Visual Loop-Closure Detection Method Using Average Feature Descriptors

Deok-Hwa Kim and Jong-Hwan Kim

Department of Electrical Engineering, KAIST
335 Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea
dhkim@rit.kaist.ac.kr, johkim@rit.kaist.ac.kr

Abstract. This paper proposes a novel visual loop-closure method using average feature descriptors. The average feature descriptors are computed by averaging the descriptors of feature points at each frame. Through GPGPU (General-Purpose computing on Graphics Processing Units) technique, the proposed method selects a frame having a minimum distance with the current average feature descriptor from the average feature descriptor history. After the minimum distance calculation, loop-closure is determined by matching feature points between the selected frame and current frame. Experiments results demonstrate that the proposed method successfully detects the visual loop-closure and is much faster than the conventional visual loop-closure method in detecting the visual loop-closure.

1 Introduction

Visual simultaneous localization and mapping (SLAM) [1][2][3][4] systems are widely used by robots and autonomous vehicles to build up a map within an unknown environment using a vision sensor. In applications of the visual SLAM, loop-closure detection is a issue that require the capacity to recognize a previously visited place from current vision sensor measurements. In a graph-based visual SLAM system, once a loop closure is detected, an additional constraint between data frames is added to optimize poses and the map.

In the field of computer vision, many researchers have conducted research to detect loop-closure using vision sensors. In 2006, Newman et al. presented outdoor SLAM system using a visual appearance and laser ranging [5]. This research detected the loop-closure in an image classification scheme. In 2008, Angeli et al. proposed a real-time visual loop-closure method using a concept of visual words to recognize a previously visited place. This research relied on the Bayesian filter to estimate the loop-closure probability [6]. In 2012, Henry et al. presented a keyframe-based loop-closure method [3]. They defined the keyframes that are a subset of the aligned frames to save unnecessary time. Once the loop closure is detected, the new correspondence between data frames can be used as an additional constraint in the graph-based SLAM [7] consisting a pose-graph. Because of the new correspondence, the pose-graph is dynamically optimized by graph optimization techniques [8][9][10].

Although the previous research have been applied to detect loop-closure and achived successful results, they have not taken computation time increasing linearly as time goes by. To detect the loop-closure, measurements of a current frame are matched with all of measurements of previously visited places. As the map size is increased, the computation time to recognize a previously visited place is increased.

In this paper, a visual loop-closure detection method using average feature descriptors is proposed. The average feature descriptors are computed by averaging the descriptors of feature points at each frame. Through the concept of average feature descriptors, parallel processing for the GPGPU is maximized and a memory efficiency is increased.

This paper is organized as follows. Section 2 proposes the novel visual loop-closure detection method using the average feature descriptors. In Section 3, experimental environment and scenarios are presented and the experimental results to detect the loop-closure in a real-time are discussed. Finally, concluding remarks follow in Section 4.

2 Visual Loop-Closure Detection Method Using Average Feature Descriptors

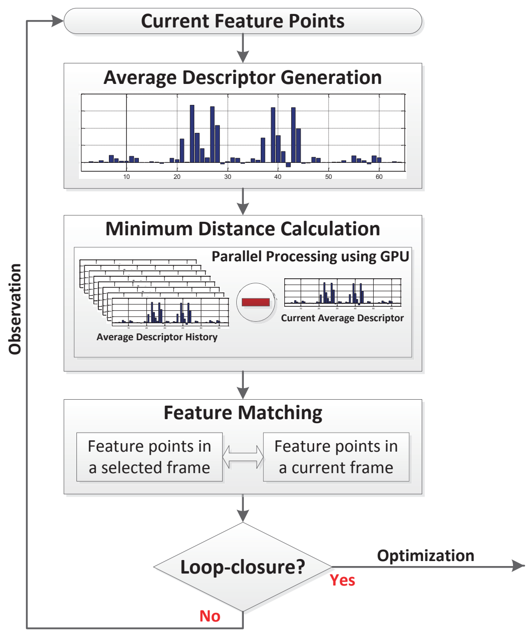


Fig. 1. A diagram of the proposed visual loop-closure detection method

The proposed visual loop-closure detection method is shown in Fig. 1. This proposed method is based on descriptors of feature points. The descriptors of the feature points are obtained by various feature point detection algorithm, such as SIFT, SURF and ORB [11][12][13]. The average feature descriptors are generated using the feature descriptors at each frame. Through the GPGPU, the proposed method selects the frame having a minimum distance with the current average feature descriptor from the average feature descriptor history. After the minimum distance calculation, the loop-closure is determined by matching the feature points between the selected frame and current frame.

2.1 Average Feature Descriptor

The usage of the feature points in all previous image frames is too slow and too large data to recognize the previously visited place. Therefore, the proposed visual loop-closure method uses the average feature descriptors to maximize the parallel processing and increase the memory efficiency. The average feature descriptors at the *i-th* is calculated from

$$m_{i,j} = \frac{1}{N} \sum_{k=1}^{N} f_{i,j,k} \tag{1}$$

where j is an index of the feature descriptor element; k is an index of the feature points; N is a number of the feature points; $m_{i,j}$ is an average feature descriptor of *j-th* element at *i-th* frame; $f_{i,j,k}$ is a *k-th* feature descriptor of *j-th* element at *i-th* frame. In this paper, the average feature descriptor is used as a frame identification at each frame. Fig. 2 shows an example of the average feature descriptor.

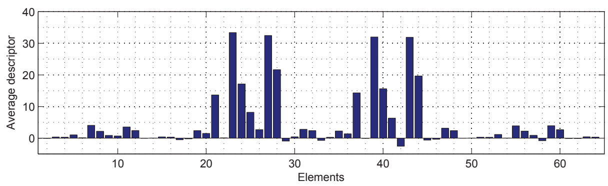


Fig. 2. An average feature descriptor of each image frame

2.2 Minimum Distance Calculation Using GPGPU

The GPGPU is the technique of using a GPU to perform computations that are usually handled by the CPU. Usually, the GPGPU is used to processing parallel tasks at high speed in the field of computer vision. Therefore, in this paper, parallel processing to detect the visual loop-closure is maximized through the concept of the average feature descriptors.

A distance of the average feature descriptors between the *i-th* frame and current frame is calculated by the Euclidean distance. A frame having the minimum distance of the average feature descriptor among the previous frames is decided by

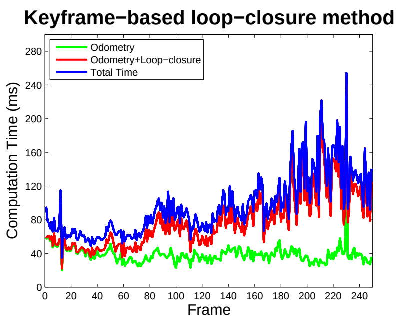
$$i_{min,cur} = \arg\min_{i} \left\{ \sqrt{\sum_{j=1}^{O} (m_{i,j} - m_{cur,j})^2} \right\} \tag{2}$$

where $i_{min,cur}$ is a frame index having the minimum distance of the average feature descriptors; O is a size of the feature descriptor elements; $m_{cur,j}$ is an average feature descriptor of *j-th* element at the the current frame.

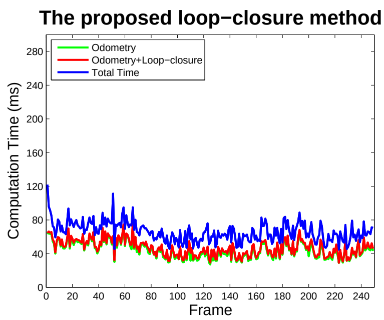
After the minimum distance calculation, the proposed method conducts matching the feature points between the selected frame and the current frame. The matching process of the feature points is performed by the Brute-force matcher. If the number of matched feature points exceeds a threshold, which is assigned heuristically, then the proposed method considers that the loop-closure is detected.

3 Experiment

The proposed visual loop-closure method was tested using an RGB-D sensor, known as the Kinect sensor. The experimental setup was composed of Gentoo OS, Intel i5 3.3GHz dual-core processor, NVIDIA GTX 560 GPU and 6GB RAM. The proposed visual odometry algorithm was tested in the Robot Intelligence Technology laboratory at KAIST. To test the proposed visual loop-closure method, Euclidean error RANSAC (EE-RANSAC) algorithm was used to estimate visual odometry of the RGB-D sensor [3]. The EE-RANSAC algorithm computed the visual odometry recursively through minimizing the Euclidean error between the two consecutive frames. Once a loop closure was detected, a pose-graph of the RGB-D sensor was optimized by iSAM algorithm [8]. Keyframe-



(a)

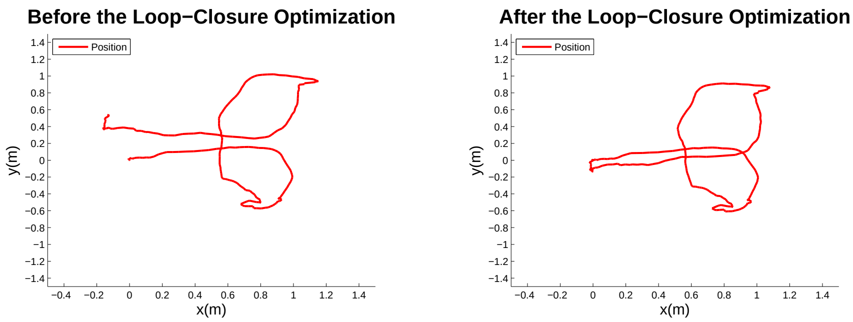


(b)

Fig. 3. Computation time results of the visual SLAM. (a) Using the keyframe-based visual loop-closure method (b) Using the proposed visual loop-closure method

based visual loop-closure method [3] was used for comparison purpose. This paper defined the keyframes based on visual overlap.

The experiment results of the computation time are shown in Fig. 3. As a result, the keyframe-based visual loop-closure method was much slow than the proposed visual loop-closure method. As time went by, while the computation time of the keyframe-based method linearly increased, the computation time of the proposed method was steady.



(a) (b)

Fig. 4. The loop-closure optimization results with the proposed visual loop-closure method. (a) Before the loop-closure optimization. (b) After the loop-closure optimization

After the loop-closure was detected, the pose-graph of the RGB-D sensor was optimized by the iSAM algorithm as shown in Fig. 4. Through the loop-closure optimization technique, an accurate pose of the RGB-D sensor could be obtained.

4 Conclusion

In this paper, a novel visual loop-closure method using average feature descriptors was proposed. The usage of feature points in all previous image frames was too slow to recognize the previously visited place. Therefore, the proposed visual loop-closure method used the average feature descriptors to maximize parallel processing and increase memory efficiency. The average feature descriptors were computed by averaging the descriptors of the feature points at each frame. Through the GPGPU technique, the proposed method selected a frame having a minimum distance with the current average feature descriptor from the average feature descriptor history. Finally, after the minimum distance calculation, loop-closure was determined by matching feature points between the selected frame and current frame. To verify the effectiveness of the proposed method, a visual SLAM experiment was conducted. Through the comparison with the conventional visual loop-closure method, our proposed method was verified to be much faster in detecting the visual loop-closure.

Acknowledgment

This work was supported by the Technology Innovation Program, 10045252, Development of robot task intelligence technology that can perform task more than 80% in inexperience situation through autonomous knowledge acquisition and adaptational knowledge application, funded By the Ministry of Trade, industry & Energy (MOTIE, Korea).

This research was also supported by the MOTIE (The Ministry of Trade, Industry and Energy), Korea, under the Human Resources Development Program for Convergence Robot Specialists support program supervised by the NIPA (National IT Industry Promotion Agency)(H1502-13-1001).

References

[1] Wang, J., Zha, H., Cipolla, R.: Coarse-to-fine vision-based localization by indexing scale-invariant features. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on **36**(2) (2006) 413–422

[2] Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: Monoslam: Real-time single camera slam. Pattern Analysis and Machine Intelligence, IEEE Transactions on **29**(6) (2007) 1052–1067

[3] Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. The International Journal of Robotics Research **31**(5) (2012) 647–663

[4] Han, S., Kim, J., Myung, H., et al.: Landmark-based particle localization algorithm for mobile robots with a fish-eye vision system. IEEE/ASME Transactions on Mechatronics **PP**(99) (2012) 1–12

[5] Newman, P., Cole, D., Ho, K.: Outdoor slam using visual appearance and laser ranging. In: Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on, IEEE (2006) 1180–1187

[6] Angeli, A., Doncieux, S., Meyer, J.A., Filliat, D.: Real-time visual loop-closure detection. In: Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on, IEEE (2008) 1842–1847

[7] Thrun, S., Montemerlo, M.: The graph slam algorithm with applications to large-scale mapping of urban structures. The International Journal of Robotics Research **25**(5-6) (2006) 403–429

[8] Kaess, M., Ranganathan, A., Dellaert, F.: isam: Incremental smoothing and mapping. Robotics, IEEE Transactions on **24**(6) (2008) 1365–1378

[9] Kuemmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g2o: A general framework for graph optimization. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA). (2011)

[10] Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J., Dellaert, F.: isam2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE (2011) 3281–3288

[11] Lowe, D.G.: Object recognition from local scale-invariant features. In: Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Volume 2., Ieee (1999) 1150–1157

[12] Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Computer Vision–ECCV 2006. Springer (2006) 404–417

[13] Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: an efficient alternative to sift or surf. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE (2011) 2564–2571