Automatic Landmark Detection for Topological Mapping Using Bayesian Surprise

Ananth Ranganathan and Frank Dellaert
{ananth,dellaert}@cc.gatech.edu
College of Computing, Georgia Institute of Technology
Technical Report No. GT-IC-08-04

***Abstract*— Topological maps are graphical representations of the environment consisting of nodes that denote landmarks, and edges that represent the connectivity between the landmarks. Automatic detection of landmarks, usually special places in the environment such as gateways, in a general, sensor-independent manner has proven to be a difficult task. We present a landmark detection scheme based on the notion of "surprise" that addresses these issues. The surprise associated with a measurement is defined as the change in the current model upon updating it using the measurement. We demonstrate that surprise is large when sudden changes in the environment occur, and hence, is a good indicator of landmarks. We evaluate our landmark detector using appearance and laser measurements both qualitatively and quantitatively. Part of this evaluation is performed in the context of a topological mapping algorithm, thus demonstrating the practical applicability of the detector.**

I. Introduction

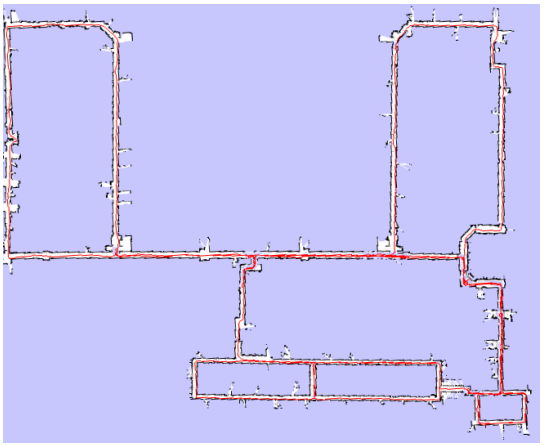
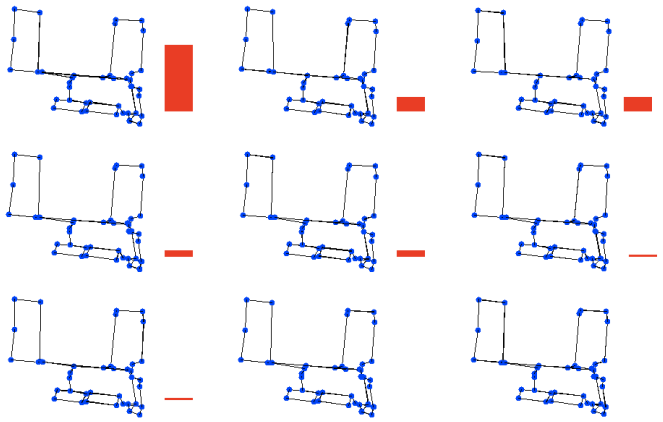
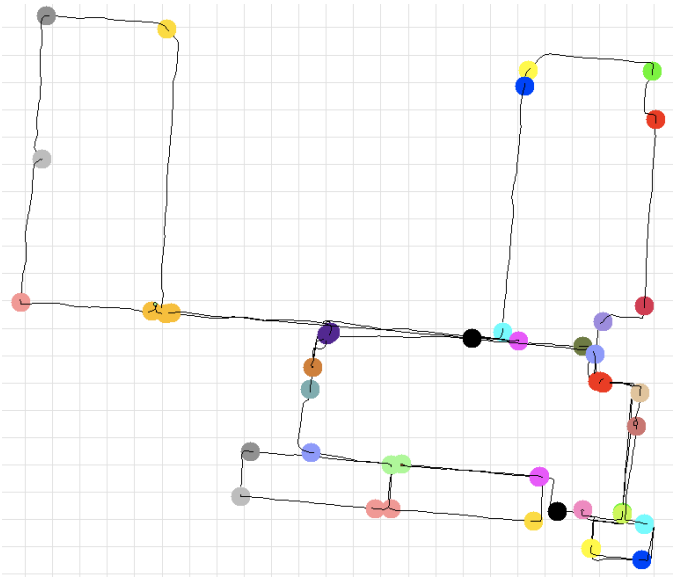


Fig. 1. Metric map of Killian Court dataset [3] produced using the iSAM algorithm [9] shown as reference for comparison with Figure 2

Topological mapping is the process of using a robot to automatically discover the topological structure of an environment. In its simplest form, this topological structure consists of a graph where the nodes denote certain distinguishable places in the environment, and edges denote connectivity. Topological maps are well suited to robotics applications since they are a sparse representation that scale well with environment size. Further, topological maps are amenable to the inclusion of higher level semantic concepts such as objects [19] and navigation techniques [11]. This is in contrast to metric maps that increase rapidly in complexity with the size of the environment, and also face brittleness due to incremental acumulation of errors.



(a)



(b)

Fig. 2. (a) PTM for the MIT Killian Court dataset with automatic landmark detection using Bayesian surprise. The topology at the top left with the maximum probability is the ground truth (b) The smoothed trajectory corresponding to the ground truth topology.

Topological mapping can be decomposed into two orthogonal sub-problems - overcoming topological ambiguity and

landmark detection. Topological ambiguity arises when multiple topologies are equally correct relative to the measurements obtained from the robot. Ambiguity occurs when multiple places in the environment look the same, called perceptual aliasing, or the same place looks different at different times. Most existing topological mapping techniques attempt to overcome topological ambiguity with methods ranging from maximum likelihood data association [22] to inference over the space of topologies [20]. This has resulted in a good understanding of the reasons for topological ambiguity and various solutions for overcoming it.

This paper deals with landmark detection. Compared to topological ambiguity, landmark detection has received relatively less attention. This is mainly because of the tenuous definition of what a landmark is; the most common being simply that landmarks are "special places" in the environment. To sidestep this issue, landmarks are frequently defined using *ad hoc* heuristics based on individual sensor characteristics. The resulting techniques are limited in being tied to a single sensor, functioning only in certain environments, and producing so many false positives as to destroy the sparsity of the topological graph. Even when landmarks are defined using invariant geometric properties of the environments [2][21], general-purpose algorithms based on these properties, in the sense of the limitations mentioned above, do not exist.

Landmark detection is a difficult problem since people often define and locate landmarks using myriad higher-level semantic concepts such as billboards and signs in outdoor environments, and objects and their relative locations in indoor environments. The detection and use of such diverse clues for automatic landmark detection is not currently possible. Hence, this rules out "perfect", human-level landmark detection. In lieu of this, we define a good landmark detector to be one that has negligibly few false negatives while producing a tolerable number of false positives, i.e. the landmark detector fires at almost all the locations that would be viewed as landmarks by a human while also firing at some locations that would not.

As the primary contribution of this paper, we propose the first ever general-purpose, Bayesian landmark detection scheme that is agnostic to the type of sensors used. This scheme also satisfies the above mentioned metric of having almost no false negatives while producing a tolerable number of false positives. Our scheme is based on the notion of "surprise", first proposed by Itti and Baldi [7]. Surprise encodes the unexpectedness of a measurement, the premise being that unexpected, and hence highly surprising, measurements arise from landmark locations. Surprise has been shown to be a good predictor of directed human attention [8]. It is, thus, naturally well-suited for landmark detection, since landmarks are places that attract human attention and are preserved in memory for use in future navigation tasks.

Sensor-independence of our surprise-based landmark detection scheme is obtained through its computation in a Bayesian framework. Bayesian surprise supports the inclusion of measurements from multiple, distinct sensor sources, the only requirement being that a measurement model is defined each of the sensors. We demonstrate the use of appearance measurements obtained from camera images, and laser range scans, for landmark detection in this context. Further, the computational framework for Bayesian surprise, which is based on KL-divergence, is exceedingly simple and computationally efficient, and is described in Section II. These advantages make surprise-based landmark detection attractive and a significant improvement to the state of the art.

As a secondary contribution, we incorporate our landmark detection scheme into the topological mapping algorithm given by Ranganathan et al. [18] to produce a complete topological mapping system. Landmark detection is evaluated in the context of this mapping system using various sensors on a number of environments, including publicly available datasets that are well-known in the robotic mapping community. An analysis of the number of false negatives and false positives output by the technique is also presented.

Among existing landmark detection techniques, many use geometric invariants of the environment such as intersection of Voronoi cells [4]. However, the use of such features may introduce a large number of landmarks in the map, thus destroying the sparse nature of the topological map. Beeson et. al. [2] overcome the problem of too many false positive landmarks by judiciously pruning the Voronoi graph so that spurious nodal points are not classified as landmarks. The use of sensor specific measures of distinctiveness for landmark detection is common, for instance Kortenkamp [10] uses range scans while Ramos et al. [17] use camera images. This leads to landmark detectors that use very specific features of the environment such as open doors and orthogonal walls, and moreover, are bound to a particular sensor [5]. Kuipers and Beeson [12] present a bootstrap algorithm for place modeling based on image clustering and learning the topology of the image locations. All these methods have the drawback of being applicable to a particular sensor or specific type of environment. Surprise-based landmark detection attempts to overcome this limitation.

II. Bayesian Surprise

"Surprise" can be said to quantify the unlikeliness of measurements according to the current model of the environment. We base our surprise computation on the method proposed by Itti and Baldi [7]. Consider the model at the current time as M and a prior distribution on the space of all possible models $P(M)$. Upon receiving a measurement z, the prior is updated to obtain a posterior on model space $P(M|z)$ using Bayes law

$$P(M|z) \quad = \quad \frac{P(z|M)P(M)}{P(z)}$$

Surprise is defined as the change in the belief in the model upon observing the measurement. Clearly if the posterior is the same as the prior, there is zero surprise. This intuitive description of surprise can be made concrete by defining it as the KL-divergence between the prior and posterior distributions on model space, i.e.

$$S(z) \quad = \quad \int_M P(M) \log \frac{P(M)}{P(M|z)} \tag{1}$$

Note that the integral is over the space of all possible models. The computation of surprise using the above equation is inherently recursive as the posterior in one step becomes the prior for the subsequent step.

This definition of surprise is intuitive in the sense that if a measurement that is surprising at first is observed repeatedly, it loses its surprising nature. Such operation is required when we apply surprise to landmark detection as the landmark detector should fire only when the robot moves into a new area.

III. Landmark Detection Using Surprise

We propose the definition of landmarks as places that yield highly surprising measurements. This implies the existence of a threshold, where a place is classified to be a landmark if its surprise value exceeds this threshold. The critical component here is, hence, a procedure to determine this threshold automatically for various environments and sensors.

The surprise threshold is defined in a general, adaptive manner by comparing with the expected surprise with the actual obtained value. Computing the expected surprise in closed form is not possible as it involves integrating (1) over all possible measurements z. Instead, we employ a Monte Carlo approximation to the integral wherein N measurements $z_{1:N}$ are sampled from the current place model $P(M)$, and the expected surprise is taken to be the average of the surprise values corresponding to these samples

$$E(z) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} S(z_i) \tag{2}$$

Further, the standard deviation σ for these surprise values is also computed. We postulate that any maximum of the actual surprise values exceeding a 3σ deviation from the expected value is a landmark. A maximum is chosen since a number of measurements in sequence may yield surprise values beyond the threshold. In this case, only the most surprising measurement need to characterized as arising from a landmark.

Before illustrating landmark detection using specific sensors, we next introduce the topological mapping algorithm used to evaluate the landmark detector in a practical scenario.

IV. Probabilistic Topological Maps

We evaluate our surprise-based landmark detection in the context of topological mapping to obtain realistic results. For this purpose, the topological mapping algorithm proposed by Ranganathan and Dellaert [18] is used. This algorithm constructs a distribution over the space of all possible topologies and hence, solves the problem of topological ambiguity. While the space of topologies is combinatorial, this is overcome by using Monte Carlo sampling techniques - in this case particle filtering - to make the algorithm tractable. The sample based posterior distribution over the space of topologies, which essentially consists of a set of topologies along with their probabilities, is called a Probabilistic Topological Map (PTM). By recording the ambiguity associated with each map in the form of its probability, a PTM provides a fail-safe mechanism to establish the correctness of the map.

PTMs are ideal in many ways for testing landmark detection. First, there are generalizable to various sensors, and hence, can be used with the appearance and laser measurements described above. Second, since PTMs are sensor independent, plugging in a landmark detector is easy, as compared to other mapping schemes. Third, PTMs only address the problem of topological ambiguity while landmark detection is unresolved. Incorporating our detection scheme results in a complete, probabilistic topological mapping system.

The topological mapping system with our landmark detection scheme alongwith the PTM works as follows. At each step, surprise computation is performed to determine if the current location is a landmark. If this is the case, a new landmark is added to the existing PTM, and the particle filtering algorithm is invoked to perform inference in the space of topologies, which results in an updated PTM.

In the following sections, we describe surprise-based landmark detection using laser and appearance measurements, though other sensors can also be incorporated similarly.

V. Laser based Surprise Computation

We now provide a landmark detection scheme using laser range scans that is based on the computation of Bayesian surprise. Firstly, we convert the laser scans to a representation that can be used to model places. Using a very simple representation, place modeling is performed using the area of laser scans as measurements.

The area contained in a laser scan can be computed by triangulation followed by computation of the areas of the triangles which are summed up to obtain the desired area. Since in most cases, only a single laser is available, the robot has a forward facing view of the world. This implies that if the robot were to approach the same place from a different direction, the place models would not match. We get around this problem by building map patches incrementally around each place as the robot moves. The areas of these patches give an omni-directional, orientation-independent model for places.

Since a place in a topology does not imply a precise metric location, the area measured by laser scans in the same place will differ slightly due to the robot not being in exactly the same location. This uncertainty is modeled using a Gaussian distribution, which is the parametric model distribution used for computing Bayesian surprise.

Given the above model, the computation of surprise is straight-forward. The actual KL-divergence between two Gaussian distributions, which is the Bayesian surprise in this case, is computed as follows

$$KL(p||q) \quad = \quad 0.5\log\frac{\sigma_p^2}{\sigma_q^2} + \frac{\mu_q^2+\mu_p^2+\sigma_p^2-2\mu_p\mu_q}{2\sigma_p^2} - 0.5 \tag{3}$$

Landmark detection using surprise computed from (3) is performed as follows. At each step, a number of measurements for the area of the place are made bsed on the current Gaussian model. The expected surprise and its standard deviation are

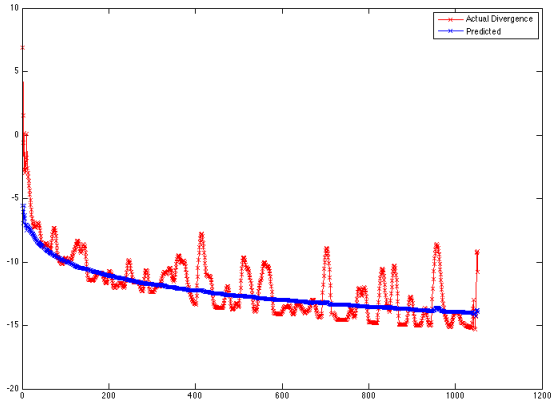


Fig. 3. Actual and expected KL-divergence (surprise) for the first experiment using laser measurements plotted on a logarithmic scale. 15 landmarks are detected in total. The standard deviation of the expected surprise is very small (~0.01) and is not shown here.

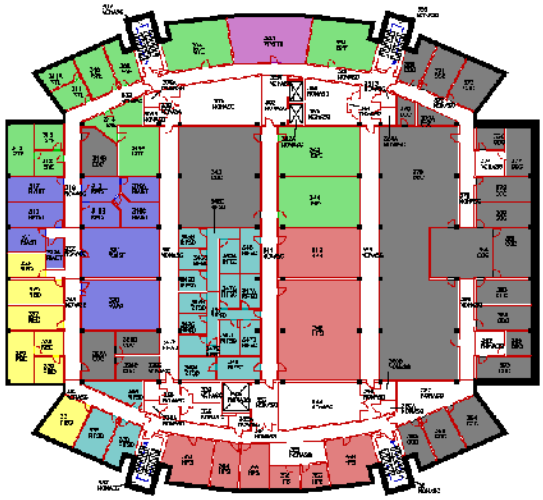


Fig. 4. Floorplan of experimental area for first experiment using laser.

computed from these sampled measurements using (2) and (3). This gives a threshold for the actual surprise computed when the real measurement is obtained. If the actual surprise is greater than the expected surprise by more than three times the standard deviation, the place is declared to be a landmark.

A. Results

We now present results using our surprise-based landmark detection scheme. Surprise was computed for laser scans using (3). The decision of whether a surprise value corresponds to a landmark is made by computing the expected surprise for a given model as described above and in Section III.

The laser-based Bayesian surprise computation was applied to the another indoor building environment, whose layout is shown in Figure 4. The dataset contains a total of 2106 laser scans. The actual and expected surprise for each step are shown in Figure 3. 15 landmarks were detected in total. The PTM obtained using these landmarks has the ground truth topology as the most likely one, receiving 64% of the probability mass, as shown in Figure 5. The smoothed trajectories corresponding to a few of the topologies in the PTM are also shown in Figure 5. Landmarks at the corners are detected when the laser sees around the corner for the first time, and hence, anticipate the actual corners slightly. The number of landmarks and their placement is almost perfect in this case.

We next apply the landmark detection scheme to the MIT Killian Court dataset [3] which is another widely used dataset in the SLAM community. The dataset consists of 1941 poses and corresponding laser scans. The ground-truth metric map with laser scans and robot trajectory is shown in Figure 1 for reference. A total of 61 landmarks were detected using laser-based surprise and the PTM obtained using these landmarks, which also contains the ground truth as the most likely topology, is shown in Figure 2. The ground truth receives 81% of the probability mass. Figure 2(b) gives the trajectory smoothed with the topological constraints and also the color-coded nodes as before. It can be seen that only a few false positives are found, and crucially, all the actual landmarks, i.e. the junctions and gateways, are accurately detected. The robot trajectory in this dataset spans an area of more than 200x200 meters and is considered challenging for metric mapping algorithms. It is however, a relatively easy sequence for performing topological mapping due to the wide separation between most landmarks, thus illustrating the advantage of a topological map over metric maps in this case.

VI. Computing Surprise Using Appearance

While laser range scanners are currently the *de facto* standard in robotic sensors, we now show how bayesian surprise can also be computed using appearance measurements obtained from camera images. However, modeling appearance measurements for this purpose is more complicated than the corresponding laser-bsed scenario.

Appearance measurements are obtained using images from an eight camera rig, shown in Figure 6. Two types of features are detected on the images; the Harris Affine features by Mikolajczyk and Schmid [15], and the Maximally Stable Extremal Regions (MSER) by Matas et. al. [14]. The reason for two types of features is their complementary nature that ensures that both affine-invariant features and regions of intensity maxima are detected, thus ensuring a relatively dense representation of the images in feature space. All the features are subsequently transformed to a 128-dimensional vector space using SIFT descriptors [13].

Each panoramic image, obtained by combining the images from the rig, is represented using a bag-of-words model [23]. Appearance "words" are obtained from the SIFT descriptors using vector quantization, where the number of bins in the vector quantization corresponds to the number of words in a text document. Vector quantization is performed using the K-means algorithm, and is done as batch process over all the features detected across all the images. Each panoramic image is, subsequently, transformed into a histogram of word counts in each of the bins. Thus, the representation of an image

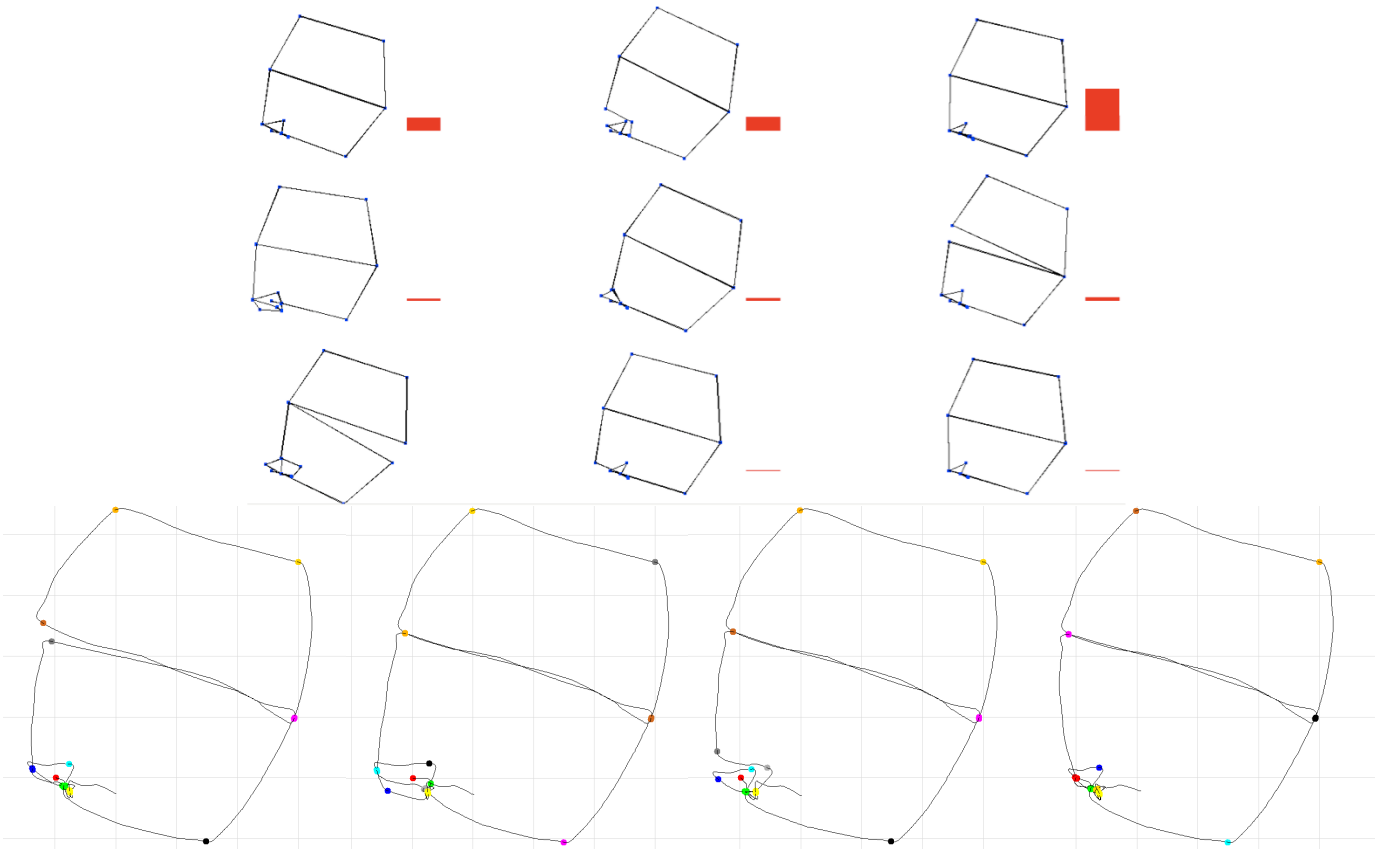


Fig. 5. PTM for the first laser experiment with automatic landmark detection using Bayesian surprise (top). The topology at the top right with the maximum probability is the ground truth. (bottom) The smoothed trajectories for the top four most probable topologies. Nodes belonging to the same physical landmark are colored similarly.

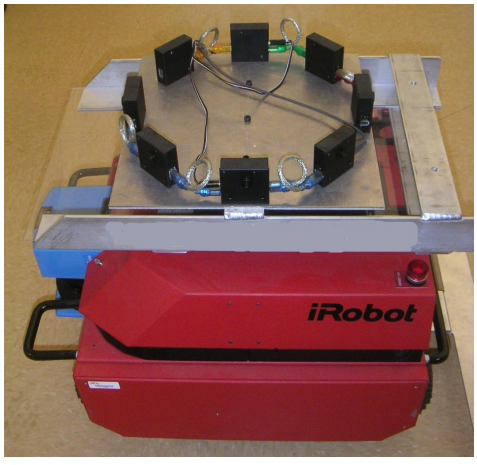


Fig. 6. The camera rig and the robot used to obtain panoramic images

in a bag-of-words model is a vector of word counts, which comprise a histogram.

A. Modeling Places Using The Multivariate Polya Model

We consider the SIFT histograms, obtained from images taken from a place, to be measurements of the appearance of the place. A model of this place can be obtained by assuming a histogram clustering model.

We model all the images arising from a landmark as having the same underlying "cause". Since the measurements are histograms of word counts, they are modeled using a multinomial distribution having dimensions equal to the number of appearance words. Further the prior over the multinomial parameter is the conjugate Dirichlet distribution to aid in ease of computation. Hence, the Dirichlet parameter is the underlying "cause" of the appearance measurements from a landmark. Given a set of appearance measurements $A = \{a\}$ from a landmark, the model $P(\alpha|A)$ can be written using Bayes law as

$$P(\alpha|A) \quad \propto \quad P(A|\alpha)P(\alpha)$$

and the likelihood of the histogram measurements $P(A|\alpha)$ can be expanded so that the above equation becomes

$$P(\alpha|A) \quad \propto \quad P(\alpha)\int_{\theta} P(\theta|\alpha)\prod_{a\in A} P(a|\theta) \tag{4}$$

where $\theta = [\theta_1, \theta_2, \ldots, \theta_W]$ and $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_W]$ are the multinomial parameter and Dirichlet prior respectively, and a denotes the SIFT histogram measurement with bin counts given as $[n_1, n_2, \ldots, n_W]$. The number of distinct appearance words is denoted as W, while the prior on α is taken to be uniform. Hence the distributions in the integrand above are

$$p(a|\theta_s) \quad = \quad \frac{n!}{n_1!n_2!\ldots n_W!}\theta_{s1}^{n_1}\theta_{s2}^{n_2}\ldots\theta_{sW}^{n_W} \tag{5}$$

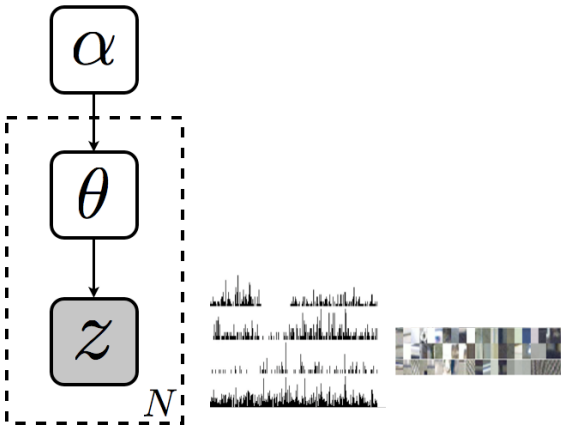


Fig. 7. Graphical model illustrating the Multivariate Polya distribution. To obtain a measurement z, which is a quantized SIFT histogram, we first sample from a Dirichlet distribution with parameter α to obtain a Multinomial vector θ . This Multinomial distribution is, in turn, sampled to obtain the measurement histogram z. Note that a different θ has to be sample for each z. For visual effect, some sample histogram measurements are shown alongside the graphical model. Image patches centered on the SIFT features corresponding to the top 20 histogram bins are also displayed.

$$p(\theta_s|\alpha_s) \quad = \quad \frac{\Gamma(\sum_{w=1}^{W}\alpha_{sw})}{\sum_{w=1}^{W}\Gamma(\alpha_{sw})}\theta_{s1}^{\alpha_{s1}-1}\theta_{s2}^{\alpha_{s2}-1}\ldots\theta_{sW}^{\alpha_{sW}-1} \quad (6)$$

The expanded likelihood model in (4), where $P(a|\theta)$ is a multinomial distribution (5) and $P(\theta|\alpha)$ is a Dirichlet distribution (6), is called the Multivariate Polya model, or equivalently in document modeling, the Dirichlet Compound Multinomial model [1]. The integration in (4) can be performed in closed form since the Dirichlet process is the conjugate prior of the multinomial distribution. Using the definitions of the distributions (5) and (6) in (4), this yields the final form of the place model as

$$P(\alpha|A) \quad \propto \quad \frac{n!}{\prod_{w=1}^{W} n_w}\frac{\Gamma(\alpha)}{\Gamma(n+\alpha)}\prod_{w=1}^{W}\frac{\Gamma(n_w+\alpha_w)}{\Gamma(\alpha_w)} \quad (7)$$

where n_w is the count of the wth appearance word across all the SIFT histograms in A and $n=\sum_w n_w$, $\alpha=\sum_w \alpha_w$. $\Gamma(.)$ denotes the Gamma function. Graphical intuition for the Multivariate Polya model is provided by Figure 7.

Given a set of D images with features detected on them, the maximum likelihood value for α can be learned by using iterative gradient descent optimization. It can be shown that this leads to the following fixed point update [16]

$$\alpha_w^{new} \quad = \quad \alpha_w\frac{\sum_{d=1}^{D}\psi(n_{dw}+\alpha_w)-\psi(\alpha_w)}{\sum_{d=1}^{D}\psi(n_{dw}+\alpha)-\psi(\alpha)} \quad (8)$$

where $\alpha=\sum_w \alpha_w$ as before, and $\psi(.)$ is the Digamma function, the derivative of the Gamma function.

B. Surprise Computation

We now apply the theory of surprise to the Multivariate Polya model discussed above. Consider the situation where the set of histogram measurements $A=\{a_i|1\leq i\leq n\}$ has been observed. The prior model for surprise computation is then simply the Multivariate Polya model learnt using A. If now a measurement z is observed, the posterior is the Multivariate Polya model learnt using the measurements $\{A,z\}$. Surprise can be computed per (1) as

$$S(z) \quad = \quad \int_a P(a|\alpha_{ML})\log\frac{P(a|\alpha_{ML})}{P(a|\alpha_{MAP})} \quad (9)$$

where α_{ML} is the maximum likelihood parameter learned using measurements A as given in (8), and α_{MAP} is the corresponding parameter learned using $\{A,z\}$

The computation of the KL divergence using (9) is still not possible in closed form due to the form of the Multivariate Polya model. We now briefly summarize the exponential family approximation to the Multivariate Polya model given by Elkan [6]. Using this approximation, surprise can be computed in closed form.

C. Exponential Family Approximation

Empirically, the learned values of α is usually such that $\alpha_w \ll 1$ in most cases. For small α, the following approximation holds

$$\frac{\Gamma(x+\alpha)}{\Gamma(\alpha)}-\Gamma(x)\alpha \quad = \quad 0$$

so that we can substitute $\frac{\Gamma(x+\alpha)}{\Gamma(\alpha)}$ by $\Gamma(x)\alpha$. Also using the fact that $\Gamma(z)=(z-1)!$ in (7) yields the exponential family approximation to the Mutivariate Polya model

$$q(a) \quad = \quad \frac{n!}{\prod_{w:n_w\geq 1} a_w}\frac{\Gamma(s)}{\Gamma(s+n)}\prod_{w:n_w\geq 1}\beta_w \quad (10)$$

where the parameters have been denoted as β instead of α following Elkan [6] to distinguish them from the exact model , and $s=\sum_w \beta_w$. More details of the exponential nature of the above distribution can be found in [6].

Given a collection of documents the maximum likelihood value of β can be learned in a similar manner to (8) using iterative fixed point equations as follows

$$s \quad = \quad \frac{\sum_w\sum_d I(n_{dw}\geq 1)}{\sum_d \psi(s+n_d)-|D|\psi(s)} \quad (11)$$

$$\beta_w \quad = \quad \frac{\sum_d I(n_{dw}\geq 1)}{\sum_d \psi(s+n_d)-|D|\psi(s)} \quad (12)$$

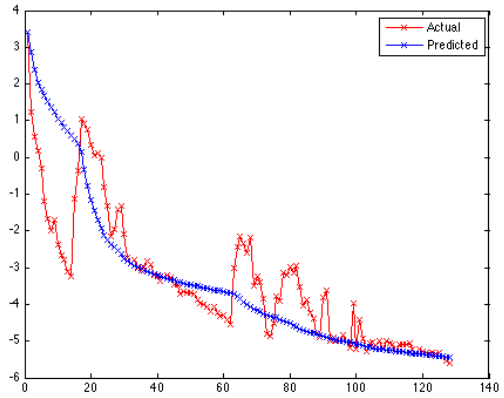
where $I(.)$ is the indicator function.

D. A Closed-form Expression for Surprise

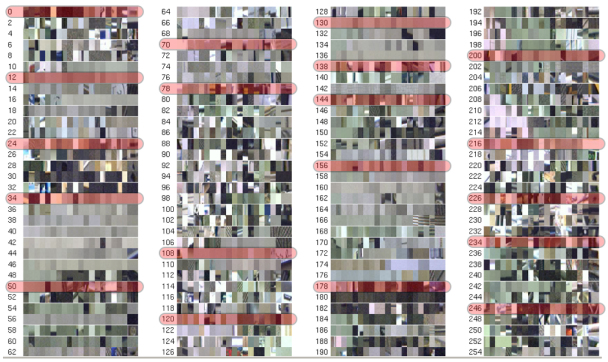
Given the above discussion, we can now compute the KL-divergence between two exponential family Polya models using the expression for the model (10). The calculation is straight-forward using basic properties of exponential family distributions and is omitted here for brevity. The final KL-divergence expression is given as

$$KL(p||q) \quad = \quad \log\frac{\Gamma(s_q+n)}{\Gamma(s_p+n)}-\log\frac{\Gamma(s_q)}{\Gamma(s_p)}- \\ \left(\psi(s_q+n)-\psi(s_q)\right)\sum_{w=1}^{W}\beta_w^q\log\frac{\beta_w^q}{\beta_w^p} \quad (13)$$

where p and q are the two distributions respectively.



(a)



(b)

Fig. 8. (a) Actual and predictive KL-divergences for the TSRB dataset plotted on a log scale. The variances for the predictive divergences are so small that 3σ curves are hard to view at this scale. (b) Top 20 SIFT features by histogram count for each location denoted by the measurement number. Only every second measurement is shown. The measurements corresponding to landmarks (i.e. where the landmark detector fires) are shown in red (shaded overlay). It can be seen that these correspond to the start of sub-sequences of measurements that differ from the preceding measurements.

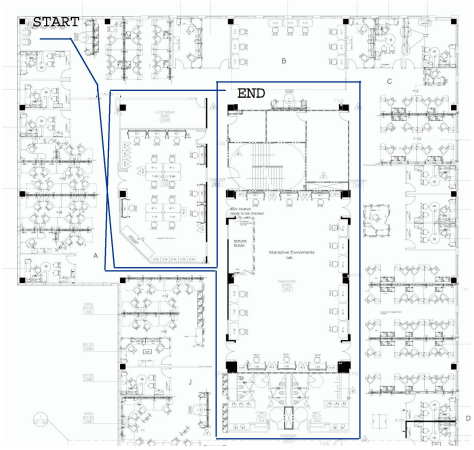


Fig. 9. Floor plan with approximate robot path overlaid for the first appearance experiment.

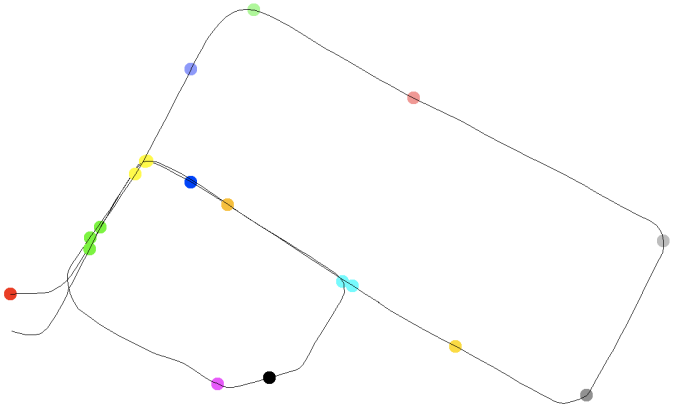


Fig. 10. Topological map, showing landmarks detected using Bayesian surprise computation. The smoothed trajectory is also shown. Nodes belonging to the same physical landmark are colored similarly.

Surprise, as defined for the Multivariate Polya model in (9), can be computed using the above equation. The parameter values are learned for the prior distribution using all the measurements observed upto the current time. The posterior parameter is learned similarly, but by also adding the current measurement to the dataset. The KL-divergence between these distributions, which is the surprise, is computed using (13).

E. Results

The above landmark detection scheme was applied to a robot run in the building where our lab is situated. The building floorplan with hand-drawn robot trajectory is shown in Figure 9. SIFT features were detected on images obtained from the camera rig and appearance words computed in exactly the same fashion as Section VI with 1024 appearance words being computed using K-means clustering. The topological mapping algorithm described in [18] was used to compute the map using these landmarks.

The expected and the actual surprise values are shown in Figure 8(a). The figure also shows the top twenty SIFT features from the appearance histogram for certain places. A total of 19 landmarks were detected in this dataset, and the topology obtained using the landmarks is the same as the ground truth topology as is shown in Figure 10 along with the smoothed trajectory. Colors of the nodes depict correspondence, so that nodes classified as being the same place are colored similarly. Note that all the decision points are classified as landmarks, while a few false positives also exist. The number of false positives is quire small since a number of landmarks that appear to be false positives are, in fact, gateway locations. Mosaics of a few of these landmarks in Figure 11 show that they indeed correspond to locations that are qualitatively different from their surrounding areas.

The operation of the landmark detector was quantitatively tested wrt the number of false negatives and false positives. For this purpose, a number of robot runs with laser were performed in the environment of Figure 4 and with cameras in the environment of Figure 9. A total of 7000 laser measurements and 1371 panoramic images were obtained. Gateways were marked manually as landmarks and the results of the landmark detectors were compared against this ground truth labeled data. Results are shown in the form of contigency tables in Figure 12. It can be seen that the number of false negatives is very

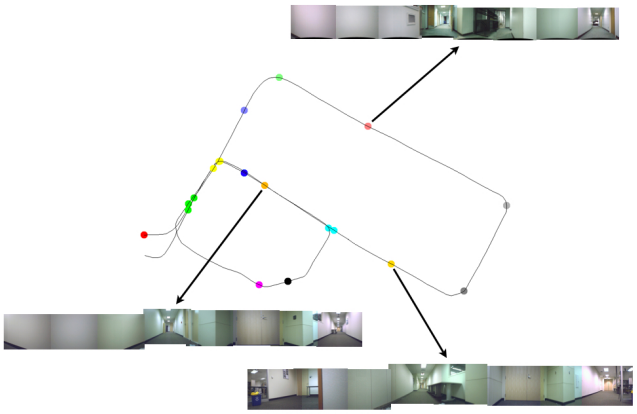


Fig. 11. Smoothed trajectory for the ground truth topology with the rig panoramas corresponding to a few landmarks. This illustrates that many of the landmarks that seem to be false positives at first glance are, in fact, genuine landmarks due to the presence of doors and gateways, even though the trajectory does not indicate this.

	Actual Positive	Actual Negative
Predicted Positive	39	8 (False Positive)
Predicted Negative	3 (False Negative)	6952

	Actual Positive	Actual Negative
Predicted Positive	32	11 (False Positive)
Predicted Negative	6 (False Negative)	1322

Fig. 12. Contingency tables showing quantitive landmark detection results for laser (left) and appearance (right). False negatives (undetected landmarks) are more important than false positives.

low as required, though the number of false positives is large but still reasonable. Further, the results using laser are better due to the very simple but highly informative nature of the measurements used, viz. the area of laser scans.

VII. Conclusion

We have proposed a new landmark detection scheme that equates the presence of a landmark with a sudden change in environment characteristics as quantified by Bayesian surprise. The computation of surprise was illustrated for appearance measurements using a bag-of-words model, and using laser range scans, thus proving the generality of the algorithm. Landmark detection was tested with a topological mapper on a number of datasets, hence demonstrating its practicality.

An important limitation of the current approach is that it maintains the same model for all time and updates it continually as measurements are received. As the KL-divergence between models computed at consecutive steps decays over time due to the increasing number of measurements incorporated into the model, the detector becomes increasingly sensitive to noise in the measurements. However, an alternate method using change point detection is possible that is superior in this respect. In this scenario, we assume that change points in the environment, where the generative model for measurements changes, are landmarks. At each time step, inference is performed to detect if the current location is a change point, and the current model is discarded if this is the case. A systematic analysis of change-point detection requires that all possible locations of change-points be considered. It is future work to compare the performance of the change-point detection scheme to the current technique.

References

[1] Modeling word burstiness using the dirichlet distribution. In *Intl. Conf. on Machine Learning (ICML)*, pages 545–552, 2005.

[2] P. Beeson, N. K. Jong, and B. Kuipers. Towards autonomous topological place detection using the Extended Voronoi Graph. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2005.

[3] M.C. Bosse, P.M. Newman, J.J. Leonard, and S. Teller. Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework. *Intl. J. of Robotics Research*, 23(12):1113–1139, December 2004.

[4] H. Choset and K. Nagatani. Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization. *IEEE Trans. Robot. Automat.*, 17(2):125 – 137, April 2001.

[5] G. Dedeoglu, M. Mataric, and G. Sukhatme. Incremental, online topological map building with a mobile robot. In *Proceedings of Mobile Robots*, 1999.

[6] C. Elkan. Clustering documents with an exponential-family approximation of the dirichlet compound multinomial distribution. In *Intl. Conf. on Machine Learning (ICML)*, pages 289–296, 2006.

[7] L. Itti and P. Baldi. A principled approach to detecting surprising events in video. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 631–637, 2005.

[8] L. Itti and P. Baldi. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1–8, Cambridge, MA, 2006. MIT Press.

[9] M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Fast incremental smoothing and mapping with efficient data association. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1670–1677, Rome, Italy, April 2007.

[10] D. Kortenkamp and T. Weymouth. Topological mapping for mobile robots using a combination of sonar and vision sensing. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 979–984, 1994.

[11] B. Kuipers. The Spatial Semantic Hierarchy. *Artificial Intelligence*, 119:191–233, 2000.

[12] B. Kuipers and P. Beeson. Bootstrap learning for place recognition. In *Proc. 19th AAAI National Conference on AI*, pages 174–180, Edmonton, Alberta, Canada, 2002.

[13] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Intl. J. of Computer Vision*, 60(2):91–110, 2004.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *British Machine Vision Conf. (BMVC)*, pages 414–431, 2002.

[15] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Eur. Conf. on Computer Vision (ECCV)*, volume 1, pages 128–142, 2002.

[16] T.P. Minka. Estimating a dirichlet distribution. 2003.

[17] F.T. Ramos, B. Upcroft, S. Kumar, and H.F. Durrant-Whyte. A bayesian approach for place recognition. In *IJCAI Workshop on Reasoning with Uncertainty in Robotics (RUR-05)*, 2005.

[18] A. Ranganathan and F. Dellaert. A Rao-Blackwellized particle filter for topological mapping. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 810–817, 2006.

[19] A. Ranganathan and F. Dellaert. Semantic Modeling of Places using Objects. 2007.

[20] A. Ranganathan, E. Menegatti, and F. Dellaert. Bayesian inference in the space of topological maps. *IEEE Trans. Robotics*, 22(1):92–107, 2006.

[21] D. Schroter, T. Weber, M. Beetz, and B. Radig. Detection and classification of gateways for the acquisition of structured robot maps. In *Proceedings of 26th Pattern Recognition Symposium (DAGM)*, 2004.

[22] H. Shatkay and L. Kaelbling. Learning topological maps with weak local odometric information. In *Proceedings of IJCAI-97*, pages 920–929, 1997.

[23] Josef Sivic, Bryan Russell, Alexei A. Efros, Andrew Zisserman, and Bill Freeman. Discovering objects and their location in images. In *Intl. Conf. on Computer Vision (ICCV)*, 2005.