

## Research Article

# Affine-Invariant Geometric Constraints-Based High Accuracy Simultaneous Localization and Mapping

**Gangchen Hua and Xu Tan**

*School of Software Engineering, Shenzhen Institute of Information Technology, Shenzhen, China*

Correspondence should be addressed to Xu Tan; [xutan@nudt.edu.cn](mailto:xutan@nudt.edu.cn)

Received 30 September 2016; Accepted 25 December 2016; Published 9 February 2017

Academic Editor: Hao Gao

Copyright © 2017 Gangchen Hua and Xu Tan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study we describe a new appearance-based loop-closure detection method for online incremental simultaneous localization and mapping (SLAM) using affine-invariant-based geometric constraints. Unlike other pure bag-of-words-based approaches, our proposed method uses geometric constraints as a supplement to improve accuracy. By establishing an affine-invariant hypothesis, the proposed method excludes incorrect visual words and calculates the dispersion of correctly matched visual words to improve the accuracy of the likelihood calculation. In addition, camera's intrinsic parameters and distortion coefficients are adequate for this method. 3D measuring is not necessary. We use the mechanism of Long-Term Memory and Working Memory (WM) to manage the memory. Only a limited size of the WM is used for loop-closure detection; therefore the proposed method is suitable for large-scale real-time SLAM. We tested our method using the CityCenter and Lip6Indoor datasets. Our proposed method results can effectively correct the typical false-positive localization of previous methods, thus gaining better recall ratios and better precision.

## 1. Introduction

Simultaneous localization and mapping (SLAM) is widely used to generate maps for localization or autonomous robotic navigation.

The appearance-based SLAM type is characterized by low-cost solutions. Moreover, SLAM based on visual features provides abundant information for use in matching and recognition.

Almost all appearance-based SLAMs are pure bag-of-words approaches that extract SIFT [1] or SURF [2] descriptors from images and then match descriptors by brute force method or NNDR [3] and so forth to calculate likelihood between two locations.

The biggest challenge for improving the precision and recall ratio of loop-closure detection is that the false-positive localization loop-closure hypothesis selection score is higher than the false-positive localization. This results in acceptance of false-positive localizations and rejection of false-negative localizations.

Likelihood calculations between two places are the most decisive factor for establishing a loop-closure hypothesis. But

in many conditions a pure bag-of-words approach cannot effectively calculate the likelihood between places.

Our proposed method attempts to improve the likelihood calculation results by appending the geometric constraints of visual words to the classic pure bag-of-words likelihood calculation. The geometric constraints include order and acreage constraints, which are designed as affine-invariant geometric constraints. Therefore, the proposed method can work well even though the viewpoint is significantly changed. This method uses a memory management approach similar to those in [4, 5] for real-time processing and uses SURF for its visual descriptors; descriptors are matched by NNDR [3].

In this paper, we describe the more accurate likelihood calculation to improve the loop-closure detection performance. By establishing an affine-invariant hypothesis, the proposed method excludes incorrectly matched visual words and calculates the dispersion of correctly matched visual words to improve the accuracy of the likelihood calculation. Section 2 reviews some previous pure bag-of-words-based approaches and their typical problems. In Section 3, we describe the proposed method. Section 4 presents our

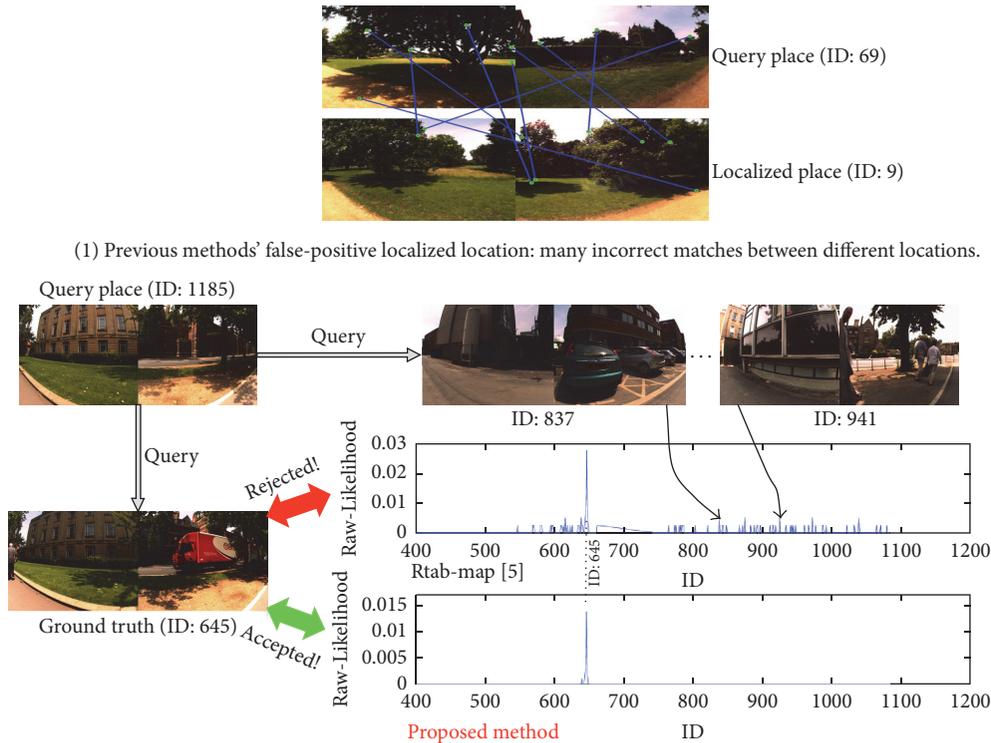


FIGURE 1: Examples of two problems occur in [4, 5, 7, 8] on the basis of CityCenter [14] dataset. (1) illustrates that, because too many words are matched between two different locations incorrectly, previous pure bag-of-words systems treat two different locations as the same place. By using the proposed method, 0 match is accepted. So the proposed method can solve this problem. (2) shows Raw-Likelihood (without normalizing) comparison between [5] and the proposed method. Although the ground truth obtains the highest likelihood, because there are incorrect matches between too many different locations, those noises result in that the ground truth is rejected by previous pure bag-of-words systems incorrectly. The proposed method prevents noises and retains the peak of ground truth and thus the proposed method can accept the ground truth correctly.

experimental results. In Section 5, we discuss our proposed method's advantages, disadvantages, and outlook.

## 2. Related Work

References [4–8] present some typical pure bag-of-words approaches.

Cummins and Newman [6] proposed a rapid method based on the probabilistic bailout condition for appearance-only SLAM. But this approach's precision and recall ratio are not satisfactory.

Kawewong et al. [9] proposed a method that tracks robust features in a sequence of images, called position-invariant robust features (PIRF). They also proposed two online-incremental-appearance-only methods for SLAM PIRF-nav [7] and PIRF-nav2 [8] based on PIRF. Regarding PIRF's robustness, the methods in PIRF-nav and PIRF-nav2 perform satisfactorily in dynamic environments. Compared with the method in [6], the precision and recall ratio also improved significantly.

However, PIRF-nav and PIRF-nav2's processing time for loop-closure detection cannot be controlled very well. The processing time increases as the map's scale increases. In

addition, because PIRF-extracted robust features persist in a sequence of images, many useful features are ignored. This can cause significant loss of visual features, particularly, in indoor low resolution datasets such as Lip6Indoor [10]. Thus, it is difficult to improve the performance of PIRF-nav and PIRF-nav2.

Labbé and Michaud [4, 5] proposed a method based on a short term memory (STM) and Long-Term Memory mechanism called RTAB-map. It can optimize the processing time of SLAM by controlling the processing speed effectively without increasing the processing time when the map's scale increases.

However, because of the problems shown in Figure 1, it is difficult to improve RTAB-map's recall ratio.

RTAB-map is the best vision-only SLAM method currently available and probably represents the limit of performance possible for pure bag-of-words approaches.

FAB-MAP3D [11] is a SLAM method that combines a pure bag-of-words approach with 3D geometric constraints. It works better than the FAB-MAP [6], but it requires 3D measurement information about each visual word.

The proposed method attempts to design geometric constraints for appearance-only SLAM without any 3D measuring.

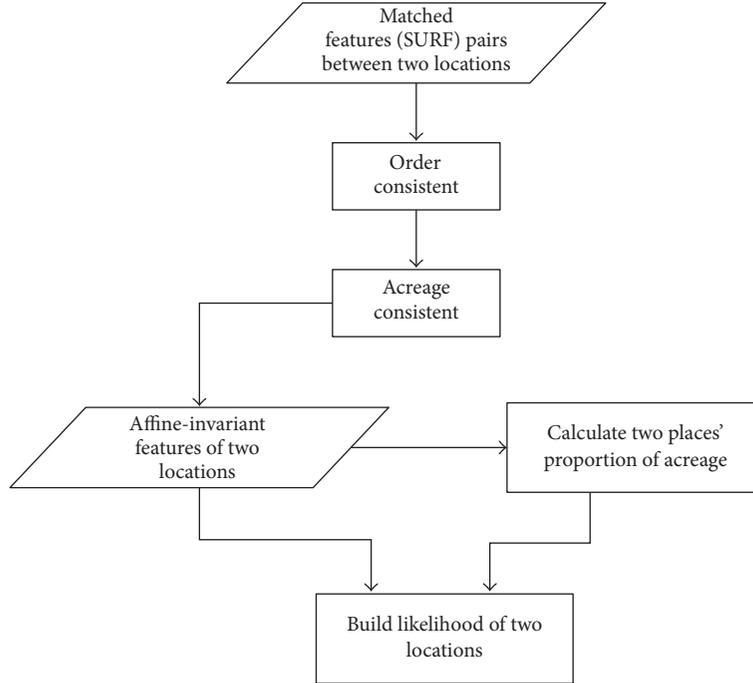


FIGURE 2: Framework of likelihood calculation on the basis of geometric constraints.

Unlike RANSAC [12] and PROSAC [13], the proposed method estimates an affine-invariant hypothesis and calculates the likelihood between two places without any random elements. Thus the proposed method is more stable and better suited for situations in which only few words match.

### 3. Proposed Method

This section presents our new likelihood calculation method having geometric constraints. We also include a brief explanation of the loop-closure hypothesis selection. Figure 2 shows the likelihood calculation of the proposed method.

**3.1. Image Undistortion.** Sometimes a camera lens will cause significant distortions; undistorted images are necessary to establish an affine-invariant hypothesis.

To produce undistorted images, we must establish the camera's requisite intrinsic parameters ( $f_x, f_y, c_x, c_y$ ), radial distortion coefficients ( $k_1, k_2$ ), and its tangential distortion coefficients ( $p_1, p_2$ ) by calibration. It is easy to calibrate and undistort a camera using OpenCV [15]. Intrinsic parameters and distortion coefficients are stable for certain cameras. More details are available in OpenCV documents.

Since the real world is not flat, real world images do not strictly abide by the affine-invariant constraint. However, for the most part, landmarks in images can be considered to be in a flat environment.

**3.2. Order Constraint.** We designed a distance order constraint to exclude incorrectly matched visual words.

As illustrated in Figure 3, ( $d \rightarrow d'$ ) is an example of incorrect matching. We first calculate  $a$ 's relative distance vector

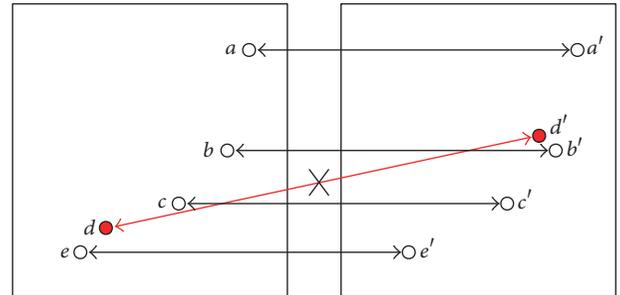


FIGURE 3:  $W = (a, b, c, d, e)$  and  $W' = (a', b', c', d', e')$  are matched visual words between two images. However, the matching ( $d \rightarrow d'$ ) is incorrect.

$D_a$ , which is sorted from nearest to farthest.  $D_a = (b, c, d, e)$  and  $D_{a'} = (d', b', c', e')$ . Except for  $d$  and  $d'$ ,  $D_a = D_{a'}$ .

Similarly,  $D_d = (e, c, b, a)$  and  $D_{d'} = (b', a', c', e')$ . Despite being matched visual words,  $D_d$  and  $D_{d'}$  are significantly different.

We designed an offset-based linear formula to calculate  $\text{diff}(D_i, D_{i'})$ . The offset's definition is shown in Figure 5. We also define  $\text{diff}(D_i, D_{i'})$  in (1). In Figure 4,  $\text{diff}(D_a, D_{a'}) = 0.25$ , and  $\text{diff}(D_d, D_{d'}) = 11/16$ . Therefore, the higher  $\text{diff}$  indicates that the probability of an incorrect matching is higher.  $\text{diff}(D_i, D_{i'})$  can be used to distinguish correctly and incorrectly matched visual words.

$$\text{diff}(D_i, D_{i'}) = \frac{\sum_{k=1}^n \text{offset}_k}{n^2}. \quad (1)$$

Please note that  $\text{diff}(D_i, D_{i'})$  is not an affine-invariant quantity and is sensitive to noise percentage. So we cannot

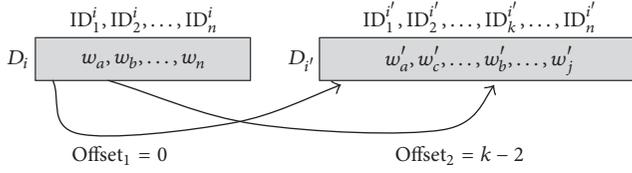


FIGURE 4: Offset's calculation. At first find a word in  $D_i$ 's corresponding matched word in  $D_{i'}$ , and then calculate the offset between corresponding IDs.

set a certain threshold eliminating incorrectly matched visual words for large-scale SLAM. Using a normalized  $\text{diff}_{\text{normal}}$  is one candidate. We normalized  $\text{diff}$ s using its mean  $\mu_{\text{diff}}$  and standard deviation  $\sigma_{\text{diff}}$ .

Our proposed method uses kd tree-based [16] FLANN [17] to establish relative distance vectors  $D$ s when descriptors are extracted. All extracted words are used for establishing  $D$ s and these vectors are retained for further queries. When required, these vectors eliminate all mismatched words by calculating expression (1) to establish new vectors for the processing of order constraint. The original vectors do not change.

In Figure 4,  $d$  and  $d'$  are excluded and we can obtain a corrected set of words  $W_{oc} = (a, b, c, e)$  and  $W'_{oc} = (a', b', c', e')$ .

However, the only order constraint is not strict enough for a highly accurate likelihood calculation. We designed an acreage constraint to establish an affine-invariant hypothesis based on  $W_{oc}$  and  $W'_{oc}$ .

**3.3. Acreage Constraint.** An example of an affine invariant is illustrated in Figure 6. Although, from  $i$  to  $j$ , the coordinates of  $W$  and  $W'$  changed significantly, the proportional relationship of the acreage illustrated in the figure did not change; that is, not only  $S_{aob}/S_{abcd} = S_{a'o'b'}/S_{a'b'c'd'}$ , but also  $S_{aob}/S_{abc} = S_{a'o'b'}/S_{a'b'c'}$ ,  $S_{aob}/S_{aoc} = S_{a'o'b'}/S_{a'o'c'}$ , and so forth.

$$\frac{S_{aob}}{S_{abcd}} = \frac{S_{a'o'b'}}{S_{a'b'c'd'}}. \quad (2)$$

Therefore, when the affine-invariant proportional relationship of acreage has been found, an affine-invariant hypothesis can be established.

We propose a method to establish an affine-invariant hypothesis based on results of the order constraint. First, we calculate a total area:

$$S_{\Sigma} = \sum_{i=1}^{n-1} S_{i,o,i-1}, \quad (3)$$

where  $o$  is the center of gravity of  $W$ :

$$o = \frac{\sum_{i=1}^n w_i}{n}. \quad (4)$$

We then define the deviation of two pairs of visual words according to an affine-invariant hypothesis:

$$\text{Dev}_{ai}^{i,j} = \left\| 1 - \frac{S_{ioj}/S_{\Sigma}}{S'_{i'o'j'}/S'_{\Sigma}} \right\|. \quad (5)$$

$\text{Dev}_{ai}^{i,j} = 0$ , if  $\text{Dev}_{ai}^{i,j} < T_{AC}$  and this establishes an affine-invariant hypothesis. Because  $T_{AC}$  is a robust affine-invariant threshold, a certain  $T_{AC}$  is suitable for large-scale SLAM.

In fact,  $o$  is important in the establishment of the affine-invariant hypothesis. The  $\text{Dev}_{ai}^{i,j}$  is meaningful only in the sense that  $o$  is built by visual words that obey the affine-invariant constraint. After processing the order constraint, the incorrectly matched words have been eliminated, but the noise remains.

**3.4. Likelihood Calculation.** After the above processing ( $W \rightarrow W_{OC} \rightarrow W_{AC}$ ), only correctly matched words remain. Now it is possible to calculate a geometric constraints-based likelihood  $s_{cg}(z_t, z_c)$  between the testing and current place.

$$s_{cg}(z_t, z_c) = S_{\text{Affine}} * S_{\text{Dispersion}}, \quad (6)$$

where  $z_t$  is the testing place and  $z_c$  is the current place.  $S_{\text{Affine}}$  is the proportion between  $W_{AC}$ 's size and the sum of the matched word pairs.

$$S_{\text{Affine}} = \frac{\text{size of } (W_{AC})}{N_{\text{pair}}}, \quad (7)$$

where  $N_{\text{pair}}$  is the number of matched word pairs between the two places.  $0 \leq S_{\text{Affine}} \leq 1$ .

$S_{\text{Dispersion}}$  is dispersion of the affine-invariant words-based parameter for estimating the likelihood between two places.

$$S_{\text{Dispersion}} = \begin{cases} \frac{D}{D'} - 1, & D' \leq D \leq 2D' \\ \frac{D'}{D} - 1, & D \leq D' \leq 2D \\ 0, & D' \geq 2D \text{ or } D \geq 2D' \end{cases} \quad (8)$$

where

$$D^{(l)} = \frac{\sum_{i=1}^n w_i^{(l)} - o^{(l)}}{n}. \quad (9)$$

Apparently  $0 \leq S_{\text{Dispersion}} \leq 1$ .

In [4, 5], the likelihood calculation formula is

$$s(z_t, z_c) = \begin{cases} \frac{N_{\text{pair}}}{N_{z_t}}, & N_{z_t} \leq N_{z_c} \\ \frac{N_{\text{pair}}}{N_{z_c}}, & N_{z_c} < N_{z_t} \end{cases} \quad (10)$$

where  $N_{z_t}$  and  $N_{z_c}$  are the total number of words of the signature  $z_t$  and the compared signature  $z_c$ , respectively. However, since this method attempts to obtain a low likelihood, it may cause a false-negative localization. But for pure bag-of-words-based approaches, because precision is hard to control there is no alternative but to choose low likelihoods.

We propose a new likelihood calculation method combined with  $s_{cg}(z_t, z_c)$ :

$$S_{\text{Combined}} = \frac{s_{cg}(z_t, z_c) * N_{\text{pair}}}{N_{z_t} + N_{z_c}}. \quad (11)$$

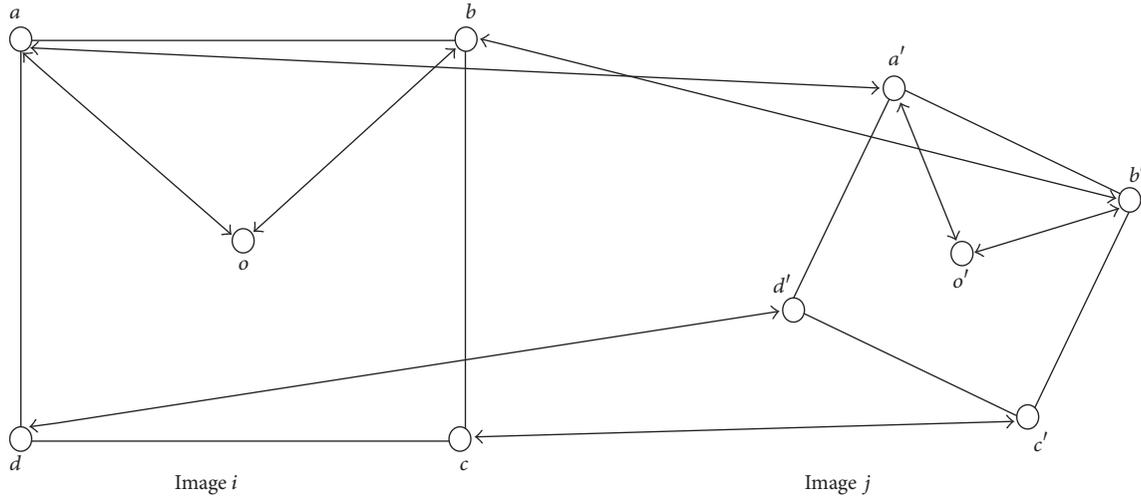


FIGURE 5: An affine-invariant's example. Affine-transfers are including rotation, shearing, translation, and scaling. This example is including rotation and scaling.  $W = (a, b, c, d)$  and  $W' = (a', b', c', d')$  are matched visual words between two images.  $o$  and  $o'$  are center of gravity of each set of words.

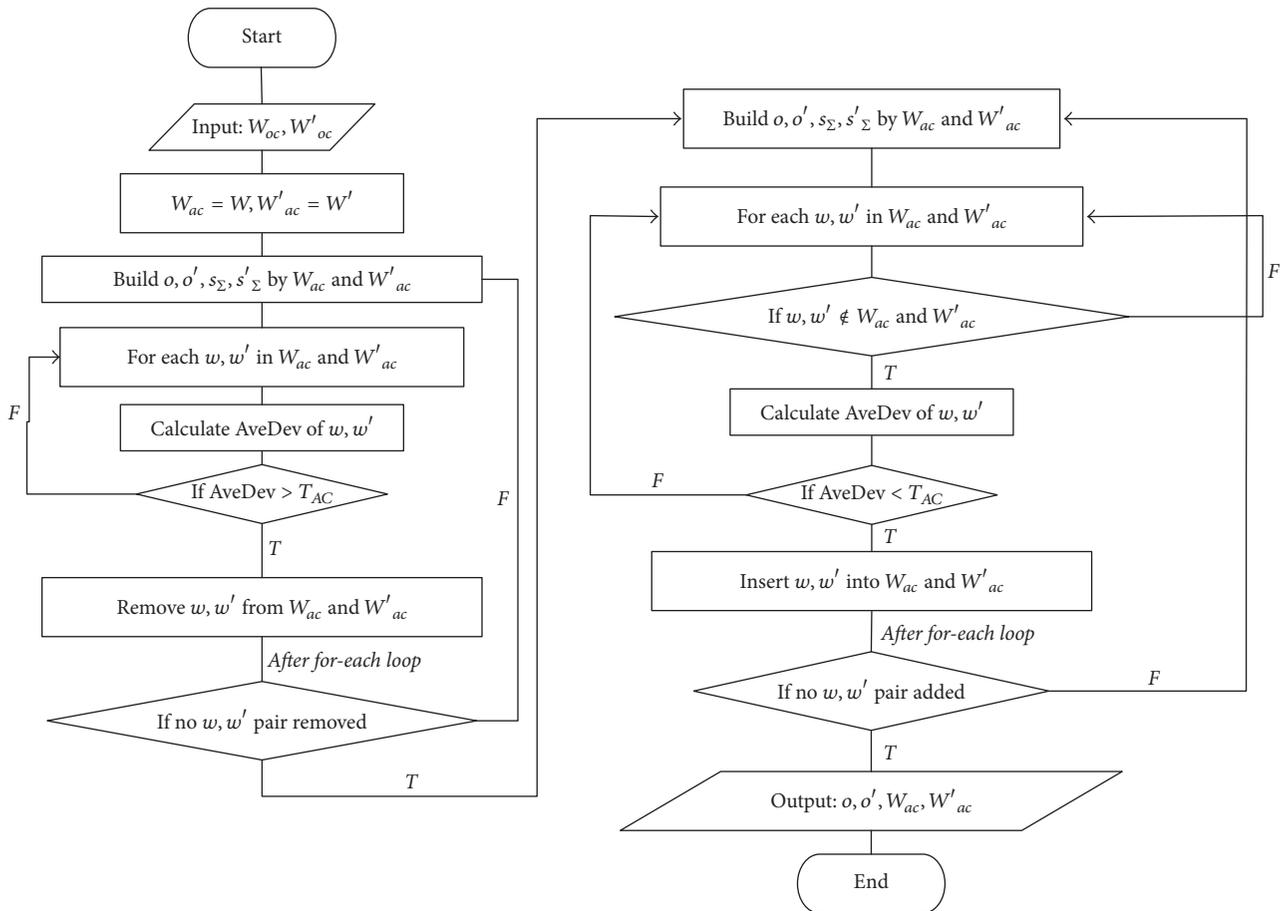


FIGURE 6: Processing flow of acreage constraint. Input:  $W_{oc}, W'_{oc}$  are corresponding pairs already flitted by order constraint. Output:  $W_{ac}, W'_{ac}$  are corresponding pairs processed by acreage constraint.  $o, o'$  are center of gravity of  $W_{ac}, W'_{ac}$ .  $W_{ac}, W'_{ac}$  and  $o, o'$  are used to establish a likelihood between two locations. Left part of figure is the first step of the algorithm. This step is obtaining credible  $o$  and  $o'$ . It will be converged when size of  $W_{tmp}$  and  $W'_{tmp}$  did not change and obtain credible  $o$  and  $o'$ . The convergence is achieved rapidly (after 2~3 loops and then it gets converged) because, during establishing credible  $o$  and  $o'$ , some correct matching may be rejected (overfitted). Right part of Figure 6 presents the processing of attempting to retrieve incorrectly rejected matching on the basis of credible  $o$  and  $o'$ .



FIGURE 7: Typical false-positive localized place in Lip6Indoor data set. Although the query place (right) matches only a part of the localized place (left), pure bag-of-words approaches do not calculate likelihood based on matched acreage.

This likelihood calculation method is fairer than [4, 5]. In addition, since geometric constraints are added to the calculation, the proposed method achieves better accuracy.

**3.5. Brief Summary of Loop-Closure Hypothesis Selection.** The proposed method uses a loop-closure hypothesis selection method similar to that in [5]. We update the Bayesian filter by the following recursion formula:

$$\begin{aligned}
 P(S_t | L^t) \\
 = \eta P(L^t | S_t) \sum_{i=-1}^{t_n} p(S_t | S_{t-1} = i) p(S_{t-1} | L^{t-1}), \quad (12)
 \end{aligned}$$

where  $S_t = i$  is the probability that  $L_t$  closes a loop with a past location  $L_i$  and  $S_t = -1$  is the probability that the current place in the STM is a new place.  $p(L^t | S_t)$  is important for this formula, being a normalized likelihood by the mean  $\mu$  and standard deviation  $\sigma$ , which the proposed method significantly affects.  $\sum_{i=-1}^{t_n} p(S_t | S_{t-1} = i) p(S_{t-1} | L^{t-1})$  briefly describes the likelihood, which, due to space limitations, we cannot describe in detail. Please refer to [4, 5].

When  $p(S_t = -1 | L^{t-1})$  is lower than the loop-closure threshold  $T_{loop}$ , the loop-closure hypothesis will be accepted.

$$P(L^t | S_t = -1) = \frac{\mu}{\sigma} + 1. \quad (13)$$

Please note that when  $\mu$  is too high the loop-closure hypothesis will be rejected, although the probability of a high loop-closure hypothesis is very high. This may cause false-positive localizations.

## 4. Experiments

We performed our calculations using a MacBook Pro, i7 with 16 GB RAM. The application is written in C++. We tested our method by two well-known datasets: Lip6Indoor and CityCenter.

**4.1. Lip6Indoor.** Figure 7 shows that a typical false-positive localized place occurs in pure bag-of-words approach such as [5, 6, 8]. After processing by the proposed method, geometric constraints for two places  $S_{Dispersion} = 0.12$ , the false-positive

TABLE 1: Results of experiment Lip6Indoor.

	Recall	Precision	Average processing time (per frame)
The proposed method	99.55%	100%	731.7 ms
RTAB-map [5]	98%	100%	678.2 ms
FAB-map [6]	23.6%	100%	588.6 ms
PIRF-nav2.0 [8]	77.3%	99.4%	103.2 ms

loop-closure hypothesis is rejected by the proposed method. The resolution of 388 images in this dataset is  $240 \times 192$ . Compared with [5], the proposed method improved the recall proportion by only 1.55%. But as the recall proportion increases in [5], precision decreases rapidly. For a 100% recall proportion of [5] precision is 63%, but the precision of our proposed method is 87.5%. Table 1 shows the results of Lip6Indoor dataset.

References [6, 8] are faster than [5], but their recall proportion is low. After comparison between the proposed method and [5], an average of 53.5 ms additional processing time is required for each frame. The maximum processing time for one frame using our proposed method is 825.3 ms. Because this dataset is captured at 1 HZ, the proposed method can be processed in real time.

**4.2. CityCenter.** In the CityCenter dataset, since our method has effective control, we obtained a higher recall proportion. The resolution of 2474 images in this dataset is  $640 \times 480$ . Every two images were captured simultaneously at the same location.

The recall proportion cannot be further increased because in some scenes (like jungles) there are too many similar words. The proposed method failed in these types of scenes. With too many incorrectly matched pairs, a bad affine-invariant hypothesis was established. Table 2 shows the results of CityCenter dataset.

The maximum processing time for one frame of the proposed method is 1780.7 ms. The dataset is captured at approximately 0.5 Hz, so the proposed method can also be processed in real time in this dataset.

TABLE 2: Results of experiment CityCenter.

	Recall	Precision	Average processing time (per frame)
The proposed method	86.2%	100%	1570.1 ms
RTAB-map [5]	81%	100%	1375.2 ms
FAB-map [6]	43.32%	100%	233.2 ms
PIRF-nav2.0 [8]	80%	99.4%	426.3 ms

## 5. Conclusion and Future Studies

These experiments showed that our proposed method can work better than pure bag-of-words-based SLAM approaches. We proved that 2D geometric constraints are an effective way to break the bottleneck and improve the accuracy of appearance-based SLAM.

Although the proposed method works well for the most part, it cannot handle some problems. In particular, one typical problem is too many similar words in the same image. Methods to solve this problem are being considered. One possible solution is to increase NNDR [3] threshold to avoid repeated features more effectively. This step should reduce false-positive ratio of descriptors matching but causes more false-negative avoiding. Then, use the proposed method to construct affine-invariant hypothesis based on features matched by higher threshold. Lastly, test avoided repeated features by affine-invariant hypothesis to retrieve potential correct matches.

Today, high-performance handheld smart phones are very popular. Because the proposed method does not require any 3D measuring to achieve high robustness while using handheld devices, it may be applied to many types of platforms, for navigation by pedestrians.

## Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this article.

## Acknowledgments

This research is supported by the Key Project of National Natural Science Foundation of China (Grant no. 51538007) and the Project of National Natural Science Foundation of China (Grant no. 71101096).

## References

- [1] D. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1150–1157, Kerkyra, Greece, 1999.
- [2] H. Bay, T. Tuytelaars, and L. Gool, "Surf: speeded up robust features," in *Computer Vision—ECCV 2006*, vol. 3951, pp. 404–417, Springer, 2006.
- [3] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [4] M. Labbé and F. Michaud, "Memory management for real-time appearance-based loop closure detection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems: Celebrating 50 Years of Robotics (IROS '11)*, pp. 1271–1276, San Francisco, Calif, USA, September 2011.
- [5] M. Labbé and F. Michaud, "Appearance-based loop closure detection for online large-scale and long-term operation," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 734–745, 2013.
- [6] M. Cummins and P. Newman, "Invited applications paper FAB-MAP: appearance-based place recognition and mapping using a learned visual vocabulary model," in *Proceedings of the 27th International Conference on Machine Learning (ICML '10)*, Edinburgh, UK, 2010.
- [7] A. Kawewong, N. Tongprasit, S. Tangruamsub, and O. Hasegawa, "Online and incremental appearance-based slam in highly dynamic environments," *International Journal of Robotics Research*, vol. 30, no. 1, pp. 33–55, 2011.
- [8] N. Tongprasit, A. Kawewong, and O. Hasegawa, "PIRF-Nav 2: speeded-up online and incremental appearance-based SLAM in an indoor environment," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV '11)*, pp. 145–152, IEEE, Kona, Hawaii, USA, January 2011.
- [9] A. Kawewong, S. Tangruamsub, and O. Hasegawa, "Position-invariant robust features for long-term recognition of dynamic outdoor scenes," *IEICE Transactions on Information and Systems*, vol. E93-D, no. 9, pp. 2587–2601, 2010.
- [10] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "Fast and incremental method for loop-closure detection using bags of visual words," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1027–1037, 2008.
- [11] R. Paul and P. Newman, "FAB-MAP 3D: topological mapping with spatial and visual appearance," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '10)*, pp. 2649–2656, Anchorage, Ala, USA, May 2010.
- [12] 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 Association for Computing Machinery*, vol. 24, no. 6, pp. 381–395, 1981.
- [13] O. Chum and J. Matas, "Matching with PROSAC—progressive sample consensus," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 220–226, IEEE, San Diego, Calif, USA, June 2005.
- [14] M. Cummins and P. Newman, "FAB-MAP: probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.
- [15] G. Bradski, "The OpenCV library," *Dr. Dobbs Journal of Software Tools*, 2000.
- [16] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp. 209–226, 1977.
- [17] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proceedings of the International Conference on Computer Vision Theory and Application (VISSAPP '09)*, pp. 331–340, INSTICC Press, 2009.

