2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Accelerated Appearance-Only SLAM

Mark Cummins and Paul Newman
Oxford University Mobile Robotics Research Group. {mjc,pnewman}@robots.ox.ac.uk

Abstract— **This paper describes a probabilistic bail-out condition for multihypothesis testing based on Bennett's inequality. We investigate the use of the test for increasing the speed of an appearance-only SLAM system where locations are recognised on the basis of their sensory appearance. The bail-out condition yields speed increases between 25x-50x on real data, with only slight degradation in accuracy. We demonstrate the system performing real-time loop closure detection on a mobile robot over multiple-kilometre paths in initially unknown outdoor environments.**

I. Introduction

This paper is concerned with speed improvements to an appearance-only SLAM system. We show that by employing a probabilistic bail-out test in the core likelihood calculation, speed improvements of between 25 and 50 times are possible, with only slight accuracy penalty. Typical filter update times are on the order of 150 ms for maps which contain several thousand locations. This enables real-time loop closure detection on a mobile robot for loops tens of kilometers in length.

Our core appearance-only SLAM system has been previously described in [1], [2]. In appearance-only systems, the robot's map consists of a set of locations, each of which has an associated appearance model. When the robot collects a new observation, its location can be determined by deciding which location in the map was most likely to have generated the observation. This approach has recently shown success in large scale global localization [3] and online loop closure detection [1], both difficult problems in more typical metric SLAM frameworks.

The limiting computational cost of appearance-only SLAM is computing the observation likelihood for each location in the map. Typically, only a small number of these places will yield non-negligible probability of having generated the current observation. The main idea of this paper is that by evaluating the appearance likelihoods in parallel, these unlikely hypotheses can be identified and discarded while the likelihood calculation is only partially complete, yielding large speed increases. Very similar ideas have been described elsewhere in computer vision, notably in the context of efficient RANSAC algorithms [4], [5]. Matas and Chum showed that for RANSAC, the sequential probability ratio test (SPRT) yields the optimal solution. The SPRT approach was originally designed for testing two hypotheses under a sequence of identical and equally informative observations [6]. Extensions exist for the multihypothesis case [7]. However, stopping boundaries for the SPRT are not easy to derive when the observations are not equally informative. We describe an alternative approach based on concentration inequalities [8]. Unlike the SPRT, this approach is straight-forward to apply even when there are multiple hypotheses and the observations are not equally informative. We have noted related ideas in other fields [9], however we believe our approach is novel in this context.

II. Appearance-Only SLAM

Our appearance-only SLAM system is described in detail in [1], [2]. Briefly, at time t the robot's map consists of n_t discrete locations, each location L_i having an associated appearance model. Our representation of appearance is inspired by the bag-of-words image retrieval systems developed in the computer vision community [10]. Sensory data is converted into a bag-of-words format; a place appearance model is a distribution over appearance words. We extend the basic bag-of-words approach by learning a generative model for the sensory data, in the form of a Chow-Liu tree [11]. This generative model captures the fact that certain combinations of appearance words tend to co-occur, because they are generated by common objects in the environment, and yields a significant improvement in navigation performance.

When the robot collects a new observation Z_t, we compute $p(L|Z_t)$, the probability distribution over locations given the observation. This can be cast as a recursive Bayes filtering problem:

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

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 likelihood of the observation given the location L_i and the previous observations $\mathcal{Z}^{t-1}$, $p(L_i|\mathcal{Z}^{t-1})$ is our prior belief about our location, and $p(Z_t|\mathcal{Z}^{t-1})$ normalizes the distribution. The normalization term can be written as a summation

$$\begin{aligned} p(Z_t|\mathcal{Z}^{t-1}) &= \sum_{m \in M} p(Z_t|L_m)p(L_m|\mathcal{Z}^{t-1}) \\ &\quad + \sum_{u \in \overline{M}} p(Z_t|L_u)p(L_u|\mathcal{Z}^{t-1}) \end{aligned} \tag{2}$$

over the set of mapped places M and the unmapped places $\overline{M}$. This summation can be approximated by sampling, where the "unmapped places" are drawn from a set of training data. This yields a probability that the observation came from a place not in the map. Using the resulting PDF over location, we can make a data association decision and either add a new location to our map, or update the

appearance model of an existing place. Essentially this is a SLAM algorithm in the space of appearance.

The core of the PDF calculation is computing the observation likelihood $p(Z_t|L_i, \mathcal{Z}^{t-1})$ for each location in the map and each sample in the training set. The following section describes an approach to increasing the speed of this likelihood calculation. By identifying locations that will have insignificant likelihood before the calculation is fully complete, many locations can be excluded quickly and large speed increases can be realized.

III. Probabilistic Bail-Out using Bennett's Inequality

Let $\mathcal{H} = \{H^1, ..., H^K\}$be a set of K hypotheses and let $Z = \{z_1, ..., z_N\}$ be an observation consisting of N features. The likelihood of the observation under the the k^{th} hypothesis is given by

$$p(Z|H^k) = p(z_1|z_2, ..., z_N, H^k)...p(z_{N-1}|z_N, H^k)p(z_N|H^k) \tag{3}$$

Define the log-likelihood of the first i features under the k^{th} hypothesis as

$$D_i^k = \sum_{j=1}^{i} d_j^k \tag{4}$$

where

$$d_i^k = \ln(p(z_i|z_{i+1}, ..., z_N, H_k)) \tag{5}$$

is the log-likelihood of the i^{th} feature under the k^{th} hypothesis. We would like to determine, as rapidly as possible, the hypothesis H^* for which the total log-likelihood D_N^* is maximized. Finding H^* with certainty requires a complete evaluation of the likelihood of each hypothesis, which may be too slow for applications of interest. Consequently, we consider the problem of finding a hypothesis $H^\#$, subject to the constraint that $p(H^\# \neq H^*) < \epsilon$, where ϵ is some user-specified probability.

In overview, our approach is to calculate the likelihoods of all hypotheses in parallel, and terminate the likelihood calculation for hypotheses that have fallen too far behind the current leader. "Too far" can be quantified using concentration inequalities, which yield a bound on the probability that a hypothesis will overtake the leader, given their current difference in likelihoods and some statistics about the properties of the features which remain to be evaluated.

Consider two hypotheses $H^x, H^y \in \mathcal{H}$ and let

$$X_i = d_i^x - d_i^y \tag{6}$$

the difference in the log-likelihood of feature i under hypothesis H^x and H^y. X_i can be considered as a random variable before its value has been calculated. This is useful because we can calculate some key statistics about X_i more cheaply than we can determine its exact value. Now define

$$S_n = \sum_{i=n+1}^{N} X_i \tag{7}$$

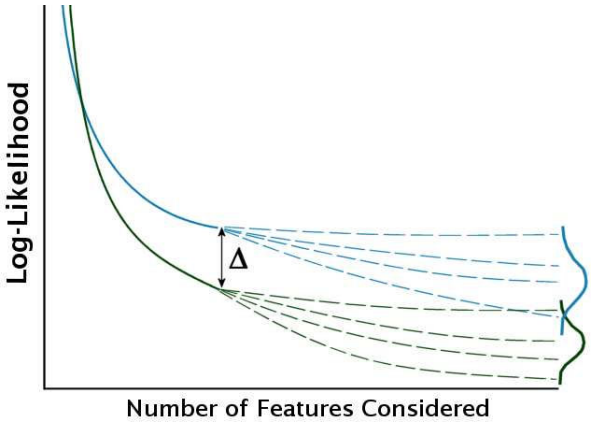


Fig. 1. Conceptual illustration of the bail-out test. After considering the first i features, the difference in log-likelihoods between two hypotheses is Δ. Given some statistics about the remaining features, it is possible to compute a bound on the probability that the evaluation of the remaining features will cause one hypothesis to overtake the other. If this probability is sufficiently small, the trailing hypothesis can be discarded.

If after evaluating n features, the log-likelihood of some hypothesis is Δ less than the current best hypothesis, then the probability of failing to locate H^* if we discard this hypothesis is given by $p(S_n > \Delta)$. Thus, knowing the distribution of S_n allows the creation of a probabilistic bail-out test for discarding hypotheses subject to an error constraint. Calculating an explicit distribution on S_n is infeasible, however concentration inequalities – which bound the probability that a function of random variables will deviate from its mean value – can be applied to yield bounds on $p(S_n > \Delta)$.

A large variety of concentration inequalities exist, many of which apply under very general conditions, including cases where the component distributions are not identically distributed, not independent, and are combined using arbitrary functions. For an overview see [8]. Typically, the more information available about the component distributions X_i, the tighter the bound. Our bail-out test applies the Bennett inequality for sums of symmetric random variables [12]. This inequality is specified in terms of two parameters — M, a bound on the maximum value of any component X_i, and v, a bound on the sum of the variances of the components X_i.

Formally, let $\{X_i\}_{i=n+1}^N$ be a collection of independent mean-zero random variables with symmetric distributions (corresponding to the log-likelihood changes due to those features not yet considered), and satisfying the conditions

$$p\left(|X_i| < M\right) = 1, \forall i \tag{8}$$

$$\sum_{i=n+1}^{N} E\left[X_i^2\right] < v \tag{9}$$

and let

$$S = \sum_{i=n+1}^{N} X_i \tag{10}$$

then the Bennett inequality states that

$$p(S > \Delta) < \exp\left(\frac{v}{M^2}\cosh(f(\Delta)) - 1 - \frac{\Delta M}{v} f(\Delta)\right) \tag{11}$$

where

$$f(\Delta) = \sinh^{-1}\left(\frac{\Delta M}{v}\right) \tag{12}$$

Note that as the calculation of the hypothesis likelihoods progresses, the number of unconsidered features (and hence the number of X_i variables) decreases, so M and v will change. As a result the bail-out threshold changes throughout the calculation.

IV. Application to Appearance-Only SLAM

Ranking Features

We now turn our attention to applying this bail-out condition to our appearance-only SLAM system. Firstly, we must define an order in which to consider the features. While the bail-out test applies to any ordering, it is natural to rank the features by information gain. That way, the hypotheses will converge most rapidly toward their final log-likelihood values and poor hypotheses can most quickly be identified (see Figure 2).

Each of our features z_i is a binary variable indicating whether or not the i^{th} word of the vocabulary was present in the current observation. The occurrence of these visual words is not independent – certain combinations of words tend to occur together because they are generated by some underlying object in the environment. To capture this structure we learn a Chow Liu tree model [11] which approximates the true distribution over the observations. Under this model, each feature z_i is conditionally dependent on one other feature z_{pi}. If we observe $z_i = s_i$ and $z_{pi} = s_{pi}$ (with $s \in \{0, 1\}$), then the information gain associated with this observation under our model is

$$I = -\ln p(z_i = s_i | z_{pi} = s_{pi}) \tag{13}$$

Typically observations of rare words are the highest ranked features, though, perhaps surprisingly, failure to observe a word can sometimes also have high information gain – for example, if two words are almost always observed together, then failure to observe one while observing the other is an informative observation.

Note that because the probabilities in Equation 13 come from the training data on which we learnt the model of our visual words, we are calculating the information gain with respect to the places in the training data. Strictly we should consider the the information gain with respect to the set of places in our current map – for example, some feature might be very rare in the training set but very common in the map. In practice we observe that the difference between the two values is usually small, so maintaining a separate set of probabilities is unnecessary.

Application of Bennett's Inequality

To apply Bennett's inequality, we must calculate v and M, the parameters in Equation 11 which depend on the component random variables X_i. In our appearance-only SLAM system

$$\begin{aligned} X_i &= d_i^x - d_i^y \\ &= \ln(p(z_i|z_{pi}, L_x)) - \ln(p(z_i|z_{pi}, L_y)) \end{aligned} \tag{14}$$

where, recalling our notation from Section II, L denotes a location (hypothesis), and x and y are random variables which specify which locations in the map are being considered. Now, given that the values of z_i and z_{pi} are known, $p(z_i|z_{pi}, L_x)$ depends only on the number of times feature z_i has previously been observed at location L_x (details in [1], [2]). Thus X_i attains its maximum value when x and y correspond to the locations where feature i has been observed most and fewest times respectively. Keeping track of these statistics allows us to easily calculate M.

Calculating v, which bounds the sum of the variances of the X_i variables, requires some information about the distribution of the index random variables x and y. We assume that these have uniform distribution, which effectively amounts to assuming that all of our hypotheses have equal a-priori probability[1]. Given this assumption, the distribution of X_i is fully specified and can be calculated directly by considering $d_i^x - d_i^y$ for all index pairs x, y. We observe that X_i has a multinomial distribution which must be mean-zero and symmetric[2].

To evaluate v we must calculate the variance of this distribution. In practice, this calculation can be fast. For example, in our appearance-only SLAM system, when the robot is first exploring the environment almost all place models have only one observation associated with them, so d_i^x can take on only a small number of distinct values. Keeping track of the possible discrete values of d_i^x and their relative proportion allows for rapid calculation of the variance of X_i. As exploration continues, the possible values of d_i^x become larger, and the calculation becomes more expensive. At some point it may be beneficial to switch from using Bennett's inequality to Hoeffding's inequality [13], a similar concentration inequality that requires knowledge only of the maximum value of each X_i. Hoeffding's inequality gives a weaker bound, but this is compensated for by the fact that by the time the variance becomes expensive to compute, the place models themselves are more differentiated, and so their likelihoods will diverge faster.

One remaining issue is that our appearance-only SLAM system requires a PDF over hypotheses, whereas our discussion so far has concerned locating only the best hypothesis. Computing a PDF requires a simple modification to the bail-out scheme. Consider that instead of locating only the best hypothesis H^*, we would like to locate all hypotheses whose

[1] If the assumption is far from the truth, then Hoeffding's inequality can be applied in place of Bennett's. See below.

[2] If $X_i = c$ for some choice of indices x, y, then $X_i = -c$ for y, x.

log-likelihood is at most C less than that of H^*. C is a user-specified constant chosen so that hypotheses less likely than this can be considered to have zero probability with minimal error. Simply increasing our bail-out distance by C will retain all those hypotheses whose final likelihood may be within this likelihood range, thus giving us a close approximation to the PDF over hypotheses.

A final note – Bennett's inequality requires that the variables X_i are independent. Our Chow Liu model captures much but not all of the conditional dependence between features. Thus the variables X_i may have weak dependence. Our experiments would appear to indicate that this is not a problem in practice.

V. Results

We tested the system on data collected by a mobile robot. The robot collected images to the left and right of its trajectory approximately every 1.5m. Each collected image is processed by our algorithm and is used either to initialize a new place, or, if loop closure is detected, to update an existing place model. Results are presented for three datasets. The first dataset – labeled City Centre – is 2km in length and was chosen to test matching ability in the presence of scene change. It was collected along public roads near the city centre, and features many dynamic objects such as traffic and pedestrians. The second dataset – New College – is 1.9km in length and was chosen to test the system's robustness to perceptual aliasing. It features several large areas of strong visual repetition, including a medieval cloister with identical repeating archways and a garden area with a long stretch of uniform stone wall and bushes. The third dataset – Parks Road – features a typical suburban environment. The robot's appearance model was built from a fourth dataset collected in a different region of the city, the area of which did not overlap with the test sets.

Navigation results for these datasets were generated using both the original SLAM system and the accelerated SLAM system incorporating the bail-out test. All datasets were processed using the same visual vocabulary and algorithm parameters. The bail-out boundary was set so that the probability of incorrectly discarding the best hypothesis at any step was $< 10^{-6}$. This value can be varied to trade off speed against accuracy.

Results are summarized in the figures below. Figure 2 illustrates the bail-out calculation on some real data. Precision-recall curves for the full and accelerated algorithms on the City Centre dataset are shown in Figure 4. The curves were generated by varying the probability at which a loop closure was accepted. Recall rates are quoted in terms of image-to-image matches. As a typical loop closure is composed of multiple images, even a recall rate of 35% is sufficient to detect almost all loop closures. The relative performance of the two algorithms on the other datasets is summarized in Table I. Figure 3 visualizes the performance of the accelerated algorithm on the City Centre dataset. The system correctly identifies a large proportion of possible loop closures with high confidence. There are no false positives that meet the

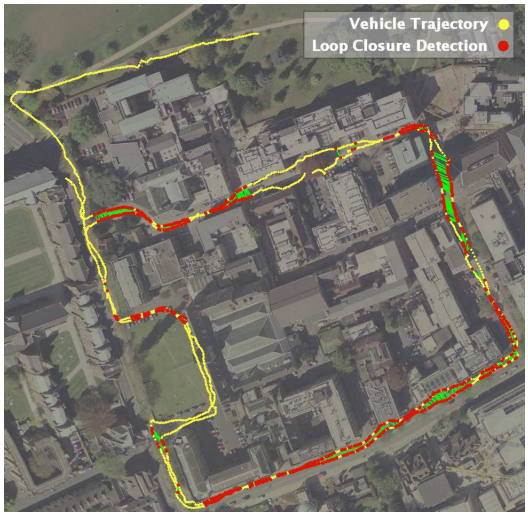


Fig. 3. Appearance-only matching results (using the accelerated algorithm) for the City Centre dataset overlaid on an aerial photograph. The robot travels twice around a loop with total path length 2km, collecting 2,474 images. Each of these images is determined to be either a new place or a loop closure. Positions (from hand-corrected GPS) at which the robot collected an image are marked with a yellow dot. Two images that were assigned a probability $p \geq 0.99$ of having come from the same location are marked in red and joined with a green line. There are no incorrect matches that meet this probability threshold.

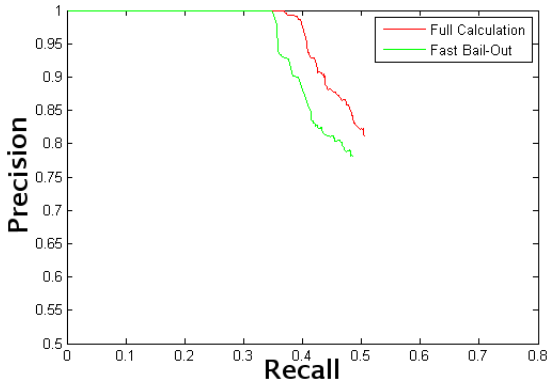


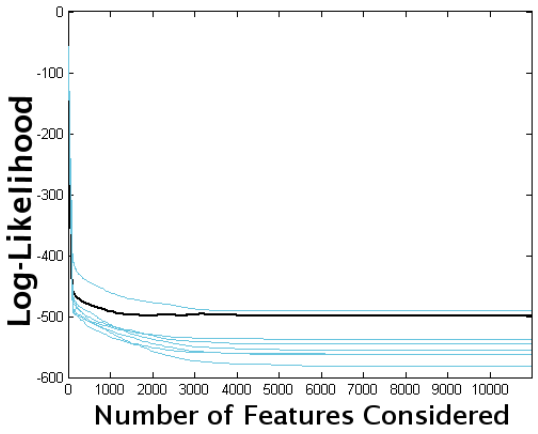
Fig. 4. Precision-Recall curves for the City Centre dataset, showing the full likelihood calculation (red) and the accelerated calculation using the bail-out test (green). Notice the scale.

probability threshold. Figures 6 and 7 show some examples of place recognition performance, highlighting matching ability in the presence of scene change and robustness to perceptual aliasing. The robustness to perceptual aliasing is particularly noteworthy. Of course, had the examples shown in Figure 7 been genuine loop closures they might also have received low probability of having come from the same place. We would argue that this is correct behaviour, modulo the fact that the probabilities in (a) and (b) are too low. The very low probabilities in (a) and (b) are due to the fact that the best matches for the query images are found in the sampling set, capturing almost all the probability mass. This is less likely in the case of a true but ambiguous loop closure, particularly because in the case of a true loop closure the ambiguity can be resolved by temporal information via the prior term in Equation 1.

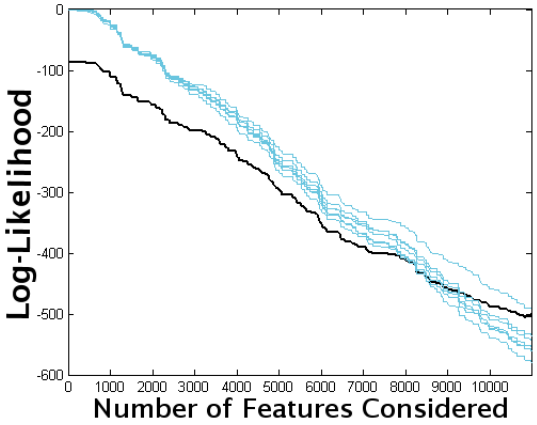
	Full Calculation		Fast Bail-Out		
Dataset	Recall	Mean Time	Recall	Mean Time	Speed-Up
City Centre	37%	5015 ms	35%	141 ms	35.5
New College	46%	4818 ms	42%	178 ms	27.0
Parks Road	44%	4267 ms	40%	79 ms	53.6

TABLE I

COMPARISON OF THE PERFORMANCE OF THE SLAM SYSTEM USING FULL AND ACCELERATED LIKELIHOOD CALCULATIONS. THE RECALL RATES QUOTED ARE AT 100% PRECISION. TIMING RESULTS ARE FOR THE FILTER UPDATE, ON A 3GHZ PENTIUM IV. FEATURE GENERATION ADDS AN EXTRA 330 MS ON AVERAGE. UPDATE TIME FOR THE ACCELERATED CALCULATION IS DATA DEPENDENT AND VARIES FROM OBSERVATION TO OBSERVATION. TIME QUOTED IS THE AVERAGE OVER THE DATASET.



(a) Features Ordered by Information Gain

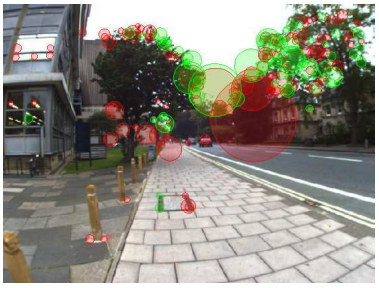


(b) Random Feature Order

Fig. 2. Bail-out test on real data. Here the blue lines show the log-likelihoods of each place versus number of features considered. Typically there are thousands of places - here only a few are shown for clarity. The black line is the bail-out threshold. Once the likelihood of a place hypothesis falls below the bail-out threshold, its likelihood calculation can be terminated (the remainder of the likelihood calculation is shown above for illustration). In (a), observations are ordered of information gain; in (b) they are ordered randomly. Note that ordering the features by information gain results in much faster convergence toward final likelihood values, and hence a much more effective bail-out test. The bail-out threshold does not converge to the leading hypothesis because of the offset constant C.

(a) p=0.996

(b) 0.999998



(c) 0.999992

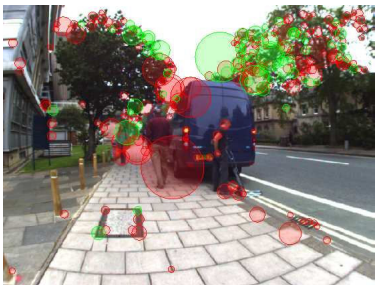
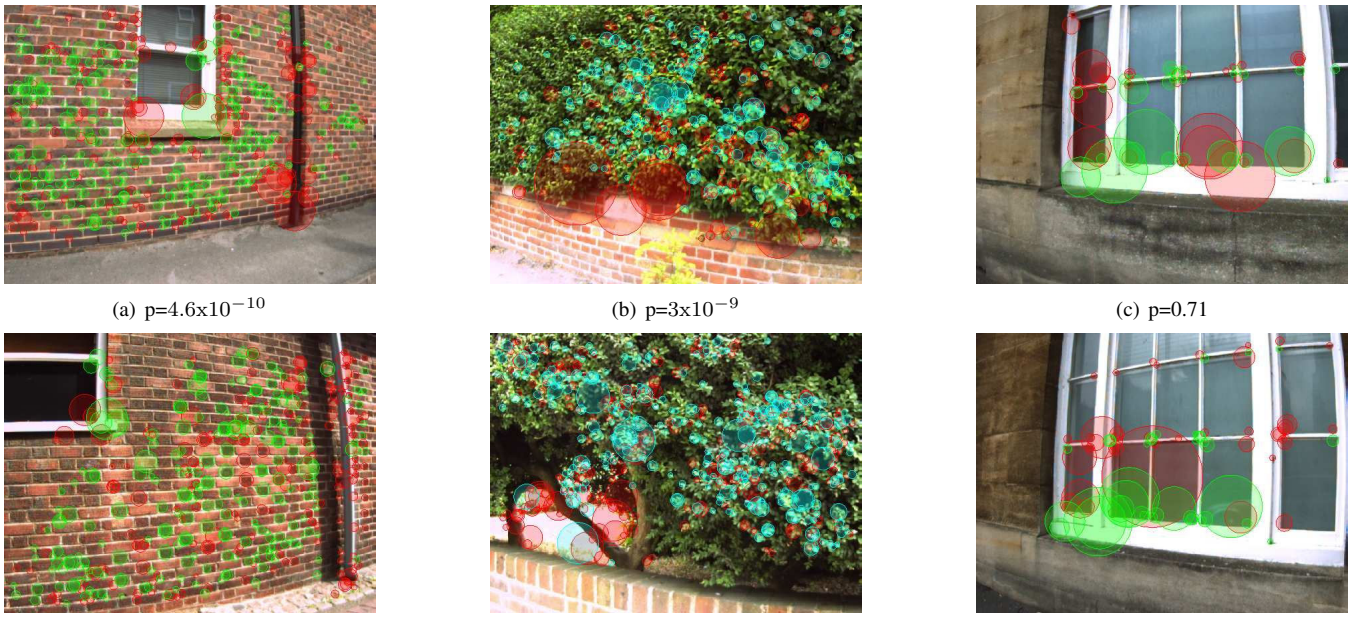


Fig. 6. Some examples of images that were assigned high probability of having come from the same place, despite scene change. Results were generated using the accelerated likelihood calculation. Words common to both images are shown in green, others in red. The probability that the two images come from the same place is indicated between the pairs.



(a) p=4.6x10^{-10} (b) p=3x10^{-9} (c) p=0.71

Fig. 7. Some examples of remarkably similar-looking images from different parts of the workspace that were correctly assigned low probability of having come from the same place. Results were generated using the accelerated likelihood calculation. We emphasize that these examples are not outliers, but represent typical system performance. This result is possible because most of the probability mass is captured by locations in the sampling set – effectively the system has learned that images like these are common in the environment. Words present in both images are shown in green, others in red. (Common words are shown in blue in (b) for better contrast). The probability that the two images come from the same place is indicated between the pairs.

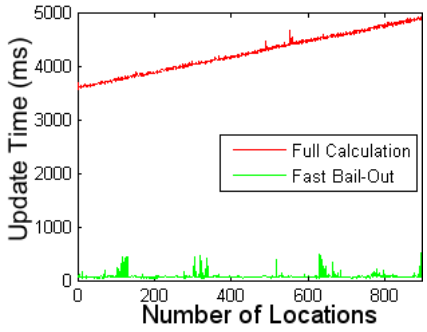


Fig. 5. Filter update time versus the number of locations in the map, for the Parks Road dataset. Update time with zero locations is non-zero due to the fixed cost of evaluating the partition function. Calculation time with the bail-out test grows linearly, however the slope is too small to be seen on this graph.

VI. Conclusions

This paper has presented a new approach to rapid multihypothesis testing using a probabilistic bail-out condition based on concentration inequalities. Concentration inequalities exist that apply under very general conditions, even for arbitrary functions of non-iid random variables, hence our basic idea should be applicable to a wide variety of problems. We have applied the bail-out test to accelerate an appearance-only SLAM system. The speed increase is data-dependent, but acceleration factors in the range 25x-50x are typical in our tests. The location recognition performance of the accelerated system is only marginally worse than the full solution, and more than sufficient for reliable online loop closure detection in mobile robotics applications. We have presented results demonstrating online loop-closure detection over 2km loops, however the system is fast enough to scale to loops of tens of kilometres in length while maintaining sub-second filter update times. Investigating system performance on this scale will be a focus of future work.

Acknowledgments: The work reported in this paper was funded by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre established by the UK Ministry of Defence and by the EPSRC.

References

[1] M. Cummins and P. Newman, "Probabilistic appearance based navigation and loop closing," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA'07)*, Rome, April 2007.

[2] ——, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Intl. Journal of Robotics Research*, to appear.

[3] G. Schindler, M. Brown, and R. Szeliski, "City-Scale Location Recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2007.

[4] D. Nistér, "Preemptive RANSAC for live structure and motion estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.

[5] J. Matas and O. Chum, "Randomized RANSAC with sequential probability ratio test," in *Proc. IEEE Intl. Conf. on Computer Vision (ICCV)*, 2005.

[6] A. Wald, *Sequential Analysis*. New York: Dover Publications, 1947.

[7] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1994–2007, 1994.

[8] S. Boucheron, G. Lugosi, and O. Bousquet, *Concentration Inequalities*. Springer, 2004, vol. Lecture Notes in Artificial Intelligence 3176, pp. 208–240.

[9] O. Maron and A. W. Moore, "Hoeffding races: Accelerating model selection search for classification and function approximation," in *Advances in Neural Information Processing Systems*, 1994.

[10] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proceedings of the Intl. Conf. on Computer Vision*, Nice, France, October 2003.

[11] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. IT-14, no. 3, May 1968.

[12] G. Bennett, "Probability inequalities for the sum of independent random variables," *J. Am. Stat. Assoc.*, vol. 57, pp. 33–45, March 1962.

[13] W. Hoeffding, "Probability Inequalities for Sums of Bounded Random Variables," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.