

Offline Navigation Summaries

Yogesh Girdhar and Gregory Dudek

Abstract—In this paper we focus on the task of summarizing observations made by a mobile robot on a trajectory. A navigation summary is the synopsis of these observations. We pose the problem of generating navigation summaries as a sampling problem. The goal is to select a few samples from the set of all observations, which are characteristic of the environment, and capture its mean properties and surprises. We define the surprise score of an observation as its distance to the closest sample in the summary. Hence, an ideal summary is defined to have a low mean and a low max surprise score, measured over all the observations.

We present three different strategies for solving this sampling problem. Of these, we show that the k Cover sampling algorithm produces summaries with low mean and max surprise scores; even in the presence of noise. These results are demonstrated on datasets acquired in different robotics context.

I. INTRODUCTION

A navigation summary is a synopsis of observations made by a robot on a trajectory. Several years ago the term *vacation snapshot problem*[1] was coined to refer to the algorithm task analogous to what many tourists face: summarize their vacation using a small set of images. Hence, the task of generating a navigation summary can be seen as a generalization of the vacation snapshot problem, as it incorporates not just image data, but any other sensor data, including location.

We pose the problem of generating navigation summaries as a sampling problem. Three different data sampling strategies to generate navigation summaries are proposed. An ideal navigation summary consists of a small set of observation samples, which are characteristic of what was observed. The size of the summary is a user definable parameter, and can be arbitrarily small. By analogy, when you show your friends a slide show of you last vacation, the length of the show is determined in advance (or we at least hope so, for your friend's sake).

The summary samples should not only represent the mean appearance of the trajectory, but also its surprises. Hence, we judge the quality of a given summary using two different criterion: mean surprise and max surprise. We define surprise of an observation as its distance to the nearest observation in the summary. Finally, we analyze the properties of the navigation summaries generated using the proposed algorithms for data collected over two different terrains. This paper focuses on observation data in the form of images and the corresponding camera locations. The techniques discussed can however be applied to observations coming from any other sensors.

Applications of this work extend into several different areas such as surveillance, automated sample collection by

Y. Girdhar and G. Dudek are with Center of Intelligent Machines, McGill University, 3480 University Street, Montreal, QC, H3A2A7, Canada. yogesh@cim.mcgill.ca, dudek@cim.mcgill.ca

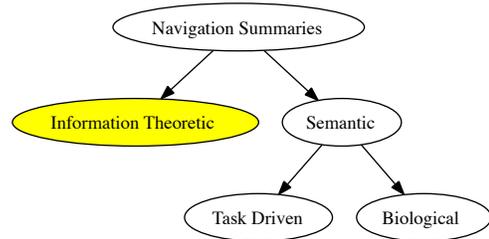


Fig. 1. Focus of this paper is to compute navigation summaries using information-theoretic and statistical tools. Generating semantic summaries requires either task specific information, or learning from a biological agent.

a planetary rover, and automated inspections of an environment. Simply saving every observation or collecting every sample is often of little utility. The approach presented in this paper can be used to produce summaries of size suitable for analysis by a human directly.

In our previous work [2] we have explored the problem of generating navigation summaries online. The goal of this paper is to formalize the offline version of this problem.

II. RELATED WORK

The problem of generating navigation summaries with image data is related to both the video segmentation problem and the problem of identifying landmark views in a view based mapping system.

A good example of work on view based maps is by Konolige et al. [3]. In this work, the goal is to identify a set of representative views and the spatial constraints among these views. These views are then used to localize the robot. With this approach we end up with a number of images proportional to the length of the robot trajectory, and hence these view images do not satisfy our size criterion.

Related is the work by Ranganathan and Dellaert [4], where the goal is to identify a set of landmark locations, and then build a topological map using them. This work is also the first to introduce the notion of “Bayesian Surprise” [5] to the robotics community. The images selected by this system, although well suited to building topological maps, are however still not suitable for generating navigation summaries. First, we are not only interested in selecting surprising landmark locations, but also images which represent the typical (i.e. mean) appearance of the world. Second, number of samples selected are a function of trajectory length and complexity.

Chang et al. [6] have used Set Cover methods to build a video indexing and retrieval scheme. We use a similar sampling strategy, however our goal of building navigation summaries is quite different from this work. First, we consider data collected using GPS and other sensors while selecting the samples. Second, we modify and extend the set cover based sampling strategy to enable it to work with noisy data.

The seminal work by Cummins and Newman [7] on loop closure involves ability to identify the common geographic origin of images, which might not be completely similar; thereby producing a meaningful clustering of images, which could be used to generate summaries. This would indeed produce very detailed summaries, however, our focus is on producing a much sparser sampling of the observed data.

Methods of video summarization have generally put emphasis on summaries for video data created by humans where significance is related to both temporal extent and audio content. For example, in Gong and Liu [8] video summaries were produced by exploiting a principal components representation of the color space of the video frames. They used a set of local color histograms and computed a singular value decomposition (SVD) of these local histograms to capture the primary statistical properties (with respect to a linear model) of how the color distribution varied. This allowed them to detect frames whose color content deviated substantially from the typical frame, as described by this model. Cai et al.[9] have proposed a technique to creating video digests from omni-direction street view video sequence.

Video segmentation, in contrast to this work, places emphasis on finding category boundaries or cut scenes almost independent of their absolute number and without regard for positional data. For example, Ngo et. al. [10] first modelled the video as a complete undirected graph, and then used the normalized graph cut algorithm to partition the video into different clusters.

III. SURPRISE

A. Set Theoretic Surprise

Itti and Baldi [5] formally define Bayesian surprise in terms of the difference between posterior and prior beliefs about the world. They showed that observations which lead to high Kullback-Leibler(KL) divergence [11] between posterior and prior visual appearance hypothesis are very likely to attract human attention.

The relative entropy or KL divergence between two probability mass functions $p(x)$ and $q(x)$ is defined as:

$$d_{\text{KL}}(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}. \quad (1)$$

KL divergence can be interpreted as the inefficiency in coding a random variable from distribution p , when assuming its distribution to be q .

In this paper we represent surprise with the symbol ξ .

$$\xi = d_{\text{KL}}(\text{posterior}||\text{prior}). \quad (2)$$

Instead of modelling the properties of the world with a single distribution, we propose to maintain a set of local hypotheses, each corresponding to an observation sample in the summary set [12]. This set of distributions can then be interpreted as the prior hypothesis of the properties of the world. Using a set of distributions to model the properties allows us to model worlds with arbitrary complexity, by simply increasing or decreasing the size of the summary set.

Similarly, to measure the surprise of a given observation, we model the posterior using the set containing all the observations in the summary and the given observation. Now, analogous to Bayesian surprise, we would like to measure the distance between these two distribution sets. The Hausdorff metric provides a natural way to compute distance between two such sets. For two sets A, B , the Hausdorff distance between the sets is defined as

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}. \quad (3)$$

However, since the prior and posterior hypothesis sets only differs by one element (the given observation), the surprise of a given observation Z given the summary set $\mathbf{S} = \{S_i\}$ is:

$$\xi(Z|\mathbf{S}) = \inf_i \xi(Z|S_i) \quad (4)$$

We call this measure of surprise ‘‘Set Theoretic Surprise’’.

Each observation can be represented using a variety of descriptors depending on the sensors used. If $d_r(Z_i, Z_j)$ is the normalized distance between two observations Z_i, Z_j using descriptor r , then we define surprise:

$$\xi(Z_i|Z_j) = \max_r d_r(Z_i, Z_j). \quad (5)$$

In this paper, the observations are in the form of an image and the associated geographic location. We use ‘‘bag-of-words’’ representation to describe the image. Sivic and Zisserman [13] have proposed this model, in which each image is described as a histogram of word counts. To build the vocabulary, we extract SURF [14] features from a subsampling of all the images in the observation set, and then cluster them using k means algorithm to get 1000 words. The normalized frequency count of these SURF words in an image are then assumed to be their respective descriptions. We use KL divergence to measure the distance between two different word distributions.

To measure distance in terms of geographic location, we simply take Euclidean distance between the coordinates.

Surprise $\xi(Z|\mathbf{S})$ can be interpreted as the amount of information gained in observing Z . Given the set of all the observations $\mathbf{Z} = \{Z_i\}$, ideally we would like to choose a summary set $\mathbf{S} = \{S_i\}$ such that the mean surprise

$$\bar{\xi}(\mathbf{Z}|\mathbf{S}) = \frac{1}{|\mathbf{Z}|} \sum_i \xi(Z_i|\mathbf{S}), \quad (6)$$

and max surprise

$$\hat{\xi}(\mathbf{Z}|\mathbf{S}) = \max_i \xi(Z_i|\mathbf{S}), \quad (7)$$

are both low.

A summary with low $\bar{\xi}$ implies that it captures the mean properties of the world well. A summary with low $\hat{\xi}$ implies that it captures the outliers well. In the following section, we will describe three different sampling techniques to choose samples in the summary set.

IV. SAMPLING STRATEGIES

We are given a set of observation \mathbf{Z} and number of desired samples k in the summary set \mathbf{S} . Goal is to come up with a set of observations called summary set $\mathbf{S} \subset \mathbf{Z}$, $|\mathbf{S}| = k$, such that the $\bar{\xi}$ and $\hat{\xi}$ are low.

The first strategy focuses purely on picking outliers by repeatedly picking the samples farthest from the samples in the summary. The second strategy focuses on minimizing the mean surprise. The third strategy finds a balance between the previous two approaches.

A. Extremum Summaries

Algorithm 1 presents a simple greedy approach to the summary problem. Here k is the number of desired samples in the summary set, \mathbf{Z} is the set of input observations, and \mathbf{S} is the set of observations in the summary. This algorithm has $O(|\mathbf{Z}|^2)$ computational complexity.

The algorithm repeatedly picks samples which have maximum surprise score, given the samples already in the summary. To initialize the summary set \mathbf{S} , we propose to use the sample farthest away from the mean of the dataset.

This algorithm is essentially picking samples at the corners of the high dimensional manifold formed by the samples in the description space.

```

S ← {Sinit}
repeat
  Zmax ← argmaxZi ∈ Z ξ(Zi|S)
  S ← S ∪ {Zmax}
  Z ← Z \ Zmax
until |S| ≥ k
return S

```

Algorithm 1: EXTREMUMSUMMARY ($\mathbf{Z}|k$). Computes a summary as a subset of input samples \mathbf{Z} , by greedily picking the samples with maximum surprise.

This algorithm is good at picking outliers, however, with a large enough summary size it approximates uniform sampling of the description space. This is shown in Fig 2.

B. k Medoids Summaries

The k Medoids clustering algorithm is similar to k Means, with difference that k Medoids uses one of the input samples to designate the cluster centres. The algorithm minimize the distance between these cluster centres and other points within the cluster. Hence, if we use the surprise function described in (5) as the distance function for k Medoids, the cluster centres then correspond to a summary which minimizes the mean surprise $\bar{\xi}$.

C. Summaries Using Set Cover Methods

The task of selecting samples in the summary set can also be modelled as an instance of the classical *Set Cover Problem*.

We define the cover of an observation Z to be all the set of all observations that are only surprising up-to a surprise threshold ξ_T :

$$C(Z|\xi_T) = \{\{Z_j\} : \xi(Z_j|Z) < \xi_T\}. \quad (8)$$

Similarly we define the cover of a summary set \mathbf{S} :

$$C(\mathbf{S}|\xi_T) = \bigcup_{S_i \in \mathbf{S}} C(S_i|\xi_T). \quad (9)$$

Our goal now is to find the minimal set of samples \mathbf{S} , which cover the entire terrain. This is essentially an instance of the classical *Set Cover* problem with $|\mathbf{Z}|$ elements in the universe and $|\mathbf{S}|$ sets which span the universe. Fig. 3 shows an example of an instance of the *Set Cover* problem.

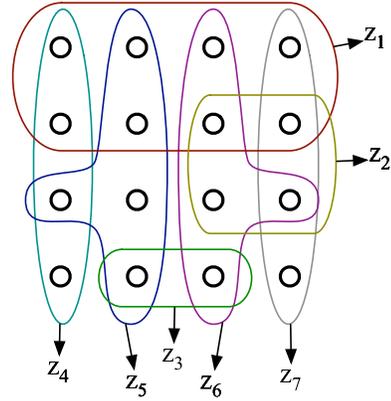


Fig. 3. An instance of the *Set Cover* problem. Goal is to find the smallest number of sets, which span all the elements. In this example, set $\{Z_4, Z_5, Z_6, Z_7\}$ is the smallest set of sets which cover all the elements in the universe.

Set Cover is known to be NP-hard [15]. Hence we use a greedy strategy to pick our samples.

```

S ← {Sinit}
repeat
  Zmax ← argmaxZi ∈ Z |C({Zi} ∪ S|ξT) \ C(S|ξT)|
  S ← S ∪ {Zmax}
  Z ← Z \ Zmax
until |C(S|ξT)| / |Z| < γ
return S

```

Algorithm 2: γ -COVERSUMMARY ($\mathbf{Z}|\xi_T, \gamma$). Computes a summary as a subset of all observations \mathbf{Z} , given the surprise threshold ξ_T , by greedily picking the observations with maximum cover. We stop when the coverage ratio is more than γ .

Algorithm 2 greedily picks $Z_{\max} \in \mathbf{Z}$ which provides maximum additional cover, and then adds it to the summary

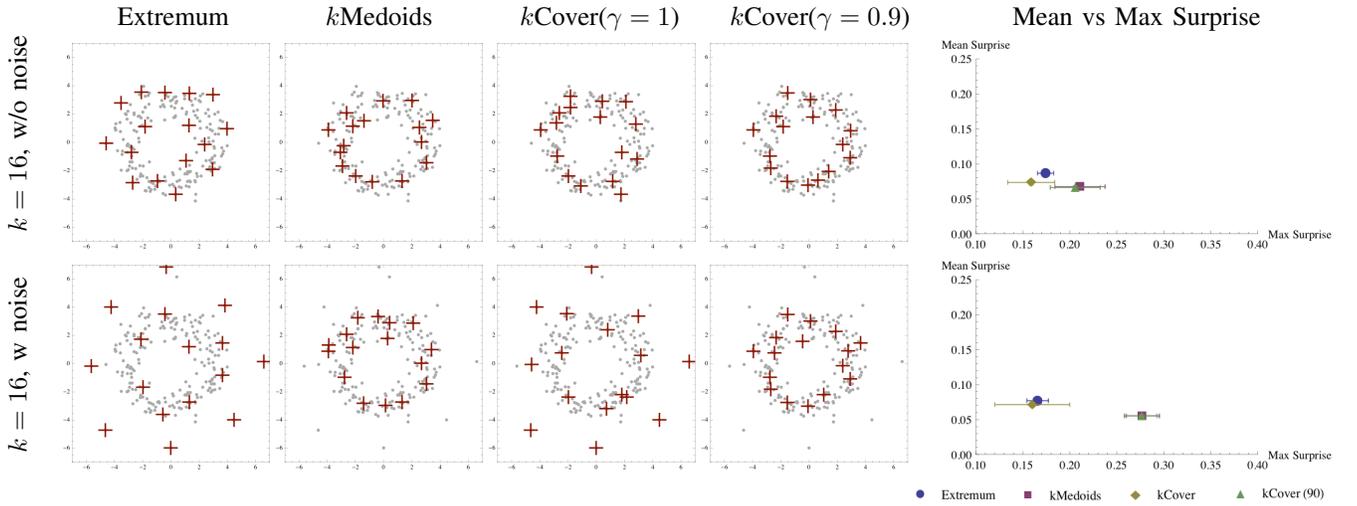


Fig. 2. Summaries for Noisy Datasets. When the data has little noise (row 1), we see that all algorithms perform similarly. When some noisy samples are added to the same data(row 2), both extremum summary algorithm and $k\text{Cover}(\gamma = 1.0)$ favour picking these outliers. However, by lowering γ , we can make $k\text{Cover}$ immune to noise and perform similar to $k\text{Medoids}$. Mean vs Max surprise plots shown are averages for 100 randomly generated datasets from the same distribution. The dataset shown above is one specific example.

set \mathbf{S} . Algorithm stops when the coverage ratio exceeds the parameter γ . Setting coverage parameter $\gamma = 1$ implies that the algorithm stops when the summary covers all the observations. Lower values of γ can be used for higher noise tolerance, however at the cost of increased likelihood of missing a truly surprising sample.

To initialize the summary set, S_{init} is chosen to be the sample with smallest mean distance to other samples.

The number of samples in the summary obtained using this greedy strategy, is guaranteed to be no more than $OPT \log |\mathbf{Z}|$, where OPT is the number of sets in the optimal summary set [16].

Often times we do not have a way of estimating a good value for the surprise threshold ξ_T . Instead, we can fix the size of the summary k , and then find the smallest value of ξ_T that gives us the desired coverage ratio.

Using Algorithm 3, we can hence define ξ_T as:

$$\xi_T(k) = \min_{\xi_{T'}} \left\{ \mathbf{S} = \text{K-COVERSUMMARY}(\mathbf{Z}|k, \xi_{T'}), \frac{|C(\mathbf{S}|\xi_{T'})|}{|\mathbf{Z}|} \geq \gamma \right\}. \quad (10)$$

Algorithm 3 greedily picks k summary samples which provide maximum combined cover for the summary set. This problem of finding the optimal set of samples with maximum cover is known as the *Max k Cover* problem. *Max k Cover*, like *Set Cover* is also known to be NP-Hard, and the greedy approach to this problem approximates the optimal solution to within a ratio of $1 - 1/e$ [17].

V. EVALUATION

We tested the summarization strategies presented in this paper on datasets collected by different classes of vehicle moving outdoors. We evaluate the summaries based on their

```

1  $\mathbf{S} \leftarrow \{S_{\text{init}}\}$ 
2 repeat
3    $Z_{\text{max}} \leftarrow \underset{Z_i \in \mathbf{Z}}{\text{argmax}} |C(\{Z_i\} \cup \mathbf{S}|\xi_T) \setminus C(\mathbf{S}|\xi_T)|$ 
4    $\mathbf{S} \leftarrow \mathbf{S} \cup \{Z_{\text{max}}\}$ 
5    $\mathbf{Z} \leftarrow \mathbf{Z} \setminus Z_{\text{max}}$ 
6    $k \leftarrow k - 1$ 
7 until  $k > 0$ 
8 return  $\mathbf{S}$ 

```

Algorithm 3: K-COVERSUMMARY ($\mathbf{Z}|k, \xi_T$). Computes a summary of size k , given the surprise threshold ξ_T by greedily picking samples with maximum cover.

mean and max surprise scores. A good summary should have both low mean and low max surprise scores.

A. \mathbb{R}^2 Ring Dataset

In Fig. 2, we shows 200 samples generated randomly around a circle in \mathbb{R}^2 . We then add 8 extra samples from a different distribution, representing the outliers. This dataset allows us to visualize the difference between the three proposed strategies clearly.

In Fig. 2(row 1), we see that when there is little noise, and size of the summary is large ($k = 16$), all algorithms perform similarly, with $k\text{Cover}(\gamma=1)$ having the lowest max surprise, and both $k\text{Cover}(\gamma=0.9)$ and $k\text{Medoids}$ having lowest mean surprise.

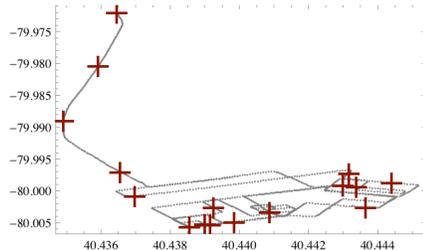
Fig. 2(row 2) shows result of adding outliers to this dataset. We see that the $k\text{Cover}(\gamma=1)$ and the extremum summaries look similar and both sample the outliers and the mean samples well. The $k\text{Cover}(\gamma=1)$ and the $k\text{Medoids}$ summary on the other hand completely ignore the outliers. As a result they have high max surprise and low mean surprise.



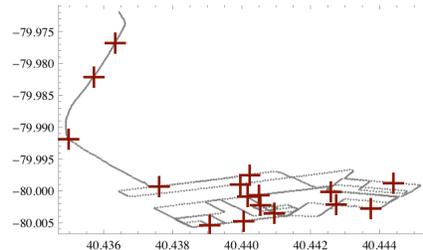
(a) Extremum Summary

(b) $k\text{Cover}(\gamma=1.0)$ Summary

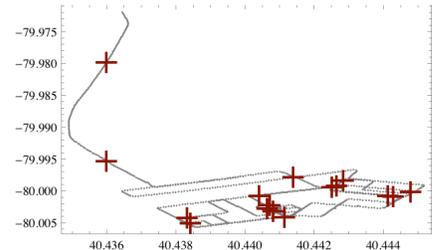
(c) $k\text{Medoids}$ Summary



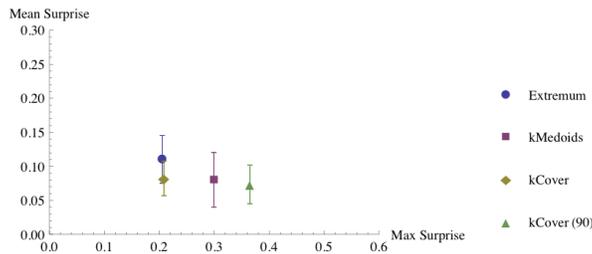
(d) Extremum Summary



(e) $k\text{Cover}(\gamma=1.0)$ Summary



(f) $k\text{Medoids}$ Summary



(g) Mean vs Max Surprise

Fig. 4. Street View Dataset. Three different summaries are show. Extremum summary algorithm focuses on finding the outliers and results in low max surprise. Most images are of unique buildings which are rare in the dataset, but are extremely different in appearance from everything else. The $k\text{Medoids}$ algorithm produces a summary with no outliers and low mean surprise. The $k\text{Cover}$ algorithm has in-between characteristics and produces a summary with low mean and max surprise. It contains some of outliers which exist in the extremum summary(5 \rightarrow 9, 9 \rightarrow 13, 10 \rightarrow 11, 12 \rightarrow 8, where $a \rightarrow b$ implies image #a matches with image #b), while not ignoring the samples representing the mean appearance. See text for a more detailed discussion.

B. Aerial View Dataset

The aerial view(AV) dataset contains 847 images taken by a GPS equipped unmanned aerial vehicle(UAV). The images in this dataset were transmitted over an analog channel before being captured digitally. As a result, some of the images in the AV dataset are noisy due to temporary signal losses, and interference with on board electronics and power lines. This results in the dataset having many undesirable outlier images.

Fig.5(a) highlights the problem of using a noise sensitive algorithm such as the extremum summary algorithm on noisy

data. Since the extremum summary algorithm favours picking outliers, the summary generated contains several images with visible noise related artifacts, and at least one image with complete loss of signal. However due to the constraints imposed by geographic data, it still manages to get good geographic coverage as shown in the map in Fig 5(b). The map shows location of each image in the summary. We see that images 3,4,5 and 7 correspond to the four corners of the map, which is the behaviour expected from the extremum summary algorithm.

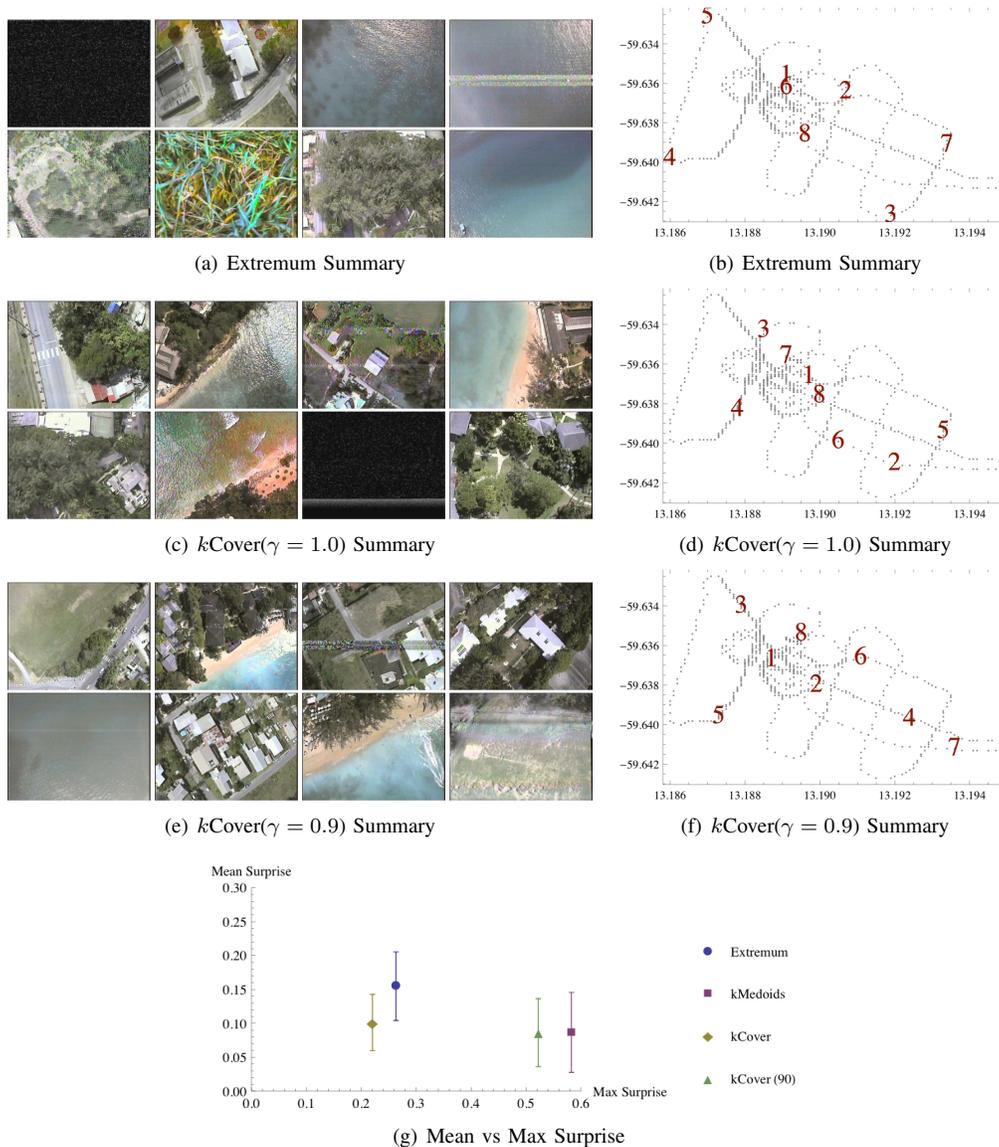


Fig. 5. Aerial View Dataset. This dataset contains images which were captured as a UAV takes off from a field, flies over over water, beaches and buildings, and lands. Images were transmitted over an analogue channel, and hence contains noise due to interference. (a) Extremum summary algorithm prefers picking outliers and hence prefers picking noisy images (#1,3,4), or images at the geographic corners (#3,4,5,7). Image 6 is of grass, taken just before the landing. (c) The $k\text{Cover}$ summary algorithm with $\gamma = 1.0$ is forced to choose summary picks which cover all samples including the noisy samples, but not necessarily pick them. Hence, it mitigates the effect of noise, while not completely eliminating it. We see the inclusion of some obvious outliers like image #7. (e) The $k\text{Cover}$ algorithm with $\gamma = 0.9$ does a good job of ignoring these outliers completely, and presents a clearer picture of what was observed during flight: ocean(#5), beaches(#2,7), buildings(#4,6), fields(#1,8). See text for a more detailed discussion.

Fig.5(c) shows images from the $k\text{Cover}$ summary with coverage ratio $\gamma=1$. Qualitatively we see that it provides a much better representation of the environment than the extremum summary. However, coverage ratio $\gamma=1$ implies that every sample in the observation set must be covered by a summary sample. The algorithm is therefore forced to pick image 7, which is the result of a temporary signal loss, and is not representative of the environment.

Fig.5(e) show the $k\text{Cover}$ summary with $\gamma = 0.9$. We see that the algorithm is able to ignore the noisy samples, and produces a navigation summary with good geographical and visual coverage. We see that all major features of the

environment: ocean, beach, buildings, fields, and trees, are represented in the summary. The $k\text{Medoids}$ algorithm produces a summary very similar to the $k\text{Cover}(\gamma=1)$ summary. We do not show it this paper due to space constraints.

C. Street View Dataset

The street view(SV) dataset contains 1255 geo-tagged images of a city centre, subsampled from the data originally collected by Google Inc., for their street view application. Unlike the AV dataset, SV dataset has negligible noise.

Fig. 4(d) shows the summary generated by the extremum summary algorithm which favours picking outliers or the corner samples. Upon closer inspection, it seems that many

of the selected images correspond to images of building with different repetitive patterns, completely occupying the camera's field of view. Such images would be represented by histograms with a sharp peaks at different locations, when using the bag-of-words representation of an image. As a result, the KL divergence between these images is very high, and hence, they are favoured by the extremum summary algorithm.

Fig. 4(f) shows the summary produced by the k Medoids algorithm. Compared to the images in the extremum summary, the images in the k Medoids summary will not have a peaky bag-of-words histogram. The images in the summary do not contain repetitive patterns and represented by many different visual words. As a result, the selected images share visual properties with many other images in the dataset, and are good candidates for characterizing the mean appearance of the environment.

Fig. 4(e) show the summary generated using the k Cover algorithm. Using k Cover with $\gamma = 1.0$ is a good choice for this dataset because there is negligible noise in the dataset, and hence the outliers correspond to truly surprising observations. The summary produced by k Cover does not completely ignore the outliers, and nor does it only pick samples representing the mean. It finds a good balance between lowering the mean and max surprise. This is shown by the plot in Fig. 4(g). The k cover summary has some of the outliers from extremum summary, while still having some images which are more representative of the environment.

VI. CONCLUSION

A navigation summary is a synopsis of observations made by a robot on a trajectory. These observations can come from any sensors. In this paper we focus on summarizing observation data in the form of images captured by a robot, and its location.

We pose the problem of generating navigation summaries as a sampling problem. Three different strategies for picking the summary samples were proposed.

We demonstrate the difference between these strategies by experimenting with different datasets, and find that the k Cover summary algorithm with the right coverage ratio parameter, performs at least as good or better than the k Medoids or the extremum summary algorithm, in several different scenarios. The performance of a summary was measured in terms of its mean and max surprise score measured over all the observations.

In general, we find that the extremum summary algorithm focuses on picking outliers. These outliers could be genuinely interesting samples like in the case of street view dataset, or they could correspond to undesirable noisy samples like the ones present in the aerial view dataset.

On the other hand, the k Medoids summary algorithm focuses on picking samples representing the mean properties of the environment. These summaries could be useful if the dataset is simple and lacks and interesting outliers.

Using the k Cover algorithm, we can produce summaries which finds a balance between representing outliers and the

mean properties of the environment. The desired coverage ratio parameter γ allows us to tune out noise in the case that the dataset is full of undesirable outliers.

In future we hope to look at the extending this work in several different directions. We would like to use these navigation summaries for planning exploration strategies for a mobile robot. Second, we would like to compare the summaries generated by the algorithms presented with human generated summaries. Third, we would like to incorporate the use of more sensors and image descriptors to enhance the surprise function.

ACKNOWLEDGEMENTS

We would like to thank Google Inc. for generously allowing us to use their street view data.

REFERENCES

- [1] E. Bourque and G. Dudek, "On the automated construction of image-based maps," *Autonomous Robots*, vol. 8, no. 2, pp. 173–190, April 2000.
- [2] Y. Girdhar and G. Dudek, "ONSUM: A system for generating online navigation summaries," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, Oct 2010, pp. 746–751.
- [3] K. Konolige, J. Bowman, J. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, "View-based maps," *RSS'09*, 2009.
- [4] 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.
- [5] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision Research*, vol. 49, no. 10, pp. 1295 – 1306, 2009, visual Attention: Psychophysics, electrophysiology and neuroimaging.
- [6] H. S. Chang, S. Sull, and S. U. Lee, "Efficient video indexing scheme for content-based retrieval," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 8, pp. 1269–1279, dec 1999.
- [7] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, p. 647, 2008.
- [8] Y. Gong and X. Liu, "Video summarization using singular value decomposition," in *Proc. of CVPR*, 2000, pp. 174–180.
- [9] H. Cai and J. Y. Zheng, "Digesting omni-video along routes for navigation," in *Proceedings of the international conference on Multimedia*, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 647–650. [Online]. Available: <http://doi.acm.org/10.1145/1873951.1874042>
- [10] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang, "Video summarization and scene detection by graph modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 2, pp. 296–305, 2005.
- [11] S. Kullback, *Information theory and statistics*. John Wiley and Sons, NY, 1959.
- [12] Y. Girdhar and G. Dudek, "Online navigation summaries," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2010, pp. 5035–5040.
- [13] J. Sivic and A. Zisserman, "Video google: Efficient visual search of videos," in *Toward Category-Level Object Recognition*, ser. Lecture Notes in Computer Science, J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, Eds. Springer Berlin / Heidelberg, 2006, vol. 4170, pp. 127–144. [Online]. Available: http://dx.doi.org/10.1007/11957959_7
- [14] L. V. G. Herbert Bay, Tinne Tuytelaars, "Surf: Speeded up robust features," *ECCV*, 2006.
- [15] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher, Eds. Plenum Press, 1972, pp. 85–103.
- [16] V. Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of operations research*, vol. 4, no. 3, pp. 233–235, 1979.
- [17] P. Slavík, "A tight analysis of the greedy algorithm for set cover," *Journal of Algorithms*, vol. 25, no. 2, pp. 237 – 254, 1997. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WH3-45M8Y07-3/2/8910a0cff5a2ce3a17fb73af435eccc>