Detecting man-made objects in unconstrained subsea videos

Adriana Olmos and Emanuele Trucco
Department of Computer and Electrical Engineering
Heriot Watt University
Riccarton EH14 4AS Edinburgh, Scotland, UK
{ceeao, mtc}@cee.hw.ac.uk

Abstract

We present a system detecting the presence of unconstrained man-made objects in unconstrained subsea videos. Classification is based on contours, which are reasonably stable features in underwater imagery. First, the system determines automatically an optimal scale for contour extraction by optimising a quality metric. Second, a two-feature Bayesian classifier determines whether the image contains man-made objects. The features used capture general properties of man-made structures using measures inspired by perceptual organisation. The system classified correctly approximately 85% of 1390 test images from five different underwater videos, in spite of the varying image contents, poor quality and generality of the classification task.

1 Introduction

Motivation. Video cameras are standard equipment for underwater remotely operated and autonomous unmanned vehicles (ROVs, AUVs) used in scientific and industrial missions at sea, many of which aim at or rely on the detection of man-made objects. This often tedious task is typically carried out manually. In scientific missions, for example, video data is recorded nearly non-stop by the ROV cameras, so that a single day of operation can generate many hours of video. Most of this material is uninteresting as it contains no targets whatsoever, and editing the videos manually is an unwelcome but currently necessary job for the scientist in the aftermath of missions. In another example scenario, AUVs or ROVs must locate benthic installations for inspection or intervention purposes. Approaching the target's area can be accomplished by geo-referenced navigation systems, but close-range confirmation of the presence of the target, as well as control of the final approach, depend on video data. For ROVs, these are constantly monitored by the vehicle's pilot, but AUVs need autonomous video detection capabilities.

There is therefore a need for automatic vision systems capable of detecting a variety of man-made objects, in a variety of underwater imaging conditions (natural or artificial illumination, target nature, turbidity, disturbances, and so on). In some applications, in particular, a good but not extremely high rate of correct detections, say 75% or better, is acceptable as long as misses (false negatives) are practically zero. For example, a false negative in mine-like object detection (ignoring a real mine-like object) is much more dangerous than a false alarm; in post-mission tape analysis, scientists would prefer

BMVC 2002 doi:10.5244/C.16.50

a selection of sequences including some uninteresting objects than one missing several, potential interesting targets.

The system proposed. We present a system detecting the presence of unconstrained man-made objects in unconstrained subsea videos. Classification is based on contours, which are reasonably stable features in underwater imagery. First, the system determines automatically an optimal scale for contour extraction by optimising a quality metric. Second, a two-feature Bayesian classifier determines whether the image contains man-made objects. The features used capture general properties of man-made structures using measures inspired by perceptual organisation. The system classified correctly approximately 85% of 1390 test images from five different underwater videos, in spite of the varying image contents, poor quality and generality of the classification task.

Related work. Much research has been devoted to the detection and recognition of *specific* objects in natural environments [17, 14, 3]. Natural scene analysis is not an easy problem at all; natural objects viewed under realistic conditions do not have uniform shapes which can be matched against stored prototypes. According to [1] the strategies developed to approach this problem can be divided in three groups: bottom-up, top-down or hybrid. *Top-down* approaches start with the hypothesis that an image contains a particular object or can be categorised as a particular type of scene. This kind of systems carry out specific tasks: a typical system like this applies specialised segmentation and recognition methods to each object [17, 3, 8]. These systems are broadly used when specific kind of object must to be found (lane detection, building detection, traffic and sign recognition, military targets), and not when the whole scene must be segmented and classified. In a *bottom-up* system a scene is partitioned into regions by using general purpose segmentation techniques. These regions are then characterised by linking the objects to each other. The consequent labelling process requires an inference engine to match each region to the best object model [10, 12]. The *hybrid* approaches are generally characterised by the use of general segmentation methods and specific procedures for labelling purposes, reaping most of the advantages of top-down, bottom-up or variants of the mentioned approaches [20, 6]. In subsea image processing, most of the work reported focuses on the detection of specific targets in sonar images [19, 4] and, to a smaller extent, on seafloor pipes in video images [5, 21]. In all these cases targets are well-defined *a priori* and detailed models can be formulated.

Detecting the presence of *generic* man-made objects in a natural, unconstrained environment is a more difficult and, to our best knowledge, open problem, exacerbated underwater by frequent clutter in the seafloor background, occlusions (e.g., only part of the target may be imaged), and the generally low image quality caused by backscattering, absorption and attenuation of the light [9]. The key limitation is that no specific models can be formulated, as no specific targets are known. One must therefore turn to general properties of the images of man-made objects, at the price of a reduced selectivity and therefore accuracy of recognition (obviously, specific models support more accurate results). Selectivity can however be improved by *perceptual grouping* [11, 7, 14], which can assist with the selection of candidate areas most likely to contain targets. These ideas are at the basis of our work.

Paper structure.
This paper is organised as follows. Section 2 introduces the main design choices and architecture adopted. Sections 3.1 and 3.2 describe, respectively, the optimal scale selection for contour extraction and the contour-based classification algorithms. Section 4 reports

the results of our tests with footage from five different videos acquired in ocean waters and in a laboratory tank, in a variety of imaging conditions. Section 5 summarises and discusses our work.

2 Design choices

The key design issues concern the choice of suitable image features, the choice of general, detectable properties of man-made objects, and the selection of the optimal spatial scale for analysis.

Image features. What features can support the general classification we aim to achieve? Underwater image formation [9, 13] makes texture and colour unsuitable as both disappear rapidly with stand-off distance. Contours, however, are reasonably preserved over a useful range of distances and imaging conditions. This choice is adopted also in other work on detecting man-made objects [7, 14].

Properties of man-made contours. What general properties can be expected of man-made object contours subsea? We use *contour length*, as long contours suggest the presence of a man-made structure; subsea objects tend to generate a multitude of short contours; *contour number*, as many short contours occur usually in natural images, whereas few, long contours in man-made environments; *contour curvature*, as underwater man-made objects contain mostly smooth, simple shapes, generating smooth curves and simple lines in images, whereas natural scenes abound in tortuous and jagged contours.

It is important to stress that these criteria do not constitute a specific, highly selective model, but a set of plausible guidelines for the detection of *general* man-made artifacts underwater.

Optimal spatial scale. The spatial scale at which the above criteria can be applied must be chosen with care. Excessively coarse scales may miss short but suggestive contours; excessively fine scales generate strong clutter, camouflaging target contours. In addition, submerged structures may be covered by algae or similar marine growth, which may camouflage the structure itself to various degrees. Instead of looking at contour length, we select an optimal scale by maximising an *image quality criterion*, in order to ensure that reliable contours are detected.

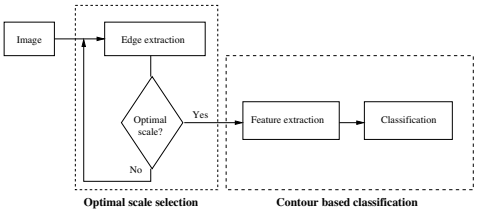


Figure 1: The two-module system architecture.

3 System description

The system is organised in two sequential modules (Figure 1): *optimal scale selection for contour extraction* and *contour-based image classification*. The former selects the scale at which contours must be extracted by optimising a uniformity metric (explained below) over the scale-space generated by the input image. The latter applies a two-feature Bayesian classifier to the contours at the selected scale in order to determine whether man-made objects are present. The two modules are detailed in the following sections.

3.1 Optimal scale estimation for contour extraction

Following the approach of [18, 16] we define the optimal scale as the one maximising a *contour localisation criterion* based on the uniformity of the intensities on both sides of the contour. The idea is that high uniformity on both sides is an indicator (not the only one, of course) of good localisation. Contrast is not considered in order to capture weak and strong edges alike. The uniformity criterion for a contour, U, is the average of the uniformities on both sides of the contour, U_l and U_r (subscripts for left and right), which are computed as shown below in regions obtained by sliding 3×3 windows along the contour:

$$U = \frac{1}{2}(U_l + U_r), \tag{1}$$

$$U_r = 1 - \frac{\sigma_r^2}{N}, \tag{2}$$

where σ_r is the standard deviation of the intensities in R, and

$$N = \frac{(I_{max} - I_{min})^2}{2}. \tag{3}$$

U_l is computed analogously. U varies from 0 (badly localised contour) to 1 (perfectly localised boundary). (The reason of choosing a 3×3 window for computing the uniformity criterion is because a bigger window would miss small contours and influence the selection for a larger scale). Edges are extracted using a Canny edge detector [2].

3.2 Classification

Once an optimal scale has been selected for contour extraction, contour-based features must be used to classify the original image as containing man-made objects or not. We achieved good performance with a standard Bayesian classifier (quadratic decision surface). We set the prior probabilities of the two classes to 0.1 for man-made and 0.9 for natural objects. These values are based on realistic, common expectations in real AUV missions. We noticed also that the sensitivity of the classification results to variations of these priors is limited, although we do not have performed a quantitative analysis yet.

It is difficult to capture good features supporting such a general classification task. The classifier uses two features, M_L and M_I, respectively defined in terms of *contour length* and *contour interest*. In particular, M_L is the average (normalised) contour length:

$$M_L = \frac{1}{N_{\Gamma_I}} \sum_{i=1}^{N_{\Gamma_I}} \frac{L(\Gamma_i)}{L_{\Gamma_{max}}}, \tag{4}$$

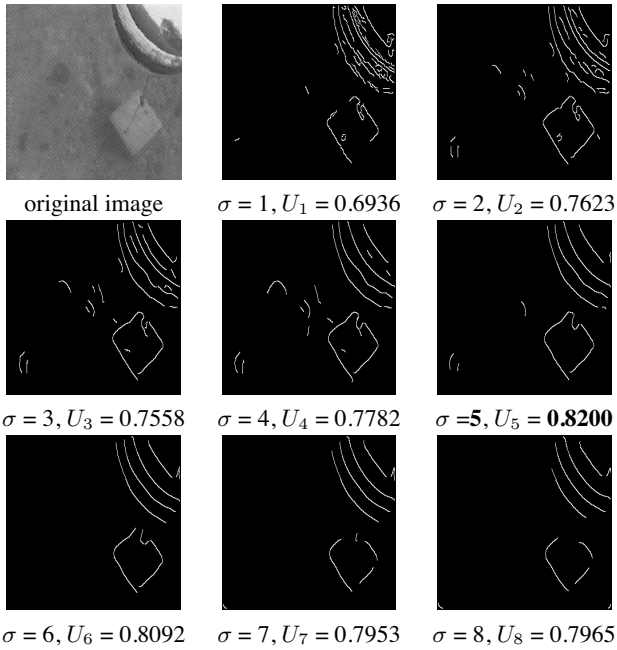


Figure 2: The result of edge detection at different scales computing the localisation measure. The scale selected is = 5. The edge-map selected and the scale are showed with bold letters.

where N_{Γ_I} is the total number of image contours found, $L_(\Gamma_i)$ is the length of the i-th contour, and $L_{\Gamma_{max}}$ is the length of the longest contour found. This feature tends to select images containing uniformly long contours.

The feature M_I is the average interest of the contours in a set composed of the N_{Γ_J} longest contours in the image (in our experiments N_{Γ_J} is typically 5):

$$M_I = \frac{1}{N_{\Gamma_J}} \sum_{j=1}^{N_{\Gamma_J}} I(\Gamma_j), \tag{5}$$

where N_{Γ_J} is the number of contours of interest analysed and $I(\Gamma_j)$ is the interest of the j-th contour. $I(\Gamma_j)$ is defined as the product of two interest measures: $I_r(\Gamma_j)$, based on *contour regularity*, and $I_p(\Gamma_j)$, based on *contour proximity*.

$$I(\Gamma_j) = I_r(\Gamma_j) \times I_p(\Gamma_j) \tag{6}$$

Interest due to regularity[1]. A regular contour is smooth and geometrically simple, as contours of subsea man-made structures often are. We capture this idea of regularity with two measures. First, the *number of curvature zero crossings*: for a fixed length, a small number of curvature zero crossings along the contour suggests a simple, regular contour; a high number, a tortuous, intricate, frequently turning contour. Second, *persistence in scale space*: scanning the curvature scale space [15] of a given set of contours from fine

[1]Notice that we use *regular* and *smooth* only suggestively, not in their precise geometric meaning.

to coarse scales, simple contours will achieve a small number of zero crossings rapidly (low persistence), complex contours later (high persistence). Combining these two ideas, we define the interest due to regularity, $I_r(\Gamma_j)$, for the j-th contour as follows:

$$I_r(\Gamma_j) = \exp^{-\frac{Zc \times S_n}{L}} . \tag{7}$$

where S_n its the maximum number of scales used in the computation of the curvature scale space, the scale is selected automatically when the number of zero crossings repeats twice before converging into a convex contour and the zero crossings disappear, Z_c is the number of curvature zero crossings of the contour Γ_j at the curvature scale selected, and L is the length of the segment.

Interest due to proximity. Some underwater natural events, such as sand ripples, generate long, simple image contours. If the classifier were based only on regularity only, such events would trigger false alarms. We add therefore a *proximity measure*. The idea is that contours of man-made structures, in addition to being long and simple over a range of spatial scales, would also appear close together in an image region, whereas natural events like sand ripples tend to be distributed over the whole image. Therefore the interest due to proximity of the j-th contour, $I_p(\Gamma_j)$, decreases with the distance from any neighbouring contours, and is penalised by the normalised distance from contours of interest:

$$I_p(\Gamma_j) = \prod_{n=1}^{N} \exp^{-\frac{d_n}{L_{max}}} - \sum_{k=1}^{K} \frac{d_k}{L_{max}} \tag{8}$$

where N is the number of the contour neighbours of Γ_j; d_n is the distance from the contour of interest Γ_j to the nth neighbour; L_{max} is the length of the image diagonal; K is the number of neighbouring contours of interest; d_k is the distance between Γ_j and the k-th neighbouring contour of interest.

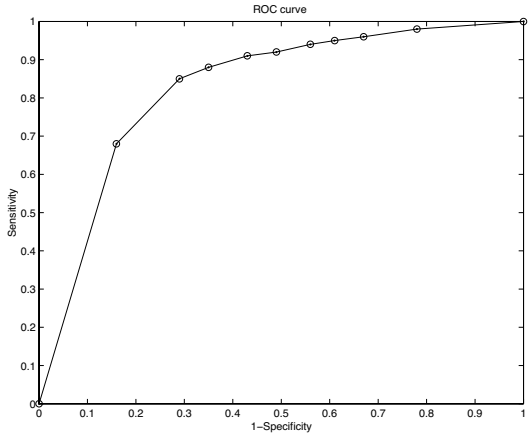
Distances between contours are computed as distances between contours' centroids. Two contours are neighbours if they are adjacent in a top to bottom, left to right scanning order.

4 Experimental results

Five real videos were used in our experiments, two for training and three for testing. The training videos are records of shallow-water AUV operations run by Florida Atlantic University (FAU), covering a variety of conditions. We selected manually 322 frames from the training videos, as representative cases, of which 161 containing man-made objects and 161 natural subsea objects only. The test set was taken from one video acquired in our instrumented tank and two further AUV videos supplied by FAU, taken under turbid water conditions in Fort-Lauderdale, Florida and Camp Pendleton, San Diego. The testing set in total consisted of 1390 frames. All images were pre-processed using our own physics-based restoration algorithm. The system is implemented in Matlab.

In summary, 85% of the test images containing man-made objects and 71% of those containing only seafloor and natural objects were classified correctly. These percentages must be considered in the light of the generally poor image quality of subsea imagery and the generality of the classification task.

Figure 4 shows, for a number of examples, the input frame, the edge map at the selected optimal scale, the same superimposed with the input frame, and the contours of interest selected. Our particular combination of interest criteria proves effective: notice for instance the correct classification of natural scenes in spite of long contours (case B and D), of a man-made target in spite of very poor image quality (case C) or small size (case E). Case F is an example of wrong result, probably due to the scattering of interesting segments across the image.



ROC curve from Test Videos

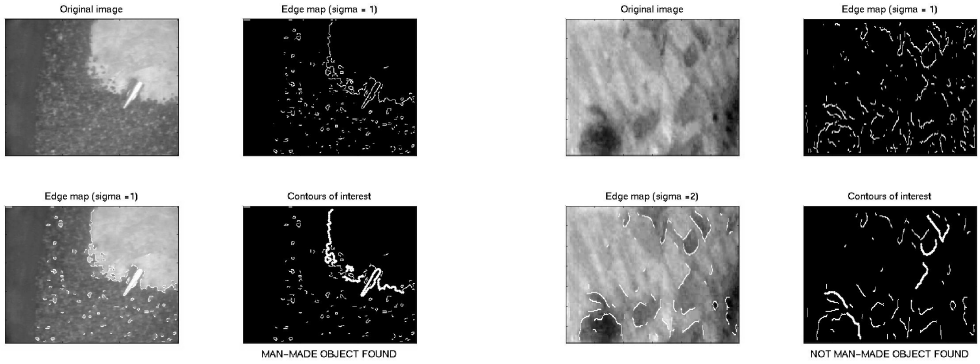
Figure 3: Results from detection of man-made objects in the testing videos.

Figure 3 presents the classification results with different classification-thresholds as a *Receiver Operating Characteristic* curve (ROC), a compact visualisation of the trade-off between the portions of false positives (1-specificity) and of true positives (sensitivity). The *sensitivity*, S_e, is the proportion of correctly classified frames containing man-made targets; the *specificity*, S_p, is the proportion of correctly classified frames containing non-man-made objects:

$$S_e = \frac{TP}{TP + FN} \tag{9}$$

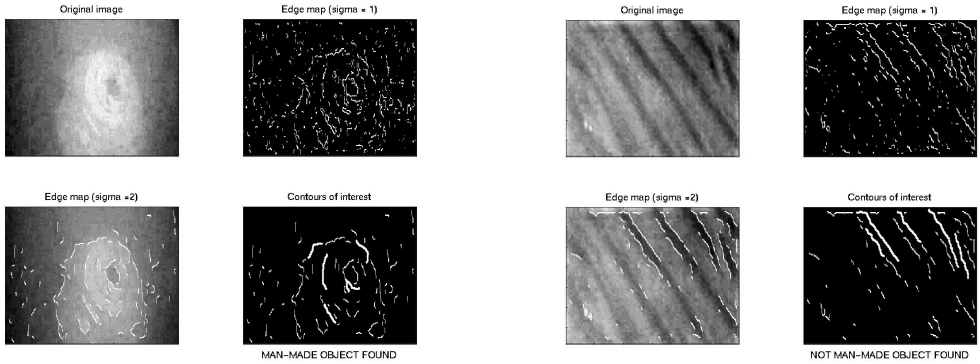
$$S_p = \frac{TN}{TN + FP} \tag{10}$$

TP is the number of true positives, or correctly classified frames containing man-made targets, and FP of false positives, or frames incorrectly classified as containing man-made objects. Similarly, FN is the number of false negatives, and TN the number of true negatives. The closer the ROC curve follows the left-hand border and then the top border of the ROC space, the more accurate the classification. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the classification. In these terms an accurate classification of man-made objects is observed in figure (3) for the test videos.



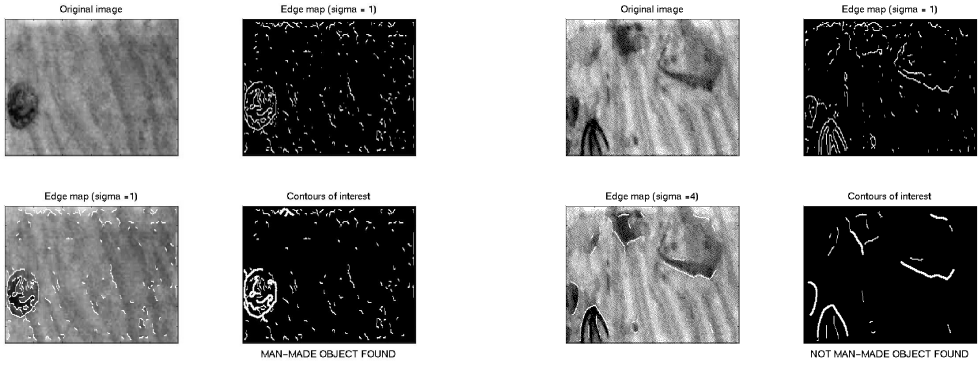
A) Frame with: Man-made object
Classification: Man-made (true)

B) Frame with: Natural objects
Classification: Natural (true)



C) Frame with: Man-made object
Classification: Man-made (true)

D) Frame with: Natural objects
Classification: Natural (true)



E) Frame with: Man-made object
Classification: Man-made (true)

F) Frame with: Man-made object
Classification: Natural (false)

Figure 4: Matlab output of the integration of the whole system. Man-made and not man-made objects found within the images. (A) Image from the OSL tank video, (B) image from Morpheus video, images (B), (E) (D), (F) from FAU video. Original images (B) to (F) courtesy of Florida Atlantic University.

5 Conclusions

We have presented a system detecting the presence of unconstrained man-made objects in unconstrained subsea videos. The generality of the problem, compounded by the varying quality of subsea imagery, is tackled with a combination of optimal scale selection and classification features inspired by perceptual organisation. The system classified correctly approximately 85% of a 1390 test images from five different underwater videos, in spite of the varying image contents, poor quality and generality of the classification task.

A reliable classifier of man-made targets would cut substantially post-mission analysis time. Real-time versions could assist AUVs with the location of man-made targets in intervention missions, e.g., on pipeline-work or scientific installations. Notice that a few seconds per frame would be an acceptable frame rate in several, low-speed operations.

Finally, the system can be regarded as a tool for general video editing, and we plan to investigate applications beyond the subsea domain. We also intend to take advantage of time correlation between frames.

5.1 Acknowledgements

We thank Francesco Isgro', Francesca Odone, Alessandro Verri and the people at the OSL for valuable discussions and inputs. The AUV videos used in our experiments were kindly supplied by the SeaTech centre of Florida Atlantic University, US. This work was supported by the Department of Computing and Electrical Engineering at Heriot Watt University.

References

[1] J. Battle, A. Casals, J. Freixenet, and J. Marti. A review on strategies for recognising natural objects in colour images of ourdoor scenes. *Image and Vision Computing*, 18:515–530, 2000.

[2] J. Canny. A computational approach to edge detection. *T-PAMI*, 8:679–689, 1986.

[3] H. S. Dabis, P. L. Palmer, and J. Kittler. An interest operator based on perceptual grouping. In *The 9th Scandinavian Conference on Image Analysis*, pages 315–322, 1995.

[4] E. Dura, J. Bell, and D. Lane. Seafloor classification using contour maps recovered from side-scan sonar. In *CAD/CAC Conference, Halifax, Cananda*, 2001.

[5] G. L. Foresti and S. Gentili. A hierarchical classification system for object recognition in underwater environments. *IEEE Journal of Oceanic Engineering*, 27(1):66–78, 2002.

[6] P. Gamba, R. Lodola, and A. Mecocci. Scene interpretation by fusion of segment and region information. *Image and Vision Computing*, 15:499–509, 1997.

[7] Q. Gao. Extracting object silhouettes by perceptual edge grouping. In *IEEE International Conference on Systems, Man, and Cybernetics*, volume 3, pages 2450–2454, 1997.

[8] Q. Iqbal and J. K. Aggarwal. Applying perceptual grouping to content-based image retrieval: Building images. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 42–48, 1999.

[9] J. S. Jaffe. Computer modelling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15:101–111, 1990.

[10] K. S. Kumar and U. B. Desai. Joint segmentation and image interpretation. In *Proceedings of Third IEEE International Conference on Image Processing*, volume 1, pages 853–856, 1996.

[11] D. G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.

[12] M. Singh M. Markou and S. Singh. Neural network analysis of minerva scene analysis benchmark. In *11th International Conference on Image Analysis and Processing*, September 2001.

[13] B. L. McGlamery. A computer model for underwater camera systems. *SPIE Ocean Optics*, 208:221–231, 1979.

[14] M. Mirmehdi, P. L. Palmer, and J. Kittler. Feedback control strategies for object recognition. *IEEE transactions on image processing*, 8(8):1084–1101, 1999.

[15] F. Mokhtarian. Silhouette-based isolated object recognition through curvature scale space. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(5):539–544, 1995.

[16] P. L. Palmer, H. Davis, and J. Kittler. A performance measure for boundary detection algorithms. *Image Processing: Image Understanding, Computer Vision and Graphics*, 63:476–494, 1996.

[17] P. Parodi and G. Piccioli. A feature-based recognition scheme for traffic scenes. In *Proceedings of the Intelligent Vehicles Symposium*, pages 229–234, 1995.

[18] G. L. Pronzato and A. M. Wallace. Adaptive control of a boundary detection algorithm. In *IEE International Conference on Image processing and its Applications*, pages 356–360, 1997.

[19] I. T. Ruiz, Y. Petillot, D. Lane, and J. Bell. Tracking objects in underwater multibeam sonar image. In *IEE Colloquium on Motion Analysis and Tracking*, 103, pages 11/1 –11/7, 1999.

[20] T. M. Strat and M. A. Fischler. Context-based vision: Recognizing objects using information from both 2d and 3d imagery. *IEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10), 1991.

[21] P. Zingaretti and S. M. Zanoli. Robust real-time detection of an underwater pipeline. *Engineering Applications of Artificial Intelligence*, 11:257–268, 1998.