Proceedings of the 2012 International Conference on Detection and Classification of Underwater Targets

Proceedings of the 2012 International Conference on Detection and Classification of Underwater Targets

Edited by

Isabelle Quidu, Vincent Myers and Benoit Zerr

CAMBRIDGE SCHOLARS

PUBLISHING

Proceedings of the 2012 International Conference on Detection and Classification of Underwater Targets
Edited by Isabelle Quidu, Vincent Myers and Benoit Zerr

This book first published 2014

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-5709-2, ISBN (13): 978-1-4438-5709-3

TABLE OF CONTENTS

FOREWORD

The International Conference on Underwater Remote Sensing (ICoURS) was held from 8-11 October 2012 at the Le Quartz Conference Centre in Brest, France as part of Sea Tech Week. It was composed of three symposia: Quantitative Monitoring of the Underwater Environment (MOQESM), Advances in Seafloor Mapping, and Detection and Classification of Underwater Targets (DCUT), and the last one being the subject of these proceedings. DCUT took place on 9-10 October and was organized by the Ocean Sensing and Mapping Team of ENSTA Bretagne and was, in spirit, the 3rd conference on what can broadly be described as Automatic Target Recognition (ATR) for underwater applications: The first was CAD/CAC 2001 organized by Defence R&D Canada and held in Halifax, Nova Scotia, Canada; and the second was the International Conference on Detection and Classification of Underwater Targets, organized by Heriot-Watt University and held in Edinburgh, Scotland.

It was noted during the Plenary Session by keynote speaker Dr. John Fawcett that during the 11 years that have passed since the original CAD/CAC 2001 conference, progress in the fields of pattern recognition, machine learning, image and signal processing, as well as the advent of high-resolution sensors such as synthetic aperture sonars have led to significant improvements in underwater ATR technology. Perhaps more salient, however, is the now ubiquitous presence of Autonomous Underwater Vehicles (AUV), making high-performing, computationally-efficient ATR no longer simply an aid for human operators, but rather a necessary technology to enable the use of unmanned systems. In addition, the applications of this technology has started to move out of military applied research programs, typically the naval mine countermeasures (MCM) community, and is being applied in civilian applications such as pipeline inspection and environmental monitoring. For this reason, the papers of these proceedings will also be of interest to researchers working the area of remote sensing (for instance, with Synthetic Aperture Radar) as well as medical imaging and robotic perception.

The increasingly interdisciplinary nature of this field is evident by the papers that were presented during DCUT: From traditional acoustics/sonar to non-acoustic methods such as ground penetrating radar, magnetic gradiometry and video; application of machine learning, pattern recognition, image processing, optimization, anomaly detection, acoustic modelling and data fusion; as well as applications such as environmental characterization and change detection.

These proceedings contain 20 papers whose abstracts were reviewed by at least two reviewers. We would like to thank all of the reviewers in the Scientific Committee listed immediately below for providing their time and effort to ensure the quality of the articles in this conference. Also included are abstracts of four papers from the Poster Session. We would sincerely like to thank Annick Billon-Coat for her help in organizing this conference, as well as the staff at the Le Quartz conference centre and Brest Métropole Océane for their support.

The discussions and collaborations that ensue from these conferences are key to moving the field forward. With a relatively small community, it is important that we come together occasionally in a specialized forum in order to share ideas, show some fresh results and obtain feedback on our work. We look forward to seeing you all again, along with some new faces, during the next incarnation of the DCUT conference, wherever and whenever it may be.

Vincent Myers, Isabelle Quidu and Benoit Zerr
Brest, France, October 2012

SCIENTIFIC COMMITTEE

Maude AMATE (DGA), Toulon, France

Didier BILLON (Thales Underwater Systems), Brest, France

Didier CHARLOT (IxBlue), Brest, France

Gerald DOBECK (NSWC), Panama City, Florida, USA

Yves DUPONT (Belgian Navy), Belgium

John FAWCETT (DRDC), Halifax, Canada

Johannes GROEN (Atlas Electronik), Bremen, Germany

Ursula HÖLSCHER-HÖBING (Atlas Electronik), Bremen, Germany

Dieter KRAUS (University of Applied Sciences), Bremen, Germany

Øivind MIDTGAARD (FFI), Kjeller, Norway

Yvan PETILLOT (Heriot-Watt University), Edinburgh, Scotland

Benoît QUESSON (TNO), The Hague, The Netherlands

Scott REED (SeeByte), Edinburgh, Scotland

David WILLIAMS (CMRE), La Spezia, Italy

PLENARY SESSION

PREFACE

AUTOMATIC TARGET RECOGNITION METHODS FOR SIDESCAN SONAR IMAGES: THE ADVANCES AND THE CHALLENGES

JOHN A. FAWCETT

Abstract

Over approximately the last decade Defence R&D Canada – Atlantic, Canada (also known as DRDC Atlantic) has been involved with research into Automated Target Recognition (ATR) algorithms for sidescan sonar imagery. In this paper, some of the past and present DRDC Atlantic work in ATR will be discussed with some illustrative experimental results. Related work by other authors is also discussed.

Keywords: Sidescan Sonar, Automatic Target Recognition, Detection, Classification.

1. Introduction

In 2001 Defence R&D Canada – Atlantic hosted the conference CAD/CAC 2001 in Halifax as an initial start into a research program for the development of automated sidescan sonar detection and classification methods in support of the Canadian Remote Minehunting System project. Now, eleven years later, there has been much progress in image-processing and pattern recognition algorithms. Synthetic aperture sonars (SAS) have significantly improved the resolution of the images of the seabed. The use of autonomous underwater vehicles (AUVs) for sonar surveys has become very common and there has been interest in making AUVs more intelligent and adaptable during a mission. For example, after an initial standard survey, an AUV could revisit a list of Automatic Target Recognition (ATR) contacts for another sonar look [1] or during the survey the AUV could perform a multi-aspect run at each potential target [12, 23]. However, such concepts rely upon accurate and robust ATR processing. The development of reliable and computationally-efficient ATR methods is more relevant than ever in minehunting.

(a) (b)

Figure 1: The semi-submersible remote minehunting vehicle DORADO (a) out of the water (b) underway

For several years, the Mine Counter-Measures group (now Mine Warfare group) at DRDC Atlantic, in collaboration with Canadian industry, was involved with the development of the remote minehunting vehicle DORADO and its associated sensors and software. This vehicle is shown in Fig. 1(a) out of the water. On the bottom of the Aurora towfish, the Klein 5500 sidescan sonar can be seen. The actively controlled towfish can be winched out to depth. The sonar data is transmitted back, in near real-time, to a mother ship at distances of up to 12 km away from the DORADO. The vehicle is shown underway in the water in Fig.1(b). The data is displayed as a waterfall on this ship and an operator and/or background ATR algorithms analyze the data for mine-sized contacts. Much of the DRDC Atlantic data used for research over the last decade was collected from various trials using this system.

In the following sections, the ATR processing stream is broken down into three basic steps: (1) normalization of sonar data (2) simple and rapid automated detection and (3) more detailed analysis of small images (mugshots or snippets extracted from the second step). This breakdown is historically the approach taken at DRDC Atlantic but the divisions are somewhat arbitrary. For example, some newer methods of automated detection/classification [34, 38] combine, to some extent, steps (2) and (3).

2. Normalization

Typically, recorded sonar data shows systematic amplitude variations with respect to range (travel time) and the sonar's beampattern. The large scale amplitude variations can be reduced by computing a local background mean amplitude and dividing through by this value. For the Klein 5500 data, we typically compute, on a per file and per side basis, an average empirical amplitude/cross-range curve and normalize the data by this curve. For some sonar data, more complicated vertical beampattern effects are observed in the data and need to be accounted for. Dobeck [39, 40] has described sophisticated normalization algorithms. In these papers, he also emphasizes that by reducing the system amplitude variations the subsequent false alarm rate in the automated detection phase can be significantly reduced.

There are also environmental features in the data which will cause significant false alarms for many automated detectors; in particular, sand ripples cause a sequence of highlights and shadows in the sonar data which can resemble a minelike structure. In [40, 41], Fourier- and wavelet-based methods are described to reduce the effect of ripples on the sonar image. In [42], Williams mitigates the effects of ripples during the detection phase by considering the distribution of elliptical descriptors of the shadow regions and eliminating those regions which are consistent with ripples (with some additional criteria to mitigate against "losing" targets). In Fig. 2(a) we show an unnormalized sonar image (Marine Sonic) from a Remus AUV. A surface echo has already been suppressed from the original image by predicting its position in the image and replacing abnormally high values with a local median value. The coloured lines (cyan and green) indicate some predicted grazing angle curves on the seabed as the altitude of the AUV varies (the red lines indicate along-track regions of turns). By integrating the amplitudes along these curves, across-track normalization curves can be computed. The resulting normalized image is shown in Fig. 2(b). Figs. 2(c) and 2(d) show the results of a simplistic segmentation of the image into 5 values representing the range of deep shadow to high highlight; first using the normalized image (Fig. 2(c)) and secondly (Fig. 2(d)) combining the segmentation of Fig. 2(c) with a segmentation [2] after filtering the image using the method of Dobeck [40]. The results of Fig.2(d) show that much of the shadow due to the sand ripples has been eliminated. More details of this processing are described in [2].

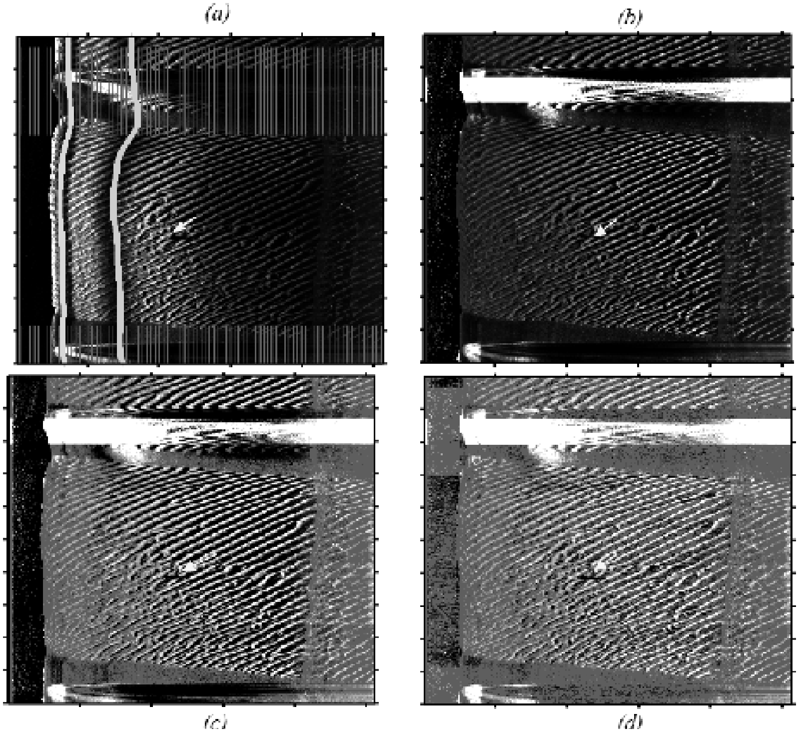

Figure 2: A Marine Sonic sonar file showing some different phases of normalization: (a) unnormalized data with representative grazing angle curves (b) normalized data (c) rebinned into shadow-highlight values and (d) rebinning in combination with Fourier filtering. The yellow arrow indicates a mine-like object. Data Source: NURC.

3. Automatic Target Detection

Given a normalized, filtered sonar image, the DRDC Atlantic detection process consists of cross-correlating the image (or a transformation of the image) with various filters. One which we have used for several years is based upon the work of [3]. As mentioned above, the sonar image is roughly segmented into 5 basic values based upon the median value or on percentiles of the pixel values. A two-dimensional filter consisting of +1 for highlight and -1 for shadow is then cross-correlated with the data and regions exceeding a threshold are taken to be detection regions. The predicted shadow length for a target of fixed height should increase linearly with range. This is difficult to implement with FFT-based crosscorrelations and, in the past, we used 3 different sized (in terms of shadow extent) filters to address this issue. The implementation we use in a structured C++ development does utilize a continuously growing shadow. There are also a variety of other filter possibilities. We have found that the local Lacunarity [4] (defined as variance of pixel values/squared mean value) can, for some environments, be a very good detection feature. Here too, this feature can be computed by using sliding windows to compute the local means and mean squared values. In Fig. 3(a) we show a sonar image (from the NURC AUV/synthetic aperture sonar vehicle, MUSCLE) (unnormalized), in Fig. 3(b) the match-filtered output, in Fig. 3(c) the Lacunarity output, and in Fig. 3(d) the detections (yellow) based upon a match-filtered threshold and those which exceed the threshold for the match filter and also another threshold for Lacunarity (cyan). Here the seabed has patches of the seagrass Posidonia. This produces “natural” pairs of highlight and shadow which can cause detector false alarms. There is a dummy target which can be observed as a high output for both the match-filter and the Lacunarity images.

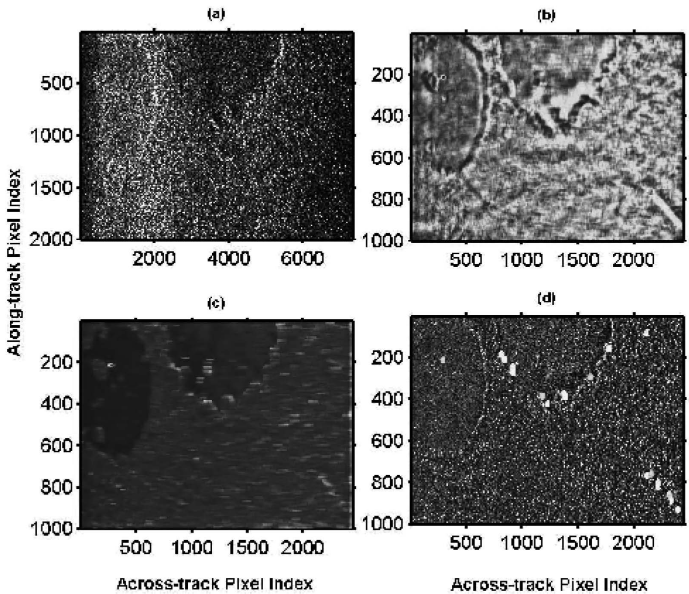

Figure 3: (a) unnormalized NURC MUSCLE data tile (b) matched-filter output (c) Lacunarity output (d) resulting detections from a matched-filter threshold (yellow boxes) – also satisfying a threshold on Lacunarity (cyan). The output images (b),(c) and (d) are computed on a reduced version of the image of (a). Data Source: NURC.

Williams [42] uses a moving window to find regions of shadow and then considers those shadow regions with an associated highlight or echo region. In general, one can compute a number of features for a detection region: various filter outputs and local statistical values, and the detection process (or a secondary detection process) can be improved by looking for combinations of features which improve the detection/false alarm ratio [2, 5, 6]. The method of Williams [42] is a simple example of a cascade: (1) a simple detection method (*e.g.,* existence of shadow) is used to eliminate much of the sonar image from consideration and then a second detection method (*e.g.,* the existence of an associated echo) is applied to those regions of the image which remain after step (1). In general, one can use a cascade of several detectors to sequentially eliminate regions of the image for further consideration. At each successive level of the cascade, the detection test used may be more complex (*e.g.* may involve more features), but this is offset by the fact that the number of image regions to process at the higher levels is smaller. Sawas et al [34] and Petillot el al. [38] used a trained Haar Cascade detection method to obtain very good detection performances. This type of detection method was first developed in the face-recognition community [7, 8] and various training and testing methods are available in the openCV [9] library. In Fig. 4 we show the results from a face detection method available in the openCV library which uses an existing trained Haar cascade for face detection. We have also used the openCV software to train our own cascade for sonar images (MUSCLE data from NURC) and utilize the same face detection algorithm (with some adjustments of the parameters). A sonar image with the resulting detections is shown in Fig. 5.

Figure 4: Two Canadian scientists and a French scientist relaxing on a trial with their faces detected by the openCV face detection routine.

Whether one wishes to use this type of detection method or not, a very powerful concept utilized in this application is that of the integral image and the rapid computation of rectangular-based features. The integral image is computed from the image $I(i,j)$ by computing its two-dimensional cumulative sum. Then the summed value of the image over a specified rectangle can be expressed as the sums and differences of the 4 corner points. If we wish to consider two adjacent rectangles, one positive and one negative, this can be expressed as 6 operations. In fact, the sliding window output from a number of different combinations of adjacent rectangles or nested rectangles can be very efficiently computed from the integral image. In addition, this concept has been extended to include rotated rectangular features [8]. This method can be applied to our simple matched-filter or Lacunarity detectors (in this case, also computing an image of squared pixel values). The increase of the shadow length with the across-track pixel index is very simply included with this approach. Once a detector, a fused set of detectors, or a cascade of detectors has determined a detection point, then a small image about this point (mugshot) is extracted. The resulting set of small images is then passed to the next stage of analysis – classification.

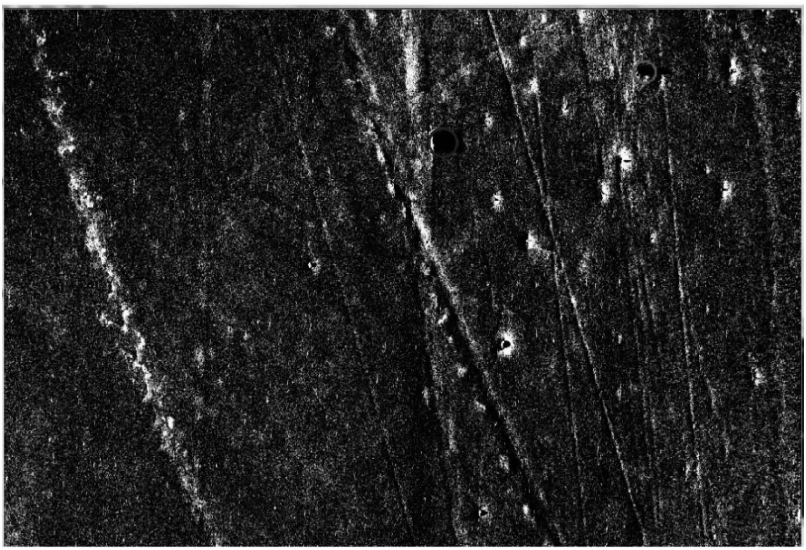

Figure 5: A sonar image with a fairly complex seabed and two detected objects (no false alarms) using a trained Haar-Cascade classifier. Data Source: NURC.

4. Classification

The boundary between classification and detection is not well-defined. In the classification phase, it is often desired to classify the mugshots resulting from the detection phase as a possible generic mine type: *e.g.*, cylinder, sphere, truncated cone, etc… However, since the detection phase has usually yielded many false alarms, it is important to eliminate more of these. We have tended to think of the detection phase of ATR as being computationally efficient, but not being particularly sophisticated. The classification phase utilizes more sophisticated methods which require more computation time (*e.g.,* feature computation, support vector machine classification and template-matching). The face-detection methods described in Section 3 are trained with a large set of data (positives and negatives) and the training time can be long. However, in the detection phase they are very computationally efficient. These methods blur the boundary somewhat between the detection and classification phases.

There are two main approaches in sonar image classification. One is shadow- and highlight-feature based. In this approach, the mugshot is first segmented into shadow and highlight regions. This is normally a fairly sophisticated segmentation approach. That is, instead of using simple hard image thresholds to define shadow and highlight, these algorithms consider the pixel values and also the neighbouring values in order to obtain accurate representations of the shadow and highlight regions. Some of these algorithms are an iterative threshold and connectivity method [10], Markov Random Fields [11], Statistical Snakes [11], and Fourier Descriptors [13]. Although the concept of shadow and highlight segmentation is straightforward, it can be surprisingly difficult to develop robust methods for complex seabed types. Once the segmentation has been performed (and the appropriate regions associated with the detected object) various features based upon these regions can be computed. These features are often geometrical or statistical in nature: for example, the estimated height of the object from the shadow length (and known range/altitude of detection), the length of the object, the ratio of the convex area/area of the shadow, the standard deviation of the shadow profile, the eccentricity of the shadow, the orientations of the shadow and highlight regions, the width of the highlight region, etc. A full description of some of the features we have used at DRDC Atlantic can be found in [14]. In [15] the height profile of an object (estimated from the shadow length) was considered as the feature vector. There are also choices of features which are invariant to scaling and rotation [16]. In Fig.6 we show a screen capture of the display from the DRDC Atlantic Sonar Image Processing

System (SIPS) showing the results for an automatic segmentation of shadow and highlight and some computed feature values. These feature values can subsequently be used for training and testing classifiers.

Given a set of features, a classifier can be trained using labelled mugshots. This can be a binary-classifier or a multi-class classifier. There are many possible choices for a classifier. We have often used a kernel-regression method [17] with an exponential kernel based at each training point. For multiple classes, our method is equivalent to solving multiple binary problems (i.e, 1 if a specific target type and -1 if not). In this approach, we also specifically consider clutter to be a class and train with it. This type of approach works well, as long as the preliminary shadow/highlight segmentation and computed features are good.

Sonar images of mine-like objects are collected during sea trials with dummy mine shapes deployed on the seabed. There are often only a few (*e.g.* 9) deployed at a site and these are repeatedly imaged at different ranges and aspects to yield a few hundred images. The danger in training and testing with such a data set is that, despite the changing sonar position, it is often the same object (and surrounding seabed) being imaged. In the Citadel trial [14, 18], targets and a rock were deployed at 2 sites. The rocks were mine sized but were taken to represent the clutter class. They were different at the 2 sites. In Table 1 we show the averaged confusion matrices from a set of training/testing runs. First, the confusion matrix data for a classifier tested using data from Site 2 when trained with Site 1 data is shown. Below these classification rates, the results for training and testing with just the Site 1 data are shown. In this second case, the classifier was able to distinguish the rock from the targets about 83% of the time at Site 1. However, when the classifier trained with Site 1 data was used at Site 2, the rock at Site 2 is most often confused with the truncated cone shape. Thus the “clutter” sample at Site 1 was not sufficiently diverse to provide good clutter discrimination at Site 2. The classification results for the other dummy target types at Site 2 are good, with the classification of the cylindrical shape being somewhat poorer. This example illustrates a fundamental concern for ATR using trained classification: its ability to perform well in a new environment.

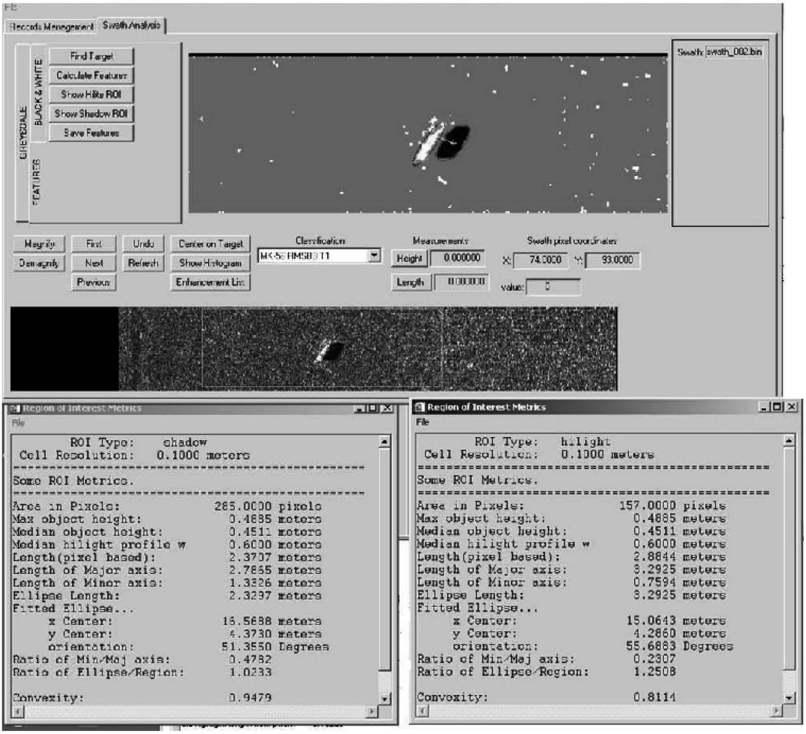

Figure 6: The segmentation and feature computation tool from DRDC Atlantic SIPS database viewer.

Table 1: The confusion matrices resulting from using (1) a classifier trained with Citadel Site 1 data and tested with Site 2 data and below (2) a classifier trained with Citadel Site 1 data and tested with Site 1 data.

	Rock	T. Cone	Wedge	Cyl
Rock	**0.26**	0.53	0.17	0.03
	0.83	0.08	0.06	0.03
T. Cone	0.08	**0.82**	0.09	0.02
	0.05	**0.92**	0.03	0.00
Wedge	0.00	0.01	**0.99**	0.00
	0.01	0.04	**0.93**	0.02
Cyl	0.15	0.03	0.06	**0.76**
	0.05	0.03	0.03	**0.89**

The other type of classification approach is to work more directly with the mugshots' pixel values. There are pros and cons to this approach. The advantages are that it avoids the preliminary shadow/background/highlight segmentation. There are often seabed or target features which can cause problems for shadow/highlight segmentation. Also, the image-based approaches do not rely on explicitly defined features. On the other hand, they may be sensitive to the range of the mugshot detection (because of the increasing shadow length), normalization effects, the size of the mugshot, etc... Some of these problems can be mitigated. For example, careful image normalization and consistent placement of the objects within the mugshots can help classification performance. One can consider the use of expansion functions such as Zernicke polynomials [16, 19, 20] which are range/aspect independent.

Once again, as with the previously-described Haar Cascade methods for detection, the image-based classification methods follow closely some of the approaches used in facial recognition. In fact, as shown in [34, 38] one can train the Haar Cascade method for specific mine types. An approach which has been enjoying popularity in the last few years is template matching [21, 22, 24, 25, 26, 27] and this is certainly an approach which is used, in general, in the object detection community. Various algorithms can be found, for example, in the openCV library. The idea is to construct for the range (the sonar range corresponding to the across-track pixel of the detection) of the mugshot detection a set of ray-trace model templates (*i.e.,* a basic highlight/shadow structures) encompassing the various possible target types and a discrete set of aspects. In our implementation, a library of precomputed templates at a discrete set of ranges is used. However, one can compute the templates "on the fly" with a ray-tracing subroutine. In many of the template approaches, the templates are then cross-correlated with the mugshot (or a rebinned version of it). The maximum value of the output-filtered image (the cross-correlation is typically performed by moving the template about the image in some neighbourhood of the detection centre) is computed for each template and the maximum of these values is taken to indicate the best target and aspect match. If this value is not sufficiently high, then the object may be deemed to be clutter. There are a variety of different cross-correlation measures which can be used. Reference [26] discusses various template-matching measures. For example, one can simply use the true cross-correlation value (as defined for normxcorr2 in the MATLAB image processing toolbox based upon the method of [37])

$$C(u,v) = \frac{\sum_{x,y} (I(x,y) - \bar{I}_{u,v})(t(x-u,y-v) - \bar{t})}{(\sum_{x,y} (I(x,y) - \bar{I}_{u,v})^2)^{1/2} N_T} \tag{1}$$

Here the template t is centred at (u,v) and x,y vary over the portion of the image contained within the template.

A template $t(x,y)$ is moved about the image and a local image mean $\bar{I}_{u,v}$ and normalization of I within the region of the template is computed. We have used N_T in Eq.(1) to denote the L_2 norm of the template. It is interesting to note that the computation of the image mean and standard deviation within the the moving template window is most efficiently accomplished using the method of integral images [37]. A simpler expression for Eq.(1) results when the mugshot and template's mean values are taken to be zero [26] (we typically first remap the mugshot and template into positive and negative values for relative highlight and shadow regions),

$$C_1(u,v) = \frac{\sum_{x,y} (I(x,y)t(x-u,y-v)}{N_T N_I} \tag{2}$$

where N_T is the L_2 norm of the template and N_I is the L_2 norm of the image within the extent of the template. In Fig. 7(a) we show a cylinder lying in a sand ripple field (NURC MUSCLE data). The template yielding the best match is shown in Fig. 7(b). As can be seen, the match is very reasonable. In Fig. 7(c) the variation of the maximum value of $C_1(u,v)$ is shown as a function of the hypothesized templates and it can be seen that there is a significant relative peak in the neighbourhood of the correct match. Although this particular result is encouraging, there can be problems with the method. Even in this example, the actual value of the output is fairly low – approximately 0.28. This is due to the fact that there is a fair amount of speckle in the shadow regions and in our remapping of the original image into [-1 1], much of the shadow region is defined as background. This means that if we had set a simple threshold to reject clutter, this target may have been lost. Also, although we do not show it here, there was a rock in this dataset which was quite mine-like and simply using the correlation value to discriminate this particular object is not reliable. Of course, images from multiple sonar aspects could help this situation. Also, simply using a single correlation value as a means of classification may not be optimal. There is more information in the entire correlation curve (*e.g.* Fig. 7(c) which is not used and one can also consider the curves from other correlation measures. We have [21, 27] considered using various sets of template features for classification.

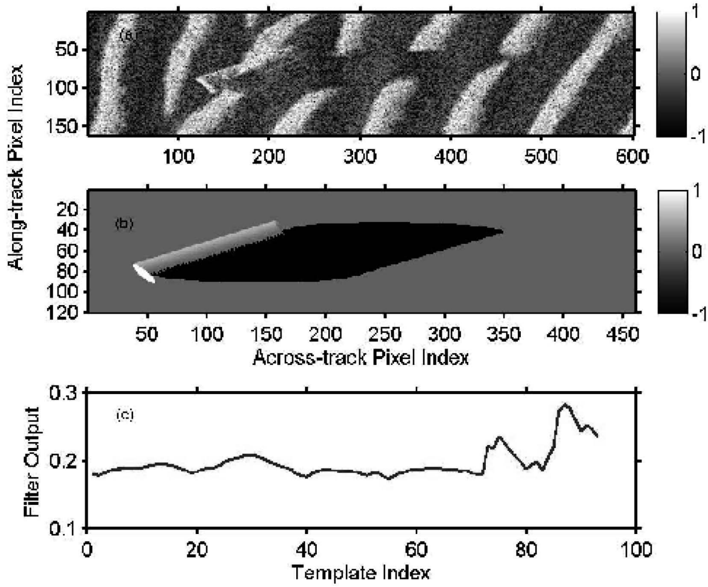

Figure 7: Template matching Eq.(2): (a) the input mugshot (a normalized, remapped version) (b) the best matching template and (c) the variation of the correlation score with template index. Data Source: NURC.

In [18, 28] we considered expressing a collection of mughots in terms of their principal component representation. That is, each mugshot can be rearranged as a one-dimensional vector of pixel values. This set of vectors has a set of principal components. Linear combinations of the first 50 or so of these principal component vectors can often yield very good approximations to the image vectors. Of course, each of the Principal Component vectors can be reshaped into a two-dimensional image or template. The Principal Component coefficients for the mugshots can be considered as classification features. Linear combinations of these features can be found which optimally discriminate between target classes and/or clutter. This is analogous to the concepts of eigen- and Fisher-faces [29] in facial recognition. In Fig. 8 we show some results taken from [28]. Here the sonar images for rock, truncated cone and small cylinder classes were simulated for a fixed range using a ray-trace model. The dimensions of these objects were varied randomly within specified limits. Some example images are shown in Fig. 8(a). A portion of the images were used to determine the best discriminating pixel features or templates which are shown in Figs. 8(b) and 8(c). The images from the testing set can then be projected onto those templates resulting in the clustering shown in Fig. 8d. Another image-based approach was used in [30]. Here, the authors used a convolutional restricted Boltzmann machine to “learn” disciminating features and the outputs from the top layer of this machine are then used by a support vector machine for classification.

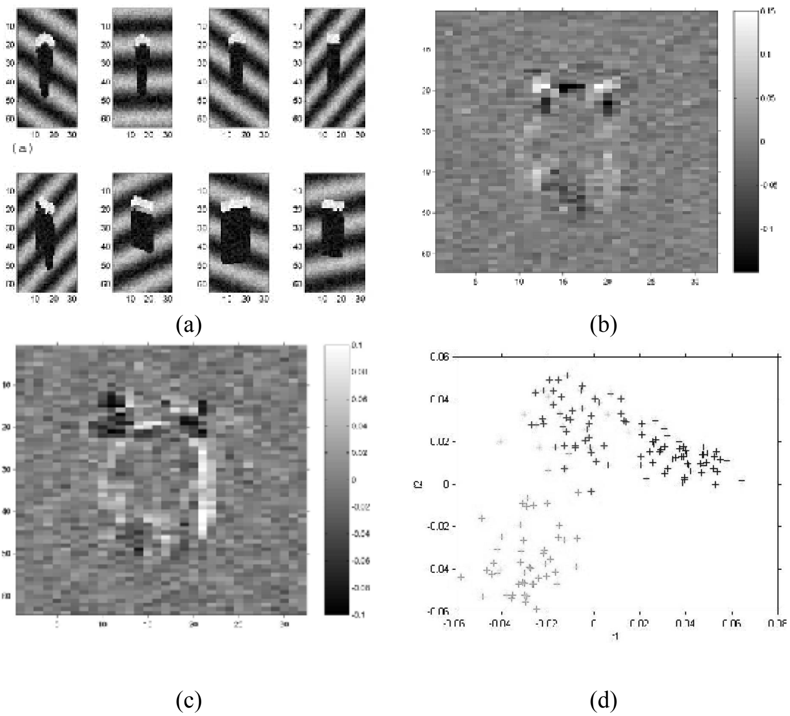

Figure 8: (a) some simulated images – truncated cone, small cylinder, rock (b) first discriminating template (c) second discriminating template (d) resulting clustering of the testing set images. In (a),(b), and (c) the horizontal indices are the across-track pixel indices and the vertical indices are the along-track indices. In (d) the axes are the two discriminating feature values. This figure is taken from [28].

We have considered only a fraction of the available image classification techniques which can be applied to the problem of automated detection and classification of mine-like objects in sidescan sonar imagery. Much of the ATR problem is concerned with rejecting false alarms. In the case that an AUV will revisit a contact at a different aspect(s), either immediately or in a later re-survey, the number of contacts must be reasonable. However, there may be seabed regions, such as boulder fields, where it will always be difficult to reduce the number of false alarms. Another issue with existing ATR methods is that they often rely on *a priori* knowledge of the objects of interest. The more specific this knowledge, the less robust the method may be. One approach which addresses these issues is change detection [31, 32]. In this approach, a very high percentage of the false alarms which would be present on a single survey are effectively eliminated because they are present on a previous survey. It is only the differences between the images which are of interest. In addition, there is no reliance on *a priori* information to find the regions of change. It is important in this approach that the 2 surveys be accurately co-located so that any differences are meaningful. This can be done by estimating relative local translations and rotations between the sonar images from the data sets themselves. Of course, this approach assumes that one is able to carry out repeated sonar surveys of a region.

It may also be possible to improve target/clutter discrimination by considering lower sonar frequencies and wider bandwidths. The ATR approaches described in this paper are based upon the analysis of features extracted from an image. The sidescan and/or SAS frequencies are usually high and any elastic/structural scattering characteristics of a target are not exploited. Man-made objects often have distinctive scattering characteristics which distinguish them from, for example, rocks. The successful use of lower frequency/large bandwidth sonars to detect/classify different types or targets has been shown in [33, 35, 36]. A hybrid system using high-frequency sonar imagery combined with lower-frequency wideband spectral information could be an effective system for lower false alarm rates.

5. Summary

Automatic Target Recognition for sidescan or synthetic aperture sonar imagery is a complex area of scientific research. It combines the disciplines of sonar sensors, image and statistical processing, fusion theory, pattern recognition and more. It plays a fundamental role in expanding the autonomous behaviour of AUVs. Despite the advances in sonar performance and algorithms, there remain fundamental problems in making ATR robust for a wide variety of seabed environments and minimizing the number of false alarms without missing the real objects of interest.

Acknowledgements

I would like to thank my colleagues over the years at DRDC Atlantic for their help and collaboration in the results shown in this paper. I have also benefited from very helpful collaborations with NATO and TTCP colleagues. Some of the data shown in this paper is courtesy of NURC.

References

[1] M. Couillard, J. Fawcett, and M. Davison, “Optimizing constrained search patterns for remote mine-hunting vehicles,” *IEEE Journal of Oceanic Engineering*, vol. 37, pp. 75–84, Jan. 2012.
[2] J. Fawcett, “Automatic target recognition for an autonomous underwater vehicle (AUV) sidescan sonar in complex environments,” in *Proceedings of the Fourth International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results*, (Kos, Greece), pp. 123–130, June 2011.
[3] G. Dobeck, J. Hyland, and L. Smedley, “Automated detection/classification of seamines in sonar imagery,” in *Proceedings of SPIE*, vol. 3079, pp. 90–110, 1997.
[4] R. Kessel, “Using sonar speckle to identify regions of interest and for mine detection,” in *Detection and remediation techniques for mines and minelike targets, Proceedings of SPIE*, vol. 4742, pp. 440–451, 2002.
[5] C. Ciany, W. Zurawski, and G. Dobeck, “Application of fusion algorithms for computer-aided detection and classification of bottom mines to shallow water test data,” in *Detection and remediation techniques for mines and minelike targets, Proceedings of SPIE*, vol. 4742, pp. 412–418, 2002.
[6] V. Myers and Ø. Midtgaard, “Fusion of contacts in synthetic aperture sonar imagery using performance estimates,” in *International Conference on Detection and Classification of Underwater Targets, Proceedings of the Institue of Acoustics*, (Heriot-Watt University, U.K.), Sept. 2007.
[7] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *IEEE CVPR*, 2001.
[8] R. Lienhart and J. Maydt, “An extended set of Haar-like features for rapid object detection,” *IEEE ICIP 2002*, pp. 900–903, 2002.
[9] Open Source Computer Vision OpenCV 2.3, http://opencv.willowgarage.com/wiki/, Access date:(January 2012).
[10] V. Myers, “Image segmentation using iteration and fuzzy logic,” in *Proceedings of CAD/CAC 2001*, (Halifax, Canada), 2001.
[11] 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, pp. 90–105, Jan. 2003.
[12] B. Zerr, J. Fawcett, and D.Hopkin, “Adaptive algorithm for sea mine classification,” in *Proceedings of the Third International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results*, (Nafplion, Greece), pp. 319–326, June 2009.

[13] I. Quidu, J. Malkass, G. Burel, and P. Vilbe, “Mine classification based on raw sonar data: An approach combining fourier descriptors, statistical models and genetic algorithms,” in *Proceeedings of OCEANS 2000 MTS/IEEE Conference*, vol. 1, (Providence, U.S.A.), pp. 285–290, 2000.
[14] 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, pp. 863–876, Oct. 2010.
[15] B. Zerr, B. Stage, and A. Guerrero, “Automatic target classification using multiple sidescan sonar images of different orientations,” *SACLANTCEN SM-309*, 1997.
[16] I. Quidu, “Classification multi-vues d’un objet immergé à partir d’images sonar et de son ombre portée sur le fond,” *Doctoral Thesis, L’Université de Bretagne Occidentale*, 2001.
[17] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
[18] J.Fawcett, M. Couillard, D. Hopkin, A. Crawford, V. Myers, and B. Zerr, “Computer-aided detection and classification of sidescan sonar images for the Citadel trial,” in *International Conference on Detection and Classification of Underwater Targets, Proceedings of the Institue of Acoustics*, (Heriot-Watt University, U.K.), Sept. 2007.
[19] A.Khotanzad and Y. Hong, “Invariant image reconstruction by Zernike moments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 489–497, May 1990.
[20] G. Tao, M. Azimi-Sadjadi, and A. Nevis, “Underwater target indentification using GVF snake and Zernike moments,” in *Proceeedings of OCEANS 2002 MTS/IEEE Conference*, vol. 3, pp. 1535–1541, 2002.
[21] J. Fawcett, “Computer-aided detection and classification of minelike objects using template-based features,” in *Proceeedings of OCEANS 2003 MTS/IEEE Conference*, vol. 3, pp. 1395–1401, 2003.
[22] S.Reed, Y.Petillot, and J.Bell, “Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information,” *IEE Proc-Radar Sonar Navig.*, vol. 151, pp. 48–56, 2004.
[23] V. Myers and D. Williams, “Adaptive multiview target classification in synthetic aperture sonar images using a partially observable markov decision process,” *IEEE Journal of Oceanic Engineering*, pp. 45–55, Jan. 2012.
[24] V. Myers and J. Fawcett, “A template matching procedure for automatic target recognition in synthetic aperture sonar imagery,” *IEEE Signal Processing Letters*, vol. 17, pp. 683–686, 2010.
[25] H. Midelfart, J. Groen, and Ø. Midtgaard, “Template matching methods for object classification in synthetic aperture sonar images,” in *Proceedings of the Third International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results*, (Nafplion, Greece), June 2009.
[26] H. Midelfart and Ø. Midtgaard, “Robust template matching for object classification,” in *Proceedings of the Fourth International Conference and*