Not to be cited without prior reference to the authors. CM 2000/T: 15

Using Autonomous Underwater Vehicles for Seabed Habitat Mapping

Stefan Rolfes[1], Maria João Rendas[1], Jon Side[2]

ABSTRACT

We present work concerning the use of autonomous sensors for the observation of natural phenomena. More precisely, we address the problem of using an underwater robot equipped of vision and sonar to produce maps of the spatial distribution of maerl, yielding indication of the concentration of living and dead material and of their mixing degrees. Instead of relying on a complete coverage of the area of interest, our approach is based on the use of probabilistic models (random closed sets) learned from a priori information, coupled with a random sampling strategy.

I Introduction

We present work in progress in the EU/IST project SUMARE[3] on the use of autonomous sensors for seabed habitat mapping. Autonomous sensors offer new possibilities for observation of natural environments, due to their ability to react to terrain ground-truth on-line, adapting the survey conditions to the characteristics of the surveyed region. Expected increased efficiency, simplicity of use and cost reduction offered by autonomous sensors may have a significant impact on the amount and quality of data that is collected, and consequently on the impact of regulating or conservation measures. Our goal is to demonstrate the importance of this new technology in two distinct applications of current interest: monitoring of sand banks in the North Sea, and mapping of maerl habitats in Orkney, Scotland.

In the paper, we focus on the last problem, assessing the use of autonomous sensors as a means for collecting information relevant for the study of the maerl distribution. More precisely, the goal of the project is to contribute to the understanding of the formation of living maerl patches and the accretion processes that lead to the presence of deposits of dead material.

[1] Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S), CNRS/UNSA, 2000 rte des Lucioles, BP 121, 06903 Sophia Antipolis cedex, France [tel: +33 4 92 94 27 14, fax: +33 4 92 94 28 96, email: {rolfes,rendas}@i3s.unice.fr]

[2] Department of Civil and Offshore Engineering, Heriot-Watt University, Orkney Campus, The Old Academy, Back Road, Stromness, Orkney, KW16 3AW [tel: 01856 850605, fax: 01856 85 1349, email: J.C.Side@hw.ac.uk]

[3] SUMARE: Survey of Marine Resources, project IST- 1999- 10836.

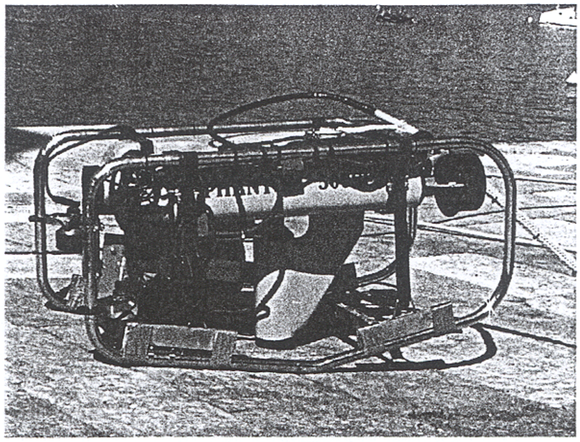

Figure 1: The ROV Phantom used in SUMARE.

In the project, field surveys will be conducted using an underwater vehicle,[4] see Figure 1, equipped of a wide suite of sensors including a Tritech profiler sonar and a SONY video camera, with the goal of producing maps of the distribution of areas of living, dead material and mixtures thereof. A first survey campaign has been done in the first week of August 2000, where a large number of maerl beds have been detected. For each one, a large volume of data has been recorded (video and sonar), with the goal of gathering learning data, for both the recognition algorithms required for automatic vehicle guidance and for the probabilistic models of maerl occurrence and spatial distribution. Once this learning step has been completed, another campaign will be conducted, with the vehicle in autonomous or semi-autonomous mode. Obviously, time and cost limitations preclude complete coverage of the entire Wyre Sound, i.e., the production of an exhaustive map of the entire region of interest to the study. Our goal with respect to this application is thus two-folded:

- Demonstrate the effectiveness of autonomous sensors as a means of intelligent (selective) observation of environmental phenomena, enabling the surveyor to *focus* on *areas* or *features of* special *interest*, for instance to track boundaries of regions of dead or living maerl, or to guide observations along regions of given geophysical parameters (constant depth, current strength, .)
- Demonstrate their ability to *infer* (coupled with appropriate models of the phenomena under study) *global characteristicsfrom local observations.*

In this communication we concentrate of the last goal: to show that autonomous sensors can be an effective mean to learn models of the possible spatial distribution patterns of a given species, and that they can later use these models on-line to recognize a given situation and extrapolate local information to tier global parameters.

The project's approach is based on the interplay of four components: modeling, signal/image processing, control and statistical inference.

Modeling will use the formalism of random closed sets, briefly presented on the next section, describing the spatial distribution patterns of maerl as a realization of a doubly stochastic process, whose control parameters statistical distributions will be learned using

[4] The ROV Phantom, produced by Deep Ocean Engineering, USA

existing a priori knowledge and field data from the first survey campaign. The problem of model learning is assessed in Section III.

During *the* actual *survey,* the vehicle uses vision/sonar to *incrementally update estimates of the control parameters of the stochastic model.*
This requires the on-line *discrimination* of the classes of interest. Preliminary work on images acquired in Orkney during August suggest that living maerl can be effectively identified from images taken at a safe distance from the bottom (around 3 meters), offering a good compromise between coverage efficiency and reliability of discrimination. The video segmentation algorithm used in these first studies is based on the technique of vector quantification, with the codewords being chosen such that the probability of confusing the two main classes of interest (living and dead maerl) is minimized [3].

Processed sensor data is used for adaptively guiding the platform, in two major ways. Firstly, it enables direct observation of important shape characteristics of the habitat (e.g. following boundaries of a class) : in this case, data is directly used to generate the control signals applied to the motors of the platform. Moreover, it enables, as discussed in Section IV, the update of estimates of the control parameters of the random closed set model.

II Modeling Maerl distribution as Random Closed Sets

Existing knowledge about the habitat of maerl [1] suggests a patchy spatial pattern. The frequency of occurence (the number of patches counted in an area) and the characteristic size of the patches are largely determined by the physical characteristics of the environment: the ocean current, depth and exposure ievel.

A maerl map can thus naturally be represented as a collection of compact sets Ξ_i (each one corresponding to a patch), spatially located at sites p_i

$$\Xi = \bigcup_{i=1}^{\infty} (\Xi_i + p_i).$$

In the previous equation, the individual compact sets Ξ_i describe the geometric properties of the patches, such as shape, size and density, and p_i indicates their spatial location. We use the theory of random closed sets (RCS) in order to build statistical models that describe the global properties of the collection of patches. See [2] for details on the theory of RCS, which are also called germ-grain models. These models describe complex spatial patterns by using a doubly stochastic mechanism. A first random point process p describes the spatial locations (germs) at which, realisations of a second stochastic process (grains) determine the local morphology of the field, i.e., the characteristics of the sets Ξ_i .

The germ model (point process) characteristics define the global dispersion of the field; which can in particular be clustered, regularly structured or uniformly distributed. We assume that the distribution of the point process belongs to a parametrized family, by considering that the associated counting measure μ is a member of G_p:

$$\mu \in G_p = \{\mu_\lambda, \lambda \in \Lambda\},$$

denoting by λ the collection of parameters that determine the statistical distribution of the locations p_i.

The grain model (shape process) constrains the possible elementary shapes, for instance to circles or lines of random size and orientation. As for the germ model, we consider also that the distribution of the shape process can be parametrized by a finite number of parameters, if required using mixture models to describe all possible appearances of the maerl patches, such that

$$p(\Xi) \in G_\Xi = \{p_\gamma(\Xi), \gamma \in \Gamma\}.$$

Once the model type (e.g. the families of distributions G_p and G_Ξ) and its parameters $\theta = (\lambda, \gamma)$ are known, the moments (average, variance) of the overall surface covered by the patches inside the area of interest can be estimated over any area of interest. In the project we try to exploit this interesting property to estimate the amount of living and dead material over extended areas. From local observations we will estimate the process control parameters θ, and use them to infer global maerl distribution over surrounding regions. The main advantage of this approach is that we don't need to map each and every patch individually. Instead, we just need to gather enough observations to accurately estimate the global characteristics of the field expressed by the low dimensional parameter vector θ.

We assume that the process control parameters θ depend on the physical characteristics of the environment at each location p, which we denote by $\varphi(p)$. This is in agreement, as we said before, with existing studies that suggest that maerl occurrence can be predicted using a reduced set of local indicators (depth, current strength and exposure level)

$$p \to \varphi(p) \to \theta(\varphi(p)).$$

We assume that over conveniently defined neighborhoods, the dependency of θ on φ can be modelled by a function

$$\theta(\varphi) = g(\varphi, \alpha)$$

dependent on a set of parameters α, whose values are drawn from a known distribution $p(\alpha)$. In this way, we can capture the field variations between areas of similar geophysical characteristics.

Note that knowing θ, i.e., the RCS model control parameters, is equivalent to knowing the parameters a. We will call α the structural parameters of the field.

It is a fundamental result on the theory of random closed set models that knowledge of their underlying distributions (germ and grain models) is completely equivalent to the specification of their ***hitting probabilities.*** The hitting probabilities specify, for each possible closed set Φ the probability of non-empty intersection with the spatial process Ξ:

$$T_{\theta(\varphi(p))}(\Phi) = \mathrm{Prob}_{\theta(\varphi(p))}(\Xi \cap \Phi \neq \varnothing).$$

To compute the right-handside of this equation, we must **know** θ, or, equivalently, the structural parameters α. We will **explicitely** indicate the dependency of the hitting probabilities on a by writing

$$T_{\theta}(\Phi) = T_{\theta(\alpha)}(\Phi).$$

The parameter of interest, the amount of maerl in a region A (compact set), S_{maerl}, can thus be obtained by integrating the hitting probability over the whole surface

$$E[S_{maerl}] = \int_A T_{\theta(\alpha)}(p)dp.$$

These probabilities can be empirically estimated **from** a large number of observations of Ξ, enabling subsequent estimation of the structural parameters α, which define the spatial dependency of the RCS model control parameters θ.

The modeling approach describe above considers only first-order **characterisation** of the field process, its spatial coherence being imposed by the function *g()* and the structural parameters *a*. A more complex statistical modeling approach would consider θ to be a stochastic (spatial) stochastic process, and impose constraints in its-spatial correlation However, we think that the approach chosen here is well adapted to biological studies, where *an* ***explanation*** *model* of a given phenomenon is sought, as a function of a number of explanatory variable (here the geophysical parameters φ). Note that if these parameters are indeed **sufficient** for predicting the occurrence and appearance of maerl, 'then the entropy of the random variable a should be small. On the contrary, a large entropy of this random variable indicates that φ is not **sufficient**, and that other parameters should be taken into account in the model.

III Model learning

Learning the distribution of maerl requires learning of *(i)* the model type, i.e., the families of distibutions G_p and G_Ξ and of the structural parameter *a* and *(ii)* learning of the corresponding process structural parameters, which determine the intensity λ function describing the spatial distribution of the parameters y of the distribution of the morphological characteristics of the patches.

We will consider two candidate modeling hypotheses.

Under the first one we assume that individual patches can be distinguished, enabling separate estimation of the intensity λ and of the distribution of the morphological characteristics.
Namely, we consider that G_p is the general **familiy** of non-stationary Poisson processes, with the intensity parameter being a smooth function of $\varphi(p)$ at each point p. The description of the shape of each patch will use the theory of shape introduced by Grenander, which enables to describe a priori shape models by defining a distribution of graphs that describe the interconnection of elementary shape forms.

Under the second modeling hypothesis we do not assume that clearly identification of each patch is possible. We describe the point process as a clustered process: each cluster center (which is itself a Poisson process with intensity parameter l_c, is the origin of a second non-isotropic Poisson process, with intensity described by l_p. In this case,the shape process can be very simple, for instance, circles of random circle (with the radius being a random variable with known a priori density).

Note that while in the first approach all the complexity of shape description is contained in the shape process, the second translates this responsibility to the definition of the local intensity of each cluster.

For both modeling approaches, the RCS model control parameters will be, as said previously, expressed at each point as a function of its geophysical characteristics φ. Learning the field model, when the model type has been fixed, amounts thus to learn the funcitonal form of $g()$ and the statistical distribution of the parameters α. To learn this distribution, we will use the the observations of an significant part of the Orkney seabed collected in the August survey campaign. This first model will of course be updated with new observations acquired during subsequent survey missions.

Using video mosaicing algorithms [3] developped at IST, Lisbon, Portugal (one of the SUMARE partners) a sequence of images recorded by the camera of the autonomous vehicle while moving over the ocean bed are 'glued' together by associating features of the partly overlapping regions. We obtain in this way a global image of a large portion of the seabed. Automatic image processing (initially supervised by an expert) performs a segmentation of the image (using vector quantification method [4]) into areas covered by dead and living maerl. It is this segmented image, together with knowledge of the physical characteristics of the observed areas, that will be used to estimate $g()$ and $p(\alpha)$.

Alternatively, we can use the ability of the vehicle to track the boundaries between living and dead material, producing in a more efficient way the global map of maerl patches inside a given area. This approach could not be used in Orkney, since one of the goals of this campaign was to acquired learning data for the classification algorithms that will automatically discriminate them automatically, but it will developped and applied in the future.

Work with simulated images has already validated our approach showing that we can identify the strctural parameters from partial observations of process realisations. The work is now in progress using the data acquired at Orkney.

IV Habitat mapping

In the previous section we addressed the problem of learning, probabilistic models for the statistical distribution of the intensity and shape parameters of random closed set models, as a function of the physical characteristics of the region observed.

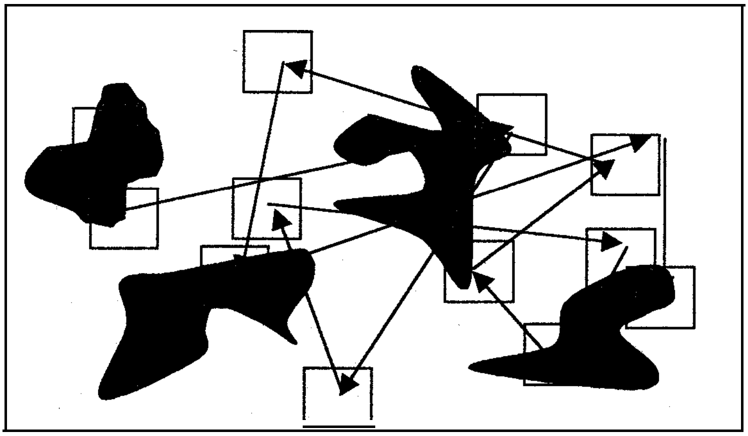

Figure 2: Random sampling.

In this section we consider that such models exist, and discuss how they can be used to efficiently produce approximate maps of the maerl distribution over extended regions. We stress that our maps are not of a common type: they do not indicate the detailed localisation of each individual maerl patch. Instead, they associate, to each point of the seabed, a value of the structural parameter a (and thus of the RCS model control parameters θ). We believe that the information they convene is more directly related to the goals of marine scientists, directly mapping indicators of spatial dispersion and agregation degree. Moreover, they can be used as a means of verifying explanatory models, by using entropy of the distribution of a as a means to assess the sufficieny of the base parameters φ.

During actual survey, the vehicle covers an extended area, constantly computing the compatibility of the observed pattern as a possible realization of the (learned) statistical model. This comparison uses ancilliary information about the physical characteristics of the region and the apriori model described in the previous section.

To be able to directly estimate the hitting probabilities of the field, a completely random sampling strategy should be adopted: the vehicle should randomly wander in the observed area, randomly determining the next observed window, see Figure 2. Assuming these probabilities can be estimated, our mapping problem can then be formulated as the one of estimating, from the partial observations inside the square sub-regions in that figure, the structural parameters of the random closed set that would, with largest probability, yield the (partial) observed pattern.

We estimate a using a Bayesian approach by maximizing. the a posteriori density, or equivalently:

$$\hat{\alpha} = \max_{a} p(I_1, \ldots, I_n | \theta(\alpha)) p(\alpha),$$

where the observations I_i are the images obtained at locations p_i. The a priori density of a is given by the learned a priori model. If the images are obtained at random locations in the region to be surveyed, such that the observations are independent, we obtain an unbiased estimation of the hitting probability, and thus a good estimate of a. *The* images as basic observations can then be replaced by a set binary indicators at a set of random locations p_i inside each square frame, indicating, whether or not maerl was observed at location p_i

$$f_{hit}(p_i) = \begin{cases} 1; p_i \cap \Xi \neq \varnothing \\ 0; p_i \cap \Xi = \varnothing \end{cases}.$$

In this case we obtain an estimation of the process control parameter by

$$\hat{\theta} = \max_{\theta} p(f_{hit}(p_1), \ldots, f_{hit}(p_n) | \theta(\alpha)) p(\alpha).$$

The actual random generation of the next visited (observed) region, implying a chaotic motion of the vehicle between points is certainly not the most efficient strategy for sampling a given region. Alternatively, we can base the computation of the hitting probabilities on rectangular areas, effectively using the ability of the vehicle to continuously observe an extended area.

We stress that the sample trajectory (e.g. a set of connected straight lines) (see figure 3) must.. be chosen at random, in order to avoid observation of 'favorite' regions. It is important that the observations do not be concentrated on areas where maerl is expected. Negative information is as important for the estimation of the process complete model as positive information.

In order to avoid violation of the independence assumption of the observations we need to choose locations p_i (the locations at which the hitting probabilities are estimated) at random along this sampling path, rather than using the whole area that is observed by the vehicle.

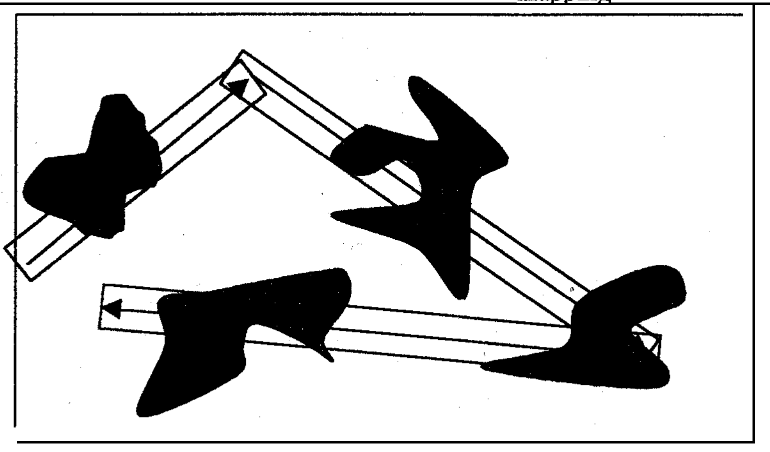

Figure 3: sampling using straight lines.

V Conclusions

We presented in this paper work in progress in the European project SUMARE, concerning the use of autonomous sensors for study of the distribution of underwater ressources, focusing on one of the applications considered in the project: mapping of maerl in Orkney. We discussed the advantages that provide an autonomous sensor, allowing periodical mapping of natural phenomena, such that evolution of the phenomena can be observed permanently. An important advantage of autonomous sensors is augmented efficiency (observations can be done permanently), while costs are reduced. We propose an approach that uses the theory of random closed sets, in order to model the distribution of maerl. The process control parameter will be estimated based on learned a priori information (obtained from large images of the seabed, taken in August). This a priori information will be updated throughout following observation campaigns during that the vehicle acquires images along a random sample path, covering a large portion of the field. Habitat mapping done by an autonomous sensor has the additional advantage, that the information gathered along this sampling path contains negative as well as positive information obtained at random locations, while human observers tend to map mainly areas where maerl is found, such that the estimated distribution of the maerl is just reliable in these observed areas.

Acknowledgments

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

References

[1] Nicola White , 'The distribution of maeri in Orkney' Master Thesis, Department of Civil and Offshore Engineering Heriot-Watt University, Orkney; 1999

[2] Bardorff-Nielsen, O.E.; Kendall, W.S.; van Lieshout, M.N.M.; 'Stochastic Geometry, Likeleehood and Computation'; Chapman & Hail/CRC; 1999

[3] Nuno Gracias, 'Application of robust estimation to computer vision: Video mosaicing and 3-D reconstruction' Master Thesis at 'Universidade Tecnica de Lisboa', 1997

[4] K. 0. Perlmutter et *al*, Bayes Risk Weighted Vector Quantization with Posterior Estimation for Image Compression and Classification, IEEE Transactions on Image Processing, Vol 5, No 2, Freb 19%.