretation of SAS imagery is similar to that of sidescan sonar imagery. Figure. 2.2 shows an example of a SAS image containing seven man-made objects. The image has a resolution of about 1.5 cm along range and 2.5 cm along track. This resolution is around 18 times higher than the resolution of the sidescan image presented earlier. The image covers up to about 150 metres range (around 5 times further than the sidescan image presented earlier). More details about the data will be provided in Section 4.5.1.

To show the level of complexity in sonar data, close snapshots of some targets and target-like clutter are presented in Figure. 2.3. Sonar data is very likely to include clutter objects which resemble real targets in shape, size, and intensity (compare the two columns on the left in Figure. 2.3). The views of real targets are likely to get distorted due to technical or/and environmental reasons (see the right column in Figure. 2.3).

For instance, the view of the target may get blurred due to some limitations in the sensor (see Figure. 2.3 (c) and (f)). The view of the target may become unclear or almost hidden due to the complexity of the seafloor (see Figure. 2.3 (f) and (i)).

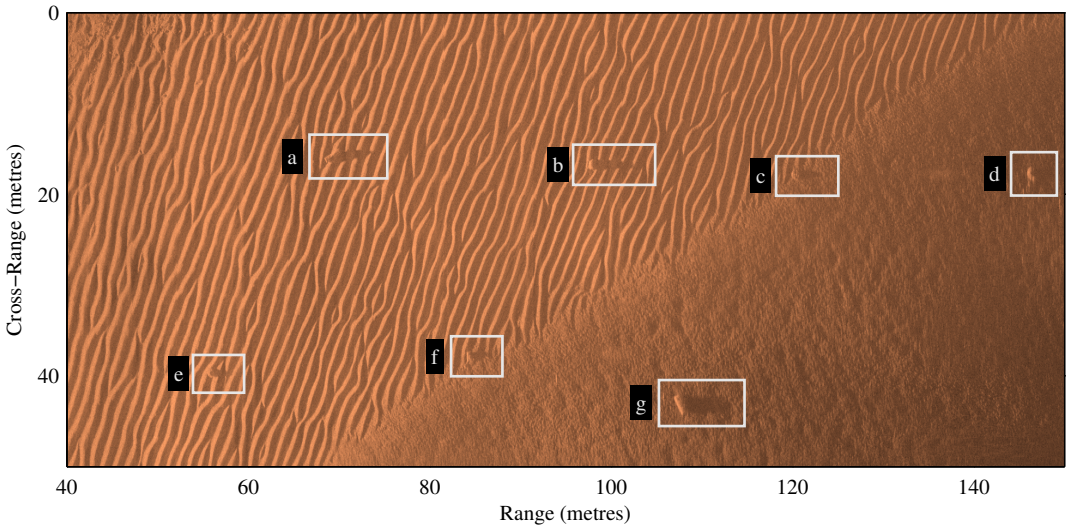

Figure. 2.2 Real SAS image including seven different underwater man-made objects. Objects are highlighted by bounding rectangles. They are truncated cones (e,c,d), cylinders (a,g), and wedges (f,b) (see Section 4.5.1 for more details about the data).

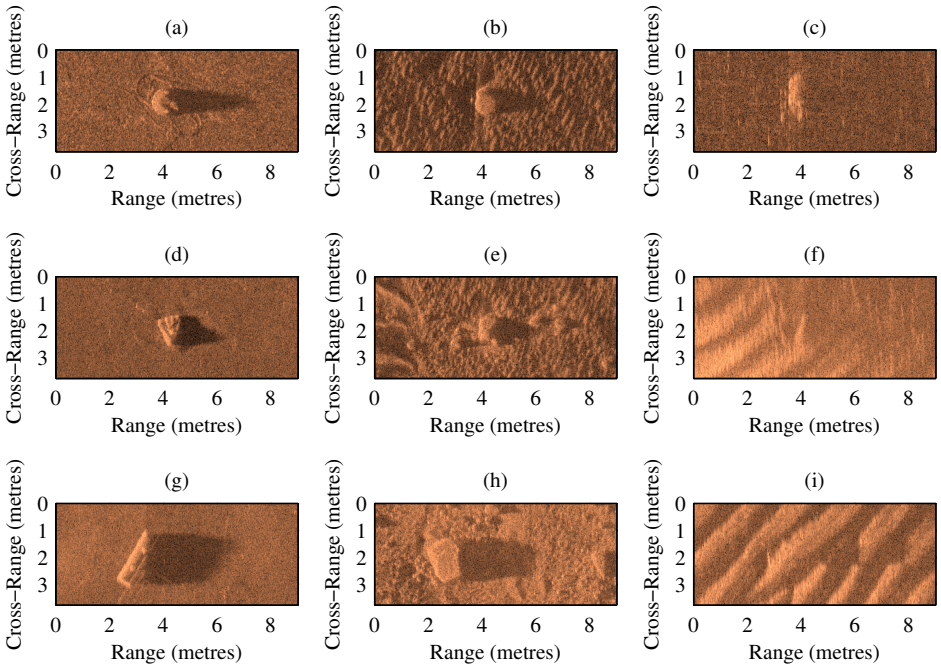

Figure. 2.3 Examples of targets and clutter from SAS data. They are typical targets (a,d,g), target-like clutter (b,e,h), and distorted targets (c,f,i).

Initial commercial SAS systems have recently become available. Their increased resolution offers the opportunity of using more conventional image processing techniques. In particular, the object highlight is much clearer in SAS in comparison with

sidescan sonar. Therefore, exploiting the highlight in traditional ATR approaches which mainly concentrate on the shadow may improve the detection performance as demonstrated by [4]. However, the high level of speckle noise within SAS images is problematic to ATR approaches. Furthermore, the shadows are often slightly altered by SAS processing which may limit the ATR performance. Nevertheless, the major challenge of SAS technology lies in the stability of the vehicle (sonar platform) and the accuracy of its navigation. For more details about SAS, the interested reader is referred to [5] for deeper discussions and [6] for a recent survey.

2.2.3 Forward Looking Sonar

Forward looking sonar is another type of sonar which is relevant to this thesis. It can produce several acoustic images per second in a video-like fashion. Unlike typical sonar systems which require motion or mechanical rotation, forward looking sonar can work on stationary and moving platforms. It quickly constructs an image by forming many small acoustic beams at once using an array of transducers.

Forward looking sonar systems are generally mounted vertically to the front of a vehicle. They are commonly used for obstacle avoidance and target recognition. Similar to other sonar systems, bright regions in the forward looking sonar image corresponds to objects reflecting sound, while dark regions are acoustic shadows resulting from an object blocking the sound.

An example of forward looking sonar image is shown in Figure. 2.4 with an illustration of its content relative to an approximate schematic diagram of forward looking sonar. The image covers a sector of approximately 15 metres radius and 120° field of view. The vertical beam width is 20°. The image shows mainly sediments from the sea bottoms. There is only one man-made object identified by the highlight/shadow pair. It is the exact same object shown earlier in sidescan sonar image in Figure. 2.1. Data will be described in more details in Chapter 4.

One of the disadvantages of forward looking sonar is the ambiguity in the location of the target in the axis normal to the sensed plane. This is a direct result of the significant vertical beam-width (20° in the example shown above).

2.3 Literature Review

Early approaches of ATR in sonar data appeared in the 1980s. In the literature, these approaches are often referred to as computer aided detection (CAD) and computer aided classification (CAC) algorithms. This term was coined due to the initial need for these algorithms to aid the human operators rather than replace them. Among the more influential authors on the subject is Dobeck [7]. In this section we will give a brief

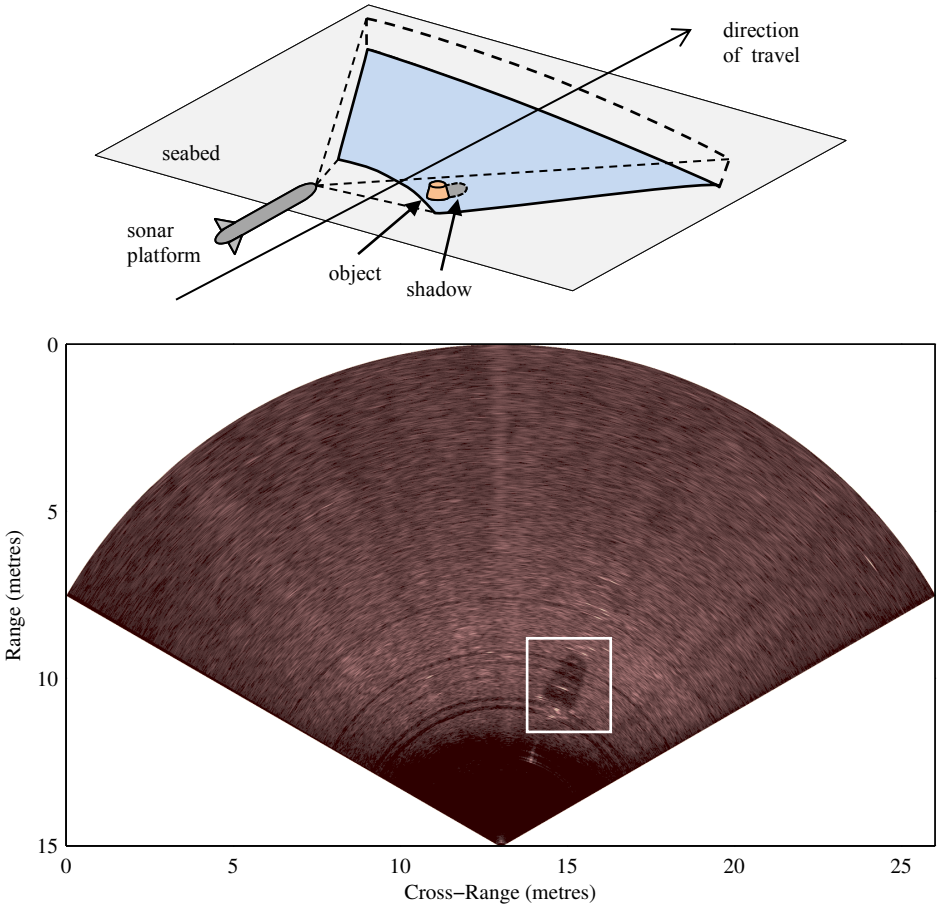

Figure. 2.4 Illustration of forward looking sonar principles. Schematic diagram of the acoustic ”footprint“ is shown on top. An example of a real forward looking sonar image is shown at the bottom. The image contains a man-made object highlighted by bounding rectangle (see Section 4.7 for more details).

review of the literature. A review of the whole field would be a formidable task due to the large number of studies done over the years. This review will therefore mainly highlight the typical paradigms in this field. Some important work has inevitably been missed. The review will mainly cover sidescan sonar and SAS techniques, as they are the main areas of interest in this thesis. This section is divided into three parts. First, we give an overview of the design of a typical ATR system. We then give a summary of existing approaches by organising them into different groups. Finally, we discuss the limitations of current approaches and introduce our solution to overcome some of these limitations.

2.3.1 The Design of Automatic Target Recognition

The design of pattern recognition techniques has been the central focus of the machine learning community for many decades and the sonar community has benefited consid-

erably from this research. Therefore, the design of traditional ATR approaches in sonar imagery follows that of traditional pattern recognition approaches and is illustrated in Figure. 2.5. The underlying operation can be divided into the following steps:

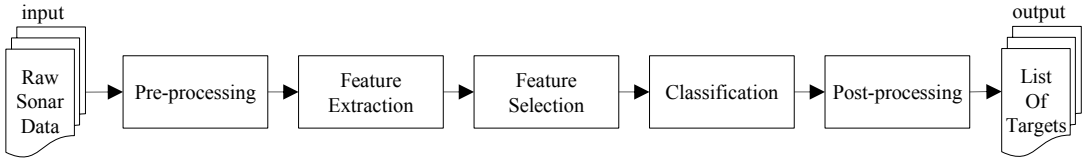

Figure. 2.5 The architecture of traditional ATR approach.

- **Preprocessing**: This step transforms the raw data produced by the sonar sensor into a form suitable for ATR processing. This step is typically sensor-specific and may include transforms such as normalization, de-noising, beam form correction, slant range correction, and surface return removal. These transforms are too many to discuss here for which the interested reader is referred to [2] inter alia. Although the preprocessing of sonar data is not what this thesis is most concerned with, the impact of some preprocessing techniques on the proposed ATR is evaluated. This includes de-noising and sand ripple suppression, as will be described in Section 3.6.1.

- **Feature extraction**: This step transforms the data pre-processed in the previous step into a new space referred to as the feature space. This space is designed to provide better separation between classes than the original space. The choice of distinguishing features is a very critical step which depends mainly on the problem being studied. Different feature descriptors have been published over the past few decades, but there are no standard features for ATR in sonar imagery. More details about this topic will be provided in the following section.

- **Feature selection**: This step identifies the relevant features and rejects any redundant or irrelevant features. Reducing the feature space does not only reduce the computational complexity but it is also frequently asserted that a simpler classifier is expected to generalise better according to the principles of Occam's razor [8]. This is of particular importance to sonar applications due to the limited amount of data available for training, according to Bellman's curse of dimensionality [8]. Despite their potential benefits to sonar ATR, feature selection methods have only occasionally been used [7, 9]. Adaptive Boosting (AdaBoost) (described in Section 3.2.2) is the feature selection method used in this thesis.

- **Classification**: This step assigns a class label to each data sample. The trend of classification techniques in sonar ATR follows that of machine learning. They

vary from Bayesian techniques [10] to neural networks [7, 11] and Support Vector Machine (SVM) [12–14]. There are no context-independent reasons to favour one classification method over another, as the No Free Lunch theorem states [8]. In other words, context-specific knowledge is required to design successful algorithms; there is no such thing as a general purpose algorithm. The ATR approach proposed in this thesis employs three different types of classification algorithms: decision trees, boosting, and the cascade classifier. They together construct an effective ATR approach as described in the following chapter.

- **Post-processing**: This step may include any techniques which verify the classifier output in an attempt to reduce the number of false alarms while maintaining high detection rate. This may be as simple as a constraint on the size of the contact. It may also include advanced techniques such as data fusion, where additional data (multiple views from the same sensor or from different sensors) is exploited in the final decision process [15]. Post-processing may also involve classifier fusion techniques, where multiple classifiers take part in making the final decision. Data fusion is often avoided due to its high complexity and cost (i.e. additional sensors, additional power, and additional time) in comparison with classifier fusion (only additional computation). Classifier fusion is discussed in more detail in the following section.

2.3.2 Existing Approaches

Human operators and ATR tools often search for possible highlight shadow pairs as a first guess of possible objects in sonar imagery (see Section 2.2.1). The majority of existing techniques rely on the description of the shadow rather than the highlight in making their decisions. This is due to the high variability in the intensity and the shape of the highlight versus the shadow in sidescan sonar imagery.

In the literature of object detection in sonar imagery, techniques may be divided based on their output into: detection techniques, classification techniques, and identification techniques. Detection techniques find all regions of interest (object-like regions) in an image. Classification techniques determine whether an object-like region contains an object or not. Finally, identification techniques determine the identity of an object (cylinder, truncated cone, etc.).

These divisions may also be recognised as the typical steps of some comprehensive ATR systems. Regions of interest in an image are first sought, then each region is checked to determine whether it contains an object or not. Only object-like regions go through more detailed analysis to work out the object identity. Traditional ATR systems often employ separate algorithm for each step. However, some recent ATR systems combine multiple steps together in one algorithm. The approach proposed in

this thesis combines all of these steps in one algorithm. However, depending on how the training dataset is partitioned, the algorithm can either be used for detection or identification.

In the literature of object detection in sonar imagery, techniques may also be categorised based on their learning models as follow:

- **Unsupervised techniques**: use a priori models and statistics and do not require training data. Since they are not trained, their decision process is typically simple (such as Bayesian) and transparent (i.e. why a decision has been made). The authors of [16] propose a complete unsupervised ATR system of three phases. In the first phase, regions of interests are found by segmenting the image using Markov Random Fields algorithm. Each region of interest is then passed to a second phase, which extracts the highlight and shadow based on the statistical snakes technique (active contour model). In the final stage, the object class (cylinder, sphere, or truncated cone) is determined based on comparing the extracted shadow with synthetic shadows based on the Hausdorff distance.

- **Supervised technique**: learn from labelled sets of data. They analyse the training data to infer a function, known as the classifier, which discriminates between different object classes. Rather than dealing directly with the data, supervised approaches often classify based on a vector of features extracted from the data. Therefore, their performance is dependent on the choice of features (discussed later in this section). When well trained, supervised approaches outperform unsupervised approaches. A range of different supervised approaches have been examined in the literature [7, 17] including the approach proposed in this thesis. In [18], the Hilbert transform is used to segment the object and shadow regions after which a curve fitting algorithm is used to extract features for classification by decision trees.

Moreover, techniques may also learn from both labelled and unlabelled data and they are referred to as semi-supervised. The method in [19] differs from most previous work in that it automatically selects a few samples which are most informative for classifier design. However, the authors of [19] assume access to a human operator or another sensor to ascertain the labels associated with the samples. The method in [14] also differs from most previous work in that it only learns the background (non-targets). The targets are then detected as anomalies.

In the literature of object detection in sonar imagery, techniques may also be divided informally based on their underlying principles as follow:

- **Segmentation techniques**: These techniques assign a class label (such as highlight, shadow, or background) to each data pixel. Any large enough blob of

highlight pixels associated with a large enough blob of shadow pixels may constitute a target. These techniques range from as simple as thresholding [20] and matched filtering [7, 14] to more sophisticated methods such as Markov random fields [16], statistical snakes [16], and Fourier descriptors [21]. Although the concept of segmentation in this context may seem straightforward, methods often break down in complex seabed types. Moreover, most segmentation algorithms require an initialization step which determines to a great extent the final results. They mostly require parameters which must be chosen heuristically. When located over sand ripples, the shadow of the object may get segmented with the shadow of the sand ripple (the same may happen to the highlight). This scenario is common in sonar imagery and represents one of the greatest challenges of object detection in sonar imagery. The false alarms produced by segmentation based ATR approaches are often due to the poor segmentation.

- **Template matching techniques**: These techniques measure the similarity between a new unlabelled sample and a library of labelled templates. They are quite popular for ATR in general and for ATR in sonar specifically [22, 23]. These techniques are useful in discriminating between different object types. Templates are often constructed by simple simulation methods (e.g. ray-tracing) which are not very accurate in modelling the actual sonar formation process. Consequently, this limits the performance of template matching techniques. Furthermore, templates need to be generated under different sonar conditions (e.g. sonar to target azimuth), which may result in a very large library of pre-stored models. This issue has been alleviated in the literature by generating a few templates on-the-fly; based on the navigational information of the sonar platform [24]. In general, template matching techniques require high computational power, which represents one of the main obstacles in using them in real-time. Their computational demand increases as the number of templates increases.

- **Feature-based techniques**: Rather than dealing with the pixels directly, these techniques extract features. The motivation may be better object representation, dimensionality reduction, and/or computational efficiency. The features used to encode the characteristics of object signatures in sonar imagery can informally be divided into geometrical features and statistical features.

  Geometrical features exploit the shape properties of the segmented regions. The assumption is that objects from different classes have different geometrical properties. Contrary to natural objects, man-made objects often cast highlights and shadows of regular shapes. Examples of simple geometrical features include area, elongation, extent, and compactness [7, 20, 25]. Geometrical features may also be represented by the properties of a generic shape fitted to the segmented

region, such as fitting the super-ellipse to the shadow in [26].

Statistical features exploit the statistical properties of sonar images. These features rely on the assumption that objects from different classes have different distributions of their pixel intensity. Examples of statistical features include: mean, variance, Kurtosis, and Skewness [7, 27, 28].

Additional features include the coefficients of some mathematical transforms such as Fourier in [21] and wavelets in [9]. In this thesis, special arrangements of Haar wavelets are used to extract features, referred to as Haar features which will be described in Section 3.2.1.

- **Fusion techniques**: These techniques combine the outputs of multiple ATR algorithms to produce a new more reliable ATR output. Fusion techniques have been shown to be effective in reducing the false alarm rate relative to that of a single ATR algorithm [29–31]. They rely on the assumption that different ATR algorithms make different mistakes. Therefore, there is often no consensus about the false alarms among the team of algorithms, which facilitates eliminating them by the fusion process. Fusion, in this context, often refers to combining classifiers which are heterogeneous and independently developed. However, in this thesis, we use two different forms of fusion-like methods which do not follow this rule. They combine homogeneous classifiers which are not independently developed. They are AdaBoost and the cascade structure which will be explained in the following chapter.

Although it does not appear to have yet been applied to the problem of ATR in sonar data, it is worth to mention here Deep Learning [32]. This is an emerging approach within the machine learning community which has recently attracted wide-spread attention. Although it is appealing to investigate this approach for ATR in sonar, it is beyond the scope of this thesis, and is considered in the chapter on conclusions as future work.

2.3.3 Limitations of Existing Approaches

Despite the remarkable advances in sonar ATR, the problem is still unsolved in all but the most trivial applications (where the environment and/or objects of interest are heavily controlled or predictable). The typical performance of existing ATR algorithms is still poor and they have therefore not gained a widespread acceptance yet, especially for critical tasks such as underwater mine recognition. The fundamental reasons for this generally poor performance are listed in this section. They are informally divided into two camps: in the first, the practicality of existing approaches is questioned, and in the second, their theoretical limitations are considered.

Practical Limitations

This section lists the limitations which relate to the practicality of existing approaches in real-world applications:

- Existing approaches are in general computer intensive; a factor which hinders the possibility of real-time processing for on-board AUVs. Traditional approaches typically include a segmentation step which might take up to several minutes to process one sonar image.
- Existing approaches result in a high number of false alarms, which constitutes one of the biggest obstacles in the way of the complete automation of underwater missions.
- Existing approaches are mainly hindered by the need to heuristically choose some parameters. Such parameters may need to be changed when some circumstances change (such as the sensor and the sensor frequency) which may be beyond the expertise of the human operator on-board.
- Existing approaches are often evaluated on simulated data, where the variability encountered in real data is not reproduced.
- Existing approaches are mostly evaluated on simple seabed regions (flat) and most likely to break down on complex seabed regions (rocky and sand ripples).
- Existing approaches are mostly evaluated on a small set of homogeneous data while it might be necessary to evaluate them on large datasets of different characteristics (different frequencies, different sensors, etc.).

Theoretical Limitations

This section lists the limitations which relate to the theoretical principles of existing approaches:

- Traditional approaches rely on segmentation as a first step to simplify the image so it becomes easier to analyse. However, if the segmentation is poor, the subsequent steps employed to further analyse the segmented regions will also be poor. For instance, some man-made object regions may look more like clutter after segmentation, and vice versa.
- The results generated by some approaches can be hard to interpret. In other words, why a particular sample/region is labelled as such. This may be attributed to the black-box solution such approaches offer to the problem.

- Unsupervised approaches rely on a priori fixed models of the objects. New models need to be added when new objects become of interest. Therefore, the list of a priori models may become large over time and consequently slow down the processing of such approaches.

- Supervised approaches rely on the similarity between the data previously used for training and any new data that becomes available. Such approaches may require re-training when new data become available if the performance is not sufficient.

- The inability to learn the clutter. Most existing approaches rely on the characteristics of the objects and do not exploit the information in the large amount of background data previously collected. Some recent approaches do exploit the background in their learning process. However, they are limited to learn only a small amount of background data due to the computational complexity involved.

- Most supervised approaches require a large amount of data for training. This data is often unavailable due to the high cost and limited scale of real underwater experiments. Trained on a small dataset, the generalisation ability of the classifier may be impaired due to over-fitting the training data. This issue has been addressed indirectly in the literature in two different ways: the use of model based approaches which require no training, and the use of simulation and augmented reality to generate large amount of training data. Both of these solutions have their own flaws. Model based approaches require accurate models and may become computer intensive with the increased number of models as discussed earlier. The success of simulated and augmented reality data rely on how closely they represent real data.

- The focus on minimizing the rate of missed objects (objects classified as clutter) rather than the rate of false alarms (clutter detected as objects). This is perhaps a direct consequence of the application of mine hunting, where the cost of missing a target is much higher than the cost of a false alarm. However, the high number of false alarms produced by existing approaches is one of the biggest obstacles in the way of the complete automation of such systems.

When described in this light it is clear that further research is required to overcome the limitations of existing approaches. The purpose here is not to criticize prior work; rather, it is to analyse it, extract the lessons learned, and use these to build on an effective approach.

Thus, in this thesis, a novel approach for ATR in sonar imagery is proposed to alleviate some of the limitations discussed above. The proposed approach is based

on an approach from the computer vision community originally developed for face detection [1]. To date, no research on applying this approach to sonar data appears to have been done.

This approach deviates from prior work in that it processes data extremely fast while achieving high detection accuracy. It is a supervised approach which learns features directly and automatically from the data. Haar features are used for object representation and AdaBoost is used for classification. Several AdaBoost classifiers are combined in a cascade structure to gradually focus the attention on regions of interest.

While most ATR approaches use a multi-tier process (i.e. segmentation, detection, classification, and identification), this approach will directly identify the objects in a sonar image. Also, unlike most ATR approaches, the proposed approach is capable of learning large amount of background data. All of these characteristics and several more will be discussed in more detail in the following chapter.

The focus was mainly on real-time processing and acceptable detection performance (the main aims of this thesis) when this approach was selected. While this approach will be shown to be capable of achieving these objectives and filling several gaps in the literature, it has its own limitations in addition to some limitations it shares with supervised approaches in general. Some of these limitations will be discussed in the following chapter, where some extensions will be proposed to alleviate them. However, other limitations are still open and will be recommended for future research in the chapter on conclusions.

2.4 Conclusion

This chapter has given the reader a brief background about the subject of this thesis. The basic principles of sidescan sonar were first presented in addition to important characteristics specific to SAS, and forward looking sonar. The content of several sonar images were illustrated and discussed where particular attention was given to the object signature. Consideration was then given to previous work in the area of ATR in sonar. The typical design of existing ATR approaches was first described. We then discussed various existing techniques and identified their limitations. We concluded by introducing our solution to this problem which will be described in greater detail in the following chapter.

Chapter 3

The Cascade of Boosted Classifiers

3.1 Introduction

The problem of Automatic Target Recognition (ATR) in sonar has been specified in the previous chapter; where the nature of sonar data has been described and prior work on this subject has been discussed. This chapter proposes a novel solution to the problem of ATR in sonar which is assumed to achieve the aims of this thesis and fill some of the gaps in the literature.

This chapter starts by presenting the architecture of the system in Section 3.2. This includes Haar features for object representation, Adaptive Boosting (AdaBoost) for classification, and the cascade structure for classifier fusion. The architecture will be discussed under the problem of ATR in general and ATR in sonar in specific.

The main thrust of the thesis is then introduced which includes several extensions to the proposed approach. This will mainly include a more robust confidence measure in Section 3.3, a faster training phase in Section 3.4, and a higher detection performance in Section 3.5.

Finally, the impact of various external effects on the performance of the proposed approach is discussed in Section 3.6 including data preprocessing and post-processing. All the proposed extensions in this chapter will be accompanied by experimental results to support the assumptions made. Complete experiments along with a full description of the data will later be presented in Chapter 4.

3.2 Architecture

The architectural overview of the proposed ATR approach is provided in Figure. 3.1 as applied to the task of mine detection. Both the training phase and the testing phase are outlined in Figure. 3.1. In the training phase, the system takes as input two sets of samples (image patches): one from the target class and another from the non-target class.

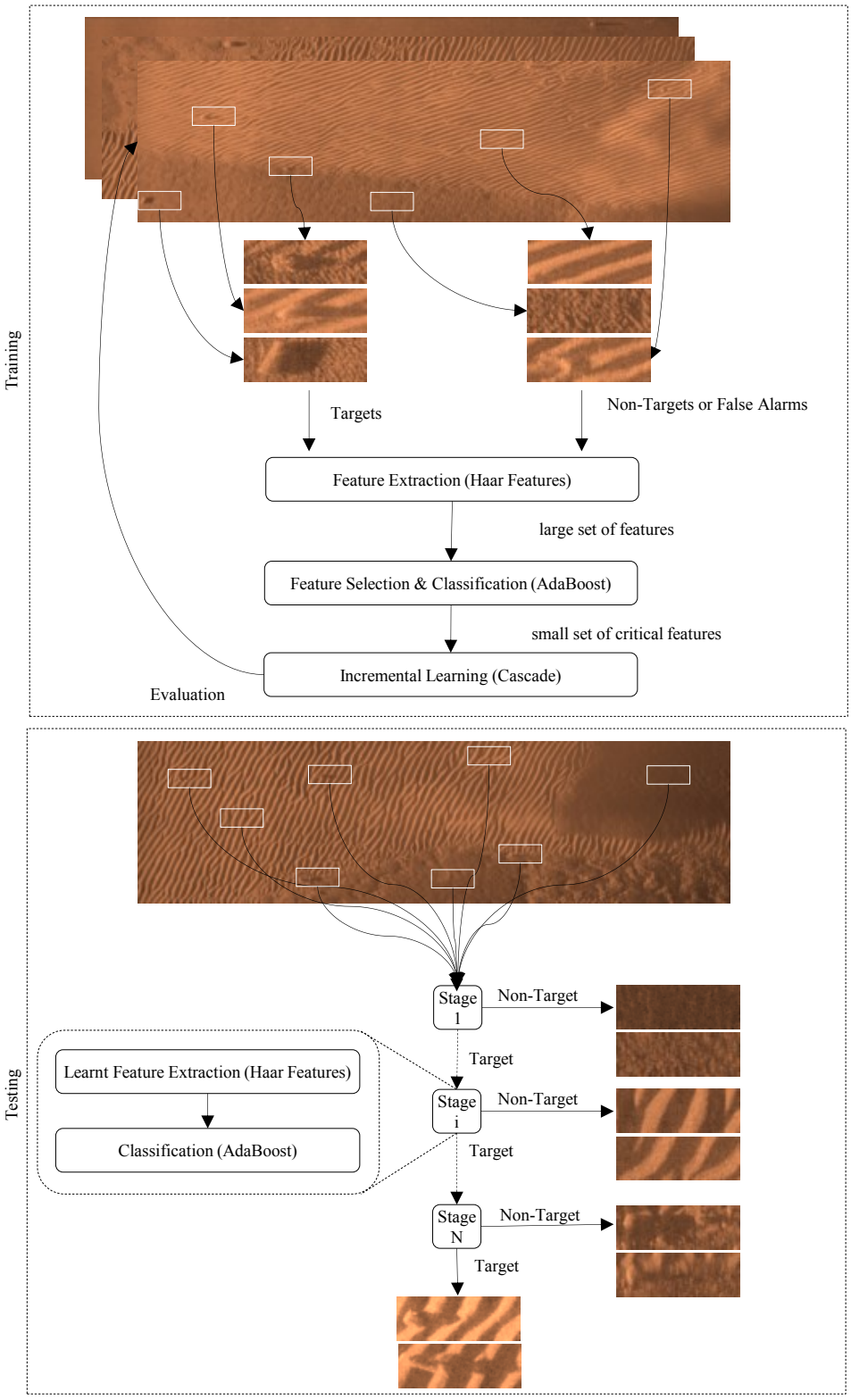

Figure. 3.1 Architecture of the proposed approach including the training and testing phases.

While target samples are aligned so that they are all in approximately the same position and the same size, non-target samples are randomly selected from the background.

An intermediate representation (Haar features) that extracts important information about the object of interest is computed for each of these samples, yielding a set of very large feature vectors which can either be positive (target) or negative (non-target). A small set of critical features are selected and combined by a classifier (AdaBoost) to detect almost all (e.g. 99%) target samples and a moderate fraction (e.g. 50%) of no-target samples. The system then evaluates the interim detector on the non-target (background) images. If the desired false alarm rate is achieved, the processing stops. If the desired false alarm rate is not achieved, a new classifier is trained similar to the previous one and added to it, creating a sequence of classifiers (the cascade). This process is repeated until the desired false alarm rate is achieved. The only exception is that subsequent classifiers are trained using false alarms of previous classifiers.

In the testing phase, we are interested in detecting targets in out-of-sample images. The system slides a fixed size window over an image and uses the trained cascade of classifiers to decide whether a window contains the object of interest. At each window position, the system extracts the small set of features learnt in the first stage and feeds them to the first stage classifier. The classifier output determines whether that patch is a target pattern. If it is not a target pattern, the processing stops. If it is a target pattern, the second stage is consulted similar to the first stage and so on. In summary, the patch is highlighted as a target if it is classified as a target by all the trained layers. On the other hand, the first time the patch is classified as a non-target by a stage, the system highlights the patch as a non-target and the processing stops.

This architecture is based on the seminal work of Viola and Jones in [1]. It is one of the most influential object detection systems in the computer vision community. However, as we will see in the following subsection, the authors of [1] did not invent any of the building blocks of their architecture; they selected them from a large body of literature and carefully put them together to build a strong approach. Their work can be credited with the widespread popularity of these building blocks. Their work has also spawned a large body of literature which investigates various aspects of the proposed architecture. In the following subsections we will briefly describe these building blocks within the context of sonar imagery. First, the object representation based on Haar features is explained. Second, feature selection and classification based on AdaBoost is introduced. Finally, the cascade structure which allows the focus of attention is described.

3.2.1 Object Representation

The success of any ATR algorithm depends on using a suitable image representation. The ultimate goal in choosing a representation for an ATR system is to find one which effectively encodes the target concepts while being tolerant to noise and intra-class variability. The representation used in this thesis is a variant of wavelet representation which will be described in this section.

Using features rather than pixels for classification can be motivated by the fact that features may provide a better encoding of the domain knowledge. This is particularly important when we deal with finite training data that is inevitable in underwater applications due to the high cost and the tremendous efforts required for data collection. In addition, a classifier built using features could run faster if only a few simple features need to be calculated.

Therefore, the question becomes: what features can indicate the existence of an object in a sonar image? To the human observer, objects lying on the seafloor share some similar characteristics in sonar images as shown in the previous chapter. These characteristics include: the object region is associated with a much darker accompanying shadow region, the object region is brighter than the seafloor, and the shadow region is darker than the seafloor. These characteristics represent significant visual information about the object signature in sonar imagery and features which can extract such characteristics will be useful.

In the context of face detection in optical imagery, faces also share similar characteristics such as: the eye region is darker than the forehead region, the cheek region is brighter than the eye region, and the mouth region is darker than the cheek region. Motivated by such qualitative relationships, the authors of [33] derived an image representation scheme called the "ratio template" and applied it to face detection. The ratio template consists of a set of relationships between the average intensities of a few different face regions. Being primarily inspired by qualitative measures, the ratio template is biologically plausible. In essence, the human visual system is much better at comparing brightness levels than at judging their absolute values.

Motivated by the work on the ratio template, the authors of [34] proposed an extension called the "wavelet template" and applied it to pedestrian detection. The wavelet template automatically captures the relationships between intensities of adjacent regions using 2-dimensional Haar wavelets which encode such relationships along different directions. Figure. 3.2 (1) shows the 3 types of 2-dimensional Haar wavelets which capture intensity differences along the horizontal, vertical, and diagonal directions.

Motivated by the work on the wavelet template, the authors of [1] used features reminiscent of Haar wavelets, called Haar features. They used more complex features

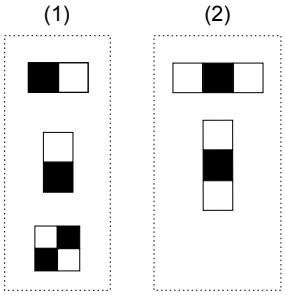

Figure. 3.2 Prototypes of Haar features. (1) 2-dimensional Haar wavelets used in [34], (2) additional Haar features used in [1].

(see Figure. 3.2 (2)) in addition to the original Haar wavelets (Figure. 3.2 (1)). Most importantly, they introduced a very efficient approach for feature evaluation based on an image representation called the integral image.

The integral image is an image representation for efficiently computing the sum of intensities within any rectangular region of an image. The integral image (also known as a summed area table) was first introduced in [35] for texture mapping and popularized later by its prominent use in [1] for object detection. As the name suggests, the integral image at any location of an image is the sum of all the pixel intensities above and to the left of that location. Once the integral image is computed, the sum of pixel intensities within a rectangle of the image can be computed rapidly as the sums and the differences of the integral image values at the four corner points of the rectangle as illustrated in Figure. 3.3. Hence, Haar features, which represent the differences of the intensity sums between rectangles as illustrated in Figure. 3.2, can be computed at any scale and at any location of the image very efficiently using the integral image.

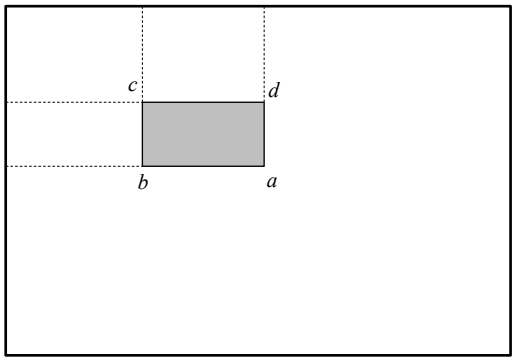

Figure. 3.3 The sum of the pixel intensities within the grey rectangle of the image can be computed rapidly using the integral image values of the four corner points as $(a+c)-(b+d)$.

For the task of mine detection, the initial Haar features selected by the classification algorithm, AdaBoost, described in the following section, are meaningful and easy to interpret. The first feature selected seems to focus on the property that the region of the highlight is often brighter than the region of the shadow (see Figure. 3.4). The second feature relies on the property that the region of the highlight is often brighter than the

background. These early features are relatively large in comparison with the sample size and therefore they should be less sensitive to sensor noise and small variations in object size, viewing angle, and location.

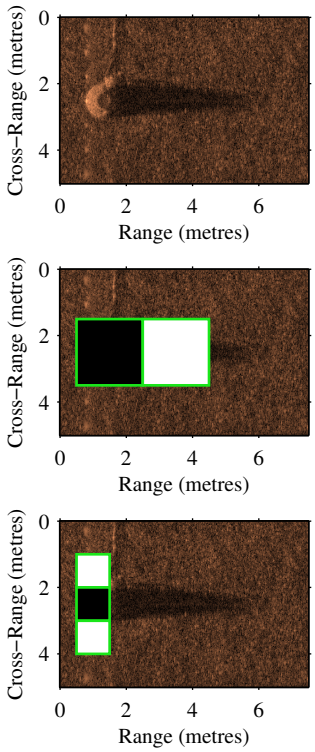

Figure. 3.4 The first two Haar features selected by AdaBoost. A typical truncated cone mine sample is shown in the top. The first feature is shown in the middle overlaid on the mine sample. It measures the difference in intensity between the object highlight and its shadow. It encodes the property that the highlight is brighter than the shadow. The second feature is shown in the bottom overlaid on the mine sample. It measures the difference in intensity between the highlight and the background. It encodes the property that the highlight is often brighter than the background.

Haar features are quite primitive in comparison with other features such as Histogram of Oriented Gradients (HOG) [36], Local Binary Patterns (LBP) [37], and Scale-Invariant Feature Transform (SIFT) [38]. Therefore, several attempts have been proposed in the literature to enhance the expressional capability of Haar features or replace them. An extended set of Haar features will be proposed and discussed in Section 3.5.2.

Once we have selected an adequate object representation and extracted the corresponding feature vectors from target samples and non-target samples, a classification algorithm is needed to learn to differentiate between the two categories. The particular learning engine we use is AdaBoost, described in the following section.

3.2.2 Feature Selection and Classification

The second key component of the proposed system is the use of a pattern classifier that learns to distinguish between samples from our target class and all other samples. Given the feature vectors of target samples and non-target samples, a number of classification approaches could be used to learn the classification function (e.g. Support Vector Machine (SVM), neural networks). They basically derive an implicit model of the domain of interest based on the labelled data.

Given the large number of Haar features which can be extracted from each sample, it is prohibitively expensive to extract all of them in the classification process. It is also frequently asserted that a simpler classifier (which uses a lower number of features) is expected to generalise better according to the principles of Occam's razor [8]. Therefore, ATR design may include a step for feature selection as we saw in the previous chapter. This step selects a subset of features which provides the most discriminatory information to the classification system.

In the proposed system, only one algorithm, AdaBoost, is used both for feature selection and for classification. AdaBoost is one of the most successful learning-based classification techniques developed in the last two decades which won the Godel prize in 2003. While classification algorithms are often application specific and require parameter tuning for an optimal performance, AdaBoost (with decision trees as weak learners) represents one of the best off-the-shelf classifiers (see discussion in [39]). AdaBoost is explained in this section along with some extensions used in this thesis.

AdaBoost

Boosting is a method for finding a strong (highly accurate) classifier by combining many weak (moderately accurate) classifiers. For an introduction on boosting, we refer the readers to [40]. AdaBoost [41], short for "Adaptive Boosting", is an adaptive extension of boosting which is generally considered the first step towards practical applications of boosting. AdaBoost is adaptive in the sense that it adapts to the error rates of the individual weak classifiers. In other words, subsequent weak classifiers learn in favour of classifying those examples misclassified by previous weak classifiers. This is achieved by re-weighting the training examples at each iteration, where examples that are misclassified get higher weights in the next iteration. The final output of AdaBoost is the weighted combination (weighted majority vote) of the weak classifiers where each weak classifier weight is inversely proportional to its training error. A pseudo code of the AdaBoost training algorithm is given in Figure. 3.5 for the two-class classification problem.

To support their algorithm, Freund and Schapire [41] provided some theory. They proved that the training error of AdaBoost drops exponentially with the number of it-

1: Given N examples $(x_1, y_1), \ldots, (x_N, y_N)$ with $x \in \Re^k, y_i \in \{-1, 1\}$
2: Initialize example weights $w_i = 1/N$
3: **for** $m = 1, 2, \ldots, M$ **do**
4: Select the weak classifier $f_m(x) \in \{-1, 1\}$ which minimizes the weighted misclassification error:
$$e_m = \sum_{i:f_m(x_i) \neq y_i}^{N} w_i$$
5: Update the example weights: $w_i \leftarrow w_i exp\left(\alpha_m . 1_{(y_i \neq f_m(x_i))}\right)$ and normalize so that $\sum_i^N w_i = 1$, where:
$$\alpha_m = \log\left(\frac{1-e_m}{e_m}\right)$$
6: **end for**
7: **return** $sign[\sum_{m=1}^{M} \alpha_m f_m(x)]$

Figure. 3.5 AdaBoost training algorithm [41].

erations while each weak classier is performing slightly better than a random guess. They also provided an upper bound on the generalisation error of AdaBoost. Their bound suggests that AdaBoost with too many weak classifiers will over-fit. However, in practice, AdaBoost achieves results much better than the bound and does not over-fit. It was also observed that AdaBoost continues to drive down the generalisation error even after the training error had reached zero. In response to these empirical findings, an alternative interpretation of AdaBoost was given in [42] based on the theory of the margins. The margin is a measure of the detection confidence of the prediction. It was shown theoretically and experimentally that AdaBoost is effective at increasing the margin of the training examples. Even after the training error reaches zero, AdaBoost continues to increase the margins of the training examples. This explains the corresponding drop in the generalisation error. Five years after the inception of AdaBoost, its equivalence to forward stage-wise additive modelling was discovered in [39]. The authors of [39] analyse AdaBoost from a statistical perspective. They derive AdaBoost as a method for fitting an additive model in a forward stage-wise manner which explains why AdaBoost tends to outperform a single weak classifier.

AdaBoost as a Feature Selection Process

In its original form, AdaBoost is used to boost the classification performance of weak classifiers by combining a set of them to form a strong classifier. However, AdaBoost can be easily interpreted as a feature selection procedure. This can be achieved by constraining each weak classifier to depend on a single feature only. As a result, in each iteration of AdaBoost, the weak learning algorithm selects the single highly selective feature for which the target examples are most distinct from the non-target examples, taking into account examples weights. Tieu and Viola [43] used such an analogy be-

tween week classifiers and features in the domain of image retrieval. They use a simple and quick weak learning algorithm, where for each feature; the optimal threshold is selected such that the minimum number of examples is misclassified, taking into account examples weights. Effectively this is a single node decision tree, known as the decision stump in the machine learning literature. Viola and Jones followed a similar approach in [1] for their object detection framework. They also proposed an optimisation for finding the optimal feature threshold. First, samples are sorted based on the feature value. Then, the optimal threshold is computed in a single pass over the sorted list (see [1] for more details).

In this thesis, each weak classifier is also constrained to one feature. Regression stumps are used rather than decision stumps due to the use of an extended version of AdaBoost (described in the following section) which requires weak classifiers of real rather than binary outputs. The optimised search for the feature threshold mentioned above is also extended to suit regression stump in this thesis.

Unlike typical classification algorithms such as SVM, AdaBoost has the feature selection step integrated into its design. In comparison with other feature selection approaches, AdaBoost learns very fast as the dependence on previously selected features is encoded in the example weights.

Gentle AdaBoost

Several variants of AdaBoost have been proposed in the literature. They attribute improved performance to their use of different boosting methods. This thesis uses a variant called Gentle AdaBoost which is an extension to another variant called Real AdaBoost. Therefore, this section will first introduce Real AdaBoost followed by Gentle AdaBoost.

Schapire and Singer in [44] observe that AdaBoost combines binary weak classifiers assigning a weight to each weak classifier reversely proportional to its classification error; while in fact the confidence of a weak classifier may vary with inputs. For instance, when trained, a decision stump classifier may have a pure leaf and an impure leaf. While all samples in a pure leaf have the same class label, samples in an impure leaf have different class labels. In this case, when tested, the decision stump classifier will be more confident in classifying samples which go to the pure leaf than samples which go to the impure leaf. To tackle this issue, Schapire and Singer in [44] propose a generalisation of AdaBoost called Real AdaBoost which uses confidence-rated classifiers rather than binary classifiers. The sign of the weak classier output gives the classification and its absolute value gives a measure of the confidence. A pseudo code of the Real AdaBoost training algorithm is given in Figure. 3.6 for the two-class classification problem.

The formula for computing the weighted class probability estimate P_w in Fig-

ure. 3.6 depends on the type of weak classifier employed. In this thesis, where the weak classifier is a decision stump, P_w depends on which leaf a given example x falls into. Within this leaf, P_w is the probability estimate of the corresponding class y given example weights.

1: Given N examples $(x_1, y_1), \ldots, (x_N, y_N)$ with $x \in \Re^k, y_i \in \{-1, 1\}$
2: Initialize example weights $w_i = 1/N$
3: **for** $m = 1, 2, \ldots, M$ **do**
4: Select the weak classifier:
$f_m(x) = \frac{1}{2} \log\left(\frac{P_w(y=1|x)}{P_w(y=-1|x)}\right)$
which minimizes the error:
$e_m = \sum_i^N w_i \, exp\left(-y_i f_m(x_i)\right)$
where P_w is the weighted class probability estimate.
5: Update the example weights: $w_i \leftarrow w_i \, exp\left(-y_i f_m(x_i)\right)$ and normalize so that $\sum_i^N w_i = 1$
6: **end for**
7: **return** $sign[\sum_{m=1}^M f_m(x)]$

Figure. 3.6 Real AdaBoost training algorithm [44].

While AdaBoost is well known to be immune to over-fitting, it is sensitive to outliers (mislabelled data or noisy data), see e.g. [45]. During the training, AdaBoost increases the weights of any mislabelled samples and consequently tries hard to classify them. A similar issue arises where some training samples of different classes are almost identical. This is in fact the case in the application of mine detection, where many clutter objects resemble real targets in shape, size, and intensity as shown in Section 2.2.2. On the other hand, the views of some targets may not look any similar to the typical target class. This is usually the case where the target view gets distorted due to technical or/and environmental reasons, as also shown in Section 2.2.2.

Therefore, to alleviate this issue we choose to use a variant of AdaBoost which is more robust to outliers, called Gentle AdaBoost [39]. Gentle AdaBoost is an extension of Real AdaBoost which puts less emphasis on misclassified samples (i.e. outliers, mislabelled samples, or object-like clutter) by only gently increasing their weights during training. A pseudo code of the Gentle AdaBoost training algorithm is given in Figure. 3.7 for the two-class classification problem. Empirical evidence in [39, 46] suggests that Gentle AdaBoost often outperforms both standard AdaBoost and Real AdaBoost, especially when stability (e.g. outliers) is an issue.

While we choose to use Gentle AdaBoost in this thesis, there exist a number of alternative variants of AdaBoost which also claim to improve the tolerance to outliers such as RobustBoost [47] and TangentBoost [48]. Although it is appealing to examine

such methods, it is beyond the scope of this thesis, and is considered in the chapter on conclusions as future work.

1: Given N examples $(x_1, y_1), \ldots, (x_N, y_N)$ with $x \in \Re^k, y_i \in \{-1, 1\}$
2: Initialize example weights $w_i = 1/N$
3: **for** $m = 1, 2, \ldots, M$ **do**
4: Select the weak classifier:
$f_m(x) = P_w(y = 1|x) - P_w(y = -1|x)$
which minimizes the weighted least squares error:
$e_m = \sum_i^N w_i (y_i - f_m(x_i))^2$
where P_w is the weighted class probability estimate.
5: Update the example weights: $w_i \leftarrow w_i \, exp\left(-y_i f_m(x_i)\right)$ and normalize so that $\sum_i^N w_i = 1$
6: **end for**
7: **return** $sign[\sum_{m=1}^M f_m(x)]$

Figure. 3.7 Gentle AdaBoost training algorithm [39].

Once we have selected an adequate object representation to extract features from target samples and non-target samples and an adequate classifier to learn to differentiate between the two categories, the typical design of an ATR system is complete. In our case, Haar features and AdaBoost may be used to design a fully-fledged ATR system. However, such an ATR may not be able to operate at the desired speed and accuracy required in this thesis. The speed deficiency of this ATR is mainly attributed to the large number of features that need to be extracted from each patch in the image regardless of its complexity (whether it is a flat seabed or an object-like rock). The accuracy deficiency of this ATR may be attributed to the limited number of non-target samples which can be learnt by a single classifier. Such a phenomenon has previously been tackled in the computer vision community for face detection using an ensemble of classifiers combined in a simple and intuitive structure called the cascade. This structure is described in the following section.

3.2.3 The Cascade Structure

The third key component of the ATR system proposed in this thesis is a simple arrangement of classifiers, called the cascade. It enables the underlying classification algorithm, AdaBoost in this thesis, to operate more efficiently in terms of processing speed and classification accuracy. The cascade is explained and discussed in this section.

The cascade is a sequence of increasingly specialized classifiers, each trained to identify almost all target samples and a moderate fraction of the false positives of previous stages. One stage classifier is trained and added to the cascade at a time until

the desired performance is achieved. Once the cascade is constructed, when a sample enters the cascade, it is examined by the first stage which either rejects it (thereby classified as a non-target sample) or passes it on to the next the stage for further scrutiny. This procedure is followed by all subsequent stages until the sample is rejected or passes all the stages and is thereby classified as a target sample. Figure. 3.1 shows the cascade within the architecture of the ATR framework presented in this thesis. Figure. 3.8 shows a schematic depiction of the cascade including AdaBoost as a stage classifier and the decision stump as a weak classifier within AdaBoost.

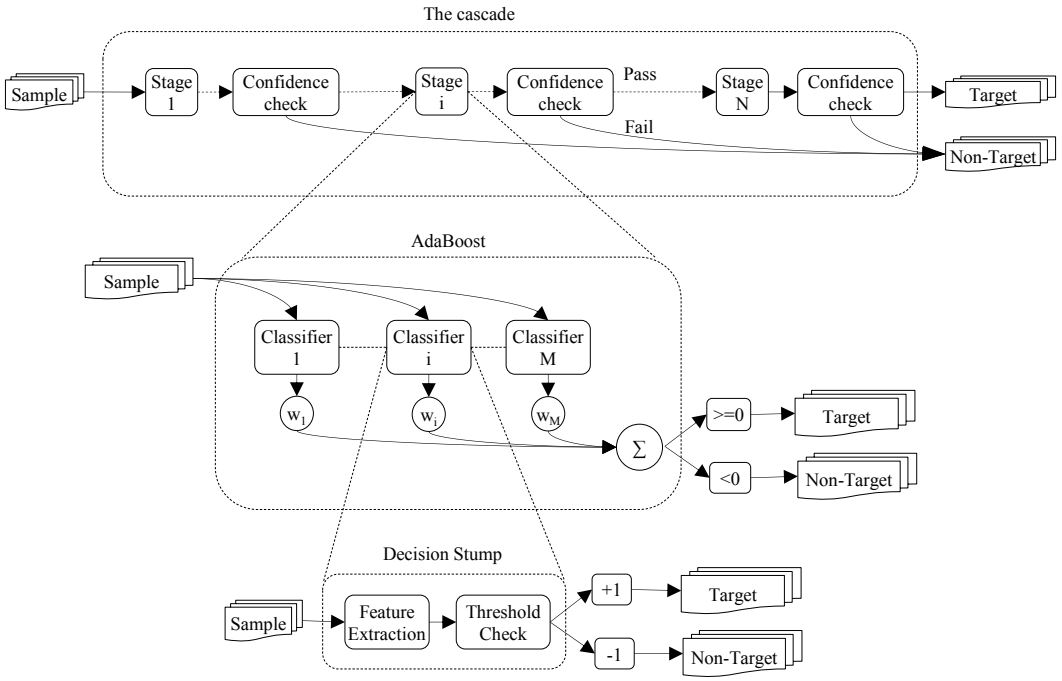

Figure. 3.8 Schematic depiction of the cascade including AdaBoost as a stage classifier and decision stump as a weak classifier within AdaBoost.

Standard classification algorithms may struggle when applied to the problem of object detection, where the discrimination is required between the object class and the rest of the world. This may be mainly attributed to the high imbalanced data, where the number of non-target samples is typically much higher than the number of object samples. Therefore, it can be computationally inefficient to use a single classifier. The computational burden appears in both the train phase and the test phase of the classification algorithm. In the train phase, the classifier is required to learn a very large number of non-target samples in order to achieve low false alarm rate, which might be prohibitively expensive. In the test phase, the classifier is required to scan across images in a brute force fashion in order to find the objects of interest, which might also be prohibitively expensive.

The cascade classifier overcomes these obstacles. It examines the clutter incrementally in the training phase, so a limited number is learnt in every stage. In the

test phase, the cascade rejects as many of the clutter samples as possible at the earliest stage possible before much time has been invested on them. The key insight is that the majority of samples in a typical image are simple non-target samples that can be identified by simple classifiers (i.e. fewer features and therefore computationally more efficient). Whenever identified, non-target samples can be rejected from the cascade process, while keeping all the remaining samples to be processed by later more complex classifiers.

The process by which the cascade classifier is trained requires some care. A generic outline of the learning algorithm of the cascade classifier is provided in Figure. 3.9. One of the critical steps in training a cascade classifier is deciding when to stop training one stage and move on to the next. The original approach of Viola and Jones [1] trains all stages under a fixed goal as shown in Figure. 3.9. This is the minimum detection rate per stage d and the maximum false positive rate per stage f. This solution can be argued because the stages of the cascade face increasingly more complex classification problems throughout the cascade. Nevertheless, this solution appears to produce efficient cascade classifiers in practice. This issue requires further research.

1: given:
   P : set of positive (target) examples
   N : set of negative (non-target) examples
   F : desired false alarm rate of the system
   S : maximum number of stages
   d : minimum detection rate of a stage
   f : maximum false positive rate of a stage
   M : number of negative samples used to train each stage
2: **repeat**
3: Add a new stage to the cascade. The stage is trained using all positive examples P and a number (M) of false positives randomly collected from the negative set N. The goal is to satisfy detection rate d and false positive rate f.
4: Evaluate the current cascade on the set of all negative samples N.
5: **until** the desired false alarm F is achieved or the number of stages reaches the maximum S

Figure. 3.9 The cascade training algorithm.

Another critical step in training a cascade classifier is how to balance the detection rate and false alarm rate within a stage. The original approach of Viola and Jones [1] simply adjusts the AdaBoost threshold after every iteration to balance the detection rate and the false alarm rate. This solution is not well founded as it is completely independent from the feature selection process. However, results are good in practice. Further research should consider this issue.

The cascade structure is biologically plausible due to its links to biological vision. In particular, there is evidence that primates appear to gradually process subsets of the

available visual information in interpreting complex scenes [49].

The idea of using cascade-like structures for object detection is not new. It has existed for decades, as pointed out in [50]. It has been used implicitly to filter out non-target samples based on various criteria [51, 52]. The overall form of the cascade is that of a degenerate decision tree [53], where subsequent classifiers are trained using examples which pass all previous classifiers. Nevertheless, the work of Viola and Jones [1] can be credited with the widespread popularity of the cascaded detectors.

While the cascade classier has demonstrated impressive results in speeding up the detection process, it has several disadvantages, some of which will be addressed in this thesis. Fundamentally, the training phase of the cascade is a very time-consuming task. The training time of the cascade was reported on the order of days or even weeks in several publications. This issue represents one of the most serious obstacles to wider use of the cascade classifier. This problem will be addressed in Section 3.4.

Another disadvantage of the cascade structure is the rejection procedure. The cascade rejects a sample when it first fails a stage. While this procedure is mainly attributed to the tremendous computational efficiency of the cascade, it might limit the detection performance as will be discussed in Section 3.5.3.

By introducing the cascade classifier in this section after AdaBoost and Haar features, the architecture of the proposed approach has been fully explained. Having this architecture well understood, some limitations start to arise and some ideas to overcome these limitations start to develop. This is what will be discussed in the remaining sections of this chapter.

3.3 Detection Confidence

ATR approaches typically assign a class label to every input sample. Some ATR approaches also return a real value, commonly referred to as the confidence value, which represents how confident the approach is in labelling the corresponding input sample. The confidence value is very important for making decisions. It assures the human operators of their decisions. It also allows the on-the-fly adaptation of Autonomous Underwater Vehicle (AUV) missions. For example, the close inspection of possible target locations could be prioritised based on the confidence values (some may be completely ignored).

The standard cascade classifier produces a binary classification rule (target or non-target). However, the number of co-located detections is traditionally used in the cascade classifier to measure the confidence. While this method appears to produce reasonable confidence values, its foundation can be argued. It is incapable of measuring the confidence in an individual sample. Empirical evidence also shows that this method is very sensitive to some factors such as:

- The strength of the classifier (level of training): the stronger the classifier the fewer the number of detections around a sample.
- The complexity of the context: simple backgrounds appear to allow more hits around a target than complex backgrounds.
- The resolution of the image: the higher the resolution the higher the number of neighbouring detections.
- The density of scanning the image: the smaller the step in shifting the window across the image, the higher the number of neighbouring detections.

Therefore, a more reliable method is required to measure the confidence. We propose a new method to measure the confidence based on combining the confidences of individual stages of the cascade. Our method measures the confidence in every sample regardless of neighbouring samples.

Due to the fact that stages of the cascade gradually become more complex, the reliability of each stage increases gradually in the cascade. Therefore, in our confidence measure, the contribution of each stage increases gradually based on its order in the cascade. This is implemented by multiplying the confidence of each stage by the stage number before combining them.

Due to the scanning nature of the detector, multiple detections often appear in the neighbourhood of a potential object. The number of co-located detections is traditionally the confidence measure. However, after introducing the confidence measure for individual detections, we select the maximum confidence within each neighbourhood instead.

To support our assumptions above, we conducted the following experiment on the real Synthetic Aperture Sonar (SAS) dataset of truncated cones (see Section 4.5 for more details about the data). We evaluated the same detector on the same data using two evaluation methods. The first method is the traditional method which is based on the number of co-located detections. The second method is our proposed method which is based on the weighted combination of the confidence values of the stages. Figure. 3.10 shows the Receiver Operator Characteristics (ROC) curves produced using each method. Within the context of object detection, the ROC curve is a graphical illustration of the performance; plotting the probability of detection against the probability of false alarm as the discrimination threshold (the confidence value in our case) is varied (see [54] for more details about ROC analysis). As Figure. 3.10 shows, the detector which uses the proposed confidence value outperforms the detector which uses the traditional confidence value.

We have on several occasions throughout the experiments conducted in this thesis compared the performance of the confidence value proposed in this section to the tra-

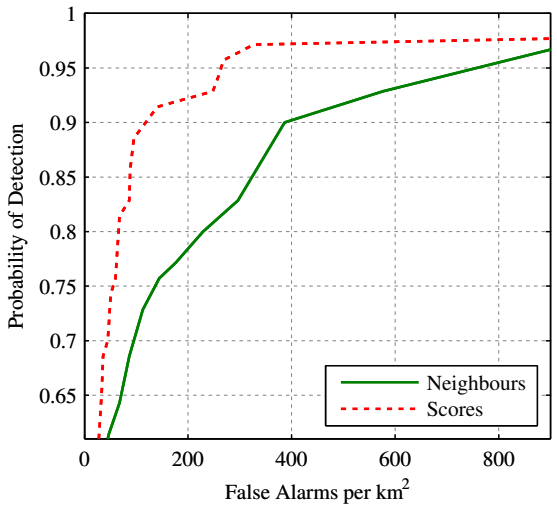

Figure. 3.10 Comparison of ROC curves for the same detector: once using the standard confidence values (Neighbours) and once using the proposed confidence value (Scores). The real SAS data of truncated cones is used.

ditional one. We found that the proposed confidence measure is often comparable to or better than the traditional one. The superiority of the proposed measure becomes clear when images are not very densely scanned. The density of the scan is traditionally reduced to increase the computational efficiency of the detector. This is not really needed when processing data of low resolution and low frame rate (e.g. sidescan sonar). However, the problem appears when dealing with sensors of high resolution (SAS) and high frame rate (e.g. Blueview, Dual Frequency Identification Sonar (DIDSON)).

Both confidence measures, the traditional one and the one proposed in this section, obviously carry useful information. Therefore, we assume that a new measure which is based on a combination of these two measures may be even better than any of them individually. We tried some basic fusion techniques, but they did not produce any better results. One interpretation could be the clear correlation between the two measures as shown in the scatter plot in Figure. 3.11.

We have also looked at this problem from a different angle. Since the cascade classifier is not binary anymore and the proposed confidence measure assigns a value to every individual detection, we assume that the confidence values of co-located detections may be fused rather than just taking the maximum value. Therefore, we studied the confidence value distribution of co-located detections. We found that the confidence value is the highest at the exact position of the target and it degrades gradually away from it as Figure. 3.12 shows. Again simple fusion techniques such as averaging did not improve the performance. Dalal in [55] faced a similar problem and proposed a solution based on kernel-density estimation. Every detection represents a point in a 3D space (2D position and scale) weighted by its SVM response. The corresponding density function is then estimated using a kernel based approach. The final detections

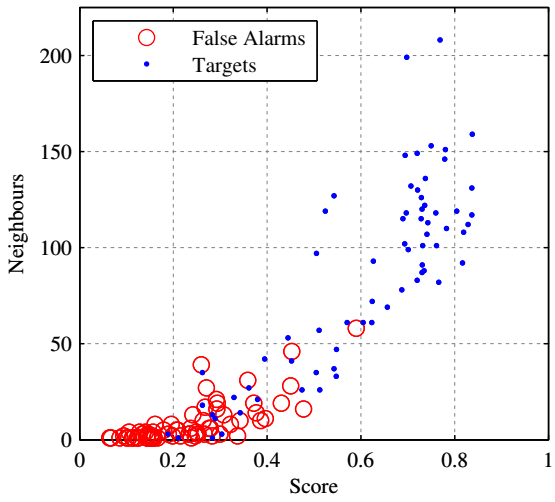

Figure. 3.11 Scatter plot of the proposed confidence value (Score) versus the traditional confidence value (Neighbours). As the graph shows, there is a clear correlation between the two values.

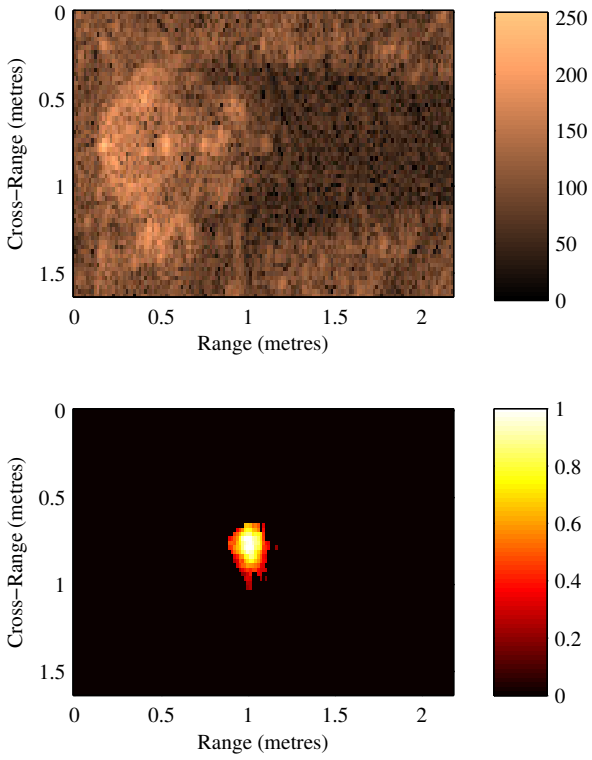

Figure. 3.12 The distribution of the confidence value within the target neighbourhood. An example of a truncated cone target from the real SAS dataset is shown on the top. The proposed confidence response is shown on the bottom. As the graph shows, the confidence is maximised at the centre of the target and drops gradually away from it.

are the peaks of the estimated density function. While this solution may be extended to suit the ATR system proposed in this thesis, we have not tried it due to the following reasons:

- Contrary to SVM which produces a response to each possible position in the image, our ATR produces responses to potential positions only. This is due to the use of the cascade which rejects the majority of positions in the early stages.
- The samples which we need this approach to update their confidence value mainly have a very low number of co-located detections.
- The approach is very sensitive to several parameters which need to be set carefully.

In conclusion, the confidence value proposed in this section is recommended against the traditional one due to the following reasons:

- Better intuitive basis.
- More reliable experimentally.
- Capable of dealing with individual samples.
- Better for dealing with high frame rate sensors.
- Opens new doors for further research to fuse co-located detections.

3.4 Training Speed Optimisation

The training phase of the proposed approach is time consuming. It may take days or even weeks as reported in the literature and experienced in this thesis. This is a serious obstacle which needs to be alleviated to make the re-training practical when new data becomes available (e.g. during or after a Mine Countermeasures (MCM) mission of a new region or new mines). Efficient training is also important to carry out the training on-board AUVs of limited resources. Moreover, fast training is important to make the work with other aspects of the algorithm practical to test from the time point of view.

The computer intensive nature of this framework training phase makes it time consuming. This is attributed mainly to two processes: the negative sampling and the feature selection. The training phase requires sampling the negative set (background) repeatedly to collect data to train each stage. While this process dominates the training cost of the cascade, it is also the step most attributed to the low false alarm rate achieved by the cascade. This issue will be discussed in detail in Section 3.4.1. We will propose an optimisation to the way of collecting negative samples which will significantly speed up the training.

Moreover, the training phase repeatedly employs a feature selection procedure, AdaBoost, which requires high processing power. This procedure is computer intensive

because it involves building a weak classifier for every possible feature, computing the misclassification error for every weak classifier, and selecting the weak classifier which has the minimum error. Sample weights are then updated and the feature selection procedure is repeated (see Figure. 3.5). Clearly, there are two main factors which affect the efficiency of the feature selection procedure:

- Weak classifier training algorithm: the simpler the algorithm the faster the feature selection procedure. We will show in Section 3.4.2 that a significant gain in the training speed can be achieved by pre-storing information about the feature values between AdaBoost iterations.

- The size of the feature space: The smaller the feature space, the faster the feature selection procedure. Two solutions will be proposed in Section 3.4.3 to reduce the size of the feature space. First, we show that a slight reduction in the feature space density speeds up the training significantly without any impact on the detection performance. Second, we will show that limiting early stages of the cascade to use coarse features and gradually introducing finer features throughout the cascade improves the training speed while also keeping comparable detection performance.

Some may argue that the number of samples used to train each stage has an impact on the training speed. While this issue is not covered in this thesis, we do not expect it to have a significant impact on the training speed because a cascade built using fewer samples to train each stage may require more stages to achieve the same training goal and consequently the same training time. It is also important to note that a small set of samples is less representative of the overall data than a larger set which may have a negative impact on the generalisation performance. Nevertheless, the optimal number of samples used to train each stage of the cascade is still an open question.

Some may also argue that the complexity of the weak classifiers used to build stage classifiers has an impact on the training speed. While this issue is also not covered in this thesis, we do not expect it to have a significant impact on the training speed because even though fewer complex classifiers will be needed, they will require longer training time each. Consequently, both cascades may require almost the same time to train. All the optimisation techniques proposed above will be discussed in more detail in the following subsections.

3.4.1 Negative Sampling

The process of collecting negative samples is computer intensive which slows down the training phase of the proposed approach. To train a stage of the cascade, negative samples are selected randomly and processed by the interim detector to check whether

they are valid to train this stage. A negative sample will only be valid to train a new stage i if it is misclassified by all previous stages (1 to $i-1$). Hence, the high processing power required to classify negative samples at each stage makes the sampling computer intensive. This is the problem addressed in this section.

The cascade classifier processes data extremely efficiently once it is trained. Therefore, we looked at the test phase, identified the principles attributed to the computational efficiency, and tried to employ them in the training phase. We found two concepts which we assume speed up the negative sampling. They are as follows:

- **The rejection procedure**: Based on the main insight of the cascade classifier, a sample is classified as non-target in the test phase when it first fails a stage of the cascade. This means no further processing will be required to classify this sample. This is where the cascade classifier mainly gets its tremendous processing speed in the test phase. We propose to employ this principle in the training phase of the cascade. This means when a negative sample is first rejected by a stage, it will never be checked again to train later stages. The main insight is that the samples rejected at a stage will never be used to train any subsequent stages. Subsequent stages are only trained on false alarms. We expect this principle to speed up the training phase of the cascade similar to the test phase.

- **The integral image**: Haar features can be computed very efficiently using the integral image representation as we saw in Section 3.2.1. To process an image by a cascade in the test phase, the integral image is calculated for the whole image and a window is shifted across the image to look for targets at all possible positions. Haar features are then extracted from the integral image based on the relative position of the scanning window. This means that only one integral image is calculated for the whole image rather than calculating an integral image at every position. This concept is not used in the training phase. This is perhaps due to the random nature of sampling. However, we assume this concept to speed up the training phase too and therefore we propose to use it in the training phase. This basically means each image from the negative training set is processed as a whole rather than dealing with each sample individually, which makes sense. This speeds up the sampling process without any impact on the detection performance.

To support our assumptions above, we conducted the following experiment on the real SAS dataset of truncated cones (see Section 4.5 for more details). We trained two detectors: one uses the traditional negative sampling procedure, and one uses a new negative sampling procedure based on the two ideas proposed earlier in this section. As a result, the proposed negative sampling procedure allowed the training to finish in less than a third of the time required by the traditional sampling procedure (from around 60

hours to around 19 hours). On the other hand, other aspects of the new detector, such as the number of stages, the number of features, and the detection performance are all very comparable to the traditional detector.

3.4.2 Weak Classifier Training

The weak classifier training algorithm is computer intensive which slows down the feature selection procedure and consequently the training phase. As described in Section 3.2.2, the weak classifier training algorithm searches for the optimal feature and feature threshold so that the misclassification error is minimized. This requires the extraction of all features from all the training samples and repeatedly sorting these samples based on their feature values.

For every feature, the list of training examples is sorted based on the corresponding feature values, so the optimal threshold can be computed with one pass over the sorted list. The optimal feature threshold is eventually the average feature value of two adjacent examples in the sorted list. This is repeated for all features to finally find the optimal combination of feature and threshold (optimal weak classifier). This is a computationally very expensive procedure which will be optimised in this section.

Features do not change value between rounds because AdaBoost only updates the weights of the samples. Therefore, features can be extracted only once, stored in memory, and reused in all rounds. However, these values need to be sorted repeatedly to select the optimal threshold. Moreover, a very large space is required to store these values (gigabytes).

We observe that once the list of examples is sorted based on a feature value, the position of the optimal threshold can be found without using the feature values. Once the position of the optimal threshold is found, only two feature values are required to compute the value of the optimal threshold. Therefore, there is no need to pre-store feature values. We only pre-store the indexes of the sorted list of samples. This means that feature values are computed and sorted only once which saves significant processing power. 50% of memory is also saved as the indexes need less space than the values (unsigned short (2 bytes) versus float (4 bytes)).

The modification proposed in this section does not require experimental proof as it only optimises the computational performance without changing the algorithm (i.e. without any impact on the detection performance). This optimisation was developed and adopted early in this research. All the experiments in this thesis including the ones presented before this section utilise this optimisation. We do believe that a significant reduction in the training time is attributed to this optimisation. This reduction is very hard to quantify because if this optimisation is disabled, the training becomes extremely long (from days to weeks). It is important to note that the idea proposed in this

section is partially inspired by an implementation of the decision tree in the OpenCV library [56]. Nevertheless, the OpenCV implementation was not documented and very hard to understand and reuse.

3.4.3 Feature Space Density

The feature space is very large which slows down the feature selection procedure and consequently the training phase of the proposed approach. This is the problem addressed in this section.

The feature space is very large because it is generated by repeatedly shifting and scaling each feature template (see Figure. 3.2) by one pixel at a time. This feature space is highly redundant and AdaBoost is therefore used to select the most relevant features. However, the large size of the feature space slows down AdaBoost itself and it therefore needs to be reduced.

Alternative feature selection methods could be used to replace or work with AdaBoost. They have their own computational burden which may not be any less than the computational burden of AdaBoost. They will also most likely have an impact on the detection performance of the proposed approach. The investigation of such methods lies outside the scope of this thesis. We here focus on the possibility of reducing the size of the feature space before applying AdaBoost without any extra computational cost or impact on the detection accuracy. We therefore propose the following ideas:

- We remove very fine scale features from the feature space. The insight is that very fine features are likely to capture very fine details, such as noise and intra-class variations, which may not characterise the object well.

- We reduce the density of the feature space which is many times over-complete. In comparison with the standard Haar wavelet transform, which generates complete basis (linearly independent), a much denser set of non-standard wavelets is used in this thesis to provide a richer representation following the work in [1]. Therefore, we do not expect a small reduction in the density to have a great impact on the detection performance.

- We propose to start the classification at coarse scales and gradually introduce finer scales. This idea is biologically plausible as it has been shown in the human visual pathway that the analysis of global features of a visual scene precedes the analysis of local features [57]. We integrate this idea into the cascade classifier by limiting early stages of the cascade to use coarse features only and gradually allowing later stages to use finer features. This idea clearly speeds up the training phase since smaller feature spaces will be used in the early stages of the cascade. We do not expect this idea to have a great impact on the detection performance

since the coarse-to-fine behaviour is already taking place in the cascade naturally (i.e. the majority of features selected in the early stages of the cascade are already coarse even when fine features are available for the selection process).

To support our assumptions about the first two ideas above, we conducted the following experiment on the real SAS dataset of truncated cones (see Section 4.5 for more details). We trained two detectors: one uses the original feature space, and one uses a feature space of reduced density. The feature space density was reduced based on several factors. First, the minimum size of a rectangle in any feature is restricted to 3x3 pixels rather than 1x1 pixels. Second, each feature is shifted by 2 pixels rather than 1 pixel. Third, each feature is scaled by 2 pixels at a time rather than 1 pixel at a time. As a result, the training time is reduced to 3% (from 19 hours to 35 minutes). As expected, the detection performance is comparable as Figure. 3.13 shows.

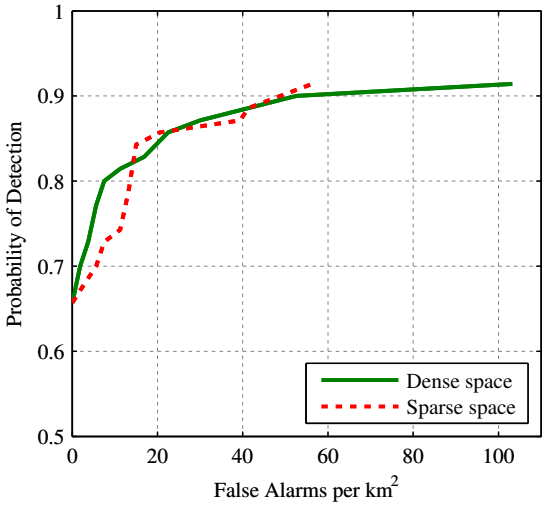

Figure. 3.13 Comparison of ROC curves for two detectors: one trained with the original space (dense space) as a benchmark, and one trained with a feature space of reduced density (sparse space). The real SAS data of truncated cones is used.

To support our assumptions about the coarse-to-fine idea above, we conducted the following experiment on the real SAS dataset of truncated cones. We trained two detectors: a standard detector which uses all features in all stages, and a new detector with some restrictions on the minimum feature size in the first five stages (27x27, 21x21, 15x15, 9x9, and 3x3 respectively). As a result, the training took 23% less time and the detection performance is comparable or slightly better as Figure. 3.14 shows.

It is worth mentioning that while we propose to integrate the coarse-to-fine idea into the cascade structure, there might also be alternative or complementary solutions which leverage this idea. For example, AdaBoost could also be modified to select its early features at high scales and gradually search in finer scales.

The computer vision literature is abundant in coarse-to-fine approaches. As the use of the term coarse-to-fine in the literature typically refers to the use of the cascade

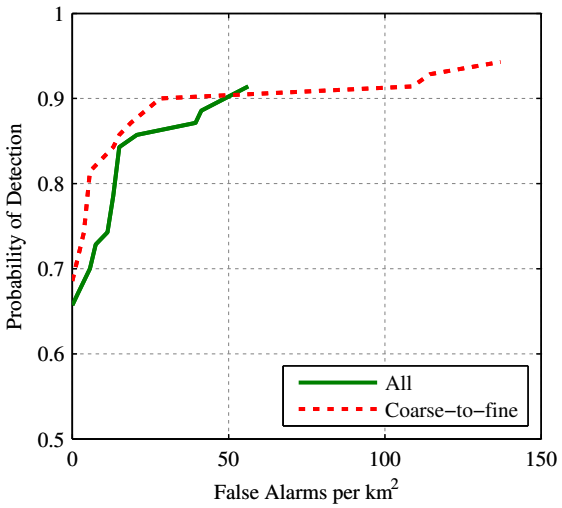

Figure. 3.14 Comparison of ROC curves for two detectors: one trained with all features as a benchmark (All), and one trained with coarse-to-fine features. The real SAS data of truncated cones is used.

classifier, it is not easy to find work similar to the one presented here which uses the coarse-to-fine concept together with the cascade classifier. While we have not found such work in our search, some works have been found which apply the coarse-to-fine idea differently. For instance, starting the classification with a large patch size (analysis window) and progressively reducing the patch size has been tried in some works such as [58]. Starting the classification at a low image resolution and progressively increasing the resolution has also been tried in some works such as [59].

3.5 Detection Optimisation

This section presents some extensions to improve the detection performance of the proposed approach. There are several factors which impact the performance of the proposed approach. Four different factors will be studied in the following subsections: context knowledge, object representation, rejection procedure, and context adaptation.

3.5.1 Context Knowledge

The poor structure within object signatures in sonar imagery limits the detection performance. The proposed approach was originally found to detect faces in optical imagery. Contrary to faces in optical imagery which include rich structure (eyes, eyebrows, nose, mouth, etc.), objects in sonar imagery do not. The highlight of the object in sonar imagery may not even include sufficient information to distinguish it from the clutter. Therefore, the shadow of the object is traditionally analysed to extract information about the object. This motivated us to include the shadow in all object samples

from the early stages of this research. However, the information within the highlight and shadow may still not be sufficient to achieve very high detection performance.

Therefore, we propose to encode some information about the region surrounding both the object and its shadow. We assume that the addition of such knowledge about the context will improve the detection performance. This can be attributed to the ability to encode new information such as:

- The contour of the object highlight.
- The contour of the object shadow.
- The brighter nature of the highlight in comparison with the surrounding region.
- The darker nature of the shadow in comparison with the surrounding region.

To support our assumptions above, we conducted the following experiment on the real SAS dataset of truncated cones (see Section 4.5 for more details). We trained two detectors: one learns from targets samples tightly cropped from the background, and one learns from target samples with additional 7 pixels from the surrounding background. We did not add context from the shadow side due to the changing length of the shadow. Samples of 125x55 pixels rather than 118x41 pixels are used (see Figure. 3.15). The addition of the context improves the performance as Figure. 3.16 shows.

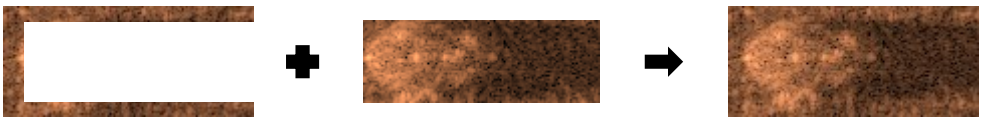

Figure. 3.15 An example target sample with an additional narrow frame from the background.

To further support our assumptions above, we also conducted a similar experiment on the Augmented Reality (AR) dataset (see Section 4.4 for more details). We trained a detector with 3 additional pixels from the context surrounding target samples. Samples of 40x12 pixels rather than 37x6 pixels are used. The addition of the context improves the performance as Figure. 3.17 shows.

Some may argue that the context varies and find it unreasonable to encode information about it. However, the approach proposed in this thesis does not use the complete representation (i.e. all features) of a sample in the classification procedure. It automatically selects the most relevant features. This means a feature related to the context will only be selected if it is somehow consistent within all or a certain group of the training samples from the same class.

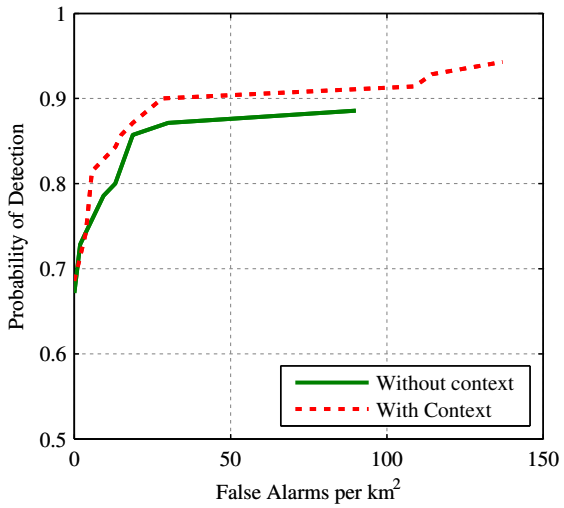

Figure. 3.16 Comparison of ROC curves for two detectors: one trained with target-only snippets (without context) as a benchmark, and one trained with targets snippets including narrow frames from the background (with context). The real SAS data of truncated cones is used.

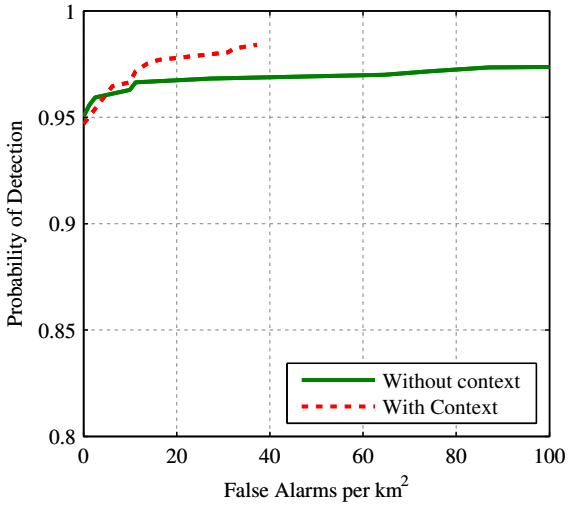

Figure. 3.17 Comparison of ROC curves for two detectors: one trained with target-only snippets (without context) as a benchmark, and one trained with targets snippets including narrow frames from the background (with context). The AR data of truncated cones is used.

3.5.2 Object Representation

Haar wavelet representation is well known to be effective for analysing the content of images. However, computing the complete set of features (wavelet coefficients) is very expensive. AdaBoost is therefore used in the proposed approach to select and combine a very small set of these features. However, the effectiveness of a small subset of Haar features can be argued. We assume that Haar features become primitive when they are used in a small subset which limits the detection performance of the proposed approach.

Therefore, we propose to use features more complex than Haar features. We as-

sume this will improve the performance by increasing the discriminative power of the classification. While Haar features could be completely replaced with complex feature such as HOG [36], Local Binary Patterns (LBP) [37], and Scale-Invariant Feature Transform (SIFT) [38], they are likely to slow down the detector. Real-time processing is one of the objectives of this thesis and therefore we constrained ourselves to extending the existing set of Haar features while keeping the computational efficiency intact.

Several attempts have been proposed in the literature to enhance the expressional capability of Haar features while keeping them computationally efficient. Extensions have mainly introduced variations in the number of rectangles and the way they are combined. For instance, the authors of [60] introduced 45 degree rotated Haar features. A generalisation of Haar features was introduced in [61] which allowed more flexible combinations of rectangle regions of different sizes and at certain distances apart. [62] proposed an extension of Haar features based on feature co-occurrence, where no change has been made to the original set of Haar features apart from combining the binarised values of multiple features.

Despite the great interest in Viola and Jones framework [1], where Haar features were first effectively used, we have not found comparative studies that address the impact of features on the performance of this framework. Conducting such a study sits outwith this thesis. We here introduce a new set of Haar like features which enriches the existing set by encoding new visual information. Two features are added to encode the diagonal information (see Figure. 3.18 (3)). Three features are added to encode the varying nature of the object shadow in sonar imagery (see Figure. 3.18 (4)).

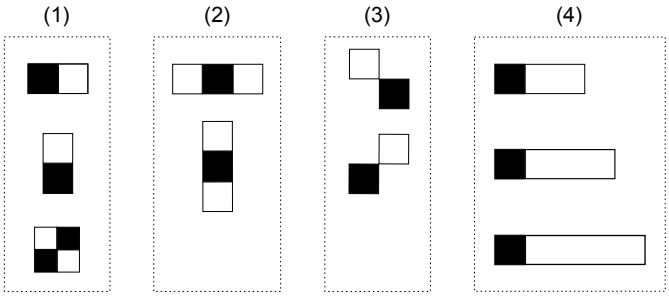

Figure. 3.18 Prototypes of the original set of Haar features and our extended set. (1) 2-dimensional Haar wavelets used in [63], (2) used in [1], (3) our diagonal features, and (4) our range features.

To support our assumptions above, we conducted the following experiment on the real SAS dataset of truncated cones (see Section 4.5 for more details). We trained two detectors: one uses the original set of features, and one uses the extended set. As Figure. 3.19 shows, the extended set outperforms the original set at very high detection rates only.

To further support our assumptions above, we also conducted a similar experiment

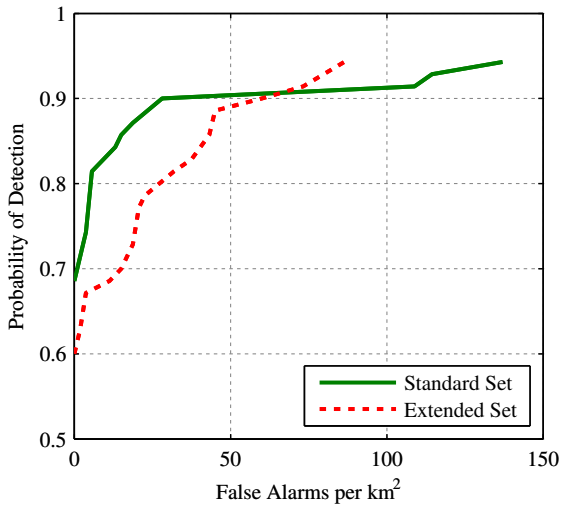

Figure. 3.19 Comparison of ROC curves for two detectors: one trained with the standard set of features as a benchmark, and one trained with the extended set of features. The real SAS data of truncated cones is used.

on the AR dataset of truncated cones (see Section 4.4 for more details). Similar to our previous experiment, we trained two detectors: one with the original set of features, and one with the extended set. As Figure. 3.20 shows, the extended set improves the performance at high detection rates.

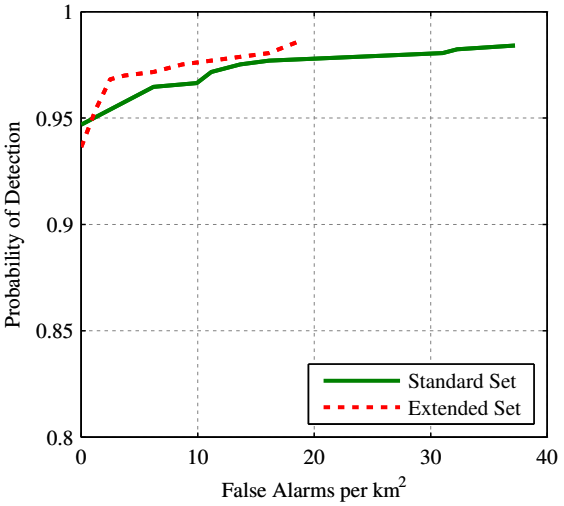

Figure. 3.20 Comparison of ROC curves for two detectors: one trained with the standard set of features as a benchmark, and one trained with the extended set of features. The AR data of truncated cones is used.

Based on the initial experimentation carried out here, the extended set of features appears to reduce the number of false alarms at high detection rates. Therefore, the extended set of features could be used for applications which require very high detection rate such as MCM.

3.5.3 Rejection Procedure

The cascade classifier rejects a sample when it first fails a stage which limits the performance of the proposed approach. Two solutions are proposed in this section to alleviate this problem.

The Exploitation of Historic Information

We propose to improve the rejection procedure by exploiting historic information. The traditional cascade rejects a sample when it first fails a stage. Even though this implies that the sample had passed all previous stages, the final decision is still made solely by the current stage. We assume previous stages to carry useful information which can be exploited to make a more reliable decision. Therefore, when a sample is rejected by a stage, we allow all previous stages to contribute in making the final decision. This is achieved by combining the confidence values of all previous stages similar to computing the confidence value of the cascade proposed in Section 3.3. This modification does not compromise the computational efficiency of the cascade because it only uses information which was previously extracted.

To support our assumptions above, we conducted the following experiment on the real SAS dataset of truncated cones (see Section 4.5 for more details). The same detector was run twice: once using the traditional rejection procedure, and once using the rejection procedure proposed above. As Figure. 3.21 shows, the proposed approach makes very high detection rates possible. However, the traditional approach outperforms the proposed approach at lower detection rates. Therefore, the proposed approach is only recommended for applications which require very high detection rates.

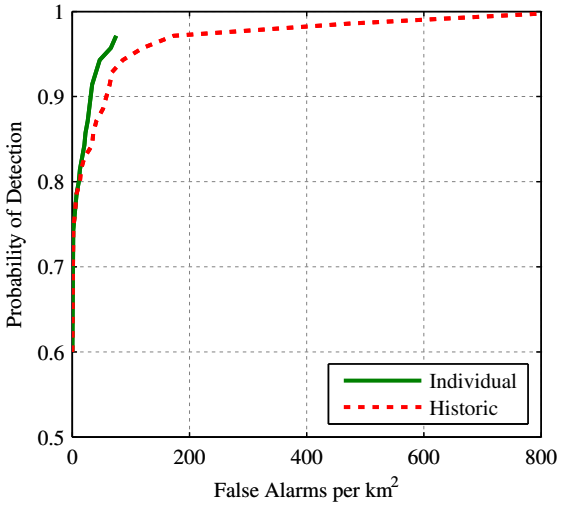

Figure. 3.21 Comparison of ROC curves for the same detector: once using the standard individual rejection, and once using a rejection based on historic information. The real SAS data of truncated cones is used.

To further support our assumptions above, we also conducted a similar experiment on the real SAS dataset of wedges. The motive is to use targets which are hard to learn because of their changing signature from different angles and their similarity to the clutter. As Figure. 3.22 shows, the modification proposed above allows higher detection rates.

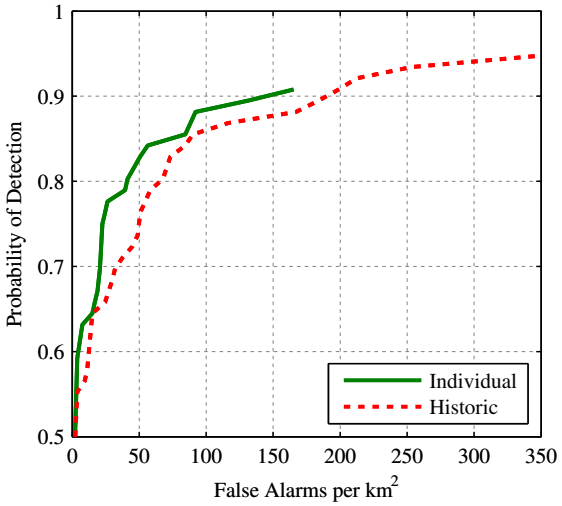

Figure. 3.22 Comparison of ROC curves for the same detector: once using the standard individual rejection and another using a rejection based on historic information. The real SAS data of wedges is used.

This solution does not require re-training. It can be easily optimised to suit the application and the underlying data. However, it can be extended to the training phase. In other words, knowledge from previous stages may be exploited in training a new stage. While this concept is not covered in this thesis, we found that some researchers have successfully tried similar concepts. For instance, in [64], rather than starting to learn each stage of the cascade from scratch, the cumulative sum of the confidence values of the previous stages is used as a prefix classifier for training the next stage. In [65], the feature value used to create the first weak classifier of a stage is the confidence value of the previous stage rather than the value of a Haar feature.

The Exploitation of Future Information

Contrary to the previous solution, which exploits historic information to improve the rejection procedure, this solution looks at exploiting future information. The insight of the cascade is to reject samples at the earliest stage possible without the need to extract all information. Therefore, any extraction of future information may compromise the well-known computational efficiency of the cascade. However, we assume that some future information can be exploited with a trivial compromise to the computational efficiency. We propose to give a sample more than one chance before rejecting it. In other words, a sample is allowed to fail multiple stages before it gets rejected. The

motive is that a negative sample is more likely to be rejected multiple times than a target sample. In other words, when a sample fails a stage it is more likely to fail a future stage if it is negative than if it is positive.

To support our assumptions above, we conducted the following experiment on the real SAS dataset of truncated cones. We ran the same detector three times: one rejects a sample when it first fails a stage, one gives a sample a second chance, and one gives a sample three chances. As Figure. 3.23 shows, the proposed approach makes very high detection rates possible. However, the traditional approach outperforms the proposed approach at lower detection rates. Therefore, the proposed approach is particularly useful for applications which strictly require very high detection rate and can tolerate some extra false alarms.

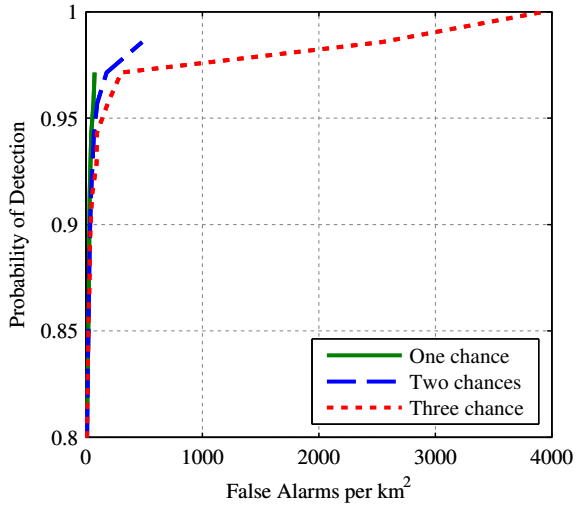

Figure. 3.23 Comparison of ROC curves for the same detector given different number of chances in the rejection procedure. The real SAS data of truncated cones is used.

To further support our assumptions above, we also conducted a similar experiment on the real SAS dataset of wedges. The motive is to use a target hard to detect because of its complex shape. As Figure. 3.24 shows, the modification proposed above allows much higher detection rates while keeping comparable performance at low detection rates.

This solution does not require re-training. It can be easily optimised to suit the application and the underlying data. However, it can be extended to the training phase, where multiple stages learn to reject the same sample.

In conclusion, the reject procedure of the cascade at any stage can be improved by exploiting information from previous or/and successive stages in the cascade. This allows very high detection rates which were not possible before. However, the gain in the detection rate is often accompanied with a higher number of false alarms. Therefore, this extension to the rejection procedure is particularly useful for applications which strictly require very high detection rate and can tolerate some extra false alarms.

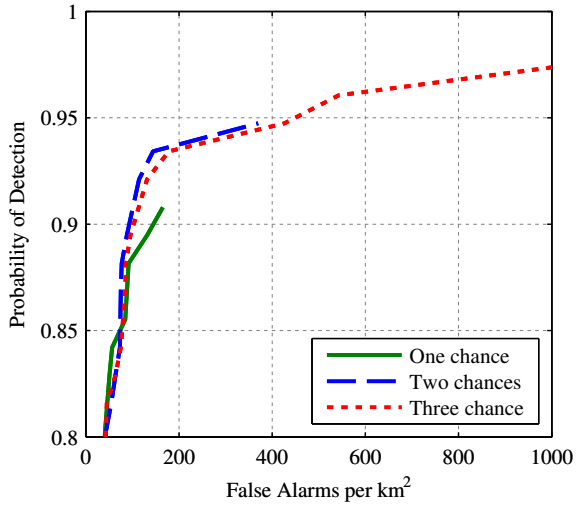

Figure. 3.24 Comparison of ROC curves for the same detector given different number of chances in the rejection procedure. The real SAS data of wedges is used.

For example, it might be required to have the attention of the human operators at every potential target location, and some extra false alarms may not be a problem. The high detection rate is also particularly useful if the potential target locations are post-processed by another classification or identification framework, where a missed target will never have the chance of post-processing.

3.5.4 Context Adaptation

The changing context represents one of the fundamental reasons for the poor performance of ATR algorithms in general. Empirical evidence has also shown that the detection performance of the proposed approach varies based on the context. We looked at the context before in Section 3.5.1 to extract more information about the object. We found that the addition of a narrow frame from the region surrounding the object improves the detection performance. As the variations in sonar data are extreme, a frame from the context may be necessary yet insufficient to achieve high detection performance.

Therefore, we propose to make our ATR approach adaptive to the changing context. This can be achieved in various ways such as: multiple classifiers each trained on a different seabed type, or one classifier which can adjust its decision boundaries based on the seabed type. The concept of adapting a classifier to changes in data characteristics is not new. It is a general problem treated by the machine learning community.

None of the datasets used in this thesis include ground truth of the seabed type. This discouraged us from training context-based detectors or allowing a detector to adapt automatically to the context. However, to prove the feasibility of context adaptation under the ATR approach proposed in this thesis, we conducted the following experi-

ment. We ran a detector on the test dataset and manually associated each response with a seabed type based on the region surrounding the response. For simplicity, we used only three labels: flat, rippled, and complex. We then studied the distribution of the real targets and the false alarms based on the confidence value discussed in Section 3.3. Figure. 3.25 shows the distributions.

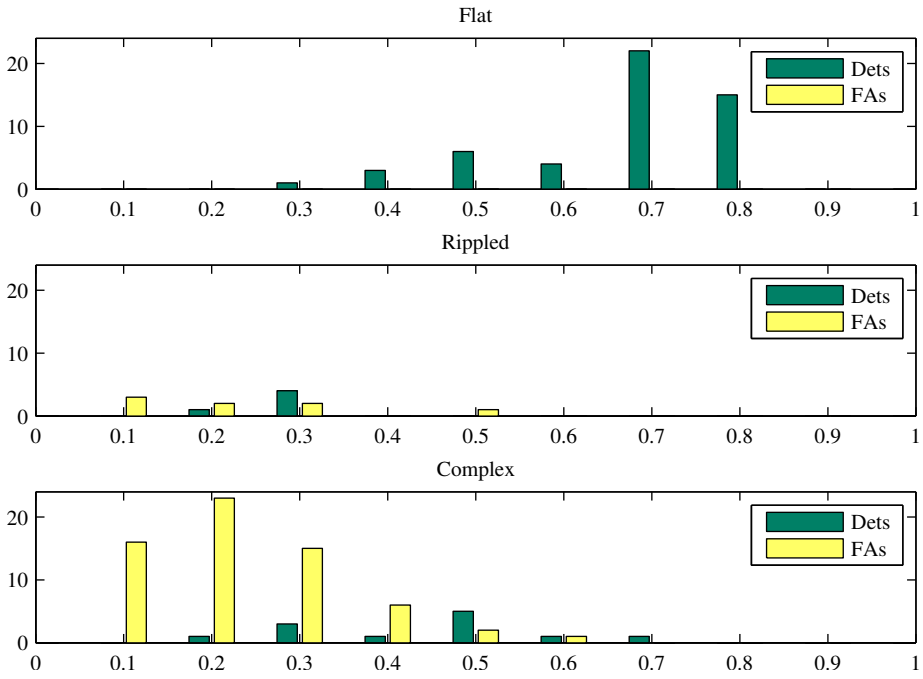

Figure. 3.25 The histograms of the ATR responses (confidence values) in three different contexts (seabed types): flat, rippled, and complex. The responses from true detections (Dets) are shown along with the responses from false alarms (FAs).

As Figure. 3.25 shows, the detector performs perfectly on flat context and there is no need to threshold the confidence value. This means if the ATR returns a potential target and a seabed classifier labels the surrounding region as flat region, it is very likely that a real target is detected.

On the other hand, the detector returns some false alarms in rippled and complex regions. Many of these false alarms can be removed based on the confidence value. We can easily notice from Figure. 3.25 that the detector mostly does not give high confidence for targets on rippled and complex regions. We assume this is a direct result of the distorted target samples in such regions. For example, the highlight of the target may be located on the highlight of the ripple and loses its contours. The highlight may also be located in the shadow of a ripple and consequently disappears partially or completely. The shadow of the target may overlap with the shadow of a ripple and loses its contours. All of such scenarios allow some weak classifiers to vote negatively which results in low confidence values.

By looking at Figure. 3.25 some may argue that the proposed approach is not very effective in rippled and complex regions. Nevertheless, by checking the data, they will

find out that the detector is able to detect targets the human operator failed to detect. They will also find out that most false alarms are anomalies within their context, and some applications such as MCM may require highlighting such locations for closer inspection. Various examples will be displayed and discussed in the following chapter.

Based on the limited evaluation carried out here, a reasonable gain in the detection performance is expected by allowing the proposed approach to adapt to the context. Further evaluation on other datasets would be required to verify the results obtained here.

An interesting research question is: what if the context information is really unknown or unavailable? For example, the seabed classification algorithm may fail to classify a completely new type of seabed which is not an unlikely situation in the underwater environment. We assume that some information about the context can be inferred indirectly from the ATR responses. ATR responses within each neighbourhood could be dealt with independently to adjust the decision boundaries within the same neighbourhood. This can be exploited differently in different applications. For example:

- Regions of too many responses could be avoided completely.
- Only the response with the highest confidence within a neighbourhood could be selected for further inspection.
- The distribution of the responses within the same neighbourhood could be analysed before making the final decision. For example, if all responses are alike, they are most likely to be false alarms.

To support our assumptions above, we conducted the following experiment on the real SAS dataset of truncated cones (see Section 4.5 for more details about the data). We studied the responses of our ATR on each image independently, under the assumption that every image is most likely to include homogeneous context. We ran the same detector two times: one returns all possible detections, and one returns the best two detections per image if any exists. Best here refers to the highest confidence value. The choice of selecting only two detections per image was made based on a priori knowledge of the highest number of real targets expected per image area. As Figure. 3.26 shows, the proposed change significantly reduces the false alarm rate while keeping the detection rate high.

To further support our assumptions above, we also conducted a similar experiment on the real SAS dataset of wedges. The motive is to use a target hard to detect. As Figure. 3.27 shows, the modification proposed above significantly improves the ATR performance by reducing the number of false alarms at high detection rates.

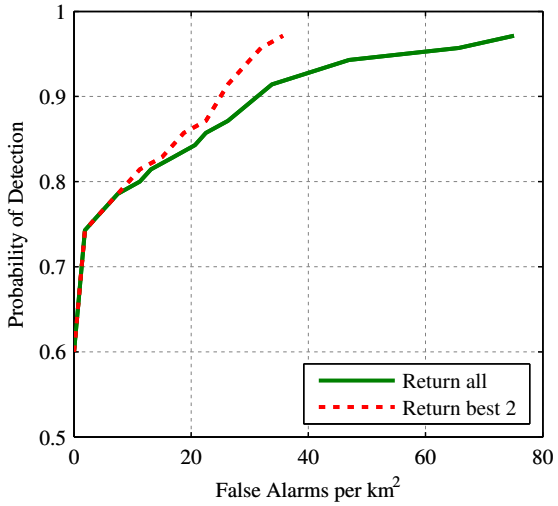

Figure. 3.26 Comparison of ROC curves for the same detectors: once returning all the detections found, and once returning the best two detections per image. The real SAS data of truncated cones is used.

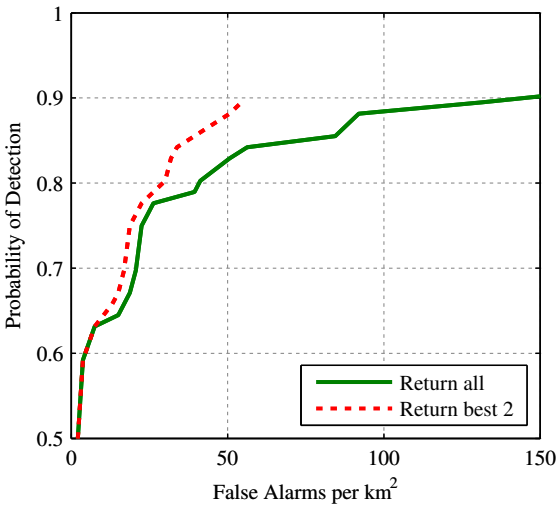

Figure. 3.27 Comparison of ROC curves for the same detectors: once returning all the detections found, and once returning the best two detections per image. The real SAS data of wedges is used.

3.6 External Effects

Various aspects internal to the proposed approach have been investigated so far in this chapter. This section investigates the impact of some external factors on the performance of proposed approach. In other words, the proposed approach will be only dealt with as a black box in this section while studying the impact of some external factors on its performance. Four different factors will be studied in the following subsections: data preprocessing, data post-processing, new data, and poor data.

3.6.1 The Impact of Pre-processing on the Performance

Sonar signals carry high levels of noise. The perturbations within the water column on the way from the sonar to the target and on the way back are one source of noise. Background noise and reverberation as the signal still propagates in the water column are other sources of noise. Noise may also come from the sensor itself or neighbouring instruments. We assume that reducing the level of noise in the data may improve the detection performance. We tried several approaches to improve the appearance of sonar imagery such as the Median filter and histogram equalization. Such approaches made sonar images easier to interpret by humans, but they did not seem to improve the performance of the computer aided ATR approach proposed in this thesis.

We also looked at the image formation procedure used to convert the raw sonar signal to the visual data (images) used by our ATR. Sonar sensors have a large dynamic range to allow the detection of weakly scattered echoes in the presence of high reflections. Due to the large dynamic range, the logarithmic scale is traditionally used to cover the wide spectrum of the data. To reduce the noise and form an image easier to interpret, we clip the signals of low strength by only taking the upper 80 dB (Decibel). We tried smaller dynamic ranges in an attempt to remove more of the noise in the signal and focus on the signals of high magnitudes which represent reflections from the target highlight. This resulted in clearer shadow regions and images easier to interpret by humans, but they did not improve the performance of the proposed approach. We have also tried to reduce the compression rate by scaling down the data before applying the logarithmic transform. We assumed this would help in preserving better information about the targets highlights. Again, this resulted in images easier to interpret by humans, but it did not improve the detection performance of the proposed ATR.

Sand ripples present a difficult challenge to existing ATR approaches including the one proposed in this thesis. This is due to the sequence of highlights and shadows sand ripples create in sonar images which may either resemble the highlight/shadow signatures of real objects or disturb them. As other authors have observed, reducing the effect of sand ripples prior to the ATR stage may improve the detection performance of the ATR. Here we consider one particular approach for sand ripple suppression due to Nelson and Kingsbury [66]. This approach is based on dual-tree wavelets and fractal dimension; see [66] for more details. The algorithm implementation was provided by O. Daniell (personal communication, November 23, 2010). It is shown in [66] that this approach improves the detection performance of a matched filtering based ATR. We have tried this approach with our ATR, but it did not improve the detection performance as Figure. 3.28 shows. One interpretation could be that our ATR outperforms the ATR used in [66] (based on the number of false alarms shown in the publication).

This is not surprising as the matched filter ATR used in [66] is simple and only chosen as a validation tool to compare the detection performance with and without ripple suppression. Within the context of our ATR, the interpretation could be that ripple suppression disturbs the object signature. Based on the limited evaluation carried out here, this ripple suppression approach does not appear to improve the detection accuracy of the ATR approach proposed in this thesis (under the caveat that the parameters of the suppression approach chosen in our experiment might not be optimal).

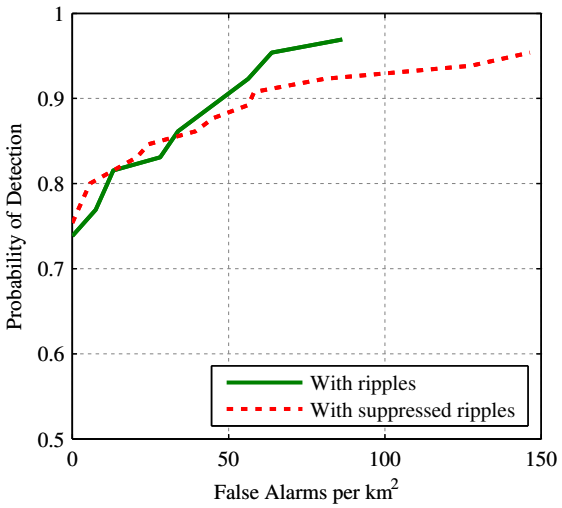

Figure. 3.28 Comparison of ROC curves for two detectors: one trained on the original data (With ripples), and one trained on the data with the sand ripples suppressed (With suppressed ripples). The real SAS data of truncated cones is used.

ATR preprocessing may potentially include techniques to measure the complexity of the data (the seabed). Such techniques may reduce the computational complexity of subsequent steps of the ATR architecture by removing very simple data (flat regions). We do not expect such techniques to improve the computational efficiency of our ATR approach because our approach has already got a focus of attention operator integrated within its architecture (the cascade). Nevertheless, the complexity measure can be useful in discarding extremely complex regions where the ATR is likely to break down (producing high number of false alarms). Human operators may then be asked to check the discarded regions.

3.6.2 Post-processing Using the Histogram of Oriented Gradients

The Histogram of Oriented Gradients (HOG) features were first introduced by Dalal and Triggs [36] in 2005 for the problem of pedestrian detection in optical imagery. Dalal and Triggs [36] show experimentally that HOG features with a linear SVM classifier outperform existing features for human detection. Since then HOG features have attracted attention and been extended in several publications [67, 68].

HOG features are reminiscent of Scale-Invariant Feature Transform (SIFT) features, but they are calculated on a dense grid of uniformly spaced regions. Contrast normalization is also carried out locally to improve the performance in HOG.

The HOG method is based on the idea that the shape of an object can be characterised well by the distribution of local gradient orientations without precise knowledge about their positions. This is achieved by dividing the image sample into small regions (cells), for each region evaluating a histogram of gradient orientations. The histograms are combined to form the HOG descriptor. Local regions bigger than cells (blocks) are normalized for better invariance to effects such as illumination and shadowing.

This approach has been identified from the computer vision community and selected for evaluation as a possible post-processing ATR component for the following reasons:

- The high resolution offered by SAS imagery now enables techniques traditionally used within the computer vision community to be evaluated: HOG features need a resolution which is high enough to create representative histograms.
- The edges of the highlight and shadow object regions show clearly in SAS: HOG features encode the existence of edges and their relative positions in object samples.
- The variable appearance of objects in sonar imagery: This is also a challenge in the application of pedestrian detection where the technique has already been successfully applied.

To evaluate the suitability of the HOG features for the particular application of ATR in SAS imagery, we ran a preliminary experiment to train and test a generic HOG-SVM detector. Our experiment follows the general form; though differ in detail, from those presented in [36]. It also employed the same real SAS dataset described in Section 4.5.1.

Figure. 3.29 compares the scores assigned to samples by both ATR approaches: HOG-SVM presented in this section and Haar-Cascade proposed in this thesis.

The figure shows that the HOG-SVM values for both the true positives (shown in blue) and the false alarms (red) are very similar and do not allow a clean separation between the two classes. Fixing a threshold on the y-axis (the HOG-SVM score) will only reduce many false alarms at the expense of removing some true positives. This is shown quantitatively in Table 3.1 below which shows the false alarm reduction achieved using the HOG-SVM as a post-processing module.

Table 3.1 shows a very modest reduction in false alarms for a given ATR detection rate using the HOG-SVM system as a post-processing module. Based on the limited

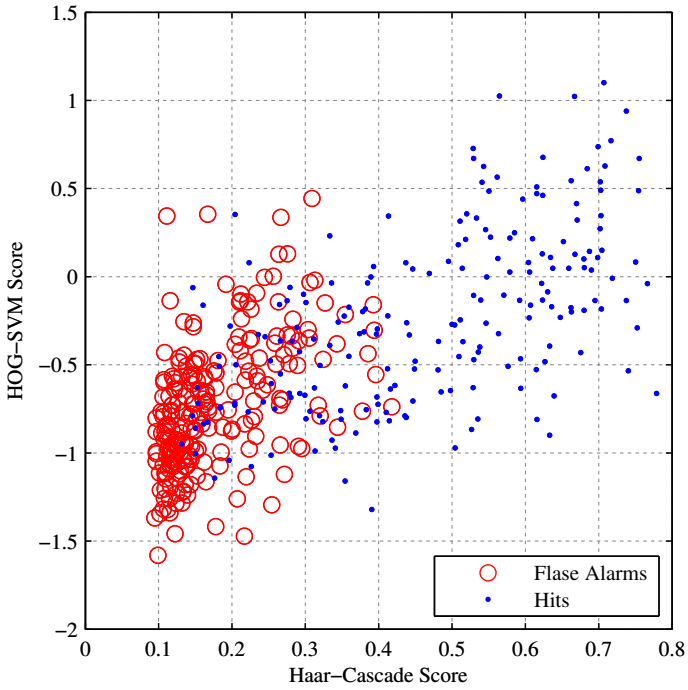

Figure. 3.29 Scatter plot of the SVM response produced by the HOG based ATR discussed in this section (HOG-SVM Score) versus the cascade response produced by the Haar based ATR discussed in Section 3.3 (Haar-Cascade Score). As the graph shows, the HOG-SVM score does not allow a clean separation between the true targets and false alarms detected by the Haar-based classifier.

Detection rate	Haar-Cascade false alarms	Haar-Cascade + HOG-SVM false alarms	Reduction in false alarm rate
95%	111	106	5%
90%	79	76	4%
85%	44	43	2%

Table 3.1 The false alarm reduction achieved through using the HOG-SVM as a post-processing module for a given ATR detection rate. This system was tested using the Haar-Cascade ATR proposed in this thesis.

evaluation carried out, the HOG features do not appear to be a robust feature for identifying and removing false alarms. The values obtained for true positives and false alarms do not allow a robust separation of the classes.

3.6.3 The Impact of New Data on the Performance

Supervised ATR systems require example training data to learn the signature of a target and build up a description of the background seafloor. When these systems are well trained and the training data matches the test data, performance results can be very good. The performance of these systems is dependent on the correlation between the

training and test samples.

The ability to deploy a supervised ATR into a new environment requires the impact of training data to be well understood and evaluated. Solutions for adaptive training / re-training and the matching of suitable training data to the particular deployment region to assure robust ATR performance need to be researched.

The training and transferability of these supervised ATR systems has been identified as a technical gap. This gap was partially addressed in this section which looked at the feasibility of ATR re-training. Preliminary research and results undertaken are presented here. Further work in the field of ATR training is required.

Supervised ATR algorithms are typically trained using all available data, but they are not typically re-trained when new data is collected because this can be costly and time consuming with no guarantee of improved results.

This section looks at the potential benefits from re-training and the impact this has on performance. Data from the real SAS dataset was again used for this evaluation. This data was collected from different sites around the NATO Undersea Research Centre which do contain different seafloor types but can be considered similar in look and appearance. Further evaluation should consider data sets from completely different regions and collected at very different times to further verify the results obtained here.

When new data becomes available, intuitively it is possible to re-train the ATR system using all the previous data along with the new data. If the data is similar in look and appearance, the inclusion of the new data allows the ATR to encode the information on the new targets and seafloor types present. The impact of doing this with very different data sets lies outside the scope of this thesis. It is likely that this would have a negative impact on performance, but this needs to be verified.

In our example, one dataset (SAS1) was used to initially train the ATR (see Section 4.5.1 for more details about the data). A second data set (SAS2) was used to carry out the re-training experiment. The new data is collected using the same sensor and has similar characteristics to the previous data. Table 3.2 compares between the two datasets used in the experiment.

Dataset	Number of images	Area (km2)	Number of target views
SAS1	201	1.071	441
SAS2	3352	18.675	Unknown

Table 3.2 Details of the data used in sets SAS1 and SAS2 for the re-training experiment. No ground truth was available for the data within SAS2.

To evaluate the impact of re-training, two detectors were built up. D1 was trained using a selection of data from training set SAS1. A second detector D2 was trained using the same set of data used for D1 along with a subset of the new data SAS2. The

subset of data from SAS2 was selected by running detector D1 on dataset SAS2 and choosing the images with the highest number of false alarms. This was necessary to overcome the problem of the very large size of dataset SAS2. 98 images from the data set SAS2 were used, which contained 20% of the false alarms.

Detectors D1 and D2 were then run on both datasets SAS1 and SAS2 for evaluation. Figure. 3.30 shows the results from this evaluation. The two lower performance curves (red and green) show the results of running detectors D1 and D2 on the entire data set (SAS1 and SAS2). As would be expected, D2 outperforms D1 due to data from both SAS1 and SAS2 being included in its training process.

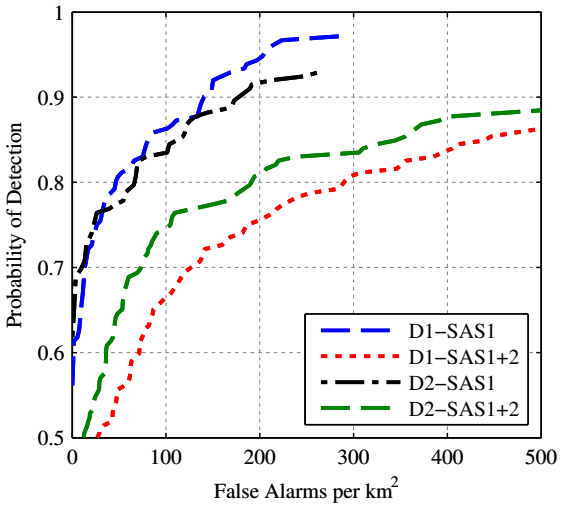

Figure. 3.30 ROC curves of detectors D1 and D2 running on data sets SAS1 and SAS2. The best performance is detector D1 running on data set SAS1 (which it has been optimally tuned for). Detector D2, which contains training data from SAS2, performs slightly less well due to the inclusion of external data. When processing the entire data set (SAS1 and SAS2), detector D2 performs better due to the inclusion of training data from both data sets. However, performance is significantly lower than the detectors running on the single data set (SAS1).

The black and blue curves shows performance results on the SAS1 dataset. Again as expected, detector D1 out-performs D2 since it has been optimally trained for data SAS1. D2, which has been trained using both data sets, still performs well but the inclusion of external data from SAS2 does impact performance, typically by somewhere between 0 and 5% for a given false alarm rate.

The ROC curves in Figure. 3.30 demonstrate the sensitivity of supervised ATR systems to the similarity between the test and training data. Both detectors (D1 and D2) perform significantly worse on the entire data set (SAS1 and SAS2) than they do when processing only imagery from data SAS1. This is likely due to the fact that most of the training data is from data set SAS1, resulting in possible model over-fitting. The outcome of this research is not conclusive; the results clearly show the impact the training data has on ATR performance. The viability of re-training the ATR for a new environment when new data becomes available has not been confirmed and requires

further work.

As a final note, no ground truth for data set SAS2 was available. This dataset contained a lot of complex areas containing mine-like objects. In the absence of ground truth, all these detections are considered as false alarms. An example image from the data set is shown in Figure. 3.31.

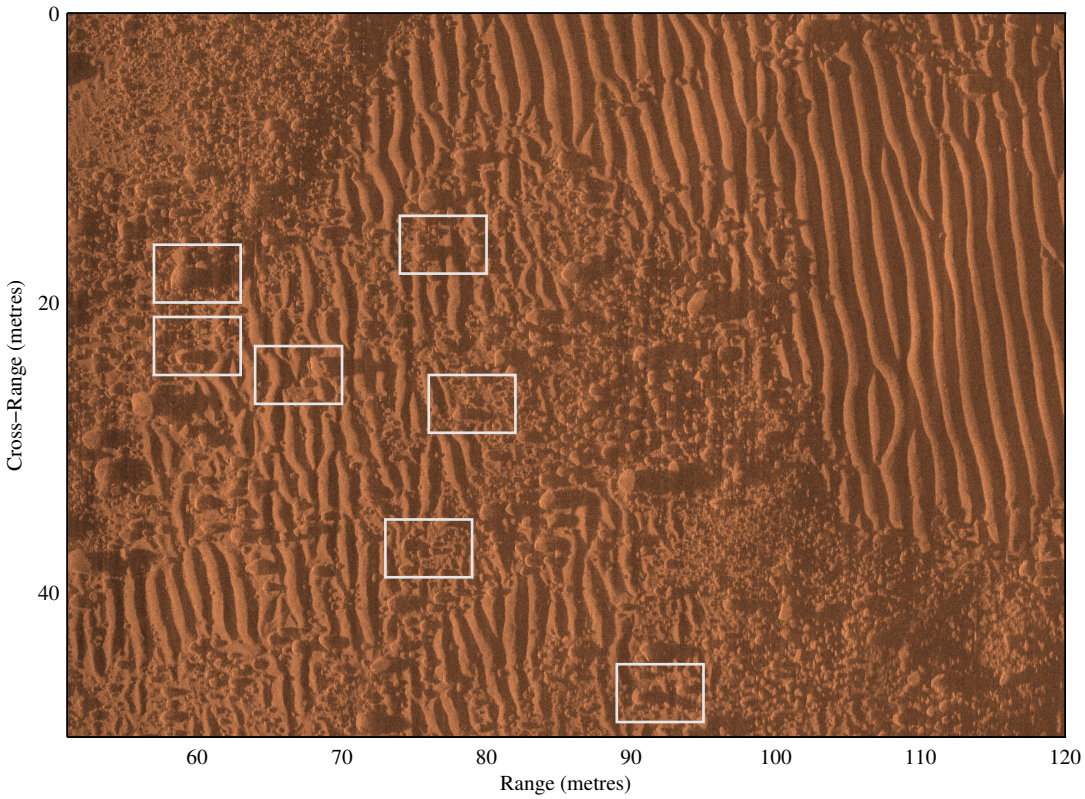

Figure. 3.31 An example image from dataset SAS2 which had no ground truth. Many of the ATR detections appear very mine-like, but they are all considered as false alarms in the analysis.

3.6.4 The Impact of Poor Data on the Performance

Within sonar data in general and the real SAS dataset described in Section 4.5.1 in specific, some target samples do not appear clearly. Some examples of these poorly contrasted targets are shown in Figure. 3.32. This is a regular occurrence in sidescan or SAS data collection; while much of the data will be of a high quality, conditions will inevitably produce target data which is of a lower quality. Inputting this data into the ATR training process may impact the performance of the algorithm.

Due to the low contrast and image quality, forcing the ATR to learn these signatures in the training phase will likely deteriorate ATR performance as it will be much more difficult to discriminate between these targets and the background clutter. To assess the impact of this data, an experiment was run where most of the poor quality training

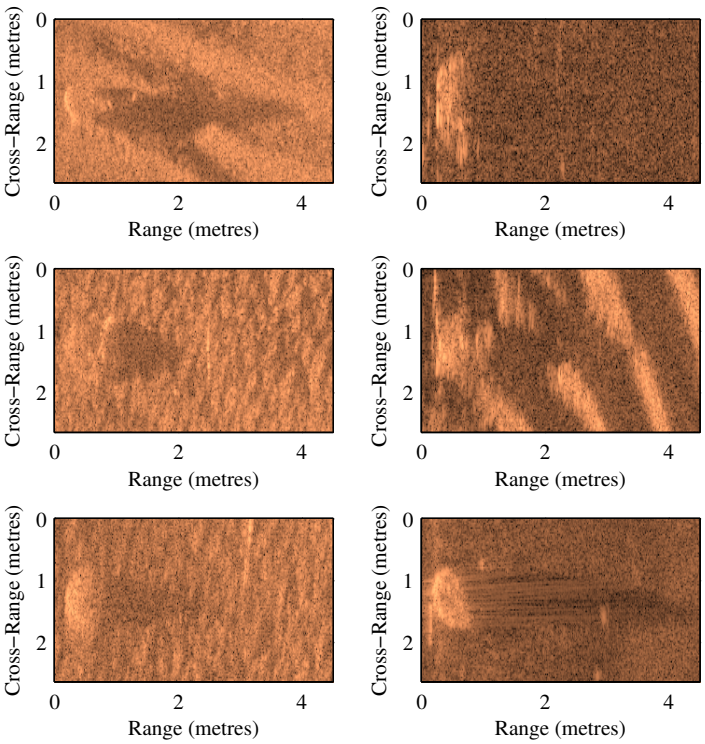

Figure. 3.32 Examples of low contrast objects from the SAS data. Some sonar data will have low quality imagery. Using this data in the training phase of an ATR will have implications for the ATR performance.

samples were removed from the training data. These targets were selected manually.

Figure. 3.33 shows the ROC results of the new detector and compares it with a detector trained using all the training data. The ATR used is the cascade classifier approach proposed in this thesis. The removal of poor data resulted in a detector which required fewer features and fewer cascade stages and consequently less time to train and test. The performance results in Figure. 3.33 do not show an obvious increase in performance when removing the lower quality samples from the training set.

Further evaluation on other data sets would be required to confirm the result shown above. Initial experimentation suggests that the inclusion of poor data in the training samples does not adversely impact ATR performance. This result is based on an experiment where the number of low quality object samples is much lower than the number of high quality samples used.

3.7 Conclusion

This chapter has presented a novel approach for ATR in sonar data. The architecture of this approach was first outlined and justified within the context of sonar data. The use of Haar features was shown effective in encoding object signatures in sonar imagery. AdaBoost was proven capable of selecting features relevant to the object of interest.

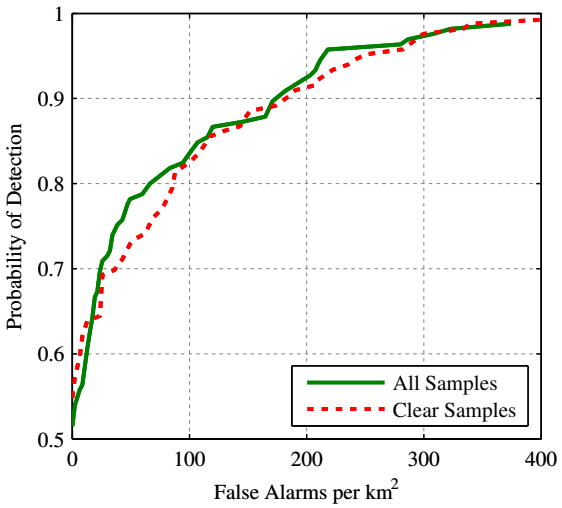

Figure. 3.33 Comparison of ROC curves for two detectors: one trained on all training data (all samples), and one trained on clear data only (clear samples). The real SAS data of truncated cones is used. As the graph shows, the removal of the lower quality samples from the training set has no impact on the detection performance.

The use of a cascade of classifiers rather than an individual classifier was shown to be responsible for the high processing speed of this approach.

Various extensions were proposed to overcome some limitations of the proposed approach. A new confidence measure was proposed and proven more robust than the traditional one. Several optimisations were proposed to speed up the training which reduced the training time from days to minutes. The impact of some internal and external factors on the detection performance of the proposed approach had then become the focus of the research. An extended set of features was shown to improve the discriminative power of the detector. The exploitation of information from previous and subsequent stages of the cascade was discussed and proven very effective when very high detection rate is desired and extra false alarms are not a burden.

An approach which reduces the effect of sand ripples in sonar data was tried, but did not improve the performance. Only modest reduction in the false alarm rate resulted from post-processing the output of the proposed approach using HOG features.

Particular attention was given to the context which may change extremely in sonar data. The inclusion of limited context surrounding the object improved the performance. Simple attempts to adapt the proposed approach to the context resulted in significant increase in the performance.

All the solutions proposed in this chapter were directly validated by experiments. Complete descriptions of the data and the experimental design will be presented in the following chapter along with a direct comparison of the results with two human experts and in-water demonstration on-board two AUVs.

Chapter 4

Results

4.1 Introduction

The Automatic Target Recognition (ATR) approach proposed and extended in the previous chapter is thoroughly evaluated in this chapter. While the previous chapter includes several experiments to support the extensions proposed and the assumptions made, this chapter will include a series of experiments on the system as a whole using various datasets. This includes full descriptions of the datasets employed, the experiments conducted, and the results acquired.

The complexity of the data used in the evaluation process will gradually increase throughout the chapter. This includes synthetic data, Augmented Reality (AR) data, and finally real data. Before starting the evaluation process, Section 4.2 will discuss some general design choices and requirements for all the experiments presented in this chapter.

Due to the scarcity of real sonar data, the proposed ATR will be first evaluated on synthetic data in Section 4.3, where a sidescan sonar simulator is used to generate a large set of sonar data including targets of different shapes.

However, synthetic data may not look realistic due to several environmental and technical factors which do not allow accurate modelling of the natural seafloor. Therefore, a compromise between using real and synthetic data is presented in Section 4.4 where a special tool is used to augment previously collected real sonar data with synthetic targets.

Although the AR data may look more realistic than the synthetic data, the evaluation of any ATR approach will not be realistic until it is done on real data. Therefore, a set of real Synthetic Aperture Sonar (SAS) data is used in Section 4.5 to place the proposed ATR under real scrutiny.

In order to replace human experts, an ATR should at least offer an equivalent performance. The performance of the proposed ATR will therefore be directly compared

with the performance of two human experts in Section 4.6.

Finally, to gain further trust in the proposed ATR, its ability to process data and recognise targets real-time will be demonstrated on-board two Autonomous Underwater Vehicles (AUVs) in real in-water trails in Section 4.7.

4.2 Experimental Design

All the experiments presented in this chapter are based on the ATR approach which has been introduced and extended in the previous chapter. They all follow the general train and test forms outlined in the previous chapter; though differ in detail.

The training goals represent one of the essential settings. In each round of boosting, one Haar feature is selected at a time until the stage training goal of minimum detection rate and maximum false alarm rate is achieved. Stages are added to the cascade until the overall training goal of maximum false alarm rate is achieved. In all our experiments, a global false alarm rate of 10^{-7} is used. The stage training goals are also fixed to 0.998 minimum detection rate and 0.333 maximum false alarm rate in all our experiments which deal with synthetic and AR data. These goals had to be slightly relaxed when dealing with real data due to the limited size and the high variation of the data. Hence, the stage training goals were changed to 0.995 minimum detection rate and 0.5 maximum false alarm rate in all our SAS experiments.

It is important to mention that the ability to achieve very high detection rate was an essential design requirement in all our experiments. This condition has been adopted because of the higher cost of missing a target than having fewer extra false alarms in most applications. Therefore, all train and test parameters were chosen carefully to fulfil this requirement. It is important to note that relaxing this requirement may result in detectors that generate Receiver Operator Characteristics (ROC) curves better at low to moderate detection rates but incapable of high detection rates.

The lack of a large dataset of sonar data is a serious problem for evaluation purposes in general and for supervised algorithms in particular. Sonar data is scarce mainly because data collection operations are time-consuming, resource-intensive, and require human intervention which may risk lives of personnel. The majority of already collected data is also classified and cannot be made public due to its relevance for military applications.

To alleviate the scarcity of sonar data, we use a sonar simulator to generate synthetic data, as will be described in Section 4.3. We also use special tools to infuse existing real background data with synthetic targets, as will be described in Section 4.4. This was motivated by the availability of large amount of sonar data which do not include targets.

Synthetic and AR data are fine for initial evaluation purposes, but no ATR will be

trusted until it is evaluated on real data. Therefore, after the evaluation on synthetic and AR data, the proposed approach will be evaluated on real data from three different sensors: SAS, sidescan sonar, and forward looking sonar. To alleviate the scarcity of real data without the use of anything synthetic, our experiments follow some tricks which will be described here.

To increase the number of target samples, we generate a new version of each target sample by flipping it vertically (cross-range). This is possible because of the special characteristics of sonar imagery. Figure. 4.1 shows a snapshot of a target and its flipped version. The flipped sample clearly represents a new sample of the same target scanned while traversing in the opposite direction. To increase the number of target samples even further we generate samples by shifting each sample by 1 pixel in the four main directions.

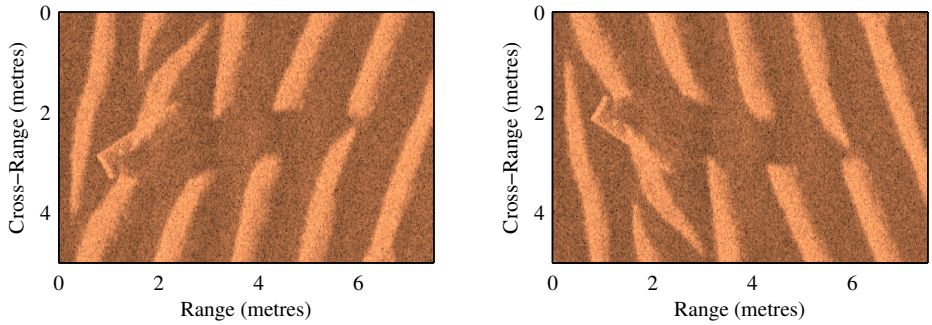

Figure. 4.1 An example of a real SAS target sample (left) and its cross-range flipped version (right). This mechanism doubles the number of target samples used for training.

The ATR algorithm proposed in this thesis learns not only the targets but also the clutter. The clutter in this context includes anything, which can be seen, which does not represent a target. Therefore, a very large set of data which does not include targets (the negative set) is required to represent the non-target (clutter) class. To increase the size of the negative set we do the following two steps:

- Images which include targets are pre-processed to remove the targets, so they can be added to the negative training set. Target regions are replaced by regions from the local context to avoid creating anomalies.

- All images are flipped vertically (cross-range) which doubles the size of the dataset. Again, this is possible because of the special characteristics of sonar imagery. The flipped image clearly represents a new image of the same region scanned while traversing in the opposite direction.

Figure. 4.2 shows an example of a real SAS image including seven targets, its non-target version (the same image after hiding the targets based on the context), and its cross-range flipped version. This mechanism significantly increases the size of the

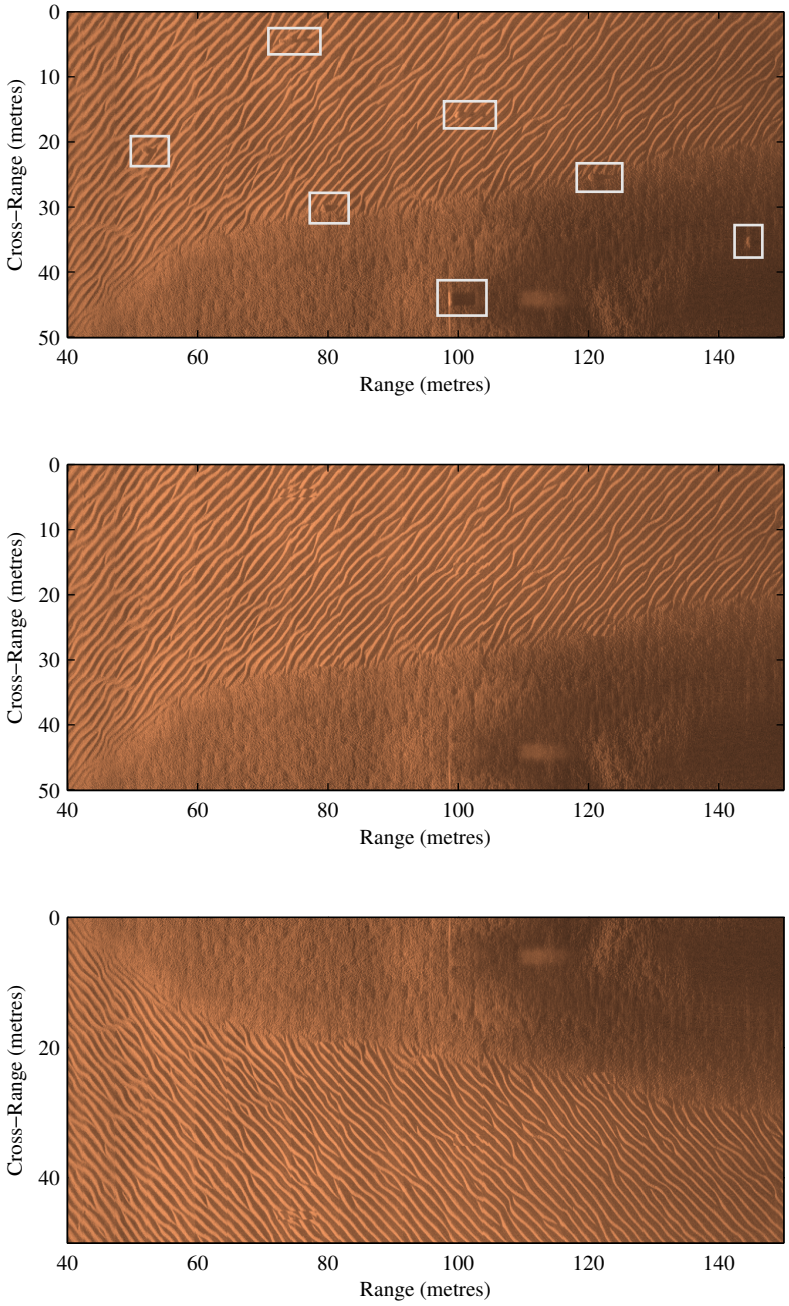

Figure. 4.2 An example of a real SAS image which includes seven targets highlighted by bounding rectangles (top), its non-target version (the same image after hiding the targets based on the context) (middle), and its cross-range flipped version (bottom). This mechanism significantly increases the number of the non-target samples used for training. It allows the use of images which include targets to generate non-target samples after hiding the targets. It also doubles the size of the dataset by flipping each image across the range.

negative set while keeping it representative. This allows the ATR to learn from a larger number of non-target examples and consequently encode better knowledge about the background.

4.3 Experiments on Synthetic Data

4.3.1 Dataset Description

A realistic sidescan simulator presented in [69] has been used to generate the synthetic data used in this section. The simulator is based on two fundamental steps: a 3D terrain generator and the sidescan generator. The 3D terrain generator synthesizes an environment with a variety of seabed types. Fractal texture models are used in this generator to represent the natural environment. From the numeric 3D seafloor, synthetic sidescan is generated according to a trajectory into the 3D environment. The sidescan generator is based on a simple ray-tracing method. Objects of different shapes and different materials can be added into the environment. More details about the simulator can be found in [69].

The simulator computer program was provided by the authors of [69] to generate the data. Based on personal discussions with the sonar experts from the Ocean Systems Laboratory at Heriot-Watt University, the simulator was configured to generate data of specific characteristics suitable for the application of underwater ATR. These characteristics are listed in Table 4.1. Three types of targets were added to the environment. These targets are truncated cones, wedges, and cylinders. The dimensions of these targets are:

- Truncated cone: 100cm lower diameter, 50cm upper diameter, and 50cm height.
- Wedge: 100cm long, 50cm width, and 50cm height.
- Cylinder: 100cm long and 30cm diameter.

Pixel dimensions	15x15 centimetre
AUV/Tow-fish altitude	[3 - 5] metres
Tile range	50 metres
Image size	334x334 pixels $\approx$ 50x50 metres
Sediment type	coarse sand, fine sand, and sandy mud
Seafloor type	flat, clustered, and sand ripples

Table 4.1 Characteristics of the synthetic data.

Figure. 4.4 shows three examples of the images generated by the simulator under the configurations specified in Table 4.1. Four targets of the same type were added to each image as Figure. 4.4 shows. Closer snapshots of these targets are shown in Figure. 4.3.

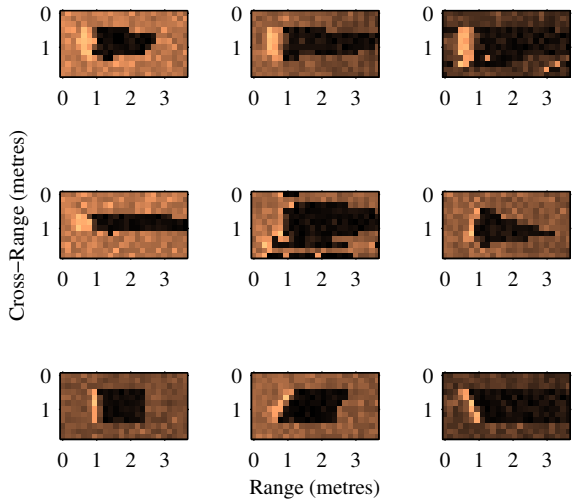

Figure. 4.3 Snapshots of targets from the synthetic data. They are truncated cones, wedges, and cylinders from top to bottom.

4.3.2 Experiments Description

To allow our ATR algorithm to detect each target type mentioned above, a separate detector was trained to detect each type. Three training datasets were generated using the simulator mentioned above, one for each target type. Each dataset consists of 1000 images. Four targets of the same type were added to each image. Table 4.2 lists the characteristics of the resulting detectors.

	Truncated cone	Wedge	Cylinder
Template size (pixels)	24x12	24x12	24x12
Training time (minutes)	3	12	12
Number of stages	6	9	10
Number of features	55	379	375

Table 4.2 Characteristics of the detectors trained on the synthetic data.

4.3.3 Results Analysis

To evaluate the performance of the trained detectors, new data was created using the same simulator mentioned above. The new data consists of the same number of im-

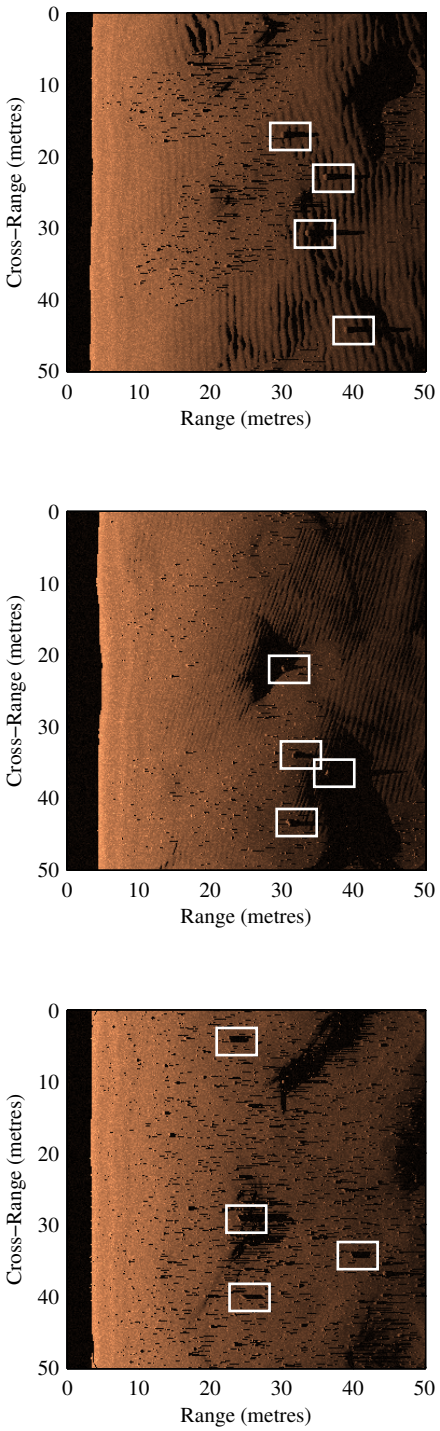

Figure. 4.4 Examples of images from the synthetic data. Each image includes four targets from the same type. They are truncated cones, wedge, and cylinders from top to bottom. The targets are highlighted by bounding rectangles.

ages and targets as the data used for training. Figure. 4.5 shows the ROC curves for the resulting detectors on the test data. The processing time required to run any of these detectors on an image of 334x334 pixels using a 3 GHz Intel Xeon processor is approximately 5 milliseconds.

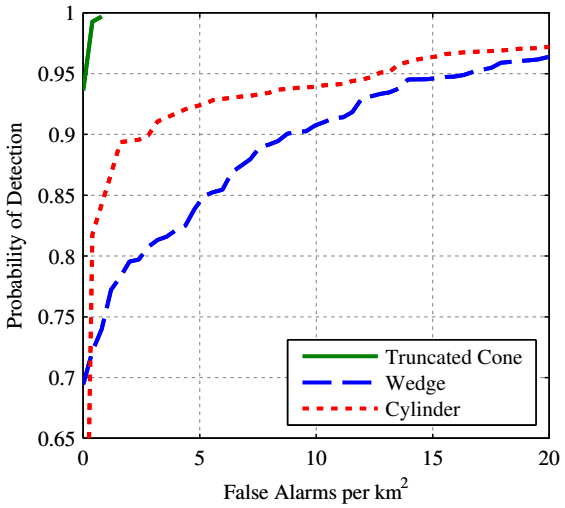

Figure. 4.5 ROC curves on the synthetic data for detectors of different shapes.

The performance of the wedge detector and the cylinder detector is not as good as the truncated cone detector, as Figure. 4.5 shows. This is mainly attributed to the asymmetric and more complex shapes of the wedge and the cylinder in comparison with the truncated cone. The performance of the wedge detector is the lowest. This is perhaps caused by the high similarity between the wedge signatures and the clutter. To show the complexity level of the problem we are dealing with, Figure. 4.6 displays examples from the test dataset. These examples can be categorised as follows:

- Targets which are easy to detect and consequently detected by our approach with high level of confidence (see Figure. 4.6 (a,e,i)).
- Targets which are hard to detect and consequently detected by our approach with low level of confidence. Figure. 4.6 (b,f,j) shows such examples. They are mainly located on complex seafloor which we assume reduces the ATR confidence. Other reasons may include some limitations in the simulator.
- Targets which are very hard to detect even by human experts and consequently missed by our approach. Figure. 4.6 (c,g,k) shows such examples which we assume missed by our approach due to reasons similar to the reasons which made the examples in the previous category hard to detect. See how the target shadow overlaps with the shadow of a natural structure of the seafloor in Figure. 4.6 (c,k).

- Clutter objects which look very similar in shape and size to the learnt target samples and consequently misclassified by our approach (see Figure. 4.6 (d,h,l)). Although we count such examples of target-like clutter as false alarms in all our experiments, we think that they represent points of interest which require further inspection.

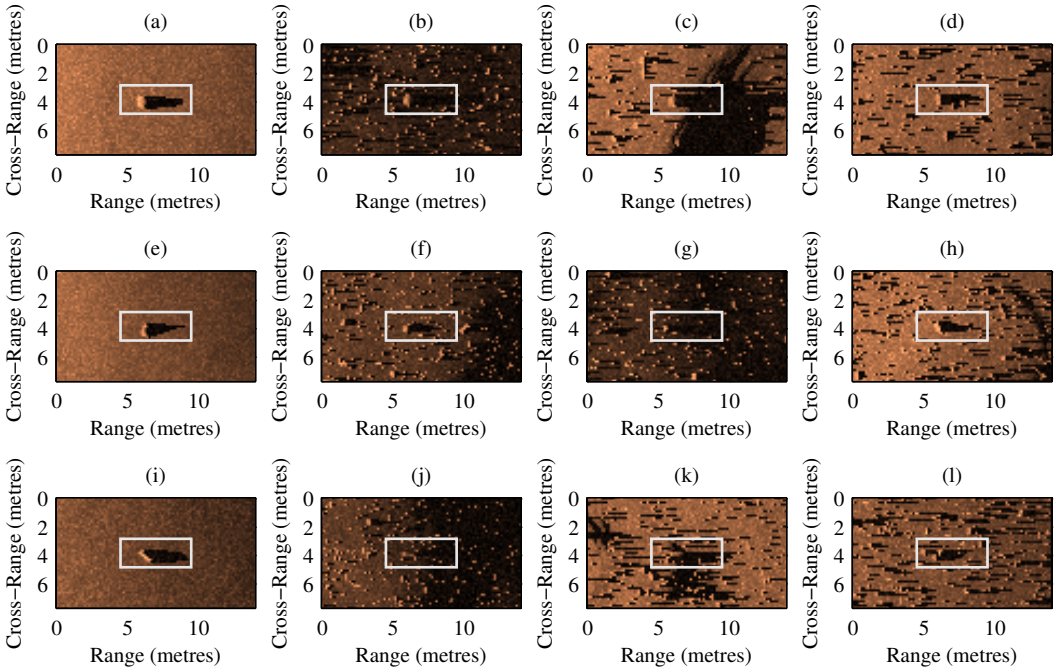

Figure. 4.6 Examples of the test results on the synthetic data. They are generated by the truncated cone detector, the wedge detector, and the cylinder detector from top to bottom. They are easy true positives, hard true positives, false negatives, and false positives from left to right. They are highlighted by bounding rectangles.

4.4 Experiments on Augmented Reality Data

4.4.1 Dataset Description

The natural seafloor is difficult to model accurately due to several factors such as system noise, environmental inhomogeneities, and sensor artefacts in real operational conditions. A compromise between using real and synthetic data can be found in the AR simulation presented in [70], where synthetic target models are embedded on a real image of the seafloor. A computer model of the seafloor is constructed from the sidescan image by an inversion process [71], which determines the parameters that characterise the observed scene. Then the computer model for the seafloor and that of the target are combined and rendered to obtain a new AR image that realistically integrates the synthetic target within the observed scene.

Unlike other simulators which paste a simulated mine on top of an existing image, this approach enables the modelling of the interactions between the topography of the seabed and the target. For instance, if a target is placed behind a 3D structure (respective to sonar), it should not be visible. The length of the projected shadow should also depend on the local elevation of the target.

The construction of the AR database starts by selecting appropriate sidescan sonar images and 3D computer models of target geometry. The sidescan images should be representative of the type or types of seabed to be expected in the test dataset or the final mission, and ideally should have been acquired by the same model of sensor that will acquire the test data. The computer models for the targets should approximate the expected types of targets as accurately as possible. With the real sonar images and the computer models of the targets, numerous simulated AR samples can be generated and stored in the database along with the ground-truth.

A real dataset of sidescan sonar images was provided by the Ocean Systems Laboratory at Heriot-Watt University as the base to add synthetic targets to. This dataset was acquired over the Framura Area (Italy, South of La Spezia) during the BP02 experiment conducted by NURC (NATO Undersea Research Centre). The data was obtained by REMUS 100 AUV using Marine Sonic sonar. The data contains some target-like clutter but no real targets. The seabed texture is varied: the majority of the mission is flat seabed, some is covered by sand ripples and some by complex seafloor (mainly posidonia, complex seafloor is described here as difficult to mine hunt). Table 4.3 lists the main characteristics of the dataset.

Pixel dimensions	0.058 (range) x 0.12 (cross-range) metres
AUV altitude	[2.9 - 3.8] metres
Range	30 metres
Image size	1024 (range) x 1000 (cross-range) pixels $\approx$ 60x120 metres
Number of images	226
Area covered	$\approx$ 1.6 square kilometres

Table 4.3 Characteristics of the real sidescan data used to generate the AR dataset.

The computer program of the AR simulator described above was provided by the authors of [70] to evaluate the ATR approach proposed in this thesis. Three types of targets were added to the environment described above. These targets are truncated cones, wedges, and cylinders. These targets are very similar in shape to the targets used for the synthetic dataset in the previous section, but slightly different in dimensions:

- Truncated cone: 100cm lower diameter, 50cm upper diameter, and 40cm height.

- Wedge: 100cm long, 100cm width, and 30cm height.
- Cylinder: 200cm long and 60cm diameter.

Figure. 4.8 shows a sector of a typical image from the original dataset used to generate the AR dataset. Figure. 4.8 also shows the exact sector after augmenting targets of different shapes to it. Closer snapshots of various targets are shown in Figure. 4.7.

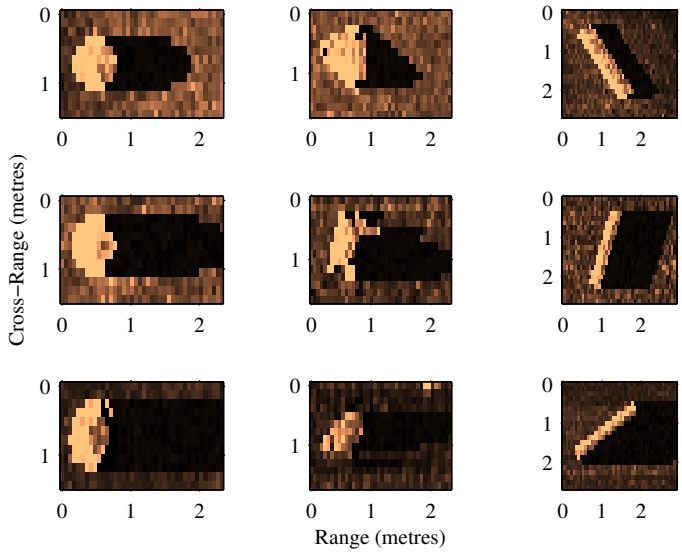

Figure. 4.7 Snapshots of targets from the AR dataset. They are truncated cones, wedges, and cylinders from left to rights.

To allow the evaluation of our ATR algorithm on the AR data, three datasets were generated based on the real sidescan data presented above, one for each target type. The truncated cone dataset contains 1130 (5 in each image) target views, while the wedge and the cylinder datasets contain 2260 (10 in each image) target views each. More wedges and cylinders were added due to their more complex signatures in sonar data in comparison with truncated cones.

4.4.2 Experiments Description

To allow our ATR algorithm to detect each target type mentioned above, a separate detector was trained to detect each target type. Each of the datasets built above was split randomly into two equal subsets; one for training and another for testing. Table 4.4 lists the characteristics of the resulting detectors.

4.4.3 Results Analysis

Figure. 4.9 shows the ROC curves for the resulting detectors on the test datasets. The processing time required to run any of these detectors on an image of 1024x1000 pixels using a 3 GHz Intel Xeon processor is approximately 80 milliseconds.

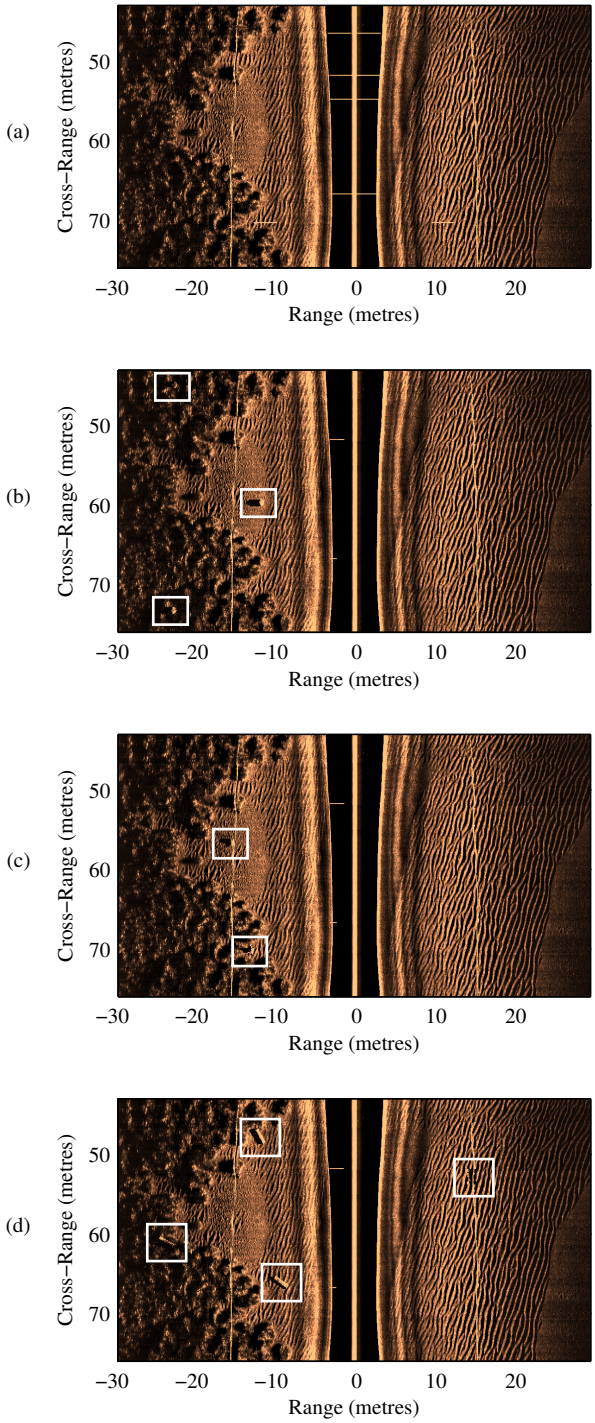

Figure. 4.8 A sector of an image from the original dataset used to generate the AR dataset and how it looks like after augmenting targets to it. (a) the original image. (b) the image augmented with three truncated cones. (c) the image augmented with two wedges. (d) the image augmented with four cylinders. The targets are highlighted by bounding rectangles.

	Truncated cone	Wedge	Cylinder
Template size (pixels)	40x12	40x14	50x22
Training time (minutes)	31	281	1802
Number of stages	11	17	13
Number of features	284	1316	806

Table 4.4 Characteristics of the detectors trained on the synthetic data.

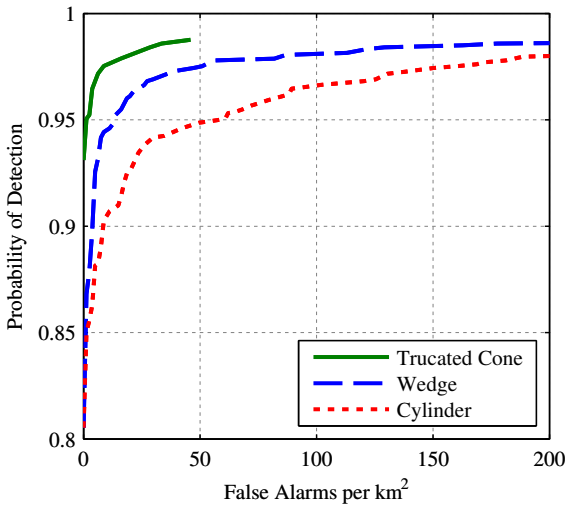

Figure. 4.9 ROC curves on the AR data for detectors of different shapes.

To show the complexity level of the problem we are dealing with, Figure. 4.10 displays examples from the test dataset. These examples can be categorised as follows:

- Targets which are easy to detect and consequently detected by our approach with high level of confidence (Figure. 4.10 (a,e,i)).
- Targets which are hard to detect and consequently detected by our approach with low level of confidence. Figure. 4.10 (b) shows such an example which we assume hard to detect because it is located on a complex seabed. Figure. 4.10 (f) shows another example which we assume hard to detect because of the overlap between the target highlight-shadow pair and the highlight-shadow pair of the sand ripple. Figure. 4.10 (j) shows another example which we assume hard to detect because of the surface return.
- Targets which are very hard to detect even by the human operator and consequently missed by our approach. Figure. 4.10 (c,k) shows such examples which we assume missed by our approach due to reasons similar to the reasons which made the examples in the previous category hard to detect. Figure. 4.10 (g) shows another example which we assume missed by our approach because it is

located at very low range.

- Clutter objects which look similar in shape and size to the learnt target samples and consequently misclassified by our approach (see Figure. 4.10 (d,h,i)). Although we count such examples of target-like clutter as false alarms in all our experiments, we think that they represent points of interest which require further inspection.

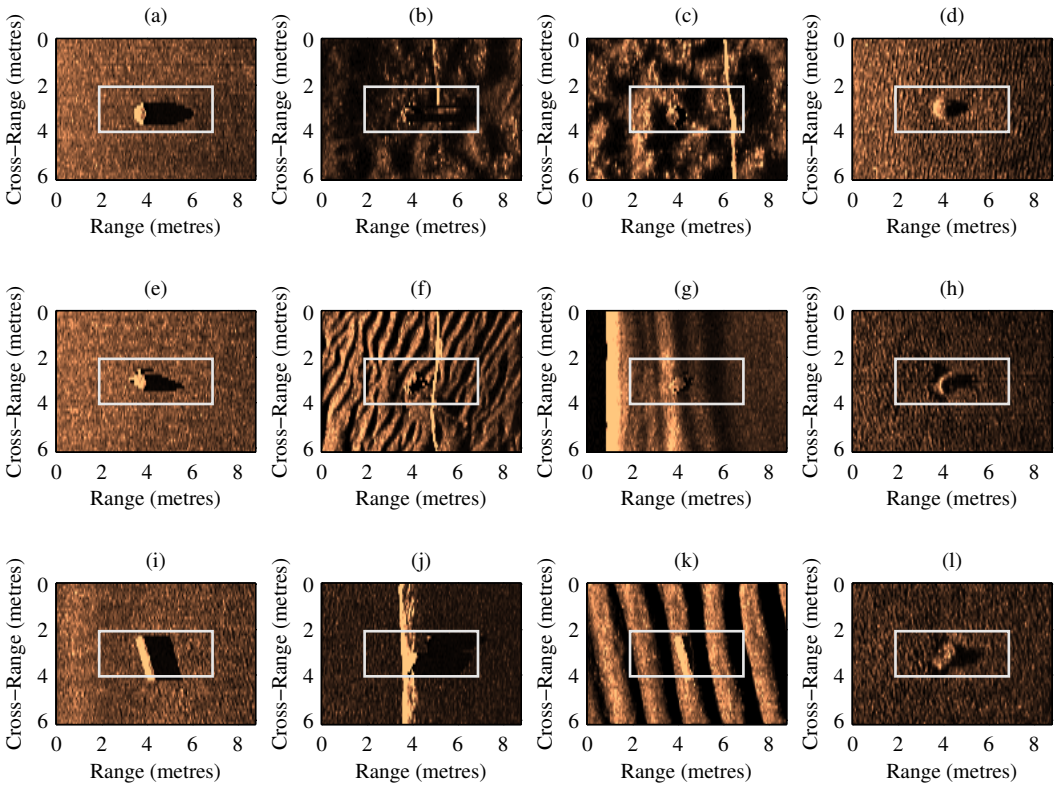

Figure. 4.10 Examples of the test results on the AR data. They are generated by the truncated cone detector, the wedge detector, and the cylinder detector from top to bottom. They are easy true positives, hard true positives, false negatives, and false positives from left to right. They are highlighted by bounding rectangles.

4.5 Experiments on Real Synthetic Aperture Sonar Data

4.5.1 Dataset Description

The real data used in this section is based on SAS technology. It was provided by DSTL (Defence Science and Technology Laboratory) to evaluate the ATR proposed in this thesis. The dataset was collected by NURC (NATO Undersea Research Centre) using a state-of-the-art AUV known as MUSCLE (Mine-hunting Unmanned underwater vehicle for Shallow water Covert Littoral Expeditions). Table 4.5 lists the characteristics of the data.

Pixel dimensions	0.015 (range) x 0.025 (cross-range) metres
AUV altitude	$\approx$ 13 metres
Tile range	[40 - 150] metres
Image size	7333 (range) x 2001 (cross-range) pixels $\approx$ 110x50 metres

Table 4.5 Characteristics of the real SAS data.

A number of mine-like targets were deployed and surveyed from different aspects. These targets are very similar in shape to the targets used in the previous sections, but slightly different in dimensions:

- Truncated cone: 1m lower diameter, 0.5m upper diameter, and 0.5m height.
- Wedge: 1m long, 0.6m width, and 0.3m height.
- Cylinder: 2m long and 0.5m diameter.

The dataset covers 3 separate regions, with similar targets and target distribution patterns but with different seabed characteristics as follows:

- Area B: uncluttered background (flat) (69 images, containing 159 target views).
- Area C: cluttered background (ripples, rocks and weed) (61 images, containing 141 target views).
- Area D: cluttered background (rocks and troweling marks) (71 images, containing 141 target views).

Each area contains 9 targets (3 from each target type) arranged in a similar pattern. Many rocks, boulders, and other clutter objects were also surveyed. Figure. 4.11 shows example images taken from the three separate regions, including views of different targets lying on the seabed. Close snapshots of some targets are shown in Figure. 4.12.

4.5.2 Experiments Description

The dataset has been split randomly into two equal subsets; one for training and another for testing. A cascade detector has been trained for each target type mentioned above (truncated cone, wedge, and cylinder) in addition to a generic detector to detect all types of targets together. Table 4.6 lists the characteristics of the resulting detectors.

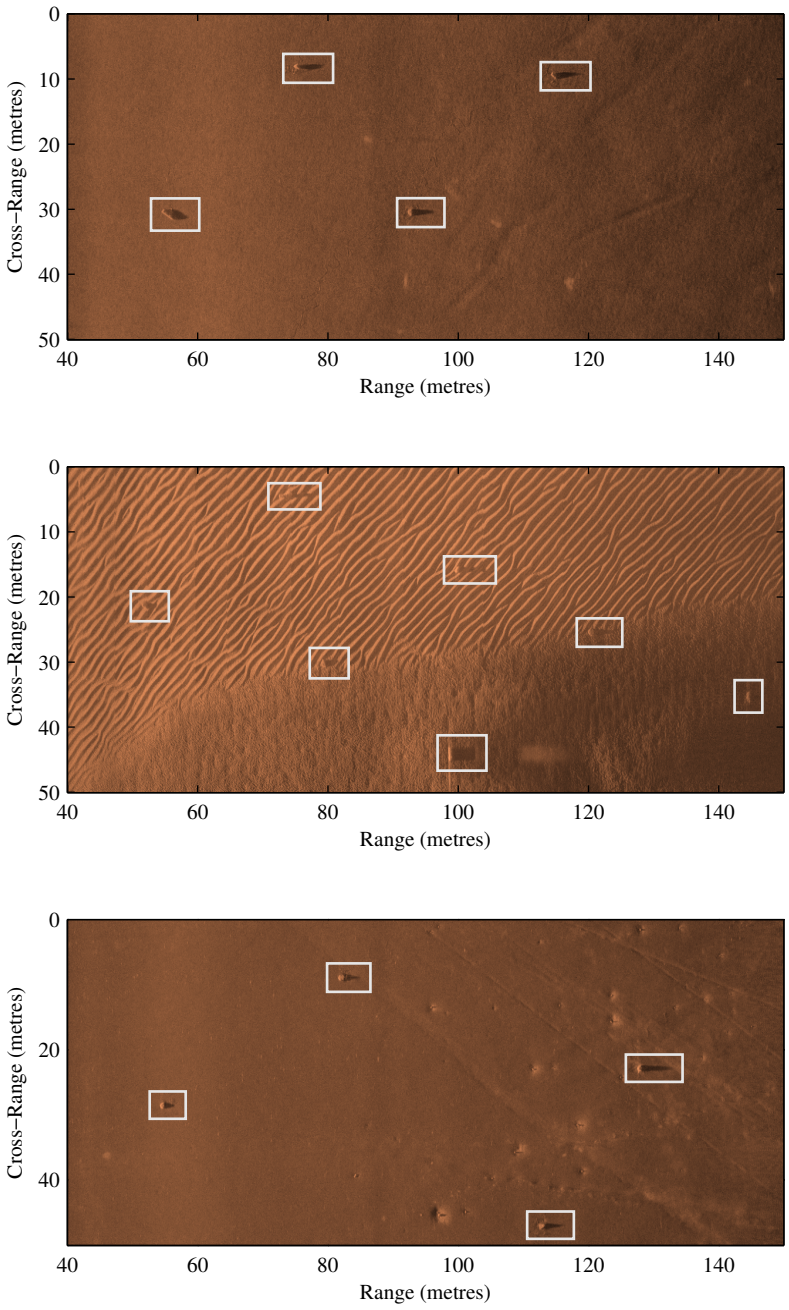

Figure. 4.11 Example images taken from the three separate regions in SAS dataset. They are from areas B, C, and D from top to bottom. They include views of different targets lying on the seabed. Targets are highlighted by bounding rectangles.

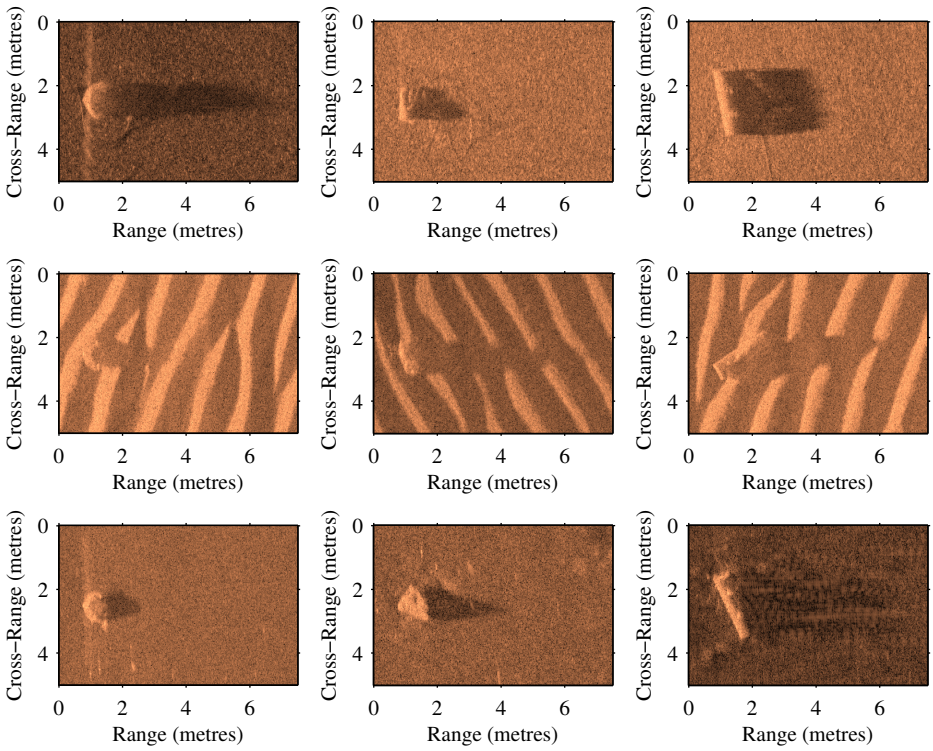

Figure. 4.12 SAS snapshots of real objects on the seafloor. They are truncated cones, wedges, and cylinders from left to right. They are taken from three different areas, B, C, and D from top to bottom.

	Truncated cone	Wedge	Cylinder	Generic
Template size (pixels)	125x55	130x60	300x105	300x105
Training time (minutes)	41	79	78	240
Number of stages	11	18	12	20
Number of features	83	196	83	289

Table 4.6 Characteristics of the detectors trained on SAS data.

4.5.3 Results Analysis

Figure. 4.13 shows the ROC curves for the resulting detectors on the test dataset. The processing time required to run any of these detectors on an image of 7000x2000 pixels using a 3 GHz Intel Xeon processor is approximately 475 milliseconds.

To show the complexity level of the problem we are dealing with, Figure. 4.14 displays examples from the test dataset. These examples can be categorised as follows:

- Targets which are easy to detect and consequently detected by our approach with high level of confidence (Figure. 4.14 (a + e + i)).
- Targets which are hard to detect and consequently detected by our approach with low level of confidence. Figure. 4.14 (b) shows such an example which we

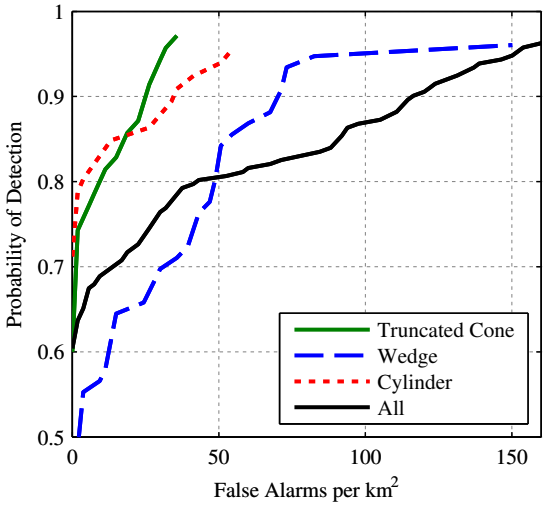

Figure. 4.13 ROC curves on the real SAS data for three detectors of three different shapes in addition to a generic detector (All) which is trained and tested on all shapes together.

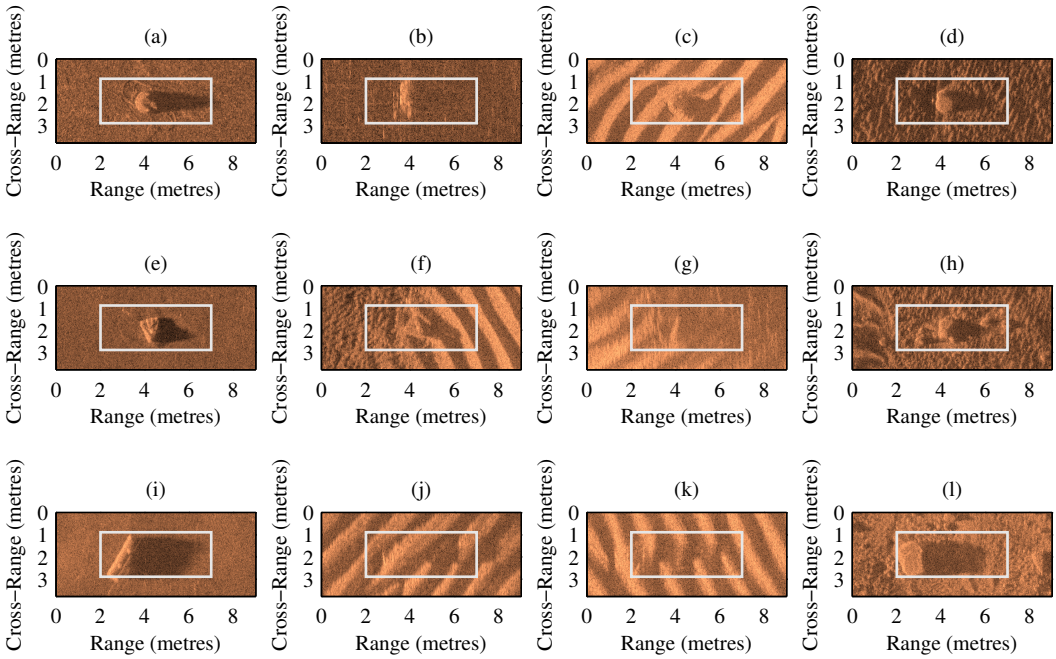

Figure. 4.14 Examples of the test results on SAS data. They are generated by the truncated cone detector, the wedge detector, and the cylinder detector from top to bottom. They are easy true positives, hard true positives, false negatives, and false positives from left to right. They are highlighted by bounding rectangles.

assume hard to detect because it is blurred and almost has no shadow (perhaps because it is located at high range). Figure. 4.14 (f) shows another example which we assume hard to detect because it is located at the border between two types of seabed. Figure. 4.14 (j) shows another example which we assume hard to detect because of the overlap between the highlight-shadow pair of the target and the highlight-shadow pair of a sand ripple.

- Targets which are very hard to detect even by the human operator and consequently missed by our approach. Figure. 4.14 (c,g,k) shows such examples which we assume missed by our approach due to reasons similar to the reasons which made the examples in the previous category hard to detect.

- Clutter objects which look similar in shape and size to the learnt target samples and consequently misclassified by our approach (see Figure. 4.14 (d,h,i)). Although we count such examples of target-like clutter as false alarms in all our experiments, we think that they represent points of interest which require further inspection.

4.6 Comparison with the Human Operators

In this section we compare the proposed approach with real world results. In order to carry out such a comparison we need human experts to look at the same dataset and identify targets. Bristol University conducted such an experiment on the dataset described in the previous section and they provided us with the results. Two human expert operators were shown around half of the dataset (30 images from each of the three separate areas) and asked to identify targets. The results of these operators are shown in Table 4.7.

	Detection rate	Number of false alarms
Operator 1	84% (155 of 184)	86
Operator 2	77% (142 of 184)	39

Table 4.7 Results of the human operators.

In order to make a direct comparison with the operator tests the dataset was split so that the test set is identical to the operator test set. The remaining data (111 images including 257 target views) was used to train a detector of all available shapes.

Figure. 4.15 shows the ROC results of this detector on the operator test set. The data points for operators 1 and 2 are also shown in Figure. 4.15. The result indicates that our approach outperforms the operators. Table 4.8 compares in detail between the operators and our approach. The results indicate that our approach reduces the false alarm rate by 39% in comparison with the first operator and by 25% in comparison with the second operator.

This comparison with the human experts did not only allow a real world evaluation of our approach, but also allowed a comparison with two recent ATR approaches presented in [72] and [14]. These approaches were both applied to the same dataset considered here and compared with the exact human expert results presented above.

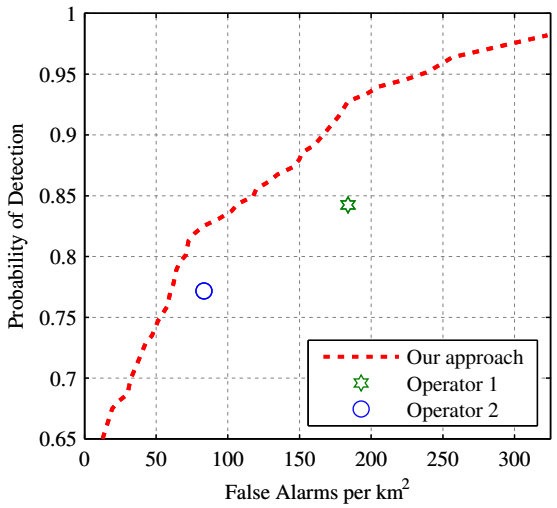

Figure. 4.15 Performance comparison between the proposed approach and two human operators. The detector is trained and tested identical to the operators. As the graph shows, a performance that is comparable or better than the operator may be obtained.

Detection rate	Operator false alarm rate (km2)	Our approach false alarm rate (km2)	Reduction in false alarm rate
84%	184	113	39%
77%	83	62	25%

Table 4.8 Performance comparison between the proposed approach and two human operators.

Hill et al. [72] used log-Gabor, matched and shaped filters together with a two-class Support Vector Machine (SVM) classifier. The work of Nelson and Kingsbury [14] includes two stages. First, sand ripple suppression together with a matched filter is used to discard any areas of the image which can be easily distinguished from the targets. Second, lacunarity-based features are extracted from the false alarms of the first stage and used to train a one-class SVM classifier (i.e. the targets need not to be learned or known, they are detected as anomalies).

In comparison with the results reported by Hill et al. [72] and Nelson and Kingsbury [14], our approach appears to outperform both of their approaches (under the caveat that the experimental set-ups might be a little different). For instance, in order to detect 95% of the targets, Nelson and Kingsbury [14] incur 337 false alarms and Hill et al. [72] incur 310 false alarms while our approach incur 112 false alarms. In a two-class SVM version of Nelson and Kingsbury [14] work, 95% of targets were recovered at the cost of 201 false alarms which appears to be better result than Hill et al. [72], but still outperformed by our approach.

4.7 Field Trials

In order to perform a real evaluation of the ATR approach proposed in this thesis, I joined a team of researchers from the Ocean Systems Laboratory at Heriot-Watt University in a project to demonstrate a fully autonomous solution for subsea survey, target detection, and intervention operations. This was applied to a typical Mine Countermeasures (MCM) scenario, which involves the survey of an area, the detection of potential targets, and their final neutralisation. This scenario could be applied to many other applications in the marine science, offshore, and archaeology domains.

This solution is based on multiple AUVs collaborating to achieve different objectives. It uses a high level of semantic interaction with the operator for describing the mission goals and domain. A non-expert operator is able to specify high level goals (such as search, identify, and neutralise) and the vehicles can jointly plan and execute these goals. Once the mission goals have been specified, a knowledge based framework finds the match between the high level goals requirements and the capabilities of the vehicles. A planner is then used to plan and coordinate the actions of the vehicles. The planner uses embedded knowledge on each platform to activate the capability required by a specific action. This planner can adapt the actions of the platforms based on online sensor data analysis or information exchange with another vehicle.

The three key modules of the proposed system are: a distributed knowledge representation of the environment, an effective autonomous decision making, and an ATR algorithm. The focus here will only be on the ATR algorithm. More details about other modules can be found in [73, 74].

The ATR algorithm is critical to the performance of the system as it provides an essential link between data and information. In our case the algorithm analyses sonar data and tries to identify potential targets of interest in the data. The algorithm has to be sufficiently accurate in the sense that all targets have to be detected and that the false alarm rate must be sufficiently low to enable meaningful replanning. Moreover, the algorithm must be computationally efficient, making real-time on-board operation on low-power hardware possible. These requirements match the aims of this thesis. Therefore, the ATR approach proposed in this thesis was chosen for this system.

Integrating the ATR results into the planner allows multi-views re-acquisition. An octagonal spiral pattern is introduced for the re-acquisition maximising the number and the variety of views and minimising the mission time. This shows that the ATR is intrinsically linked to the operational side of the MCM mission.

The performance of the system was evaluated on the Ocean Systems Laboratory platforms (see Figure. 4.16) in a set of integrated in-water field trial demonstration days at Loch Earn, Scotland (56°23.1′ N,4°12.0′ W) (see Figure. 4.17). Initially, the plan was to use the three platforms, but only two (Remus 100 and Nessie 5) were

eventually used due to hardware problems with one vehicle.

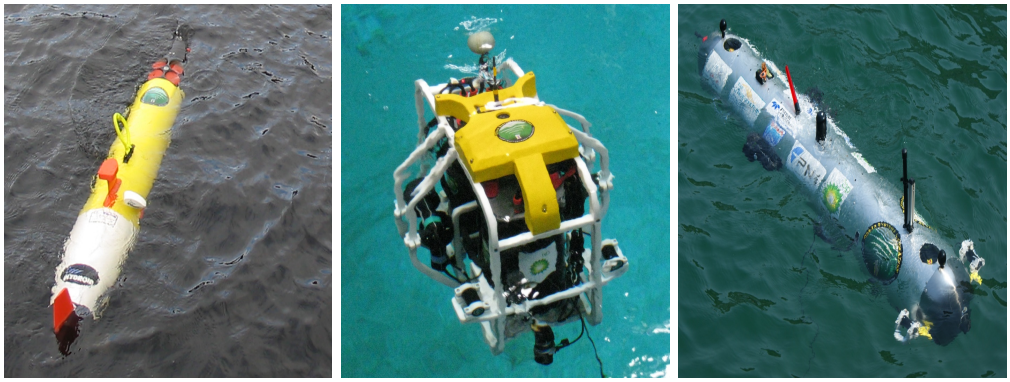

Figure. 4.16 The Ocean Systems Laboratory AUV platforms. They are REMUS 100, Nessie 4, and Nessie 5 from left to right.

Figure. 4.17 Loch Earn and operational area (highlighted red square) where the mission was performed. The mission included surveying the area, finding targets (unknown number and location) and identifying them.

An area of approximately 300m by 300m was defined for inspection. The area contains clutter on the seabed and several man-made objects, such as sunken rowing boats, anchors and other features such as sand ripples, submerged tree branches, and rocks of different sizes. A target of truncated cone shape was deployed in the area (see Figure. 4.18). It has a similar shape and dimensions as well known types of classical sea mines.

A high level goal was assigned to the system: survey, detect, and identify this area. The final stage (intervention) was not included. The mission was deemed complete when every possible target in the area had been identified and mapped accurately. The various phases were assigned to the two available platforms based on their capabilities:

- REMUS 100 AUV was assigned as a platform type A for detection.
- Nessie 5 AUV was a platform type B for identification.

The two vehicles started their respective missions at the same time. REMUS first performed an initial survey of the zone as REMUS is the only vehicle that has the capability of performing this service. The survey was not preprogrammed by the operator but calculated online by the adaptive mission planning module. The ATR module

Figure. 4.18 Dummy target of truncated cone shape deployed in the area of operations.

was running live and the vehicle was constantly identifying potential mine like objects. Once the survey was complete, the REMUS vehicle performed a closer inspection of the potential targets using an octagonal spiral motion around the target to maximise the number of viewpoints at different angles and ranges on the targets. Compared to the classical daisy pattern, the number of views is maximised (from 12 possible views for the daisy pattern to 24 for the octagonal pattern). The time for reacquisition is also reduced drastically (from 30 min. to 6 min.). The potential targets were then ranked by order of priority and Nessie 5 was used to perform the identification of the most likely target. It is important to note that Nessie 5 could have performed part of the secondary inspection of the targets and indeed this was done in simulation. However, it was not done during the final trials due to time constraints.

Figure. 4.19 shows the results of a collaborative mission. The spiral reacquisition pattern enables many ATR hits on the target which quickly disambiguates the false alarms from the real target; as demonstrated in Figure. 4.20. The number of false alarms is in line with classical operator's performances and our own evaluation of the mission data.

The ATR approach proposed in this thesis was integrated within the software of two AUVs: REMUS for ATR in sidescan sonar (the detection task), and Nessie 5 for ATR in forward looking sonar (the identification task). The remaining part of this section is concerned with describing the data used and analysing the results of the ATR. It is divided into two folds: one for sidescan sonar data and another for forward looking sonar data.

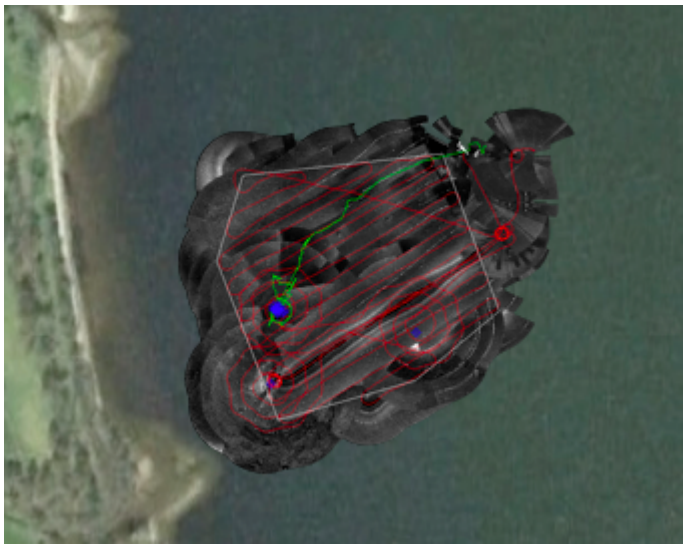

Figure. 4.19 The results from an example mission. Mission track with geo-referenced sidescan mosaic over-imposed. The area identified by the operator is described by the white polygon. The REMUS 100 AUV track is red and the Nessie 5 track is green. The hits from the ATR module are shown in blue.

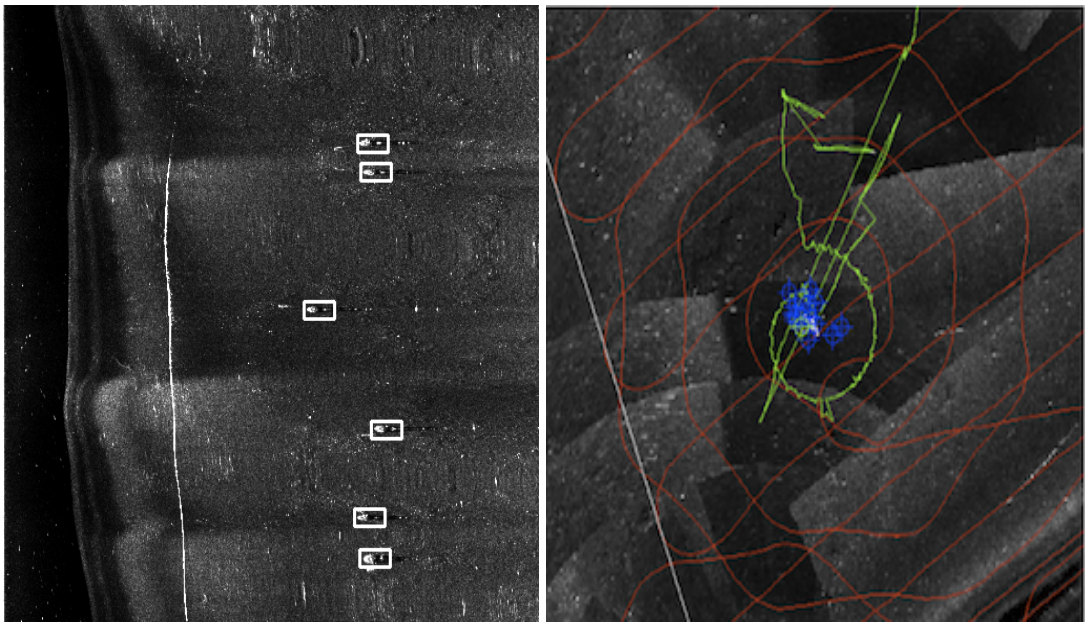

Figure. 4.20 Series of target detections of the embedded ATR over sidescan data (left) while performing the octagon pattern with the REMUS 100 over the truncated cone target (right). The maximum error in the localisation of the target was 8m.

4.7.1 Sidescan Sonar Experiments

Dataset Description

Sidescan sonar data was collected by REMUS 100 AUV using the Marine Sonic sensor. Table 4.9 lists the main characteristics of the data. Figure. 4.21 shows example images taken from separate missions, including views of the truncated cone target lying on the seabed. Close snapshots of the target are shown in Figure. 4.22.

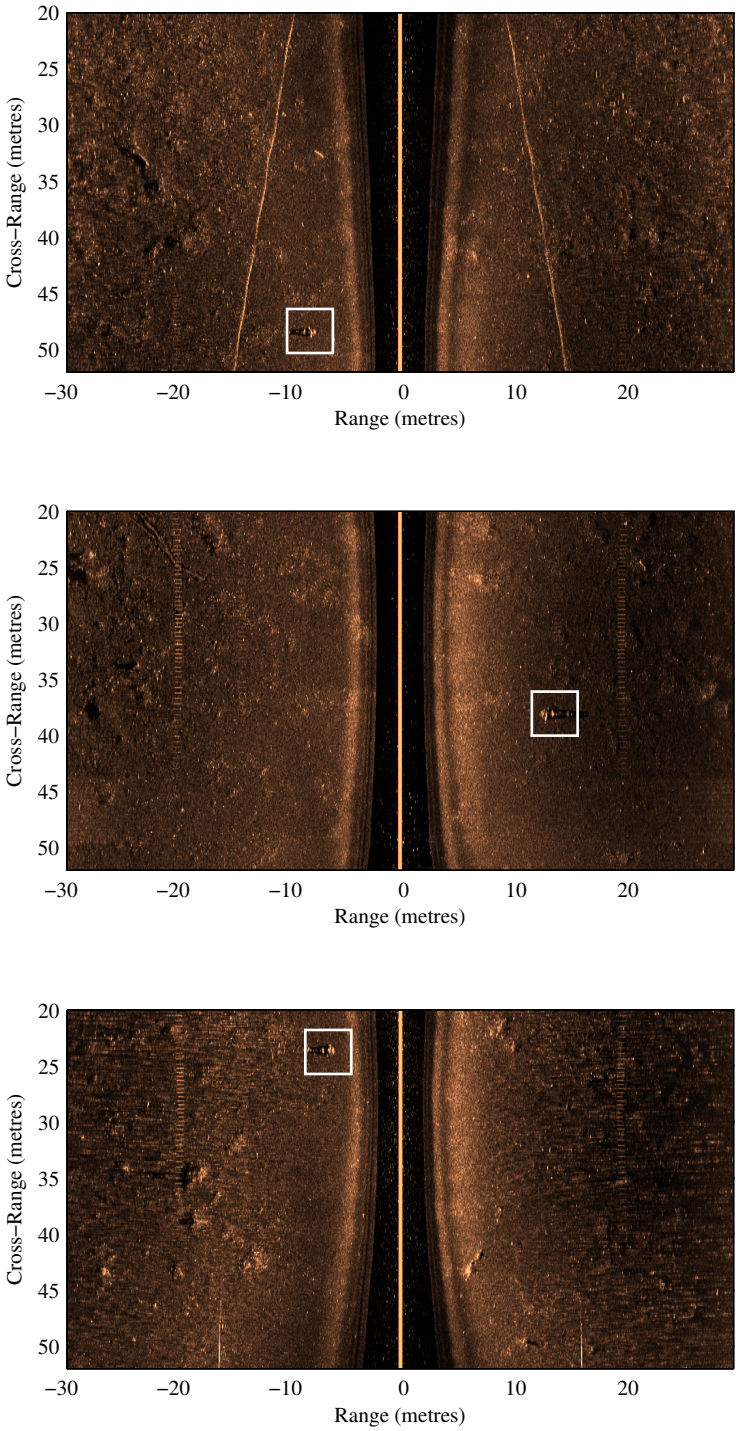

Figure. 4.21 Example images from the sidescan sonar data taken from separate missions. Each image includes a view of the truncated cone target. Targets are highlighted by bounding rectangles.

Pixel dimensions	0.058 (range) x 0.12 (cross-range) metres
AUV altitude	$\approx$ 4 metres
Range	30 metres

Table 4.9 Characteristics of the sidescan sonar data collected by REMUS 100 using Marine Sonic sensor.

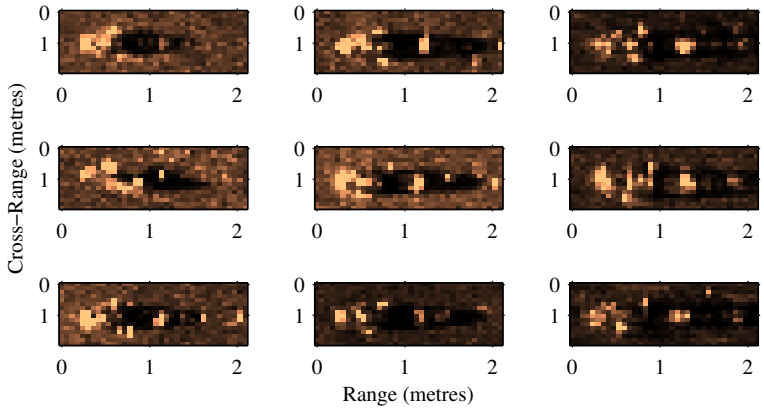

Figure. 4.22 Sample snippets of truncated cone target from sidescan data.

Experiments Description

The data for training was gathered during several test missions where 44 views of the target and a large amount of seabed data were recorded. Table 4.10 lists the characteristics of the resulting detector.

	Truncated cone detector
Template size (pixels)	36x16
Training time (minutes)	8
Number of stages	7
Number of features	92

Table 4.10 Characteristics of the truncated cone detector trained on the field trials sidescan sonar data.

Results Analysis

After the detector was trained, it was tested real-time in three complete missions. The target was subsequently moved to another area for a more realistic evaluation and further three missions were conducted. The ATR results are listed in Table 4.11. The processing time required to run this detector on an image of 1024x500 pixels using a 3 GHz Intel Xeon processor is approximately 24 milliseconds.

Mission	Target location	Survey target detections	Survey false alarms	Inspection target detections	Inspection false alarms
1	A	6 (of 6)	3	8 (of 12)	1
2	A	5 (of 5)	2	15 (of 19)	0
3	A	5 (of 5)	2	0 (of 0)	0
4	B	3 (of 4)	1	11 (of 16	0
5	B	5 (of 5)	2	1 (of 3)	0
6	B	4 (of 6)	2	18 (of 25)	1

Table 4.11 Experimental results on sidescan sonar data of the field trials.

Table 4.11 does not only show the results of the typical survey, but also the results during the inspection. The detection rate during the spiral inspection is low in comparison with the detection rate during the usual lawn mower survey. This is because of the distortion in the target view during the spiral movement (the detector was not trained on such samples). The spiral inspection was supposed to be done using a camera sonar rather than the sidescan sonar which is not very suitable for this task. However, we have to do it this way due to time constrains and some faults in the inspection vehicle.

To show the complexity level of the problem we are dealing with, Figure. 4.23 displays examples from the test missions. These examples can be categorised as follows:

- A target which is easy to detect and consequently detected by our approach with high level of confidence (Figure. 4.23 (a)).
- A target which is hard to detect and consequently detected by our approach with low level of confidence. Figure. 4.23 (b) shows such an example which we assume hard to detect because it was scanned while the vehicle was turning.
- A target which is very hard to detect and consequently missed by our approach. Figure. 4.23 (c) shows such an example which we assume missed by our approach because of a defect in the sensor which distorted the highlight of the target.
- A clutter object misclassified by our approach. Figure. 4.23 (d) shows such an example. However, this example is not very similar to a typical target sample and consequently our approach associates it with low confidence value which makes it easy to reject by simple thresholding.

In conclusion, our ATR approach successfully found the real target in all test missions with very low number of false alarms (2 per mission on average). Given the

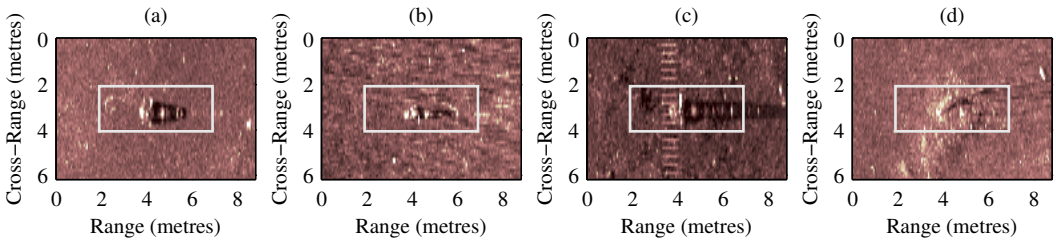

Figure. 4.23 Examples of the test results on the sidescan sonar data from the field trials. They are generated by the truncated cone detector. They are easy true positive, hard true positive, false negative, and false positive from left to right. They are highlighted by bounding rectangles.

fact that the target was detected at least three times in each mission, we ran a simple multi-view analysis which cleared all the false alarms.

4.7.2 Forward Looking Sonar Experiments

Dataset Description

Forward looking sonar data was collected by our hover capable vehicle, Nessie 5, using Tritech Gemini 720i sonar sensor. Table 4.12 lists the main characteristics of the data. Figure. 4.24 shows example images taken from two separate missions, including views of the truncated cone target lying on the seabed. The examples are displayed in the original polar coordinate system along with the Cartesian coordinate system for clarity. Close snapshots of the target in the polar coordinate system are shown in Figure. 4.25.

Pixel dimensions	0.008 metres (range) x 0.5 °(cross-range)
Field of view	120 °
Vertical beam-width	20 °
Range	15 metres
Frame rate	$\approx$ 1 Hz

Table 4.12 Characteristics of the forward looking sonar data collected by Nessie 5 using Tritech Gemini 720i sonar sensor.

Experiments Description

Due to hardware faults we were not able to collect as much forward looking sonar data as sidescan sonar data. We have managed to run two missions where Nessie 5 was able to hover around the truncated cone target in position B and collect forward looking sonar data. The target appears in 153 frames out of a total of 494 frames in

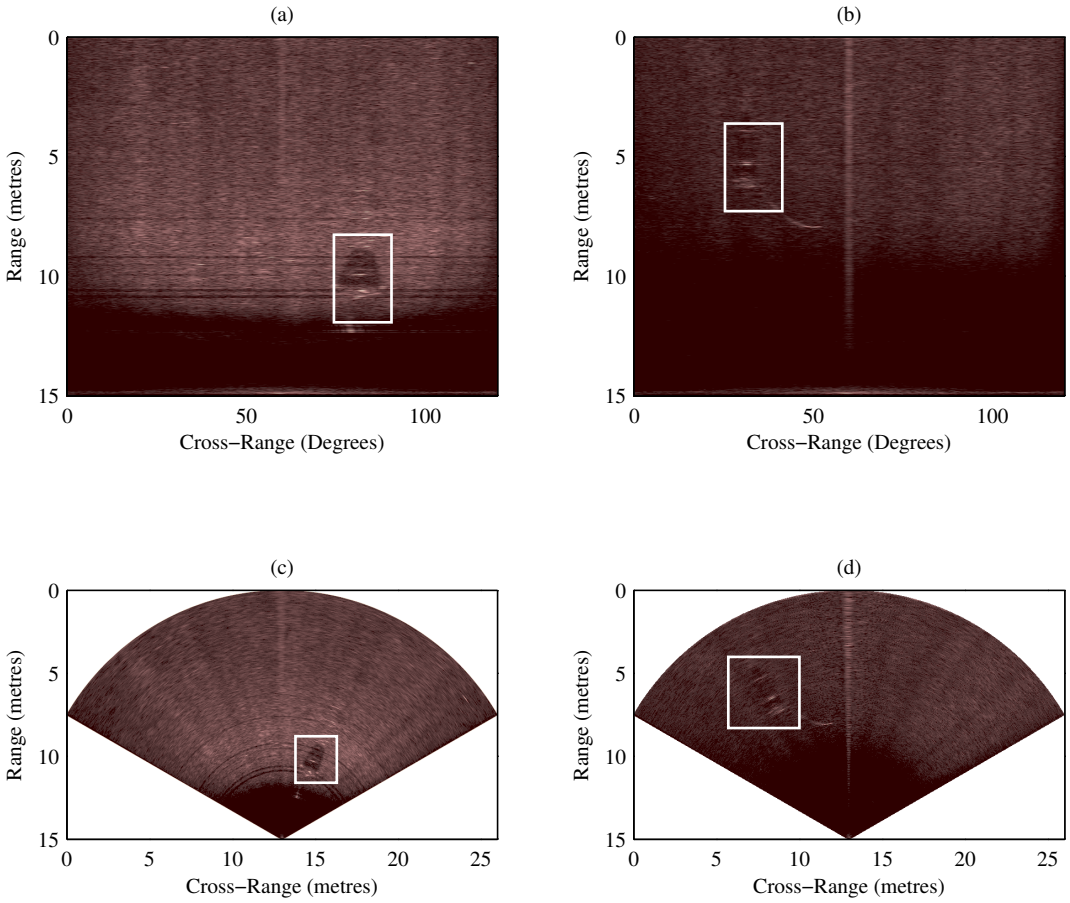

Figure. 4.24 Two examples from the forward looking sonar data. They are taken from two separate missions. (a) and (b) display the examples in the original polar coordinate system. (c) and (d) display the same examples in the Cartesian coordinate system for clarity. Each image includes a view of the truncated cone target. Targets are highlighted by bounding rectangles.

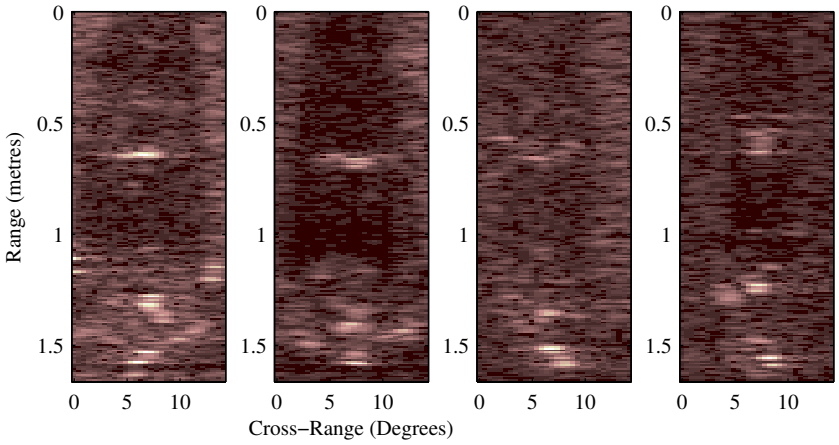

Figure. 4.25 Sample snippets of the truncated cone target from forward looking data. They are taken from the original polar coordinate data.

the first mission and in 91 frames out of 492 frames in the second mission. The first mission was used to train the ATR, while the second mission was used to evaluate its performance. Table 4.13 lists the characteristics of the resulting detector.

	Truncated cone detector
Template size (pixels)	30x200
Training time (minutes)	33
Number of stages	7
Number of features	62

Table 4.13 Characteristics of the truncated cone detector trained on the field trials forward looking sonar data.

Results Analysis

After the detector was trained on the data collected in the first mission, it was tested real-time in the second mission. The detection rate was 65% with 1 false alarm every 20 frames. This means in this mission, on average, we detected the object in two of every three scans that sensed the area in which it was located. The processing time required to run this detector on an image of 256x1809 pixels using a 3 GHz Intel Xeon processor is approximately 16 milliseconds.

To show the complexity level of the problem we are dealing with, Figure. 4.26 displays examples from the test mission. These examples can be categorised as follows:

- A target which is easy to detect and consequently detected by our approach with high level of confidence (Figure. 4.26 (a)).
- A target which is hard to detect and consequently detected by our approach with low level of confidence. Figure. 4.26 (b) shows such an example which we assume hard to detect because it is located at very low range which distorts its highlight.
- A target which is very hard to detect and consequently missed by our approach. Figure. 4.26 (c) shows such an example which we assume missed by our approach because of a defect in the sensor which distorted the view of the target.
- A clutter object misclassified by our approach. Figure. 4.26 (d) shows such an example. However, this example is not very similar to a typical target sample and consequently our approach associates it with low confidence value which makes it easy to reject by simple thresholding.

In conclusion, our ATR approach was successful in detecting the real target in the majority of the frames in the test mission with very low number of false alarms (1 every 20 frames). The majority of the missed views of the target are caused by defects in the sensor and its software driver. These defects also caused false positives in some occasions.

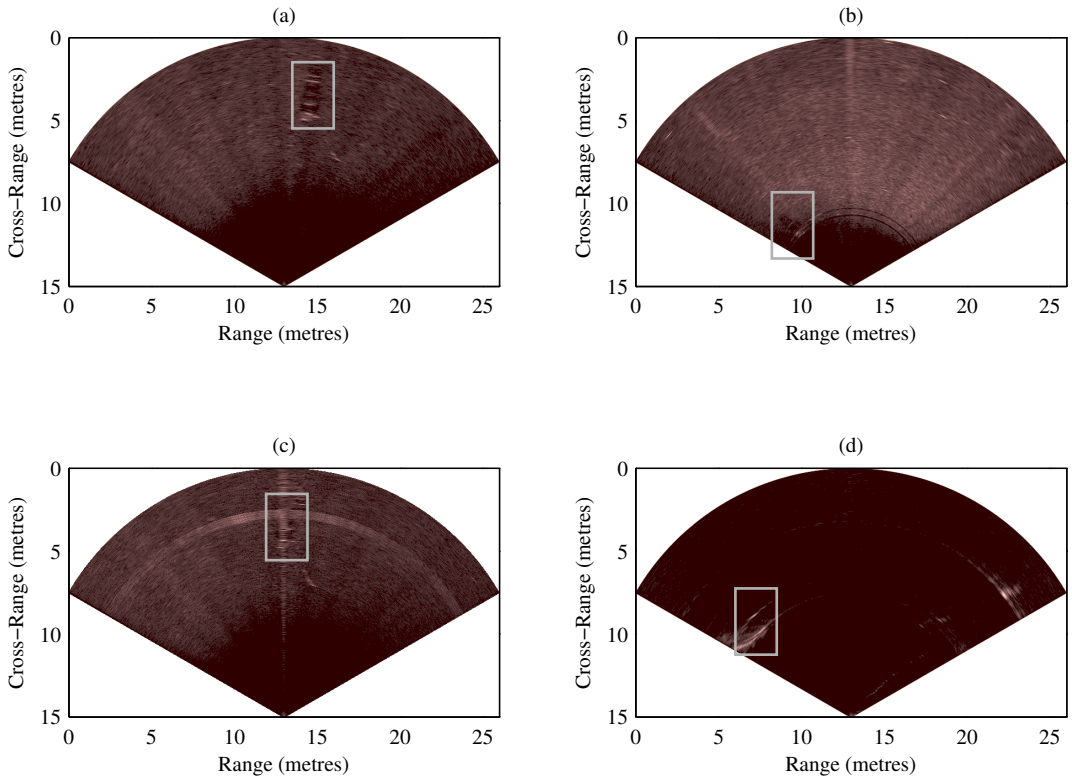

Figure. 4.26 Examples of the test results on the forward looking sonar data from the field trials. They are generated by the truncated cone detector. They are: (a) easy true positive, (b) hard true positive, (c) false negative, and (d) false positive from left to right. They are highlighted by bounding rectangles

Given the fact that the forward looking sensor generates multiple frames per seconds and the target appears in many frames when the vehicle hovers around the target, we recommend post-processing the results of the ATR by techniques such as tracking and multiple view analysis. We assume such techniques to remove the majority of false alarms and increase the detection accuracy.

4.8 Conclusion

A thorough evaluation of our ATR approach has been conducted in this chapter. The proposed approach was initially tested on two large synthetic datasets. The first dataset was completely synthetic where both the seafloor and the objects were generated by simulation. The second dataset was partially synthetic where only objects are generated by a simulator and then augmented to real sonar data. Good results were obtained on both datasets which encountered more realistic evaluation. Therefore, the proposed approach was then tested on a set of real SAS data. Results were also good which proves the effectiveness of this approach on real data of higher variations and smaller quantity to learn from in comparison with synthetic data. These results also prove the

suitability of this method for SAS data of high resolution and blurred nature.

Results showed problems relating to complex seafloors (rocks and sand ripples). These regions exhibit multiple scenarios which resemble the man-made objects of interest. Indeed, it is often hard to classify these regions even by a human expert. To gain human trust in our approach, we directly compared our approach with two human experts. Results showed that our approach outperforms human operators.

Finally, the proposed approach was tested in a real world scenario of in-water field trials. The algorithm was implemented and integrated within two AUVs. It was capable of processing data real-time and successfully recognizing objects in sidescan sonar and forward looking sonar.

Taken together the results show a detection accuracy which is consistently high. This provides evidence that our approach is robust to both sonar conditions and sensor modality. One interpretation could be that these variations may be automatically accounted for by the supervised nature of the proposed approach. Furthermore, the object representation using Haar features, the voting scheme of Adaptive Boosting (AdaBoost), and the focusing capability of the cascade structure all contribute to the success of this approach.

Chapter 5

Conclusion

This study was set out to explore the problem of automating the interpretation of underwater environment. This is crucial to allow subsea robots to become completely autonomous, so time and money can be saved and humans can be kept out of harm's way. Autonomous robots are needed for a variety of applications, such as unexploded ordnance clearance, oil pipeline inspection, and ship/plane wreck hunting.

Two attributes of the processing are crucial if automated interpretation is to be successful. First, the interpretation should run real-time on-board Autonomous Underwater Vehicles (AUVs). Second, the interpretation should perform equal to or better than a human operator. Approaches in the open literature do not appear to have these attributes. These attributes were therefore made the objectives of this thesis.

This thesis has proposed a novel approach capable of analysing sonar data extremely rapidly while achieving high detection accuracy. In field trials, this approach was shown to process data real-time on-board AUVs. In direct comparison with human experts, this approach offered comparable detection performance while reducing the false alarm rate by around 40%. Therefore, the stated aims of this thesis were successfully achieved.

This chapter concludes this thesis. First, the research outcomes and the contributions are summarised. Second, the limitations are discussed and some areas are recommended for further research.

5.1 Research Outcomes and Contributions

This thesis presented a feature-based supervised approach to the problem of Automatic Target Recognition (ATR) in sonar imagery. The proposed approach builds upon an approach from the computer vision community originally developed for face detection [1]. To date no research on applying this approach to sonar data appears to have been done. It uses Haar features for object representation, Adaptive Boosting (AdaBoost)

for classification, and the cascade for classifier fusion. This approach deviates from prior work of sonar ATR in the following main points:

1. The proposed approach learns directly and automatically from the data, independently from any navigational information or assumptions such as whether the mine is tethered or lies on the seafloor. This makes the proposed approach immune to any errors in such information.

2. The proposed approach is capable of learning large amounts of previously collected clutter information (background data). This significantly reduces the number false alarms.

3. The proposed approach offers a complete solution to the problem of ATR which performs all the traditional tasks of existing ATR solutions at the same time: detection (regions of interest), classification (object/non-object), and identification (object type). This makes the proposed approach very easy to use.

This approach was studied extensively in this thesis. First, the approach was described within the context of sonar data. Haar features were shown to be effective to encode the highlight-shadow signature of objects in sonar imagery. A variant of AdaBoost, called Gentle AdaBoost, was chosen to overcome the problem of outliers in sonar data. The ability of the cascade to achieve real-time processing and learn the large amount of clutter information was also discussed.

After this approach was fully described, several extensions were proposed to overcome some of its limitations. A novel confidence measure, which fuses the confidence measures of the cascade stages, was shown to be more effective than the number of co-located detections. The time consuming nature of the training phase was then carefully investigated. Several steps were optimised to speed up the training phase which resulted in a significant reduction in the training time (minutes versus days).

After the training time was made reasonable to test other aspects of the system, the focus of the thesis turned to improving the detection performance. The inclusion of a narrow frame from the context surrounding the object was showed to improve the performance. The primitive nature of Haar features was argued and an extended set of features was proposed which appears to improve the discriminative power of the classification and enable very high detection rates.

The rejection procedure of the cascade, which rejects a sample when it first fails a stage, was then argued. Results showed that the rejection procedure of the cascade at any stage can be improved by exploiting information from previous or/and successive stages in the cascade. This allowed very high detection rates that were not possible before. However, the gain in the detection rate was accompanied with a higher number of false alarms. Therefore, this extension to the rejection procedure is particularly

useful for applications which strictly require very high detection rate and can tolerate some extra false alarms.

After several approach-specific attempts to improve the detection performance, the focus was turned to deal with the proposed approach as a black box and attempt to improve its performance by processing its input or output. A recent technique to reduce the effect of sand ripples, where most ATR approaches break down, was tried with the ATR proposed in this thesis. Based on the limited evaluation carried out, the detection performance of our ATR did not appear to improve after the ripple suppression. One interpretation could be that the target signatures get distorted while trying to suppress sand ripples.

In an attempt to reduce the number of false alarms produced by the proposed approach, we tried a detector based on the Histogram of Oriented Gradients (HOG) features as a possible post-processing step. Limited experimentation suggested that this approach can only offer little reduction in the false alarms.

Analogous to supervised approaches, the performance of the proposed approach is dependent on the correlation between the training and test samples. Therefore, we looked at the potential benefits of re-training and the impact this has on the performance when new data becomes available. The sensitivity of the proposed approach to the similarity between train and test data was demonstrated. The viability of re-training when new data becomes available has not been confirmed and requires further work.

Finally, the proposed approach was thoroughly evaluated on various datasets. Good results were obtained on synthetic data and Augmented Reality (AR) data which encountered more realistic evaluation on real Synthetic Aperture Sonar (SAS) data. As expected, the performance dropped slightly when dealing with real data. This is perhaps due to the limited amount of data available for evaluation and the high variability of this data. However, in comparison with best published results on real SAS data, the proposed approach performs well. This proves its suitability for SAS data of high resolution and blurred nature.

Results in general showed problems relating to complex seafloors (i.e. rocks and sand ripples). These regions sometimes include natural clutter which resembles the man-made objects. Such clutter objects are hard to distinguish from man-made objects even by human experts. This has been proven by directly comparing the proposed approach with two human experts. Results suggest that the proposed approach outperforms the human experts in discriminating between clutter and man-made objects.

To gain trust in the proposed approach it was integrated within two AUVs and tested in real in-water missions which proved the proposed approach capable of processing data real-time and successfully recognizing objects in sidescan sonar and forward looking sonar.

Taken together the results show a detection accuracy which is consistently high with a reasonable number of false alarms. This provides evidence that the proposed approach is robust to various sensor conditions and modalities. However, it is not ideal and it has limitations. While some of these limitations have already been tackled in this thesis, some are still open for future research which will be discussed in the following section.

5.2 Current Limitations and Future Recommendations

The best obtained overall performance for the proposed approach is beyond what is required for most real world applications in terms of processing speed and detection accuracy. However, the false alarm rate may still be unacceptable for critical applications in complex environments.

Care was taken during this study to demonstrate when the proposed approach did not operate well. This occurred in very complex regions including rocks and sand ripples. This is mainly attributed to target-like clutter in such regions which results in many false alarms. Moreover, the views of real targets in complex regions are often distorted by the clutter. Therefore, it is argued whether such regions should be automatically avoided using seafloor classification techniques.

The false alarms in general (in complex and non-complex regions) were often clutter objects which are classified as targets due to their visual and dimensional similarity to the considered classes. It is also argued whether these are true false alarms as they fulfil almost all the necessary criteria for a real object and they are not obvious clutter even to the human eye.

In scenarios like this, further analysis may be required either by the intervention of a human operator or by the use of another sensor (e.g. Dual Frequency Identification Sonar (DIDSON), wide-band sonar, video). However, such solutions are typically more complex, expensive, and time-consuming. Therefore, the question becomes what can be improved with the data at hand. This indicates that more research is needed to answer this complex but important question. Exploring the following areas as future research strategies is recommended to facilitate the attainment of this goal.

5.2.1 Complex Features

The proposed approach appears to have some limitations due to the use of Haar features. These features are quite primitive in comparison with other features. However, the computational efficiency of Haar features often compensates for their limited discriminative power. In this research, we found that Haar features are good in distinguishing objects from the clutter but to a certain extent; they break down when the

clutter is visually very similar to the considered objects. In an attempt to enrich the discriminative power of Haar features, an extended set was proposed in this thesis (in Section 3.5.2), but the improvement was modest. One interpretation could be that the extended set is still primitive. Therefore, the use of more complex features (e.g. Local Binary Patterns (LBP) [37] and Scale-Invariant Feature Transform (SIFT) [38]) is recommended for future research.

Complex features could either replace Haar features or complement them where the feature selection procedure decides which to use. However, complex features are typically computer intensive and may slow down the ATR if they are used all over the cascade. Therefore, it is recommended that complex features are only used in the final few stages of the cascade where the classification problem becomes harder.

The majority of the image is discarded in the early stages of the cascade and a relatively low number of regions remain for further processing in the final stages. This fact justifies the use of computationally expensive features in the final stages of the cascade where evaluating such complex features is not a burden anymore.

Moreover, this research showed several techniques to allow the proposed approach to achieve very high detection rate (almost 100%) at the cost of some extra false alarms. Therefore, if those adjustments are adopted before complex features take over, it will eliminate the risk of missing objects in the early stages of the cascade before they are analysed by the complex features in the final stages.

5.2.2 Asymmetric Learning

Symmetry here means the equal treatment of all classes in the learning process. While each stage of the cascade is required to achieve an asymmetric goal (i.e. a very high detection rate and a moderate false alarm rate); the underlying learning algorithm (AdaBoost in this thesis) is designed to achieve a symmetric goal (i.e. low classification error). AdaBoost treats target samples and clutter samples equally. This means that the cost of missing a target is not any different form the cost of a false alarm. To trade-off between the detection rate and false alarm rate, AdaBoost threshold is adjusted as we saw in Section 3.2.3. Even though this solution appears to work well, it may not be optimal because it is independent from the learning process. This problem has already been investigated in a number of publications. Viola and Jones in [75] propose to increase the weights on positive samples in every round of AdaBoost such that the error criterion biases towards having very high detection rate. Wu et al. [76] decouple the feature selection step from the classification step. They first use AdaBoost to select features and then use a form of linear discriminant analysis to achieve the stage training goals. Further research is required on this topic.

5.2.3 Outlier-Tolerant Learning

Datasets are not unlikely to include outliers. Outliers here may refer to noisy samples, mislabelled samples, or most importantly naturally ambiguous samples. Examples of outliers from the SAS dataset used in this thesis have been discussed in Section 2.2.2, where target-like clutter and clutter-like targets have been shown in Figure. 2.3. One of the problems with classification methods in general and boosting methods in particular is their sensitivity to outliers [45]. Therefore, a variant of AdaBoost called Gentle AdaBoost [39], known to be less sensitive to outliers, has been chosen in this thesis.

However, there exist a number of alternative solutions to this problem. For example, some variants of AdaBoost have been specifically designed to improve AdaBoost tolerance to outliers such as RobustBoost [47] and TangentBoost [48]. While such methods handle outliers implicitly, other solutions offer explicitly ways to handle the problem. For instance, Kobetski and Sullivan [77] argue and show that pruning the training dataset by excluding hard-to-learn examples can improve the performance of outlier-sensitive algorithms. They propose a two-round learning method, where the outliers are identified in the first round based on their classification score. A second round of training is then performed, where the outliers are either excluded or subjected to a much softer loss function. Kobetski and Sullivan show experimentally that their extension improves the performance of several boosting algorithms. Applying such extensions to the ATR approach proposed in this thesis is recommended for future research.

5.2.4 Parameter Optimisation

The process by which the cascade classifier is trained requires some care. The overall training process of the cascade is supposed to build the optimum classifier in terms of processing speed and detection accuracy. In other words, the training algorithm is supposed to find an optimum set of the following parameters:

- The number of stages.
- The number of features in each stage.
- The threshold of each stage.
- The threshold of each feature.

This set should minimize the computational cost and maximises the performance. This optimisation is a tremendously difficult problem. In practice, some assumptions are made to simplify the training algorithm. In [1], Viola and Jones proposed a simple solution for this problem which seems to produce effective cascade classifiers. They assume the user to specify the following parameters:

- the maximum overall false alarm rate of the cascade.
- the minimum detection rate of a stage (fixed for all stages).
- the maximum false alarm rate of a stage (fixed for all stages).

Features are added to each stage until the stage training goal (the minimum detection rate and the maximum false alarm rate) is met. Stages are added to the cascade until the cascade training goal (the maximum overall false alarm rate) is met. Although this solution seems to produce efficient cascade classifiers in practice, it is not optimal as pointed out in [1]. It is also unclear how to choose these parameters. This issue was addressed in the literature. Various approaches were used to determine the training goals either during the training such as the work in [78] or after the training such as the work in [79]. Further research is required to tackle this issue.

5.2.5 Multi-View Analysis

Currently the ATR approach proposed in this thesis considers each view of an object individually and does not consider the possibility that views in different images could correspond to the same object. This information is typically available in sidescan data due to the lawn-mower nature of the surveys. This information could be exploited in future research to improve the accuracy of the proposed ATR.

The detection probability reported in this research, which describes the number of individual views detected, actually equates to a higher probability of the actual objects detected. Alternatively, some false alarms, reported in this research, could be removed based on their absence of multiple views.

The idea of combining several views of an object in sonar data to increase the classification accuracy has already been addressed in the literature. For instance, fusing the outputs of a single-view classifier from multiple views has been investigated using approaches such as Dempster-Shafer in [24] and [80]. The construction of an extended feature vector based on the features extracted from multiple views before applying the classification method has also been examined in recent work such as [81] and [80]. Applying such fusion techniques to the ATR approach proposed in this thesis is recommended for future research.

5.2.6 Range-Based Analysis

Traditional ATR algorithms in sonar imagery depend mainly on the target shadow for detection and classification. The assumption made usually is that the information relative to the target is mostly contained in its shadow. However, the acoustic shadow

which is cast by an object on the seabed varies in length at different ranges and altitudes. Target samples in our examples have all been cropped to a fixed size. This means that the shadow is not completely included in all target samples. This problem could be alleviated using multiple detectors; each of which covers a limited range. This observation has also been made in previous works such as the work in [54] where three matched filters were used depending upon the cross-range of the data. This problem could also be tackled by adjusting the template in rage to compensate for the lengthening of the shadow in slant range such as the work in [82]. Further research is recommended in this area; where the features used in this thesis may be made invariant to the grazing angle, allowing more accurate detection across the full sonar range.

5.2.7 Context Adaptation

The changing context represents one of the fundamental reasons for the poor performance of ATR algorithms in general. This issue becomes more serious when the test (operational) context is not represented within the training (historic) context. This is common in underwater applications in general and military applications in specific. Therefore, we assume that ATR algorithms should be able to adapt to the context. In this thesis, we proposed some basic solutions to this problem. A gain in the performance was achieved by including a frame from the region surrounding the object to the training as shown in Section 3.5.1. Further gain in the performance was achieved by adapting the classification confidence to the context as shown in Section 3.5.4. Context adaptation is strongly recommended for future research.

5.2.8 Deep Learning

Recently, new methods referred to as Deep Learning have started to emerge within the machine learning community [32]. They are attracting much attention both from the academic and industrial communities. This is mainly due to their empirical success in several traditional artificial intelligence applications such as computer vision, outperforming alternative machine learning techniques in a number of official international pattern recognition competitions [83].

Deep learning methods are typically artificial neural network-like models with many layers of neurons, inspired by the the architecture of the mammalian neocortex. While the theoretical foundations of deep learning have existed for many years [84, 85], only until recently deep learning started to make great success. This could be mainly attributed to the great advances in the computational resources including the graphical processing units (GPUs). The availability of large amount of data also contributes to the recent success of deep learning methods.

One of the key advantages of deep learning is that it does not require a problem-specific hand-engineered feature extraction step. It learns automatically and directly from the data and discovers multiple levels of distributed representations, with higher levels representing more abstract concepts. Another key advantage of deep learning is the ability to learn from unlabelled data, making advantage of the large amount of unlabelled data which is very expensive to label.

Deep learning methods do not appear to have yet been applied to the problem of ATR in sonar. The applicability of these recently introduced methods for the purpose of ATR in sonar needs to be investigated. This is recommended for future research.

5.2.9 3-D Sonar Imagery

Backscatter amplitudes are typically the acoustic measurements collected by most sidescan sonar devices, and translated into imagery. This imagery may have misleading or incomplete information. For example, a bright region in a sonar image could be bright because it is facing the sonar, because it is extremely rough, or because of the saturation of the radiometric values [2]. Bathymetric measurements collected by recent interferometric sonar devices, at the same resolution and in the same locations as the imagery, provide powerful aid to sonar data interpretation. Sonar imagery can be combined with the bathymetry which offers a 3-D representation of the seafloor. Such representation, if available, should be exploited in future ATR solutions for more reliable results.

Furthermore, given the fact that the approach proposed in this research is a feature-based supervised approach, it is worth fusing the results with a template-based unsupervised approach as both approaches will be independent from each other. Finally, the proposed approach could also be extended for use in other pattern recognition applications such as Radar and medical imagery.

Appendix A

Publications and Other Scientific Activities

A.1 Publications

This section lists the publications which have arisen from the work presented in this thesis.

- J. Sawas, Y. Petillot, and Y. Pailhas, "Cascade of boosted classifiers for rapid detection of underwater objects," in *Proceedings of the European Conference on Underwater Acoustics (ECUA)*, Istanbul, Turkey, 2010.

  This is the first paper to introduce the use of the cascade of boosted classifiers in sonar imagery. It proved the applicability of this approach in this field using synthetic sidescan sonar data.

- Y. Petillot, Y. Pailhas, J. Sawas, N. Valeyrie, and J. Bell, "Target recognition in synthetic aperture and high resolution side-scan sonar," in *Proceedings of the European Conference on Underwater Acoustics (ECUA)*, Istanbul, Turkey, 2010.

  This paper briefly reviewed traditional Automatic Target Recognition (ATR) techniques in sonar data and presented new possible solutions to the problem of detection and classification in Synthetic Aperture Sonar (SAS) and high resolution sidescan sonar.

- Y. Petillot, Y. Pailhas, J. Sawas, N. Valeyrie, and J. Bell, "Target recognition in synthetic aperture sonar and high resolution side scan sonar using AUVs," in *Proceedings of the International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, Lerici, Italy, 2010.

  This paper also gave a brief review of the state-of-the-art of ATR in sonar data and presented the results of our initial experiments on real SAS data.

- J. Aulinas, A. Fazlollahi, J. Salvi, X. Llado, Y. R. Petillot, J. Sawas, and R. Garcia, "Robust automatic landmark detection for underwater SLAM using sidescan sonar imaging," in *Proceedings of the International Conference on Mobile Robots and Competitions (Robotica)*, Lisbon, Portugal, 2011.

  This paper introduced a novel solution to the problem of simultaneous localization and mapping (SLAM) by automatically detecting landmarks using the approach proposed in this thesis. This work has been conducted in collaboration with the computer vision and robotics group at the University of Girona in Spain.

- Y. Pailhas, P. Patron, J. Cartwright, F. Maurelli, Y. Petillot, J. Sawas, and N. Valeyrie, "Fully integrated multi-vehicles mine countermeasure missions," in *Proceedings of the International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results (UAM)*, Kos, Greece, 2011.

  This paper analysed the benefits and requirements of heterogeneous fleets of Autonomous Underwater Vehicles (AUVs) for enabling fully integrated Mine Countermeasures (MCM) operations. It also presented the results of a fully integrated MCM mission, a demonstration showing all the multidisciplinary capabilities fully integrated in a truly autonomous distributed sensing-decision-act loop. The ATR approach proposed in this thesis was successfully integrated into two AUVs for real-time object detection in sidescan sonar and forward looking sonar data.

- F. Maurelli, P. Patron, J. Cartwright, J. Sawas, Y. Petillot, and D. Lane, "Integrated MCM missions using heterogeneous fleets of AUVs," in *Proceedings of the IEEE Oceans Conference*, Yeosu, Korea, 2012.

  This paper also presented results from the in-water trials explained in the previous paper.

- J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," in *Proceedings of the European Conference on Underwater Acoustics (ECUA)*, Edinburgh, Scotland, 2012.

  This paper contributed two novel extensions to the ATR approach presented in our earlier publications. These are mainly a new technique to measure the confidence in the classification of the cascade, and a new cascade structure which allows higher detection rates. This paper also demonstrated the effective use of the proposed approach on a large set of real SAS data.

- J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," in *Proceedings of the Inter-*

national Conference on Underwater Remote Sensing (ICoURS), Brest, France, 2012.

This paper also presented novel extensions to our ATR approach similar to the previous paper.

- J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," *Proceedings of Meetings on Acoustics*, vol. 17, 2013.

  This is a journal version of the previous conference paper.

- S. Reed, Y. Petillot, J. Nelson, P. Hill, J. Sawas *et al.*, "Assessment of automatic target recognition and fusion processes for cluttered environments," SeeByte, Tech. Rep., 2013.

  This is an internal report which presented some work conducted in this thesis in collaboration with academic and industrial partners as part of a DSTL (Defence Science and Technology Laboratory) programme which looked at the state-of-the-art ATR techniques before investigating technology gaps that offered the potential for a step increase in performance.

- I. Quidu, V. Myers, and B. Zerr, *Proceedings of the 2012 international conference on detection and classification of underwater targets*. Cambridge Scholars Publishing, 2014.

  This book includes a chapter about some of the work conducted in this thesis and previously published in a conference.

A.2 Technology Transfers

Some of the knowledge and the expertise gained during this PhD tenure were transferred to SeeByte under two projects. In 2010, the ATR approach proposed in this thesis was introduced to SeeByte and consultancy was provided around the best practice in using this approach. In 2012, some extensions to the proposed approach were presented to SeeByte and consultancy was provided around integrating them to SeeByte software.

A.3 Research Projects

- Scientific collaboration in a Competition of Ideas project (RT/COM/5/059). The project mainly researched semantic world models for multi-platforms collaboration. As part of this project, a fully autonomous solution for subsea survey,

target detection, and intervention operations was demonstrated. The ATR approach proposed in this thesis was successfully exploited in this demonstration. It was integrated into the software of two AUVs: REMUS for ATR in sidescan sonar and Nessie for ATR in forward looking sonar.

- Scientific collaboration in a DSTL (Defence Science and Technology Laboratory) research programme (MHP Sonar - Task 37 - Activity 3). This programme involved academic and industrial partners including Heriot-Watt University, University of Bristol, University College London, SeeByte, and Atlas Elektronik. The programme mainly looked at the benefits and limitations of current ATR algorithms, highlighted technology gaps and proposed new approaches to improve algorithm performance where appropriate. As part of this programme, the ATR approach proposed in this thesis has been independently evaluated by DSTL on various large sets of sonar data. In direct comparison with some existing approaches available to DSTL, our approach has provided the best processing speed and detection accuracy (Y. Petillot, personal communication, February 12, 2014).

References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[2] P. Blondel, *The handbook of sidescan sonar*. Springer, 2009.

[3] J. P. Fish and H. A. Carr, *Sound underwater images: a guide to the generation and interpretation of side scan sonar data*. Lower Cape Pub., 1990.

[4] V. Myers, A. Fortin, and P. Simard, "An automated method for change detection in areas of high clutter density using sonar imagery," in *Proceedings of the International Conference on Underwater Acoustic Measurements*, 2009.

[5] X. Lurton, *An introduction to underwater acoustics: principles and applications*. Springer, 2002.

[6] M. Hayes and P. Gough, "Synthetic aperture sonar: A review of current status," *IEEE Journal of Oceanic Engineering*, vol. 34, no. 3, 2009.

[7] G. J. Dobeck, J. C. Hyland, and L. Smedley, "Automated detection and classification of sea mines in sonar imagery," in *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets II*, vol. 3079, Jul 1997.

[8] R. Duda, P. Hart, and D. Stork, *Pattern classification*, 2nd ed. Wiley, 2001.

[9] M. Azimi-Sadjadi, D. Yao, Q. Huang, and G. Dobeck, "Underwater target classification using wavelet packets and neural networks," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, May 2000.

[10] B. Calder, L. Linnett, and D. Carmichael, "Bayesian approach to object detection in sidescan sonar," in *Proceedings of the International Conference on Image Processing and Its Applications*, vol. 2, Jul 1997.

[11] S. Perry and L. Guan, "Pulse-length-tolerant features and detectors for sector-scan sonar imagery," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 1, 2004.

[12] G. J. Dobeck and J. T. Cobb, "Fusion of multiple quadratic penalty function support vector machines QPFSVM for automated sea mine detection and classification," in *Proceedings of the SPIE AeroSense Conference*, 2002.

[13] N. Ma and C. Chia, "False alarm reduction by LS-SVM for manmade object detection from sidescan sonar images," in *Proceedings of the IEEE Oceans Conference*, Jun 2007.

[14] J. Nelson and N. Kingsbury, "Fractal dimension, wavelet shrinkage and anomaly detection for mine hunting," *IET Journal of Signal Processing*, vol. 6, no. 5, Jul 2012.

[15] D. Williams and J. Groen, "Multi-view target classification in synthetic aperture sonar imagery," in *Proceedings of the International Conference on Underwater Acoustic Measurements*, 2009.

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

[17] J. A. Fawcett, "Image-based classification of side-scan sonar detections," in *Proceedings of the Computer-Aided Detection/Computer-Aided Classification Conference*, Nov 2001.

[18] J. Del Rio Vera, E. Coiras, J. Groen, and B. Evans, "Automatic target recognition in synthetic aperture sonar images based on geometrical feature extraction," *Journal on Advances in Signal Processing*, 2009.

[19] E. Dura, Y. Zhang, X. Liao, G. Dobeck, and L. Carin, "Active learning for detection of mine-like objects in side-scan sonar imagery," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 2, 2005.

[20] I. Quidu, J. Malkasse, G. Burel, and P. Vilbe, "Mine classification using a hybrid set of descriptors," in *Proceedings of the IEEE Oceans Conference*, 2000.

[21] I. Quidu, J.-P. Malkasse, G. Burel, and P. Vilbé, "Mine classification based on raw sonar data: an approach combining fourier descriptors, statistical models and genetic algorithms," in *Proceedings of the IEEE Oceans Conference*, vol. 1, 2000.

[22] J. A. Fawcett, "Computer-aided detection and classification of minelike objects using template-based features," in *Proceedings of the IEEE OCEANS Conference*, vol. 3, 2003.

[23] V. Myers and J. Fawcett, "A template matching procedure for automatic target recognition in synthetic aperture sonar imagery," *IEEE Signal Processing Letters*, vol. 17, no. 7, Jul 2010.

[24] S. Reed, Y. Petillot, and J. Bell, "Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information," *Journal on Radar, Sonar and Navigation*, vol. 151, no. 1, Feb 2004.

[25] S. Johnson and M. Deaett, "The application of automated recognition techniques to side-scan sonar imagery," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 1, Jan 1994.

[26] E. Dura, J. Bell, and D. Lane, "Superellipse fitting for the recovery and classification of mine-like shapes in sidescan sonar images," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 4, Oct 2008.

[27] I. Quidu, N. Burlet, J.-P. Malkasse, and F. Florin, "Automatic classification for MCM systems," in *Proceedings of the IEEE Oceans Conference*, vol. 2, 2005.

[28] F. Maussang, J. Chanussot, A. Hetet, and M. Amate, "Higher-order statistics for the detection of small objects in a noisy background application on sonar imaging," *Journal on Advances in Signal Processing*, 2007.

[29] G. Dobeck, "Algorithm fusion for automated sea mine detection and classification," in *Proceedings of the IEEE Oceans Conference*, vol. 1, 2001.

[30] T. Aridgides, M. Fernandez, and G. Dobeck, "Fusion of adaptive algorithms for the classification of sea mines using high resolution side scan sonar in very shallow water," in *Proceedings of the IEEE Oceans Conference*, vol. 1, 2001.

[31] C. M. Ciany and W. C. Zurawski, "Enhanced ATR using fisher fusion techniques with application to side-looking sonar," in *SPIE Defense, Security, and Sensing*, 2010.

[32] J. Schmidhuber, "Deep learning in neural networks: An overview," *Journal of Neural Networks*, vol. 61, 2015.

[33] P. Sinha, "Qualitative representations for recognition," in *Proceedings of the International Workshop on Biologically Motivated Computer Vision*, 2002.

[34] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, 2000.

[35] F. C. Crow, "Summed-area tables for texture mapping," in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, 1984.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005.

[37] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, 2002.

[38] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, 2004.

[39] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The Annals of Statistics*, vol. 28, no. 2, Apr 2000.

[40] R. Meir and G. Rätsch, "An introduction to boosting and leveraging," in *Advanced Lectures on Machine Learning*. Springer, 2003.

[41] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Proceedings of the European Conference on Computational Learning Theory*, 1995.

[42] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the margin: a new explanation for the effectiveness of voting methods," *The Annals of Statistics*, vol. 26, no. 5, Oct 1998.

[43] K. Tieu and P. Viola, "Boosting image retrieval," *International Journal of Computer Vision*, vol. 56, no. 1-2, 2004.

[44] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Journal of Machine Learning*, vol. 37, no. 3, 1999.

[45] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Machine Learning*, vol. 40, no. 2, 2000.

[46] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *Proceedings of the Annual Symposium of the German Association for Pattern Recognition (DAGM)*, 2003.

[47] Y. Freund, "A more robust boosting algorithm," *arXiv preprint arXiv:0905.2138*, 2009.

[48] H. Masnadi-Shirazi, V. Mahadevan, and N. Vasconcelos, "On the design of robust classifiers for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[49] J. K. Tsotsos, S. M. Culhane, W. Y. Kei Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Journal of Artificial Intelligence*, vol. 78, no. 1, 1995.

[50] H. Schneiderman, "Feature-centric evaluation for efficient cascaded object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[51] Y. Amit, "A neural network architecture for visual selection," *Journal of Neural Computation*, vol. 12, no. 5, 2000.

[52] F. Fleuret and D. Geman, "Coarse-to-fine face detection," *International Journal of Computer Vision*, vol. 41, no. 1-2, 2001.

[53] J. R. Quinlan, "Induction of decision trees," *Journal of Machine Learning*, vol. 1, no. 1, 1986.

[54] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, 2006.

[55] N. Dalal, "Finding people in images and videos," Ph.D. dissertation, Institut National Polytechnique de Grenoble, Jul 2006.

[56] "Open Source Computer Vision OpenCV 2.1.0," http://opencv.org/, accessed: Jan 2011.

[57] D. Navon, "Forest before trees: the precedence of global features in visual perception," *Journal of Cognitive Psychology*, vol. 9, no. 3, 1977.

[58] P. Blanchart, M. Ferecatu, S. Cui, and M. Datcu, "Pattern retrieval in large image databases using multiscale coarse-to-fine cascaded active learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, 2014.

[59] M. Pedersoli, A. Vedaldi, J. Gonzàlez, and X. Roca, "A coarse-to-fine approach for fast deformable object detection," *Journal of Pattern Recognition*, vol. 48, no. 5, 2015.

[60] R. Lienhart and J. Maydt, "An extended set of Haar-like Features for rapid object detection," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, 2002.

[61] S. Z. Li and Z. Zhang, "Floatboost learning and statistical face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, 2004.

[62] T. Mita, T. Kaneko, and O. Hori, "Joint Haar-like features for face detection," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, 2005.

[63] C. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 1998.

[64] R. Xiao, L. Zhu, and H.-J. Zhang, "Boosting chain learning for object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2003.

[65] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast rotation invariant multi-view face detection based on real adaboost," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[66] J. Nelson and N. Kingsbury, "Fractal dimension based sand ripple suppression for mine hunting with sidescan sonar," in *Proceedings of the International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, 2010.

[67] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian detection using infrared images and histograms of oriented gradients," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2006.

[68] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[69] Y. Pailhas, Y. Petillot, and C. Capus, "High-resolution sonars: What resolution do we need for target recognition?" *EURASIP Journal on Advances in Signal Processing*, 2010.

[70] Y. Petillot, S. Reed, E. Coiras, and J. Bell, "A framework for evaluating underwater mine detection and classification algorithms using augmented reality," *IEEE Journal of Oceanic Engineering*, 2006.

[71] J. Bell, "A model for the simulation of sidescan sonar," Ph.D. dissertation, Heriot-Watt University, 1995.

[72] P. Hill, A. Achim, and D. Bull, "Underwater target detection in synthetic aperture sonar data," in *Proceedings of Sensor Signal Processing for Defence*, 2010.

[73] Y. Pailhas, P. Patron, J. Cartwright, F. Maurelli, Y. Petillot, J. Sawas, and N. Valeyrie, "Fully integrated multi-vehicles mine countermeasure missions," in *Proceedings of the International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results (UAM)*, Kos, Greece, 2011.

[74] F. Maurelli, P. Patron, J. Cartwright, J. Sawas, Y. Petillot, and D. Lane, "Integrated MCM missions using heterogeneous fleets of AUVs," in *Proceedings of the IEEE Oceans Conference*, Yeosu, Korea, 2012.

[75] P. Viola and M. Jones, "Fast and robust classification using asymmetric adaboost and a detector cascade," *Advances on Neural Information Processing Systems*, 2001.

[76] J. Wu, M. D. Mullin, and J. M. Rehg, "Linear asymmetric classifier for cascade detectors," in *Proceedings of the International Conference on Machine Learning*, 2005.

[77] M. Kobetski and J. Sullivan, "Improved boosting performance by explicit handling of ambiguous positive examples," in *Pattern Recognition Applications and Methods*, 2015.

[78] S. C. Brubaker, J. Wu, J. Sun, M. D. Mullin, and J. M. Rehg, "On the design of cascades of boosted ensembles for face detection," *International Journal of Computer Vision*, vol. 77, no. 1-3, 2008.

[79] L. Bourdev and J. Brandt, "Robust object detection via soft cascade," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005.

[80] J. Fawcett, V. Myers, D. Hopkin, A. Crawford, M. Couillard, and B. Zerr, "Multiaspect classification of sidescan sonar images: Four different approaches to fusing single-aspect information," *IEEE Journal of Oceanic Engineering*, vol. 35, no. 4, 2010.

[81] M. Couillard, J. Fawcett, V. Myers, and M. Davison, "Optimizing time-limited multi-aspect classification," in *Proceedings of the Institute of Acoustics*, 2007.

[82] Y. Petillot, S. Reed, and V. Myers, "Mission planning and evaluation for minehunting AUVs with sidescan sonar: Mixing real and simulated data," NATO Undersea Research Centre, Tech. Rep., 2005.

[83] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012.

[84] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, 2006.

[85] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, 2006.

[86] J. Sawas, Y. Petillot, and Y. Pailhas, "Cascade of boosted classifiers for rapid detection of underwater objects," in *Proceedings of the European Conference on Underwater Acoustics (ECUA)*, Istanbul, Turkey, 2010.

[87] Y. Petillot, Y. Pailhas, J. Sawas, N. Valeyrie, and J. Bell, "Target recognition in synthetic aperture and high resolution side-scan sonar," in *Proceedings of the European Conference on Underwater Acoustics (ECUA)*, Istanbul, Turkey, 2010.

[88] Y. Petillot, Y. Pailhas, J. Sawas, N. Valeyrie, and J. Bell, "Target recognition in synthetic aperture sonar and high resolution side scan sonar using AUVs," in *Proceedings of the International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, Lerici, Italy, 2010.

[89] J. Aulinas, A. Fazlollahi, J. Salvi, X. Llado, Y. R. Petillot, J. Sawas, and R. Garcia, "Robust automatic landmark detection for underwater SLAM using side-scan sonar imaging," in *Proceedings of the International Conference on Mobile Robots and Competitions (Robotica)*, Lisbon, Portugal, 2011.

[90] J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," in *Proceedings of the European Conference on Underwater Acoustics (ECUA)*, Edinburgh, Scotland, 2012.

[91] J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," in *Proceedings of the International Conference on Underwater Remote Sensing (ICoURS)*, Brest, France, 2012.

[92] J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," *Proceedings of Meetings on Acoustics*, vol. 17, 2013.

[93] S. Reed, Y. Petillot, J. Nelson, P. Hill, J. Sawas *et al.*, "Assessment of automatic target recognition and fusion processes for cluttered environments," SeeByte, Tech. Rep., 2013.

[94] I. Quidu, V. Myers, and B. Zerr, *Proceedings of the 2012 international conference on detection and classification of underwater targets*. Cambridge Scholars Publishing, 2014.