Visual summaries for low-bandwidth semantic mapping with autonomous underwater vehicles

Visual Summaries for Low-Bandwidth Semantic Mapping with Autonomous Underwater Vehicles

Jeffrey W. Kaeli[1], John J. Leonard[2], and Hanumant Singh[1]

Abstract— **A fundamental problem in autonomous underwater robotics is the high latency between the capture of image data and the time at which operators are able to gain a visual understanding of the survey environment. Typical missions can generate imagery at rates orders of magnitude greater than highly compressed images can be transmitted acoustically, delaying that understanding until after the robot has been recovered and the data analyzed. We present modifications to state-of-the-art online visual summary techniques that enable an autonomous robot to select representative images to be compressed and transmitted acoustically to the surface ship. These transmitted images then serve as the basis for a semantic map which, combined with scalar navigation data and classification masks, can provide an operator with a visual understanding of the survey environment while a mission is still underway.**

I. INTRODUCTION

Seventy percent of the Earth's surface is covered by water, below which lie diverse ecosystems, rare geological formations, important archeological sites, and a wealth of natural resources. Understanding and quantifying these areas presents unique challenges for the robotic imaging platforms required to access such remote locations. Low-bandwidth acoustic communications prevent the transmission of images in real-time, while the large volumes of data collected often exceed the practical limits of exhaustive human analysis. As a result, the paradigm of underwater exploration has a high *latency of understanding* between the capture of image data and the time at which operators are able to gain a visual understanding of the survey environment.

A robotic vehicle capturing one still image every few seconds can easily generate thousands of images within a matter of hours. This sheer volume of data presents a formidable obstacle to any individual attempting to gain an understanding of the survey environment. Often, when a vehicle operator obtains a dataset for the first time, their instinct is to quickly scan thumbnails of the images for any that "pop out." While this can be useful, it is not necessarily the best or fastest way to obtain images that "represent" the data in a meaningful way. In this work, we explore the use of navigation summaries [1], [2], [3] to obtain a small subset of images that can serve as the basis for low-bandwidth semantic maps to give an operator a fast, high-level understanding of the survey environment while a mission is still underway.

[1]Jeffrey W. Kaeli and Hanumant Singh are with the Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA {jkaeli,hsingh}@whoi.edu

[2]John J. Leonard is with the Massachusetts Institute of Technology, Cambridge, MA 02139, USA jleonard@mit.edu

II. RELATED WORK

A. Underwater Communications

Without a physical link to the surface, AUVs rely on acoustic signals to communicate with shipboard operators. These channels have very limited bandwidth with throughput on the order of tens of bytes per second depending on range, packet size, other uses of the channel (for instance, navigation sensors), and latencies due to the speed of sound in water [4], [5]. While much higher data rates have been achieved using underwater optical modems for vehicle control [6] and two-way communication [7], these systems are limited to ranges on the order of 100 meters and are inadequate for long-range communication [8]. In the absence of mission-time operator feedback, an AUV must either navigate along a preprogrammed course or use the data it collects to alter its behavior. Examples of the latter, termed adaptive mission planning, include detecting mines so potential targets can be re-surveyed in higher-resolution [9] and using chemical sensors to trace plumes back to their source [10], [11]. The overarching implication is that, with the exception of low-bandwidth status messages, data collected by an AUV is not seen by operators until after the mission is completed and the vehicle recovered.

B. Clustering Data

Clustering can be viewed as an unsupervised compression strategy that allows multidimensional data to be quantized to one of several discrete distributions by defining a distance metric between samples and minimizing some measure of that distance. We can think of each image as a data point characterized by some distribution of features, such as a quantized descriptor (which itself could have been obtained through clustering). One of the most well-known clustering algorithms is the K-means algorithm which seeks to find a set of cluster centers that minimize the within-class distances between each cluster center and the members of its representative class [12]. While this method has been extremely useful in generating texton dictionaries for texture analysis [13], [14], the fact that the cluster centers are not guaranteed to occur at a data point makes mapping back to a single representative image for each class difficult. A similar algorithm, k-medoids, only considers data points as potential cluster centers, and is more useful for generating representative images. Both of these methods require the number of cluster to be set *a priori*.

Other methods seek to determine the number of clusters based on the natural structure of the data. Affinity propagation accomplishes this by picking "exemplars" that are suggested by nearby data points [15] and has found use in building texton vocabularies [16]. Hierarchical methods have also been used to learn objects [17], scenes [18], and underwater habitats [19] based on topic models using Latent Dirichlet Allocation (LDA) [20]. However, a drawback of all methods mentioned thus far is that they operate upon a static dataset. This "offline" approach is ill-suited to real-time robotic imaging because it offers no way to characterize the dataset until after all the data has been collected.

Clustering data in an "online" fashion provides two important benefits. Firstly, it allows data to be processed continuously throughout the mission, reducing the overall computational load. Secondly, at any point in time it provides a summary of the imagery captured thus far by the vehicle. A drawback to online methods is that they offer less guarantees of stability and are ultimately dependent upon the order in which images are presented to the algorithm [21]. The worst-case scenario for online approaches would be for the most extreme data points to occur first, followed by interior points which become poorly represented. Luckily, natural underwater environments are highly redundant with habitat domains that persist across many frames. One possible approach uses incremental clustering of topic models using LDA [3]. We are particularly interested in recent work on navigation summaries [1], [2] which operate on the concept of "surprise."

C. Surprise-Based Summaries

An event can be said to be "surprising" because it happens unexpectedly. The idea of what is expected can be modeled as a probability distribution over a set of variables and considered as *prior* knowledge about the world. When a novel event occurs, it augments this body of knowledge and creates a slightly different *posterior* knowledge of the world. If the amount of knowledge added by any single event is large enough, that event can be said to be unexpected and thus is "surprising."

This concept has been formalized in a Bayesian framework as the difference between the posterior and prior models of the world [22]. For measuring this difference, the Kullback-Leibler divergence, or relative entropy, was shown to correlate with an attraction of human attention,

$$d_{KL}(p \parallel q) = \sum_x p(x) log \frac{p(x)}{q(x)} \tag{1}$$

where $p(x)$ is the posterior model, $q(x)$ is the prior model, and x is some observed variable over which distributions can be computed. Rather than modeling the prior knowledge Π^- as a single distribution $P(F)$ over a set of features F, we follow [1] and model it over each member of summary set $\boldsymbol{S}$ containing M members.

$$\Pi^- = \{P(F|S_1), \cdots P(R|S_M)\} \tag{2}$$

The posterior knowledge Π^+ is simply the union of prior knowledge with the new observation Z

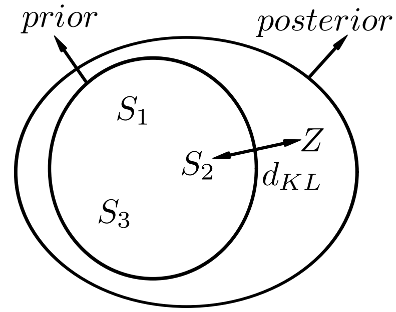

Fig. 1. Visual illustration of surprise, defined as the distance of the nearest member of set $\boldsymbol{S}$ to Z measured by the KL divergence. The surprise is a measure of information gain from a new observation based on a summary set of existing hypotheses.

$$\Pi^+ = \{P(F|S_1), \cdots P(R|S_M), P(F|Z)\} \tag{3}$$

The set theoretic surprise ξ can be defined as the Hausdorff distance between the posterior and prior distribution using the KL divergence as a distance metric [1]. The Hausdorff metric is a measure of the distance between two sets based on the greatest possible difference between one point in the first set to the nearest point on the other sets. Since the prior and posterior sets differ only by Z, the surprise can be simply expressed as the KL distance between observation Z and the nearest summary image in $\boldsymbol{S}$. This distance is illustrated graphically in Figure 1.

$$\xi(Z|\boldsymbol{S}) = \inf_{\pi^- \in \Pi^-} d_{KL}\left(P(F|Z) \parallel \pi^-\right) \tag{4}$$

When a new observation's surprise exceeds a threshold, it is added to the summary set. The threshold is generally set as the lowest value of surprise in the current summary. That member of the old summary set with the lowest surprise is then removed and replaced by the new observation, and the surprise threshold set to the next least-surprising member of the summary set. In this manner, a temporally global summary of the images is maintained at all times [1].

III. MODIFIED VISUAL SUMMARIES FOR SEMANTIC MAPPING

A. Basic Implementation

We implemented the aforementioned navigation summary on a 3000+ image dataset collected by the towed camera system SeaSLED in the Marguerite Bay area of the west Antarctic Peninsula in December 2010. The system was towed from a ship at an average of 3 meters altitude from the seafloor. However, due to ship motion and unpredictable topography, there is a large variation in the altitude, much more than an AUV usually might experience. We truncated the data to only include about 2500 images captured at altitudes between 1.5 and 4 meters, approximately the range within which additive scattering can be neglected [23].

For each image, we computed 1000 keypoints and quantized each to one of 14 binary QuAHOG patterns [23]. A global histogram was then computed for the entire image.

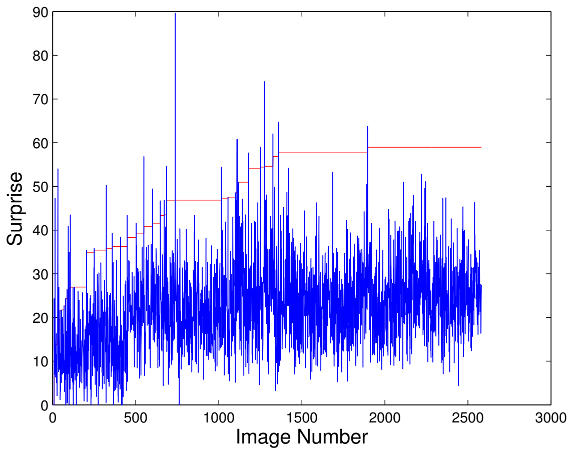

Fig. 2. Surprise as a function of image number. The threshold of surprise grows as more images are incorporated into the summary set.

Considering that images are captured every 3 seconds, the total mission time to capture 2800 images is over 2 hours. With the current state of the art in acoustic image transmission being approximately one full-resolution 1-megapixel image every 15 minutes [24], we estimate that about 8 images could be transmitted back within the course of a mission. Therefore, we set the summary set size to 8.

The summary set is initialized with the first 8 images and their corresponding surprise values are set to the smallest surprise measured relative to the set of images before it. Progress then continues throughout the rest of the data until the surprise threshold is exceeded by a novel image. When this happens, the novel surprising image is incorporated into the summary set, the least surprising image removed, and the surprise threshold augmented to the new lowest surprise value within the set as previously described. Figure 2 plots the surprise value and threshold throughout the course of the mission. As more of the environment is surveyed, the more surprising a new image must be to become incorporated into the summary set. The set of 8 summary images is shown in Figure 3. Images correspond to a spectrum of sandy and rocky areas.

There are several drawbacks to this approach that make it ill-suited in its current form for picking images to transmit during a mission. First, the summary represents a dynamic set of images, so there is no guarantee that an image that is transmitted will remain a member of the summary set throughout the rest of the mission. Second, simply transmitting images based on the highest "surprise" value can result in a handful of "outlier" images that are not representative of the dominant habitats in a survey. Lastly, if our goal is to use these summary images as the bases for building a semantic map to spatially characterize the survey environment, we need a means of reliably classifying non-summary images online as well. In the next section, we discuss several modifications to the online summary scheme of [1], [2] that enable mission-time visual feedback using low-bandwidth semantic maps.

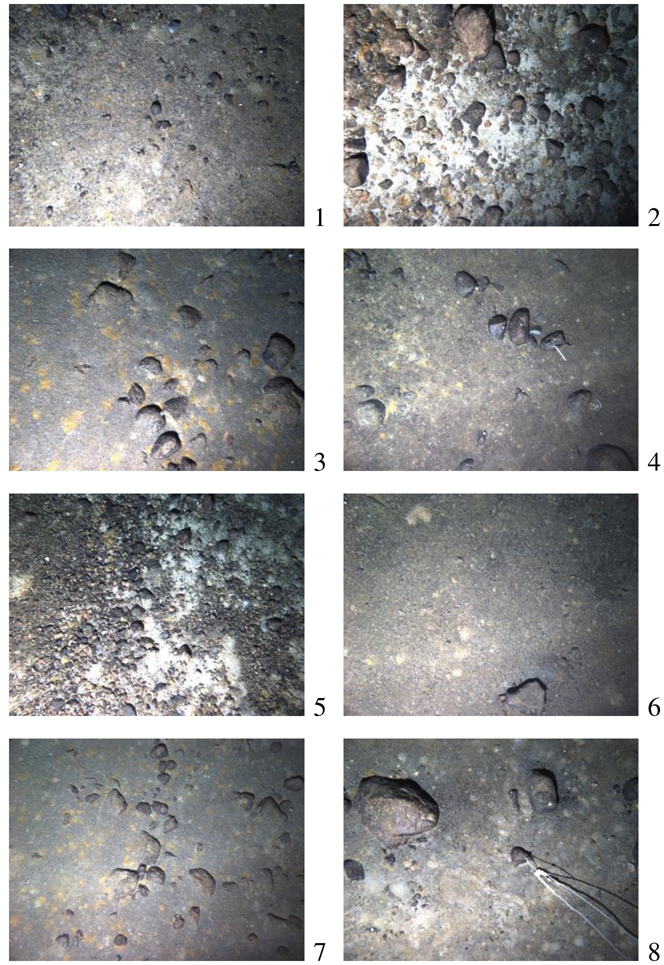

Fig. 3. The 8 summary images produced by the algorithm.

B. Modified Visual Summaries

Our first modification is to represent each non-summary image with a member of the summary set. Assuming that we have navigation data available to be transmitted as well, we can combine these representations with the approximate vehicle position to create spatial coverage maps based on the summary set. Intuitively, a non-summary image should be best represented by the summary image that is most similar. However, our current definition of surprise is not a true distance metric because it lacks symmetry. Therefore, we follow [2] and use a symmetric measure of surprise

$$d_{KL,sym}(p \parallel q) = \frac{1}{2}\big(d_{KL}(p \parallel q) + d_{KL}(q \parallel p)\big). \quad (5)$$

Representing a non-summary image by its nearest neighboring summary in this way can be thought of as minimizing the surprise one would have when looking through all the non-summary images represented by a given summary image.

We next must determine which summary images to transmit. Obviously, it is desirable to transmit the first image as soon as possible to minimize the latency of understanding for the operator. However, early in the mission the surprise threshold grows rapidly as the algorithm determines which images best represent the data, as seen in Figure 2. Thus, we propose to wait until the surprise threshold does not change for a specified number of images, implying that the vehicle is imaging consistent terrain that could be represented well by a single image. Using the summary set size M as a threshold is a simple and natural choice.

For subsequent images, we assume that the vehicle will be ready to transmit another image after a set number of frames. If imagery is captured every 3 seconds and can be transmitted every 15 minutes, this means that one summary image can be transmitted approximately every 300 frames. We would like to choose a summary image that is different enough from the previously transmitted summary images while at the same time represents enough non-summary images to make it a worthwhile choice for a map basis. Figure 4 illustrates two extreme cases. If the summary set that represents the most non-summary images is chosen, the blue circle, there is no guarantee that it is different enough from the previously transmitted summary images. As before, we can formulate our choice to minimize the surprise one would have when looking through the other summary images. We are effectively choosing a summary subset within the summary set. However, simply choosing the summary image that minimizes this surprise does not guarantee that it represents enough non-summary images to make it a useful basis for the map. Hence, we select the summary set that both minimizes the Haussdorff distance when the summary set is partitioned into subsets as well as represents enough non-summary images to exceed a given threshold. As before, we simply use the summary set size M as a minimum acceptable value.

Selecting good summary images to transmit is important because these images will be used to represent the entire dataset for the duration of the mission. Furthermore, this means that, as new summary images are added to the summary set, previously transmitted summary images should not be removed from the summary set given the high cost of transmitting an image. Subsequently, after a summary image is transmitted, it becomes "static," as opposed to the other "dynamic" summary images. To ensure this at runtime, both the surprise value and the number of non-summary images that "static" summary image represents are set to infinity.

Online summary methods do not require distances to be recomputed for all existing data points when new data appears which is one quality that makes them attractive for power-limited underwater robots. Thus, when a new summary is added to the set, we would rather not lose the information we have gained by simply removing the least-surprising summary image and the non-summary images that it represents. Instead, we propose to merge it with the nearest summary image so that it and its non-summary images all become non-summary images represented by the nearest summary image.

In practice, we found that straightforward merging can result in summary images representing large groups of non-summary images being absorbed by new summary images that do not represent many non-summary images. Such an occurrence is less than ideal for creating consistent maps. Thus, we advocate a greedy approach whereby, when merging two summary images, the one that represents more non-summary images remains a summary image. In the case of the least surprising summary image being chosen, the surprise threshold will not increase. To show that our overall approach preserves distance information, we plot the symmetric surprise distance between all 3000+ images in Figure 5. At left, distances have been calculated between each image. At right, the distances for each image have been replaced by their representative summary image's distances. Remarkably, the structure within the dataset is preserved quite well given the almost 30,000:1 compression ratio.

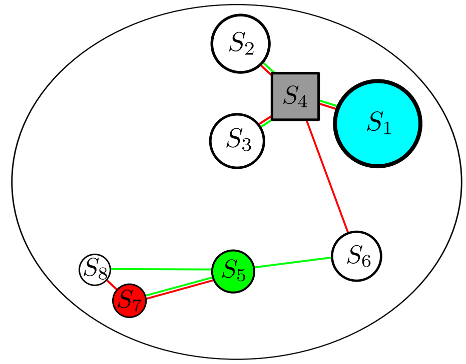

Fig. 4. Visual illustration of our symmetric surprise-based method for choosing summary image to transmit. The grey square indicates a previously transmitted or "static" set. Circles indicate untransmitted or "dynamic" sets whose size is proportional to the number of non-summary images they represent. Choosing the largest summary set, the blue circle, is not guaranteed to full characterize the diversity within the dataset. Choosing the summary set that minimizes the Haussdorff distance, the red circle and lines, between the subsequent classes will often pick the most surprising summary image, but this set is not guaranteed to represent enough non-summary images to contribute usefully to the semantic map. We elect to use the Haussdorff distance, but threshold the minimum number of non-summary images, shown by the green circle and lines.

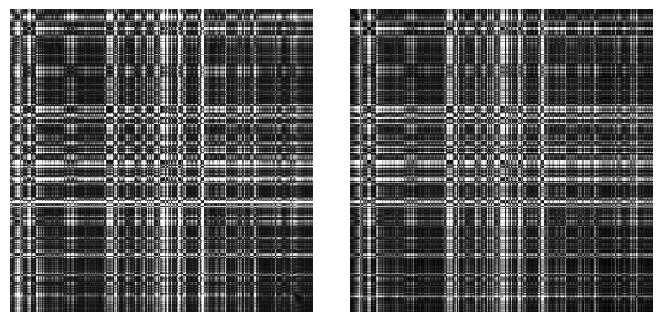

Fig. 5. Symmetric surprise between all images (left) and symmetric surprise between all images (right) using only the surprise values from their representative summary set.

C. Generating Semantic Maps

We have described modifications which enable us to select summary images to transmit that characterize the diversity in the dataset and will not change as additional summary images are added and merged. After the first image is transmitted and received, an operator has an initial understanding of the survey environment. After the second image is transmitted and received, additional scalar data containing navigation and classification information can be compressed and transmitted as well, providing the operator with ample information to begin to construct a spatial map of the survey environment. The classification masks exhibit high redundancy and covariance so they can be compressed at high rates. These data can be transmitted using very little bandwidth with the techniques presented in [25] and [24].

We implemented this new approach on a 2800 image dataset collected by the SeaBED AUV [26] in 2003 in the Stellwagen Marine Sanctuary. The survey consisted of multiple track lines over various habitats composed of boulders, rubble, sand, and mud. Imagery was captured every 3 seconds from approximately 3 meters altitude. Figure 6 shows the resulting progressive semantic maps created after each subsequent image and corresponding data are transmitted. Because the transmitted summary images become static, to allow freedom in the dynamic summary images we set the summary set size M to approximately twice the number of images we anticipate to transmit, in this case 16. The first image (red) was transmitted when the surprise threshold stabilized after 147 images. Each subsequent transmitted image was chosen after 300 frames had elapsed, simulating a realistic 15 minute transmission interval [24]. The first map is based on the first (red) and second (green) images, the second on the first three, and so on, until all 9 images are used.

Some of these classes are similar and the operator may wish to merge them for visual clarity. In Figure 7 the 9 transmitted images have been heuristically merged into 5 distinct classes: (from top to bottom at right) sand, piled boulders, lone boulders in sand, mud, and rubble. From the complete mosaic and the bathymetric map, it is clear that the piled boulders correspond to the tops of ridges. Depths in the bathymetric map range from 60 meters (warmer hues) to 70 meters (colder hues). Between these ridges are sandy areas, all of which are bordered by mud and smaller rubble.

This level of dataset understanding would be extremely valuable for an operator to possess during a mission. For instance, if the boulder fields were of particular interest to a scientist, the vehicle could be issued a redirect command to resurvey that area at higher resolution. Conversely, if a particular substrate of interest is not being imaged, the mission can be terminated and the vehicle recovered and relocated to another area. Furthermore, upon recovery of the vehicle, the operator has a fully classified dataset with additional summary images as well. The non-summary images represented by each summary images can be browsed to check the class validity. Several randomly selected non-summary images have been chosen from each of the 5 summary sets in Figure 7 and are shown in Figure 8.

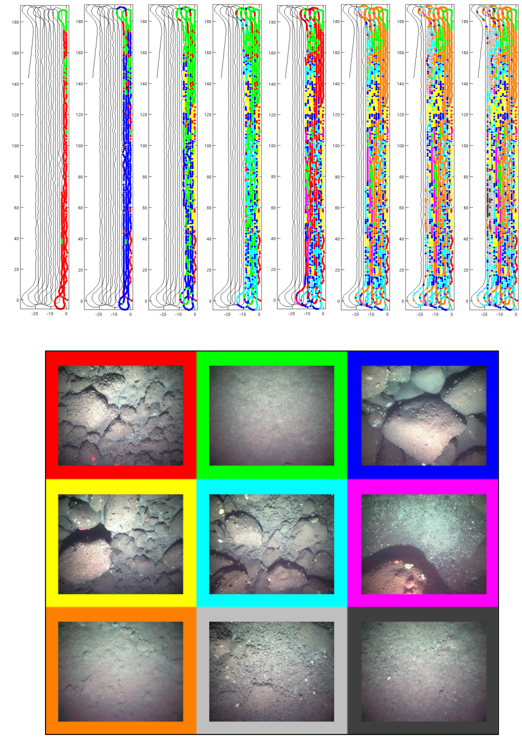

Fig. 6. Semantic maps created after each subsequent image is transmitted (top) with summary images and respective color codes (bottom).

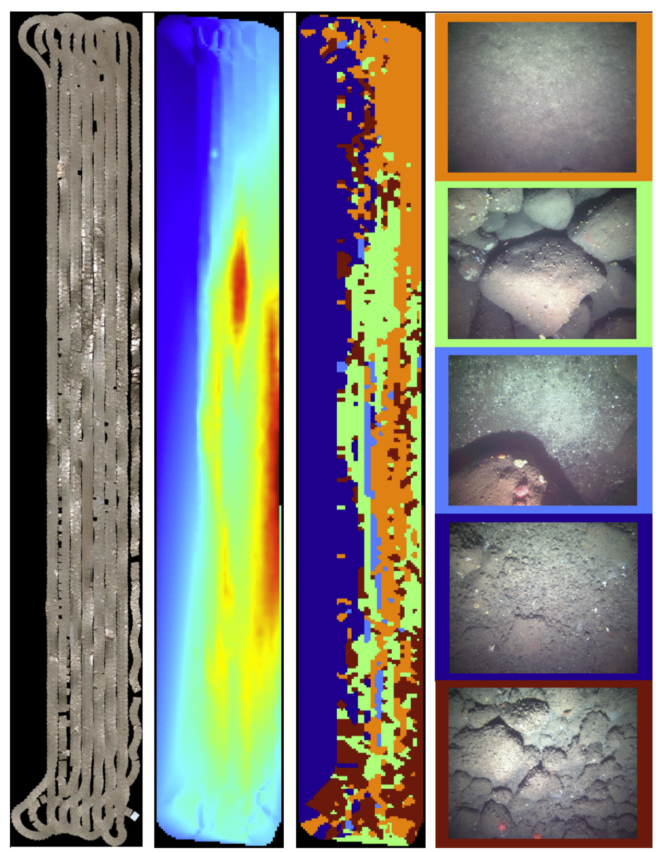

Fig. 7. Photomosaic (left) and bathymetry (middle left) of the entire mission. The final semantic map (middle right) using 9 images which have been heuristically merged into 5 distinct classes (right) and color coded.

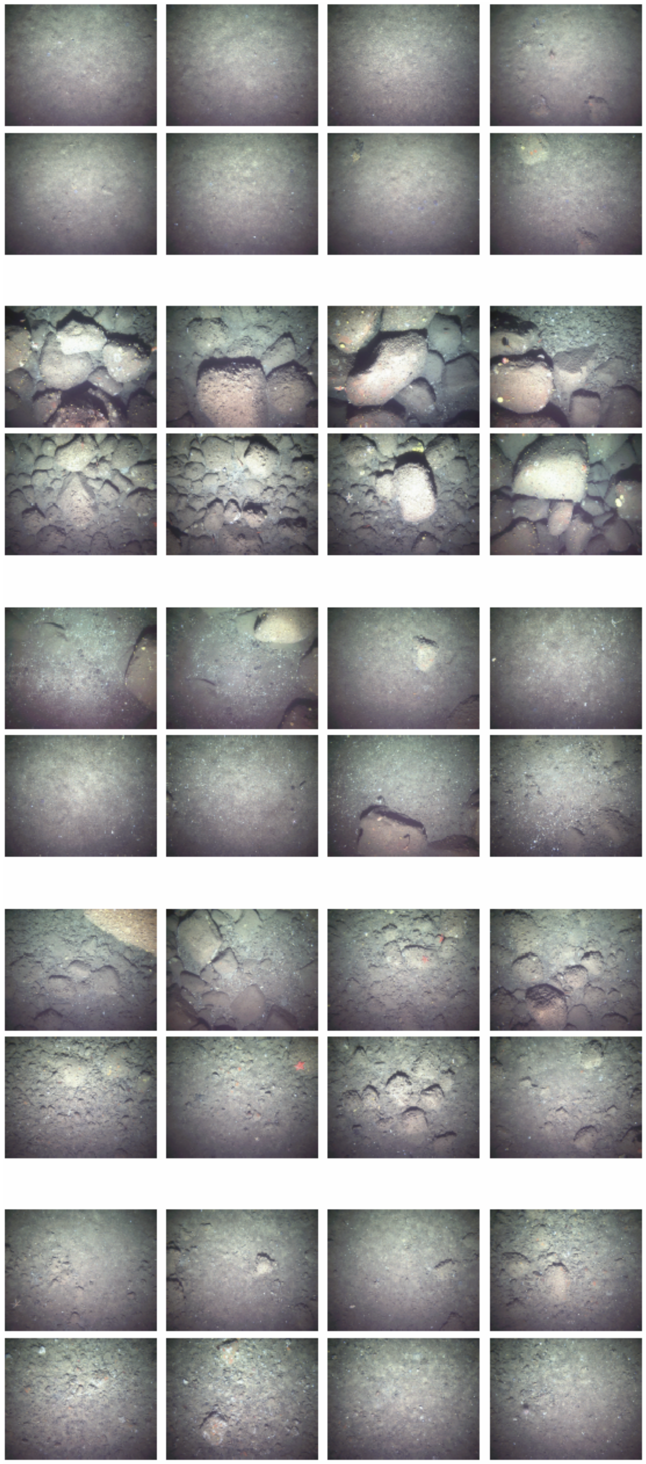

Fig. 8. Example imagery from each of the five heuristically merged classes.

IV. DISCUSSION

This work makes contributions to the field of autonomous underwater robotics by describing a framework that can be used to reduce the "latency of understanding," or the time delay between when an image is captured and when it is finally "understood" by an operator. This latency is propagated from two sources: first, from the low-bandwidth of the acoustic communication channel which greatly restricts the throughput of data; second, from the large volume of image data that must be analyzed. The second source has been addressed by numerous automated classification algorithms designed to annotate image data in an offline post-processed sense. The first source has been addressed by recent compression work allowing a small set of images to be transmitted over the course of a mission. We have addressed both of these sources by describing a lightweight framework designed to run in real time aboard a robotic vehicle that can produce environmental maps based on a subset of summary images.

Our approach is unique because it demonstrates that a simple, forward-mapped pattern vocabulary can be used to produce meaningful results without relying on complex descriptors that must be first learned and then subsequently quantized into a dictionary [23]. Furthermore, this work augments the visual summary literature under the assumption that summary images, navigation data, and classification masks can all be transmitted back at some rate during a mission. While existing techniques approach the visual summary problem strictly as a visual summary problem, we approach it from a compression standpoint in the context of a robot vehicle's ability to communicate a high-level understanding of its environment given the limitations of acoustic modems. Our work represents an enhancement of the capabilities of robotic vehicles to explore new environments and improve the quality of operator involvement during vehicle missions.

We hope to implement this framework on a physical embedded system aboard a vehicle. One realization is to operate directly off the image buffer in the sealed camera pressure housing. If images were processed and stored in the same housing as the camera, this scenario would limit the required information transfer between pressure housings on vehicles, which are often highly modular, only sharing compressed summary images and the accompanying semantic map. This implementation also makes the camera unit more modular and applicable to other monitoring applications such as moored or cabled observatories that are continuously collecting image data where storage constraints become problematic over long timescales.

Another way to utilize acoustic modems and image compression techniques to reduce the latency of understanding is to continuously transmit a low-bitrate descriptor for each image as it is captured. This concept has roots in the mobile visual search paradigm [27], [28] where a descriptor is sent to a server in place of the query image itself. In this scenario, the online clustering (or even repeated offline clusterings) would be performed on the ship where power and computational resources can be virtually unlimited and thus a better set of representative cluster centers can be obtained. The vehicle can then be queried to compress and transmit these representative images during the mission.

ACKNOWLEDGMENT

We are grateful to Clay Kunz for his generous help in generating photomosaics of the data.

References

[1] Y. Girdhar and G. Dudek, "Onsum: A system for generating online navigation summaries," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 746–751.

[2] ——, "Efficient on-line data summarization using extremum summaries," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3490–3496.

[3] R. Paul, D. Rus, and P. Newman, "How was your day? online visual workspace summaries using incremental clustering in topic space," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4058–4065.

[4] L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, and K. Ball, "The whoi micro-modem: an acoustic communications and navigation system for multiple platforms," in *OCEANS, 2005. Proceedings of MTS/IEEE*. IEEE, 2005, pp. 1086–1092.

[5] M. Stojanovic, "Recent advances in high-speed underwater acoustic communications," *Oceanic Engineering, IEEE Journal of*, vol. 21, no. 2, pp. 125–136, 1996.

[6] M. Doniec, C. Detweiler, I. Vasilescu, and D. Rus, "Using optical communication for remote underwater robot operation," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 4017–4022.

[7] M. Doniec and D. Rus, "Bidirectional optical communication with aquaoptical ii," in *Communication Systems (ICCS), 2010 IEEE International Conference on*. IEEE, 2010, pp. 390–394.

[8] N. Farr, A. Bowen, J. Ware, C. Pontbriand, and M. Tivey, "An integrated, underwater optical/acoustic communications system," in *OCEANS 2010 IEEE-Sydney*. IEEE, 2010, pp. 1–6.

[9] L. Freitag, M. Grund, C. von Alt, R. Stokey, and T. Austin, "A shallow water acoustic network for mine countermeasures operations with autonomous underwater vehicles," *Underwater Defense Technology (UDT)*, 2005.

[10] J. A. Farrell, S. Pang, and W. Li, "Chemical plume tracing via an autonomous underwater vehicle," *Oceanic Engineering, IEEE Journal of*, vol. 30, no. 2, pp. 428–442, 2005.

[11] M. V. Jakuba, "Stochastic mapping for chemical plume source localization with application to autonomous hydrothermal vent discovery," Ph.D. dissertation, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2007.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd ed. John Wiley & Sons, 2001.

[13] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, vol. 43, pp. 29–44, 2001, 10.1023/A:1011126920638. [Online]. Available: http://dx.doi.org/10.1023/A:1011126920638

[14] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, pp. 61–81, 2005, 10.1023/B:VISI.0000046589.39864.ee. [Online]. Available: http://dx.doi.org/10.1023/B:VISI.0000046589.39864.ee

[15] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *science*, vol. 315, no. 5814, pp. 972–976, 2007.

[16] N. C. Loomis, "Computational imaging and automated identification for aqueous environments," Ph.D. dissertation, MIT/WHOI Joint Program in Oceanography / Applied Ocean Science & Engineering, 2011.

[17] J. Sivic, B. Russell, A. Zisserman, W. Freeman, and A. Efros, "Unsupervised discovery of visual object class hierarchies," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, june 2008, pp. 1 –8.

[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. IEEE, 2005, pp. 524–531.

[19] O. Pizarro, S. Williams, and J. Colquhoun, "Topic-based habitat classification using visual data," in *OCEANS 2009 - EUROPE*, may 2009, pp. 1 –8.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944937

[21] S. Zhong, "Efficient online spherical k-means clustering," in *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, vol. 5. IEEE, 2005, pp. 3180–3185.

[22] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention," in *Advances in neural information processing systems*, 2005, pp. 547–554.

[23] J. W. Kaeli, "Computational strategies for understanding underwater optical image datasets," Ph.D. dissertation, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2013.

[24] C. Murphy, "Progressively communicating rish telemetry from autonomous underwater vehicles via relays," Ph.D. dissertation, MIT/WHOI Joint Program in Oceanography / Applied Oceans Science and Engineering, 2012.

[25] T. E. Schneider, "Advances in integrating autonomy with acoustic communications for intelligent networks of marine robots," Ph.D. dissertation, Massachusetts Institute of Technology, 2013.

[26] H. Singh, R. Eustice, C. Roman, and O. Pizarro, "The seabed auv–a platform for high resolution imaging," *Unmanned Underwater Vehicle Showcase*, 2002.

[27] B. Girod, V. Chandrasekhar, D. M. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. S. Tsai, and R. Vedantham, "Mobile visual search," *Signal Processing Magazine, IEEE*, vol. 28, no. 4, pp. 61–76, 2011.

[28] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1349–1380, 2000.