Published by :
http://www.ijert.org
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 09, September-2021

Review of Loop Closure Detection Methods for Mobile Robot Visual SLAM

Jinhai Jiang
School of Mechanical Engineering
Tianjin University of Technology and Education
Tianjin, 300222, China

***Abstract*—Simultaneous Localization and Mapping (SLAM) is a hot topic in the field of mobile robots. In this paper, the traditional loop closure detection methods in visual SLAM are introduced, their advantages and disadvantages are described, and the loop closure detection methods based on deep learning and their latest progress are introduced. Finally, the loop closure detection methods based on deep learning is summarized and its development direction is prospected.**

***Keywords*—Mobile robot; SLAM; loop closure detection; deep learning; convolutional neural network**

I. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) is a robot that moves in an unknown environment, relies on its own sensors for localization, and gradually draws a map of the environment. In recent years, with the improvement of Graphics Processing Unit (GPU) computing power and the continuous development of computer vision technology, visual SLAM has attracted extensive attention from academia and industry due to its rich visual information and low price compared with lidar [1].

Loop closure detection plays a crucial role in SLAM systems, which can solve the problem of location estimation drifting over time and eliminate accumulated errors to ensure the reliability of map construction, as shown in Figure 1.Although the back-end optimization can reduce the influence of the wrong closed-loop, a correct closed-loop can significantly reduce the cumulative error of the system, and a wrong loop closure may cause the back-end optimization algorithm to converge to completely wrong values. Therefore, relying on back-end optimization cannot fundamentally eliminate the influence brought by the wrong closed-loop, and how to correctly detect the closed-loop is the key. However, because loop closure detection provides the correlation between current data and all historical data, the large amount of data and strict evaluation criteria make it a difficulty for SLAM.

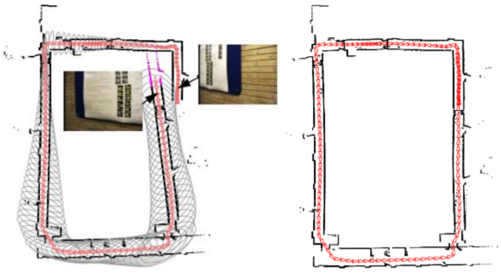

Fig.1. Correction results of cumulative error and loop closure detection

This paper mainly introduces the loop closure detection of visual SLAM for mobile robots. By introducing the common methods and the latest progress of loop closure detection, the relevant difficulties and solutions are discussed.

II. LOOP CLOSURE DETECTION AND BAG OF WORDS

Loop closure detection refers to the robot judging whether it has been to this position before and whether the trajectory has formed a closed-loop according to the current observation information and all historical observation information of its sensor. In visual SLAM system, robots perform this task according to images, which is similar to image retrieval in computer vision [2].

Abstract description of image and similarity calculation are the key techniques of loop closure detection. In traditional methods, researchers rely on artificial features to describe an image. These features are divided into local features and global features. Local features include SURF, ORB, SIFT, etc. Bag of words (BoW) methods use local features to construct a dictionary, and the whole image can be described by words. The image is converted into a sparse vector, which greatly reduces the subsequent use of Term Frequency-Inverse Document Frequency (TF-IDF) is the calculation amount of similarity calculation; Global features include GIST, Vector of localized Aggregated Descriptors (VLAD), and Fisher Vector (FV), etc. to describe the features of the whole image with different calculation methods [3].

Adrien Angeli [4] et al. not only used SIFT descriptor, but also applied the color histogram method to extract image features. Based on the assumption that each feature space is independent of each other, they also use Bayesian formula to update the occurrence probability of loop closure detection and BoW dictionary simultaneously based on the input image.

When the loop closure detection probability is high enough, the camera motion is judged to be continuous and consistent by combining the results of front-end geometric operation, so as to determine whether the current motion trajectory has a loop. In this method, BoW is updated in real time for the first time, making dictionary building an online process and making loopback detection better match the current image environment. The FAB-MAP developed by Cummins [5] et al., uses a different bag of words model. They directly use simple image blocks as the characteristic "words" of bag of words, which is called bag of visual words (BoVW). In their method, the whole input image is described as a series of fixed size image block "words" set, so as to omit the calculation and quantification process of artificial image features. In the SeqSLAM method proposed by Milford [6] et al., image feature extraction is not limited to the single image currently input, but adopts the time-domain continuous image sequence input in a past period of time to make correlation matching, so as to achieve loop closure detection.

The accuracy and recall rate of traditional loop closure detection methods will be significantly reduced due to the influence of illumination change, dynamic scene and angle change. Deep learning is applied to loop closure detection because it can automatically and comprehensively extract multiple features of images, which makes up for the shortcomings of traditional methods and improves the robustness.

III. DEEP LEARNING AND LOOP CLOSURE DETECTION

In recent years, with the development of deep learning , good progress has been made in image classification, object detection, semantic segmentation and other fields. More and more researchers begin to focus on how to apply deep learning to loop closure detection to achieve better results[7].

Arandjelović [8] et al. used neural network to fit artificial features. They selected VLAD, an image local feature aggregation descriptor with good performance in scene recognition tasks, to directly solve the recognition tasks in an end-to-end manner. AlexNet and VGG16 were used as the backbone convolutional neural networks(CNN) respectively, and a NetVLAD layer was connected in series to fit the generation process of original VLAD image descriptors. They rewrote the expression form of VLAD feature output, using Softmax function to rewrite it, so that all parts of the network can be trained. The author combined the training data into triples and trained the network in the way of weakly supervised learning, thus reducing the training difficulty. By adjusting the structure of CNN, this method makes it easier for the network to extract the features of static objects such as buildings, and makes the descriptors generated by the network more suitable for loop closure detection.

Merrill N [9] et al. proposed CALC framework, which is a lightweight loop closure detection framework. In this method, a random projection transformation is performed on the image to simulate the perspective transformation of the robot, and then the HOG descriptor is fitted by the self-coding network model to measure the similarity of the image by Euclidean distance. This method has geometric information, illumination invariance and high speed, and can reliably perform real-time loop closure detection without reducing dimension.

Tang [10] et al. combined RNN (Recurrent Neural Network) with CNN to form a GCN network, and the key points and descriptors extracted from this network can be used for pose update and loop closure detection. Image pairs are used in the training of GCN network, that is, the corresponding images of two adjacent frames in SLAM, and the feature points are Harris corner points. Although the method achieves better results than related deep learning and manually designed features, it only runs on GPU. In order to solve the real-time problem, Tang [11] et al. released GCNv2 next year, which can run in real time on Jetson Tx2. GCNv2 is running faster but with the same accuracy and performance as the original network, and the improved feature points have the same binary format as ORB, making it easy to replace features in ORB-SLAM. In order to shorten the operation time, ResNet50, the convolutional neural network originally used for feature extraction, was replaced by SuperPoint [12] network, and the network was clipped, and then the original cyclic neural network was removed. Experimental results show that GCNv2 has higher accuracy than ORB, SIFT, SURF and other traditional methods, and better accuracy than SuperPoint. GCNv2 based GCN SLAM has better accuracy than ORB SLAM in some scenarios, but ORB SLAM loses feature points in some scenarios, so GCNv2 has better robustness.

Li [13] et al. used HF-Net [14] to fit image features. The image first went through a shared encoder and then entered three parallel decoders to obtain key point detection scores, local descriptors and global descriptors. The first two decoders have the same structure as SuperPoint and the third decoder has the same structure as the NetVLAD layer. This method is robust and efficient by using Intel OpenVINO tool kit for feature fitting and Fast BoW for loop closure detection. It is the first method based on deep features that only needs CPU to run in real time. The DX SLAM framework based on this approach is similar to ORB SLAM2.

Zhang [17] et al. extracted image features through the pre-training model OverFeat on ImageNet and obtained a high-dimensional vector. In order to significantly improve their image expression ability and make calculation more efficient, principal component analysis (PCA) and whitening were used to reduce the dimensionality of the vector, and euclide distance was calculated between the obtained vectors. The similarity score was calculated to judge the image similarity. In the open data set, the recall rate is higher than FAB-MAP. Bai [18] et al. proposed a loop closure detection by combining the output of the middle layer of AlexNet based on Places pre-training with the output of traditional sequence matching. Hu [19] et al. proposed a loop closure detection method integrating semantic information based on the Faster R-CNN model. The semantic information of images is extracted by the pre-training model on CoCo dataset, the cosine distance between them is calculated and fused with the similarity weighted fusion of feature points based on bag of words method for loop closure detection.

IV. SUMMARY AND PROSPECT

With the development of computer vision and deep learning, loop closure detection methods have also made good progress. Since neural network models are mostly applied to image classification, existing networks need to be adjusted when applied to loop closure detection. In order to fit the artificial features, the existing networks are usually redesigned or combined. Using neural network to extract image features usually select a certain layer as the output, or the output of a certain layer is weighted fusion. In order to improve the precision of loop closure detection, deep learning methods are often combined with traditional methods such as sequence matching and bag of words method. In order to improve real-time performance, it is generally necessary to combine the methods of data reduction and image retrieval.

Although the loop closure detection methods based on deep learning can effectively deal with illumination and perspective changes compared with traditional methods, how to generate better image description has always been a problem to be solved. Deep learning methods rely on data sets, and how to effectively use training data and optimize the training process to improve the generalization ability of the model is also a direction that can be further studied. In view of the time consuming problem of feature extraction by neural network, methods such as parameter pruning and knowledge distillation can be considered to compress the model.

Visual SLAM puts forward new requirements for deep learning theory, and the development of deep learning theory will in turn promote the development of visual SLAM.

REFERENCES

[1] Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., et al, "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age," in IEEE Transactions on Robotics, vol. 32, no. 6, pp. 1309-1332, Dec. 2016.

[2] LIU Qiang, DUAN Fuhai, SANG Yong, ZHAO Jianlong, A Survey of Loop-Closure Detection Method of Visual SLAM in Complex Environments[J]. ROBOT, 2019, 41(1): 112-123,136.

[3] Y. Chen, Y. Zhou, Q. Lv and K. K. Deveerasetty, "A Review of V-SLAM*," 2018 IEEE International Conference on Information and Automation (ICIA), 2018, pp. 603-608.

[4] A. Angeli, D. Filliat, S. Doncieux and J. Meyer, "Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words," in IEEE Transactions on Robotics, vol. 24, no. 5, pp. 1027-1037, Oct. 2008.

[5] M. Cummins, P. Newman. "FAB-Map: Probabilistic localization and mapping in the space of appearance," Int. J. Robot. Res, vol.27(6), 2008, pp.647-665.

[6] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 1643-1649.

[7] ZHAO Yang, LIU Guoliang, TIAN Guohui, LUO Yong, WANG Ziren, et al, A Survey of Visual SLAM Based on Deep Learning[J]. ROBOT, 2017, 39(6): 889-896.

[8] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla and J. Sivic, "NetVLAD: CNN Architecture for Weakly Supervised Place Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1437-1451, 1 June 2018.

[9] Merrill N , Huang G . Lightweight Unsupervised Deep Loop Closure[C]// 2018. https://arxiv.org/pdf/1805.07703.pdf

[10] J. Tang, J. Folkesson and P. Jensfelt, "Geometric Correspondence Network for Camera Motion Estimation," in IEEE Robotics and Automation Letters, vol. 3, no. 2, pp. 1010-1017, April 2018.

[11] J. Tang, L. Ericson, J. Folkesson and P. Jensfelt, "GCNv2: Efficient Correspondence Prediction for Real-Time SLAM," in IEEE Robotics and Automation Letters, vol. 4, no. 4, pp. 3505-3512, Oct. 2019.

[12] D. DeTone, T. Malisiewicz and A. Rabinovich, "SuperPoint: Self-Supervised Interest Point Detection and Description," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 337-33712.

[13] D. Li, Xuesong Shi, Qiwei Long, Shenghui Liu, Wei Yang et al., "DXSLAM: A Robust and Efficient Visual SLAM System with Deep Features," 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020, pp. 4958-4965.

[14] P. Sarlin, C. Cadena, R. Siegwart and M. Dymczyk, "From Coarse to Fine: Robust Hierarchical Localization at Large Scale," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 12708-12717.

[15] X. Zhang, Y. Su and X. Zhu, "Loop closure detection for visual SLAM systems using convolutional neural network," 2017 23rd International Conference on Automation and Computing (ICAC), 2017, pp. 1-6.

[16] Dongdong BAI, Chaoqun WANG, Bo ZHANG, Xiaodong YI, Xuejun YANG ,"CNN Feature Boosted SeqSLAM for Real-Time Loop Closure Detection." Chinese Journal of Electronics v.27.03(2018):48-59.

[17] M. Hu, S. Li, J. Wu, J. Guo, H. Li and X. Kang, "Loop Closure Detection for Visual SLAM Fusing Semantic Information," 2019 Chinese Control Conference (CCC), 2019, pp. 4136-4141.