Self Help: Seeking Out Perplexing Images for Ever Improving Navigation

Rohan Paul and Paul Newman
Oxford University Mobile Robotics Research Group. {rohanp,pnewman}@robots.ox.ac.uk

***Abstract*— This paper is a demonstration of how a robot can, through introspection and then targeted data retrieval, improve its own performance. It is a step in the direction of lifelong learning and adaptation and is motivated by the desire to build robots that have plastic competencies which are not baked in. They should react to and benefit from use. We consider a particular instantiation of this problem in the context of place recognition. Based on a topic based probabilistic model of images, we use a measure of perplexity to evaluate how well a working set of background images explain the robot's online view of the world. Offline, the robot then searches an external resource to seek out additional background images that bolster its ability to localise in its environment when used next. In this way the robot adapts and improves performance through use.**

I. Introduction

This paper is about having a robot actively seek data to improve its understanding of the world. The big picture motivation of the work is to enable robot longevity and in this paper we consider a specific instantiation of this problem - that of asymptotically improving scene recognition with a camera. We shall make use of the FAB-MAP algorithm (Cummins et al. [4], [5]) which probabilistically associates a current view of the world (image) taken by a robot with a previously visited or a new place. FAB-MAP requires priming with a set of images (which we shall refer to as a sample set) which in concert, statistically represents the appearance of the robot's workspace. For operation in urban settings, for example, one equips it with a sample set containing random images of cities and towns. There is an obvious shortcoming here - the robot is constrained to work in settings in which its sample set has sufficient explanatory power. If moved into surroundings quite different from those represented by its sampling set performance drops - nothing is as expected and everything is astounding. In this paper, we show how by producing a generative model of the underlying topics present in observed images we can actively grow a customised sample set by incorporating well chosen examples from an external corpus which is more representative of the workspace the vehicle is experiencing. In this way, we replace the inflexibility of a static *a-priori* sample set with a plastic, dynamic one and we show this affords an improvement in performance over time. One could think of this as a robot actively seeking to widen its experience, better understand its surroundings and becoming less perplexed with time.

Our problem setup is as follows. A mobile robot must maintain a compact on-board sample set summarizing the visual appearance of its environment. The robot explores the environment collecting image data and identifies the most perplexing images based on its current sample set. It then searches the least explained images in a large repository of image data (or past datasets from the robot) finding images with similar thematic content. Next, the robot retrieves examples (based on their likelihood in the environment) and assimilates them into the sample set, thereby improving its representation and performance. The rest of the paper details this framework and presents the following components:

- Use of Latent Dirichlet Allocation (LDA) topic model to extract a low-dimensional thematic representation for images incorporating word co-occurrence statistics in an unsupervised manner, Section III.
- Identifying most novel images seen by the robot given its current representation using a perplexity-based measure, Section IV.
- Finding images similar in thematic content from an external repository applying language-model based information retrieval approach, Section III.
- Application to FAB-MAP, presenting an algorithm for constructing a representative sampling set, Section V.
- Experimentation on real datasets collected by a mobile robot, Section VI.

II. Related work

This paper builds on research in the areas of long-term topological mapping, topic-based document modeling and information retrieval.

In [13], Milford and Wyeth present a biologically inspired system for persistent mapping that can learn long-term changes in workspace appearance. For topological mapping, Konolige et al. [11] present *view based maps* based on geometric feature matching in stereo views. Angeli et al. [1] discuss an incremental loop-closure detection scheme with epipolar geometry checks.

Topic models based on Latent Dirichlet Allocation (LDA) were developed by Blei et al. [3] in the context of statistical text analysis. In computer vision, Sivic et al. [15] employed LDA for discovering object categories in image corpora and present a hierarchical model for unsupervised discovery of object class hierarchies in [16]. Fei Fei et al. [7] present an application for learning natural scene categories. Recently, Philbin et al. [14] introduced geometric LDA, where trasformations from the latent spatial model are additionally estimated.

Within robotics, Endres et al. [6] applied LDA for unsupervised discovery of object classes in 3D laser range data. In the information retrieval community, Wei and Croft [19] used LDA-based topic models for ad-hoc document retrieval

extending a previous work by [12]. Hörster et al. [10] applied this technique for image retrieval tasks. In another related work Zhang et al. [21] discuss novelty detection for adaptive filtering applications.

III. Probabilistic Topic Models

Probabilistic topic models represent documents as a mixture of intermediate latent topics. Given a collection of documents such as scientific abstracts, each represented as a bag-of-words vector, the model is able to learn common topics such as *ecology*, *astronomy* etc. in an unsupervised manner [8]. Topics are distributions over words and each document is a distribution over topics. Different documents can have varied mixing proportions of each topic. By mapping documents to a low-dimensional thematic representation, the model can semantically associate ones with similar topics, even though the documents themselves might have few words in common. Using the approach by Sivic et al. [17], images can be represented as a vector of visual words. Hence, we use the terms documents and images interchangeably.

Latent Dirichlet Allocation (LDA) is a widely used probabilistic topic model [3]. LDA is a hierarchical bayesian generative model for a collection of discrete data possessing tractable inference to estimate topics and topic proportions. The following sub-sections review the LDA generative model, inference and its application for retrieving images similar in thematic content. For detailed description on LDA please refer to [3], [8] and [9].

A. LDA Generative Model

A document d from a corpus of D documents consists of a set of words $(w_1, w_2, \ldots, w_{N_d})$, where w_i is a single word occurrence from a vocabulary of size W. The model postulates T topics, each characterized by a distribution over words $P(w|z)$. The generative process (Figure 1) for each word begins by first sampling a topic label z_i from the multinomial distribution over T topics for the given document $P(z|d)$ followed by sampling a word w_i from the distribution over words $P(w|z)$ for the sampled topic label. Hence, the likelihood of a word in a document can be obtained via marginalization over intermediate topics.

$$P(w_i|d) \quad = \quad \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j|d) \qquad (1)$$

The topic-proportion $P(z|d)$ for each document d and word likelihoods $P(w|z = j)$ for each topic j are abbreviated as $\theta^{(d)}$ and $\phi^{(j)}$ respectively [8]. Symmetric Dirichlet priors are placed on $\theta^{(d)}$ and $\phi^{(j)}$, with $\theta^{(d)} \sim Dirichlet(\alpha)$ and $\phi^{(j)} \sim Dirichlet(\beta)$, where hyper-parameters α and β control the sparsity of these distributions.

B. Inferring Topics and Topic Proportions

Estimating the set of topics and topic-proportions from observed word tokens requires reversing the generative process. For each obsered word w let z be the topic indicator variable. The goal is to estimate the topic distributions that best describe the data by evaluating the posterior distribution $P(\mathbf{z}|\mathbf{w}, \alpha, \beta) \propto P(\mathbf{w}|\mathbf{z}, \beta)P(\mathbf{z}|, \alpha)$. Exact inference is intractable and approximated via Markov Chain Monte Carlo (MCMC) using collapsed Gibbs sampling in the state space of topic labels (Griffiths et al. [8]). The Markov chain is initialized by a random assignment of topic labels $\mathbf{z}$. Subsequent states are reached by sequentially sampling each variable z_i from a distribution conditioned on observed words and current assignment of all other topic labels. The desired conditional distribution is expressed as:

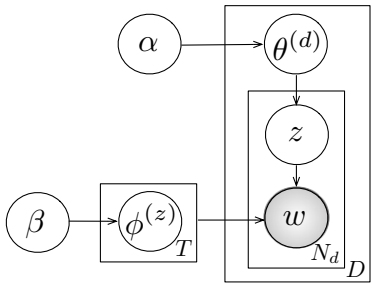

Fig. 1: LDA Generative Model. Topics, $\phi^{(z)}$ are multinomial distributions over vocabulary words (with dirichlet prior β). The generative process for a document, d begins by sampling a distribution over topics, $\theta^{(d)}$ (with Dirichlet prior α). Document words are generated by first drawing a topic label z from topic-proportion $\theta^{(d)}$ and then sampling a word w from $\phi^{(z)}$.

$$P(z_i = j|\mathbf{z}_{-i}, \mathbf{w}) \quad \propto \quad \left[\frac{n_{-i,j}^{(w_i)} + \beta}{n_{-i,j}^{(\cdot)} + W\beta}\right]\left[\frac{n_{j}^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + T\alpha}\right] \qquad (2)$$

Here $\mathbf{z}_{-i}$ refers to the current topic assignments of all other word tokens and $n_{-i}^{(\cdot)}$ is the count excluding the current assignment z_i. Equation 2 expresses the conditional distribution for topic label z_i assigned to word w_i as a product of the likelihood of word w_i under topic j and the probability of topic j in document d_i. Upon convergence after sufficient iterations topic labels are recorded and the maximum likelihood multinomial estimates for topics and topic proportions are obtained as:

$$\hat{\phi}_j^{(w)} \quad = \quad \frac{n_j^{(w)} + \beta}{n_j^{(\cdot)} + W\beta} \qquad (3)$$

$$\hat{\theta}_j^{(d)} \quad = \quad \frac{n_j^{(d)} + \alpha}{n_{\cdot}^{(d)} + T\alpha} \qquad (4)$$

Topics are typically estimated once for a large corpus. For a new document, topic proportions $\theta^{(d)}$ can be inferred using the learned topic distributions $\hat{\phi}^{(w)}$ via Gibbs sampling using the following conditional distribution:

$$P(z_i = j|\mathbf{z}_{-i}, \mathbf{w}) \quad \propto \quad \left(\hat{\phi}_j^{(w)}\right)\left(\frac{n_j^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + T\alpha}\right) \qquad (5)$$

C. Retrieving Similar Images

Topic models provide a low-dimensional representation of bag-of-words data capturing their thematic content via word co-occurrences. We employ this representation to find images similar to a query image from a repository. Please note that our application does not require precise geometric image matches. Instead, we seek images similar in semantic content.

Given a large repository of images we learn topics ϕ and topic proportions θ^d for images in the corpus, thereby forming a topic-based model $P(w|d, \theta^d, \phi)$ for each document. We define query-document similarity using the LDA generative model. A document is similar to a query if the document model has a high predictive likelihood of generating the query. This formulation is also termed as language-model based retrieval [19]. Formally, for a given query image d_a with N_a words, the likelihood of originating from document d_b is expressed as:

$$P(d_a|d_b, \theta^b, \phi) \quad = \quad \prod_{j=1}^{N_a} P(w_j^a|d_b, \theta^b, \phi) \tag{6}$$

where w_j^a represents the j^{th} word in d_a. Word likelihood $P(w_j^a|d_b, \theta^b, \phi)$ is evaluated using Equation 1. Further, as discussed in [20], topic model based estimates must be smoothed for retrieval with a unigram model and Dirichlet smoothing ([12], [10]). Additionally, in [10], authors provide quantitative results to show that an LDA based approach performs better than simpler measures like cosine distance or Jenson-Shannon divergence on image retrieval tasks.

IV. Is an Image Perplexing?

We now address the task of determining the novelty (or redundancy) of an observed image given a corpus. As discussed in the previous section, the LDA generative model allows us to calculate the document likelihood $P(d_a|d_b, \theta^b, \phi)$ via a topic model representation (Equation 6). Note that the computed likelihood is dependent on the query length. We seek a length normalized measure and hence estimate perplexity [9] of the observed image given a document model from the corpus as:

$$Perplexity(d_a|d_b, \theta^b, \phi) \quad = \quad exp\left\{\frac{-logP(d_a|d_b,\theta^b,\phi)}{N_a}\right\} \tag{7}$$

Perplexity indicates the uncertainty in predicting a single word. Chance performance results in the maximum possible value of perplexity which equals the vocabulary size. A model that better captures word co-occurrences requires fewer possibilities to pick words, yielding lower perplexity per word for new data.

The next step is determining the redundancy of a new unseen document in relation to a corpus. A new document is redundant, if its information content is covered by documents present in the corpus. Similarly, documents highly dissimilar to the ones seen previously contain new information and hence considered novel. We adapt the framework proposed by Zhang et al. [21] in the context of adaptive information filtering and define redundancy of document d_i vis-à-vis a document d_j as:

$$R(d_i|d_j) \quad = \quad -Perplexity(d_i|d_j, \theta^j, \phi) \tag{8}$$

where θ^j is the topic proportion estimated for d_j. The redundancy of an observation given a document corpus, D is expressed as:

$$R(d_i|D) \quad = \quad max_{d_j \epsilon D} R(d_i|d_j) \tag{9}$$

Finally, given an observed image set Q, the most novel image $d_{Novel(Q)}$ pertaining to an existing corpus D is selected as:

$$d_{Novel(Q)} \quad = \quad argmin_{d_j \epsilon Q} R(d_i|D) \tag{10}$$

Intuitively, the novel images thus identified are the ones least explained by images in the corpus. Hence, we seek images similar to the perplexing ones and augment those to the corpus forming an improved representation. Next, we apply this technique to improving appearance based navigation.

V. Application to Topological Mapping

We now apply our framework to an appearance based mapping algorithm FAB-MAP (Cummins et al. [4], [5]). FAB-MAP is a loop closure detection system that allows a navigating robot to determine if the current observation comes from a known location in its map or from a new one. Our goal will be to show that over time we can improve its performace with experience. Next, we present a brief summary of FAB-MAP followed by a procedure to build a sampling set over time to improve performance.

A. FAB-MAP Overview

At time t, the robot's workspace has n_t locations $\mathcal{L}^t = \{L_1, \ldots, L_{n_t}\}$ where each location L_i has an associated appearance model represented by a distribution over appearance words. When the robot collects a new observation Z_t, we compute the distribution over locations given the observation $p(L_i|\mathcal{Z}^t)$ formulated as a recursive Bayes estimation problem:

$$p(L_i|\mathcal{Z}^t) = \frac{p(Z_t|L_i, \mathcal{Z}^{t-1})p(L_i|\mathcal{Z}^{t-1})}{p(Z_t|\mathcal{Z}^{t-1})} \tag{11}$$

where $\mathcal{Z}^t$ is the set of all observations up to time t, $p(Z_t|L_i, \mathcal{Z}^{t-1})$ is the observation likelihood of the observation given the location L_i. The normalization term $p(Z_t|\mathcal{Z}^{t-1})$ is the total likelihood of the observation, Z_t. An observation can come from the set of locations currently in the robot's map (M) as well as the set of all previously unknown locations ($\overline{M}$). Hence, the denominator is expressed as:

$$\begin{aligned} p(Z_t|\mathcal{Z}^{t-1}) \quad &= \quad \sum_{m \in M} p(Z_t|L_m)p(L_m|\mathcal{Z}^{t-1}) \\ &+ \quad \sum_{u \in \overline{M}} p(Z_t|L_u)p(L_u|\mathcal{Z}^{t-1}) \end{aligned} \tag{12}$$

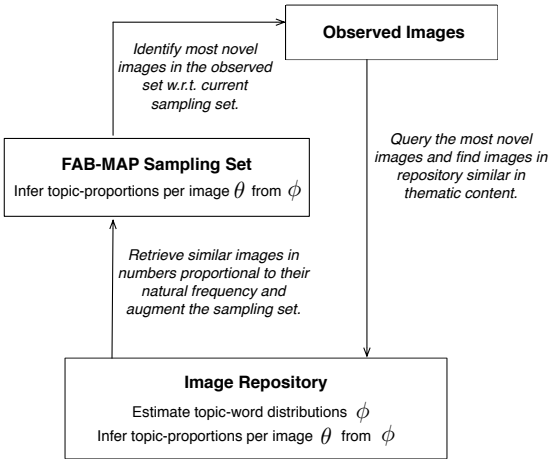

Fig. 2: Sampling set evolution. Topics ϕ are estimated from the image repository and used to infer topic proportions θ for sampling set and repository images. Online, the robot collects observations from which most novel images are identified. These perplexing images are searched in the repository for thematically similar images which are retrieved (respecting the underlying distribution) and augmented to the evolving sampling set.

Algorithm 1 FAB-MAP Sampling Set Evolution

```
// Sampling Set, SS
// Database, DB
// Observed Image Set, ObS
// Topics, φ
// Topic-proportions, θ
SS[0] ← Random image from DB.
for each improvement cycle, t
  // Operate
  ObS[t] ← Collect data from robot.
  Use FAB-MAP on ObS[t] with SS[t].
  // Introspect
  for each image d_i ∈ ObS[t]
     //Estimate redundancy
     R(d_i|SS[t]) eq:9
  end
  Get novel image d_Novel ∈ ObS[t] eq:10
  // Retrieve
  for each image d_j ∈ DB
     //Estimate redundancy
     Mult_dist[j] ← p(d_Novel|d_j, θ^j, φ) eq:6,7
  end
  Ret_samples ← Sample from Mult_dist
// Improve sample set.
SS[t+1] ← SS[t] ⋃ Unique(Ret_samples)
end
```

The second term involves summation over all unmapped places and is approximated via sampling location models L_u [4]. This is instead approximated by sampling observations Z and using them to form location models. Observation likelihood $p(Z_t|L_u)$ is evaluated for each sample and Equation 12 is expressed as:

$$p(Z_t|\mathcal{Z}^{t-1}) \approx \sum_{m\in M} p(Z_t|L_m)p(L_m|\mathcal{Z}^{t-1}) + \frac{p(L_{new}|\mathcal{Z}^{t-1})}{n_s}\sum_{u=1}^{n_s} p(Z_t|L_u) \quad (13)$$

Here, n_s is the number of samples used and $p(L_{new}|\mathcal{Z}^{t-1})$ is the prior probability of being at a new place, uniformly distributed among samples. A reasonable number of samples is 2,800. This yields the total probability of the observation originating from a place not in the map. The resulting PDF over locations is used to decide whether to add a new location to the map or not.

B. The Sampling Set

The sampling set forms the robot's compact representation of the workspace visual appearance. The sampling set is critical for performance since it ameliorates the perceptual aliasing problem: the fact that different parts of the environment appear the same to robot's sensors. e.g., similar looking foliage and brick walls appear commonly while navigating outdoors. The sampling set captures such commonly seen visual features and hence the normalization step distributes the probability mass preventing a false loop closure declaration. In [4], the sampling set was constructed by randomly picking 2,800 images from data sets obtained from previous runs of the robot. Specific examples of common features (potential perceptual aliasing cases) were explicitly added through inspection to make the sampling set representative of the application environment.

Recently, FAB-MAP has been successfully demonstrated to create maps exceeding 1000km [5]. Consequently, the map generated (visual experience of the robot) can far exceed the size of the sampling set (onboard workspace representation). Hence, for long term operation, there is a need to form a compact and representative sampling set that improves over time with the robot's experience.

C. Sampling Set Evolution

We now apply techniques introduced in previous sections to identify gaps in the sampling set representation and improve performance through introspection.

The procedure begins by estimating topics ϕ on a large repository of images (Figure 2). The sampling set is initialized with a single image picked randomly from the database. Topic proportions θ are inferred for all images in the database and the sampling set. From the images collected online by the navigating robot, most novel images given the current sampling set are identified using Equations 9 and 10.

The most perplexing images are queried in the large image repository and images similar to the queried images are retrieved and augmented to the evolving sampling set. For each selected novel image, topic model based similarity to each database image is estimated, yielding a multinomial distribution over database images. Similar images in the repository must be retrieved respecting their natural occurrence frequency in the environment (as represented in the database). For example, images containing foliage features

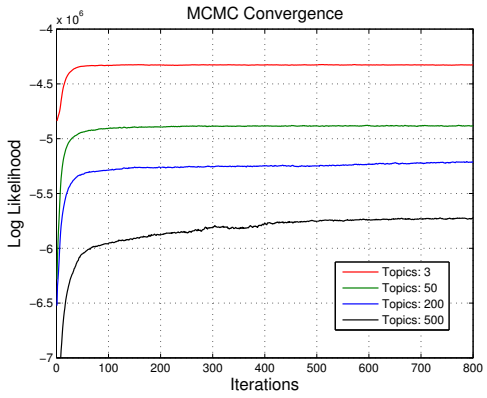

Fig. 3: Iterations for MCMC convergence. Data loglikelihood plotted after each MCMC iteration for varying number of topics. For topics 3, 50 and 200 the chain converges within 200 iterations. Convergence for 500 topics is much slower and stabilizes after 600 iterations. Hence, the number of Gibbs sampling iterations was set to 600.

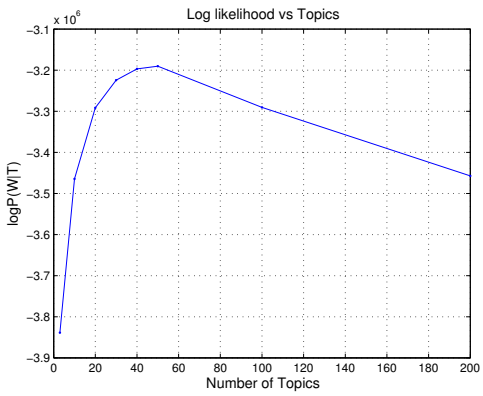

Fig. 4: Selecting the number of topics, T for the urban dataset. The plot shows the data loglikelihood (upon convergence) for varying number of topics. The data loglikelihood peaks for 50 topics. Hence, we set $T = 50$ (assuming a uniform prior on topics).

are very common compared to rare images of sign boards. Hence, we retrieve a greater number of images when the likelihood distribution is more uniform. This is accomplished by sampling the likelihood multinomial a fixed number of times and selecting only the unique samples, thereby accepting fewer samples when the distribution is peaked. The retrieved samples are then augmented forming new sampling set. The steps are summarized in Algorithm 1. Hence, we create a compact sampling set targeted to the operating environment, eliminating the need for a very large sampling set including all past datasets.

Please note that in the present formulation, relevant images are only added to the sampling set. Over time, it might be desirable to remove images no longer relevant and restrict the sampling set size. One possible solution is to remove the most perplexing images in the sampling set given the currently observed set of images. However, we wish to formally explore replacement strategies as part of future work.

VI. Results

We tested the system on data collected from a mobile robot. A collection of 2,800 images from 28km of urban streets and parks using the robot's camera formed the database. Images were captured perpendicular to robot's motion and did not overlap. Images were converted to a bag-of-words representation [17] by first extracting SURF features [2] and later quantizing them against a fixed vocabulary of size $11K$.

Topic models were estimated on this dataset using Gibbs sampling (Section III). In order to determine the number of iterations required to ensure MCMC convergence to the target distribution we used data loglikelihood as a measure [8]. Figure 3 plots loglikelihood with each iteration for varying number of topics T. Dirichlet priors were set to $\alpha = 50/T$ and $\beta = 0.1$. For the runs with topics 3, 50 and 200 the loglikelihood stabilizes within 200 iterations. However, convergence for the run with 500 topics is much slower and stabilizes after 600 iterations. We also experimented with observing multiple runs of the Markov chain and convergence rates were similar. Hence, the number of sampling iterations was set to 600.

(a) (b)

Fig. 5: Illustrative visual topics discovered on the urban dataset. Five most probable visual words for each topic appear along rows and common instances along columns. In (a) a topic capturing visual words co-occurring on foliage and trees is shown. In (b) a topic representing words appearing on walls and houses is illustrated.

The LDA model require the number of topics to be specified. This can be cast as a model selection problem. The appropriate number of topics modeling the dataset was determined by maximising the data likelihood given topics $P(\mathbf{w}|T)$ (assuming a uniform prior on the number of topics). As suggested in [8], $P(\mathbf{w}|T)$ can be approximated using $P(\mathbf{w}|\mathbf{z}, T)$ where topic-labels $\mathbf{z}$ are obtained from the Gibbs sampler upon convergence. Figure 4 plots the result. The data loglikelihood (upon convergence) for the dataset peaks with 50 topics.

Figure 5 shows two illustrative visual topics discovered on the urban dataset. The most probable visual words oc-

(a)

(b)

Fig. 6: Illustrative results for topic-model based image retrieval. Query image is shown top left and five most similar images from the database are shown subsequently. Note that the retrieved images represent a common visual theme.

(a) Typical images from the New College Dataset used for topic learning.

$R = -1.17 \times 10^3$ $R = -1.32 \times 10^3$ $R = -1.18 \times 10^3$

(b) Least novel (most redundant) images in urban dataset.

$R = -7.86 \times 10^3$ $R = -6.54 \times 10^3$ $R = -1.05 \times 10^4$

(c) Most novel (least redundant) images in the urban dataset.

Fig. 7: Detecting novel images. Topics were learnt on the New College Dataset with typical images of medieval buildings and parklands, shown in (a). For images in the urban dataset, redundancy values were computed and listed below each image. (b) Brick wall and foliage images were least novel due to high similarity to visual themes in the New College Dataset. (c) Images of road signs, cars and modern buildings were found most novel.

cur along rows and typical occurrences are shown along columns. Co-occurring words frequently get assigned to a particular topic. Figure 5a shows a topic for words commonly occurring on foliage and trees and Figure 5b illustrates a topic modeling words appearing on walls and houses.

Figure 6 shows examples of topic-model based image retrieval. The query image is highlighted and the most similar images in the database are shown. Note that the system returns images that are thematically similar as opposed to strict geometric matches. This is key to our approach. While querying to find images similar to a perplexing image (say an image of a building) we are not looking for exact instances of the same building. Rather, we wish to retrieve examples of the class of similar looking buildings which better represents a common mode of visual appearance and is accomplished via a low-dimensional topic representation. By mapping visual features on buildings to a topic that probabilistically models co-occurring words on buildings (e.g. Figure 5b), other relevant images containing similar visual features are retrieved.

To test our approach for detecting most novel images, we slightly modified the setup. In this experiment, we learnt topic models on the New College Dataset [18], where visual appearance is restricted to medieval buildings and parkland areas. Typical images are shown in Figure 7a. A restricted set of 564 images were used from the dataset (removing loop closure pairs) and the number of topics was set to 50. We then estimated the redundancy values for images in the urban dataset introduced previously. Images of brick walls and foliage were found to be most redundant due to high similarity to visual themes in the New College Dataset (Figure 7(b)). Images of sign boards, road vehicles and modern, regular shaped buildings were found to be most novel (least redundant). Since, no such examples are present in the New College Dataset, the learnt topic model is highly perplexed to encounter these feature sets.

Finally, we bring the above components together and test performance on the FAB-MAP algorithm. We use the urban dataset as the database to construct the sampling set, initialized with a single randomly-selected image from the database. We used the City Centre dataset as the observed dataset [4]. The City Centre dataset is 2km in length, possessing 2474 images and provides a challenging setup for image matching due to considerable scene change in images. This dataset does not overlap with the urban dataset used for learning topics.

To simulate days of operation (improvement cycles) the dataset was split into 10 sections and presented sequentially to the sampling set construction algorithm. In each iteration we identify 25 most novel images and sample the similarity multinomial distribution of database images 50 times and pick the unique image samples to be retrieved (parameters set empirically). After each iteration, the sampling set (incrementally developed) was used for loop closure detection

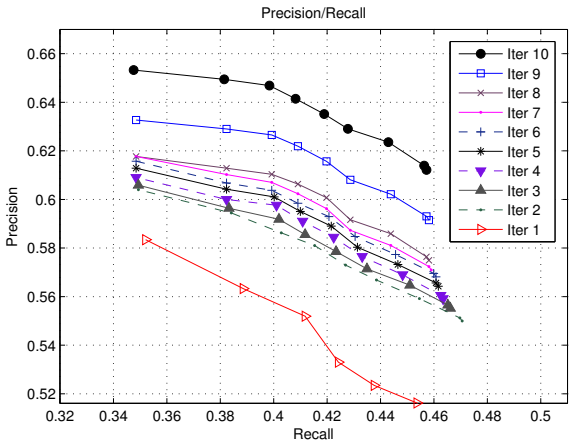

Fig. 8: Asymptotic increase in FAB-MAP loop closure detection performance. Precision-recall curve is plotted after each improvement cycle. The curve moves upwards with each iteration (indicating improved performance) as a better sampling set is evolved. Maximum precision increases from 58.3% to 65.3%.

using FAB-MAP on the City Centre dataset. The visual detector parameters [4] were set to $p(z_i = 1|e_i = 0) = 0$ and $p(z_i = 0|e_i = 1) = 0.39$. Ground truth was generated via visual inspection. The prior probability over locations in the map was left uniform to test the core inference component independent of a motion model. Precision-recall was calculated by varying the probability threshold at which loop closures are accepted. Figure 8 plots the precision-recall curves after each iteration. The curve moves upwards after each improvement cycle indicating an improved performance as a better sampling set is evolved by the algorithm. Maximum precision increases from 58.3% to 65.3%.

VII. Conclusions

We have shown how a robot can, through introspection and then targeted data retrieval, improve its own navigation performance. It is a step in the direction of lifelong learning and adaption and was motivated by the desire to build robots that have plastic competencies which are not baked in. They should react to and benefit from use. We have considered a particular instantiation of this problem in the context of place recognition using the FAB-MAP algorithm. We used LDA based topic models to enable the calculation of a measure of image perplexity viz-a-viz a sample set which is supposed to be representative of all scenes experienced by the robot. This measure guides a retrieval of additional images that share topics with the confusing ones. The sample set is thus extended to better explain the images that the robot is seeing at run time - in a sense, the robot is customising itself to its surroundings. Although we have demonstrated adaption in the context of a particular problem this is part of a bigger picture - one in which lifelong learning is achieved by robots actively and continually seeking out training data or experience as a result of on going use.

VIII. Acknowledgements

We are grateful to Mark Cummins and James Philbin for helpful discussions and insights. We thank anonymous reviewers for valuable suggestions. This research was supported by the Rhodes Trust, Oxford, UK.

References

[1] A. Angeli, D. Filliat, S. Doncieux, and J.A. Meyer. A Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words. *IEEE Transactions On Robotics, Special Issue on Visual SLAM*, 24(5):1027–1037, 2008.

[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. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008.

[5] 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.

[6] F Endres, C Plagemann, and C Stachniss. Unsupervised discovery of object classes from range data using latent dirichlet allocation. *Proceedings of Robotics: Science and Systems*, 2009.

[7] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition, CVPR 2005*, volume 2, pages 524–531, 2005.

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

[9] G Heinrich. Parameter estimation for text analysis. *Technical Report. Web: http://www. arbylon. net/publications/text-est. pdf*, Jan 2005.

[10] E Hörster, R Lienhart, and M Slaney. Image retrieval on large-scale image databases. *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, page 24, Jan 2007.

[11] K. Konolige, J. Bowman, J.D. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua. View-based maps. In *Proceedings of Robotics: Science and Systems (RSS)*, 2009.

[12] X Liu and W Croft. Cluster-based retrieval using language models. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 193, Jan 2004.

[13] M Milford and G Wyeth. Persistent navigation and mapping using a biologically inspired slam system. *The International Journal of Robotics Research*, Jan 2009.

[14] J Philbin, J Sivic, and A Zisserman. Geometric lda: A generative model for particular object discovery. *Proceedings of the British Machine Vision Conference*, Jan 2008.

[15] J Sivic, B Russell, A Efros, and A Zisserman. Discovering object categories in image collections. *Proc. ICCV*, Jan 2005.

[16] J Sivic, B Russell, A Zisserman, and WT Freeman. Unsupervised discovery of visual object class hierarchies. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Jan 2008.

[17] 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.

[18] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman. The new college vision and laser data set. *International Journal for Robotics Research (IJRR)*, 28(5):595–599, May 2009.

[19] X Wei and W Croft. Lda-based document models for ad-hoc retrieval. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, page 185, Jan 2006.

[20] C Zhai and J Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems (TOIS)*, 22(2):214, Jan 2004.

[21] Y Zhang, J Callan, and T Minka. Novelty and redundancy detection in adaptive filtering. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 81—88, Jan 2002.