High Performance Online Loop Closure Detection for Topological Mapping

Zhaowei Shi¹, Jingyu Luo¹, Yunfeng Wang¹* and Jinfeng Liu²
¹School of Electronic Science and Technology, Xiamen University, Xiamen, China
²Guangdong Polytechnic of Science and Technology, Zhuhai, China
Email: yunfengwang@xmu.edu.cn

Abstract. In simultaneous localization and mapping (SLAM) system, loop closing is defined as the correct identification of a previously visited location. Loop closing is essential for the precise self-localisation of the robot; however, the performance of loop closure detection is seriously affected by dynamic objects and perceptual aliasing in the environment. In the traditional likelihood matching methods, the number of matching words and the difference between them are not considered. This paper proposes a method based on mixed similarity to calculate the similarity score, thereby improving the performance of closed-loop detection. Experiments are performed on datasets from dynamic environments and visual repetitive environments, and then this method can produce a higher recall rate with 100% accuracy compared to the latest methods.

1. Introduction
Simultaneous localization and mapping (SLAM) is an important technique for autonomous mobile robots navigation [1]. Loop closure detection within SLAM is required to create a consistent topological or metric map [2]. Robots can recognize previously visited places accurately and correctly when they are revisited by Loop closure detection.

This paper focuses on the application of appearance image based method to detect closed loop topology mapping. The bag-of-words (BoW) approach [3] is one of the most popular image-based methods, which consists in representing each image by visual words taken from a vocabulary. The visual words are commonly made from images local feature descriptors, such as SIFT, SURF, ORB or BRISK [4]. The vocabulary can be generated offline or on line. Probably the most well-known solution that generates the dictionary offline is fast appearance-based mapping (FAB-MAP) [5]. Labbe at al utilizes NNDR method to construct online a visual dictionary in real-time appearance-based mapping (RTAB-MAP) [6]. In order to solve the problem of increased memory and time requirements for large SLAM systems. [6] proposed a novel memory architecture to meet the growing demand for computing and memory. Recently Yin [7] established an end point and linear visual dictionary describing environmental texture information. In Ref. [8], the author uses the multi-index hash (MILD) method to calculate the image similarity instead of BOW.

The BoW approach utilizes image similarity to identify previously visited places. The similarity is evaluated by the number of matched word pairs between tow images [6], or the TF-IDF score [2, 9, 4]. When the most pixels represented by matched words are concentrated in small fields of two images, the number of matching words does not reflect the similarity of the two images well. Moreover, the evaluation method based on the number of matching words does not consider the difference of words. Especially if some words appear in all images, they do not play any roles for closed-loop detection.
TF-IDF considers the different roles of each word for closed-loop detection. However, if two images have few matched words with large weight, the false loop closure could be detected. By incorporating the number of matching words into TF-IDF, we propose a novel approach for images similarity evaluating to reduce the error rate of closed loop detection.

The main contributions made in this paper are as follows: (1) A mathematical formula for computing images similarity is developed by incorporating the number of matching words into TF-IDF, which could be applied for all BoW approaches to reduce the error rate of closed loop detection. (2) An online loop closure detection algorithm based on topology mapping is proposed. In addition, the algorithm also incorporates other classic concepts, such as K-D tree, Bayesian framework and memory management methods.

2. Online Appearance-Based Loop Closure Detection

In this section, we explain our proposed appearance-based loop closure detection method. This method reduces the error rate of closed loop detection by merging the number of matching words into TF-IDF, mechanical energy image similarity evaluation.

2.1. The Mixed Image Similarity

The similarity of two images is an important basis for judging whether a closed loop occurs. Therefore, it is very important that how to compute similarity measure.

There are two kinds of compute similarity measure: a based on matching word pairs method and tf-idf method. The based on matching word pairs method is shown in equation (1):

\[
\text{sim}(Z_t, Z_i) = \begin{cases} 
\frac{N_{\text{pair}}}{N_{Z_t}}, & \text{if } N_{Z_t} \geq N_{Z_i} \\
\frac{N_{\text{pair}}}{N_{Z_i}}, & \text{if } N_{Z_t} \leq N_{Z_i} 
\end{cases}
\]  

(1)

N_{\text{pair}} represents the number of visual word that are common between the image \(I_t\) and the image \(I_i\). \(N_{Z_t}\) is the number of the all words in the image \(I_t\). At the same time, \(N_{Z_i}\) is the number of the all words in the image \(I_i\). \(\text{sim}(Z_t, Z_i)\) is the similarity between image \(I_t\) and image \(I_i\). Each visual word has a weight TFIDF in tf-idf method. It can be calculated by equation (2):

\[
\text{TFIDF} = \frac{N_{a_i}}{N_{Z_t}} \log \frac{N}{N_{a_i}}
\]  

(2)

The number of occurrences of the visual word \(W_a\) in image \(I_t\) is \(N_{a_i}\). \(N_{Z_t}\) is the number of the all words in the image \(I_t\). \(N\) is the total number of nodes in the map at time \(t\). \(N_{a}\) is the number of images containing the word \(W_a\). The similarity between image \(I_t\) and image \(I_i\) is the sum of the weight of word that both in image \(I_t\) and image \(I_i\). It is calculated by the following equations (3):

\[
\text{sim}(Z_t, Z_i) = \sum \frac{N_{a_i}}{N_{Z_t}} \log \frac{N}{N_{a_i}}
\]  

(3)

where \(a\) is the visual word that are common between the image \(I_t\) and the image \(I_i\).

Both of the above methods have disadvantages. In the method of matching word pairs, when the matching words are concentrated in a small area, or the matching words frequently appear in the scene, the number of matching words does not reflect the similarity of the two images well, which may lead to a false closed loop as shown in figure 1a. The tf-idf method sets a weight for each word. A word that often appear in scenes has small weight, while the word appears in few images has large weight. If two images have few matched words with large weight, the false loop closure could be detected as shown in figure 1b.
A new method based on mixed similarity is proposed to solve the shortcomings of the above two methods. This method considers both the number of common words and the degree of discrimination of different words. We normalize equation (1) to get (4) to reduce the influence of the difference in the total number of words between two frames.

\[
\text{sim}(Z_t, Z_i) = \frac{2 \times N_{\text{pair}}}{N_{Z_i} + N_{Z_t}}
\]

Then we calculate the sum of all word weights H for each frame of image according to equation (5). If the ratio of the H of the two frames is large, it means that the two images is very different. We combination equation (4) and (5), based on the mixed similarity method, the similarity of two images can be obtained by equation (6).

\[
H = \Sigma_{i=1}^{N} \log \frac{\Sigma_{i=1}^{N} \log \frac{N}{N_{a}}}{N_{a}}
\]

\[
\text{sim}(Z_t, Z_i) = \begin{cases} 
2 \times N_{\text{pair}} & \Sigma_{i=1}^{N} \log \frac{N}{N_{a}} , \text{ if } H_{Z_t} \geq H_{Z_i} \\
N_{Z_i} + N_{Z_t} & \Sigma_{i=1}^{N} \log \frac{N}{N_{a}} , \text{ if } H_{Z_t} < H_{Z_i}
\end{cases}
\]

2.2. Loop Detection

Figure 2 illustrates the algorithm. When a new image is acquired, SURF features are extracted to create an image signature. If few SURF features are extracted, they will not be processed for closed-loop detection. For good signatures, we calculate the nearest neighbor distance (ND) of each feature to the word already in the vocabulary, and then sort, and use the Nearest Neighbor European Distance Ratio (NNDR). To update the visual dictionary. At the same time, a new location \( L_t \) is created by the sensory memory (SM) and sets the \( L_t \) the threshold \( T_{\text{sm}} \). At this time, the foremost position in STM is transferred to the working memory (WM). According to Bayesian filter, the closure hypotheses are obtained by estimating the probability that the current location matches one of an already visited location stored in WM. If a loop is closed, the neighbors of the loop position in long term memory (LTM) are transferred to WM while updating the visual dictionary. If the loop closure
time is detected to be greater than the threshold $T_{\text{sim}}$, the position with the smallest weight in WM is transferred to the LTM and the word created by it should be deleted in the dictionary.

![Flowchart of loop closure detection algorithm.](image)

**Figure 2.** Flowchart of loop closure detection algorithm.

### 3. Experimental Results

The datasets used for the experiment are the used by FAB-MAP [5] and IAB-MAP [2]. In this experiment, a single image for each location is composed of the NC and CiC datasets stitched together for each pair of stereo images. Each threshold setting in this experiment is shown in Table 1. Threshold $T_{\text{ND1}}$ and $T_{\text{ND2}}$ are obtained by counting a large number of SURF algorithm matching results. The rest is the same as in RTAB-MAP [6].

| $T_{\text{sim}}$ | $T_{\text{loop}}$ | $T_{\text{NNDR}}$ | $T_{\text{STM}}$ | $T_{\text{ND1}}$ | $T_{\text{ND2}}$ |
|-----------------|------------------|------------------|-----------------|-----------------|-----------------|
| 0.2             | 0.11             | 0.8              | 30              | 0.001           | 0.08            |

The maximum value of the posterior probability of the loop closure between the current frame image and the historical frame image is shown in Figure 3.

Figure 3a is the detection results of RTAB-MAP [6]; Figure 3b is the result of the method in this paper. The method has more 90% posterior probability from 690th image and detect the loop closure. While RTAB-MAP [6] needs to detect the loop closure from 730th image. It can detect more loop closure because of it improves the accuracy of the image description. An example of accepting a loop closure is shown in figure 4. It can be seen that this method has certain robustness to the dynamic environment.
Figure 3. Comparison of loop closure detection posterior probability.

The precision-recall rate curve of our method is shown in figure 5. The recall performance of our method is compared with other methods at 100% precision on same data sets. The results are shown in table 2. The symbol ‘*’ means the recall performances were not available. As can be seen from the table, the method proposed this paper has a better recall ratio in most data sets than other methods. What calls for special attention is that most methods have significantly reduced performance on dynamic datasets, but the method in this paper has high recall rate in all datasets. This proves that our method is more robust.

Figure 4. A loop closure.

Figure 5. Precision-recall rate curve.
Table 2. Recall (%) at 100% precision.

| Datasets  | NC  | CiC | L6I | L60 |
|-----------|-----|-----|-----|-----|
| IAB-MAP [2] | *   | *   | 80  | 71  |
| RTAB-MAP [6] | 89  | 81  | 98  | 95  |
| FAB-MAP [5] | 46  | 37  | *   | *   |
| DoWP [10]   | 77  | 86  | 94  | 92  |
| HOVD [11]   | 74.6 | 52.36 | 42.32 | 49.55 |
| PAL [7]     | 57.14 | 67.59 | 80.45 | 69.65 |
| MILD [8]    | 87  | 83  | 94.5 | 93.4 |
| OUR         | 94.2 | 88.7 | 98.6 | 91.9 |

4. Conclusion
As an important part of SLAM, loop closure detection is crucial to reduce the accumulated errors of mobile robots and achieve the consistency of map. The challenge is how to obtain a higher recall performance at 100 precision under reasonable computing requirements. The similarity is evaluated based on the method of mixed similarity that improved the performance of loop closure detection. Compare with most of the existing methods, the proposed method has better recall performance on datasets from dynamic environments and visually repetitive environment demonstrated.

Acknowledgments
The project was supported by the Scientific and Technological Program of Xiamen city, China (No.3502Z20193013), and the Science and Technology Program of Guangzhou, China (No.201804010253).

References
[1] Filliat D 2007 A visual bag of words method for interactive qualitative localization and mapping *IEEE International Conference on Robotics and Automation* (IEEE) pp 3921-3926.
[2] Angeli A, Filliat D, Doncieux S and Meyer J A 2008 Fast and incremental method for loop-closure detection using bags of visual words *IEEE Trans. Robot* 24 1027-1037.
[3] Botterill T, Mills S and Green R 2011 Bag-of-words-driven, single-camera simultaneous localization and mapping *Journal of Field Robotics* 28 204-226.
[4] Mur-Artal R and Tardós J D 2014 Fast relocalisation and loop closing in keyframe-based SLAM *IEEE International Conference on Robotics and Automation* pp 846-853.
[5] Cummins M and Newman P 2008 FAB-MAP: Probabilistic localization and mapping in the space of appearance *Int. J. Robot. Res.* 27 647-765.
[6] Labbe M and Michaud F 2013 Appearance-based loop closure detection for online large-scale and long-term operation *IEEE Transactions on Robotics* 29 734-745.
[7] Yin J, Li D and He G 2018 Mobile robot loop closure detection using endpoint and line feature visual dictionary *IEEE International Conference on Robotics & Automation Engineering*.
[8] Han L and Fang L 2017 MILD: Multi-index hashing for appearance based loop closure detection *IEEE International Conference on Multimedia & Expo*.
[9] Galvez-López D and Tardos J D 2012 Bags of binary words for fast place recognition in image sequences *IEEE Transactions on Robotics* 28 1188-1197.
[10] Morioka N and Satoh S 2010 Building compact local pairwise codebook with joint feature space clustering computer vision *ECCV 2010* (Berlin Heidelberg: Springer).
[11] Bampis L, Amanatiadis A and Gasteratos A 2017 High order visual words for structure aware and viewpoint-invariant loop closure detection *IEEE/RSJ International Conference on Intelligent Robots & Systems*. 