International Core Journal of Engineering Volume 6 Issue 10, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202010_6(10).0033

Research on Closed Loop Detection Algorithm

Min Liao, Guangping He and Li Zhang

Chengdu University of Technology, Chengdu 610059, China.

Abstract

With the rapid development of computer and artificial intelligence technology, technologies such as mobile robots, unmanned driving, and virtual technology have been widely used in various fields of people's lives. Among them, the simultaneous localization and mapping (SLAM), which belongs to the field of robotics, has become a major research focus in recent years. This article focuses on the three main problems of visual SLAM closed loop detection, image description, main methods and future outlook. First, introduce the descriptors and feature points in image description, and analyze their advantages and disadvantages. Secondly, it outlines two commonly used methods in loop detection algorithms, based on the bag of words model and deep learning, expounds their advantages and disadvantages, and introduces the latest research progress. Finally,Combined with the current research on loop detection, make an outlook on the development of loop detection in the future.

Keywords

SLAM, Closed Loop Detection, Bags Of Word, Deep Learning.

1. Introduction

Simultaneous Localization and Mapping (SLAM) is a robot that incrementally creates a continuous map of the surrounding environment in the location environment, and uses the created map to navigate itself.

Visual SLAM (vSLAM) is a SLAM technology that uses visual sensors. With the rise of visual sensor technology and the enhancement of computer computing power, vSLAM technology has also been progressing day by day. It has been applied in various fields and has excellent performance in indoor, outdoor and underwater environments. In vSLAM, loop detection is an important modules, what we hope to get is an estimate of where loopbacks may occur. This article focuses on the problem of vision-based loop detection. Loopback is an important basis for mapping and back-end optimization. Accurate loop detection can ensure the correctness of vSLAM trajectory estimation and map construction during long-term operation. In the case of loss of tracking algorithm, loop detection can also be relocated. Closed loop detection has a huge impact on the performance of vSLAM. If the closed loop detection can not be detected in a timely and correct manner, the backend optimization module and mapping module will be affected, resulting in low mapping accuracy and even the robot losing position.

Early closed loop detection methods are mostly based on the assumption of appearance invariance. These methods can still operate normally in a stable indoor environment, but face long-term autonomous navigation tasks in outdoor environments, such as changes in illumination, seasonal changes, dynamic scenes, and changes in perspective. It will greatly reduce the detection accuracy and recall rate. This article focuses on the three main problems of visual SLAM closed loop detection: iamge description, main methods of closed loop detection and the future development of loop detection.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202010_6(10).0033

2. Image Description

In this section, we mainly discuss how to reasonably describe the scene. In SLAM, we always judge the pose of the camera based on the image, and judge whether the camera has reached a certain position. The image sequence represents a series of locations on the map. Therefore, we need to describe the scene that is converted to describe the image. In SLAM, we mainly rely on feature points and descriptors to describe the image.

2.1 Global descriptor

The global descriptor is to directly calculate the descriptor of the entire image. These descriptors usually calculate the velocity block, which can simplify matching and reduce the calculation of mapping and positioning tasks. An existing widely used global descriptor is the Gist descriptor proposed by Oliva et al [1]. It was originally mainly for scene recognition. It uses Gabor filters to extract image information in different directions and frequencies and compress it into a vector to get the image description of Krose [2] directly used PCA (principal component analysis) dimensionality reduction method to generate linear image features, and then used the features to establish an observation model based on Gaussian distribution. Ulrich [3] et al. used the histogram of panoramic color pictures combined with nearest neighbor learning to perform image matching and so on. R.Arroyo [4] proposed a loop detection technology based on global descriptors, which tends to match sequences rather than single images, and shows strong adaptability to changes in factors such as illumination In addition, there are many good descriptors based on deep learning, which are not listed here.

2.2 Local image features

Global descriptors use the entire image content to describe the image. It can capture the overall structure of the scene very well, but they cannot handle some visual problems, such as partial occlusion of lighting changes. These existing problems can be solved by local image features, where the local image features are also key points. The following mainly introduces several commonly used local feature points and descriptors.

2.2.1 SIFT

Scale-Invariant Feature Transform (SIFT) is an algorithm developed by Lowe [5] to detect and describe distinctive keypoints in images which was originally created for object recognition. The SIFT feature is a local feature of the image, which maintains invariance to rotation, scale scaling, and brightness changes, and maintains a certain degree of stability to viewing angle changes, affine transformations, and noise. The main process of the algorithm: First, the original image and the Gaussian kernel are convolved at different scales to generate the Gaussian difference pyramid (DOG pyramid), and then find the extreme points of the DOG function, which is the spatial extreme point detection, and then stabilize the key points Accurate positioning and direction information distribution, and finally generate feature points. SIFT has scale invariance, but the huge amount of calculation takes a long time.

2.2.2 SURF

The pyramid constructed by SURF is different from SIFT. SIFT uses DOG images, and the second SURF uses Hessian matrix determinant approximation images which can accelerate the detection speed. Then use non-maximum value suppression to preliminarily determine the feature points, use 3D linear interpolation to obtain sub-pixel feature points, accurately locate the extreme points, and finally select the main direction of the feature points to construct the SURF feature point descriptor. SURF has scale-invariant features and high computational efficiency.

2.2.3 FAST

Features from Accelerated Segment Test (FAST) is a corner detector proposed by Rosten and Drummond [6].FAST's detection process: Firstly,mainly choose a pixel p in the image arbitrarily, see Figure 1.With p as the center, 3 pixels as the radius, 16 pixels are selected, and a brightness difference

ISSN: 2414-1895 DOI: 10.6919/ICJE.202010_6(10).0033

threshold t is defined. In order to speed up the inspection, first calculate the brightness difference between the four pixels on the top, bottom, left, and right of p and p. If three of the absolute differences are greater than the threshold t, then point p is the corner point. For each pixel, calculate the difference of all pixels in the field. If the absolute value of the difference is greater than 9 (mainly by your own choice, the algorithm here is called FAST-9) is greater than the threshold t, the center pixel is at the corner , Otherwise it is not a corner point. FAST detects the speed block, but the accuracy is not very accurate and does not have rotation invariance.

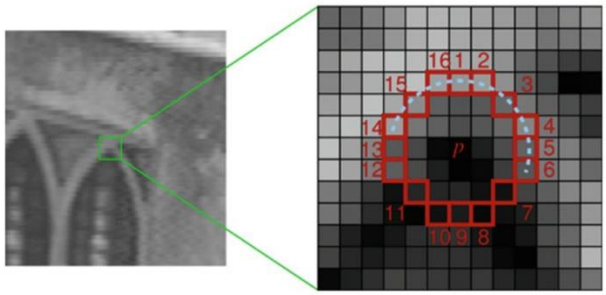

Figure 1. FAST corner detection

2.2.4 BRIEF

BRIEF is proposed in a 2010 article named "BRIEF:Binary Robust Independent Elementary Features". BRIEF is a description of the detected feature points. It is a binary-coded descriptor and discards the use of regional gray. The traditional method of describing feature points by degree histogram greatly accelerates the establishment of feature descriptors, and also greatly reduces the time of feature matching. It is a very fast and promising algorithm. BRIEF is a binary descriptor whose description vector consists of many 0s and 1s. Here 0 and 1 encode the size relationship of two pixels near the key point (such as p and q): if p is smaller than q, take 1; otherwise, take 0. Brief has the advantages of fast speed, convenient storage and high quality. The disadvantage is that it is sensitive to noise, not several times the rotation invariance and scale invariance.

2.2.5 ORB

ORB (Oriented FAST and Rotated BRIEF) algorithm is a fast image feature extraction algorithm proposed by Ethan Rublee et al [7].The algorithm is divided into two parts: feature point extraction and feature point description. oFAST mainly defines the direction at the FAST feature points and has rotation invariance. rBRIEF mainly adds a rotation factor based on the BRIEF descriptor, so that the feature descriptor has rotation invariance.

3. Common Methods Of Closed Loop Detection

There are generally two implementations of loopback detection: Odometry Based and Appearance Based. Based on the odometer, the camera's position is mainly used to determine whether to reach the previous position, but it cannot give accurate results due to the existence of accumulated errors. The appearance-based loop detection is based on the similarity between images. It has no direct relationship with the previous pose and can give more accurate results. It has become the mainstream method of VSLAM. There are currently two main types based on appearance: loop detection based on bag of words model and loop detection based on deep learning.

3.1 Bags Of Words

The bag-of-words model is proposed by D.Nister [8]. It constructs a word by extracting feature points from each frame, then treating the descriptors of the feature points as words, and then classifying all the feature descriptors through cluster analysis. Bag model, for each frame a "sentence" described by the bag of words model is established. Then the new key frame is compared with the bag of words model to match the closest "sentence". This method can effectively establish closed loop detection in

ISSN: 2414-1895 DOI: 10.6919/ICJE.202010_6(10).0033

SLAM in larger-scale scenarios. At the same time, according to the different construction process of vocabulary, it is mainly divided into offline and online.

3.1.1 Offline

The vocabulary tree used in the Bow model is usually generated offline. In the offline category, M. Cummins [9] proposed that the FAB-MAP algorithm and its extension FAB-MAP 2.0 is a more critical algorithm, which uses a Chow-Liu tree to approximate the probability of visual words. In 2012, Galvez-Lopez and Tardos [10] trained a visual vocabulary based on the improved BRIE descriptor . Based on these vocabularies, they proposed a closed loop detection method based on the concept of island. Similar images are grouped in real time, and the image sequence is divided into fixed-size intervals to prevent images with similar appearances from competing with each other. Mur-Artal and Tardos [11] enhanced the D.Galvez-Lopez algorithm by using ORB, which is more robust to scaling and rotation changes. L.Bampis [12] proposed an extension of the bag-of-words model. For two consecutive frames, visual words with similar optical flow are grouped. These groups are called structure-aware groups and view-invariant higher-order visual words (Structure-Aware). and Viewpoint-Invariant High-Order Visual-Words), they naturally include the environmental structure into the image description. Rihem El Euch [13] proposes a loop detection algorithm based on superpixels ,mainly combining color, texture, and structure to construct a descriptor, and at the same time introduces the concept of dynamic islands, allowing us to group image places in time to avoid images from the same place Compete with each other as loop candidates.

3.1.2 Online

Angeli et al [14] proposed an incremental bag-of-words model scheme, and used the Bayesian filtering framework to detect loops. Inspired by the work of Angeli, Labbe and Michaud presented a solution called Real-Time Appearance-Based Mapping (RTAB-Map) [15, 16] a loop closure detection approach for large-scale and long-term SLAM. The main contribution of this solution was that they provided memory management mechanis msforcaching a subset of the online learnt visual words in the main memory (called Working Memory), and this subset was used for detecting loop closures. The rest were stored in a database stored in an external memory called Long Term Memory. The transition of words between memories was ruled by the time taken for processing images in an adaptive way. In addition, Khan et al [17] proposed an incremental bag-of-words model algorithm based on binary descriptors in 2015, but its disadvantage is that it does not use an effective search scheme, which reduces the scalability of the algorithm. Based on the algorithm of D.Galvez-Lopez, Emilio Garcia-Fidalgo et al. [18] proposed to allow different types and dynamic sizes of islands to meet online requirements. Nicosevici and Garcia [19, 20] introduced Online Visual Vocabulary (OVV), where the words were generated at the same time that the robot was exploring the environment using a modified version of a nagglomerative clustering algorithm.Sheraz Khan et al. [21] resents an appearance based loop closure detection mechanism titled IBuILD (Incremental bag of BInary words for Appearance based Loop closure Detection), The presented approach focuses on an online, incremental formulation of binary vocabulary generation for loop closure detection. The proposed approach does not require a prior vocabulary learning phase and relies purely on the appearance of the scene for loop closure detection without the need of odometry or GPS estimates. The vocabulary generation process is based on feature tracking between consecutive images to incorporate pose invariance.

3.2 Deep Learning

In recent years, deep learning and Convolutional Neural Networks (CNN) have developed rapidly, bringing huge breakthroughs to the development of the computer field. Sünderhauf [22] and R. Arandjelovic [23] and others proposed CNN-based solutions, which have become an effective method to combat environmental changes. Sünderhauf and others have made a pioneering work to evaluate the effect of ConvNets in location recognition. They combined object proposal technique and CNN features to match the location of extreme appearance changes. Arroyo et al. [24] proposed a method of fusing different convolutional layer information for topological positioning. In a recent work.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202010_6(10).0033

Arandjelovic et al. [23] introduced a CNN architecture, which is mainly based on a layer in the VLAD image representation for weakly supervised location recognition. Through reasonable selection of training data, there are changes in perspective and appearance between a group of pictures corresponding to the same location, so that the descriptors generated by the final training network can adapt to changes in environmental conditions. Ability. Mingyue Hu [25] proposes a closed loop detection algorithm that integrates semantic information, which mainly realizes semantic similarity based on the bag-of-words model. It mainly uses FAST R-CNN to extract, segment, obtain semantic information, and then merge with the similarity of feature points, and use the fused similarity Degree judgment closed loop.

4. Development Directions

Visual SLAM has always been a hot research topic in recent years, and deep learning also shows more prominent advantages in image processing. Many SLAM problems are deeply related to deep learning, such as loop detection. In the future development, whether it is a two-dimensional image or a three-dimensional point cloud, the amount of data will increase as the scale of the scene becomes larger. When it comes to compression and extraction, Parisotto compresses the relative pose information through the pose aggregation method, and then relies on the network. The output of the relative pose and the global absolute pose, extract key information by extracting the Soft Attention model, generate a similarity matrix between frames, and use the similarity matrix to complete the closed loop detection in SLAM, which proves that the Soft Attention model is in closed loop detection effect. In the future, SLAM also needs to compress the perceptual information and divide the search space to complete loop detection. At the same time, under the influence of deep learning, the target detection technology has also made great developments, which will greatly help the cognitive ability and adaptability of the scene in dynamic scenes, and the closed loop detection in the future. At the same time, deep learning can also be used for semantic segmentation to extract image semantics, so that only the ability of the algorithm to detect loops in the same scene from different perspectives is required to be compared, which is very helpful for closed loop detection on more advanced features. In short, the combination of deep learning has a lot of room for development in the future.

5. Conclusion

As an important part of visual SALM, closed loop detection has achieved many results under the development of computer vision and deep learning. There are still problems with the two mainstream loop detection methods. The loop detection algorithm based on the bag-of-words model has limitations in dynamic environment, perceptual aliasing, and redundant information. Although many existing algorithms have been improved, there are still some limitations. The loop detection algorithm based on deep learning relies on a large number of labeled data sets, real-time performance, scalability, etc., which require further research and discussion. In short, visual SLAM still has great research value in the future, and it needs to continue to work hard to solve the closed loop detection of complex environments.

References

[1] Oliva A, Torralba A. Building the gist of a scene: The role of global image features in recognition[J]. Progress in Brain Research, 2006, 155(2): 23-36.

[2] Krose B J A, Vlassis N, Bunschoten R, et al. A probabilistic model for appearance-based robot localization[J]. Image & Vision Computing, 2001, 19(6): 381-391.

[3] Lowry S M, Wyeth G F, Milford M J. Unsupervised online learning of condition-invariant images for place recognition[J]. Procedia – Social and Behavioral Sciences, 2014, 106: 14181427.

[4] Arroyo R, Pablo Fernández Alcantarilla, Bergasa L M, et al. Fusion and Binarization of CNN Features for Robust Topological Localization across Seasons [C]// IEEE/RSJ.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202010_6(10).0033

[5] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004).

[6] Rosten, E.,Drummond,T.:Machinelearningforhigh-speedcornerdetection.Eur.Conf.Comput. Vision, 430–443 (2006).

[7] Rublee E, Rabaud V, Konolige K, et al. 2011. ORB: An efficient alternative to SIFT or SURF[C]. international conference on computer vision, 2564-2571.

[8] David Nistér, Henrik Stewénius. Scalable Recognition with a Vocabulary Tree[C]// Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006.

[9] Cummins M, Newman P. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance[M]. Sage Publications, Inc. 2008.

[10] Galvez-LoPez D, Tardos J D. Bags of Binary Words for Fast Place Recognition in Image Sequences[J]. IEEE Transactions on Robotics, 2012, 28(5):1188-1197.

[11] Raúl Mur-Artal, Juan D Tardós. Fast Relocalisation and Loop Closing in Keyframe-Based SLAM[J]. Proceedings - IEEE International Conference on Robotics and Automation, 2014:846-853.

[12] Bampis L, Amanatiadis A, Gasteratos A. High order visual words for structure-aware and viewpoint-invariant loop closure detection [C]// IEEE/RSJ International Conference on Intelligent Robots & Systems. IEEE, 2017.

[13] R. El Euch, E. Garcia-Fidalgo, A. Ortiz, F. Chaabane and A. Ghazel, "Superpixel Description and Indexing for Visual Loop Closure Detection," 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 2019, pp. 1591-1594, doi: 10.1109/ ETFA. 2019. 8869091.

[14] Angeli, A., Doncieux, S., Meyer, J.A., Filliat, D.: Incremental vision-based topological SLAM. IEEE/RSJ Int. Conf. Intell. Robot. Syst. 1031–1036 (2008).

[15] Labbe, M., Michaud, F.: Memory management for real-time appearance-based loop closure detection. IEEE/RSJ Int. Conf. Intell. Robot. Syst. 1271–1276 (2011).

[16] Labbe, M., Michaud, F.: Appearance-based loop closure detection for online large-scale and long-term operation. IEEE Trans. Robot. 29(3), 734–745 (2013).

[17] Khan S, Wollherr D. 2015. IBuILD: Incremental bag of Binary words for appearance based loop closure detection[C]. international conference on robotics and automation, 5441-5447.

[18] Garcia-Fidalgo E, Ortiz A. iBoW-LCD: An Appearance-based Loop Closure Detection Approach using Incremental Bags of Binary Words[J]. IEEE Robotics & Automation Letters, 2018.

[19] Nicosevici, T., Garcia, R.: On-line visual vocabularies for robot navigation and mapping. IEEE/RSJ Int. Conf. Intell. Robot. Syst. 205–212 (2009).

[20] Nicosevici, T., Garcia, R.: Automatic visual bag-of-words for online robot navigation and mapping. IEEE Trans. Robot. 28(4), 886–898 (2012).

[21] S. Khan and D. Wollherr," IBuILD: Incremental bag of Binary words for appearance based loop closure detection," 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, 2015, pp. 5441-5447, doi: 10.1109/ICRA.2015.7139959.

[22] Sünderhauf, Niko, Dayoub F, Shirazi S, et al. On the Performance of ConvNet Features for Place Recognition[J]. 2015.

[23] Arandjelovic R, Gronat P, Torii A, et al. NetVLAD: CNN architecture for weakly supervised place recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017:1-1.

[24] Arroyo R, Pablo Fernández Alcantarilla, Bergasa L M, et al. Fusion and Binarization of CNN Features for Robust Topological Localization across Seasons[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016.

[25] Mingyue Hu, Sheng Li, Jingyuan Wu, Jiawei Guo, Haiyu Li,Xiao Kang. Loop Closure Detection for Visual SLAM Fusing Semantic Information[C], 2019:1084-1089.