# A comparison of loop closing techniques in monocular SLAM

Brian Williams†, Mark Cummins†, José Neira⋆, Paul Newman†, Ian Reid† and Juan Tardós⋆
⋆Universidad de Zaragoza, Spain, †University of Oxford, UK

***Abstract*— Loop closure detection systems for monocular SLAM come in three broad categories: i) map–to–map, ii) image–to–image and iii) image–to–map. In this paper, we have chosen an implementation of each and performed experiments allowing the three approaches to be compared. Using these insights we go on to describe an extension to the image–to–map matching approach which makes more use of the available information to improve the algorithm.**

## I. Introduction

Loop closure detection is an important problem for any SLAM system and, since cameras have become a common sensor in robotics applications, more people are turning towards vision based methods to achieve it. In this paper, we compare three quite different approaches to loop closure detection for a monocular SLAM system. The approaches essentially differ in where the data association for detecting the loop closure is done – in the metric map space or in the image space. The three approaches are as follows:

- **Map–to–map** – Correspondences are sought between features in two submaps taking into account both their appearance and their relative positions. In this paper we look at the method of Clemente *et al.* [2], who applied the variable scale geometric compatibility branch and bound (GCBB) algorithm to loop closing in monocular SLAM. The method looks for the largest compatible set of features common to both maps, taking into account both the appearance of the features and their relative geometric location.
- **Image–to–image** – Correspondences are sought between the latest image from the camera and the previously seen images. Here, we discuss the method of Cummins *et al.* [4] [3]. Their method uses the occurrences of image features from a standard library to detect that two images are of the same part of the world. Careful consideration is given to the distinctiveness of the features – identical but indistinctive observations receive a low probability of having come from the same place. This minimises false loop closures.
- **Image–to–map** – Correspondences are sought between the latest frame from the camera and the features in the map. We examine the method of Williams *et al.* [12] who find potential correspondences to map features in the current image and then use RANSAC with a three–point–pose algorithm to determine the camera pose relative to the map.

该方法考虑到特征的外观及其相对几何位置，寻找两个地图共有的最大兼容特征集

在当前图像中找到与地图特征的潜在对应关系，然后使用RANSAC与三点姿态算法来确定相对于地图的相机姿态

First, we briefly describe the underlying monocular SLAM system used during the experiments. Then, we describe in more detail the chosen implementation of each of the different approaches to loop closure. Results are then given on the performance of each algorithm at closing a loop and comparisons are made between the methods. The bulk of this work on comparing the methods is covered in [12] where the image–to–map method is introduced. Finally, we describe an extension to the image–to–map method which makes use of more of the available image information.

## II. The Monocular SLAM System

The monocular SLAM system used is derived from Davison's original system [5] where the pose of a handheld camera is tracked, while simultaneously building a map of point features in 3D using the EKF. The underlying system is essentially the same as the system described in [2], but with a relocalisation module [13] to recover from situations where the system becomes lost.

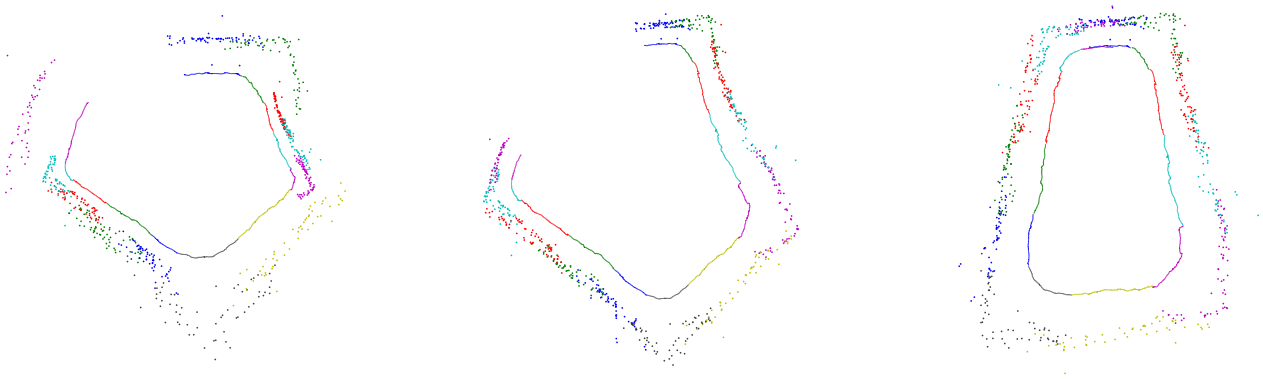
The Hierarchical SLAM [6] submapping technique is used to both reduce linearisation errors and to allow the system to make larger maps in real time. The system creates a series of submaps while determining the relative scale differences between the maps which result from using a bearing only sensor. For more details of this Hierarchical SLAM technique in monocular SLAM see [2]. The scale correction can be seen in Fig. 1(a) and (b).

When loop closure is detected, the global hierarchical map can be updated by adjusting the transformations between submaps in a non-linear constrained optimisation. The result of the optimisation after the loop closure has been detected is shown in Fig. 1(c). This loop closure can be detected in many ways though as will be discussed in the next section.

## III. Detecting Loop Closure

In order to close loops in a map, the system must recognise when it has returned to a previously mapped region of the world. Essentially, at this point two regions in the map are found to be the same region in the world even though their position is incompatible given the uncertainty estimate in the map – the classic loop closure problem. The system must then be able to calculate the transformation needed to align these two regions to 'close the loop'.

In the following sections, we describe three methods for detecting loop closure based on three quite different approaches. We will later test the performance of all three algorithms.



(a) Local maps obtained with pure monocular SLAM

(b) Local maps auto-scaled

(c) After loop closing

Fig. 1. Map made of a university courtyard. Twelve submaps with a total of 848 features were made during the 70m trajectory. The loop closure was detected using the image–to–map method [12].

### A. *Map–to–Map Matching: Clemente* et al.

Clemente *et al.* [2] presented a method to close loops in monocular SLAM maps based on finding correspondences between common features in different submaps. The algorithm used is a variable scale version of the original geometric compatibility branch and bound algorithm (GCBB) [10]. The system uses both similarity in visual appearance (unary constraints) and relative distances between features (binary constraints) to find the largest compatible set of common features between two submaps. Once a consistent set has been found, the relative scale, rotation, and translation needed to align the two submaps can easily be determined.

The system was shown to work in [2] where it found a set of five common features between the first and last submaps in a large loop.

### B. *Image–to–Image Matching: Cummins* et al.

Cummins *et al.* [4] have developed a method to detect loop closures based on recognising the visual appearance of previously seen places. The matching is performed by detecting in each image the presence or absence of features from a visual vocabulary [11] based on SURF features [1], which is learned off–line from training data. Note that the training data consists of generic images not collected in the environment where loop closure detection is performed. The system takes into account the probabilities of features appearing together, and is able to work out the probability that two images show the same region of the world. This method does not depend on a metric map being created since it only compares images directly. However, it can be used with a metric map if the camera pose relative to such a map can be found for each image as well as the relative pose between two images for the loop closure. Much work has been done on this problem in the field of computer vision [8].

### C. *Image–to–Map Matching: Williams* et al.

In [12] a loop closure detection method is proposed which is based on a relocalisation technique used to recover from tracking failures [13]. This relocalisation module determines the pose of the camera relative to a map of point features by finding correspondences between the image and the features in the map. The pose is then determined from the correspondences using RANSAC and the three–point–pose algorithm [7].

The relocalisation module is able to run faster than framerate through the use of a fast matching algorithm [13] based on the randomised fern classifier [9]. While the features are being tracked, each successful observation is used to train the classifier. This classifier is fast but it has a high false positive rate. Incorrect classifications are handled using RANSAC.

To detect loop closures, the system uses the module to attempt relocalisation in distant regions of the map according to the feature covisibilities. When a relocalisation is successful, it gives a correspondence between the current pose being tracked, and the pose given by the relocalisation elsewhere in the map. This gives the translation and rotation needed to align the two regions, but a single pose is not enough to determine the scale difference. To achieve this, the camera is tracked for some time in both regions (while freezing one of the maps so information is not counted twice), and this common trajectory can be used to find the transformation between the two regions including the relative scale difference (Fig. 2).

## IV. RESULTS

We have used the monocular SLAM system to build a map of a university courtyard. Due to the size of the environment, the system built twelve submaps as the camera was moved around the 70m trajectory facing the wall. Each new submap was begun by initialising new features in the same image locations as those just observed as the last submap finished. These common features can then be used to fix the relative scale between submaps as shown in Fig. 1.

Even after the scale between submaps has been corrected, the map still exhibits a common problem, that although it has returned to the same region in the world, this is not reflected in the map. A loop closure detection system is needed to recognise that the system has traversed a loop so the map can be corrected accordingly.

We have used all three algorithms to try to detect the loop closure in this sequence. We have also evaluated the

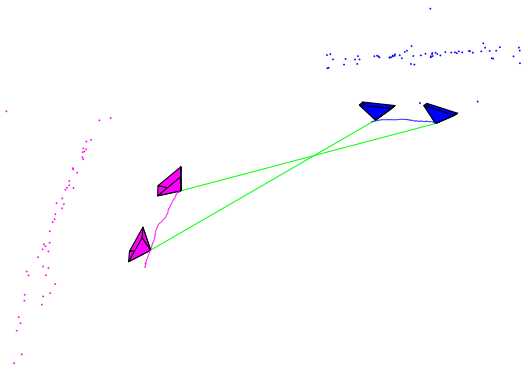


Fig. 2. While tracking in the twelfth map (left), the system relocalises in the first submap (right) using our image–to–map algorithm. The two supmaps are merged by first aligning the common trajectories, and then enforcing the constraint that the two sets of corresponding camera poses (linked by green lines) are equal.

performance of the algorithms further by checking their susceptibility to false positives and their run time.

### A. *Map–to–Map Matching: Clemente* et al.

When the system comes to close a loop using the map-to-map method, it is able to find the common features between the two maps as shown in Fig. 4(a). Unfortunately, during the loop closure, there is no guarantee that the system will have initialised features in the exact same place in two different maps. In fact, in our experiments to date, we have found submaps with sufficient common features to detect the loop closure to be rare. Fig. 3 shows an example of the same frame being tracked in two different maps. Despite the large number of features visible, only two features are common to both maps.

事实上，在我们迄今为止的实验中，我们发现具有足够共同特征的子映射很少检测到环路闭合

Even getting a corresponding set of features does not guarantee a true correspondence between the two submaps. Fig. 4(b) shows that the GCBB algorithm also found sets of five “common” features between eight other pairs of submaps. We were unable to find a threshold able to reliably distinguish between true positives and false positives for the maps created by our SLAM system.

During our tests, the variable scale GCBB algorithm took around 100ms[1] to compare two maps. When the SLAM system finishes one submap, there is easily time to compare this submap to all previous submaps before the next one is completed.

### B. *Image–to–Image Matching: Cummins* et al.

The image–to–image matching method of Cummins *et al.*, is designed to work with non-overlapping key frames. When run on a robot, the odometry is used to trigger key frame capture. Without odometry, we simply used every 40th frame of the video to test the system. Ideally though, an automatic key frame detector should be used.

The loop closure detection system determines for each of these input images if it is a new place or a loop closure.

[1]Tests were done on a Dual Core 3GHz machine.

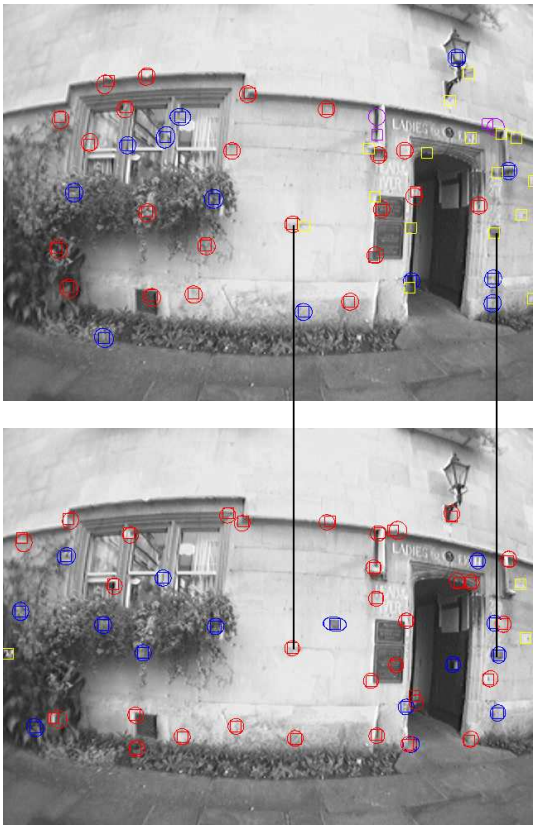


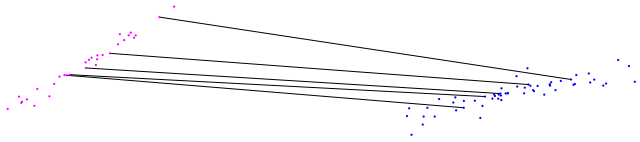
Fig. 3. During the overlap in the sequence, the system tracks the camera in two submaps. The colours indicate if an observation was successful (red), unsuccessful (blue), rejected by JCBB (purple), or not attempted, (yellow). Only two of the features are actually common to both submaps. This makes it impossible for the map–to–map method to detect the loop close.

此时，系统给出的最新图像与序列开始时的图像对应的概率很高(99.9%)(图4(c))

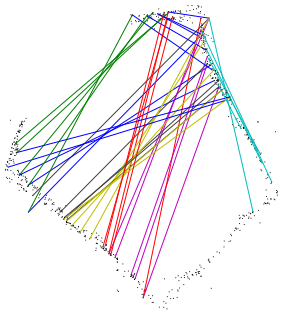
The algorithm correctly gave high probability that each image was a new place until the camera had traversed the loop and returned to the start of the loop. At this point, the system gave high probability (99.9%) that the most recent image corresponded to an image at the start of the sequence (Fig. 4(c)).

To test the reliability of the loop closure detection, we computed loop closures for every frame from a second lap of the courtyard, against the set of images from the first lap. This simulates the ‘kidnapped robot situation’, a sudden transition from the end of the first loop to a random part of the courtyard. It is a way to test if the algorithm would be able to detect a loop closure at each position. The results are shown in Fig. 4(d) where frames that matched an image in the previous loop are marked. A threshold was chosen that removes all false positives to allow comparison with the image–to–map method. The system found matches that met this probability threshold in 8% of attempts indicating that the system would be able to close the loop at these positions. The precision–recall curve in Fig. 5 shows the effect of the probability threshold on the reliability of the system.

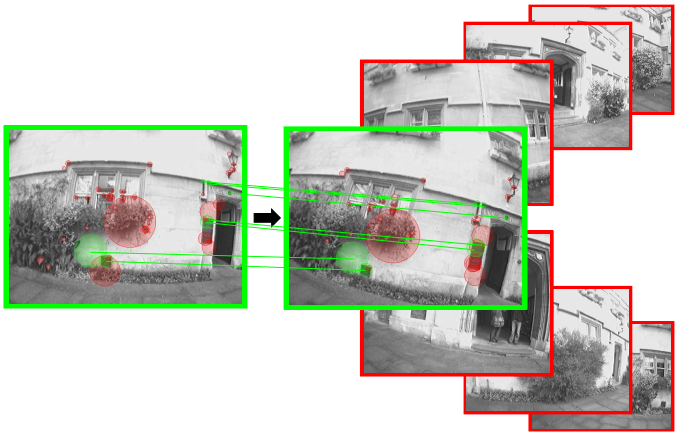
On each image, the algorithm takes on average 283ms to run. Much of this time (73ms) is taken up by SURF feature detection. This method relies on this descriptor which is richer



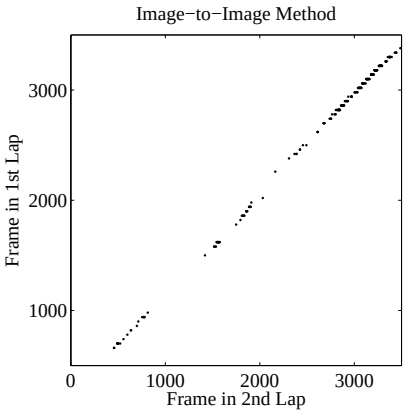
(a) **Map–To–Map:** Loop closure detected using the method of Clemente *et al.* [2]. The system finds a set of features consistent in both geometry and appearance between the first and last submaps. It is only successful if the SLAM system has initialised common features in the two submaps.



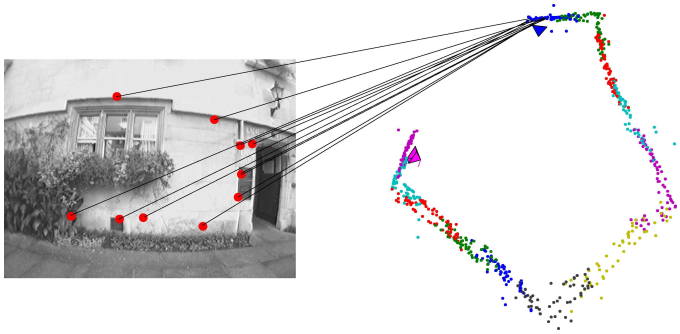
(b) **Map–to–Map Reliability:** Matching was attempted between every pair of non-consecutive submaps. Shown here are the eight false positives sets with five correspondences. The true positive was not found in this run since only two features were shared between the first and final submaps (See Fig. 3).



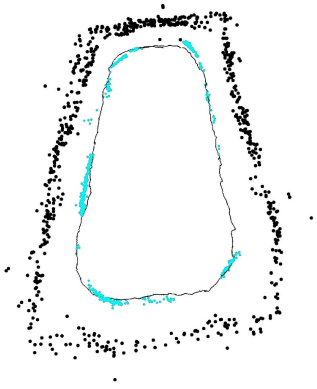
(c) **Image–To–Image:** Loop closure detected using the method of Cummins *et al.* [4]. The system detects visual words in each image and the cooccurence of these words is used to calculate the probability of loop closure. The system finds a high probability that the most recent image matches one seen earlier in the sequence. Visual words are detected in the two images are indicated in green if they match in the other image. Note that interest point geometry is not considered.




(d) **Image–To–Image Reliability:** Correspondences were found between every frame in a second lap and every 40th frame in the first lap. A threshold was chosen to remove all false positives. At this threshold, the system was successful in 8% of attempts. To see the effect of the threshold on performance see Fig. 5. Gaps are in regions of the world with lots of foliage (where the image–to–map method also struggles).



(e) **Image–To–Map:** Loop closure detected using the method of Williams *et al.* [13]. While tracking in the last submap, the system finds a camera pose consistent with the features in the first submap. The common trajectory is used to determine the relative rotation translation and scale needed to align the submaps.



(f) **Image–to–Map Reliability:** Relocalisation was attempted on every frame of a second lap. The light dots show the camera pose recovered relative to the map and trajectory created on the first lap (black). This indicates that loop close would be successful for these frames. Successful in 20% of frames. No false positives.

Fig. 4. The results of experiments on all three loop closing methods. The left column shows a successful loop closure for each method. The right column shows tests on the reliability of each method.

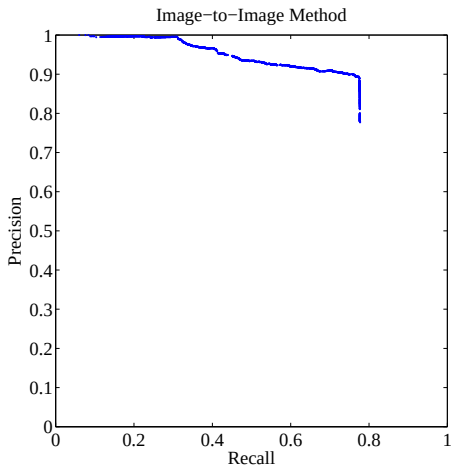



Fig. 5. This precision–recall curve for the image–to–image method [4] shows the algorithm performs well. Quite a high number of true loop closure are detected with few false positives.

总体速度比帧速率慢，但是，循环关闭算法不需要在每一帧上运行

yet slower than the randomised fern classifier. The overall speed is slower than the framerate, however, the loop closing algorithm does not need to be run on every frame.

This method was also tested on the benchmark dataset for this workshop and successfully detected the loop closures (Fig. 7 and 8).

### C. *Image–to–Map Matching: Williams* et al.

At every frame, there is usually enough remaining time after tracking to attempt relocalisation in one other submap. The system cycles through submaps until a relocalisation is successful, indicating a loop closure. For the university courtyard sequence, the system successfully detected the loop closure as the features in the original map came back into view (Fig. 4(e)). Note that for this method, no common features are needed between submaps as they are for the map–to–map method.

The reliability of this loop closure method was tested using the same 'kidnapped robot' situation we used to test the image-to-image method. The system was allowed to continue searching for loop closures as the camera continued around the courtyard for a second lap. For the test, the system attempts relocalisation in every submap for every frame. The results of this test can be seen in Fig. 4(f).

The method takes 10-15ms to find potential matches to map features in each image. The remaining time is used to run RANSAC on the matches to determine the pose. This is usually found within a few milliseconds if a valid pose exists for those matches. This is fast enough to allow the algorithm to run on a single submap after the system has finished tracking in each frame.

## V. DISCUSSION

We have tested three quite different approaches to detecting loop closure for monocular SLAM systems. We found the map–to–map matching technique of Clemente *et al.* to be unsuitable for these sparse maps since it relies on common features being initialised by the system. The image–to–image matching technique of Cummins *et al.* works well since it can be tuned to remove all false positive while still detecting 8% of true positive for this sequence but the image–to–map matching technique of Williams *et al.* was able to achieve a higher true positive rate of 20%. The image–to–map is able to prune more false positives than the image–to–image method by making use of the geometry information of the features detected in the image (see Fig. 6). In general, it is best to take as much information as is feasible into account when detecting loop closures. In the next section, we discuss recent work to extend the image–to–map method to allow more of the image information to be used.

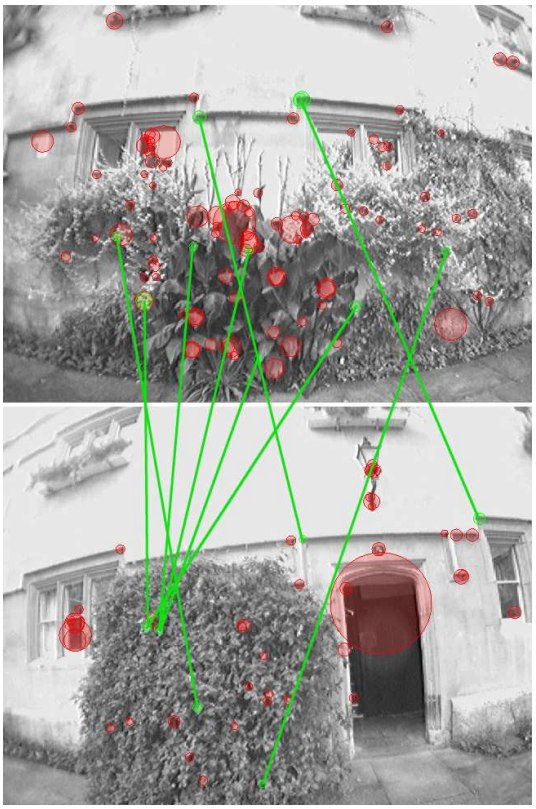


Fig. 6. Image-to-image method: False positive with matching probability of 99.9935%. The detected visual words are indicated in each image in green if they match the other image. This false positive could easily be discarded if the geometric information were known for the detected visual words.

通过利用图像中检测到的特征的几何信息，image - to - map方法比image - to - image方法能够剔除更多的误报(见图6)

## VI. EXTENSION TO THE IMAGE–TO–MAP METHOD

In the results presented so far, the image–to–map method used a separate randomised ferns classifier for each submap and had to cycle through submaps when attempting loop closure. We have recently been exploring a way of using a single classifier which can attempt loop closure with all submaps simultaneously. However, as the number of features in the map increases, the randomised ferns classifier returns a greater number of possible correspondences for the corner points in each image. RANSAC has to work harder to find a

set of true correspondences amongst the much larger number of combinations.

To guide RANSAC into favouring more likely correspondences, we look at the image context surrounding the features as well as their classification. This context is described by the presence of features from a standard vocabulary in the whole image in a method similar to the method of Cummins *et al.*. However, here we use a faster but less rich vocabulary from a second randomised ferns classifier.

Every time a map feature is observed by the SLAM system, the frequency of standard features from the vocabulary is noted. Later, for loop closure or relocalisation, RANSAC gives higher weight to correspondences where the current frequency of standard features in the image closely matches the distribution observed when that map feature was visible during tracking. The initial results for this method are promising but more work remains to be done to choose the best distance metric for measuring which features best match the current context.

## VII. Conclusion

我们发现地图对地图匹配技术不适合单眼SLAM，因为稀疏地图包含的信息太少，无法可靠地检测出真实的对应

We have tested three quite different approaches to detecting loop closure for monocular SLAM systems. Experiments were performed in a university courtyard using the Hierarchcal SLAM technique to build a sequence of submaps of the environment.

We found the map–to–map matching technique to be unsuitable for monocular SLAM because the sparse maps contain too little information to reliably detect true correspondences.

The image–to–image method was shown to work well in this sequence. However, the method is not complete if the relative pose between corresponding images is needed for correcting the metric map. The method would benefit from making some use of the relative positions of the detected visual words to remove some obvious false positives.

The image–to–map method works well and returned the highest number of true positives with no false positives. We predict even better performance can be achieved by taking more of the image into account as outlined in our proposed extension to the method.

## VIII. Acknowledgements

We gratefully acknowledge the financial support of the EPSRC (grant GR/T24685, EP/D037077, and a studentship to BW) and the Royal Society (International Joint Project).

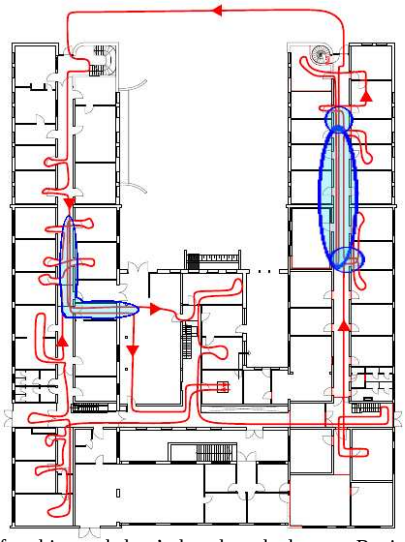


Fig. 7. Map for this workshop's benchmark dataset. Regions of potential loop closure where the robot faced the same direction are circled in blue. The image–to–image method [4] was able to detect loop closures in all four of these regions.

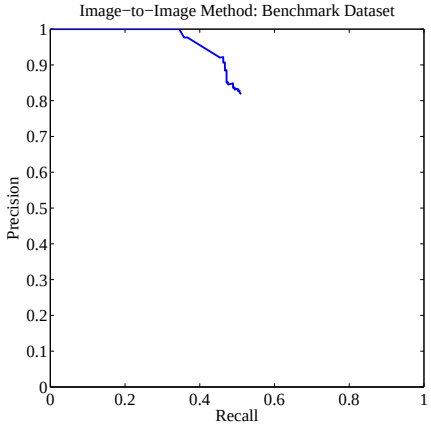



Fig. 8. Precision–recall curve for this workshop's benchmark dataset.

## References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Proc. European Conference on Computer Vision*, 2006.

[2] L. Clemente, A. Davison, I. Reid, J. Neira, and J. D. Tardós. Mapping large loops with a single hand-held camera. In *Robotics Science and Systems*, 2007.

[3] M. Cummins and P. Newman. Accelerated appearance-only SLAM. In *Proc. IEEE International Conference on Robotics and Automation*, 2008.

[4] M. Cummins and P. Newman. FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008.

[5] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proc. IEEE International Conference on Computer Vision*, 2003.

[6] C. Estrada, J. Neira, and J. D. Tardós. Hierarchical SLAM: Real-time accurate mapping of large environments. *Transactions on Robotics*, 1(4), 2005.

[7] M. A. Fischler and R. C. Bolles. RANdom SAmple Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[9] Vincent Lepetit and Pascal Fua. Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1465–1479, 2006.

[10] J. Neira, Tardós J. D., and J. A. Castellanos. Linear time vehicle relocation in SLAM. In *Proc. International Conference on Robotics and Automation*, 2003.

[11] J. Sivic and A. Zisserman. Video google: a text retrieval approach to object matching in videos. In *Proc. IEEE International Conference on Computer Vision*, 2003.

[12] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and Tardós J. D. An image–to–map loop closing method for monocular SLAM. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, 2008.

[13] B. Williams, G. Klein, and I. Reid. Real-time SLAM relocalisation. In *Proc. International Conference on Computer Vision*, 2007.