2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Towards Training-Free Appearance-Based Localization: Probabilistic Models for Whole-Image Descriptors

Stephanie M. Lowry, Gordon F. Wyeth, *Member, IEEE*, and Michael J. Milford, *Member, IEEE*

***Abstract*—Whole image descriptors have been shown to be remarkably robust to perceptual change especially compared to local features. However, whole-image-based localization systems typically rely on heuristic methods for determining appropriate matching thresholds in a particular environment. These environment-specific tuning requirements and the lack of a meaningful interpretation of arbitrary thresholds limit the general applicability of these systems. In this paper we present a Bayesian model of probability for whole-image descriptors that can be seamlessly integrated into localization systems designed for probabilistic visual input. We demonstrate this method using CAT-Graph, an appearance-based visual localization system originally designed for a FAB-MAP-style probabilistic input. We show that using whole-image descriptors as visual input extends CAT-Graph's functionality to environments that experience a greater amount of perceptual change. We also present a method of estimating whole-image probability models in an online manner, removing the need for a prior training phase. We show that this online, automated training method can perform comparably to pre-trained, manually tuned local descriptor methods.**

I. Introduction

Vision-based techniques can be used to perform robot localization over paths thousands of kilometers long [1] and for navigation lasting weeks at a time [2]. However, visual place recognition often fails under perceptual changes such as those caused by lighting, weather or seasonal changes [3]. This is particularly significant for outdoor robot operation, where natural lighting variation and weather phenomena can cause drastic visual changes, often over very short time-scales.

Many visual localization systems [4-6] utilize local features such as SIFT [7] and SURF [8]. In FAB-MAP [4], a state of the art probabilistic visual mapping system, local features are combined with a visual bag-of-words [9, 10] and the robot location can be calculated based on the joint probability of the image words. However, as local features are susceptible to perceptual change [3], they cannot be depended on for robust localization in many outdoor environments.

Description methods that represent the whole image rather than extracting local features are more suited to perceptually changing environments [11-14]. Furthermore, whole-image descriptors can be compared to produce single-valued difference values between images, which allow simple statistics to be calculated [13]. However, many whole-image localization systems are not probabilistic – place recognition is determined using heuristic methods rather than based on probabilistic models.

In this paper we present a method for estimating probability models for whole-image descriptors in an unknown environment. We show that this probability model performs comparably to FAB-MAP, despite the lack of a training phase. We also compare the probabilistic model to a heuristic model, and show that an incremental probabilistic model can localize effectively over environments that experience significant perceptual change without the same requirements for re-tuning of place recognition thresholds. The work presented here extends preliminary work presented in [15] to perceptually varying environments, and provides further analyses of performance including comparison between probabilistic and non-probabilistic methods.

We also present a method for integrating whole-image descriptors into localization systems designed for probabilistic input. We demonstrate the effectiveness of the whole-image probability model using CAT-Graph [5, 16], an appearance-based localization system originally designed for a FAB-MAP-like [4] probabilistic visual input. We show that using whole-image descriptors rather than local features allows CAT-Graph to operate successfully on environments that exhibit perceptual change. The paper proceeds as follows. Section II reviews whole-image localization and the role of probability in appearance-based localization. Section III describes the theory behind probabilistic models for whole-image descriptors, introduces whole-image CAT-Graph, and outlines the method for creating online probability models for whole-image descriptors. Section IV describes the experimental setup. Section V presents results showing probabilistic localization using whole-image CAT-Graph, and using training-free probability models based on whole-image descriptors. The paper concludes in Section VI with discussion and future work.

II. Background

In this section we briefly review the two key themes of this paper – the use of whole-image descriptors for localization and the application of probabilistic models to whole-image appearance-based localization.

Many whole-image descriptors can be used for appearance-based localization. A popular choice is GIST [17, 18] as used in [19-21], and local feature descriptors such as SURF [8] and BRIEF [22] have been converted into whole-image descriptors for the purpose of localization [13, 23]. The whole-image localization systems RatSLAM and SeqSLAM [11, 24] use the images themselves as descriptors and are compared via the sum of absolute pixel differences (SAD).

S.M. Lowry, G.F. Wyeth and M.J. Milford are with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia. stephanie.lowry@student.qut.edu.au. This work was in part funded by the Australian Research Council through Discovery project DP1113006. S.M. Lowry is supported by an Australian Postgraduate Award and a QUT Vice-Chancellor's Scholarship.

Performance of whole-image localization systems based on GIST [19, 21] and BRIEF-Gist [23] have been shown to perform comparably to established state-of-the-art local feature systems such as FAB-MAP [4]. Whole-image localization systems also perform well across perceptually different environmental conditions such as day and night [11], and seasonal changes [13, 14]. To increase robustness over perceptual change, the whole-image visual input is generally combined with other techniques such as range sensing [13] or sequence filtering [11].

Whole-image localization systems may use heuristic measures to perform place recognition [24] or may approximate the probability via an appropriate function [19, 20]. In [13], a training phase is undertaken to model the probability based on environmental data similar to that used for testing.

In this paper we use an approach, similar to [13], of modeling the probability using environmental data. Furthermore, as whole-image descriptors are single-valued, the probability distribution is simple enough that it can be estimated during localization, rather than requiring a manual pre-training phase. The approximated probability distributions are used to perform localization whilst they are simultaneously being generated. By generating the probability distribution using the robot's current observations, the models will be as relevant as possible to the environment and removes the need for manual training when either the environment or other factors (such as the image descriptor method) are changed.

III. WHOLE-IMAGE PROBABILISTIC LOCALIZATION

In this section we introduce the theoretic and algorithmic basis of probabilistic whole-image localization. In Parts A and B we briefly summarize the relevant theory of appearance-based probabilistic localization. In Part C we describe the extension of CAT-Graph to whole-image visual input, and in Part D present our algorithm for computing a probability model for appearance-based localization online.

A. Probabilistic Appearance-Based Localization

Probabilistic localization, whether using local features or whole-image methods, can be presented as a recursive Bayes filter [4, 19]. If a robot makes observation Z_k at time k, and Z^k is the sequence of observations made at each time step thus far, then the probability that the current location L_k is the same some previously seen location L_i, is given by:

$$P(L_i = L_k \mid Z^k) = \frac{P(Z_k \mid L_i = L_k) P(L_i = L_k \mid Z^{k-1})}{P(Z_k \mid Z^{k-1})} \quad (1)$$

For a bag-of-words localization system such as FAB-MAP [4], $P(Z_k \mid L_i = L_k)$ is calculated from the joint distribution of the vocabulary words. In a whole-image descriptor system, the probability depends not on individual words, but on the distribution of the differences between the images function $d(Z_k, Z_i)$ used to compare the descriptors, which we will represent by the shorthand, d_{ki}. The choice of difference function depends on the whole-image descriptor in question – real-valued descriptors such as GIST or WI-SURF can be compared using sum of squared differences, whereas the Hamming distance is a natural comparison measure for a binary descriptor like BRIEF-Gist.

Converting (1) to an appropriate format for whole-image descriptors produces:

$$P(L_i = L_k \mid d_{ki}, Z^{k-1}) = \frac{P(d_{ki} \mid L_i = L_k) P(L_i = L_k \mid Z^{k-1})}{P(d_{ki} \mid Z^{k-1})} \quad (2)$$

The key difference between these two representations is that for (2) the denominator depends on i as well as k. This change is significant as it means the denominator cannot be treated as a normalizing constant across all i for a given k.

In the following section we describe how to calculate the probability distributions $P(d_{ki} \mid L_i = L_k)$ and $P(d_{ki} \mid Z^{k-1})$, which we refer to in the remainder of the paper as $P(d \mid L)$ and $P(d)$. We also use $P(d \mid \neg L)$ to denote the distribution $P(d_{ki} \mid L_i \neq L_k)$. The term $P(L_i = L_k \mid Z^{k-1})$ represents the prior belief about location [4, 19]. In our implementation this term is inherently calculated by CAT-Graph and is therefore not discussed further in this section.

B. Whole-Image Probability Distributions

The probability distribution for $P(d)$ is calculated via a frequency-based method. Over a sample environment containing N locations, we can calculate the set of all difference values

$$\{d_{ki} \mid k \leq N, i < k\} \quad (3)$$

There are $N(N-1)/2$ samples in such a set. The top plot in Figure 1 displays an example histogram for a set of image difference values, on an environment first presented in [25].

To calculate $P(d)$, we discretize the range of d, splitting it into s bins. Then, for any $d \in [b^k_{\min}, b^k_{\max}]$, we can compute:

$$P(d) = \frac{b^k}{\sum_{i=1}^{s} b^i} \quad (4)$$

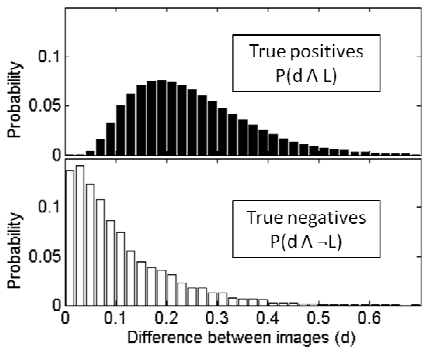

Figure 1. Example of a distribution of image differences for images taken at different locations (top plot) and at the same location (bottom plot), using GIST image descriptors.

If we have multiple samples from each location in an environment, we can calculate the distributions $P(d \wedge L)$ and $P(d \wedge \neg L)$ in a similar manner to $P(d)$. These calculations are identical to that described for $P(d)$ above. The difference values between images from the same location are used to compute the bin values for $P(d \wedge L)$, and difference values between images from different locations are used to calculate $P(d \wedge \neg L)$. Figure 1 shows an example of these probability distributions across an environment; the top plot shows $P(d \wedge L)$, whilst the bottom plot shows $P(d \wedge \neg L)$.

All other probabilities can be calculated from the distributions $P(d \wedge L)$, $P(d \wedge \neg L)$ and $P(L)$. The conditional probabilities $P(d \mid L)$ and $P(d \mid \neg L)$ can be calculated using the standard definition:

$$P(X \mid Y) = \frac{P(X \wedge Y)}{P(Y)} \tag{5}$$

$P(d)$ can be described in terms of $P(d \mid L)$ and $P(d \mid \neg L)$ via the marginal distribution:

$$P(d) = P(d \mid L)P(L) + P(d \mid \neg L)P(\neg L) \tag{6}$$

Finally, $P(L \mid d)$ and $P(\neg L \mid d)$ can be calculated using Bayes Theorem similar to that presented in (2).

Using the simple sample-based method described here we can generate the probability distribution for an environment. In the next section we present an application of these distributions, where a whole-image probability function is integrated into the appearance-based visual localization system CAT-Graph.

C. Whole-Image CAT-Graph

The localization system selected to evaluate the effectiveness of combining whole-image descriptors with existing probabilistic methods was CAT-Graph [5], a well-established appearance-based localization system. CAT-Graph uses a particle filter, which weights particles based on appearance matching and robot motion. At time step k, the weight of particle number i is updated according to:

$$\hat{w}_k^{(i)} = w_k^{(i)} P(Z_k \mid x_k^{(i)}) P(x_k^i \mid x_{k-1}^{(i)}, u_k) \tag{7}$$

where Z_k is the current observation, x_k is the location of the particle, and u_k is the most recent control input. For a whole-image descriptor, this equation is updated to:

$$\hat{w}_k^{(i)} = \frac{w_k^{(i)} P(d_{k(i)} \mid x_k^{(i)}) P(x_k^i \mid x_{k-1}^{(i)}, u_k)}{P(d_{k(i)})} \tag{8}$$

where $d_{k(i)}$ is the difference function between the current observation and the observation at particle i. As in [5], the observation at particle i is based on a linear interpolation between the nearest viewed images. The key difference in this calculation is that the equation includes a denominator term, as discussed in Section III.A. Because the denominator depends on the probability of $d_{k(i)}$, not the probability of the observation Z_k, the denominator will differ for each particle and must be explicitly calculated.

CAT-Graph also includes a new location particle that represents the probability that the robot is visiting a previously unseen location. The new location particle is weighted according to samples selected from the training data to determine Z_{avg} and u_{avg}:

$$\hat{w}_k^{new} \propto P(Z_k \mid Z_{avg}) P(u_{avg} \mid u_k) \tag{9}$$

For a whole-image descriptor method, the probability that the robot is visiting a new location is weighted according to n samples from the distribution $P(\neg L \mid d)$:

$$\hat{w}_k^{new} = \frac{1}{n} \sum_j^n P(L_j \neq L_k \mid d_{kj}) P(u_{avg} \mid u_k) \tag{10}$$

All other aspects of CAT-Graph remain unchanged.

D. Online Calculation of Probability Models

The underlying process is based on the frequency-based probability theory described in Section III.B. However, in order to incrementally generate the distributions we must also decide which samples to add at each time step as the system makes observations and new data becomes available. The algorithm has two stages – an initialization stage and a localization / update stage. The initialization stage is run at the start of the localization process for a fixed number of time steps. The initialization length is a parameter that can be chosen by the user; in our experiments we found 100 to be a suitable number of initialization time steps.

1) Initialization Stage

The initialization stage creates a distribution for $P(d \wedge \neg L)$, prior to localization occurring. If no initialization phase was performed, it is likely that spurious localization matches would occur in the early phases of operation. During this phase no samples are added to the distribution $P(d \wedge L)$. To $P(d \wedge \neg L)$ we add, at each time step *k*, the value:

$$d_{\min} = \min\{d_{ki} \mid i \in 1,...,k-1\} \tag{11}$$

2) Update Stage

Post-initialization, the distribution $P(d \wedge L)$ is updated at time step k with the value $d_{\min}$ as defined in (11), while the distribution $P(d \wedge \neg L)$ is updated with the second-smallest difference value that does not come from the same location as $P(d \wedge L)$. To determine what this value is, a recursive test is performed. More details on how to calculate this value are provided in [15], and a pseudo-code outline of the algorithm is included in the Appendix.

After each update step, a localization calculation is performed, calculating (2) using the estimated probability distribution.

IV. Experimental Setup

The probabilistic localization systems presented in Section III were evaluated against existing pre-trained probabilistic localization systems, and tuned non-

probabilistic systems.

- Whole-image CAT-Graph was tested against conventional local feature CAT-Graph.
- The training-free probabilistic localization system was compared to FAB-MAP, a pre-trained local feature localization system.
- The training-free whole-image system was compared to a heuristic, threshold-based localization method that received identical visual input.

As well as evaluating the performance of the presented probabilistic models these experiments also evaluated the portability of whole-image localization across multiple datasets and the use of multiple descriptor types. For this reason, three separate datasets and three different descriptor types were used. Comparison between the performance of different whole-image descriptors on the same dataset can be found in [15].

- Whole-image CAT-Graph was tested on a large-scale dataset [25] that included circuits undertaken at multiple times of day. The descriptor type was based on the local feature descriptor BRISK [26].
- The training-free localization system was compared to FAB-MAP on a benchmark FAB-MAP dataset [4] to ensure a fair comparison between whole-image and local feature methods. The descriptor type GIST [17] was used.
- The probabilistic vs. non-probabilistic experiment was performed on a dataset that demonstrated highly variable conditions [27] including sun, rain and darkness. Images were compared using sum of absolute differences (SAD) between pixels.

A. *Whole-Image vs. Local Feature Descriptors*

Whole-image CAT-Graph as described in Section III.C, was compared against local feature CAT-Graph on the St Lucia [25] dataset. This dataset covers a 15km path which was repeated 5 times during the day between early morning and late afternoon, and differences in sun position and shadows created visual change between each circuit (see Figure 2).

The circuit captured at midday was chosen as the base dataset to which the other 4 circuits were compared. Both versions of CAT-Graph used a probability model generated on the base dataset. For whole-image CAT-Graph, a 100-bin probability distribution was generated on the base dataset as described in Section III. The local descriptor CAT-Graph closely matched that presented in [5]. This version uses SURF features [8] extracted using the OpenCV [28] implementation. A codebook and Chow-Liu tree [29] were generated using training data from the base dataset.

Figure 2. Sample locations from St Lucia dataset.

To process the St Lucia images, the whole-image descriptor BRIEF-Gist [23] was used, but using the descriptor BRISK [26] in place of BRIEF. Each image was partitioned into 5×5 tiles, each of size 48×48 pixels, and a 512-bit BRISK feature descriptor was calculated around the center of each tile using the OpenCV [28] implementation. All CAT-Graph parameters not directly related to the image descriptors were selected to match the parameters used on this dataset in [5].

B. *Training-Free vs. Trained Localization*

In this experiment, the incremental probability algorithm was compared to FAB-MAP [4], a pre-trained localization system based on local features and a visual bag-of-words [9, 10]. The dataset used was the Oxford City Centre dataset originally presented in [4]. It consists of 2,474 images captured by a mobile robot along approximately 2km of city roads and parks, along with hand-corrected GPS ground truth. For each image, a 512-element GIST [17, 18] vector was computed using the implementation from [18]. Results were compared against published FAB-MAP results [4].

C. *Probabilistic vs. Non-Probabilistic Localization*

In this experiment, the incremental probability localization system was compared against a non-probabilistic model using identical visual input. In particular, this experiment compared the performance of a globally and manually optimized threshold-based approach to the training-free probabilistic model over changing environmental conditions.

The dataset used was originally presented in [27] and consisted of images captured by a webcam mounted on a small remote controlled car that was driven around a varied environment of parkland and university campus. The route was repeated four times under quite different and challenging environmental conditions: sunrise, sunny weather, rainy weather and nightfall (see Figure 3 for sample images). The sunrise dataset was chosen to be the base dataset against which the three traversals were tested, using the incremental probabilistic model and also using a non-probabilistic image comparison model.

For online localization operation, a heuristic approach requires a pre-tuned threshold level to determine whether place recognition has occurred. However, this threshold is environment dependent, and must be selected before operation begins. This experiment tested the significance of a threshold value by selecting threshold values to maintain specific precision levels across all three traversals, and testing how recall was affected.

Figure 3. Examples of some challenges in the Botanic Gardens dataset. From left to right: sun glare, motion blur, rain, and low light.

The images were processed in the same manner as described in the original paper [27]. The images were cropped to the upper half only and resolution reduced to 48×24 pixels. The images were patch normalized over

discrete 8×8 blocks, and sum of absolute differences on the pixel intensity values was used as the image comparison technique. As ground truth data was not available, the images were manually compared to determine correct and incorrect matches.

V. Results

Section A of the results presents precision-recall results for whole-image CAT-Graph and local-descriptor CAT-Graph across 5 different times of day, Section B compares the incremental probability algorithm to published FAB-MAP results on the City Centre dataset, and Section C compares the incremental probability algorithm to a non-probabilistic localization method, focusing particularly on the effect of threshold selection across different environmental conditions.

A. Whole-Image vs. Local Feature Descriptors

The recall achieved by whole-image and local-descriptor CAT-Graph at 100%, 99%, and 90% precision are shown in Table I across each of the 5 times of day. The 12:10pm traversal was used as the base and the other traversals were compared to it. Localization was also performed within the 12:10pm dataset to demonstrate how each system performed when localizing at a single time of day.

TABLE I. Precision and Recall – St Lucia

Time of Day	Whole-image CAT-Graph			Local descriptor CAT-Graph		
	100%	***99%***	***90%***	***100%***	***99%***	***90%***
8:45 am	3.6%	32.3%	85.7%	0.04%	0.04%	0.04%
10:00 am	18.7%	69.0%	100%	1.9%	1.9%	3.7%
12:10 pm (self comparison)	41.3%	61.0%	80.0%	48.9%	69.0%	80.5%
14:10 pm	6.8%	36.9%	96.3%	0.10%	0.10%	0.10%
15:45 pm	2.5%	2.5%	55.5%	0.3%	0.3%	0.3%

CAT-Graph combined with whole-image descriptors performs significantly more reliably across different times of day than when local feature descriptors are used. Local descriptor CAT-Graph performs well when comparing to locations seen at the same time of day (the 12:10pm self-comparison example) and in fact out-performs whole-image CAT-Graph. However, it drops off drastically across different times of day, even when the difference is as small as a 2 hour period.

Furthermore, whole-image CAT-Graph exhibits a consistent increase in recall as precision is lowered, for all times of day. Whole-image CAT-Graph achieves at least 55% recall at 90% precision on all tested datasets (see shaded columns in Table I). Local descriptor CAT-Graph does not exhibit the same increase in recall as precision decreases, as generally recall values are equally poor at 100% and 90% precision.

B. Training-Free vs. Trained Localization

The precision and recall achieved by the incremental probability algorithm on the City Centre dataset is shown in Figure 4. The maximum recall achieved by the incremental probability algorithm at 100% precision is 34%. At 99% precision, recall is 49%. This performance is comparable to that of FAB-MAP [4], which achieves maximum recall of 37% at 100% precision. However, unlike FAB-MAP, the incremental probability results are achieved without the requirement of prior training or environment-specific tuning.

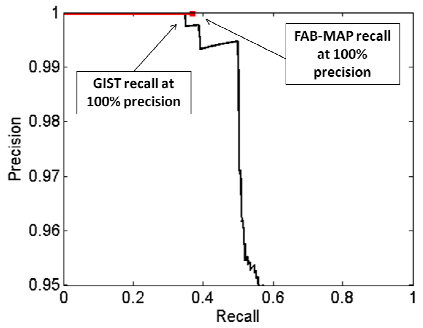

Figure 4. Precision-recall curve for the incremental probability algorithm on the Oxford City Centre dataset. The red line with square terminal represents FAB-MAP recall at 100% precision. The black line represents probabilistic GIST performance.

C. Probabilistic vs. Non-Probabilistic Localization

This experiment evaluated the effect of threshold selection on probabilistic and non-probabilistic localization. Localization algorithms need to be able to operate online, using only data that has already been observed to draw conclusions about place recognition matches. In practice, this means that a heuristic model such as [27] must pre-select a threshold value t to determine whether two images represent the same location or not. Figure 5 displays the precision against threshold for the incremental probabilistic model (red lines) and a heuristic model (blue lines). The incremental probability is plotted against 1-t so both methods can be displayed on the same plot.

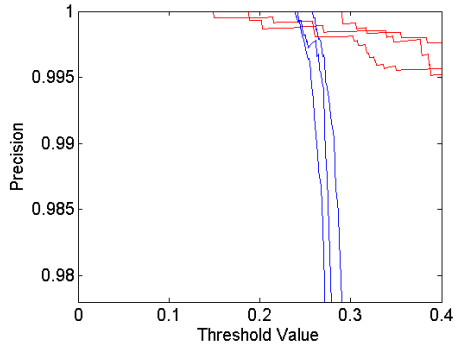

Figure 5. Precision versus threshold value for probabilistic (red lines) versus non-probabilistic method (blue lines) for the Botanic Gardens datasets.

The non-probabilistic model is very sensitive to threshold choice - a threshold variation of as little as 0.05 (0.24 to 0.29) can drop the precision from 100% to 92%. On the other hand, for the probability model the precision only decreases from 100% to 99.52% over a threshold variation of 0.11.

For both models, threshold values were selected to maintain specific precision levels over all three datasets. As the dataset ground truth was limited, no data about

precisions lower than 99.52% were available for the probabilistic model, so only precision above this value were tested. The results are displayed in Table II.

TABLE II. PRECISION AND RECALL WITH THRESHOLDS– BOTANIC GARDENS

Desired Precision	Threshold required	Recall at desired precision		
		Dataset 2	Dataset 3	Dataset 4
		Probabilistic		
100%	0.851	26.09%	37.97%	**22.02%**
99.52%	0.6	**47.5%**	53.6%	49.19%
		Non-probabilistic		
100%	0.239	**18.4%**	36.71%	24.2%
99.52%	0.256	**33.8%**	53.6%	39.07%

For each precision level and localization method, the dataset with the lowest recall has been shaded. For a threshold value that is selected to maintain 100% precision on all three circuits, the performance of the non-probabilistic model on the worst performing dataset is 18.4% and the lowest recall for the probability model is 22.0%. If the threshold value is selected to maintain 99.52% precision on all three circuits (the lowest precision we have data for), the lowest recall of the heuristic model is 33.8% on Dataset 2, and the lowest recall for the probabilistic model is 47.5%, also on Dataset 2. In other words, when global threshold values are taken into account, the probabilistic model can achieve 13.7% higher recall at the same precision, without the requiring manual tuning to produce globally optimal performance across each dataset.

VI. DISCUSSION AND FURTHER WORK

This paper presents a method for creating probability models for whole-image descriptors online. The method does not require a prior training phase.. The goal is to automate probability model creation so that appearance-based localization can occur without a human in the loop. Consequently, there is the potential for "out-of-the-box" operation in different and changing environments without the need for either a training or parameter tuning phase. Furthermore, the method is flexible and can likely be utilized with any choice of whole-image descriptor.

We have also demonstrated the integration of the generated probability models into probabilistic localization systems, such as CAT-Graph, that were originally intended for local feature descriptors, thereby allowing CAT-Graph to exploit the benefits of whole-image descriptors and extending the functionality of CAT-Graph to environments that experience perceptual change.

We note that there is a trade-off between the high recall of whole-image comparison and the high precision of local feature comparison. As can be seen in Experiment 1, whole-image descriptors can perform impressively at 99% or 90% precision but may fail to achieve good recall at 100% precision. However, methods such as [30, 31] allow correction of occasional false positive loop closures using topological information. Future work will incorporate these additions into the whole-image probabilistic localization system. We are also adapting the system to allow scaling to larger environments and more drastically changing environmental conditions.

APPENDIX

We present the algorithm for simultaneous localization and probability estimation. The following notation is assumed:

$\mathbf{b_m}$	one of the $\boldsymbol{s}$ bins representing $P(d_{ki}\|L_k=L_i)$. Bin $\boldsymbol{b_m}$ is the bin which holds value $\boldsymbol{m}$.
$\mathbf{c_m}$	one of the $\boldsymbol{s}$ bins representing $P(d_{ki}\|L_k\neq L_i)$.
$\mathbf{d_k}$	a vector containing the difference values between the current observations $\boldsymbol{I_k}$ and all previously seen observations $\boldsymbol{I_1,\ldots,I_{k-1}}$
match	the best matching locations for stages $\boldsymbol{1,\ldots,k-1}$.

```
begin
  function mainFunction(m,z,dk) :
    (m,z):=min(dk);
    if k < startFrame :
      initialize(m);
    else :
      calcProb(m);
      updateProb(m,z,dk);
    end
  end

  function initialize(m) :
    cm := cm + 1;
  end

  function calcProb(m) :
    return bm / (bm + cm);
  end

  function updateProb(m,z,dk) :
    mr := recursiveUpdate({z},dk,match);
    bm := bm + 1;
    cmr := cmr + 1;
    match(k) := z.
  end

  function recursiveUpdate(Mr,dk,match) :
    (mr,zr):=min(dk);
    if (match(zr) ∈ Mr OR match(min(Mr)) = zr) :
      dk(zr):= ∞;
      Mr+1 := Mr U {zr};
      recursiveUpdate(Mr+1,dk,match);
    else
      return mr;
    end
  end

end
```


ACKNOWLEDGMENT

We thank Will Maddern for access to the CAT-Graph code, and Arren Glover for the St Lucia dataset.

REFERENCES

[1] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *International Journal of Robotics Research,* vol. 30, pp. 1100-1123, 2011.

[2] M. Milford and G. Wyeth, "Persistent Navigation and Mapping using a Biologically Inspired SLAM System," *The International Journal of Robotics Research,* vol. 29, pp. 1131-1153, 2010.

[3] C. Valgren and A. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems,* vol. 58, pp. 157-165, Feb 2010.

[4] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *International Journal of Robotics Research,* vol. 27, pp. 647-665, Jun 2008.

[5] W. Maddern, M. Milford, and G. Wyeth, "CAT-SLAM: Probabilistic Localisation and Mapping using a Continuous Appearance-based Trajectory," *International Journal of Robotics Research,* vol. 31, pp. 429-451, 2012.

[6] T. Nicosevici and R. Garcia, "Automatic Visual Bag-of-Words for Online Robot Navigation and Mapping," *Robotics, IEEE Transactions on,* vol. 28, pp. 886-898, 2012.

[7] D. Lowe, "Object recognition from local scale-invariant features," in *IEEE International Conference on Computer Vision (ICCV)*, 1999, pp. 1150-1157.

[8] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding,* vol. 110, pp. 346-359, Jun 2008.

[9] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 524-531.

[10] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *IEEE International Conference on Computer Vision (ICCV)*, 2003, pp. 1470-1477.

[11] M. Milford and G. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1643-1649.

[12] M. Milford, I. Turner, and P. Corke, "Long exposure localization in darkness using consumer cameras," in *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*, 2013.

[13] H. Badino, D. Huber, and T. Kanade, "Real-time topometric localization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1635-1642.

[14] N. Sünderhauf, P. Neubert, and P. Protzel, "Are we there yet? Challenging SeqSLAM on a 3000 km journey across all four seasons," in *Proc. of Workshop on Long-Term Autonomy, IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[15] S. Lowry, M. Milford, and G. Wyeth, "Training-Free Probability Models for Whole-Image Based Place Recognition," in *Australian Conference on Robotics and Automation (ACRA 2013)*, 2013.

[16] W. Maddern, M. Milford, and G. Wyeth, "Towards Persistent Indoor Appearance-based Localisation, Mapping and Navigation using CAT-Graph," presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.

[17] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Progress in brain research,* vol. 155, pp. 23-36, 2006.

[18] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision,* vol. 42, pp. 145-175, 2001.

[19] A. Murillo, G. Singh, J. Kosecka, and J. Guerrero, "Localization in Urban Environments Using a Panoramic Gist Descriptor," *Ieee Transactions on Robotics,* pp. 1-15, 2013.

[20] C. Siagian and L. Itti, "Biologically inspired mobile robot vision localization," *Robotics, IEEE Transactions on,* vol. 25, pp. 861-873, 2009.

[21] Y. Liu and H. Zhang, "Visual loop closure detection with a compact image descriptor," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 1051-1056.

[22] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Computer Vision–ECCV 2010*, ed: Springer, 2010, pp. 778-792.

[23] N. Sunderhauf and P. Protzel, "BRIEF-Gist - closing the loop by simple means," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 1234-1241.

[24] M. Milford and G. Wyeth, "Mapping a Suburb With a Single Camera Using a Biologically Inspired SLAM System," *Robotics, IEEE Transactions on,* vol. 24, pp. 1038-1053, Oct 2008.

[25] A. Glover, W. Maddern, M. Milford, and G. Wyeth, "FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 3507-3512.

[26] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 2548-2555.

[27] M. Milford and A. George, "Featureless visual processing for SLAM in changing outdoor environments," in *Proceeings of 8th International Conference on Field and Service Robotics*, 2012.

[28] G. Bradski, "Programmer's tool chest: The OpenCV library," *Dr Dobb's Journal of Software Tools,* vol. 25, pp. 120-126, 2000.

[29] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory,* vol. 14, pp. 462-467, 1968.

[30] N. Sunderhauf and P. Protzel, "Towards a robust back-end for pose graph SLAM," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 2012, pp. 1254-1261.

[31] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," presented at the Robotics: Science and Systems, Sydney, Australia, 2012.