Visual Loop Closure Detection with Scene Mutual Information for Mobile Robot

Ning Liu
Jinan University/ College of Information Science and Technology, Guangzhou, China
Email: tliuning@jnu.edu.cn

Junjun Wu
Guangdong Food and Drug Vocational College/ Department of Software, Guangzhou, China
Email: junjun-wu@hotmail.com (Corresponding author)

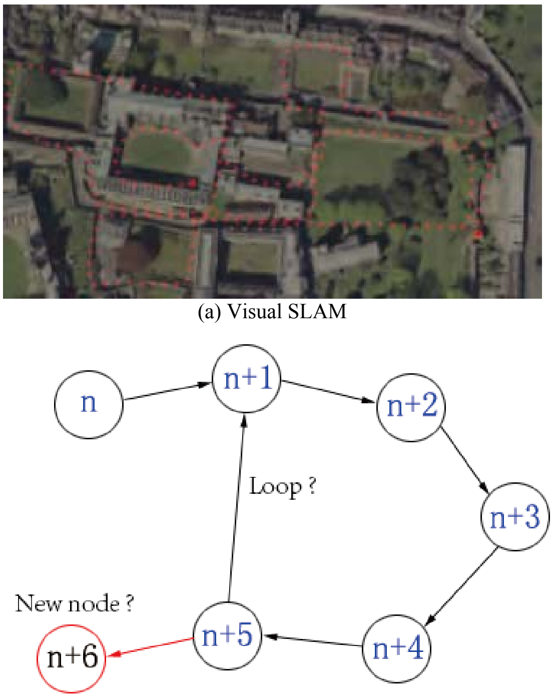
***Abstract*—In this paper, an efficient approach is proposed for loop-closure detection in robot visual SLAM. The method uses mutual information to measure similarity between current view and key frames in an appearance map, and evaluates candidate loop-closure locations in particle filter framework. Specially, the implementation of particle filter is accelerated through updating a set of weight vector of particles, and three threshold indicators are used to select loop-closure candidates and verify loop-closure location. The comparative experiments on a popular dataset verify the high efficiency of our method which is more simple and accurate than the popular bag-of-words (BoW) for loop-closure detection.**

***Index Terms*—Mobile robot, visual SLAM, loop-closure detection, mutual information, particle filter**

I. Introduction

Visual SLAM (simultaneous localization and mapping) [1] refers to a process of mobile robot that iteratively performs visual self-locating and mapping starting from one unknown location in an unknown environment through the visual sensor. This technology can be applied to the robot navigation in large-scale unstructured environment (shown in Fig. 1 a), which is currently a hot research topic [2-5]. During the process of visual SLAM, the mobile robot keeps judging whether it is in a position that has been visited previously, so as to ensure the correctness of the environment topological map, i.e., the visual loop-closure detection [6]. The core of loop-closure detection is that the mobile robot determines the spatial relevance of locations at different time according to robot observed images. If loop-closure occurs, the robot will reconstruct the environmental map, otherwise, it will add the current detected key frame to the environmental map as a new visual node [7], and shown in Fig. 1(b). Visual loop-closure detection is a key step to correct environmental map for robot visual SLAM. Therefore, the issue of loop-closure detection is considered to be a special issue in visual SLAM by researchers [9-12].

Compared with other sensors (e.g., sonar and laser), the visual sensor can capture richer environmental information (e.g., color, veins and shape). However, it is also challenged due to disturbances of observed noise. For example, the scale, view and other factors of the images observed by the robot can affect the precision of the loop-closure detection. In addition, the large amounts of visual data when robot moves along a long route decrease the efficiency of loop-closure detection.



(a) Visual SLAM

(b) Loop-closure detection

Figure .1 Visual SLAM and loop closure detection in large-scale environment

Visual loop-closure detection aims to determine the probability of the occurrence of loop closure between the current location of the robot and the visited previously locations in the environment map. Distinguished from the traditional image retrieval problem, loop-closure detection does not allow manual intervention and cannot depend on pre-training classifier that does not handling well the further images observed by robot. Specially, the detection speed must satisfy the process of SLAM. Thus, robot visual SLAM not only requires a high precision and

doi:10.4304/jcp.9.11.2642-2647

recall of loop-closure detection, but also keeps high real time.

Currently, visual loop-closure detection is mainly based on the similarity of image descriptors generated by the features in images [8]. The most popular approach is BoW (Bags of words) [9-12]. For example, in order to speed up loop-closure detection, the visual dictionary tree is utilized to index the high-dimensional image descriptor [9], or bag of binary words is proposed [12]. BoW method can be summarized into three steps: (1) Extracting the invariant local features from images such as SIFT [13], SURF [14], PCA-SIFT [15], and the BRIEF feature with binary descriptor [16]. (2) Utilizing clustering technique (e.g., K-mean clustering) to cluster the local features, and then generate the BoW by considering the center of each cluster as each word. (3) Estimating the loop-closure probability in the topological map based on the indexing technique such as the visual vocabulary tree with pyramid TF-IDF scoring math scheme for loop-closure detection [17].

BoW approach achieves a good precision, however, it has to extract a large amount of local features, maintain a large-scale visual dictionary, and the detection precision is greatly influenced by the number of visual words. In addition, BoW method suffers from perceptual aliasing (the similar scenes in different locations may be considered mistakenly to be the same one) due to the vector quantization of the raw visual feature descriptor.

To address the mentioned above problem, we proposed a new approach using mutual information (MI) between images for the loop-closure detection. Our method does not have to extract the local feature descriptors in images, and without spending much time to construct visual dictionary. Furthermore, the proposed approach exhibits a decent online performance, and can meet the requirement of precision and efficiency for visual SLAM.

II. Visual Loop Closure Detection Method

A. Visual SLAM and Loop-closure Detection

SLAM is a discrete-time nonlinear dynamic process, where the robot estimates continuously the loop-closure probability when adding a new observation to the current environment topological map. This process can be described by Bayesian filter, as is shown in (1).

$$p(x_{1:t}, l_{1:m} \mid z_{1:t}, u_{0:t-1}) = p(x_{1:t} \mid z_{1:t}, u_{0:t-1})\, p(l_{1:m} \mid x_{1:t}, z_{1:t}) \\ = p(x_{1:t} \mid z_{1:t}, u_{0:t-1}) \prod_{i=1}^{m} p(l_i \mid x_{1:t}, z_{1:t}) \quad (1)$$

Where $x_{1:t}$ denotes the location of the robot at time t; $l_{1:m}$ denotes the map size m; z_1 denotes the observation at time t; $u_{0:t-1}$ denotes the movement control variable at time t-1. The performance of loop-closure detection is mainly affected by the two variables $x_{1:t}$ and $l_{1:m}$.

B. Similarity Measurement

Robot moves from the current location to the next, the observed scene at a new location is considered to be a candidate keyframe by implementing keyframe detection method. The similarity between the candidate keyframe and the visual nodes are calculated for estimating the probability of the occurrence of loop closure. In this paper, MI (mutual information) in information theory is firstly introduced to measure similarity for loop-closure detection. The entropy h(x) of an image describes the average information of this image, it satisfies Equation (2). MI of image pairs is calculated by the image entropy h(x) and the joint information entropy h(x, y) , as shown in (3) which satisfies (4).

$$h(\mathrm{x}) = -\sum_{x} p(\mathrm{x}) \log_2[\mathrm{p(x)}] \quad (2)$$

$$h(\mathrm{x,y}) = -\sum_{x}\sum_{y} p_{xy} \log[\mathrm{p}_{xy}(\mathrm{x,y})] \quad (3)$$

$$MI(\mathrm{x,y}) == \sum_{x}\sum_{y} p_{xy} \log_2 \frac{[\mathrm{p}_{xy}(\mathrm{x,y})]}{p_x(\mathrm{x})\,\mathrm{p}_y(\mathrm{y})} \quad (4)$$

Where $p(x)$ and $p(y)$ are the histogram of image x and image y respectively, $p(x; y)$ is the joint histogram of the two images.

To improving the efficiency of similarity calculation, the proposed method firstly process the images by down-sampling, and binarizes the image using OTSU, and then calculates the similarity according to MI, as described by Equation (5).

$$S(I_x, I_y) = MI(I_x^{d-b}, I_y^{d-b}) \quad (5)$$

Where, $\mathrm{I_x}$ and $\mathrm{I_y}$ represent the images observed by the robot at two different locations; I_x^{d-b} and I_y^{d-b} denote the corresponding down-sampled binarized images.

C. Loop-closure Candidate Probability

Our method runs particle filter for evaluating loop-closure candidates. In this section, particle set and its initial distribution are defined, and then details candidate loop-closure selection with the proposed method, especially, how to re-sampling with boosted weight of particles.

During the continuous loop-closure detection, the particle set is used to dynamically estimate the probability of loop-closure location. The convergence of the particle set at certain location will be labeled in a binary candidate loop-closure vector ψ_t^n, which is defined in Equation (6).

$$\psi_t^n = \{\mathrm{x}_t^{[1]}, \mathrm{x}_t^{[2]}, \ldots, \mathrm{x}_t^{[i]}, \ldots, \mathrm{x}_t^{[n]}\} \quad (6)$$

Where n denotes the total number of the particles and $\mathrm{x}_t^{[i]}$ represents the status of the i-th particle at time t.

The initial distribution of the particles is uniform, which satisfies Equation (7).

$$\psi_0^n = \{\mathrm{x}_0^{[i]} \mid \mathrm{x}_0^{[i]} = \mathrm{round}(\frac{m*i}{n})\} \quad (7)$$

Where $\mathrm{x}_0^{[i]}$ denotes the topological location of the i-th particle at initial time; m denotes the size of current environment map; n denotes the size of the particle set.

With the movement of the robot, the particles move from the initial distribution. The weights of the particles are recalculated based on the similarity *S* between the previous environment scene and the new observed environment scene. The initial weights vector of particle V_m^n reflects the probability of current loop closure in the environment map with size of m, as shown in (8).

$$\begin{cases} V_m^n = [v_m^{[i]}] \\ v_m^{[i]} = S(I_i, I_{m+1}) \end{cases} \quad 0<i<=n \tag{8}$$

Where $v_m^{[i]}$ denotes the initial weight of the i-th particle at the m-th loop-closure detection; I_i is observed image described by the i-th particle, and I_{m+1} is newly observed image.

Robot detects loop-closure location before updating the topological map. The corresponding particle weight vector V_m^n is generated shown in Fig. 2.

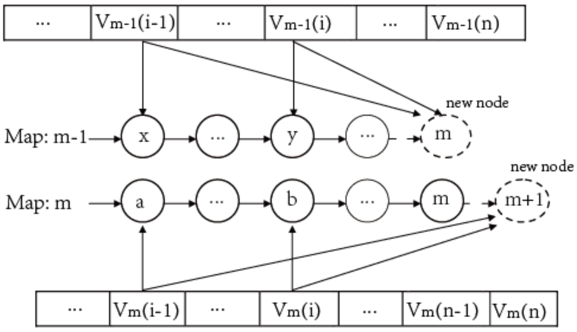


Figure. 2 The initial weight vector sequence of particles

Since the robot observes a continuous sequence of images, the neighboring images of two similar scenes are also very likely to be similar, thus the subsequences of the scenes tend to have a high similarity if a loop closure occurs, shown in Fig. 3.

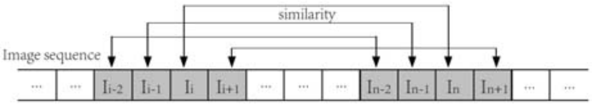


Figure. 3 The sub-sequence similarity between loop scenes

As is shown in Fig. 2, after the (m-1)-th loop-closure detection, if one new node is added to the map, node x and node a are adjacent, and node y and node b are adjacent; otherwise, node x and node y locate at the same position of node a and node b respectively. Therefore, according to the similarity between two subsequences scenes shown in Fig. 3, the initial weight vector V_{m-1}^n at the (m-1)-th detection can be used to enhance the initial weight vector V_m^n at the m-th detection. It is beneficial to the convergence of particles and discriminate loop-closure location. The variable $\widehat{V_m^n}$ satisfies Equation (9).

$$\begin{cases} \widehat{V_m^n} = [\widehat{v_m^{[i]}}] \\ \widehat{v_m^{[i]}} = v_m^{[i]} + v_{m-1}^{[i]} \end{cases} \quad 0<i<=n \tag{9}$$

Where $v_m^{[i]}$ and $\widehat{v_m^{[i]}}$ denote the initial weight of the i-th particle and its boosted weight at the m-th loop-closure detection, respectively.

Before the particles are re-sampled, their weights $\widehat{V_m^n}$ are normalized to [0, 1]. The new weight $\overline{V_m^n(\mathrm{i})}$ is calculated according to Equation (10).

$$\overline{V_m^n(\mathrm{i})} = \frac{\widehat{V_m^n(\mathrm{i})} - \min(\widehat{V_m^n})}{\max(\widehat{V_m^n}) - \min(\widehat{V_m^n})}, i \in [1, \mathrm{n}] \tag{10}$$

The re-sampling of particle $x_t^{[i]}$ is performed with the threshold T_1 of the normalized weight vector $\overline{V_m^n(\mathrm{i})}$ for overcoming particle degradation. A new particle set $\{x_t^{[i]}\}^q$ is formed, where q denotes the set of particles whose weight is larger than T_1. With the movement of the robot, the distribution of the newly formed particle set Q_t gradually converges to the loop-closure location for the current observed image. The location where there is particle is marked as 1, otherwise 0, thus generating the binary vector L_c for describing the loop-closure candidates of the current observation.

D. Loop-closure Verification

Based on the probability of loop-closure location discussed in the previous subsection, the candidate loop-closure set L_c is further converged by using T_2 and T_3 for loop-closure verification. Here, T_2 is the threshold with respect to the distance of image ID in image sequence; T_3 is the threshold with respect to the rate of particle. The loop-closure convergence result is described by the binary loop-closure vector L_m as is shown in Equation (11).

$$L_m(\mathrm{i}) = \begin{cases} L_c(\mathrm{i}) & if\ d > T_2, p > T_3 \\ 0 & else \end{cases} \tag{11}$$

Where i denotes the key frame representing the i-th topological node in the environment map.

TABLE I
THE BINARY LOOP VECTOR SEQUENCE

Node	i−1	i	···	m−1	m
…	···	···	···		
m-1	$\overline{L_{m-1}(\mathrm{i}-1)}$	$\overline{L_{m-1}(\mathrm{i})}$	···	$\overline{L_{m-1}(m-1)}$	
m	$L_m(\mathrm{i}-1)$	$L_m(\mathrm{i})$		$L_m(\mathrm{m}-1)$	$L_m(\mathrm{m})$

Subsequently, the proposed method further verifies loop-closure location by analyzing the sub-sequence similarity of loop-closure scenes (shown Fig. 3). Finally, the result of loop-closure detection is described by $\widetilde{L_m}$, which satisfies Equation (12).

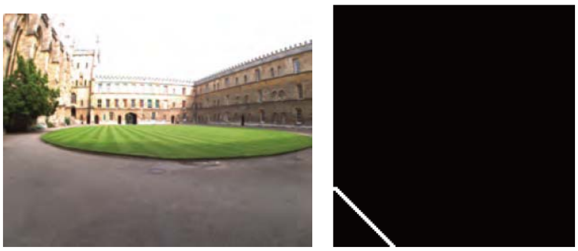
$$\overline{L_m(\mathrm{i})} = \begin{cases} L_m(\mathrm{i}) & if\ \lambda = 1 \\ 0 & else \end{cases} \tag{12}$$

Where $\lambda = \overline{L_{m-1}(\mathrm{i}-1)} \,|\, \overline{L_{m-1}(\mathrm{i})} \,|\, L_m(\mathrm{i}-1)$, variables are shown in Table I.

III. Experimental Validation

A. Experimental Dataset

We ran our experiments on the popular New College dataset introduced in FAB-MAP [18] from the robotics research group of Oxford University, which includes pairs of image sequences at the resolution of 640x480, GPS coordination information, and ground truth, shown in Fig.4. In a typical campus environment, the mobile robot uses two cameras in both sides to continuously observe images with a step size of 1.5m. Meanwhile, the GPS module records current geographic coordinate information so as to track the trajectory of the robot, and then label the loop-closure position manually to generate ground truth. In the experiment, each image is considered to be a new topological node in the experiment.

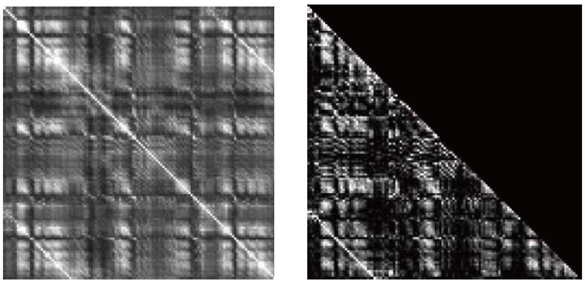


(a) Experimental scene (b) Ground truth

Figure. 4 Experimental environment and dataset

B. Similarity Measurement Efficiency Evaluation

In this section, the accuracy of similarity measurement was evaluated by calculating MI of images. Fig. 5(a) and Fig. 5(b) show the similarity matrix and the normalized similarity matrix for down-sampled and binarized images respectively.



(a) Similarity matrix (b) Normalized similarity matrix

Figure.5 Similarity matrix of 250 images

As is shown in Fig. 5(a), some regions where white lines appears clearly, which indicates the probability of occurrence of loop closure is high at the corresponding location. Considering the fact that the diagonal represents the self-similarity of each image, and that the similarity matrix is symmetric, only the lower (or upper) triangular matrix needs to be analyzed.

It can be seen from Fig. 5 (b), compared with similarity matrix M1, the white line in the lower left corner of similarity M2 is more consistent and clear. This phenomenon indicates that, after processing the images with down-sampling and binarization, the discrimination is increased, which benefit the subsequent process of loop-closure detection.

Since the dataset is generated from the pairs of image sequences acquired by the left and right cameras and each location corresponds to two images, only the images from the left (or right) camera are needed to estimate the loop-closure probability in the experiment. Here, take the observed images at location 99 and 249 for example. The specific analysis is as follow.

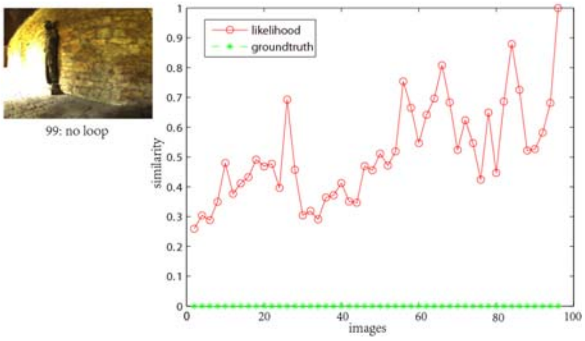


Figure. 6 The relationship between the probability of loop Q_{99} when scene is 99 and ground truth

When the newly 99th image is observed by the robot, the loop-closure probability Q_{99} at each visited previously location, is estimated by our approach. The number of images observed previously by the left camera is 49. The 50th row of matrix M2 reflects the loop-closure probability. It can be seen from the probability curve in Fig. 6. The similarity between the 98th and 99th images is the highest, but obviously, it cannot be considered as a loop-closure since they are neighboring images. With the movement of the robot, the observed scene becomes more and more similar to scene 99, and no loop-closure occurs that is fully consistent with ground true in our experimental dataset.

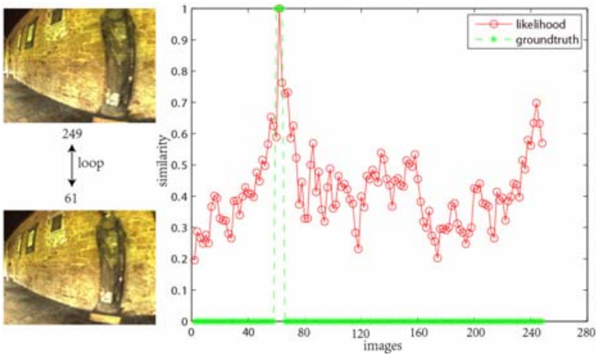


Figure. 7 The relationship between the probability of loop Q_{249} when scene is 249 and ground truth

As is shown in Fig. 7, when the robot moves to the 249th location, the loop-closure probability between the new observed scene and the previously observed scenes is calculated, which is denoted as Q_{249}. It can be seen from the probability curve that the probability of loop-closure is fully consistent with ground truth, i.e., one loop-closure occurs at scene 61.

C. Performance Evaluation and Discussion

We analysis our method with different parameters, and then compared it with BoW method, which is a popular approach for loop-closure detection.

Robot SLAM has a relative high requirement in terms of precision and recall for loop-closure detection. Therefore, in the experiment, F_1 is used to evaluate the performance of loop-closure detection, which satisfies Equation 13.

$$F_1 = \frac{2PR}{P+R} \quad (13)$$

Our experiment implements the re-sampling of particles with the particle weight threshold T_1, and the parameters T_2 (the absolute difference of image ID) and T_3 (the rate of particle at one location) are utilized to select loop-closure candidates. The curves describe the relationship of T_1 and F_1, and that of T_2 and F_1 in Fig. 8 and Fig. 9 respectively. In addition, since the particles tend to accumulate at the position where the loop-closure occurs, it is assumed that the candidate environment scene is labeled as loop-closure only when the particle ratio is greater than 90%, i.e., T_3 is 0.9.

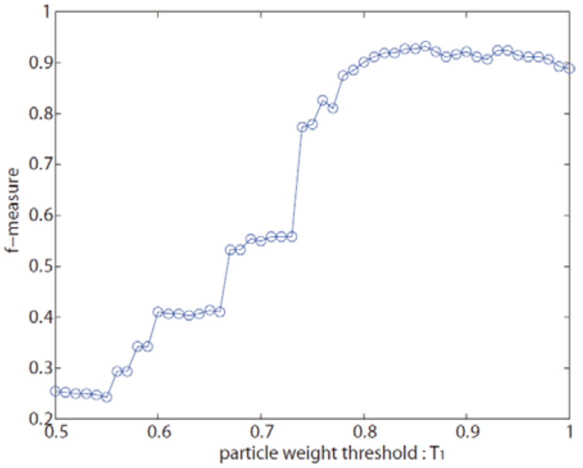


Figure. 8 The relationship between particle weight $\overline{V_m^n}(i)$ and F_1

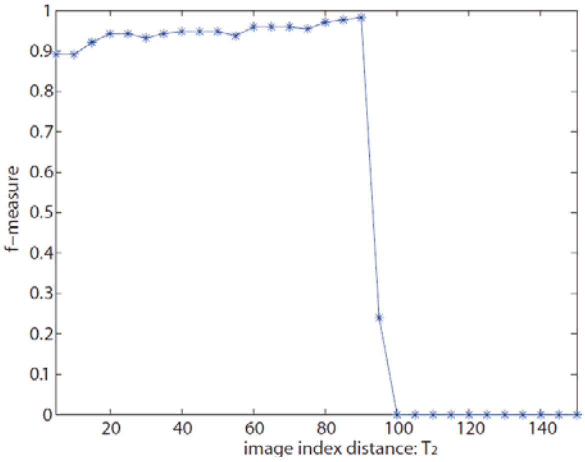


Figure. 9 The relationship between the distance of image index and F_1

Fig. 8 shows that the curve of F_1 rises when the threshold T_1 of particle weight increases. When T_1 locates in the range [0.8, 1.0], the fluctuation of curve F_1 is relatively small and exhibits a stable state basically. When T_1 and F_1 are selected to be 0.86 and 0.93, the precision of loop-closure detection and the recall achieve 92.2% and 94.3% respectively.

As shown in Fig. 9, when T_1 is 0.86, T_3 is 0.9, T_2 increases in the range [5, 90]. F_1 rises steadily at the first stage and dramatically drops afterwards. This trend is consistent with what happens in reality: when T_2 exceeds a certain level, there must be more loop-closure scenes that cannot be recalled; and with the further increase of T_2, the recall will continue to decrease until reaching 0. This is why F_1 dramatically drops when T_2>90. Hence, the threshold T_2 is selected to be 40 in the experiment. This approach exhibits a high performance where F_1=0.95, P=95.3% and R=94.3%.

D. Performance Comparison

The performance of the proposed approach (MI-VLD) and the classic approach (BOW-VLD) are compared and the discussion is as follows.

For MI-VLD, the parameter T_1, T_2, and T_3 are set to be 0.86, 40 and 0.9 respectively; for BOW-VLD, SURF is extracted to construct the image descriptor by running VLFeat [19]. The average number of the local features in each image is 369 and the number of vocabulary is set to be 700, 41000 and 7500 respectively.

TABLE II.
THE COMPARISON OF EFFECTIVENESS BETWEEN MI-VLD AND BOW-VLD

Method	F_1	P	R	Vocabulary
BOW-VLD	0.80	76.1%	86.3%	700
	0.89	82.1%	97.2%	4100
	0.92	95.1%	93.0%	7500
MI-VLD	0.95	95.3%	94.3%	-

As is shown in Table II, the performance of the BOW-based loop-closure detection approach is greatly affected by the number of vocabulary. For achieving high F_1, tens of thousands of vocabulary are needed to construct by clustering the large amount of local features. What is more, as SLAM proceeds, the size of the topological map constructed by the robot keeps increasing, hence more vocabulary are needed to guarantee the accuracy of loop-closure detection. Our method does not extract any local feature or without generating visual words, meanwhile, the higher precision and recall are achieved for visual loop-closure detection.

E. Efficiency Testing

We test the proposed method for loop-closure detection on experimental dataset on PC (Windows 64-bit OS, MATLAB 2010a, Intel i7-3770 CPU @3.40GHz (3.90GH, turbo), and Memory 8G). Due to down-sampling and binarization on the original images, the average time for calculating the similarity between each pair of images is reduced to 0.1ms, so that at least 10,000 image pairs can be processed within in 1s. In addition, because particle filter can accelerate the convergence speed, it is not necessary to use Greedy search for loop-closure detection. Therefore, the proposed approach can meet the performance requirement of visual SLAM in terms of real time and accuracy.

IV. Conclusion

In this paper, we present a new approach that applies firstly mutual information (MI) from information theory for loop-closure detection in robot visual SLAM. The proposed method avoids offline pre-training, and is independent of visual dictionary, but obtains high performance in terms of precision and recall due to high accuracy of MI for similarity calculation. Although our method offers only a linear complexity, it shows good efficiency through down-sampling and binarizing image, thus the proposed method is promising for real-time applications in visual SLAM. In the future, we plan to research semantic visual SLAM, and adopt the propose method, automatic semantic annotation to construct semantic topological map.

Acknowledgment

This work was supported in part by a grant from Science and Technology Program of Guangzhou (2014J4100109), China.

References

[1] Bailey T, Durrant W H. “Simultaneous localization and mapping (SLAM): part II”. *IEEE Robotics and Automation Magazine*, 2006, 13(3): 108-117.

[2] Aram k, Noppharit T, Sirinart, et al. “Online and incremental appearance-based SLAM in highly dynamic environments”. *The International Journal of Robotics Research*, 2011, 30(1): 33-55.

[3] Jianxian Cai, Lixin Li. “Autonomous navigation strategy in mobile robot”. *Journal of Computer*, 8(8): 2118-2125.

[4] Suhyeon Kim, Hyungrae Kim, Tae-Kyu Yang. “Increasing SLAM performance by integrating grid and topology map”. *Journal of Computer,* 2009, 4(7): 601-609.

[5] S. Hauke, J. Montiel, Andrew J D. “Visual SLAM: Why Filter?” *Image and Vision Computing*, 2012, 30(2): 65-77.

[6] B Williams, M Cummins, et al. A comparison of loop closingtechniques in monocular SLAM. *Robotics and Autonomous Systems*, 2009, 57(12): 1188-1197.

[7] H Zhang, B Li, D Yang. “Keyframe detection for appearance based visual SLAM”. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, October 2010: 2071-2076.

[8] J S Klippenstein, H Zhang. “Performance evaluation of feature extractors for visual SLAM”. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, USA, October 2009: 1574-1581.

[9] D Nister, H Stewenius. “Scalable recognition with a vocabulary tree”. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2006: 2161-2168.

[10] Kin L H, Paul N. “Detecting loop closure with scene sequences”. *International Journal of Computer Vision*, 2007,74(3): 261-286.

[11] A. Angeli, D Filliat, S Doncieux, et al. “A fast and incremental method for loop-closure detection using bags of visual words”. *IEEE Transactions on Robotics*, 2008, 24(5): 1027-1037.

[12] Galvez L D, Tardos J D. “Real-time loop detection with bags of binary words”. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011: 51-58.

[13] D Lowe. “Distinctive image features from scale-invariant keypoints”. *International journal of computer vision*, 2004, 60(2): 91-110.

[14] H Bay, T Tuytelaars, L Van Gool. “SURF: Speeded up robust features”. *The European Conference on Computer Vision*, 2006: 404-417.

[15] Y Ke, R Sukthankar. “PCA-SIFT: a more distinctive representation for local image descriptors”. *IEEE Conference on Computer Vision and Pattern Recognition*, 2004, Vol.2: II-506 - II-513.

[16] M Calonder, V Lepetit, C Strecha, “BRIEF: binary robust independent elementary features”. *The European Conference on Computer Vision,* 2010: 778-792.

[17] B. Li, D. Yang and L. Deng. “Visual vocabulary tree with pyramid TF-IDF scoring match scheme for loop closure detection”. *ACTA AUTOMATICA SINICA*, 2011, 37(6): 665-673.

[18] New College dataset: *http://www.robots.ox.ac.uk/mobile /IJRR_2008 Dataset/dataset.html.*

[19] VLFeat0.9.17, *http://www.vlfeat.org/.*

Ning Liu received Ph.D. from South China University of Technology in 2007. His research interest covers mobile robot, machine vision.

Junjun Wu received the master degree in 2008 and Ph.D. degree in 2013 respectively, in computer application from South China Normal University and in mechanical engineering automation from South China University of Technology. His research interest covers mobile robot visual navigation and simultaneous localization and mapping (SLAM).