Closing Loops Without Places

Christopher Mei, Gabe Sibley and Paul Newman
Mobile Robotics Group
University of Oxford, UK

Abstract—This paper proposes a new topo-metric representation of the world based on co-visibility that simplifies data association and improves the performance of appearance-based recognition. We introduce the concept of dynamic bag-of-words, which is a novel form of query expansion based on finding cliques in the landmark co-visibility graph. The proposed approach avoids the - often arbitrary - discretisation of space from the robot's trajectory that is common to most image-based loop closure algorithms. Instead we show that reasoning on sets of co-visible landmarks leads to a simple model that out-performs pose-based or view-based approaches. Using real and simulated imagery, we demonstrate that dynamic bag-of-words query expansion can improve precision and recall for appearance-based localisation.

I. INTRODUCTION

Robotic mapping is the task of building a representation of the world in view of providing robots with the capability of achieving high-level goals. It is one of the key components to autonomy. A full mapping system should provide accurate pose and map estimates, be scalable and provide a way of relocalising when mapping fails.

Maps are said to be metric when they represent the position of the robot and landmarks in a particular coordinate frame and aim to provide accurate estimates. Topological maps on the other hand typically provide a representation in the form of a graph with places or landmarks as nodes and edges encoding some form of connectivity such as traversability.

Topo-metric maps combine both aspects and typically provide locally accurate estimates with a connected graph of regions built for example from the robot's trajectory and a loop closure mechanism. The work presented in this article falls within this category. However it distinguishes itself from the recent trend of view-based maps by proposing a new representation that works directly on landmarks and not on poses. The proposed idea is to build a representation based on co-visibility (i.e. the fact that landmarks were seen together) that also provides metric information through a set of relative frames in which the landmark coordinates are represented. This relative co-visibility graph can be seen as an extension to our previous work [1], [2] with a more natural way to manage landmark matching and leads to a new approach to loop closure.

In this representation we infer the notion of locality through the observed landmarks and not through the robot's trajectory. This can be beneficial when reexploring previously visited areas to reduce the dependency on loop closures for data association (Section II). Furthermore this
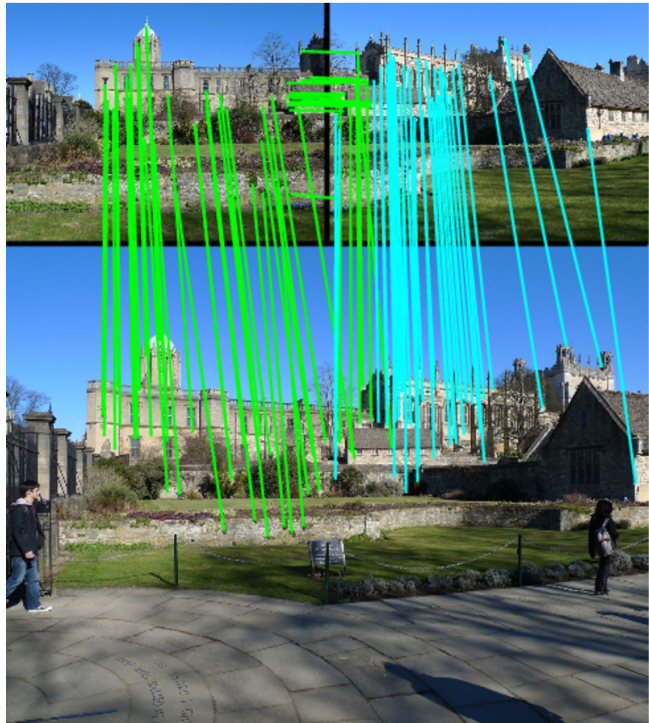

Fig. 1. Applying state-of-the-art pose-based loop closure solutions such as FABMAP [3] to these three images, the results will depend on the order. If the top two images are provided first, then the bottom image is considered to be the same location as the top right image with high probability ($p > 0.96$). If the bottom image and one of the two top images are provided first, each will be considered as a new location with high probability ($p > 0.99$). Neither of these solutions are very satisfactory as all three images overlap. This reflects a more fundamental issue with associating the notion of place to an image. In this article, we propose a solution to this problem by reasoning on landmarks directly, thus closing loops without places.

representation leads to a loop closure mechanism that no longer relies on the - often arbitrary - notion of place chosen by the discretisation of the robot's path but directly on the underlying map itself. This is achieved by finding possible local sets of observed landmarks that could have come from a particular place at *query* time and not beforehand as is typically done in the literature, where the notion of place is inherited from the notion of document from the natural language processing literature (Fig. 1). As such the proposed approach can be described as a *dynamic* bag-of-words approach. The method shares similarities with the query expansion literature [4] but in this work we benefit from the knowledge of a map and a set of co-visible landmarks

that avoids the complex task of building a latent model at query time and applying subsequent new queries.

After introducing the relative co-visibility graph and how it relates to a relative pose-graph approach in Section II, we discuss the concept of virtual location and provide an efficient way to compute possible loop closures. Finally, in Section III-B, we will show experimental evidence of the advantage of dynamic bag-of-words.

A. Related work

Visual SLAM has seen an increasing number of real-time solutions [5], [6], [7], [2], [8] in recent years. Scalability has been shown to be achievable using a topo-metric representation that provides precise local estimates [9], [8], [2]. Furthermore, most recent works combine a form of relocalisation or loop-closure mechanism [6], [7], [2], [8] for better robustness and increased precision.

When representing the world for mapping, we have the choice of mapping using a global or a relative representation [2] but also to represent the state as the trajectory and the map combined or just the trajectory. The latter approach leads to the popular pose-graph or view-based maps [10], [11], [12] for which efficient optimisation methods exist. The structure is then represented implicitly. Loop closure for pose-graph maps are often based on the poses taken individually (similar to the notion of document for text recognition). We believe that it can be beneficial to work directly on the landmarks for loop closure. This motivates the relative co-visibility graph presented in this work.

A recent survey of loop closure methods can be found in [13], [12]. The authors classify loop closures in the following way: (i) image-to-map that consists in mapping the landmarks of a given image to the map directly [14] (ii) image-to-image [3], [12] where poses are compared based on their observations and (iii) map-to-map [15] where larger portions of the reconstructed environment are used to provide reliable location matches.

The current article attempts to combine the advantages of these different methods to achieve high-quality loop closing using the single underlying map representation. The closest related work is [7] and [16]. In [7] the authors provide a loop-closure mechanism that relies on local maps to compute a matching score. In [16] a similar approach is used where these "local maps" are replaced by image nodes expanded to contain neighbouring landmarks. In the current work, however, we build locations at *query* time based on all likely visible landmarks and thus reason on "virtual locations" based on co-visibility (Fig. 1). This provides a stronger support to assess correct loop closures than bag-of-words built on individual images. It should be noted that the proposed approach is different to the notion of "scan patches" proposed in [17] that improves the matching when a potential loop closure has been proposed but does *not* improve the finding of potential loop closure candidates.

II. THE RELATIVE CO-VISIBILITY GRAPH

In this section we describe how the relative co-visibility graph can be used for data association between contiguous frames and for metric pose estimation, thus providing a topo-metric representation of the world.

In this work, we move away from the notion of place, that requires a discretisation of space, to work on landmarks directly. The relative co-visibility graph is a representation that captures which landmarks were observed at the same time. In this sense, it is a topological map where the connectivity corresponds to co-visibility.

To represent the landmarks in the 3-D world, we require a metric representation of the environment. We choose to represent landmark positions in frames that are connected to other frames by an estimated transform. Typically - but not necessarily - these frames will be robot poses and the links will correspond to the estimated inter-frame motion. In this sense, it is also a *relative* metric map. We will refer to this representation as the relative co-visibility graph (RCG).

These two notions - topology and metric estimation - interact in a specific way: when the robot explores the environment, it needs to associate data to estimate its pose. To decide which landmarks to associate with observations in the image, we use the topological connectivity: we expect to see landmarks that were seen recently or have been seen with the ones seen recently. The expected position of these landmarks can then be estimated through the kinematic chain relating the current pose to the pose representing a given landmark's position. This representation can be seen as an extension of the recent work on the continuous relative representation (CRR) [1], [2] and we will now compare the two methods.

Like most work on relative representations, the CRR is a pose-based or view-based representation. The connected poses are used to choose which landmarks to associate during the exploration of the robot. Specifically, the CRR defines the notion of *active region* that corresponds to all the poses and measured landmarks at a given depth in the relative graph (obtained for example through a breadth-first-search). Data associations do not explicitly influence which landmarks are matched.

Figure 2 illustrates the difference between using connected poses and co-visibility for data association. The relative co-visibility graph improves over the CRR by no longer using the robot poses to choose which landmarks to match but the fact that landmarks were viewed together (i.e. co-visible) at a given time. Before any loop closure occurs, the representations are similar, Fig. 2(d). The advantage of this approach becomes clear after a loop closure, Fig. 2(f). In the CRR case, if landmarks are matched but no new links are added in the graph, features that we might expect to see are ignored (Fig.2(c)). In the co-visibility graph, as long as the features are observed, they will belong to the active region. This avoids a strong dependency of the results of loop closure. Reliable loop closures are not always possible due to spatial aliasing for example [3]. It could be argued that a link could be added in the CRR for each new observation but this would greatly increase the complexity of the graph. Another option is to make links close to the last frame (this possibility is discussed in [2]). However these issues
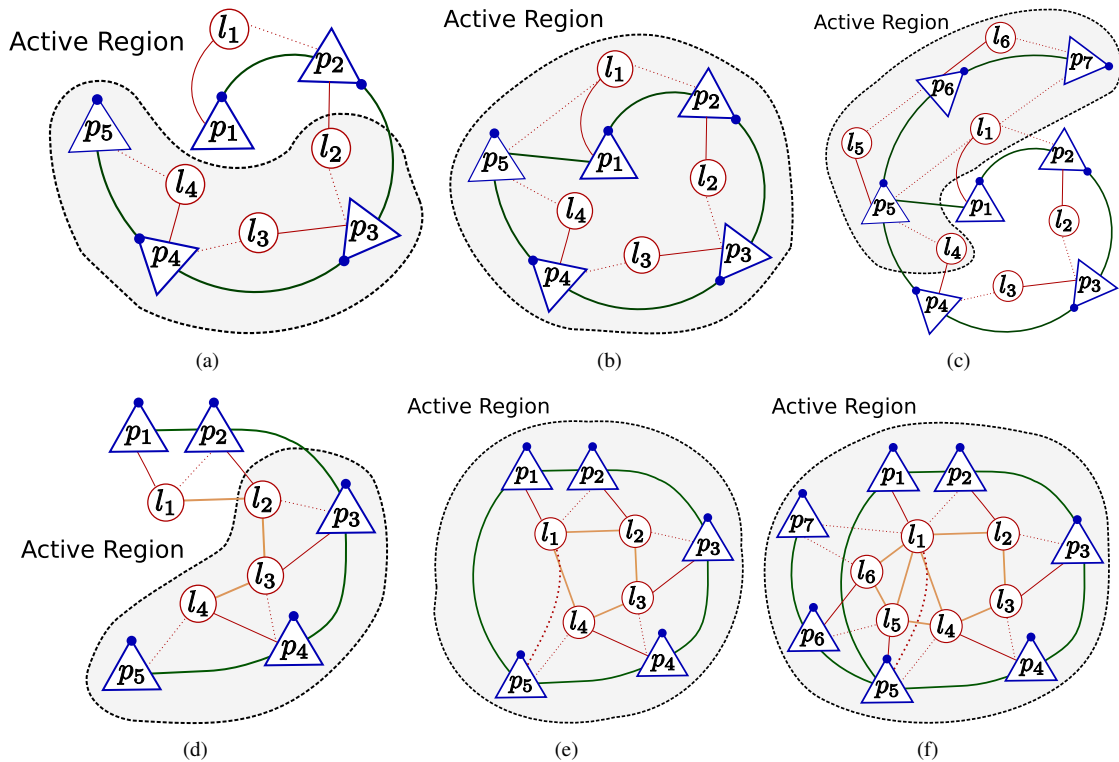
Fig. 2. Comparison between the continuous relative representation (CRR) on the top row and the relative co-visibility graph (RCG) on the bottom row on a spiral motion example with a loop closure. p_i corresponds to the robot position at time i. l_j is the jth landmark. The active region is defined as the set of poses and landmarks that are at a given graph distance (two in this example) from the current pose. The active region is used to decide what landmarks we will try to associate at the following time step. Figures 2(a) and 2(d) show the CRR and RCG representations before a loop closure has occured. The active regions have the same size and the underlying representations are similar. After a loop closure (Fig. 2(b) and 2(e)) triggered by the observation of l_1 from p_5, in both the CRR and RCG, the active regions encapsulate the entire map and a new transform is estimated between p_5 and p_1 symbolised by a link. However, in the RCG, a **new link** is also added to the co-visibility graph between l_1 and l_5. The importance of this co-visibility link becomes apparent when the robot explores the environment further (Fig. 2(c) and 2(f)). In the case of the CRR, despite the observation of l_1 from p_7, the active region of depth two does not encapsulates landmark l_2 even though it is likely to be visible because of the observation of l_1. In contrast, the active region deduced from the RCG encapsulates the entire map as each landmark is less than a distance of two in the co-visibility graph. Reasoning on co-visibility avoids the data association being dependent on the arbitrary choice of the transform links in the relative representation.
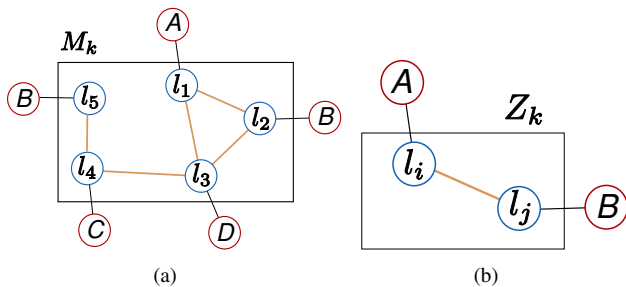

Fig. 3. A co-visibility map is a set of landmarks connected by an edge if they were observed jointly (Fig. 3(a)). Given a vocabulary $V = \{A, B, C, D\}$, each landmark has an associated word describing it. An observation is a count of words (Fig. 3(b)). The aim of loop closure can be stated as finding the correct association between landmarks based on the observed words.

illustrate a deeper conceptual difference: co-observability is a property of the underlying world (and sensor) in contrast to the robot trajectory that is to a great extent arbitrary.

III. IDENTIFYING LOOP-CLOSURES IN THE CO-VISIBILITY GRAPH

Currently, the most successful approaches to vision-based loop closure match a given pose to previous poses in the map [3], [12], [13]. One of the advantage of pose-based mapping is the efficiency that can be obtained for recognition [18], [3]. However, as illustrated by Fig. 1, these approaches dependent strongly on the chosen discretisation of space. We will now propose a solution that avoids this issue and improves recognition by exploiting co-visibility explicitly.

We will start by describing the framework and notations used and then introduce the concept of "virtual location" (VL) before describing how efficient loop closures can be computed at run-time. As we will see, VLs correspond to co-visible landmarks that share words with the query. We will not use the position estimates of landmarks or poses and in this sense the approach is purely topological.

A. Framework and notations

The proposed approach to loop closure falls within the framework of bag-of-words for place recognition [19]. A vocabulary, that we will note V, is built by kd-tree clustering

of SURF features[1]. A different descriptor such as SIFT or efficient compact signatures [12], [20] could also be used.

A map M_k at a given time k is described by an undirected graph of landmarks with edges indicating co-visibility. Each landmark has an associated word from the vocabulary V (Fig. 3(a)). The aim of loop closure is to find the correct data association between the landmarks represented by the words Z_k at time k and the map M_k (Fig. 3(a)).

B. Virtual locations

Given words Z_k extracted from a single image, instead of matching them directly with the most likely previously observed pose, we build clusters of landmarks from the map that could have generated this observation. The clusters are generated at query time, with the following steps:

1) Find each landmark in M_k that has a word belonging to the bag-of-words of Z_k.
2) Cluster the landmarks that are co-visible and extend these clusters with all co-visible landmarks (these are the landmarks we expect to see). These non-overlapping clusters are dubbed virtual locations (VLs).
3) Associate a score to each each VL as described in Section III-D. (In practise, we would then typically use geometric checking to improve the precision but this was not done in this work.)

In the case of relocalisation or kidnapped robot loop closures, Z_k is simply the words found in a given image. However during the normal operation of the system, the landmarks from the map are associated with the observations at each time step through tracking. Discovering new loop closures can then be done by augmenting Z_k with the words from all co-visible landmarks to provide a stronger support for loop closure.

In the case of Fig. 3, Z_k will induce two VLs: $\text{VL}_1 = \{l_1, l_2, l_3\}$ and $\text{VL}_2 = \{l_4, l_5\}$.

A naive implementation for building virtual locations at run-time would be computationally expensive. We will now explore how to reduce this cost by noting the important role of cliques in the co-visibility graph.

C. Clique matrix representation of the co-visibility graph

Computing co-visible clusters corresponds to finding neighbours of particular nodes in the graph. Cliques play an important role: if a word belongs to Z_k and a clique, all landmarks of the clique will belong to the same VL. Computing maximal cliques is computationally expensive, however as observations are made simultaneously at a given time, we have a natural clique factorisation of the landmarks through the poses of the graph. For this reason, in the current work, cliques coincide with poses. However, it should be noted that poses or images are just a support for computation but no specific meaning is associated to them.

In this work we represent the co-visibility graph through its clique factorisation using a (sparse) clique matrix, a

generalisation of the incidence matrix [21]. Lets assume the following sequence of landmark observations lead to M_k in Fig. 3: $C_1 = \{l_1, l_2, l_3\}$, $C_2 = \{l_1, l_3\}$, $C_3 = \{l_3, l_4\}$ and $C_4 = \{l_4, l_5\}$. Based on the sequence of observations, we obtain the following adjacency matrix $\mathbf{A}$ (with added self-connections) and associated non-maximal clique matrix $\mathbf{C}_c$:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \mathbf{C}_c = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$\mathbf{C}_c$ can be interpreted as a factorisation as $\mathbf{A} = H(\mathbf{C}_c \mathbf{C}_c^\top)$ with $H(.)$ the element-wise Heaviside function ($H(\mathbf{A})_{ij} = 1$ if $a_{ij} \geq 0, 0$ otherwise). This factorisation is not unique (maximal cliques would provide a $\mathbf{C}_c$ with fewer columns) but in this work they coincide with poses and each row corresponds to the observed landmarks for a given pose.

When building the clique matrix representation of the co-visibility graph, we also build an inverted index linking words to cliques as in [19].

Building virtual locations at run-time can be done in the following way:

1) Compute the observed words Z_k from the image.
2) Find the list of cliques that contain the words using the inverted index, $L_c = \{C_1, C_2, \ldots, C_M\}$.
3) Cluster L_c into virtual locations based on co-visibility. Co-visibility can be easily detected in $\mathbf{C}_c$ by checking for 1s on the same row for two different columns (i.e. cliques).

Two parameters can be used to control the number and size of VLs:

1) Minimum number of words required to select a given clique for L_c. Low values will increase the number of cliques but make the choice sensitive to common words and outliers. With large values, relevant locations might be missed. This parameter can be set based on the number of landmarks we expect to see: (probability of finding the correct word)×(number of visible landmarks). This probability will be denoted by P in Section IV.
2) Minimum number of co-visible landmarks required to consider another clique as co-visible, referred to as M in Section IV. Low values will lead to large VLs that are not sufficiently discriminative for loop closure. Large values will increase the fragmentation of VLs with large local variability of loop closure scores; at the limit we obtain pose-based loop closure. It should be noted that pose-based loop closure often relies on smoothing to avoid this strong variability. (Alternatively, this term can be written as a percentage of the number of words belonging to the query.)

D. Matching score

In this work, we use the standard Term Frequency Inverse Document Frequency score (tf-idf) to rank the VLs

for loop closure [19]. A vector is associated to each VL, $(t_1, t_2, \ldots t_{|V|})$ with:

$$t_i = \frac{n_{il}}{n_l} \log \frac{N}{n_i}$$

$\frac{n_{il}}{n_l}$ is the term-frequency for the specific VL computed as the number of occurrences n_{il} of word i in the VL divided by the number of words n_l in VL. $\frac{n_i}{N}$ is the document-frequency computed over all the VLs as the total number of occurrences n_i of word i divided by the total number N of VLs for the given query. The final score for each VL is then given by the inner product of its tf-idf vector with the query tf-idf vector computed from Z_k (considered as a VL). The scores are divided by the largest VL score to obtain values between 0 and 1.

E. Complexity

The complexity of the proposed approach is related to the inverted index. The inverted index look-up can be made constant-time (hash table). However the number of returned cliques (or poses in the current implementation) can be linear the number of cliques if the same word appears in each pose. The virtual locations only add an extra factor in the computation as they require a complexity linear in the number of cliques returned by the query. To summarise, the proposed algorithm has a similar complexity to a standard inverted index scheme.

## IV. EXPERIMENTAL RESULTS

A. Evaluating loop closure accuracy

To evaluate the results provided by the proposed approach, we need to define a metric to decide if a given loop closure is correct. The relevant poses for a given query are chosen as the locations that have at least one word in common with the query and are within a distance t (in this work we choose $t = 4m$).

To compute the precision and recall, the result (set of poses) returned by the system is taken to be the union of the poses belonging to the virtual locations above a given value of the score. Figure 3(b) illustrates an example of a query and loop closure result.

In Section III-C, two parameters were introduced that influence loop closure: (1) P the probability of measuring the correct word (for $P = 0$ we use at least one word to generate the query) and (2) M, the minimum number of co-visible landmarks required to consider a neighbouring clique as co-visible. When $M \to \infty$, we obtain standard pose-based loop closure. With small values for M, we typically get large VLs (potentially the entire map). Trivial solutions are checked for based on distance (if the query size is bigger than a fixed value it is discarded, here we use 20m) and not counted as a contribution to precision/recall. The value P can typically be calibrated for a particular detector as in [3]. We can expect that larger values of P will lead to better precision with lower recall. In this section, we evaluate its effect on performance.
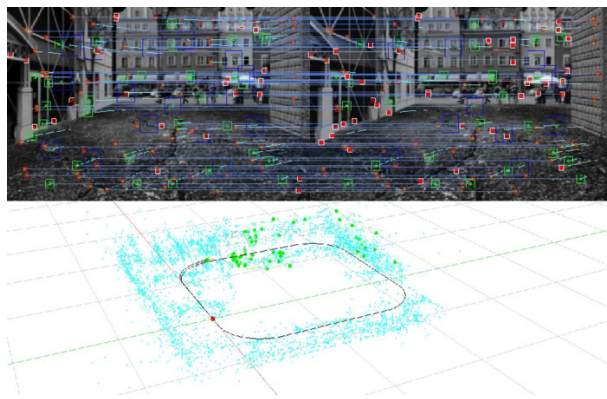
(a) Stereo images and trajectory of the simulation sequence.
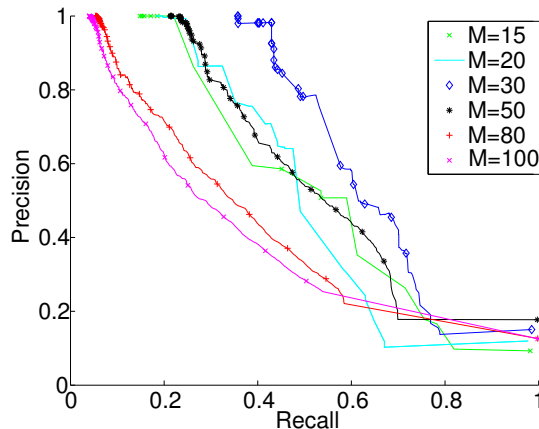
(b) Precision-recall curve on the simulation sequence. Different values of M were tested. Larger contexts are beneficial when common words are avoided when computing virtual locations.

Fig. 4.   Simulation sequence.

In these examples, SURF features and SURF descriptors provided the image-to-image matching to build the topological co-visibility map. We did not compute any pose estimation nor did we reject any outlier matches. Both example sequences have stereo images but we only processed the left image in this work. For loop closure and relocalisation, the SURF descriptors were quantised into words using the outdoor vocabulary available on the FabMap web-page.

B. Simulation sequence

The simulation sequence consists of a short loop of 500 frames (Fig. 4(a)) built from a rendered 3-D model. The precision-recall curve (Fig. 4(b)) was computed for $P = 0$ and $M = \{15, 20, 30, 50, 80, 100\}$. The best value of context size for this data set is around $M = 30$. For smaller values of M, the context is detrimental as it returns large, poorly-defined regions. The performance drops more significantly for $M > 30$ as the context size decreases and tends towards pose-based loop-closure.

C. Begbroke sequence

Loop closure performance was evaluated on the 1km (23K frames) Begbroke outdoor sequence (Fig. 5). This sequence
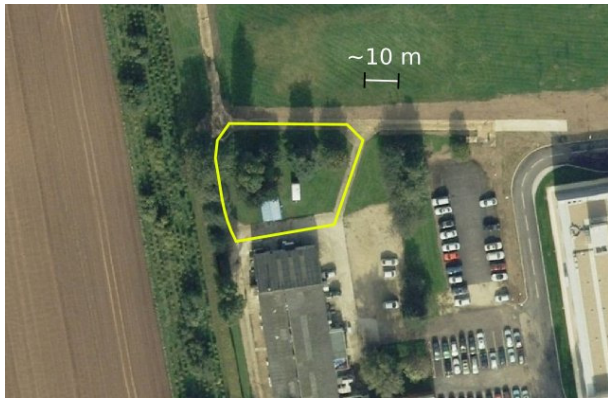
(a) Begbroke sequence aerial view.
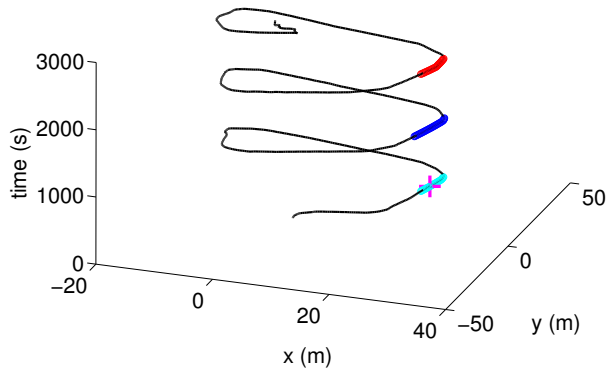
(b) Trajectory in time with results of a loop closure query. The cross indicates the pose used for the query and the circled coloured regions are the loop closure results.

Fig. 5. Begbroke sequence with three loops. Figure 5(b) shows the planar trajectory with time on the z axis.

is quite challenging for loop closure as it contains large stretches with an open space and a paved path where few SURF features were extracted. The ground truth used to evaluate the correctness of the loop closures was provided by the stereo system [2] that gives an accuracy of less than 1m on this sequence.

Figure 6(a) shows the precision-recall curve with $P = 0.1$ for varying levels of $M = \{1, 10, 20, 50, 100\}$. As M increases, the performance of the retrieval diminishes which confirms the importance of context for retrieval provided by the co-visibility. For very low values of M, large unstable regions are created that impact the precision.
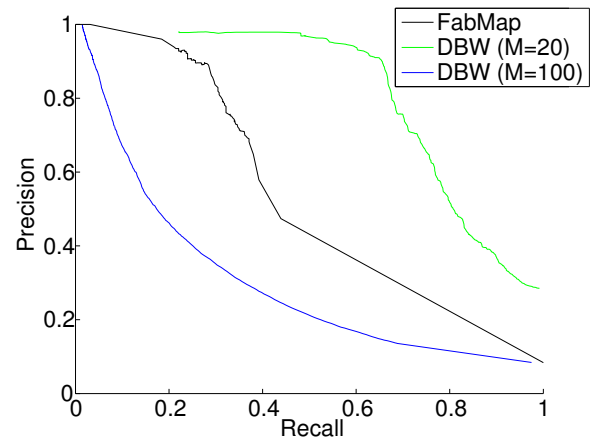
Figure 6(b) shows the precision-recall curve for $M = \{20, 60\}$ with $P = \{0.1, 0.3, 0.5\}$. The precision is partly improved for very low recall in the case of large contexts but the overall performance is degraded. Context reduces the effect of incorrect words; reducing the number of virtual locations on which to compute scores is generally detrimental. In the pose-based case, the precision does improve which probably stems from the computation of the inverse document frequency. In both cases the highest achievable
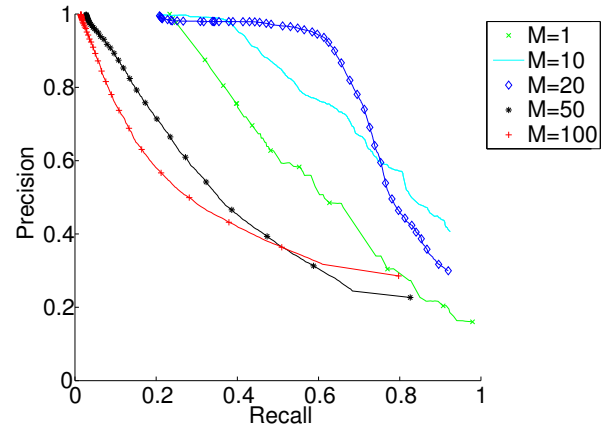
(a) Precision-recall curve for varying virtual location sizes (through M). Co-visibility generally improves the performance of loop closure except when common words are used that leads to large unstable VLs.
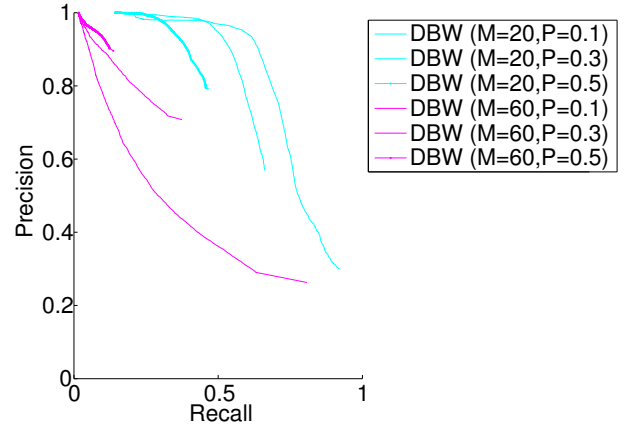
(b) Precision-recall curve for varying values of P and $M = \{20, 60\}$. Context provides resilience to incorrect data association so high values of P are not beneficial for improved precision. Without context, retrieval does benefit from reducing incorrect data association at the detriment of the highest achievable recall.

(c) Precision-recall comparison between FabMap and dynamic bag-of-words (DBW). DBW shows an improvement over the current state of the art.

Fig. 6. Precision-recall curves for the Begbroke sequence.

recall rate is reduced.

Figure 6(c) compares FabMap performance to dynamic bag-of-words with $M = \{20, 100\}$ and $P = 0.1$. In the case of FabMap, only every 100 poses was processed to reduce the bias due to the re-observation of landmarks. This is a typical limitation of pose-based methods where overlap affects the retrieval performance. For this sequence, dynamic bag-of-words shows an improvement over the state of the art.

In conclusion, these results show that using co-visibility generally improves precision and recall. It can also have the detrimental effect of creating large unstable sets of poses if common words are not taken into account which is consistent with literature on information retrieval.

V. Conclusion

In this article, we investigated the use of co-visibility as a way to represent the position of landmarks and poses for mapping but also as an approach to manage landmark matching and represent important contextual cues. We introduced a way to find potential loop closures at query time using the notion of co-visibility and virtual locations to avoid the often arbitrary discretisation of space imposed by pose-based mapping.

Future work will investigate how loop closure can be improved through the use of metric information. Another research direction is the use of the context provided by the graph to help object recognition.

Acknowledgement

References

[1] G. Sibley, C. Mei, I. Reid, and P. Newman, "Adaptive Relative Bundle Adjustment," in *RSS*, Seattle, USA, June 2009.

[2] C. Mei, G. Sibley, M. Cummins, P. Newman, and I. Reid, "A constant time efficient stereo SLAM system," in *BMVC*, London, UK, 2009.

[3] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *The International Journal for Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[4] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in *ICCV*, Rio de Janeiro, Brazil, 2007.

[5] A. J. Davison, "Real-Time Simultaneous Localisation and Mapping with a Single Camera," in *ICCV*, 2003.

[6] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," in *ISMAR*, Nara, Japan, 2007.

[7] E. Eade and T. Drummond, "Unified Loop Closing and Recovery for Real Time Monocular SLAM," in *BMVC*, Leeds, UK, 2008.

[8] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1066–1077, Oct. 2008.

[9] M. Bosse, P. Newman, J. Leonard, and S. Teller, "Simultaneous Localization and Map Building in Large-Scale Cyclic Environments Using the Atlas Framework," *The International Journal for Robotics Research*, vol. 23, no. 12, pp. 1113–1139, 2004.

[10] E. Olson, J. Leonard, and S. Teller, "Spatially-adaptive learning rates for online incremental slam," in *RSS*, Atlanta, GA, USA, June 2007.

[11] G. Grisetti, C. Stachniss, S. Grzonka, and W. Burgard, "A tree parameterization for efficiently computing maximum likelihood maps using gradient descent," in *RSS*, Atlanta, GA, USA, June 2007.

[12] K. Konolige, J. B. and. JD Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, "View-based maps," in *RSS*, 2009.

[13] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós, "A comparison of loop closing techniques in monocular slam," *Robotics and Autonomous Systems*, 2009.

[14] B. Williams, G. Klein, and D. I. Reid, "Real-time slam relocalisation," in *ICCV*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.

[15] L. Clemente, A. Davison, I. Reid, J. Neira, and J. Tardos, "Mapping Large Loops with a Single Hand-Held Camera," in *RSS*, 2007.

[16] E. Jones, "Large scale visual navigation and community map building," Ph.D. dissertation, University of California, 2009.

[17] P. Newman and K. Ho, "SLAM Loop Closing with Visually Salient Features," in *ICRA*, 2005.

[18] D. Nistér and H. Stewenius, "Scalable recognition with a vocabulary tree," in *CVPR*, vol. 2, New York, USA, 2006, pp. 2161–2168.

[19] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *ICCV*, vol. 2, Nice, France, Oct. 2003, pp. 1470–1477.

[20] M. Calonder, V. Lepetit, K. Konolige, P. Mihelich, J. Bowman, and P. Fua, "Compact signatures for high-speed interest point description and matching," in *ICCV*, Kyoto, Japan, September 2009.

[21] D. Barber, "Clique Matrices for Statistical Graph Decomposition and Parameterising Restricted Positive Definite Matrices," in *UAI*, Helsinki, Finland, 2008.