# A robust visual attention system for detecting manufactured objects in underwater video

Christian Barat
Lab. I3S, Sophia Antipolis, France
Email: barat@i3s.unice.fr

Maria-João Rendas
Lab. I3S, Sophia Antipolis, France
Email: rendas@i3s.unice.fr

***Abstract*— The paper presents an approach to detect objects in underwater images. The method is composed of two main steps. The first step detects the motion of contrasting neighboring regions in consecutive frames, based on the Minimum Description Length test. The second step adapts a statistical snake to the boundary between the adjacent regions. The method is evaluated on real images and compared to classical snakes.**

## I. Introduction

This paper is a contribution to the definition of processing chains for the automatic recognition of specific classes of objects, which generally involve three steps:

1) detection of the objects
2) active observation of their shape,
3) identification of the object class.

This communication addresses the (first) detection step, considering the problem of using the apparent motion induced in the image to detect objects and extract their contours.

Classical approaches for object detection in video streams use geometric features (e.g. edges, corners,...), texture, or color information. However, problems specific to underwater images induce spurious variations that may affect the robustness of these approaches: insufficient level of natural light often imposes the use of onboard artificial light sources, which induce systematic patterns of non-uniform illumination; the frequency selective filtering action of the water mass perturbs the perceived colors (for RGB images useful information is in most cases limited to the green band); suspended matter (marine snow) clutters the images when the robot is operating close to the seabed. On top of these environment induced problems, the quality of the video system is often poor and the acquired images are noisy, displaying periodic interference patterns. In these conditions, some of the common techniques for object detection may fail to extract the boundaries corresponding to the true object surfaces. For instance, application of usual edge detectors (Sobel or Canny filters, etc.) to images with the high noise levels common in underwater video, leads to detection of many false edges, being very sensitive to the particular choice of detection thresholds [3].

The processing chain proposed comprises three steps: *(i)* object detection (using apparent motion in consecutive frames) that initializes object's contours in regions of high confidence; *(ii)* contour expansion (in each image), using the local characteristics of the image in the regions detected in the first step; *(iii)* border detection, which terminates the iterative expansion of the contours.

In all steps, we use a probabilistic approach to design the algorithms used. Steps *(i)* and *(iii)* are implemented as MDL (Minimum Description Length, see [1]) tests. We presented before the MDL test for a binary decision problem between two distinct distributions over finite sets [4], in the context of image segmentation. Being derived from the MDL principle, the algorithms reported do not require the definition user-specified parameters, such as noise levels, intensity distributions, etc. All thresholds are automatically defined from image data. This presents obvious advantages in underwater robotics, where the operating conditions can vary significantly from one mission to another. Step *(ii)* uses a snake formalism using a force field derived from locally fitting a mixture model to the image. Being based on the characteristics of the image neighborhood around each point, the method is robust with respect to high frequency noise, and can accommodate smooth variations of the object and background characteristics. The proposed methodology presents some advantages with respect to approaches previously presented in the literature: it is robust with respect to non-uniform illumination induced by robot-transported lights (as it is common in underwater robotics), since consecutive images are compared over *the same region* of the images, and it is also robust to environmental noise such as marine snow (since it contributes in the same manner to the statistics of corresponding image regions in consecutive frames). We stress that with our approach we *don't need to correct for non-uniform illumination* or to *filter noisy artifacts*. The paper is organized as follows. Section II briefly presents the MDL principle and its application to the binary tests considered in the paper. Section III presents the attention mechanism (step *(i)*) based on detection of apparent motion. Section IV presents the initialization of contours at image sites indicated by the attention mechanism and learning of the statistical models (object and background) required by subsequent steps. The extraction of the object's contours is then assessed in Section V. Finally, Section VI illustrates the performance on real underwater images. Section VII summarizes our results.

## II. MDL as a self-tuned test for homogeneity

We introduce now some nomenclature and notation. Let $X$ be a discrete random variable (rv) with probability space $(\Omega, \mathcal{A}, P)$ where $\Omega = \{a_1, a_2, \ldots, a_M\}$ is the (finite) re-

alization space, $\mathcal{A}$ is a sigma-field of subsets of $\Omega$ and $P$ is a probability measure defined over the elements of $\mathcal{A}$. We denote by lowercase letters $x$ the realizations of the rv $X$. Consider a sequence $x^{(N)} = \{x_1, x_2, \ldots, x_N\}$ of $N$ independent realizations of $X$. The type of $x^{(N)}$, which we denote by $\nu_{x^{(N)}}$, is the empirical estimate of the probability law (pl) of $X$, and is given by:

$$\nu_{x^{(N)}}(a_j) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{a_j}(x_i), \qquad j = 1, \ldots, M \ , \quad (1)$$

where

$$\mathbf{1}_{a_j}(x_i) = \begin{cases} 1, & x_i = a_j \\ 0, & x_i \neq a_j \end{cases}$$

is the indicator function.

Let $x_i^{(N)} = \{x_{i_1}, x_{i_2}, \ldots, x_{i_N}\}, i = 1, 2$, be two sequences of length $N$. The MDL (Minimum Description Length, see [1]) test for choosing between the two following composite hypotheses[1]:

$$H_0 \quad : \qquad x_1^{(N)} \sim p_0^N, \quad x_2^{(N)} \sim p_0^N \quad (2)$$

$$H_1 \quad : \qquad x_1^{(N)} \sim p_1^N, \quad x_2^{(N)} \sim p_2^N, \qquad p_1 \neq p_2 \quad (3)$$

where the probability laws $\{p_j\}_{j=0}^2$ are unknown – i.e., for deciding whether the two sequences were generated by the *same* probability law or if they are samples from *distinct* distributions – has been derived in [4]:

$$\frac{M-1}{M}\log\left(\frac{(N+1)^2}{2N+1}\right) \underset{H_1}{\overset{H_0}{\gtrless}} D\left(\nu_1||\hat{\mu}\right) + D\left(\nu_2||\hat{\mu}\right) \ . \quad (4)$$

In the previous expression, $\hat{\mu}$ is the balanced mixture of the types $\nu_1, \nu_2$ – see eq. (1) – of the two observed sequences $x_1^{(N)}$ and $x_2^{(N)}$, $D\left(\cdot||\cdot\right)$ is the Kullback-Leibler divergence between probability laws:

$$\hat{\mu} = \frac{1}{2}\left(\nu_1 + \nu_2\right), \qquad D(\nu||\mu) = \sum_{i=1}^{M} \nu(a_i)\log\frac{\nu(a_i)}{\mu(a_i)} \ ,$$

and $M$ is the size of the realization space $\Omega$.

Equation (4) shows that under the hypothesis of statistically independent samples, the types of the observed sequences are sufficient statistics for the decision problem.

## III. Attention Mechanism: motion detection

We apply the decision test presented in the previous section to identify the image regions inside which there are (apparently) moving objects. Let $I(t)$ denote the (intensity of the) image acquired at instant $t$, of size $(n \times m)$. Consider a fixed partition of the image plane, in rectangular windows of size $n_1 \times n_2$, see Fig. 1. Let $B_{ij}(t)$ denote the $i,j$-th block of $I(t)$.

To detect image motion between images $I(t)$ and $I(t+\Delta t)$ we apply the MDL test (4) to corresponding blocks of the two images, $B_{ij}(t)$ and $B_{ij}(t+\Delta t)$, see Fig. 1. If a moving object is present in one of them (for instance in $B_{ij}(t)$) and not in the other ($B_{ij}(t+\Delta t)$), the associated types will be different, and test (4) applied to the types $\nu_1 \leftrightarrow \nu_{ij}^t$ and $\nu_2 \leftrightarrow \nu_{ij}^{t+\Delta}$ should flag this motion.

[1]Notation $x \sim p$ means that the rv $x$ is drawn from pl $p$.

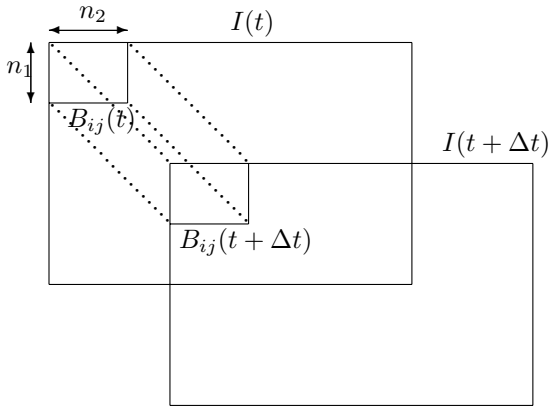


Fig. 1. Definition of image partition for motion detection test.

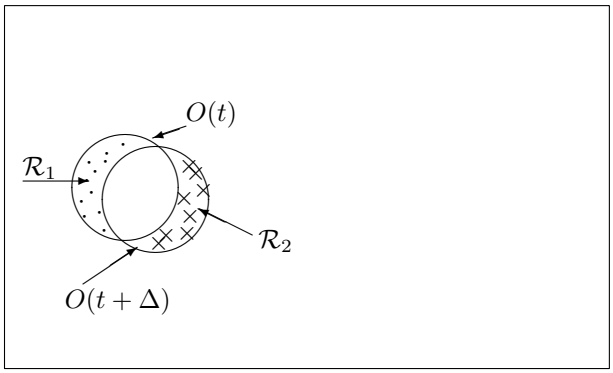


Fig. 2. Detected regions.

Let $(u_t, v_t)$ be a pixel of $I(t)$ corresponding to a point in object $O$ – denoted by $I_{ij}(t) \in \mathcal{O}$ – and $(u_{t+\Delta t}, v_{t+\Delta t})$ be its location in image $I(t+\Delta t)$.

Application of the MDL test (4) will identify two regions in the camera frame (see Figure 2): $\mathcal{R}_1$ corresponding to the pixels $i,j$ for which $I_{ij}(t) \in \mathcal{O}$ and $I_{ij}(t+\Delta t) \notin \mathcal{O}$; and $\mathcal{R}_2$ gathering the pixels such that $I_{ij}(t) \notin O$ and $I_{ij}(t+\Delta t) \in O$:

$$\mathcal{R}_1 = \overline{O}(t) \cap O(t+\Delta t), \qquad \mathcal{R}_2 = O(t) \cap \overline{O}(t+\Delta t) \ , \quad (5)$$

where we defined the notation $O(t) = \{(u,v) \ : \ I_{uv}(t) \in O\}$, and $\overline{A}$ denotes the complement of set $A$.

Since test (4) is performed over *the same* regions of the image plane, the method is robust to spatial illumination variations that are associated to the observing platform: aside from the variations induced by the 3D structure of the perceived scene, the shading factors associated to blocks $B_{ij}(t)$ and $B_{ij}(t+\Delta t)$ will be essencially the same. The fact that motion detection is based on the image types also induces robustness with respect to temporally stationnary spatially uncorrelated noise, that will contribute in the same manner to the empirical distributions $\nu_{ij}^t$ and $\nu_{ij}^{t+\Delta}$.

Figure 3 shows two real underwater images acquired at

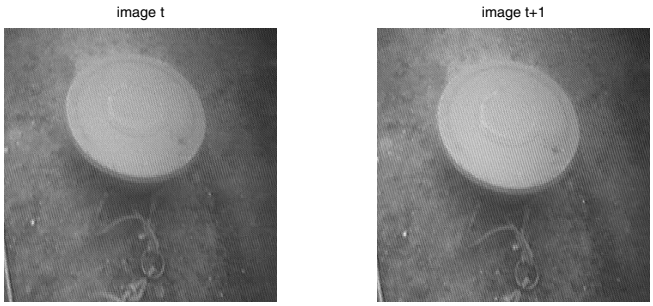


Fig. 3. Original images.

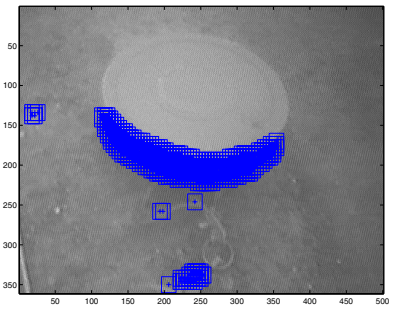


Fig. 4. Motion detection results.

consecutive times, and Figure 4 the result of applying test (4) to these images. The blue windows (coinciding with the windows used for the computation of the types $\nu_{ij}$ around each pixel) are centered at the pixels $ij$ for which test (4) decides for hypothesis $H_1$. As we can see, the bottom border of the circular object is well detected, as well as the close chain laying in the ocean bottom. A small spurious detection can be seen in the upper left region of the image, corresponding to a local variation of the sea-bottom.

## IV. Contour Initialization and Model Learning

### A. Initialization

In the previous section we addressed the problem of using *motion* in the video sequence to detect the regions that *can* belong to object edges. We address now the problem of using this information to initialize extraction of the complete object's contours. We work in the first image $I(t)$ where motion has been detected.

Our algorithm first identifies the rectangular regions $W_k(t), k = 1, \ldots, K$ containing all neighboring pixels passing test (4) - $K$ is the number of regions obtained by merging adjacent windows in image $I(t)$. The upper image in Figure 5 shows the regions obtained for the image in Figure 4, for which $K = 5$. Regions $W_k$ of small size (less than 3 times the width of the windows $B_{ij}$) are discarded (for Figure 5, the blue, yellow and brown regions near the border of the circular object).

Inside each $W_k(t)$ we then initialize a contour $c_k^0(\ell)$ by selecting the point $(i,j)_k \in W_k(t)$ at which the *image intensity gradient* $\nabla I_{(ij)_k}(t)$ is maximum (we low-pass filter the image before the computation of the gradient). Contour $c_k^0(\ell)$ is a linear segment going through $(i,j)_k$, orthogonal to $\nabla I_{(ij)_k}(t)$ and with length $\left(n_1^2 + n_2^2\right)^{1/2}$ equal to the size of diagonals of the windows $B_{ij}$ used by the motion detector of section III, see the red segment in the lower image of Figure 5.

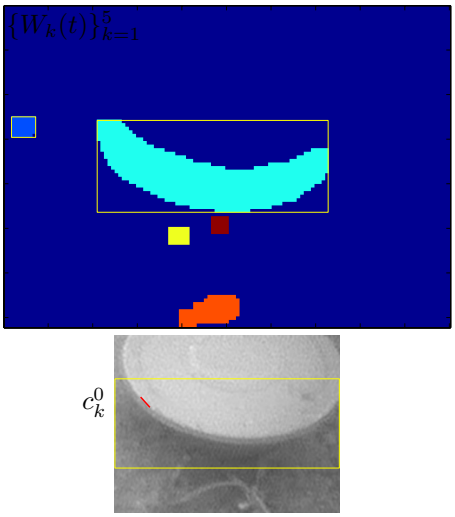


Fig. 5. Window merging and contour intialization.

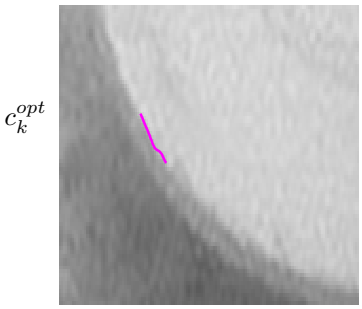


Fig. 6. Local contour adaption.

### B. Adaption based on gradient

The small *linear* segment $c_k^0(\ell)$ is next adjusted to fit to the local contour of the object. We rely on a classical (open) snake formalism (see e.g., [5]) that maximizes the intensity gradient along the optimizing contour $c_k^{opt}(\ell)$ . Figure 6 shows the contour obtained by deforming the linear segment in the bottom image of Figure 5.

### C. Model learning

Before extending the contours $c_k^{opt}(\ell)$ obtained in the previous step, we learn the distributions $h_1^k, h_2^k$ of the pixel intensities of the adjacent regions that it separates, as the types of the pixel intensities in rectangular windows on each side of $c_k^{opt}(\ell)$, see Figure 7. These distributions will drive the algorithm of contour expansion presented in the next section (see [6]).

## V. Contour Expansion

In this section we iteratively extended the local contours $c_k^{opt}(\ell)$ by "tracking" the object's boundary. We work locally with each contour, and we drop contour index $k$ in this

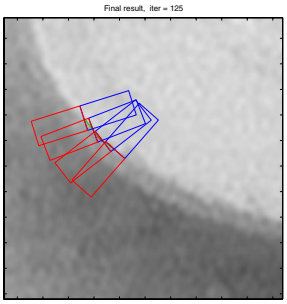


Fig. 7. Windows used to learn the two regions.

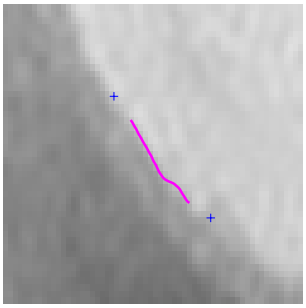

Fig. 8. Contour extension.

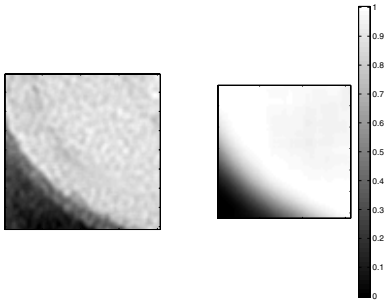

Fig. 9. Estimates of mixture coefficient ($\alpha_{ij}$).

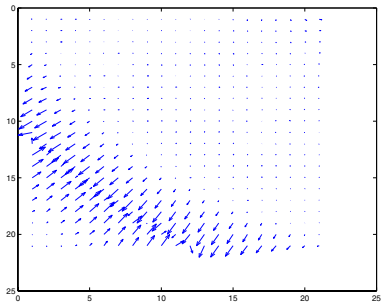

Fig. 10. Force field.

section, using the simpler notation $s^0(\ell) \equiv c_k^{opt}(\ell)$. We start by iteratively extending the contour for which the region $k$ inside which the motion detector flagged a stronger variation and inside which no detected contour passes yet. This means that we discard the regions that are crossed by a previously extracted contour.

Let $n = 0$. At iteration $n$ we extend the contour $s^n(\ell)$ by adding extra points $p_i^n, p_o^n$ along the tangent at each extremity, see the blue crosses in Figure 8. We assume that the object's contours are in general open curves with end points at the *interior* of the image, allowing for smooth transition regions at object's boundaries. The relevance of the added points is tested in a later step.

Each new end point $p_a^n, a \in \{i, o\}$ is then locally displaced according to a force field $F_{ij}$ derived from the mixture coefficient $\alpha^n$ of the following mixture model

$$\mu_{ij}^n = \alpha_{ij}^n h_1 + (1 - \alpha_{ij}^n) h_2, \qquad \alpha^n \in [0, 1], \tag{6}$$

that is locally fit to the image at each pixel $(i, j)$. In the previous equation $\mu_{ij}^n$ is the type of the pixel intensities in a window centered at the pixel $(i, j)$ and of size $n_1 \times n_2$, and $h_1, h_2$ are the pl's previously learned. We denote by $p_a^\infty, a \in \{i, o\}$ the resulting end points.
The force $F_{ij}$ is given by

$$F_{ij} = \begin{bmatrix} \frac{\partial}{\partial j} |0.5 - \alpha_{ij}| \\ \frac{\partial}{\partial i} |0.5 - \alpha_{ij}| \end{bmatrix}. \tag{7}$$

Figure 9 shows (on the right) the coefficient $\alpha_{ij}$ for the image region shown in the left, Figure 10 shows the corresponding local force field $F_{ij}$, and Figure 11 the fitted contour on the same image after 16 expansions.

### A. End condition

Each new adjusted end point $p_a^\infty, a \in \{i, o\}$ is tested for local contrast around it by using the the MDL test (4) in windows adjacent to the added contour extremity, see Figure 12. For instance, in the Figure 12 the segment 1 will not pass the test, while the segment 2 does.

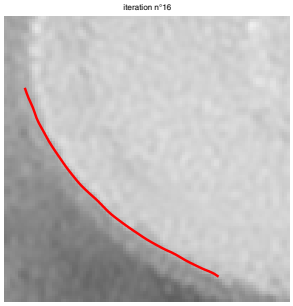


Fig. 11. Contour $s^{16}(\ell)$.

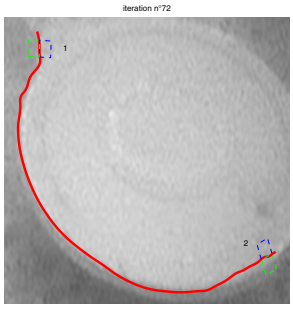


Fig. 12. End test $s^{72}(\ell)$.

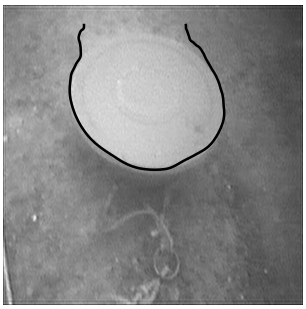

Fig. 13. Final snake for our method.

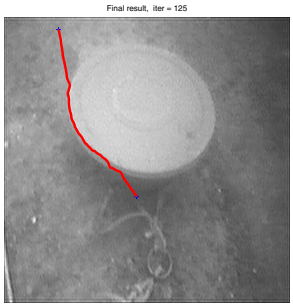


Fig. 14. Final snake for Gradient method.

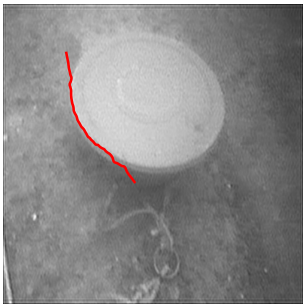

Fig. 15. Final snake for Gradient Vector Field method.

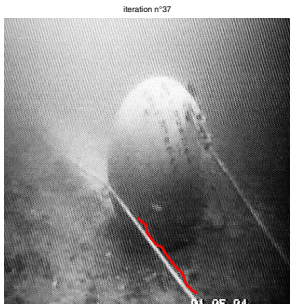


Fig. 16. Example 1, Snake 1.

## VI. Experimental results

In this section we present results of real images [2] with acquisition noise and variable illumination. Our results are compared to two other methods (Gradient Vector Field and classical Snakes based on the image Gradient [5] ). Our first result (see Figure 14) considers the already seen example (see Figure 13). The snake is stopped near the upper part of the mine due to the illumination effect and the weak contrast between the two regions. In Figure 14 we show the result in the same image with snakes driven by classical External forces based on the gradient described in [5]. Roughness of the object edges in the image determine quick loss of contour tracking, with the snake diverging from the object. The same kind of results are obtained with the Gradient Vector Field, see Figure 15, because the force is based only on the Gradient (The method presented in [5] has been modified by us to enable open snakes). The classic closed snakes leads to results very dependent on the snake initialization. These methods require to filter the images to obtain a Gradient map compare to our method which can be applied to very noisy images, since the forces deforming the snake are based on the characteristics of the local regions around each point. Figures 16, 17, 18, 19 illustrate results for different images. The method succeeds in detecting correctly some parts of the object even with the high acquisition noise.

These examples are representative of the diversity of underwater images and of the problems encountered with this kind of images. Once we have detected some interesting parts of the contours, we must determine if they correspond to a manufactured object. We obtain long contours for manufactured object and very short contours for natural objects. The form of the contour and the texture of the area adjacent to the contour are also good criteria for declaration of the presence of manufactured object. These approaches will be developed in the future.

[2]The information contained in this publication are derived from data property of the French State that have been provided by the GESMA (Groupe d'Etudes Sous-Marines de l'Atlantique) within TOPVISION project coordinated by Thales Underwater Systems SAS. This project is related to Techno-Vision Programme launched by french Ministry of Research and french Ministry of Defense

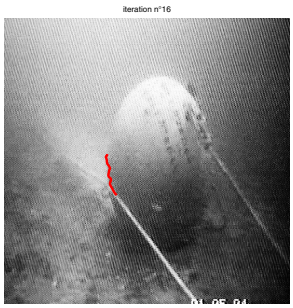


Fig. 17. Example 1, Snake 2.


Fig. 18. Example 2.

## VII. Conclusions

In this paper we proposed a new method to detect objects in underwater images using the apparent motion in consecutive frames. The approach uses MDL test to determine regions of interest, like a visual attention mechanisms. Inside these regions small snakes are initialized based on local image contrast.

We then use a statistical snake adapted from our previous work in [6] to expand the snake, tracking the contour between the adjacent regions. We present results on very noisy images demonstrating the good performance of our method and we compare it to other approaches. In the future, we must determine if the detected contour is a part of a manufactured object.

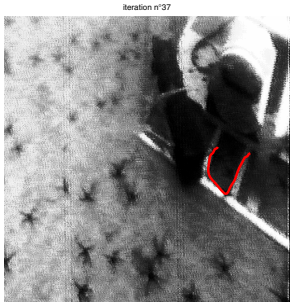


Fig. 19. Example 3. Detection of a robot part.

## References

[1] Jorma Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Series in Computer Science-Vol. 15, 1989.

[2] S. Rolfes *Stochastic geometry : an approach to featureless perception-based robot navigation*, PhD thesis, December 2002.

[3] A. Olmos, E. Trucco, "Detecting man-made objects in unconstrained sub sea videos," In Proc. British Machine Vision Conference, pages 517–526, 2002.

[4] A. Tenas, M-J Rendas and J-P Folcher, "Image Segmentation by Unsupervised Adaptive Clustering in the Distribution Space for UAV Guidance Along Sea-bed Boundaries Using Vision," Oceans'2001.

[5] C. Xu, J. Prince, "Snakes, Shapes amd Gradient Vector Flow," IEEE Trans. Image Proc., Vol. 7, No. 3, March 1998.

[6] C. Barat, M. J., Rendas,"Tracking Benthic Boundaries Using a Profiler Sonar : a mixture model approach," Proc. Oceans 2003, San Diego, USA, 2003.