Monitoring Marine Environments using a Team of Heterogeneous Robots

Yogesh Girdhar[1], Anqi Xu[1], Florian Shkurti[1], Juan Camilo Gamboa Higuera[1], Malika Meghjani[1], Philippe Giguère[2], Ioannis Rekleitis[1], Gregory Dudek[1]

***Abstract*—We present a novel approach for monitoring marine environments by a team of heterogeneous robots, comprising of a fixed-wing aerial vehicle, an autonomous airboat, and a legged underwater robot. The goal is to receive a region of interest from a remote human supervisor, and then using the coordinated effort of the robot team, produce a concise summary consisting of a small number of images, which capture the visual diversity of the region of interest. The summary could then be used by a human supervisor to plan for further exploration.**

I. Introduction

Monitoring marine environments is a challenging problem for several reasons. Being underwater for long periods of time is inherently dangerous for humans, and there are strong limitations on duration and depth of dives. Even if an under water robot is used, the lack of high bandwidth wireless communications underwater makes it difficult to remotely operating a vehicle underwater. Furthermore, many of the regions which need monitoring are far away from civilization, which makes continuos human presence around the region of interest very expensive.

We propose a solution to these problems by using a team of heterogeneous robots, which first helps in identifying a region of interest (ROI) by remote human supervisor, and then given the location of this ROI, returns a summary consisting of a small number of images that include any surprises that were encountered. This system addresses drawbacks of conventional reef monitoring methods [1] by automating the data collection process. Our work is complementary to work by Smith et al. [2], who have proposed a path planning technique for underwater gliders, given some regions of interest.

Our multi-robot system takes inspirations from related works in automated underwater data muling [3], target tracking using aerial and surface robotic vehicles [4], and marine monitoring using multiple heterogeneous robots [5, 6].

As shown in Fig. 2, our heterogeneous robot team is comprised of three vehicles that operate in diverse domains: the Aqua AUV platform [7], the Unicorn fixed-wing UAV, and the Marine Autonomous Robotic Explorer (MARE) [8] catamaran ASV.

During a monitoring session, our UAV carries out repeated aerial coverage [9] of the target reef region, and sends live bird's-eye view footage of the coral reef to the home base. Next, particular inspection sites are identified based on this live footage, either using our automated extremum summary algorithm, or obtained from human experts. These coordinates are then relayed to Aqua AUV via Mare ASV, which then proceeds to build a summary of the region of interest, and relay it back to the human supervisor. Building a good summary is an essential part of the system as wireless bandwidth is often the most limiting constraint while exploring remote regions. A summary consisting of surprising observations could be quickly used to determine whether a given region should be explored further or not.

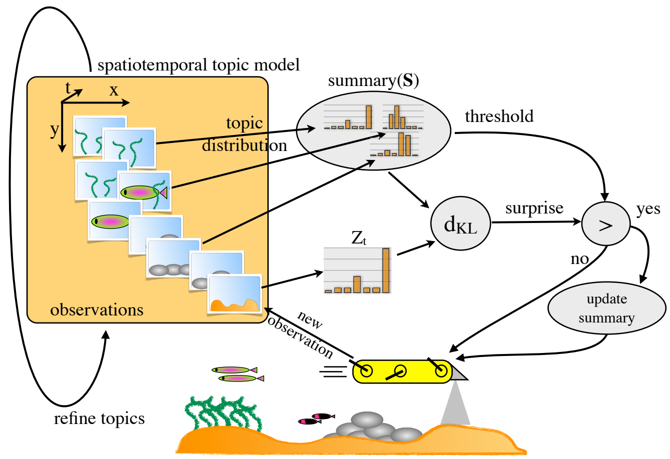


Fig. 1. *Summarizing complex marine environments*. We use a realtime online topic modelling framework, which describes the incoming observations using a low dimensional semantically sensitive descriptor. Then, using an online summarizer we build a summary which covers all observations this far, in the topic space. Surprise of a new observation is its distance to the closest sample in the summary. If this surprise is above a threshold we update the summary with the new observation.

Figure 1 gives an overview of the proposed summarization strategy. The online summarization task can be broken down into two sub problems: first, how can one describe what is being observed by the robot in a meaningful way, which is sensitive to thematic scene changes, while being immune to low level noise in the sensor data; and second, given a semantic description of the observations, how can one decide if the observation is surprising and should be part of the summary.

For the first problem, we use an online topic-modeling framework to describe the world being observed by the robot in realtime, using a low dimensional descriptor, which attempts to be sensitive to semantics of what is being observed. This descriptor is a probability distribution over the presence of various objects in the scene, which in an underwater environment might correspond to a distribution over rocks, different

[1]School of Computer Science, McGill University, Canada.
[2]Département d'informatique et de génie logiciel, Université Laval, Canada

(a) Aqua AUV (b) Unicorn UAV (c) MARE ASV

Fig. 2. Our robot team consists of (a) the Aqua AUV, an agile legged underwater robot, (b) the Unicorn, a fixed-wing UAV with an on-board autopilot and gimbal-mounted camera, and (c) the Marine Autonomous Robotic Explorer (MARE) catamaran ASV, which can operate in turbulent open water regions.

coral species, and sand etc... Given such a descriptor, we then use an online summarization algorithm which tries to identify samples, which cover the set of all observations, in this low dimensional semantic space. Since all observations must be covered, the summarizer encourages outliers to be part of the summary, which can be interpreted as surprising observations.

II. The Robot Team

Our robot team consist of three vehicles: a fixed wing unmanned aerial vehicle (UAV), an autonomous surface vehicle (ASV), and an underwater vehicle (AUV).

The UAV in the heterogeneous robot team is first used for scouting out regions of interest, and then for relaying this location information to the MARE ASV. Mare ASV and the Aqua AUV then move to target location, and start the survey. The ASV or the UAV can only maintain radio link with Aqua while it is on the surface. Hence, the ASV acts as buffer for locations of regions to be explored, and relays them to Aqua whenever it is on the surface. Since both the ASV and the AUV use very little power, while they are not actively moving, they can be left in the field for days without recharging the batteries.

A. Unicorn UAV

The Unicorn UAV is a kite-sized fixed-wing aerial vehicle with a 1 m wingspan. It operates at an average ground speed of 14 m/s for up to 30 minutes of flight time. This vehicle is equipped with multiple sensors, including an IMU, a GPS, and pressure-based altitude and speed sensors. These devices are integrated with an on-board autopilot micro-processor, which uses them to navigate autonomously based on waypoint directives issued from the home base. Communication between the autopilot and the home base is established using a high-power radio modem, which allows the UAV to be controlled at multi-kilometer ranges. The Unicorn is also equipped with a CCD camera mounted on a pan-tilt gimbal, which allows the home base to receive live aerial feed through an analog radio frequency channel.

B. MARE ASV

The Marine Autonomous Robotic Explorer (MARE) [8] is a robotic airboat developed to explore coral reefs and shallow seabeds. Its two-pontoon catamaran chassis provides sufficient hydrodynamic stability to operate in turbulent open water environments. MARE is actuated using two air propellers in a differential drive configuration. MARE is capable of conducting autonomous visual exploration [8] using its suite of sensors, which includes a downward-facing camera, an IMU, and a GPS device.

C. Aqua AUV

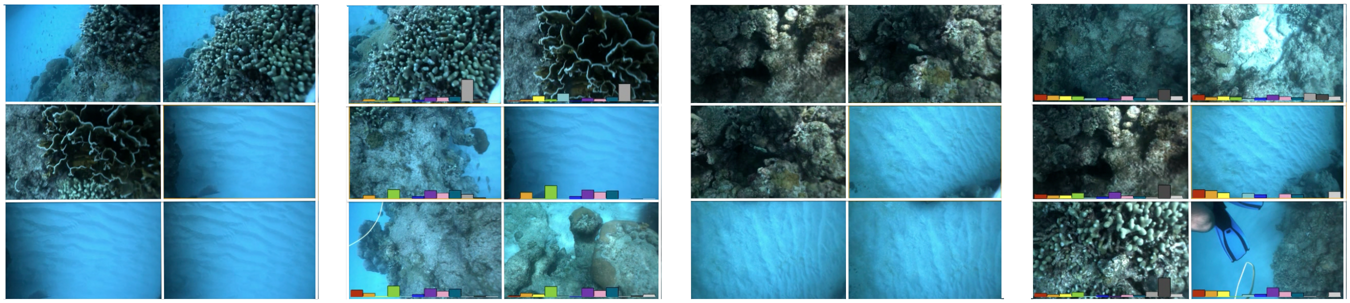
The Aqua AUV [7] is a six-legged amphibious robot that can both swim underwater and walk on land. It maneuvers underwater by synchronously actuating its six flippers, and its aluminum shell is designed to operate at depths up to 40 m. This AUV is powered by high-capacity Lithium-Ion batteries, and can operate under full load underwater for more than five hours. Aqua is equipped with a variety of sensors within its waterproof shell, including: three cameras, an inertial measurement unit (IMU), a pressure-based depth sensor, and an XBee digital radio transceiver. We also used an externally-mounted sensor kit to facilitate wireless communications, since the AUV's metallic shell acts as a Faraday cage and thus greatly attenuates the transmission of radio signals. In particular, this detachable and self-powered pack contains a GPS receiver and an XBee module, which are used to augment Aqua's sensing capabilities by relaying GPS readings and transmission from the MARE ASV wirelessly using the XBee communication channel.

III. Online Extremum Summaries

Although are many ways in which observations made by a robot could be summarized [10, 11, 12], we are interested in online summaries, which cover the range of what has been observed, including outliers. Online extremum summaries [13] provides a novel way to maintain a set of observations which cover the space of all observations, while minimizing the radius. Given a summary $\mathbf{S} = \{S_1, \cdots, S_k\}$, and all observations so far $\mathbf{Z} = \{Z_1, \cdots, Z_n\}$, extremum summaries minimize the cost function

$$\mathrm{C}ost(\mathbf{S}|\mathbf{Z}) = \max_i \min_j d(Z_i, S_j), \tag{1}$$

where d is the distance function, which measures distance as the symmetric KL divergence between the corresponding topic distributions. Minimizing this distance is essentially minimizing the distance of the worst outlier to the summary. The novelty or surprise of a new observation $\xi(Z_t)$ is then defined as its Hausdorff distance from the summary. When



(a) Dataset 1 - bag of words (b) Dataset 1 - topics (c) Dataset 2 - bag of words (d) Dataset 2 - topics

Fig. 3. Example of summaries generated by our system. We see that bag-of-words descriptor (a,c) is not able to differentiate between different images of sand, and hence the corresponding summaries contain similar images. The proposed topics based summaries (b,d) do not have such problems.

the summary size is fixed, the threshold surprise score for inclusion in the summary is $\min_i d(S_i, \mathbf{S}_{-i})$, where $\mathbf{S}_{-i}$ is the summary with ith sample removed.

IV. Topic Modeling

We use an online topic modelling framework [14] to represent observations made by a robot in a low dimensional thematic space. Topic modeling methods were originally developed for text analysis. Hofmann [15] introduced the idea of probabilistic Latent Semantic Analysis (PLSA) for text documents. PLSA models the probability of a observing a word w_i in a given document d_m as:

$$\mathbb{P}(w_i|d_m) = \sum_{k=1}^{K} \mathbb{P}(w_i|z_i = k)\mathbb{P}(z_i = k|d_m), \tag{2}$$

where z_i is the hidden variable, or topic label for w_i, which takes a value between $1 \cdots K$.

The central idea being the introduction of a hidden (latent) variable z, which models the underlying topic, or the context responsible for generating the word.

Bei et al. [16] developed the idea of pLSA further and introduced Latent Dirichlet Allocations (LDA). LDA proposes improvement over pLSA by the use of Dirichlet priors for both document and topic models. Success of LDA based topic modeling methods for semantic clustering and classification of text documents has led to their use in computer vision domain. The general idea being that a textual word could be replaced by visual words, such as ones described by Sivic et al. [17]. Works by Fei Fei et al. [18], have demonstrated the use of LDA to model visual scenes.

To make topic-modeling relevant for robotics use, we depart from the traditional idea of modeling the data as a discrete set of documents, and instead model the data as points or words in a continuous space and time. This allows for a more natural interpretation of topics as context, regions, or objects with continuity in space and time.

Each observations is a tuple (w_i, x_i) consisting of an observed visual word $w_i = 1 \cdots V$ from a fixed vocabulary of size V, and its associated spatiotemporal coordinate x_i. Dimensionality of the coordinates depend on the problem and the hardware setup. We observe this data in a streaming manner, and our goal is then to compute the topic labels $z_i = 1 \cdots K$, for each of these incoming observations, and update the labels of previous observations in the light of this new incoming data. For the system to work in real-time, we must guarantee that the update iterations happen in a constant time.

Compared to the standard document-topic generative model used by LDA, we replace the idea of documents with neighborhoods. We assume the following generative model. Given a location x, we define its spatiotemporal neighborhood by $G(x)$. The visual appearance of the location is then produced by independently sampling a topic z for each word in its neighborhood $\theta_{G(x)}$, and then independently sampling the word from ϕ_z.

The neighborhood $G(x)$ of an observation can either be defined using a k-nearest neighbor search or a radius search. Hence, the posterior distribution of a topic assignment, given the Dirichlet prior and the other assignments is

$$\begin{aligned} \mathbb{P}(z_i = k|\mathbf{z}_{-i}, \mathbf{w}, \mathbf{x}) \propto \frac{n^t_{k,-i} + \beta}{\sum_{t=1}^{V}(n^t_{k,-i} + \beta)} \cdot \\ \frac{n^k_{G(y),-i} + \alpha}{\sum_{k=1}^{K}(n^k_{G(y),-i} + \alpha)}. \end{aligned} \tag{3}$$

Several different strategies exist in the literature to do online refinement of the topic assignment in a given streaming dataset [19]. The general idea is to initialize the topic label of the current observation with random labels, and then do a batch refine of the entire dataset. This allows for previous topic assignments to be updated in the light of new observed data. Convergence is guaranteed because in the limit of time going to infinity, the algorithm behaves like a batch Gibbs samples.

Since we have a fixed amount of time between two observation, we can only do a constant number of refinements. Hence, we used a biased Gibbs sampler, which randomly picks spatiotemporal regions, giving higher picking probability to recently observed regions, and then refines their topic labels.

V. Results

Figure 3 shows examples of extremum summaries generated using the proposed topics based descriptor, by the Aqua AUV. For comparison, we show extremum summaries produced by using the classic bag-of-words descriptor. We see that topics based extremum summaries clearly outperform the bag-of-words based summaries, as they are better able to capture visual appearance of the environment. Bag-of-words based extremum summaries contain multiple images of sand, which although have very different descriptors, do not correspond to different semantics content. Topics based descriptor is not confused by such images.

VI. Conclusion

Continuous monitoring of remote marine environments is a challenging task for both humans and robots. Constraints related to battery life, bandwidth, and multiple environments make it difficult for a single robot to monitor a remote underwater region for long periods of time. We have proposed a novel use of a team of heterogeneous robots, which is efficiently able to aid a remote human supervisor in identifying and inspecting an underwater region such as a coral reef. In our trials we see that the proposed summarizer is able to recognize different coral species and include them in the summary, and hence is useful for marine monitoring. In our ongoing future work, we hope to design a combined topic model that learns correlated topics using observations from different robots. Such a model would be useful, for example, to match an underwater view of a location to its aerial view.

References

[1] C. Rogers, G. Garrison, R. Grober, Z. Hillis, and M. Frankie, "Coral reef monitoring manual for the caribbean and western atlantic," *Virgin Islands National Park, 110 p. Ilus.*, 1994.

[2] R. N. Smith, M. Schwager, S. L. Smith, B. H. Jones, D. Rus, and G. S. Sukhatme, "Persistent Ocean Monitoring with Underwater Gliders: Adapting Sampling Resolution," *Journal of Field Robotics*, vol. 28, no. 5, pp. 714–741, 2011.

[3] M. Dunbabin, P. Corke, I. Vasilescu, and D. Rus, "Data muling over underwater wireless sensor networks using an autonomous underwater vehicle," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA '06)*, 2006, pp. 2091–2098.

[4] M. A. Hsieh, A. Cowley, J. F. Keller, L. Chaimowicz, B. Grocholsky, V. Kumar, C. J. Taylor, Y. Endo, R. C. Arkin, B. Jung, D. F. Wolf, G. S. Sukhatme, and D. C. MacKenzie, "Adaptive teams of autonomous aerial and ground robots for situational awareness: Field reports," *J. Field Robot.*, vol. 24, no. 11-12, pp. 991–1014, 2007.

[5] J. Das, F. Py, T. Maughan, T. O'Reilly, M. Messie, J. Ryan, K. Rajan, and G. Sukhatme, "Simultaneous Tracking and Sampling of Dynamic Oceanographic Features with Autonomous Underwater Vehicles and Lagrangian Drifters," in *Intnl. Symp. on Experimental Robotics (ISER)*, New Delhi, India, 2010.

[6] G. Podnar, J. Dolan, A. Elfes, S. Stancliff, E. Lin, J. Hosier, T. Ames, J. Moisan, T. Moisan, J. Higinbotham, and E. Kulczycki, "Operation of robotic science boats using the telesupervised adaptive ocean sensor fleet system," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA '08)*, 2008, pp. 1061–1068.

[7] G. Dudek, M. Jenkin, C. Prahacs, A. Hogue, J. Sattar, P. Giguère, A. German, H. Liu, S. Saunderson, A. Ripsman, S. Simhon, L. A. Torres-Mendez, E. Milios, P. Zhang, and I. Rekleitis, "A visually guided swimming robot," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2005)*, Edmonton, Alberta, Canada, August 2005, pp. 3604–3609.

[8] Y. Girdhar, A. Xu, B. B. Dey, M. Meghjani, F. Shkurti, I. Rekleitis, and G. Dudek, "MARE: Marine Autonomous Robotic Explorer," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS '11)*, San Francisco, USA, 2011, pp. 5048–5053.

[9] A. Xu, C. Viriyasuthee, and I. Rekleitis, "Optimal complete terrain coverage using an unmanned aerial vehicle," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA '11)*, Shanghai, China, May 2011, pp. 2513–2519.

[10] A. Ranganathan and F. Dellaert, "Bayesian surprise and landmark detection," in *Proceedings of the 2009 IEEE international conference on Robotics and Automation*. Institute of Electrical and Electronics Engineers Inc., The, 2009, pp. 1240–1246.

[11] K. Konolige, J. Bowman, J. D. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, "View-based maps," *RSS'09*, 2009.

[12] R. Paul, D. Rus, and P. Newman, "How was your day? Online Visual Workspace Summaries using Incremental Clustering in Topic Space," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[13] Y. Girdhar and G. Dudek, "Efficient on-line data summarization using extremum summaries," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[14] Y. Girdhar, P. Giguère, and G. Dudek, "Autonomous Adaptive Underwater Exploration using Online Topic Modelling," in *International Symposium on Experimental Robotics*, 2012.

[15] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, no. 1, pp. 177–196, 2001.

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[17] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *In Proc. ICCV*, 2003, pp. 1470–1477.

[18] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, Jun. 2005, pp. 524 – 531 vol. 2.

[19] K. R. Canini, L. Shi, and T. L. Griffiths, "Online Inference of Topics with Latent Dirichlet Allocation," *Proceedings of the International Conference on Artificial Intelligence and Statistics*, vol. 5, no. 1999, pp. 65–72, 2009.