Using context to detect underwater objects

Duncan P. Williams [1], Gary L. Davies, Andrew J. Evans
Dstl Physical Sciences, Porton Down, Salisbury, SP4 0JQ UNITED KINGDOM
[1] dpwilliams@dstl.gov.uk

***Abstract*—In the absence of sufficient information about the identity of an object to detect it among other objects there is often extra information in the relationship *between* objects that can be exploited. This contextual information can help us to characterise the natural set of surrounding objects or even separate objects into different classes directly. Context is widely used to understand and characterise scenery. This paper introduces the use of context to detect underwater objects. Context has not previously been studied for complex sonar imagery and our aim is to design and test measures of context to demonstrate how it can help us to detect and/or identify "unauthorised" objects that are not part of the natural set of objects present in the imagery. This is motivated by some of the challenges that are faced in cluttered sonar environments, or when the objects to be detected are hidden or disguised, for which extra information can be decisive.**

I. Introduction

Object identification tasks have traditionally treated objects and their surroundings separately and chosen to isolate the object so that it can be interrogated on its own. In the real-world, objects will typically co-vary with other objects and their surroundings. This implicit relationship between objects can provide *contextual* information that can help us to understand a scene or identify an object. This context is often a generalised, abstract representation, based on properties from the set of objects that are present. Context has already been shown to facilitate the identification of such objects as faces, cars and pedestrians, as well as characterising scenery.

In the following we consider the use of contextual information in complex imagery, such as sonar or radar. Finding even simple objects can be limited by the information that is available about the likely identity or type of object e.g. when there is limited resolution or the properties of the object are not known in advance. It becomes more challenging to compensate for changes in viewing conditions or items such as position, size, and illumination. Our aim is to show that there is an advantage to using context to detect underwater objects in these types of conditions and, in particular, to detect so-called "unauthorised" objects that do not belong to the natural set of objects that could be present such as rocks, ridges and fish.

A. Using context in imagery

In the absence of sufficient information about the identity of an object, the surrounding objects and background and prior knowledge can often provide the additional (contextual) information needed to recognise the objects. There exists a huge body of work on contextual information and its role in object identification, see for instance [1]. Much of this research is associated with the visual system and contextual cuing whereby attention is quickly guided to regions of interest. This has its own applications and is studied for data fusion and for fast categorisation or recognition tasks. Less is known about which features carry the information context and how these features (once selected) should be built into a procedure that exploits context to detect and/or identify objects. Nevertheless, the capability of humans to recognise thousands of objects in cluttered environments, despite changes in position, size, illumination, or occlusions is a compelling starting point for using context in imagery.

Recent studies have shown that context-awareness is identity-related (semantic) and location-related (spatial). For example, one would expect to find a table and chair in the same image, but not a fish and a bicycle; and a cup is expected to be on a table and not vice-versa. Much of this has focussed on a choice between local or global representations of contextual information. In a global representation, low-level features are used to characterise the objects and its surroundings without encoding individual objects. Locally, context can be represented in terms of the relationships between objects or by means of a low-dimensional global description. In the following, we consider such statistical properties as the mean size and variance of the objects and more complex structural information such as the amount of background clutter.

The challenge is in selecting the features: global features can be defined as the (weighted) combination of local features, e.g. based on spatial or multi-scale filters at different locations. Other global representations can be based on histograms of local features such as multi-resolution methods or spatial pyramid matching. Informally, pyramid matching works by placing a sequence of increasingly coarser grids over the feature space and taking a weighted sum of the number of matches that occur at each level of resolution. This has been used, for example, to separate images into different categories [2]) or other segmentation tasks (e.g. [3], [4]). Sparse template matching can also be used to eliminate the sensitivity of global features to weak relationships between objects.

It is worth noting that there can be many different ways to separate categories or classes of objects in an image. For example, separating scenes into city or landscape, i.e. separating man-made scenes from natural scenes like [5] using histograms of local features, or into indoors or outdoors [6].

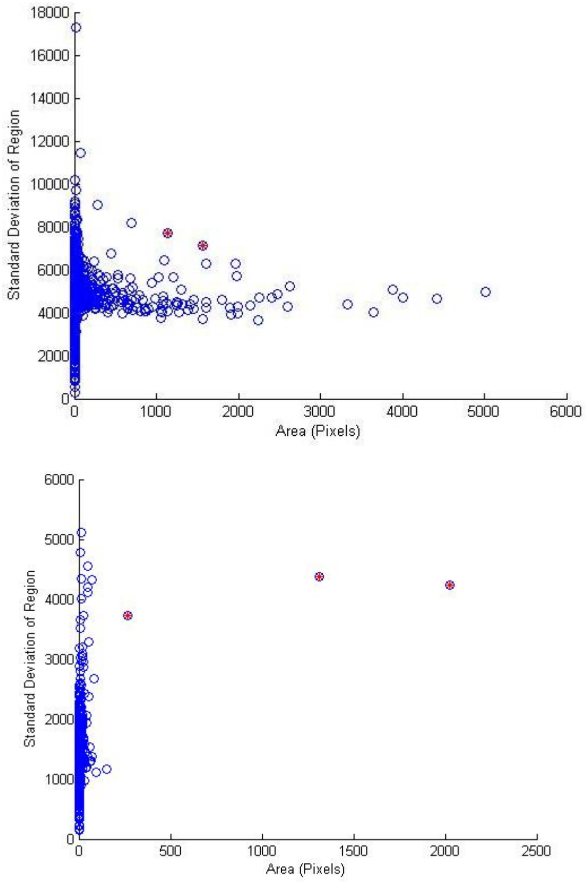

Fig. 1. Comparison between the area and standard deviation of the objects found on a seabed that contains sand ripples (top) and on a flat seabed (bottom). The outliers are marked in red and correspond to the detections in the lower panels of Fig. 3 and Fig. 4.

This, in turn, can favour different combinations of features and means of representing the selected features.

Low-level background information, like clutter or objects scattered throughout an image, could also be exploited; clutter will typically depend on the environment and is already used to characterise the background, such as statistical distributions of the seabed (e.g. [7])), and help to detect underwater objects.

II. Detecting underwater objects

Our aim is to design and test measures of context that can be used to detect, identify and/or localise underwater objects. In particular, we take advantage of high-resolution Synthetic Aperture Sonar (SAS) data, as good as about 25x25 mm/pixel, to demonstrate how information about *all* the objects present in sonar imagery can be used to detect "unauthorised" objects that are not part of the natural set of objects, and other features, that are found on the seabed. To the author's knowledge, the context in complex sonar imagery has not been previously studied, and this preliminary investigation provides a useful addition to the evidence that has been presented elsewhere for other imagery, such as natural scenes e.g. [8]. This is motivated by applications that will be familiar to the reader, such as

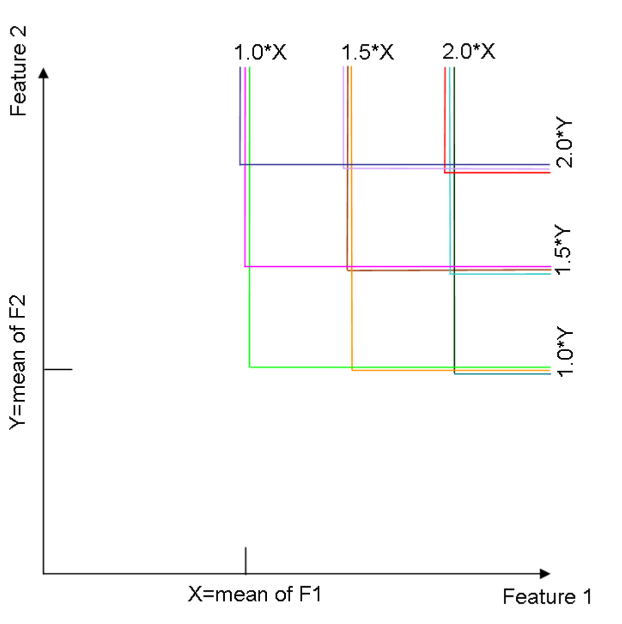

Fig. 2. Example threshold scheme based on multiples of the mean of two different features.

minehunting sonar or the detection of unexploded ordnance. Current schemes to detect underwater objects work well in normal sonar conditions when the object is known *a priori* but are less effective in the presence of clutter and variability or when the "unauthorised" objects are obscured or not already known e.g. when the objects are disguised or improvised.

A. Procedure

SAS data provided by the NATO Undersea Research Centre (NURC) is used in this section to generate a small set of measures that generalise to a wide set of objects on two seabed types: flat and sand ripples.

In the following we have outlined a simple procedure using pairs of statistical features, that describe the physical properties of *all* the objects in images that are approximately 7200 pixels across track and 1200 pixels along track in size. The first step is to convert each image to a binary representation of the objects by applying a threshold level of ten times the mean (pixel intensity) level. In this binary representation, all pixels now have a value of unity (above the threshold) or zero (below the threshold). A morphological closing operation (a smoothing function equivalent to dilation followed by erosion of the image) is used to fill in small holes in objects. Following this operation, each contiguous area of pixels having a value of unity is labelled as an individual object and different measures of each object are selected.

Fig. 1 shows example pairings of measures for two images with different seabed types. In this case, the points corresponding to two objects in the upper panel and three objects in the lower panel have been isolated (marked in red) from all the other points representing the set of objects located on the seabed. In general, features (such as physical measures)

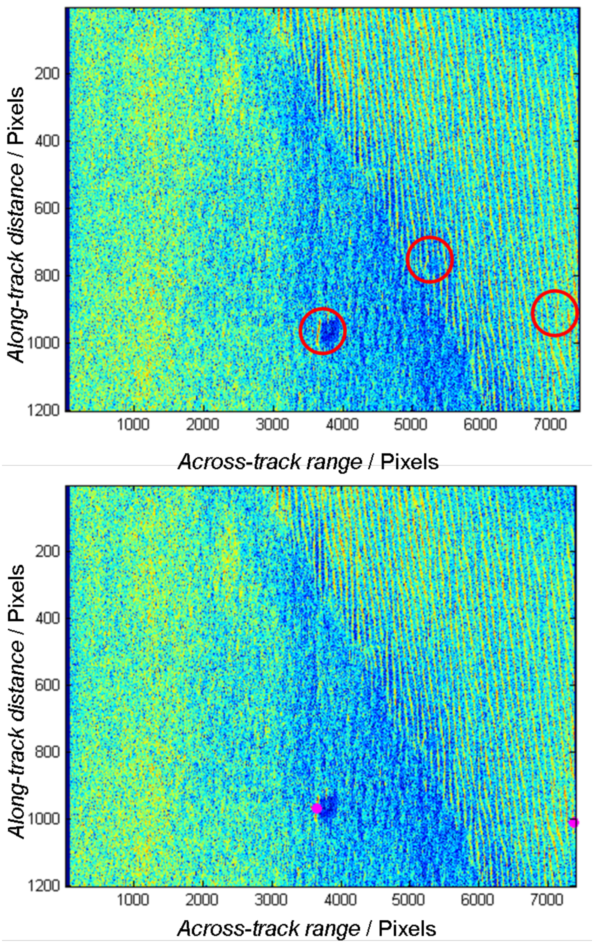

Fig. 3. Example target detections when the seabed contains sand ripples c.f. top panel in Fig. 1. The true positions are circled in the upper panel and the detections are marked in red in the lower panel; the two detections correspond to the two outliers marked in the top panel in Fig. 1.

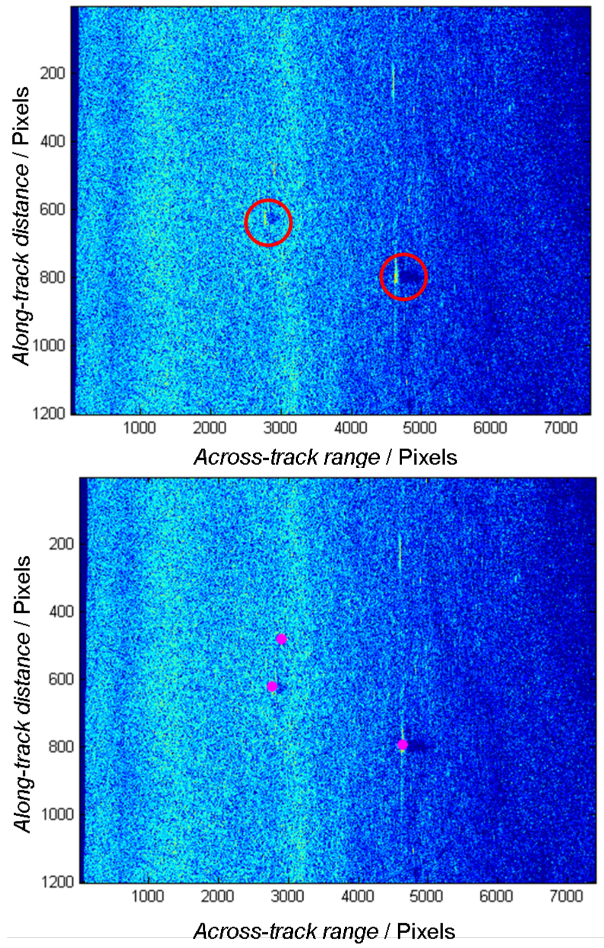

Fig. 4. Example target detections when the seabed is flat c.f. bottom panel in Fig. 1.

are needed that form clusters for different classes of objects that need to be detected. One simple procedure to detect these objects is to set a threshold for each feature based, for example, on multiples of the mean value of each feature; this is illustrated in Fig. 2 for which any objects above and to the right of the threshold line would be declared. Formally, the set of detections is defined as

$$\begin{aligned} D = \{d \mid \psi_1(d) > X\langle\psi_1(a)\rangle \\ \text{and} \quad \psi_2(d) > Y\langle\psi_2(a)\rangle,\, a \in A\}, \end{aligned} \tag{1}$$

for some X, Y, where $\psi_{1,2}$ are one pair of measures and A is the set of closed objects in the binary image; this definition could generalise to any number of measures ψ_j for $j = 1, \ldots, N$.

While this procedure is unlikely to be sufficient for poor sonar conditions, such as textured or clutter backgrounds, it is employed below in order to introduce the concept of using contextual information in sonar imagery - this is the main aim of our paper.

B. Results

Some preliminary results are shown in Figs. 3–4 based on comparisons between the area of each object and the standard deviation of the pixel intensity of each object in examples of each seabed type.

In each case, an object is declared (as "unauthorised") following this simple procedure, whenever the area of the object was greater than X times the mean area of all objects *and* the standard deviation of the object was greater than Y times the mean standard deviation of all objects.

Fig. 3 and Fig. 4 show the results when $X > 1$ and $Y > 1.5$ in (1) for the same two example images considered in the top and bottom panels in Fig. 1 respectively; the true positions of the targets are circled in the upper panel in each of Fig. 3 and Fig. 4 and the detections are marked in red in the lower panels in the same way as Fig. 1. It is worth noting that this procedure detects the two objects on a flat seabed (Fig. 4) but only one of three objects in the presence of sand ripples. Similar results were found for all other images in the data that was available but estimates of the number of positive contacts and false contacts for different values of X, Y have not been calculated.

C. Comments

Traditional methods for detecting the objects in sonar imagery would try to focus on each object in isolation and seek ways to filter and remove the background or other surrounding objects. The simple procedure developed here has effectively shown that the reverse - explicitly looking at the information about all the objects in the imagery - could help to detect underwater objects without needing detailed information about the target. There is much more to be done to select the most relevant measures but context should provide useful information in sonar imagery. The same should be true for other imagery but, in general, one may expect some measures to be needed on a case-by-case basis while other measures could be found that are tolerant to, for example, different resolution data or when the objects contained in the images have different properties.

III. Conclusion

This preliminary investigation of contextual information in sonar imagery has shown that relatively simple physical measures can be used to characterise the relationship between natural objects on the seabed and other "unauthorised" objects. This is unlikely to be effective in detecting objects in poor sonar conditions but could provide useful information to assist detection or identification of man-made objects. The next stage is to develop other measures of context, such as higher order statistical moments, that can separate different classes of objects, and an effective procedure that can exploit such measures.

Acknowledgment

This work has been carried out in the Platform Sciences Group of Dstl Physical Sciences as part of the University Defence Research Centre (UDRC) on Signal Processing. The data used in this work was provided by the NATO Undersea Research Centre.

References

[1] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, "Context-based vision system for place and object recognition," in *Proc. Ninth IEEE International Conference on Computer Vision*, 2003, pp. 273–280.

[2] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2169–2178.

[3] J. Li and R. M. Gray, "Context based multiscale classification of images," in *Proc. International Conference on Image Processing*, 1998, pp. 566–570.

[4] E. G. P. Bovenkamp, J. Dijkstra, J. G. Bosch, and J. H. C. Reiber, "Multi-agent segmentation of IVUS images," *Pattern Recognition*, no. 37, pp. 647–663, 2004.

[5] A. Vailaya, A. Jain, and H. J. Zhang, "On image classification: City images vs. landscapes," *Pattern Recognition*, no. 31, pp. 1921–1935, 1998.

[6] M. Szummer and R. W. Picard, "Indoor-outdoor image classification," in *Proc. IEEE International Workshop on Content-Based Access of Image and Video Databases*, 1998, pp. 42–51.

[7] M. Datum, D. Lerro, and F. McMullen, "Target detection enhancements using in-situ environment adaptive clutter modeling," in *Proc. IEEE Aerospace Conference*, 2001, pp. 1757–1770.

[8] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, no. 42, pp. 145–175, 2001.