Underwater Robot Navigationn Using Benthic Contours

M.-J. Rendas and S. Rolfes
Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis
2000 Rte des Lucioles, BP 121
06903 Sophia Antipolis cedex, France
{rendas,rolfes}@i3s.unice.fr

***Abstract*-This paper presents results on the navigation of mobile underwater robots using maps of contours of distinct habitats of the sea-floor. The contour maps are acquired by autonomously tracking the boundaries of contrasting regions of the sea bed using a video camera mounted on the robot. Recognition of previously seen regions enable the robot to reset dead-reckoning errors, enabling consistent position estimates to be maintained.**

I. Introduction

One of the major technical issues in autonomous underwater vehicles (AUVs) is the positioning navigation problem. While the majority of operational AUVs rely on acoustic support systems placed at calibrated positions, other approaches, in particular the use of GPS, have been proposed in an effort to decrease the large costs and limited operational range imposed by use of acoustical beacons. GPS has the advantage of allowing extended operational ranges, but the fact that precise positioning information is available only at the ocean surface, leads to increased positioning errors when operating near the bottom. Locating, or re-locating structures or objects in the sea floor may be an important requirement for many applications. Many efforts have been devoted to navigating with respect to existing environmental features, but the majority of the work presented so far has mainly concentrated in either very local navigation (for instance, station keeping) or in navigation with respect to man-made structures, such as pipelines or cables.

In this paper we present results on the navigation of autonomous underwater robots using natural features of the environment, more precisely, contours between distinct natural benthic habitats. Our approach builts upon the possibility of autonomously driving an underwater vehicle along contours between regions of distinct occupation of the sea-bed. This guidance mode is used to directly produce a map of the contours of the sea-bed regions, which the robot uses later to navigate. In the work presented here, we use vision as the basic sensing modality, although the work could be easily extended to use of sonar instead.

We shortly present the unsupervised image segmentation and contour guidance on which our approach is based [6] and concentrate afterwards on the problems of: *(i)* contour mapping and *(ii)* positioning with respect to the mapped contours. Our contour mapping algorithm recursively produces a simplified representation of the (2 ½ D) curve that separates the identified regions, as a concatenation of line segments and arcs of circle, producing thus a structure that is easy to use by the upper stages of processing. We present the mapping algorithm, which is based on the paradigm of propagation of multiple hypothesis, and the recognition algorithm, which is based on an association graph that measures the plausibility of associating the mapped representation and the contour representation obtained from current observations.

The recognition algorithm yields, as a by-product, the correction of the current estimate of the vehicle position that is required to obtain spatial co-registration of the observed and mapped shape, bringing the positioning error level to the one associate to the map. One of the major characteristics of underwater perception-based navigation is the fact that the range of the perception sensors is small, and the system must be able to cope with a complex ambiguity structure, since distinct, widely separated regions may (locally) look alike. Our recognition algorithm deals with this fact by requiring that unambiguous correspondence between the mapped and observed shapes be established.

II. Map acquistion

In this section be briefly present the methodology used for the autonomous acquisition of an internal representation of the boundaries between distinct regions of the sea-bed. It is fundamentally based on the ability of identifying, using vision, regions of the sea floor that are occupied by distinct habitats. Note that we do not claim that we identify the precise nature (species or sea floor type) of each region of the observed sea bed. This would require a supervised approach, relying on extensive sets of training data to correctly associate a meaningful label to each region. Instead, our maps, whose main purpose is to serve as navigation aids, will rely on segmenting the observed portion of the sea bed into regions of distinct visual appearance. It is, as we discuss below, an unsupervised segmentation which does not rely on any previous training step.

The primary goal of video segmentation is to be able to detect homogenous regions in each video frame, and to track them across the entire sequence of images. The first difficulty in performing this task is the fact that the sense of *homogenous* is not well defined. While for video sequences consisting of artificial objects, such as the scenes perceived by a robot moving inside an office building, homogenous can most of the time be understood as "constant intensity" (color or gray level), for natural scenes this definition becomes meaningless. An obvious alternative is to define homogenous regions as ones

0-7803-7534-3/$10.00©2002 IEEE

displaying "homogeneous texture". While this is the intuitively acceptable definition of homogeneity when dealing with natural scenes, its mathematical definition is not clear, rising additional questions of scale, allowed degree of spatial variability, etc. Moreover, for images of natural environments often no clear-cut separation exits between the distinct regions, and even human classification yields different results depending on the person that draws the separation between the classes. Our approach aims at finding the regions of "same texture." The exact sense of our definition of texture will become clear below. While we do not claim optimality of our present algorithm, we think that it casts the problem in a sufficiently rich formal framework that enables its adaptation to other more complex scenarios that those considered here, or to distinct sensing modalities. A non-parametric approach to the modelisation of the distinct classes present in the sea bottom is used, under the basic assumption that each class induces in the observed images a probability distribution of the intensity levels of the image's pixels.

A. Problem formulation (image segmentation)

We consider that the observed image $\mathbf{I}$ is dependent on a discrete image $\mathbf{E}$ in the following way:

$$\mathbf{I}=\mathbf{F}(\mathbf{E})$$

$\mathbf{E}$ is the image that we want to determine. Let K be the number of distinct classes present in the image, and denote by $\mathbf{E}_{ij}$ the value of $\mathbf{E}$ at site ij (line i and column j). Denote by I the subset of $\mathbf{Z}^2$ in which the indexes ij take values. Then $\forall ij \in I, \mathbf{E}_{ij} \in \{1,2,...,K\}$ where we have arbitrarily numbered the classes. We assume that the image pixels of the image $\mathbf{F}$ are statistically independent realization of unknown probability laws $P_k, k=1,...,K$, which are solely dependent on the class to which the pixel belongs:

$$\Pr\{\mathbf{F}(\mathbf{E}) \mid \mathbf{E}\} = \Pr\{\mathbf{F}_{ij}(\mathbf{E}_{ij}), i,j \in I \mid \mathbf{E}\} = \prod_{i,j\in I} P_{\mathbf{E}_{ij}}(\mathbf{F}_{ij})$$

The preceding equation defines the conditional probability of the observed image, given the segmented image $\mathbf{E}$ and the probability laws $\{P_k\}_{k=1}^K$, and is our observation model (actually, the likelihood function for the estimation problem formulated below).

Our goal is to estimate $\mathbf{E}$ *from the observations* $\mathbf{I}$, *without knowledge of the probability laws* $P_k, k=1,...,K$ *associated to each class (unsupervised segmentation).*

B. Segmentation algorithm

The probability laws of the unknown classes play, with respect to the estimation of the segmented image, the role of unwanted parameters. To handle the presence of these nuisance parameters, we consider the determination of the Generalized Maximum Likelihood of the label field $\mathbf{E}$. Together with our statistical model, this leads to the following criterion that the optimal estimate must maximize:

$$\ln \Pr\{O \mid E, \{P_k\}_{k=1}^K\} = -\sum_{k=1}^K N_k \left[D(\nu_n^E \mid\mid P_k) + H(\nu_n^E)\right]$$

where ν_k is the type associated to all the pixels with label k in the label field $\mathbf{E}$, $D(.||.)$ is the Kullback-Leibler divergence and $H(.)$ is the Shannon entropy [1,2], defined by

$$D(p||q) = -E_p \log\frac{q}{p} \quad \text{and} \quad H(p) = -E_p[\log p].$$

For the Generalized Maximum Likelihood, we must replace the unknown probability laws P_k by their optimal estimates for each possible label field, which can be shown to be

$$\hat{P}_k^E = \nu_k^E,$$

leading to the simple criterion

$$\ln \Pr\{O \mid E, \hat{P}_k^E\} = -\sum_{k=1}^K N_k H(\nu_k^E).$$

The optimal estimates are thus those that lead to a minimal value of the weighted sum of the classes' entropy. Optimization of this criterion is a combinatorial optimization problem which cannot be solved in real time. We rely on a sub-optimal algorithm that can be shown to converge to a local minimum of the criterion above. Moreover, instead of associating a distinct label to each pixel of the observed image, we partition the image in square windows, and assume that the pixels inside each small window all have the same label. The algorithm, which presents strong analogies with the well known K-means algorithm, but working on the space of discrete probability laws, instead of intensity of pixel values. It is initialized by a novel procedure which, starting from the histogram associated to each image, recursively builds the required number of classes' types by "splitting" one probability law at a time. Type splitting is constrained by the requirement that the current set of types are the components of a mixture model that accurately represents the image histogram, while at the same time optimizing the entropy criterion above. Details can be found in [3]. After the initialization step has fixed the initial probability laws P_k^0, the algorithm recursively alternates two steps. In the first, the site labels are determined by searching for the probability law that is closest, in the sense of the Kullback divergence, to the histogram of a window centered at the site,

$$e_{ij}^{n+1} = \arg\min_k D(\nu_{ij} \mid\mid P_k^n), n = 1,...$$

while in the second step, the classes' probability laws are updated as the histograms correspondent to the current segmentation:

$$P_k^{n+1}(i) = \frac{N_k(i)}{N_k}, i = 1,...,L$$

where L is the number of gray levels in the image, N_k is the total number of pixels with label k, and $N_k(i)$ is the number of pixels with intensity i in class k. The algorithm stops when no site changes class in one iteration.

C. *Video segmentation*

The algorithm briefly described below is used for video segmentation by considering, for each frame, the set of classes' histograms obtained in the previous image. Since no initialization step is required for the subsequent frames, the processing time requirements are minimal, enabling its use for real-time guidance of the platform. Moreover, this allows gradual adaptation to the local characteristics of the classes, which is frequent in natural media.

III. Mosaicing

Our mapping approach is based on the ability of establishing the correspondence between successive image frames, i.e., of building a mosaic of the observed sequence. Contrary to other approaches, our goal is not to produce a high quality video mosaic, but rather to infer some crude information about the geometry of the observed boundaries, while keeping the ability to recognize later the mapped regions. We assume that the robot is equipped of a compass, providing attitude information, and that a control law is implemented enabling observation of the sea floor at constant altitude. The robot camera is pointing directly downwards to the sea bottom, eliminating the need for perspective correction between frames, under the assumption that the bottom is locally flat.

Most mosaicing algorithms are based on pair-wise association of point features (usually the output of corner detectors). Identification of such features in natural environments may be difficult, and their pairwise association ambiguous, leading to important computing times, and precluding use of most of these algorithms in real time. Instead, we based our mosaicing approach on the associating of the segmented images. This greatly reduces the complexity of the matching phase (since we work of a reduced size image, whose "pixels" are the windows of the original images, while at the same time using the global structure to eliminate ambiguities. Obviously, the method will break down in regions where the geometry of the regions determined by the segmentation algorithm is invariant with respect to translations (such as straight lines).

The metric used for determining the intra-frame displacement is adapted to the statistical model used for segmentation. Let I^n and I^{n+1} be two successive images, $\{P_k^n\}_{k=1}^K$, $\{P_k^{n+1}\}_{k=1}^K$ the probability laws, and e^n, e^{n+1} the label fields found by the algorithm of the previous section. Under our simplifying assumptions, the association algorithm searches for the displacement vector Δ such that

$$C(\Delta) = \frac{\left(\sum_{s \in I(\Delta,\delta\theta)} Q(s,\Delta)\right)}{\mu(I(\Delta,\delta\theta))},$$

where $I(\Delta,\delta\theta)$ is the set of overlapping pixels for a given intra-frame rotation $\delta\theta$ and displacement Δ,

$$Q(s,\Delta) = \exp - W^2 / 2\left(D(P_{e_s^n}^n \| \nu_{s,\Delta}) + D(P_{e_{s+\Delta}^{n+1}}^{n+1} \| \nu_{s,\Delta})\right)$$

where W is the window size, and

$$\nu_{s,\Delta} = \frac{1}{2}\left(P_{e_s^n}^n + P_{e_{s+\Delta}^{n+1}}^{n+1}\right)$$

is the balanced mixture of the classes of the sites in the two images. The function $Q(s,\Delta)$ measures the probability that the two probability laws, associated to the sites being compared, be types of sequences produced by the same underlying probability laws. The expression inside the parentheses in the definition of Q is called the Jensen-Shanon divergence. Note that use of this metric, instead of abusive use of correlation of the arbitrary label fields, has the advantage of increasing the robustness of the mosaicing algorithm with respect to the "correct" identification of the number of classes. Spurious class subdivision results in probability laws which are very similar, when compared with the Kullback divergence, and spatial frame association will be correct by using the Q metric, even if there is frame-to-frame discontinuity in the classes labels.

IV. Contour extraction

As we said previously, our goal is to map contours between regions of high visual contrast. Of all the regions detected by the segmentation algorithm in the video sequence, only those with low probability of corresponding to independent observations of the same underlying distribution will be retained for mapping.

An internal representation of the geometry of the contours is recursively updated using a modification of the algorithm Ifshades[1]. This algorithm is based on the paradigm of multiple hypotheses. It incrementally builds an approximate description of 2D curves as a concatenation of fuzzy line segments and arcs of circle. These fuzzy parametric shapes are formally defined [4] by considering their parameters as fuzzy sets in the corresponding parameter space, instead of point sets, with zero measure. For instance, for line segments, their orientation, length and origin are fuzzy sets, properly defined by a membership function over $\Re^4$, instead of single values, while for fuzzy arcs of circle, their center, radius and angular opening are a fuzzy subset of $\Re^4$.

The input to the algorithm are noisy measures of points in the curve. For each input point, its degree of association to each fuzzy shape (as we said before, presently only line segments and arcs of circle are considered) is computed. In

[1] *Ifshades* is an acronym for *Incremental Fuzzy Shape Description.*

a manner consistent with the representation of the uncertainty of the curve geometry, the uncertainty associated to the measures is also characterized by a membership function. In this way, the degree of association is simply computed by computing the fuzzy truth value of the assertion "the point is a subset of the line (circle)". When no internal shape can be found with a sufficiently high degree of association, new fuzzy shapes, covering all possible sets of shapes containing the observed point, are created. Since local observations of an arc of circle cannot be discriminated from observations of its tangent, initially, only lines are created. When a new point is successively associated to existing shapes (eventually to more than one), their parameters are updated. This update assumes that the fuzzy sets defining each type of shape have a constant shape, differing only on the point where they are centered on the parameter space, and on scale parameters (location/scale model). Update of the central point is based on the observed (fuzzy) moments of all points associated to the shape, while scale parameters (that capture the uncertainty associated to the shape) are updated on the basis of the observed residuals. Due to the mechanism of multiple hypotheses, several shapes can converge to the same shape (if they are systematically updated by the same set of points) or empty (spurious hypothesis, to which no further data has been associated). Regularly, the existing representation is managed, by merging similar shapes (based on a criterion of intersection) and elimination spurious ones.

The initial *Ifshades* algorithm has been designed for distance measures, for applications in terrestrial robotics, using a mini-robot equipped of infra-red sensors. In the context of the underwater mapping of this paper, more semantic information is available for each measure, in particular the classes on each side of the detected contour, and the depth of the sea bottom. Note that, as we said, the classes' probability law can slightly vary from frame to frame. For this reason, the internal representation of each shape is now augmented with the probability laws of the classes separated by the contour. The measure of association of each contour point to the existing frame is also modified, to take into account, besides spatial co-registration, the similarity of the corresponding distributions using the Q measure defined in the previous section, and depth.

The version of the *Ifshades* algorithm presented in [4] considered that the uncertainty of each pint measure is constant. For the visual information used in this paper, the uncertainty affecting each contour point depends on the visual contrast between the classes on each side of the contour. By formulating the problem of estimating the mixing parameter that allow us to express the type of a window centered at each contour point ν_s as a mixture of the histograms associated to the classes separated by the contour, ν_{left}, ν_{right}

$$\nu_s = \alpha \nu_{left} + (1-\alpha)\nu_{right}$$

and computing the Fisher matrix for parameter α, we can approximate the covariance of the error associated to the determination of the mixing parameter as:

$$\sigma_\alpha^{-2} = \sum_{i=1}^{L} \frac{\left(\nu_{left}(i) - \nu_{right}(i)\right)^2}{\nu_{left}(i) + \nu_{right}(i)}.$$

The error on the determination of the contour point is a scaled version of the parameter a (neglecting the curvature of the contour inside the analysis window), yielding

$$\sigma_c^2 = \sigma_\alpha^2 \left(W/2\right)^2.$$

This value is used to defone the mebership function of each data point in the modified version of *Ifshades*.

V. Contour Recognition

We designed in the past [5] a contour recognition method which is able to match distinct representations of the same 2D curve produced by Ifshades. This algorithm is based on the creation of an association graph between pairs of fuzzy shapes (of the memorized and current representation). The nodes of this graph correspond to the association of two particular shapes. To each node is associated a membership function of the set of possible rotations and displacements, compatible with the corresponding association. Association between distinct representations is detected by searching for an unique path along which the intersection of the membership functions of the traversed nodes is not empty. If no such single path exists, it means that the geometry of the contour portion observed so far does not enable unambigous association of the observed curve to the mapped contour. In this case we assume that the robot is equipped of a compass, reducing the set of rigid motions that can map the current observation to the map to translations. The hypothesis of constant observation altitude enables us to neglect also the need for searching over scale parameters.

VI. Experimental Results

We present below results of the algorithms described previously to real data collected using the ROV Phantom[2] at Villefranche-Sur-Mer (South of France). In this area, several boundaries can be identified between the sandy bottom and colonies of posidonia. The data sets presented below have been acquired using the visual tracker presented in [6].

On each image of the acquired data set we applied the segmentation algorithm described in section II.A. A maximum number of three classes have been found. Each segmented image was registered on the previous image in order to determine the displacement (the orientation change

[2] *Phantom is a Remotely Operated Vehicle (ROV)* produced by Deep Ocean Eng. (USA). It has been acquired using the funding of the EPSRIT project NARVAL (Navigation of Autonomous Robots Via Active Environmental Perception), under a special educational agreement.

was given by the compass). The resulting mosaic is shown in figure 1. The quality of the mosaic is well illustrated on the top. The lower part of the figure shows the mosaic of the segmented images and it should be noted that the number and the labeling of the background classes changed several times, precluding the use of class labels in order to determine the association. It is obvious that the error of the estimated displacement between consecutive images is integrated over the entire sequence. This effect is well illustrated in figure 2.

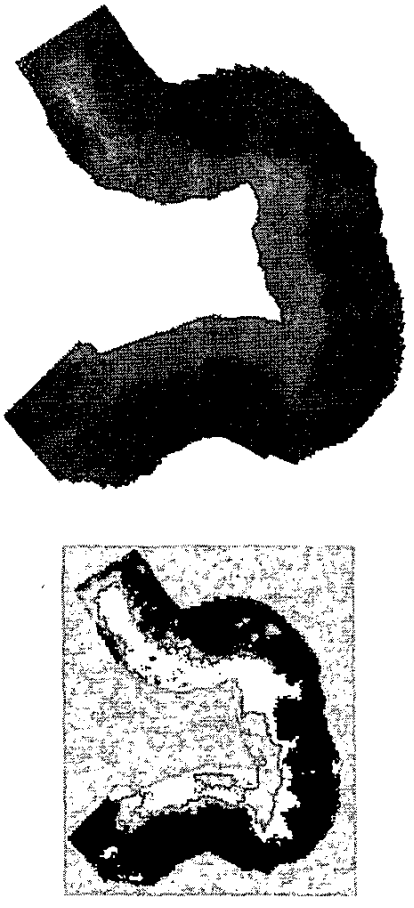

Figure 1: Mosaic of the sequence of images that were obtained during the contour tracking. The lower image corresponds to the mosaic of the segmented images.

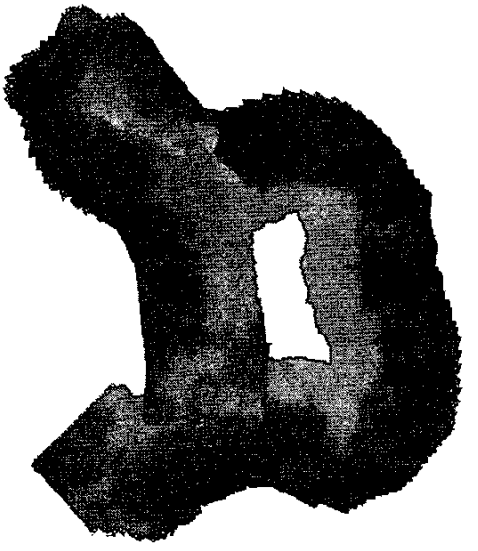

Figure 2: accumulation of the error during the acquisition of the mosaic is illustrated in this figure.

The contour that was mapped corresponds to the contour with the largest contrast of the adjacent classes and is the one that separates posidonia (dark portions of the mosaic) and other background material (mostly sandy regions). The extracted contour points (expressed in the frame of the first image) are shown in figure 3. A higher level representation of this set of contour points by 2D curves (line segments and arc of circles) is then produced by *Ifshades*. This collection of shapes produced during the first observation serves as a map and is shown in figure 4. The description is not very 'clean', but represents the advantage that due to the existence of multiple plausible descriptions of ambiguous portions of the contour recognition is made more stable.

In figure 5 we represent the contour points and their geometrical descriptions obtained during a second observation of a small portion of the contour (corresponding to the lower part of image 1). From shapes that were produced during the second pass at least two (the arc of circles traced in bold) can be matched to similar shapes contained in the map. Association of these two shapes allows then to determine without ambiguity the accumulated error (e.g. during deliberate motion) and to reset this error.

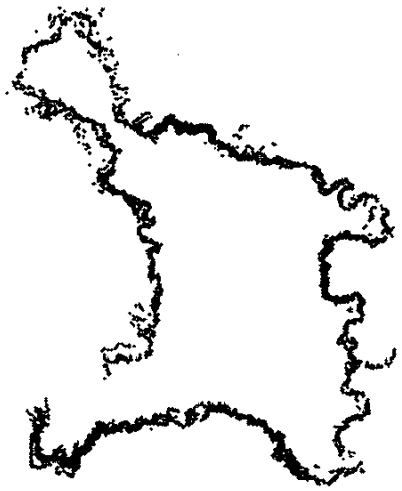

Figure 3: Sequence of contour points extracted from the contour that separates posidonia from other 'background' classes.

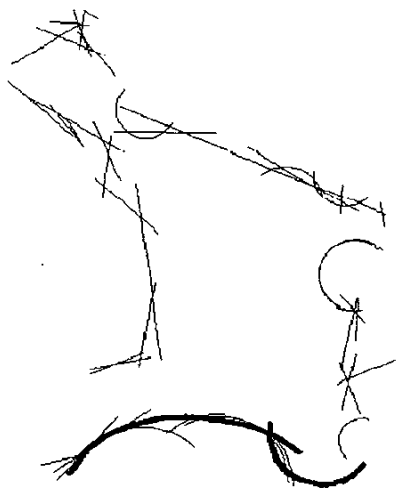

Figure 4: *Ifshades* applied on the point sequence of figure 3 results in an uncertain description of the contourpoints. Multiple 'plausible' descriptions are obtained at several locations.

Figure 5: Contour description obtained during a second pass (lower part of the contour of figure 1.

This example shows that there is enough information to establish the association between the representation obtained during the second observation and the shapes contained in the simplified observation obtained during mapping.

VII. Conclusions

We presented in this paper a methodology allowing the navigation of an autonomous underwater robot based on maps of contours between distinct regions present in the sea bottom. Several problems had to be solved in order to produce maps that are suitable for navigation. We presented the segmentation algorithm that produced coarse image by associating a label to region that present similar texture. From the segmented images we determined all contours between adjacent areas and select the contour with the largest contrast. The tracking algorithm uses this visual information in order to generate commands that keeps the robot on the contour. We then determined the displacement between the consecutive images which allows along with the information provided by the compass to produce a large mosaic of the observed region. In order to obtain a simplified description of the complex shaped contour we extracted a sequence of contour points (expressed in the same coordinate system) and created a geometrical description of the contour as a concatenation of 2D curves. We showed that this simplified contour description contains sufficient information that allows to correct positioning errors accumulated during deliberate motion between several observations of the benthic contours.

Acknowledgments

This work has been partially funded by the European Union through the IST project SUMARE (IST-1999-10836, *Survey of Marine Resources*).

References

[1] S. Kullback, *Information Theory and Statistics, 1969.*

[2] T. Cover and Joy A. Thomas, *Information Theory, John Wiley & Sons, Wiley Series in Telecommunications, 1991.*

[3] Sumare Project, Deliverable 3.3, *Video Segmentation, September 2002.*

[4] Maria-João Rendas, Stefan Rolfes, "Propagating Multiple Hypothesis for Unknown Environment Learning", Symposium of Intelligent Robotic Systems, Lisbon, Portugal, July 96.

[5] Maria-João Rendas, Stefan Rolfes, "Shape Recognition: A fuzzy Approach", IEEE Proc. of the Int. Conf. on Robotics and Automation , Leuven, Belgium, May 98.

[6] Maria-João Rendas, Albert Tenas, "Image Segmentation by Unsupervised Adaptive Clustering in the Distribution Space for AUV guidance along sea-bed boundaries using Vision," Proc. OCEANS'2001. Honolulu, Hawaii, USA, November 2001