

Object Detection at Different Resolution in Archaeological Side-scan Sonar Images

Louis Atallah
The British University in Dubai
University of Edinburgh
PO BOX 502216, Dubai, UAE
Email: latallah@inf.ed.ac.uk

Changjing Shang
School of Informatics
University of Edinburgh
Edinburgh EH8 9LE, UK
Email: shang@inf.ed.ac.uk

Richard Bates
School of Geography & Geosciences
University of St Andrews
St Andrews KY16 9AL, UK
Email: crb@st-andrews.ac.uk

Abstract—Side-scan sonar is now considered ‘the instrument-of-choice’ for underwater archaeological surveys. However, much work is required to understand the factors that may affect the surveyed data, including the effect of sonar resolution on the detection of objects on the seafloor and the potential confusion between the presence of recent objects on the seafloor and that of actual ‘archaeological’ material. To aid in addressing these issues, this paper presents an approach using scale saliency features for object detection in archaeological side-scan sonar images. Experimental results show that the techniques introduced here are capable of determining scales and saliency features for object regions. More importantly, the work is robust with respect to image intensity, scale and contrast. These factors often change in real sonar surveys due to collecting data at different depths and using different sonar parameters.

I. INTRODUCTION

Side-scan sonar is now widely adopted in industry, academic research programmes and pre-development surveys for reconnaissance scale mapping of the seafloor [1]. However, more work is needed to understand a number of important factors that typically affect the surveyed data. One of these is the effect of sonar resolution on the detection of objects on the seafloor. Another is the presence of recent objects, both natural (e.g. rocks) and anthropogenic (e.g. fishing gear, creel pots) on the seafloor which could be mistaken for archaeological material given the low resolution of commercial sonars. Computer-based interpretation, understanding and analysis of such survey data becomes essential due to its sheer volume alone. Furthermore, resolutions of the recorded data are dependent upon sonar specifications and the conditions under which the data were acquired. Therefore, object detection and classification of various types of man-made objects forms a particularly difficult challenge for manual processing, and it is indeed a

difficult task in the underwater acoustic field in general. This task becomes harder when side-scan sonar images are variable in terms of intensity, scale and rotation, and are blurred with noise.

Recently, there has been an increased use of local appearance based features to perform pattern-matching for object detection [3], [4], [8] in camera images. One such approach is to use saliency feature descriptors to support image retrieval and scene analysis, including object localisation and tracking [5], [10]–[12]. This approach has proved to be robust to planar rotation, scaling, intensity shifts and translations. These are common problems in underwater acoustic surveys, where the image is affected by sonar motion and resolution. This work applies the scale-saliency techniques to develop an automatic method for detecting saliency regions which correspond to image regions representing parts of certain objects of interest. The investigation and evaluation are carried out over different parameter settings, such as the range and fish height of the sonar.

Sonar data was acquired for this work from a control experiment carried out in Smelt Mill Bay, Belfast Lough, where a test site of material types was set out on the seafloor and repeat surveys were conducted over the control site using three different side-scan systems. From the control experiment, a large database of sonar images were obtained, and used to test the algorithm proposed here.

In section 2, the sonar images obtained from the control experiment are briefly described. Then the idea of visual saliency, the algorithm for selecting scale, and the method for computing saliency features and identifying salient regions are presented in section 3. Section 4 describes and discusses experimental results by applying the approach to detect objects within the selected sonar images. Finally, the paper is concluded in section 5.

II. DESCRIPTION OF DATASET

The Centre for Maritime Archaeology (University of Ulster), the Archaeological Diving Unit (University of St. Andrews) and Management for Archaeology Ltd. conducted a control experiment in Smelt Mill Bay, Belfast Lough, during July and August 2001. Smelt Mill Bay is located in 10m of water in a sheltered cove with a uniform, fine-grained planer sand substrate. A test site of material types was set out on the seafloor and repeat surveys were conducted over the control site using three different side-scan systems (EdgeTech 272-TD: 100/500 kHz, Imagenex 885; 675 kHz, and Geoacoustics 159-A; 100/500 kHz) run by the respective organisations.

Repeat side-scan survey lines were acquired at increasing grid to assess the resolving capability of each instrument. Targets included diverse materials such as water logged ship-timbers, 'green wood', leather, flint, glass, iron, aluminum piping of varying diameter, skeletal remains, gravel, car-tyres and ceramic-material abundant (sometimes unfortunately) on the seafloor in coastal waters. Note that only part of these materials are studied in the work reported here at this stage, which are listed in table I. In addition to the material type, this table also lists information regarding the size and position of the objects (offset from (0,0), the ship position, along the horizontal direction (x)).

TABLE I
DESCRIPTION OF THE OBJECTS

ID	Object Type	Dimension	Offset(m)
1	0.65m diameter car tyre	Vertical: H=0.60	45
2	0.65m diameter car tyre	Horizontal: H=0.20	46
3	'Amphora' shoulder and neck	H=0.35; Diameter at mouth=0.18m	47
4	Ceramic ball	Diameter=0.26m	48
5	Upturned hemispherical basket	Diameter=0.17m	49
6	Willow 'fish basket'	H=0.38m, L=1.2m	50-51.2
7	Woven grass basket	0.40x0.24x0.16m	51.4-51.8
8	Upstanding leather jacket	H=0.70m; L=1.0m	52.3-53.4

Four subsets of selected original sonar images are used here captured by EdgeTech 272-TD (100/500 kHz) sonar as shown in figure 1. The description of each image captured under different conditions is given in table II. This table details the fish height and the range from the sonar of each image. The size of each image (in pixels) is also given in table II. There are eight objects within each image where reflections from each object typically indicate the objects which are required to be detected and recognised. Those objects are further illustrated and labelled (in terms of their ID in table I) in figure 1.

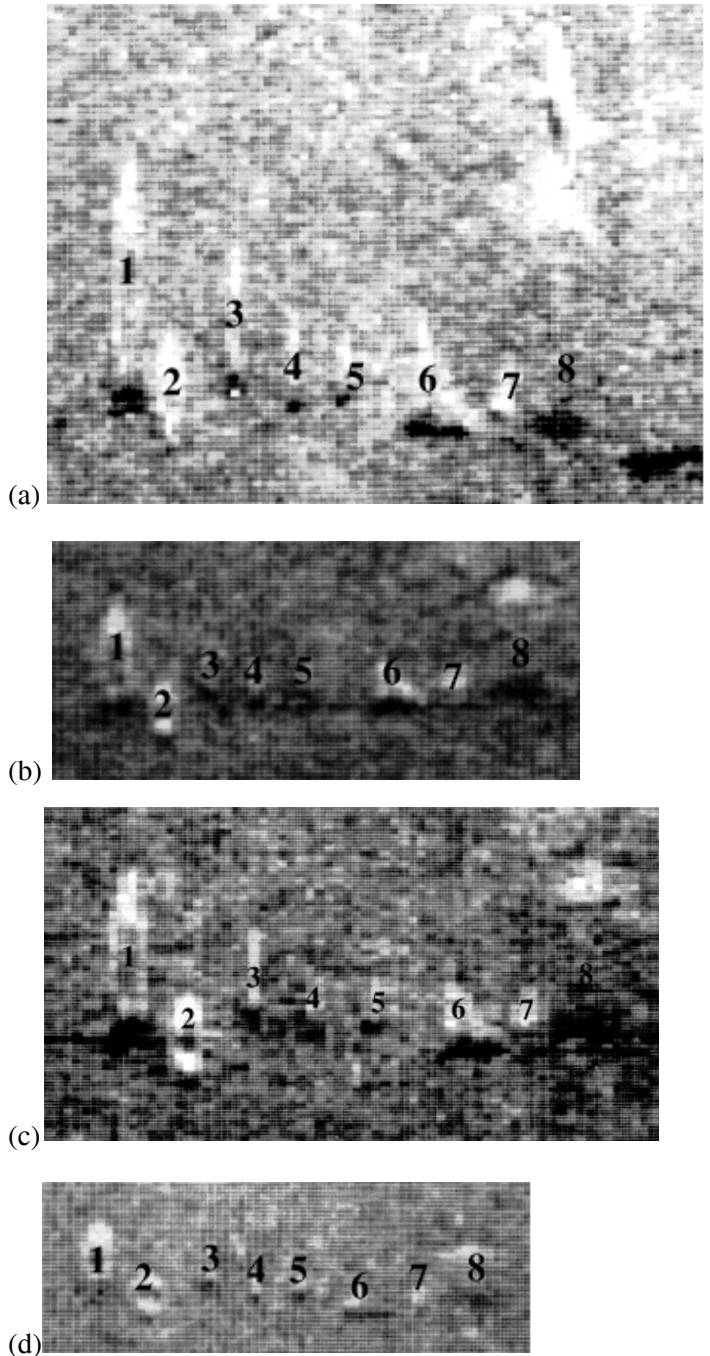


Fig. 1. Sonar images

III. SONAR IMAGE DENOISING

Sonar images are often corrupted by noise during the process of their acquisition and transmission. The quality of the original images shown in figure 1 is non-exceptional. It is affected by various kinds of noise, including the different parameters that are presumed to be fixed (e.g., the range and fish height) during the process of their capturing. Image denoising is therefore

TABLE II
DESCRIPTION OF SONAR IMAGES

Image	Fish Height	Range	Size
Figure 1(a)	2m	25m	1384 × 1054
Figure 1(b)	2m	50m	800 × 362
Figure 1(c)	4m	25m	1338 × 726
Figure 1(d)	8m	50m	766 × 266

necessary to remove the additive noise while retaining as much as possible the important image features.

There exists a fair amount of research in this area with many noise reduction techniques developed for handling noisy (sonar) images [2]. In particular, a spatially pixel-wise Wiener filter [6], [7] is used in this work. The filter is suitable for intensity images that have been degraded by additive noise of a constant power (assumed to be white). It performs 2-D noise reduction without destroying essential details contained within the images. Experiment results have demonstrated that using this filter before image analysis, say applying an algorithm for object detection is able to improve the algorithm accuracy. When using such a filter, care has to be taken in selecting an appropriate size of the local filtering window. Based on a range of experimental simulations, the window size is herein empirically set to be 5×5 in order to maximise the benefit of utilising this filter.

IV. SALIENCY FEATURE DETECTION

A. Saliency and scale saliency

A number of different definitions of saliency exist in the literature, depending on the application domain, image types and on feature descriptors used [8], [12], [13]. For the task of detecting interesting objects within an image, an object is said to be salient if it stands out from its background and all other objects. Intuitively, saliency regions in an image should thus correspond to objects of interest. Within this work saliency is defined in terms of local signal complexity or unpredictability. This is generally how an observer would note a ‘different’ or ‘salient’ object.

The degree of saliency can be computed over different scales. In this work, a scale is defined as the size of the local neighbourhood within which individual pixels are used to estimate the feature descriptors. The scale that yields the maximum saliency degree is typically chosen for use if the corresponding entropy measure reaches a peak over a range of alternative scales. This scale refers to the ‘best size’ of the object selected.

Essentially, the method searches for a scale with high entropy, assuming that it is isotropic. The scale-saliency algorithm is outlined next.

B. Algorithm for scale selection and saliency feature calculation

The following summarises the algorithm for selecting scale and computing saliency features. Further details can be found in [10].

Given a point \mathbf{x} , the algorithm starts by calculating the local Shannon entropy $\mathcal{H}_D(s, \mathbf{x})$ over a range of scales s varying between s_{min} and s_{max} . To calculate the entropy, a local feature descriptor is defined around point \mathbf{x} (in this work, this is selected as the gray level amplitude $d \in D$, where D is the set of all descriptor values). The PDF (probability density function) of the local descriptor (amplitude) $p(d/s, \mathbf{x})$ is calculated over a set of scales s (circular windows, where the radius corresponds to ‘scale’). The Shannon entropy of point \mathbf{x} and scale s is defined as:

$$\mathcal{H}_D(s, \mathbf{x}) = - \sum_{d \in D} p(d/s, \mathbf{x}) \log_2 p(d/s, \mathbf{x}) \quad (1)$$

Areas corresponding to high signal complexity tend to have higher entropy [5], [10]. Thus the next step is to select scales s_p where the entropy is peaked. However, there could be several peaks in the entropy function corresponding to different features at different scales. To select the most ‘salient’ one, statistics of the local descriptor are used to ‘weigh’ the entropy function $\mathcal{H}_D(s, \mathbf{x})$ for each of these peaks s_p .

A measure that can be used is the magnitude change of the PDF as a function of the scale on the peak point:

$$\mathcal{W}_D(s_p, \mathbf{x}) = \frac{s_p^2}{2s_p - 1} \sum_{d \in D} |p(d/s_p, \mathbf{x}) - p(d/(s_p - 1), \mathbf{x})| \quad (2)$$

The final saliency $\mathcal{Y}_D(s_p, \mathbf{x})$ is obtained as the product of $\mathcal{H}_D(s_p, \mathbf{x})$ and $\mathcal{W}_D(s_p, \mathbf{x})$ (the weighted Shannon entropy):

$$\mathcal{Y}_D(s_p, \mathbf{x}) = \mathcal{H}_D(s_p, \mathbf{x}) \times \mathcal{W}_D(s_p, \mathbf{x}) \quad (3)$$

The above equations are used to determine the scale s_p and the final saliency feature $\mathcal{Y}_D(s_p, \mathbf{x})$ of a given pixel in position \mathbf{x} , resulting in a R^3 space (expressing the 2D position and scale of individual pixels) sparsely populated with scalar saliency values. The final measure of ‘scale-saliency’ is a useful indicator for certain significant features in an image at a large range of scales.

For each pixel within a given (real sonar) image, its corresponding salient feature and scale can be obtained as described in the previous sub-section. It is, however, more interesting to extend this salient feature measurement to identify regions of saliency, corresponding to objects within the images. For this purpose, it is necessary to group together those salient pixels which possess similar salient features and scales (and in general, all the relevant neighbouring pixels). To meet this requirement, a clustering algorithm has been developed in the literature [10], which is intended for use in the aforementioned R^3 spaces. The outline of this algorithm is listed as follows (see [9] for further details):

- 1) Apply a global threshold T , to restrain the computation complexity of the clustering algorithm. (It is used to remove the least saliency features whose values are smaller than T . It also affects the performance of the object detection as discussed in section 5B.)
- 2) Choose the highest salient point in saliency-space \mathcal{Y} .
- 3) Find the K nearest neighbours (of the highest salient point found in step 2).
- 4) Test the support of the K neighbours using the variance of their centre point.
- 5) Find the distance D in R^3 between the support (calculated in the last step) and each salient region already clustered.
- 6) Accept, if $D > scale_{mean}$, with $scale_{mean}$ being the mean of the scales in the K nearest neighbours of the region, and if sufficiently clustered (i.e. variance becomes less than a pre-set value).
- 7) Store the mean scale and mean spatial location of K points.
- 8) Repeat from step 2 with next highest salient point.

The above cluster algorithm groups together similar neighbouring pixels to form individual salient regions, with each region being assigned a saliency scale to the mean of the scales of those pixels within the region. The resulting salient regions represent parts of the objects to be detected. Consequently, this method is potentially very useful for object detection and is adopted in the present investigation. As indicated in [9], [10], choosing salient regions this way rather than using single salient points also helps reduce the noise which would otherwise significantly affect the shape of the local entropy, hence the final salient features and the positions of detected objects.

This section provides the experimental results by applying the approach described above to a set of real sonar images as illustrated in figure 1. The algorithm is expected to help detect the objects of interest (which are marked in figure 1(c)).

A. Sonar image pre-processing

From the given (real sonar) images in figure 1, it can be seen that the objects contained within them vary in terms of their size and intensity. This reflects the fact that different parameter settings were employed in capturing these images. To reduce the effect of such variations, the sonar geometry parameters are used to correct the images for scaling effects between them. Depth and range are used to re-scale the images to be of an ‘approximately’ similar scale. This allows better comparison between images once the objects are detected, and also reduce computation complexity. In particular, images shown in figures 1(a) and 1(c) are rescaled to 20% of the original size and those of 1(b) and 1(d) to 50%.

The Wiener filter is then applied to each image with the size of filtering window being 5×5 , to reduce the effect of noise.

B. Effects of thresholding in clustering

As indicated in section 4C, the selection of the value of the salient threshold T in the clustering algorithm affects the accuracy of object detection. To reflect this, figure 2 shows the results of detecting objects in the image of figure 1(c) from using different values of T , where the patches of the red circles represent parts of detected objects. Within the simulation, the range of scales s is set to vary between 3-15 (scale refers to the radius of the circle used to calculate the local scale-saliency at a given point). The variance of the clusters and the number of nearest neighbours in a cluster are fixed to 10 and 5 respectively, when applying the clustering method. Those pre-defined parameters are also used in the simulation provided in the next section. To facilitate comparison between results, all images with detected objects shown are presented in the same width in figure 2 and figure 3.

It can be seen from these results that for a small value of T (e.g. $T = 3.5$ as shown in figure 2(a)), there are too many patches, some of them are actually not parts of any object although all object are detected and fully covered (by red circles in the figure). This is caused by retaining some “saliency” features which are not significant enough to reflect actual objects. Selecting

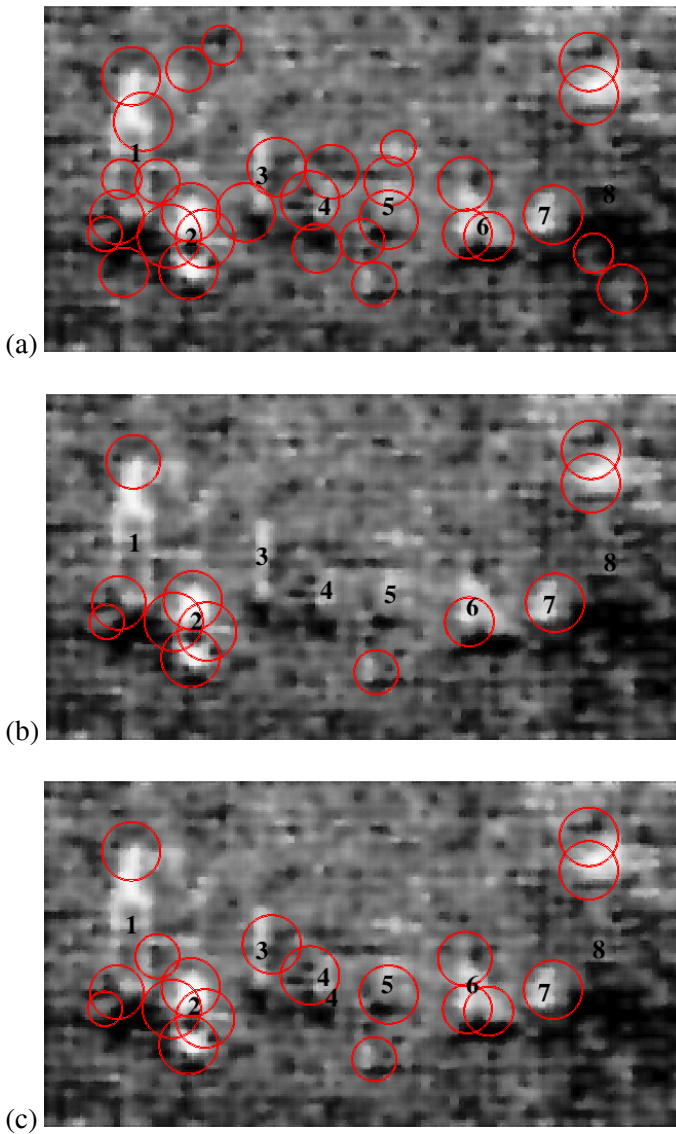


Fig. 2. Detected objects using different salient thresholds (a) $T = 3.5$, (b) $T = 4.9$, (c) $T = 4.5$.

a value of T that is too large (e.g. $T = 4.9$, as presented in figure 2(b)) may result in insufficient coverage of interesting saliency features and hence losing objects of interest (e.g. objects 3, 4 and 5). Clearly, a trade-off is required to balance the choice of an appropriate T value. This will have to be determined by a labelled training dataset for real applications, where the value of T is selected to be the one that leads to the lowest number of misclassifications.

An experimental value of $T = 4.5$ is selected for this work. The results are shown in figure 2(c). With this threshold, the saliency feature detector has been able to

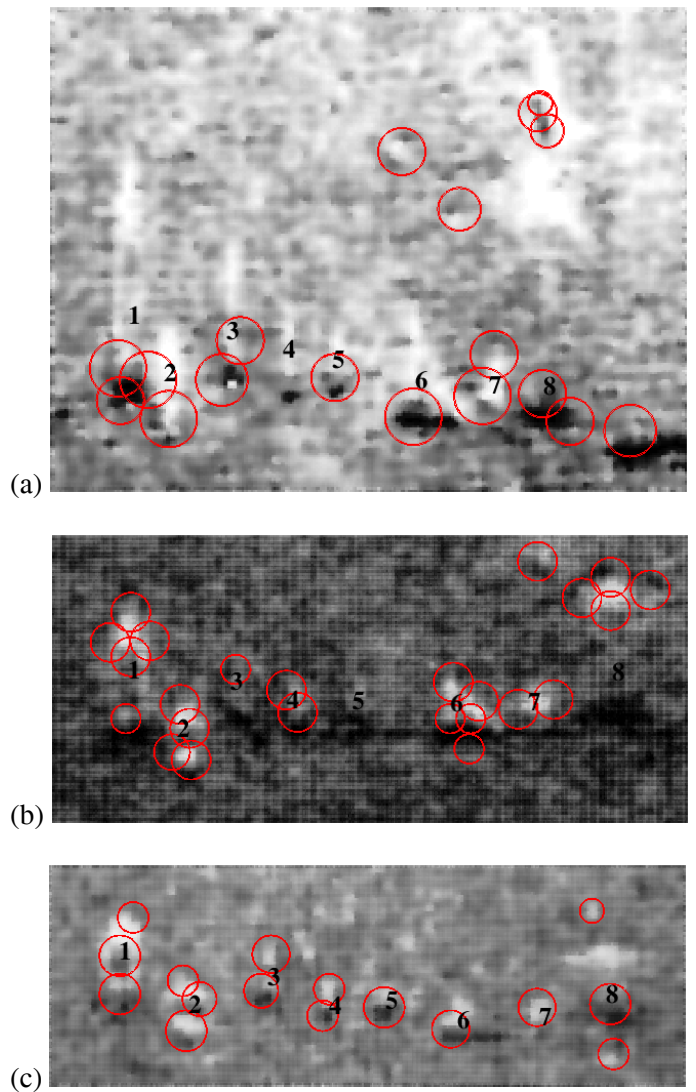


Fig. 3. Detected objects

successfully detect all object parts of interest, in terms of their locations (centers of red circle) and scale (size of radius, see table III for details).

TABLE III
RESULTS OF OBJECT DETECTION

Sonar Image	Fig. 3(a)	Fig. 3(b)	Fig. 3(c)	Fig. 2(c)
Min. Scale	5	9	7	7
Max. Scale	12	12	12	12
Mean Scale	10	11	10	11
No. Miss. Obj	0	1	0	0
No. Inc. Parts	3	1	2	2
Total No. Parts	18	24	16	18

C. Outcomes of object detection

Figure 3 presents the results of object detection for the three sonar images provided in figures 1(a), (b) and (d), with the outcome of detecting objects within figure 1(c) already shown in the last sub-section. From these, table III lists the minimum, maximum and mean scales obtained for comparison. It also gives the number of objects that are missed by the detection algorithm (*No. Miss. Obj*), the incorrectly detected object parts (*No. Inc. Parts*), and the total number of detected object parts (*Total No. Parts*) as judged by visual inspection.

These experimental results are very encouraging: The parts of seven out of eight objects placed on the seafloor in image 1 (b) and all objects in images 1(a), (c) and (d) are successfully detected. It is worth particular mentioning that with increasing ranges, the smaller objects were not imaged, whilst the larger objects showed up as amorphous lower resolution masses. Yet, the present object detection algorithm is still able to perform under such circumstances. For instance, even for human eyes it is difficult to recognise some of the objects from the sonar images when the range is far (from a range of 50 meters), objects 4 and 5 in figure 2(c) and objects 3 and 4 in figure 3(b) are nevertheless correctly detected by the algorithm. This is particularly surprising given that images of 3(b) and (c) have very low contrast.

D. Robustness of the detection algorithm

It is clear that the original images vary in terms of their intensity, contrast, scale and also in terms of the amount of noise contained. However, from the resultant images after object detection (figures 3(a),(b),(c) and 2(c)), it is shown that the scale saliency feature detection algorithm does not seem to be very sensitive to such variations. This demonstrates its robustness which is obviously very desirable in object detection applications. Note that this algorithm is also robust with regard to image rotation as demonstrated in [9], which could be an attractive feature if it were required to compare objects viewed from different locations.

VI. CONCLUSION

This paper presented a method to detect objects in (archaeological) side-scan sonar images by using scale saliency features. In addition to locating objects of interest, simulation results proved that the algorithm is robust with respect to changes in image intensity, scale and contrast, which often vary in real surveys.

Experimental results have so far been very encouraging, setting an initial foundation for carrying out further

investigations into detecting various types of objects placed on the seabed under different operating conditions (or parameter settings) of the sonar. Future research will include classifying objects seen in different surveys, the separation of significant ‘archaeological’ objects from other types and developing methods of ‘learning’ object types in a certain survey. This work could be of particular interest for maritime archaeological surveys in an attempt to locate shipwreck-sites, and to assess their condition in relation to the natural environment.

REFERENCES

- [1] V. Chandran, S. Elgar and A. Nguyen, Detection of Mines in Acoustic Images Using Higher Order Spectral Features. *IEEE Journals of Oceanic Engineering*, vol. 27, no. 3, pp. 610-618, 2002.
- [2] P. Cervenka and D. de Moustier, Sidescan Sonar Image Processing Techniques. *IEEE Journals of Oceanic Engineering*, vol. 18, pp. 108-122, 1993.
- [3] R. Deriche and T. Giraudon, A Computational Approach to Edge Detection, *International Journal of Computer Vision*, vol. 10, no. 2, pp. 101-124, 1993.
- [4] R. Fergus, P. Perona and A. Zisserman, Object Class Recognition by Unsupervised Scale-Invariant Learning. *Proceedings of International Conference in Computer Vision*, 2003.
- [5] S. Gilles, *Robust Description and Matching of Images*. PhD thesis, the University of Oxford, 1998.
- [6] R. C. Gonzalez, *Digital Image Processing*. Addison-Wesley. 1992.
- [7] W. K. Pratt, *Digital Image Processing*. Addison-Wesley. 2001.
- [8] L. Itti, C. Koch and E. Niebur, A Model of Saliency-based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, 1998.
- [9] T. Kadir and M. Brady, Scale, Saliency and Image Description. *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83-105, 2001.
- [10] T. Kadir, *IEEEhowto:Scale, Saliency and Scene Description*. PhD thesis, the University of Oxford, 2002.
- [11] T. Lindeberg, Detecting Salient Blob-like Image Structures and Their Scales with a Scale-space Primal Sketch: A Method for Focus-of-attention. *International Journal of Computer Vision*, vol. 11, no. 3, pp. 283-318, 1993.
- [12] D. G. Lowe, Object Recognition from Local Scale-invariant Features. *Proceedings of International Conference on Computer Vision*, pp. 1150-1157, 1999.
- [13] C. Schmid, R. Mohr and C. Bauckhage, Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, vol. 37, pp. 151-172, 2000.