Realtime Online Spatiotemporal Topics for Navigation Summaries

Yogesh Girdhar, Raheem Adam, and Gregory Dudek

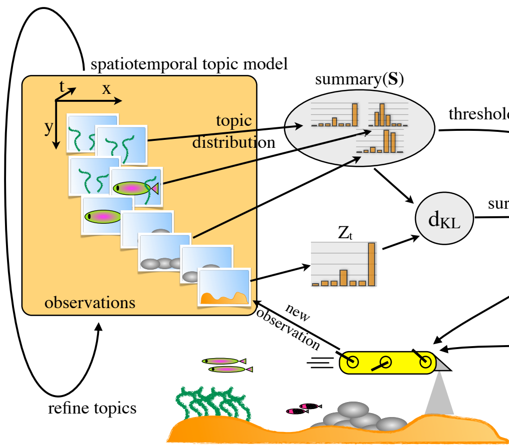


Fig. 1. Each incoming observation made by the robot, is a set of visual words, and their corresponding locations. We continuously refine topic labels for each word using an online Gibbs sampler, taking into account the spatiotemporal correlation of topic labels. Each incoming observation is hence represented using a distribution over these spatiotemporal topics, which the summarizer uses to build the summary. Surprise of a new observation is its distance to the closest sample in the summary. If this surprise is above a threshold we add the new observation to the summary.

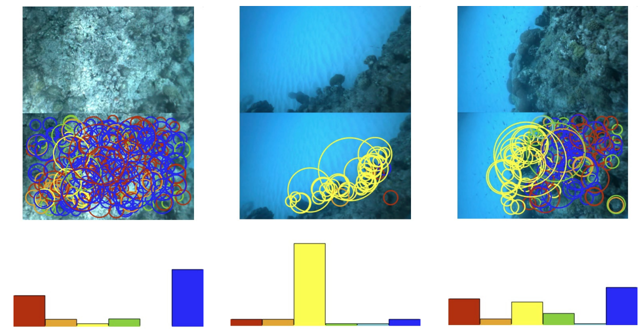
I. Introduction

Using mobile robots for scientific data collection is appropriate for several different domains, ranging from planetary exploration to undersea. We are interested in the particular challenge of generating a summary using an online anytime algorithm, so that results can be made available even while the data is being collected. This is important for real data collection scenarios since mission scientists want to be able to monitor progress. Moreover, the online capability to detect a location as surprising can be used to trigger further data gathering behaviors such as sensor dropping, or adaptive refinement of the resolution of the data being collected.

One major obstacle in implementing such an approach is the need for a low dimensional, perceptually relevant descriptor for the observations made by the robot. This descriptor should encode what is being observed in a meaningful way, and should be sensitive to scene changes. We propose a novel online topic-modeling framework, which takes into account the spatiotemporal continuity of the world to compute topic distribution for streaming visual data with spatiotemporal information; and most importantly, works online in realtime.

Given the computed topic distribution of the current observation, we use an online summarization algorithm which tries to maintain a subset of observations, which cover the set of all observations using minimum cover radius[3]. This summary, sometimes referred to as an extremum summary, encourages outliers to be part of the summary, which can be interpreted as surprising observations.

The authors are at: Center for Intelligent Machines, McGill University, Montreal, QC H3A0E9, Canada {yogesh,radam,dudek}@cim.mcgill.ca



(a) Rocks: red and blue (b) Sand boundary: yellow (c) Sand boundary with rocks: red, blue, and yellow

Fig. 2. Examples of spatiotemporal topics learned online from a dataset. Each visual word is marked by a circle, the size of which corresponds to the size of the visual feature, and the color to the topic label.

II. ROST: Realtime Online Spatiotemporal Topics

Latent Dirichlet Allocation (LDA)[1] is a technique for describing a set of documents as a mixture of topics, which are themselves a distribution over words. Our work extends LDA for robotics use. We model the observed visual words to be continuous in space and time. This allows for a more natural interpretation of topics as context, regions, or objects with continuity in space and time. Each observation is a tuple $(w_i, \mathbf{x}_i, t_i)$ consisting of an observed visual word $w_i = 1 \dots V$, associated coordinates $\mathbf{x}_i$, and timestamp t_i. The neighborhood of an observation at $(\mathbf{x}, t)$, denoted by $G(\mathbf{x}, t)$, is the set of observations in its spatiotemporal neighborhood. This could either be defined using k nearest neighbors, or using a radius search. We use the following generative model for the corresponding observed word w_i:

1) word distribution for each topic k: $\phi_k \sim \text{Dirichlet}(\beta)$,
2) neighborhood for an observation at $(\mathbf{x_i}, t_i) : G(\mathbf{x}, t) \sim$ uniformly from all neighborhoods which contain $(\mathbf{x_i}, t_i)$,
3) topic distribution the neighborhood $G(\mathbf{x}, t) : \theta_{G(\mathbf{x},t)} \sim \text{Dirichlet}(\alpha)$,
4) topic label for location $(\mathbf{x}_i, t_i)$: $z_i \sim \text{Discrete}(\theta_{G(\mathbf{x},t)})$,
5) word observed at location $(\mathbf{x}_i, t_i)$: $w_i \sim \text{Discrete}(\phi_{z_i})$,

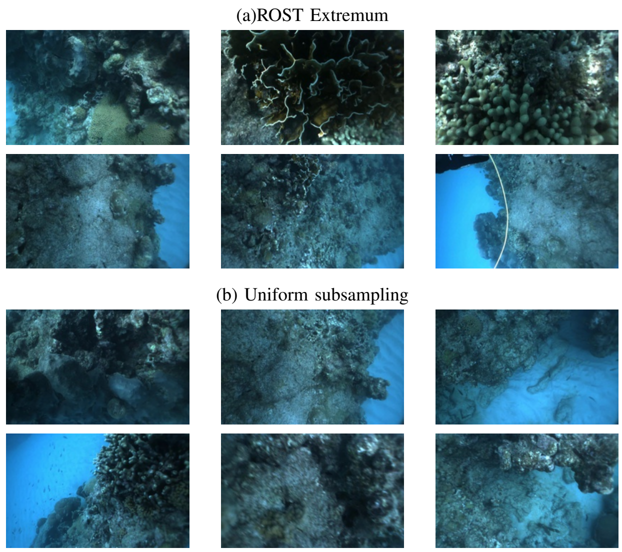


Fig. 3. (a) Extremum summary generated using ROST descriptor. (b) Uniformly subsampled images from the dataset, shown here for comparison.

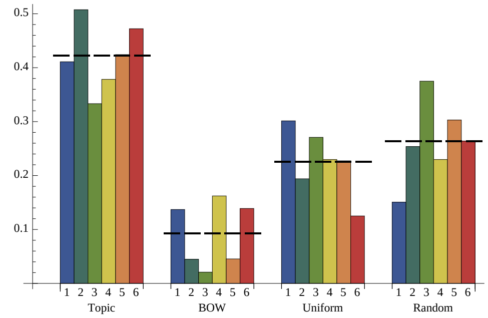


Fig. 4. Summary Evaluation by Human Subjects: Six different datasets, each containing a 2000 images taken by an underwater robot, and 4 different corresponding summaries were shown to participants. They were then asked to rank the summaries according to how well they summarize the dataset. The bar-graph shows the percentage of times an algorithm was voted as the best. Each colored bar corresponds to a dataset, and the dashed line indicates the mean percentage score for the algorithm. Overall we see that the spatiotemporal topics based summaries performs consistently the best in most of the datasets.

where $x \sim Y$ implies that random variable x is sampled from distribution Y.

At time t, we get a set of observations in the form of visual words and their locations. Let M_t be this set of word indices. We compute the topic distribution θ_t for observations at time t by integrating over all the observations and neighborhoods to get:

$$\theta_t(k) \propto \sum_{i \in M_t} \sum_{\mathbf{x}',t'} \mathbb{P}(z_i | \mathbf{z}_{-i}, \mathbf{x}_i, \mathbf{w}, G(\mathbf{x}', t')) \times \mathbb{P}(G(\mathbf{x}', t') | \mathbf{x}_i, t). \tag{1}$$

Several different strategies exist in the literature to do online refinement of the topic assignment in a given streaming dataset [4], [2]. The general idea is to initialize the topic label of the current observation with random labels, and then do a batch refinement of the entire dataset. This allows for previous topic assignments to be updated in the light of new observed data. For realtime use, the number of refinements between two observation needs to be constant. Hence, we propose to sample the neighborhoods from a $\mathrm{Beta}(a, 1)$ distribution, with $a > 1$, giving higher picking probability to recently observed regions. This ensures that new observations quickly converge, while older observations are less likely to change their labels. In this work, we set $a = 2$ for all experiments, however, increasing the value of a with time might lead to better results for long experiments.

III. Extremum Summaries

Online extremum summarization [3] maintains a set of observations which cover the space of all observations, while minimizing the coverage radius of a summary sample. 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

$$Cost(\mathbf{S}|\mathbf{Z}) = \max_i \min_j d(Z_i, S_j), \tag{2}$$

where d is the distance function, which measures distance as the symmetric KL divergence between the corresponding topic distributions computed using Equation 1. The overall algorithm for updating the summary given a new observation Z_t, and the current summary $\mathbf{S}$, is illustrated in Figure 1.

IV. Results and Discussion

Figure 2 shows examples of some of the topics learned by ROST. Each visual word is marked by a circle, size of which corresponds to the size of the visual feature, and color corresponds to the topic label. We see that the topics are representative of underlying physical phenomenon, and do well in describing scenes where a mixture of these exist. Figure 3 shows examples of summaries generated by our system. Figure 4 presents a summary of the results from a human validation experiment. The data is from 530 rank orderings, by 307 human subjects. We see that in 5/6 datasets, the proposed spatiotemporal topics based extremum summaries do the best. The bad performance of the bag-of-words summaries is likely due to the high dimensionality of its descriptor. Note that the proposed topic model uses the same vocabulary as the bag-of-words descriptor, and uses the same summarizer.

References

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

[2] 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*, 5(1999):65–72, 2009.

[3] Y. Girdhar and G. Dudek. Efficient on-line data summarization using extremum summaries. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3490–3496, 2012.

[4] X. Song, C.-Y. Lin, B. L. Tseng, and M.-T. Sun. Modeling and predicting personal information dissemination behavior. In *Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining - KDD '05*, page 479, New York, New York, USA, Aug. 2005. ACM Press.