How was your day? Online visual workspace summaries using incremental clustering in topic space

Citation: Paul, Rohan, Daniela Rus, and Paul Newman. "How Was Your Day? Online Visual Workspace Summaries Using Incremental Clustering in Topic Space." 2012 IEEE International Conference on Robotics and Automation (May 2012).

As Published: http://dx.doi.org/10.1109/ICRA.2012.6224762

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/90841

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

How was your day? Online Visual Workspace Summaries using Incremental Clustering in Topic Space

Rohan Paul, Daniela Rus† and Paul Newman
Mobile Robotics Group. Oxford University, UK
†Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA
{rohanp, pnewman}@robots.ox.ac.uk and rus@csail.mit.edu

***Abstract*— Someday mobile robots will operate continually. Day after day, they will be in receipt of a never ending stream of images. In anticipation of this, this paper is about having a mobile robot generate apt and compact summaries of its life experience. We consider a robot moving around its environment both revisiting and exploring, accruing images as it goes. We describe how we can choose a subset of images to summarise the robot's cumulative visual experience. Moreover we show how to do this such that the time cost of generating an summary is largely independent of the total number of images processed. No one day is harder to summarise than any other.**

I. Introduction

Consider the following: a robot is sent out into the world day after day continually taking pictures of its environment, implicitly accruing an ever richer picture of its world. It is gaining experience. The question we ask in this paper is how should that robot summarise its day, its week or even its working "life time" when asked? Immediately, it is interesting to think of this as the flip side to the vast amount of research which exists on metric workspace mapping. That corpus of work summarizes the experience of a mobile robot metrically - it produces crisp, sometimes almost architectural drawings of the robot's workspace. In this work, however, we swap metric summaries for visual summaries. We want the robot to produce a story board of canonical images which capture the essence of the robot's visual experience - illustrating both what was ordinary and what was extraordinary. Here, we systematically address this question in a way that scales well with time and variation of experience. We seek a summary that evolves incrementally with the novelty of data - it should grow with saliency of experience and not merely duration. To be sure, if the robot stood still for a year in static world we would not welcome a lengthy precis!

At a high level we proceed in the following way. Each image is characterized as a mixture of visual topics, mapping to a point in topic vector space. We incrementally organize these images using an online graph clustering technique. The structure of this graph is used to generate a visual summary of a robot's experience. Importantly, the graphical organization evolves over time as new imagery is collected by the robot. We show that this naturally yields an ever-improving workspace summary.

II. Related work

The problem of generating visual summaries has been explored within the computer vision community. Gong and Liu [7] employ singular value decomposition to extract key-frames for video summarization applications. The procedure operates offline and requires batch access to the entire video stream. Pritch et al. [11], present a method for generating synopsis videos from static surveillance cameras. The procedure extracts moving objects tracks through background subtraction and selects the optimal summary by minimizing temporal and backgound consistency costs for object tracks.

In mobile robotics, Girdhar and Dudek [6], present a method for online extraction of k-most novel images from an image corpus using set-theoretic surprise, measuring the fitness of an image as a summary image. The approach requires the summary size to be specified and is aimed towards identifying salient aspects of data. Note, in the context of summary generation both the common and salient aspects must be represented and learnt online. In [12], Ranganathan et al. use bayesian surprise for identifying salient landmarks for topological mapping with vision and laser features. In another related work [10], Konolige et al. present view based maps, an online large-scale mapping technique for constructing topological maps with stereo data. The map is pruned by extracting relevant keyframes using a distance based heuristic causing the number of selected keyframes to scale with map length.

III. Image Representation in Topic Space

In this section, we discuss how an image can be encoded as a vector in topic space. The techniques employed here have their genesis in information retrieval and as such we will leverage an anology between documents and images. At the lowest level we can describe an image as a list of visual words using the approach by Sivic et al. [13]. We can now think of an image as a document of visual words represented as a point in vector space where each dictionary word represents an orthogonal axis.

Visual words in an image are not independent and arise from objects characterizing the scene. Features emanating from a common object, frequently co-occur across multiple images. Topic models [8] represent documents as a mixture of intermediate latent topics. Topics are distributions over words and probabilistically capture co-occurring features.

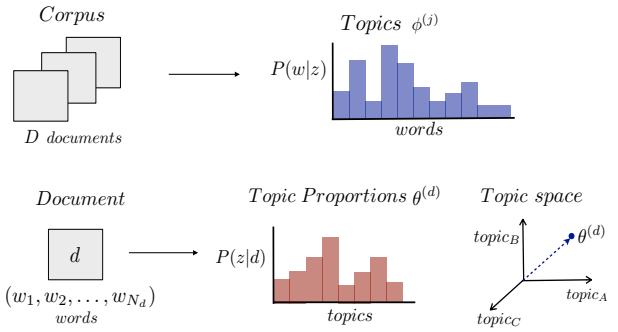


Fig. 1: Topic estimation and inference. Topics are distributions over words and estimated once from an image corpus. Learnt topics are used to estimate the vector of topic proportions for a perceived image, mapping to a point in topic vector space. Using a cosine similarity metric, online star clustering is used to organize images into topical clusters.

Each document or an image is a distribution over topics and different documents can possess varied topic proportions. Topic distributions are estimated once offline from a large corpus. Online, topic proportions are estimated for each image, see Figure 1. The vector of topic proportions maps an image to a point in topic-space. Typically, the number of topics is much less than the vocabulary size leading to considerable dimensionality reduction. Image similarity can be measured via cosine distance in topic space. Since topics provide a lower-dimensional thematic representation, images with common topics can get associated even if they have few words in common.

Latent Dirichlet Allocation (LDA) is a widely used probabilistic topic model [3] for which topic estimation is tractable. LDA is a hierarchical bayesian generative model and describes document formation as: (i) picking a multinomial distribution over topics specifying the likelihood of each topic in the document and (ii) generating constituent words by sampling topic proportions to obtain a topic label followed by sampling the word from the selected topic distribution over words. Inference involves reversing the generative process to recover the topics and the topic proportions per document. This is approximated using an MCMC Gibbs sampling procedure in the state space of topic labels for observed words with the update rule given in Equation 1. Here, z variable is a topic indicator variable, one for each observed word, w and α, β parameterize Dirichlet priors placed on topic and topic proportion distributions. The number of topics and vocabulary sizes are referred to as T and W. After sufficient sampling iterations, topic labels are recorded and used to form maximum likelihood multinomial estimates for topic and topic proportion distribution, [8].

$$P(z_i = j|\mathbf{z}_{-i}, \mathbf{w}) \quad \propto \quad \left[\frac{n_{-i,j}^{(w_i)} + \beta}{n_{-i,j}^{(\cdot)} + W\beta}\right]\left[\frac{n_j^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + T\alpha}\right] \quad (1)$$

After obtaining a suitable representation of images in topic space, our next task is to incrementally organize the robot's imagery and generate a visual summary of the traversal.

IV. Star Clustering and Online Organization

We use the star clustering algorithm [1] to compute a topic-driven organization of the robot's image collection. The star cluster algorithm is an efficient clustering algorithm that identifies the underlying thematic structure of a document collection and organizes it using topic clusters, as long as the documents can be compared using a similarity metric. When the similarity metric is the cosine distance between two feature vectors, the star clustering algorithm guarantees a minimum similarity between any pair of documents in the collection. Unlike the k-means algorithm, where the user has to specify in advance k, the number of final clusters, the star algorithm does not require as input the number of expected clusters; instead it discovers this number depending on the desired minimum similarity between the documents in the cluster. The star clustering algorithm can be run online and is computationally very efficient. The ability to incrementally determine the topic organization of an image collection makes it especially suitable for our problem setting, where the data collection from a mobile robot is incremental in nature. Next, we present a brief overview.

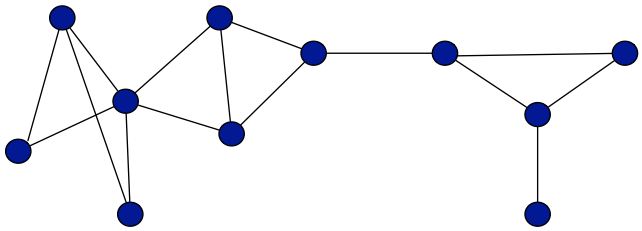
The image corpus is represented via a similarity graph, $G = (V, E, w)$ where vertices correspond to images and weighted edges represent cosine similarity in topic space. The similarity graph can be studied at various thresholds of pair-wise document similarity, σ. The thresholded graph, G_σ is obtained from G by removing edges with pairwise similarity less than σ, Figure 2a. The clustering algorithm covers G_σ with star-shaped subgraphs. A star-shaped subgraph on $m + 1$ vertices consists of a star center and m satellite vertices, where edges exist between the star center and each of the satellite vertices, Figure 2b.

The optimal clustering is obtained by forming a mimimal vertex cover for the graph with maximal star-subgraphs, Figure 2c, resulting in the following properties for each vertex: (i) a star center is not adjacent to another star center and (ii) every satellite vertex is adjacent to at least one center vertex of equal or higher degree. The number of clusters is naturally induced by the dense cover. For each cluster in the graph, the cluster center acts as its exemplar. Satellite vertices displaying multiple themes can be associated with multiple clusters.

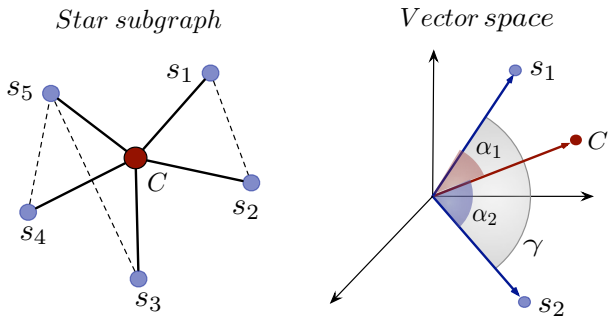
By examining the geometry of the star-subgraphs in the topic vector space, Figure 2b, the expected similarity between satellite vertices can be obtained as Equation 2. Here, $\cos\alpha_1$ and $\cos\alpha_2$ are the center-satellite similarities for any two satellites in the star and $\cos\gamma$ represents the expected satellite-satellite similarity. The expected pairwise similarities are high and imply dense clustering of data.

$$\cos\gamma \quad \geqslant \quad \cos\alpha_1\cos\alpha_2 + \frac{\sigma}{\sigma+1}\sin\alpha_1\sin\alpha_2 \quad (2)$$

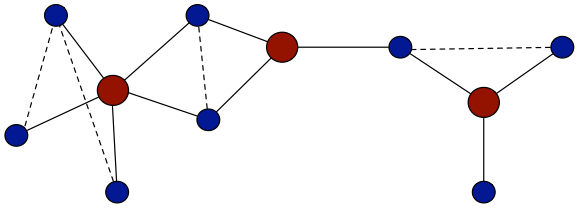
Star clustering is computationally efficient, asymptotically linear in the size of the input graph. Further, the star cover can be formed incrementally with each arriving data point with potential re-arrangement of existing stars, see Figure 3. For each inserted vertex, its degree and adjacency list is



(a) The clustering procedure begins by computing a similarity graph, G_σ with each image as a node with links indicating similarities exceeding a specified threshold σ.



(b) An example of a star-shaped subgraph with center C and five satellite vertices (left). Each node in the graph maps to a point in a vector space where pairwise similarity is endowed using cosine distance metric. By construction, center-satellite similarities are atleast σ. The vector space geometry with cosine distance ensures that expected satellite-satellite similarities are also high, leading to dense clusters.

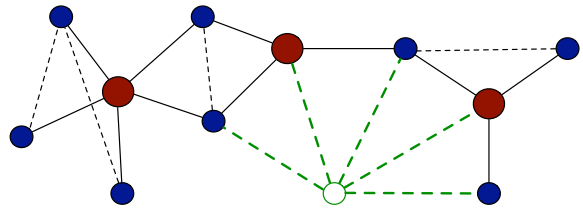


(c) The graph organized into clusters using a minimal cover with star-shaped sub-graphs. The cluster centers compactly summarize the visual experience of the robot. Note that an image (possessing varied themes) can belong to mulitiple clusters.

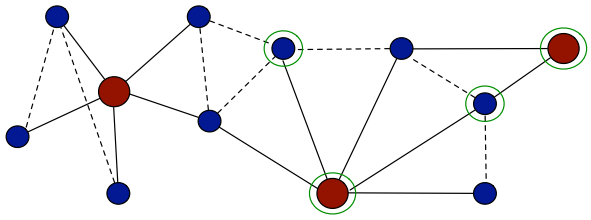
Fig. 2: Incremental clustering with star-shaped subgraphs.

computed and the following cases are examined: If the new vertex is not adjacent to a star center, then the inserted vertex is added as a star center forming a new cluster. If the inserted vertex is adjacent to a center vertex with higher degree, then the inserted vertex becomes the satellite for the center vertex. The graph is re-arranged in two cases: (i) when all centers adjacent to the inserted vertex have degree lower than the new vertex or (ii) vertex insertion increases the degree of an adjacent satellite beyond the degree of its associated star center. Under these conditioned, existing stars are broken and satellites are re-examined. However, the number of re-arrangement operations required are usually small (verified experimentally). Further, we employed an optimized version of the algorithm that saves operations by predicting the future status of a satellite vertex or other star-satellite status changes induced by the inserted vertex.

As images perceived the mobile robot are organized into



(a) A new data point may introduce additional links in the similarity graph (green) affecting adjacency and hence the validity of current minimal star cover.



(b) Inconsistent stars are broken and re-arranged to incorporate the new point. The green circles indicate positions where graph modifications took place. The number of stars broken determine the running time of insertion. On real graphs, the avg. number of stars broken is small, thereby yielding an efficient and incremental clustering approach.

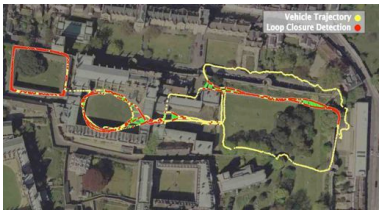
Fig. 3: Star clustering re-organization upon insertion of a new data point.

star clusters, at any time instant, the cluster centers form a visual summary of the robot's traversal. For each image in the corpus, the associated cluster centers provide a thematic annotation in terms of current summary images. Hence, the robot's trajectory can be understood as a combination of segments, each annotated by summary images to which the image are presently assigned. Since clusters adapt with each new collected image, the summary and the thematic annotation improves over time with increasing experience. Next, we bring the described components together and present experiments on data collected from a mobile platform.

V. Results

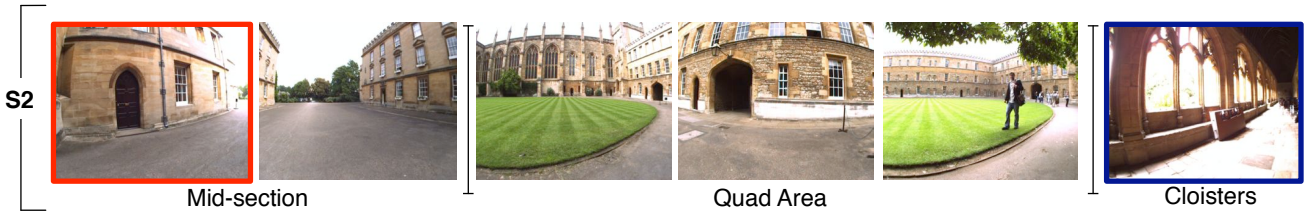
A. Vocabulary and Topic Learning

A data set traversing streets and park areas was collected consisting of 2874 images, recorded 10m apart and perpendicular to the robot's motion. Image samples were non-overlapping, excluded loop closures pairs and hence approximated independent samples from the observation distribution and used for vocabulary and topic learning. A visual vocabulary [13] of approximately 11k visual words was generated by clustering SURF features [2] extracted from this data set. Each image was represented as a multinomial of visual words by first extracting SURF features and then quantizing against the learnt vocabulary. Topic distributions were estimated using a Gibbs sampling procedure outlined in section III. The Markov chain was randomly intialized and was run till convergence for varying number of topics: ranging from 3 till 100. Dirichlet priors were set to $\alpha = 50/T$ and $\beta = 0.1$. Iterations required to ensure MCMC convergence was experimentally obtained to be atleast 200 and was found consistent across multiple re-starts. The number of topics

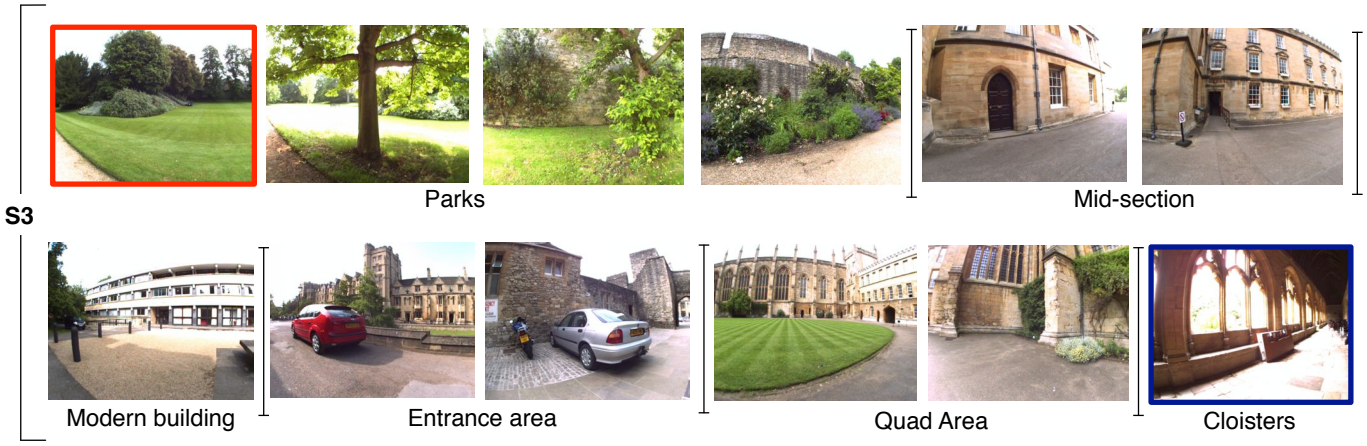


(a) Aerial view of New College with GPS plots for the robot's trajectory.

(b) Summary after traversing the cloisters with images of large windows, medieval walls etc.



(c) Summary at the start of the mid-section after traversing the quad and cloisters area.



(d) Summary at the end of the traversal including recent examples of parks, building and entrance areas.

Fig. 4: Incremental visual summary generation while traversing New College shown at three instants. The most recently added cluster center image and the first encountered are marked with red and blue borders respectively. Parameters: $\sigma = 0.6$, $T = 50$. The sections where the images were taken have been hand-labeled to facilitate interpretation. Note that clusters evolve over time and capture the dominant visual themes encountered by the mobile robot.

was selected through the bayesian model selection approach of maximizing the data-loglikelihood given topics [8] which was found to peak for 50 topics.

B. Visual Summaries

The visual summarization algorithm was run on two data sets: (i) New College data set consisting of 1355 images from cloisters, quad area, parks and facades characteristic medieval buildings in Oxford and (ii) City Center data set comprising of 1683 images taken in dynamic urban environments including roads, buildings, vehicles and pedestrians. There was no geographical overlap with the urban data set used for vocabulary and topic learning. For online experiments, images from the entire data set were presented sequentially and were incrementally organized into star clusters. Topic proportions for each image were estimated using topic distributions learnt from the urban data set.

The visual summary at three time instants for the New College data set is shown in Figure 4. The robot began operation in the cloisters area and after covering two loops took an exit into the adjacent quads. Figure 4b shows the summary at that instant consisting of images of dominant windows and stone walls seen in the area. The robot then traversed the quad area and reached the middle section. New clusters emerged as shown in Figure 4c containing varied views of the explored area. Note that images of walls found in the earlier summary are now absent. They were found similar to walls in the quad area and hence now appear as satellites. Figure 4d presents the summary at the end of the data set, containing new images from the parks and modern buildings seen later by the robot. The summary images are shown with labels indicating the geographical region where they were recorded. Note that the cluster centers summariz-

(a) Satellite image with overlaid GPS tracks for the City Center data set.

(b) Summary while exiting parklands after collecting 180 images

(c) Visual summary when the robot reaches Science Area buildings. New cluster centers for building features become promiment. Iteration 310.

(d) Visual summary at the end of traversal, iteration 900.

Fig. 5: Incremental visual summary generation for City Center data set shown at three instants. The most recently added cluster center image and the first encountered are marked with red and blue borders respectively. Parameters:$\sigma = 0.6$, $T = 50$. The sections where the images were taken have been hand labeled to facilitate interpretation. Note that clusters evolve over time and capture the dominant visual themes encountered by the mobile robot.

ing the traversal appear visually distinct indicating that the star covers capture different appearance modes present in the traversed environment. Figure 5 illustrates the summary for the City Center data set. The initial summary consists of foliage, parks and railings, Figure 5b. The robot then explores roads and building areas and hence the summary is refined with new representative clusters including vehicles, roads, buildings etc.

Every image collected by the robot is assigned to clusters in the collection, which can be considered as an annotation in terms of topical themes (cluster centers) learnt from exploration till now, see Figure 6. Since clusters evolve over time, the thematic annotation also improves with increasing experience. Note that star clustering permits multi-cluster membership accounting for images that can be explained via multiple themes in the data set.

C. Topical Clusters

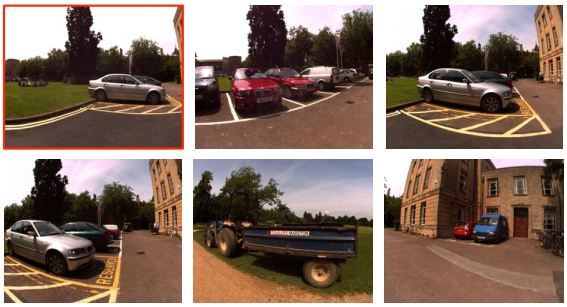
Figure 7 illustrates three representative clusters obtained at the end of the traversal. Cluster center image (indicated in red) and five randomly picked satellite images are shown. The cluster shown in Figure 7a typically consist of vehicles which generate a large number of features. The topic model learns that these feature co-occur and maps them to a common theme. The cluster shown in Figure 7b consists of similar images of foliage and trees in parks. Figure 7c presents a cluster containing images of buildings and trees as viewed from a sidewalk. Note that clusters possess a common visual theme as opposed to exact matches and hence topically organize images collected by the robot. Secondly, a relatively small number of topics (50) yielded topical clusters

Fig. 6: Example of images taken at three time instances during traversal shown with the associated cluster centers (using the final clustering at the end of traversal). The assigned cluster centers accurately capture the visual theme in the selected images (left column). Examine the first row. The observed parkland image (top left) is assigned to two cluster centers. The first center (middle) image was captured prior to the observed image and the second cluster center was collected later, indicating that the topical annotation for each image on the trajectory improves with time.

compared to the dictionary size of 10k, indicating significant dimensionality reduction.

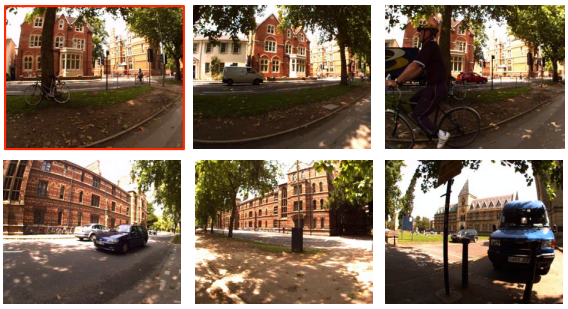
Using a higher similarity threshold, σ causes higher intra-cluster similarity, and generally results in smaller but more numerous clusters. Figure 8 compares two clusters obtained at $\sigma = 0.6$ and $\sigma = 0.7$, selected such that their respective cluster center images were taken at the same location in the New College quad. Both clusters possess a coherent visual theme consisting of medieval buildings with some foliage features. Cluster images for $\sigma = 0.7$ display higher similarity and are primarily from the same quad area, compared to the cluster at a lower threshold that consisted of images from the quad, mid-section and other parts of the college, hence possessing greater variability.

Next, we explored the cluster quality obtained at a specified σ. For each cluster, the distribution of all pair-wise similarities between satellite vertices was determined and probability histograms (bin size 0.025) were plotted vertically in Figure 9 using threshold $\sigma = 0.6$ and $\sigma = 0.7$ for the New College data set. To mitigate the effect of variable cluster sizes and sampling error, probability estimates were smoothened [5]. As discussed in Section IV, Equation 2 gives the expected similarity between satellite vertices in a star-subgraph. For a clustering at threshold, σ, the center satellite similarities are atleast σ. Hence, from Equation 2, the expected satellite-satellite similarity is σ, and is plotted as a horizontal line in Figure 9. Empirically, the expected similarity values for clusters were found close to σ indicating that star clusters are reasonably dense and imply high expected pairwise similarities between satellites.



(a) Clustered images consist of feature sets on vehicles and the horizon. City Center data set: Cluster 46, $\sigma = 0.8$.

(b) Thematic cluster of parkland images. New College data set: Cluster 8, $\sigma = 0.6$.



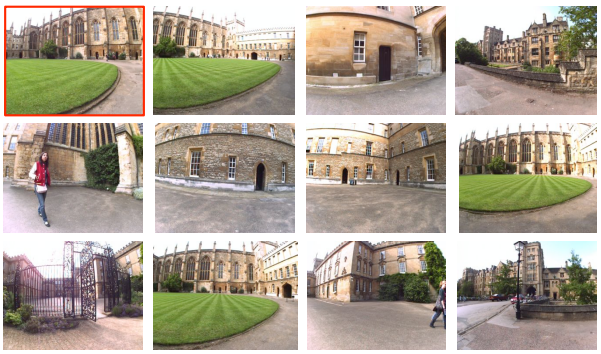
(c) Images in the cluster display common feature sets appearing on buildings and trees. City Center data set: Cluster 29, $\sigma = 0.7$.

Fig. 7: Representative clusters from City Center and New College Data sets. Images in a cluster possess similar visual topics. Cluster center is indicated in red. Images shown are randomly sampled from the total cluster images.

D. *Efficiency and Timing*

Table I presents online clustering statistics for the data sets with thresholds: 0.5, 0.6, 0.7 and 0.8. A higher similarity threshold reduces the number of edges in the graph (increasing sparsity) resulting in an increase in the number of clusters (size of the mimimal cover). The number of clusters obtained varied from 12 to 328 for the New College and from 14 to 454 for the city center data set for $\sigma = 0.5$ and $\sigma = 0.8$ respectively.

The average number of stars broken during insertion indicates the work done to re-arrange the existing graph when a data point is incorporated. Notably only a small number of stars are broken per insertion on average. For example, while inserting 1355 images in the New College data set at $\sigma = 0.6$, a total of 707 stars were broken - approximately



(a) Cluster 29, $\sigma = 0.6$. Displaying 12 of 202 cluster images.

(b) Cluster 29, $\sigma = 0.7$. Displaying 8 of 55 cluster images.

Fig. 8: Two clusters with centers in the New College Quad obtained with varying thresholds of $\sigma = 0.6$ and $\sigma = 0.7$. Cluster centers are marked with red. (a) Clustering at lower thresholds results in larger clusters with less specific visual themes. (b) Increasing the threshold results in smaller clusters with higher similarity.

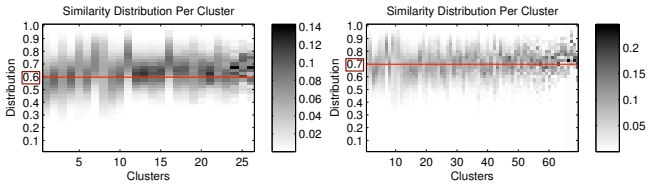
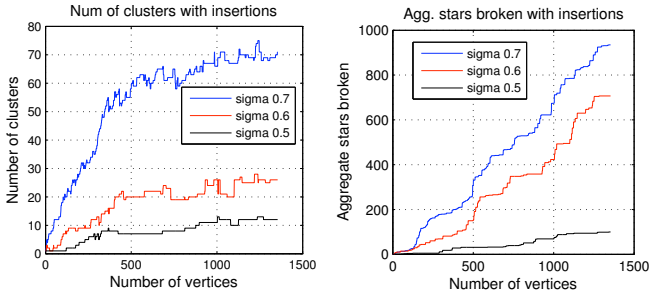


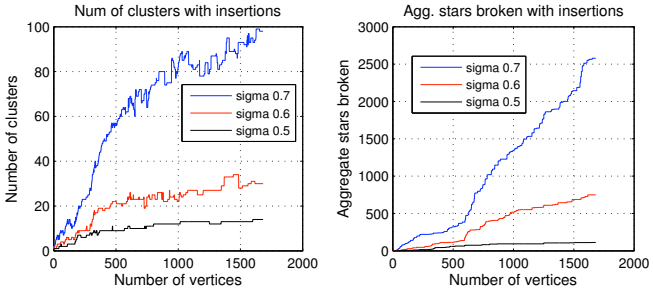
Fig. 9: Distributions of all pair-wise similarities between satellite vertices, (histogram plotted vertically) for each cluster obtained thresholds $\sigma = 0.6$ (left) and $\sigma = 0.7$ (right), indicated with horizontal red line, for the New College data set. Expected similarity values are close to σ indicating that star clusters are reasonably dense.

0.52 broken stars per insertion. The average number of stars broken were less than 0.78 for all runs except for $\sigma = 0.8$ experiment with City Center data set where a total of 2579 stars were broken (1.53 per iteration) while inserting 1689 images. The running time depends on the size of the graph, stars broken and the underlying similarity distribution for the data set. Total insertion time ranged from 0.24sec to 13.44sec yielding a small average insertion time of less than 10msec, making the approach practical for online operation.

Figure 10 plots the number of clusters and aggregate stars broken during each insertion iteration. Overall, the number of clusters increase over time as images are incrementally added. The cluster count grows rapidly as the robot begins exploring the environment. Over time, the clusters capture



(a)



(b)

Fig. 10: Number of clusters and aggregate number of stars broken during insertion iterations for (a) New College and (b) City Centre data sets.

the topical modes in the visual data and hence the growth rate shows a decline. Significant periods are observed when perceived images are added to existing clusters without increasing the total count and are interspersed with occasions when the count marginally increases or decreases when clusters are refined due to new vertices. The graph for a run with a lower threshold shows a more prominent saturation effect and always lies below the graph with a higher threshold.

Figure 11 plots the running time components for the City Center data set with $\sigma = 0.6$. The topic proportion inference time varies with the number of words in the scene and was on average 10.6msec per scene. For each vertex to be inserted the adjacency list is determined by computing its similarity to all existing nodes in the graph. The similarity computation time grows linearly with number of vertices and did not exceed 12msec during the experiment. The overall running time is dominated by the clustering algorithm and is low for most insertions (under 30msec). A few large peaks are observed during insertions when a large number of stars are re-arranged.

Figure 12 highlights the advantage of topic space representation over a basic bag-of-words representation. The image pair was taken during two visits to the same location. The second image shows a large number of features appearing on a bicycle which was absent during first visit. These images were found to be in the same clusters using the topic model representation ($\sigma = 0.5$) and in different clusters using visual words representation (even for a threshold as low as $\sigma = 0.05$). Since, the bag-of-words representation considers words independently, a large number of features observed

TABLE I: Statistics for online insertion with varying thresholds (topics, $T = 50$).

	New College Data set				City Center Data set			
Threshold	$\sigma = 0.5$	$\sigma = 0.6$	$\sigma = 0.7$	$\sigma = 0.8$	$\sigma = 0.5$	$\sigma = 0.6$	$\sigma = 0.7$	$\sigma = 0.8$
Number of clusters	12	26	71	328	14	30	98	454
Number of edges (x10^5)	3.96	1.87	0.69	0.20	7.22	3.54	1.20	0.26
Insertion time/iter (msec)	4.12	9.92	0.95	0.17	5.19	5.53	4.59	0.27
Total insertion time (sec)	5.59	13.44	1.28	0.24	8.74	9.31	7.74	0.45
Avg. stars broken/iter	0.07	0.52	0.69	0.70	0.06	0.44	1.53	0.78
Total stars broken	100	707	935	954	112	749	2579	1309

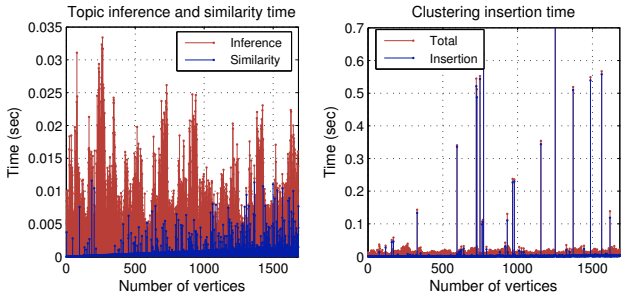


Fig. 11: Topic inference, similarity matrix computation time (sec) during insertion (left) and insertion time, total running time plots (right) for the City Center data set, $\sigma = 0.6$. Running time is primarily determined by the online insertion operation peaks are recorded in iterations when a large number of stars are broken. Note the scale in both graphs.

Fig. 12: Image pair captured while revisiting a location which was assigned to the same clusters using topic model representation and different clusters with the bag-of-visual words representation.

on the cycle makes the image pair highly dissimilar. As a contrast, employing the topic representation probabilistically captures co-occurring features and maps both images to common topics, assigning images as highly similar and consequently to the same clusters.

VI. Future Work

Future research will focus on scaling the algorithm to larger data sets and online adaptation to incoming data. While summarizing large topological maps [4] a major bottleneck is the adjacency computation for an inserted image using similarity to all previous images. We can approximate by comparing the cosine distance of an incoming image with only the cluster centers and using Equation 2 to infer the expected similarity to satellite vertices. thereby discarding clusters whose centers are highly dissimilar. Presently, visual topics are learnt *a-priori* from an image collection. We would like to adapt topics using online topic models [9] during silent periods where the robot is not collecting new imagery.

VII. Conclusions

In this paper we demonstrated an online incremental approach for generating visual summaries of a robot's workspace. We employed a topic vector space representation for images and an efficient graph-based star clustering algorithm for online organization into thematic clusters forming a compact summary for the robot's visual experience. Importantly, the thematic organization improves with new data collected by the robot resulting in an ever improving workspace summary.

VIII. Acknowledgements

Daniela Rus was supported for this work in parts by the MAST Project under ARL Grant W911NF-08-2-0004 and ONR MURI Grants N00014-09-1-1051 and N00014-09-1-1031. Paul Newman was supported by an EPSRC Leadership Fellowship, EPSRC Grant EP/I005021/1.

References

[1] J Aslam, E Pelekhov, and D Rus. The star clustering algorithm for static and dynamic information organization. *Journal of Graph Algorithms and Applications*, 8(1):95–129, Jan 2004.

[2] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded Up Robust Features. In *Proceedings of the 9th European Conference on Computer Vision*, volume 13, pages 404–417, Graz, Austria, May 7 2006.

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

[4] M. Cummins and P. Newman. Highly scalable appearance-only SLAM - FAB-MAP 2.0. In *Proceedings of Robotics: Science and Systems*, Seattle, USA, June 2009.

[5] J. Cussens. Bayes and Pseudo-Bayes Estimates of Conditional Probabilities and Their Reliability. *Lecture Notes in Computer Science*, pages 136–136, 1993.

[6] Y Girdhar and G Dudek. Online navigation summaries. *International Conference on Robotics and Automation*, Jan 2010.

[7] Y Gong and X Liu. Video summarization and retrieval using singular value decomposition. *Multimedia Systems*, Jan 2003.

[8] T Griffiths and M Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl 1):5228, Jan 2004.

[9] M.D. Hoffman, D.M. Blei, and F. Bach. Online learning for latent dirichlet allocation. *Advances in Neural Information Processing Systems*, 23:856–864, 2010.

[10] K Konolige, J Bowman, and J D Chen. View-based maps. *International Journal of Robotics Research*, Jan 2010.

[11] Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg. Webcam synopsis: Peeking around the world. In *IEEE 11th International Conference on Computer Vision, 2007*, pages 1–8, 2007.

[12] A Ranganathan and F Dellaert. Bayesian surprise and landmark detection. *International Conference on Robotics and Automation*, Jan 2009.

[13] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision*, Nice, France, October 2003.