Appearance-Based Place Recognition Using Whole-Image BRISK for Collaborative Multi-Robot Localization

Jung H. Oh, Gyuho Eoh, and Beom H. Lee
Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea
Email: {bulley85, geni0620, bhlee}@snu.ac.kr

***Abstract*—This paper deals with the problem of recognizing places based on their appearance for collaborative mobile robot localization. A whole-image descriptor and BRISK are combined in order to extract information from images collected by multiple mobile robots. The bag-of-words method is then adopted to calculate similarity scores between obtained images, and this enables each robot to find other robots' previously visited locations. Such detections make it possible to increase the precision of the actual pose estimates and achieve a precise collaborative localization. Experiments are performed to verify the effectiveness of the proposed method in outdoor environments.**

***Index Terms*—place recognition, loop closures, BRISK, image descriptor, multi-robot, Simultaneous Localization and Mapping (SLAM)**

I. Introduction

Simultaneous Localization and Mapping (SLAM) is one of the most widely researched areas in robotics. Recently, vision-based SLAM has become an active field as cameras have become more compact and accurate while providing the rich qualitative information of the environment. One of the most significant requirements for vision-based SLAM is robust place recognition that provides correct data association to obtain correct robot poses. In particular, finding a place that has already been visited in a cyclical excursion or arbitrary length is referred to as a *loop-closure detection* problem, which is crucial for enhancing the robustness of localization and mapping.

The *bag-of-words* method has been a popular way to perform visual loop-closure detection [1]-[3]. Each image is quantized into a set of visual words, and can be represented by histograms that can be compared efficiently using histogram comparison methods. In [1], FAB-MAP framework based on the bag-of-words method with probabilistic reasoning worked robustly over a long trajectory. The bag-of-words method was extended to incremental conditions in [2]. It also relied on Bayesian filtering to estimate loop-closure probability. Both works used SIFT [4] or SURF [5] to extract features from images, as they are robust to lighting, scale, or rotation changes. In [3], a method for visual place recognition using bag of words is proposed, using the FAST [6] keypoint detector and BRIEF [7] features. In particular, this work demonstrated the effectiveness of the binary features such as BRIEF, BRISK [8] or FREAK [9], which outperform the computation time of SIFT and SURF, maintaining rotation and scale invariance. Instead of using the locally extracted imaged descriptor, loop-closure detection using the whole-image descriptor was proposed in [10]. This descriptor uses whole information of the image and does not require keypoint detection step. In general, it is more susceptible to change in the camera's view than local descriptor methods. However, if we assume that the camera motion is planar, it is more robust to false positive errors and fast to compute the similarity.

In this work, we propose a whole-image descriptor combined with a binary descriptor, BRISK, which is more invariant to scale and rotation than BRIEF. By combining the whole-image descriptor and the binary descriptor, we can exploit the advantages of both descriptors, and significantly improve the efficiency and performance of the place recognition. This allows each robot to correct its positions, which improves the accuracy of collaborative multi-robot localization.

II. Place Recognition Using Whole-Image BRISK

A. Bag-of-Words Framework

Images can be represented as the set of visual words that is generated from the feature descriptors. Let ${}^{k}\mathcal{I}$ be an obtained image query from the robot k and ${}^{k}\mathbf{Z}$ be the representations of these images in the feature space. There are many descriptors to represent images, such as SIFT or SURF. In this paper, the whole-image BRISK is used to describe the images.

Then, the dictionary is built by clustering these visual descriptors, and the representative descriptors are called *visual words*. Given the dictionary, extracted features from each image can be quantized to the nearest visual words and can be represented by the histogram of visual words in the dictionary. Finally, each image can be

Manuscript received April 10, 2015; revised July 12, 2015.

doi: 10.18178/ijmerr.4.3.264-268

represented by a histogram. The set of these histograms from the robot k are denoted by ${}^{k}\mathbf{H}$. Then, we can calculate the similarity score between the histograms ${}^{u}H_i$ and ${}^{v}H_j$ using the histogram intersection function S as the following.

$$S({}^{u}H_i, {}^{v}H_j) = \sum_{k} \min({}^{u}H_i(k), {}^{v}H_j(k)) \qquad (1)$$

The overall procedure of place recognition using the bag-of-words method is shown in Fig. 1. From the bag-of-words framework, we can calculate similarity scores between obtained images, and this enables each robot to find other robots' previously visited locations from similarity scores. The most important thing in this framework is the appropriate image descriptor that is robust and efficient.

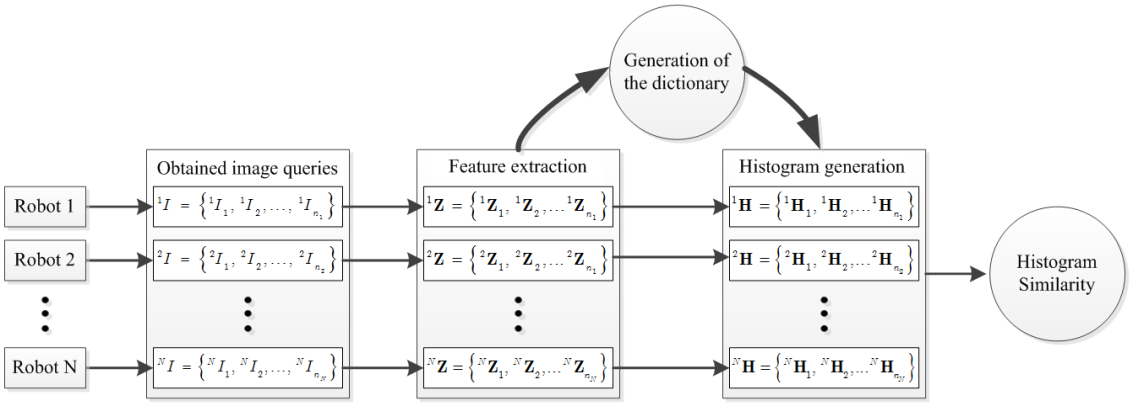
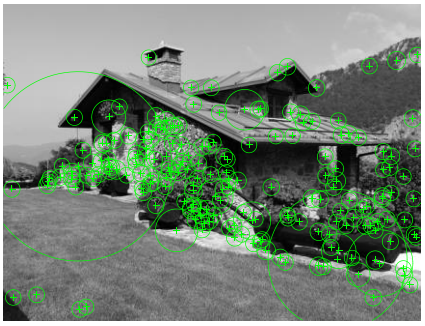


Figure 1. The procedure of place recognition using the bag-of-words method.

B. Whole-Image BRISK

The BRISK descriptor is a local binary descriptor invariant to scale and rotation. We choose this descriptor for generating the whole image descriptor as it shows high quality performance in place recognition at a lower computational cost. To extract features using BRISK, a two-step process is required; the first step is a *keypoint detection* and the second is a *keypoint description*.



(a)

(b)

Figure 2. The comparison of keypoints' locations. (a)BRISK (b)the whole-image BRISK

The keypoint detection step is to find the interest points from the image. Edges, corners, or blobs can be the interesting part of an image. As these points can show the unique characteristics of the image, local image descriptors extract features from these points. A whole image descriptor, on the other hand, does not extract features from the keypoints, but extract them from the predefined points in the image as shown in Fig. 2 (b). Therefore, the keypoint detection step that is usually computationally expensive can be omitted in the whole-image BRISK. These predefined keypoints are selected using a $d_r \times d_c$ grid-based pattern. Therefore, if the size of the image is $m \times n$, we can extract $mn / d_r d_c$ features from these keypoints.

As the keypoint detection step is not required for generating the whole image descriptor, the algorithm for extracting features becomes simpler, and the computation time is reduced. Another thing is we can use not only the partial information near the keypoints, but also the background information as it extracts features from a whole image.

The keypoints' locations between the original BRISK and the whole image BRISK are as shown in Fig. 2(a) and (b). In general, the whole image descriptor is more susceptible to change in the camera's view than local descriptor methods. However, if we assume that the camera motion is planar, it is more robust to false positive errors and fast to compute the similarity.

III. Experimental Results

A. The Evaluation of Image Descriptor Performance

To evaluate the proposed descriptor whole-image BRISK, we compared the execution time for other three

descriptors, SIFT, SURF, and BRISK. The total execution time is composed of three steps. The first step is keypoint extraction time, and the second step is the description time, and the final step is the matching time. The results are summarized in Table I.

TABLE I. DESCRIPTOR EXTRACTION AND MATCHING TIME RESULTS

	SIFT	SURF	BRISK	WI-BRISK
Keypoint extraction time [ms]	25.71	26.05	26.14	0
Description time [ms]	423.85	57.83	5.87	5.87
Matching time [ms]	37.48	18.26	5.82	5.82
Total [ms]	487.04	102.14	37.83	11.69

As shown in the Table I, the BRISK and the whole-image BRISK showed the higher performance than SIFT and SURF, as they are binary descriptors. Moreover, the whole-image BRISK does not consume the keypoint extraction time, as the keypoint detection step is not required for generating the whole image descriptor. As a result, the proposed descriptor showed the less computation time than other descriptors

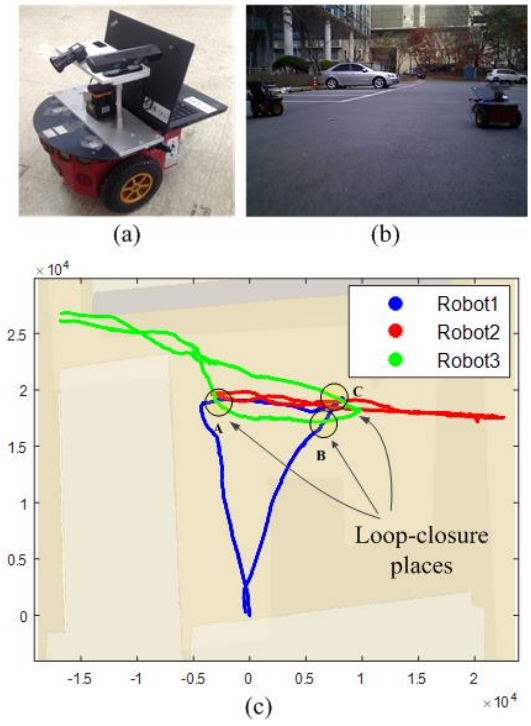


Figure 3. Experimental environment for place recognition (a)The mobile robot equipped with a camera and a laptop (b)An sample image obtained from the camera (c)Trajectory of the three robots.

B. Outdoor experiment

An experiment was conducted with three mobile robots in an outdoor environment to evaluate the proposed method. The environment of the experiment is shown in Fig. 3. The pioneer-3dx mobile robot is used for the experiment, and each robot was equipped with a camera and a laptop to collect images of 640 × 480 resolutions at 1Hz (Fig. 3 (a) and (b)). The robots started at different locations, and visited appointed places where other robots had already visited as shown in Fig. 3 (c). These places should be recognized for the loop-closure places.

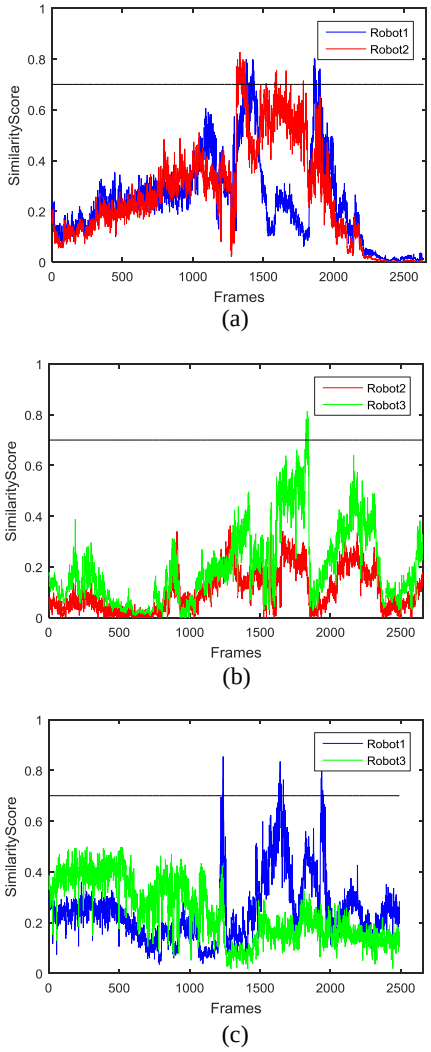


Figure 4. The similarity score results between robots (a)Robot 1 and Robot 2 (b)Robot 2 and Robot 3 (c) Robot 1 and Robot 3.

The experimental results of calculating the similarity score between robots are shown in Fig. 4. As we performed the experiment using three robots, the pairwise similarity scores for each robot are depicted in the plots, respectively. The threshold for loop closure places is defined as 0.7 in this experiment; above the threshold score is the recognized loop closure places.

For the case of the R1 and the R2, the loop closure the place A is first visited by the R1. As shown in Fig. 4 (a), the score of the R2 is above the threshold at around the frame 1300 and detect the place A as a loop closure place.

Similarly, the place C is first visited by the R2. The score of R1 is high around the frame 1800, and detect the place C as a loop closure place.

As shown in Fig. 4 (b), the R2 and the R3 have only one loop closure place C. This place is first visited by the R2. Therefore, there are no scores above the threshold for the view of the R2. On the other hand, the score of R3 is above the threshold score around the frame 1800, as it visited the place C. As a result, the loop closure place C can be detected as a loop closure place.

Finally, we can find the place A, B, and C as loop closure places for the case of the R1 and the R3 as shown in Fig. 4 (c). These places are first visited by the R3. Therefore, the scores of the R3 is below the threshold for the all frames. However, we can find three peaks which means the place A, B, and C for the view of the R1. As a result, we could conclude that the loop closure places are correctly recognized although there are different camera viewpoints between robots.

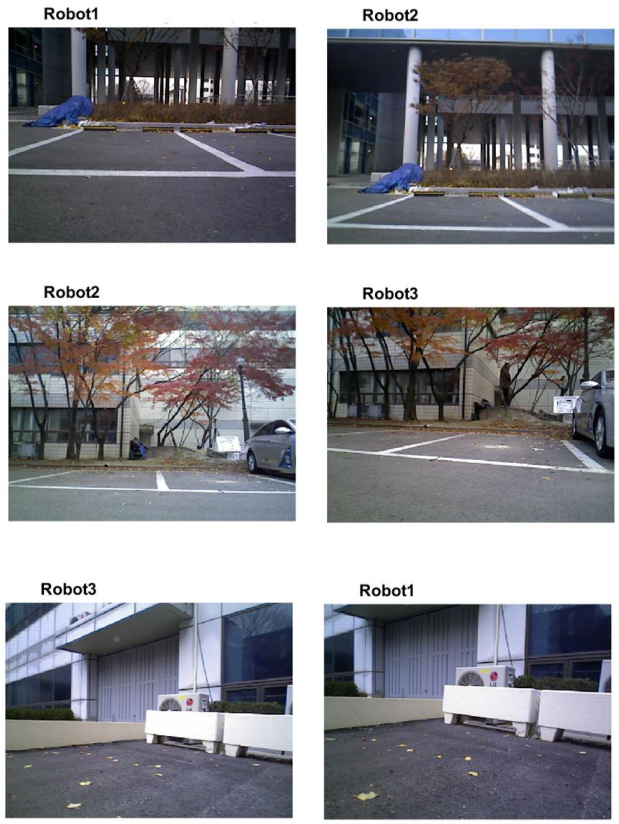


Figure 5. The detected loop closure places between robots.

The example results of detected loop closure places are shown in Fig. 5. Although the camera views are different, each robot found other robots' previously visited locations. From the results, we verified the effectiveness of the proposed method. If we use these detections for localization, it is possible to improve the precision of the actual pose estimates and achieve a precise multi-robot collaborative localization.

IV. Conclusion

In this paper, we proposed a whole-image BRISK to extract the information from the images. The bag-of-words method was used to calculate similarity scores based on this features, and could recognize the loop-closure places. By combining the whole-image descriptor and BRISK, we could significantly improve the efficiency and performance of the place recognition. This allows each robot to correct its positions, which improves the accuracy of collaborative multi-robot localization. We proved the effectiveness of the proposed method by performing the experiment in an outdoor environment.

Acknowledgment

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2013R1A2A1A05005547), in part by ASRI, and in part by the Brain Korea 21 Plus Project.

References

[1] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Int. J. Rob. Res.*, vol. 27, no. 6, pp. 647–665, 2008.
[2] A. Angeli, D. Filliat, S. Doncieux, and J. A. Meyer, "Fast and incremental method for loop-closure detection using bags of visual words," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1027–1037, 2008.
[3] D. Galvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1188–1197, 2012.
[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[5] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
[6] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, 2010.
[7] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: Computing a local binary descriptor very fast," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1281–1298, 2011.
[8] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2548–2555.
[9] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast Retina Keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 510–517.
[10] Y. Liu and H. Zhang, "Performance evaluation of whole-image descriptors in visual loop closure detection," in *Proc. IEEE Int. Conf. Inf. Autom.*, 2013, pp. 716–722.

Jung H. Oh received the B.S. and M.S. degrees in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea in 2012 and 2014. He is currently a Ph. D. candidate in the Department of Electrical and Computer Engineering at Seoul National University. His research interests include vision-based robotics applications, machine learning, simultaneously localization and mapping, and multi-agent system coordination.

Gyuho Eoh received the B.S and M.S degree in electrical engineering and computer science from the Seoul National University in 2011. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering at Seoul National University. His research interests include the multi-robot cooperation; fault tolerance system; and swarm robotics

Beom H. Lee received his B.S. and M.S. degrees in Electronics Engineering from Seoul National University, Seoul, Korea in 1978 and 1980, respectively, and his Ph.D. degree in Computer, Information and Control Engineering from the University of Michigan, Ann Arbor, in 1985. From 1985 to 1987, he was with the School of Electrical Engineering at Purdue University, West Lafayette, IN, as an Assistant Professor. He joined Seoul

National University in 1987, where he is currently a Professor at the School of Electrical Engineering and Computer Sciences. Since 2004, he has been a Fellow of the Robotics and Automation Society. His research interests include multi-agent system coordination, control, and application.