EasyChair Preprint

№ 11134

Semantic-Pixel Associative Information Improving Loop Closure Detection and Experience Map Building for Efficient Visual Representation

Deng Yufei, Xiao Rong, Jiaxin Li and Lv Jiancheng

October 23, 2023

Semantic-pixel Associative Information improving Loop Closure Detection and Experience Map Building for Efficient Visual Representation *

Yufei Deng[1], Rong Xiao[1,2], Jiaxin Li[1,3], and Jiancheng Lv[1,2]

[1] College of Computer Science, Sichuan University, Chengdu, China
[2] Engineering Research Center of Machine Learning and Industry Intelligence, Ministry of Education, Chengdu, China
[3] PetroChina Southwest Oil and Gasfield Company
[dengyufei, lijiaxin1993]@stu.scu.edu.cn
[rxiao, lvjiancheng]@scu.edu.cn

Abstract. RatSLAM is a brain-inspired simultaneous localization and mapping (SLAM) system based on the rodent hippocampus model, which is used to construct the experience map for environments. However, the map it constructs has the problems of low mapping accuracy and poor adaptability to changing lighting environments due to the simple visual processing method. In this paper, we present a novel RatSLAM system by using more complex semantic object information for loop closure detection (LCD) and experience map building, inspired by the effectiveness of semantic information for scene recognition in the biological brain. Specifically, we calculate the similarity between current and previous scenes in LCD based on the pixel information computed by the sum of absolute differences (SAD) and the semantic information extracted by the YOLOv2 network. Then we build an enhanced experience map with object-level information, where the 3D model segmentation technology is used to perform instance semantic segmentation on the recognized objects. By fusing complex semantic information in visual representation, the proposed model can successfully mitigate the impact of illumination and fully express the multi-dimensional information in the environment. Experimental results on the Oxford New College, City Center, and Lab datasets demonstrate its superior LCD accuracy and mapping performance, especially for environments with changing illumination.

Keywords: Brain-inspired simultaneous localization and mapping · Loop closure detection · Semantic information.

1 Introduction

Simultaneous Localization and Mapping (SLAM) localizes a robot and builds a map as it explores an unknown environment [1]. This map defines the robot's orientation by pose (position and orientation). Recently, the literature has grown

* Corresponding author: Rong Xiao.

up around introducing mathematical probability filtering into SLAM algorithms, such as Kalman Filter (KF), Particle Filter (PF), and their improved algorithms [2]. Since only considering the recursive relationship before the data, it is easy to cause error accumulation problems and map inconsistencies. In large-scale and complex environments, it is a big challenge for the mapping performance of the algorithm. To settle this problem, brain-inspired SLAM models have become a promising alternative method for robot spatial cognition development by transforming neuroscience research into engineering solutions [3–5]. Among them, the rodent-inspired simultaneous localization and mapping algorithm (RatSLAM) has achieved promising performance by emulating the spatial awareness of the hippocampal system [6]. RatSLAM works indoors and outdoors and requires only a monocular vision sensor. However, multiple problems still need to be solved before the RatSLAM model could be used as a practical solution in complex environments. Specifically, Loop closure detection (LCD) is the critical process for robots to relocalize themselves and correct accumulative errors. The cognitive map constructed by RatSLAM is composed of many connected experiences, which describe the spatial structure of the environment. Recently, many different methods have been proposed for detecting correct loop closure and building experience map accurately [7–11]. Although the existing methods can achieve effective LCD and experience map building in RatSLAM systems, they often result in relatively low efficiency and poor robustness to lighting changes by only using pixel information in template matching.

An emerging trend seeks help from semantic information for more accurate LCD. Imerge semantic information with pixel information not only increases the computational depth of the LCD but also enhances the brain-inspired feature of the entire system. People identify the place where they are not only spatially, but also conceptually [12]. Semantic information is a higher-level concept obtained through vision, which can successfully provide landmarks for the map. Although semantic information enables scene recognition and human-intelligible map building, it is still lacking in LCD and experience map building in RatSLAM. Further research on them may help us encode information perceived by humans and refine brain-inspired SLAM.

In this paper, inspired by the effectiveness of semantic information in the brain, semantics is added to LCD and experience map building in RatSLAM. The view template records the experienced scene. We use YOLOv2 to extract the object information from the vision, and store them in the view template together with the pixel information. For template matching in LCD, we calculate the template similarity by fusing the semantic similarity computed by the number and position of recognized objects with the pixel similarity computed by SAD. Comparing the template similarity with the threshold detects whether a closed loop is generated, the detected closed loop will correct the experience map. To improve the legibility of the map, we project the recognized objects to the corresponding positions on the experience map through 3D model segmentation technology, thus constructing a map with object-level information. By detecting semantic information, the proposed model can successfully mitigate the impact of illumination and fully express the multi-dimensional information in the environment. The experiment results show that with the fusion of semantic information, our system has higher LCD accuracy and more stable mapping performance, especially for environments with changing illumination. The main contributions of this paper can be summarized as:

- For increasing the representation capacity of the system, we store the object information recognized from the visual scene into the view templates, and fuse it with the pixel information for template matching in LCD;
- For improving the map display capability and establishing a multi-dimensional experience map, we project objects identified in the scene onto the map through 3D model segmentation technology;
- We also have verified the advantages of the semantic-pixel associative representation method and comprehensively compared the impact of various parameters and lighting conditions on performance.

2 Related Work

In traditional RatSLAM system, LCD usually downscales and vectorizes camera image vectors and then compares them. Current research on LCD can be divided into two categories: improvement of LCD detection and visual processing. Gu et al. proposed to replace the templates search module in RatSLAM with the multi-index hashing-based loop closure detection (MILD) [7] algorithm, building a more accurate experience map [8]. Xu et al. used the Bag of Words (BoW) model and the dynamic island mechanism to achieve quick and efficient image retrieval [13]. In contrast, another type of LCD innovation lies in the processing of visual information. Zhou et al. used ORB as the feature when matching [9]. Accordingly, Kazmi et al. used the Gist descriptor as the feature matching method, which can also reduce the matching error [10]. To be closer to the biological model, Wu et al. improved the visual space of the image, and deal with the characteristics of image brightness and saturation from the perspective of the visual model [11]. It has gradually become a trend to extract visual information close to biological models. But it still has a long way to brain-inspired SLAM, and the application of semantics in it needs to be further studied.

Besides, there is only a single topology environment display method for map in RatSLAM. In traditional SLAM, multidimensional representation using semantics has become a hotspot. Nakajima et al. integrated object information to associate geometric map points with semantic labels [14]. Iqbal et al. inserted complete object instances into the map instead of classifying already obtained map elements [15]. Yet the semantic map for RatSLAM is still being explored.

3 Methodology

In this section, we will first review the RatSLAM and its limitations. The proposed novel SLAM system is then introduced. Finally, we will display how to integrate semantics in LCD and build a multi-dimensional experience map.

3.1 Revisiting RatSLAM

RatSLAM is composed of three main modules: pose cell, local view cell, and experience map. The local view cells store the view template in it [16]. If a new visual scene is encountered, a new local view cell would be created, and a link would also be built to connect this local view cell with the active pose cell. When a familiar scene is sensed, the corresponding local view cell would be activated and inject energy to the connected pose cell. We detect familiar scenes by calculating the similarity between the visual templates in LCD. The pose cells are represented the robot's perception of its current pose (x', y', θ'). Their dynamics are regulated by the path integration and the energy from local view cells. Finally, the experience map groups the information from local view cells and pose cells to represent the robot pose uniquely. Due to the alone visual information in LCD, conventional RatSLAM, despite being simple and effective, struggles with low precision and poor stability.

3.2 Semantic-pixel Associative RatSLAM

The overview of the proposed semantic-pixel association RatSLAM is presented in Fig. 1, which includes two major components: LCD and experience map building. Our system will use semantic information and pixel information in LCD to eliminate trajectory offset caused by accumulated errors. Here, YOLOv2 network is introduced to obtain semantic information. We then estimate the pose through pose cells, build an experience map fused with semantic information. 3D models will be projected in the corresponding position according to semantic information, making a comprehensive representation of the environment.

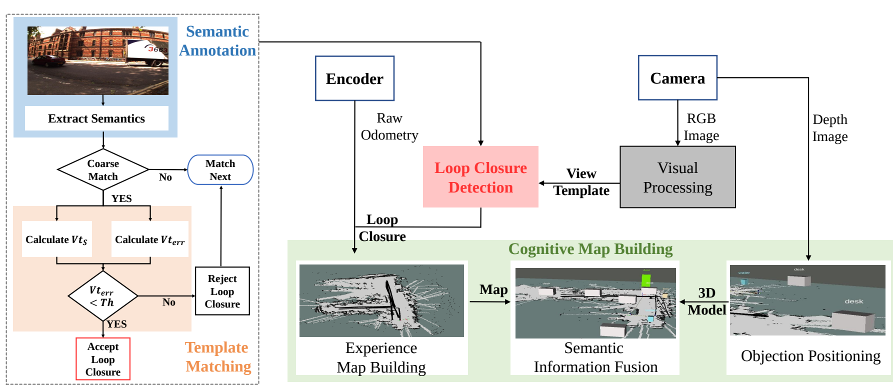


Fig. 1: The architecture of our semantic-pixel association RatSLAM system.

3.3 Integrate Semantic Information in LCD

To improve the accuracy of LCD, our system adds semantic information in conventional template similarity calculation. When a visual scene is an input, it is annotated by YOLOv2 and stored in the view template. In the LCD, semantic information is firstly used for rough matching, and templates that are less similar to the current scene are quickly excluded to reduce unnecessary calculations. Afterwards, we calculate the similarity between the current scene and the remaining templates and treat the template with the highest similarity as the matching template. In addition, we set it not to match with the recently saved templates to reduce the false closure caused by the short interval between two images. Our proposed LCD method mainly consists of two parts: (1) Semantic annotation part. We identify objects and store the semantic information and pixel information in the view template. (2) Template matching part. We compute template similarity using multiple levels of information. The architecture of the proposed method is shown in the LCD module in Fig. 1.

Semantic Annotation To perform stable and accurate LCD in complex scenes, it is necessary to perform object detection on RGB images and find out images' discriminative landmarks, including their categories and positions. YOLOv2 is a target algorithm introduced in the YOLO family. Its detection framework can be roughly divided into two parts: feature extraction network and action network. The feature extraction network is modified based on the DarkNet-19 network. The detection network consists of 4 convolutional layers, a transfer layer, and a detection layer [17]. In contrast to other YOLO detectors, YOLOv2 is relatively basic, but with better performance and a simple structure. It is proved by experiments that even with the basic detector, we can have better performance than traditional RatSLAM. By using the YOLOv2 network on the images, we obtain a set of labeled semantic regions $S = \{s_1, s_2, \ldots, s_n\},\ s_i = [type_i, x_i, y_i, w_i, h_i]$, where $type, x, y, w, h$ are the label category, center point coordinates, width, and height values respectively. The semantic information s_i of each image is stored in the label set S_k. To facilitate the use of semantic information in LCD, S_k of the image is added to the original visual template V_k.

Template Matching After semantic annotation, an image is abstracted into a semantic label set S. Taking the center of the image as the origin, we divide an image into four quadrants: upper left, lower left, upper right, and lower right. Any label s_i in S is classified into one quadrant through its center coordinates. Count the number of all regions according to the type of semantic label, and construct a scene descriptor $U = \left[Z^{lu}; Z^{ld}; Z^{ru}; Z^{rd}\right]$, $Z = [z_1, z_2, \ldots, z_n]$, where Z^{lu}, Z^{ld}, Z^{ru} and Z^{rd}represent the descriptor information of the four quadrants of a picture respectively, n is the number of semantic label categories, and z_i represents the number of a certain semantic tag in a quadrant.

After obtaining U, the algorithm calculates the similarity between the current image and the four quadrants of the template. For the current image a and the template b to be matched, according to its scene descriptors, we define the quadrant similarity calculation method as follows:

$$\zeta({Z_i}^{(a)}, {Z_i}^{(b)}) = \frac{\sum \omega \cdot |{Z_i}^{(a)} - {Z_i}^{(b)}|}{\sum \omega \cdot |{Z_i}^{(a)} + {Z_i}^{(b)}|} \tag{1}$$

where ${Z_i}^{(a)}$ is the descriptor of an image in the region i and w_j represents the similarity weight of j-type semantic labels. To avoid the impact of the few semantic labels on template matching, its w is set to 0. w is suppressed so that the greater the number of labels, the lower the similarity weight. The algorithm designs label similarity weight is as follows:

$$\omega_j = \begin{cases} 0, c_j \to 0 \\ 1 - \frac{c_j}{\sum_j^n c_j} \end{cases} \tag{2}$$

where c_j is the total number of labels of j-type. Sum the similarities of the four quadrants to obtain the semantic similarity between the image and the template:

$$Vt_s(a, b) = \sum_{i=1}^{4} \zeta({Z_i}^{(a)}, {Z_i}^{(b)}) \tag{3}$$

Combined with the Vt_s with Vt_p calculated by the SAD algorithm, the template similarity Vt_{err} between the image and the template is defined as follows:

$$Vt_{err} = \alpha \cdot Vt_s + \beta \cdot Vt_p \tag{4}$$

where α and β are the weight parameters of semantic similarity and pixel similarity, respectively. The most similar template to the current image, which with the highest similarity is preferentially selected as the best-matched closed-loop template. If a closed loop is detected, the system will correct the cognitive map, otherwise, a new template will be created.

3.4 Experience Map Building with Semantics

In this part, we project the recognized objects onto the map, displaying semantic information in the environment. Then we construct a topology map, 1displaying the physical information of the environment.

Objection Positioning When the objects are detected, we perform object instance segmentation by adopting primitive 3D shape priors. Then we reconstruct and find all 3D points inside the detected box, where the primitive model of the objects is fitted. The Euclidean region growing segmentation technique [18] is introduced as the clustering technique adopted in the shape model fitting, returning the centroid and respective convex hull dimensions of all classes except doors. Whenever doors are recognized, we fit a planar patch using RANSAC [19] for estimating the pose. The projected pose of each object, denoted by $y \in R^3$, is then represented by the 2D projected centroid from the camera coordinate system to the global map coordinate system and its orientation.

Experience Map Building Experience map in RatSLAM system is a semi-topology map composed of many experiences. When the system is running, the subsequent experiences are calculated based on the position of the previous experience and the robot's motion. When the closure is detected, the update procedure for experiences will be executed. Then the active pose cell and the local view cell will be reset to the matched experience. The map will be adjusted by the matched experience and a new experience will be created.

4 Experimental Analysis

In this section, we verify the proposed method from the following four perspectives. Firstly, LCD with semantic annotation is compared with traditional LCD in terms of Precision and Recall. Next, we discuss how the weight parameters of Vt_s and Vt_p affect the performance of our proposed method on specific datasets. We then evaluate the performance of the proposed LCD method facing noisy data. Finally, for showing the effectiveness of the entire RatSLAM system with our proposed method, the experiment is performed in real environment datasets.

4.1 Datasets

Experimental verification is performed using two datasets that are widely used to evaluate LCD algorithms because these datasets provide ground truth loop closures, which is convenient to measure the correctness of the results. Lab datasets are recorded to verify the effectiveness of multidimensional cognitive map construction with LCD. The datasets used in experiments are listed:

- Oxford New College (NC) dataset [20]. The left and right cameras each capture 1073 images, with 423 and 430 real closed loops respectively.
- City Center (CiC) dataset [20]. The left and right cameras each capture 1237 images, each with 561 real closed loops.
- Recorded Lab datasets. The two indoor datasets are gathered from a laboratory with images recorded by the RGB-D sensor, which is controlled to do semi-automatic environmental exploration. The sensor respectively captures 2893 and 1784 image pairs (RGB and depth).

4.2 Performance Comparison of LCD

Generally, the Precision-Recall (PR) curve is used to evaluate the performance of the LCD algorithm. Precision reflects the ability of the algorithm to detect correctly, and Recall reflects the ability to detect comprehensively. By changing the threshold of the similarity, different Precision and Recall can be obtained, thereby obtaining the PR curve. Fig. 2 shows the comparison of the PR curves obtained by our proposed method and the original Ratslam, DBoW2 [21], and SeqSLAM [22] algorithms for LCD on four datasets, including NC's, and CiC's left and right image datasets. For NC, we set the label type total is 17, while

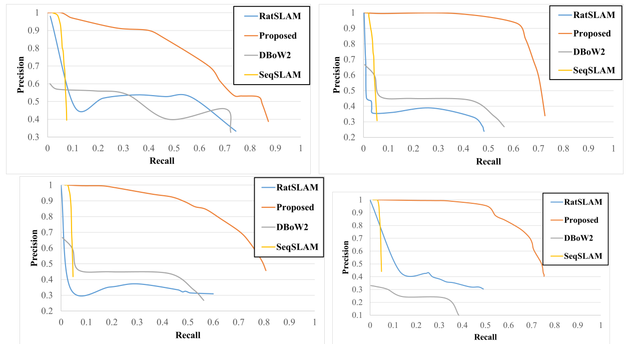


Fig. 2: Comparison results of the proposed LCD method, RatSLAM, DBoW2 and SeqSLAM on different datasets.

40 for CiC. α is set to 0.9, 0.8 0.9 0.8 for NC's, and CiC's left and right image datasets respectively, and β is set to 0.09 0.01 0.05 0.09 respectively.

Fig. 2 shows that as the Recall increases, the Precision of the method we propose drops slowly, while the others drop rapidly. When the Recall is 0.5, the Precision of our method can still maintain 0.9. This is because in these datasets, dynamic objects such as vehicles and pedestrians have great interference with traditional LCD algorithms, and their robustness is relatively poor in the face of similar structures and local changes. The proposed algorithm calculate the pixel and semantic similarity of different quadrants, which increases the amount of matching information, performs stricter similarity calculations according to areas, and improves the accuracy of the LCD.

4.3 Performance Trade-offs for LCD

To express the proposed method more clearly, we discuss the relationship between different α and β under the same Th. 9 values of α and β are tested in the NC's left and CiC's right camera image dataset respectively. We set $\alpha \in [0.1, 0.9]$ with an interval of 0.1, and set $\beta \in [0.01, 0.09]$ with an interval of 0.01. Under the same α condition, Fig. 3 indicates that the smaller the β value, the lower the Precision, and the higher the Recall. This is because the single semantics will generate a large number of false matches when the use of pixels in the image-matching process is greatly weakened. And it also shows that the greater the β, the increase of the Precision, and the decrease of the Recall. Since the SAD is sensitive to noise, it will match the image and the template too strictly. Therefore, when the use of pixels is greatly enhanced, closed loops that should be matched are not matched. When α is larger, the average performance of LCD is better, indicating that more semantic information is used in the process of template comparison, and better LCD results; when α is less than 0.5, the average performance of LCD crosses, and the difference not much, indicating that when the α is less than 0.5, its influence on the comprehensive similarity is lower than that of the pixel similarity, and the matching is mainly determined by the SAD. Therefore, the selection of parameters α and β requires a trade-off between Precision and Recall for LCD.

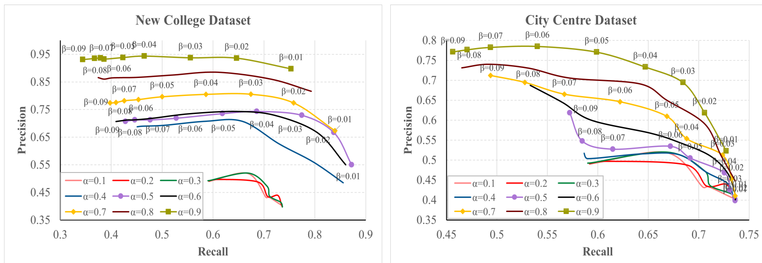


Fig. 3: PR curves under different similarity parameters.

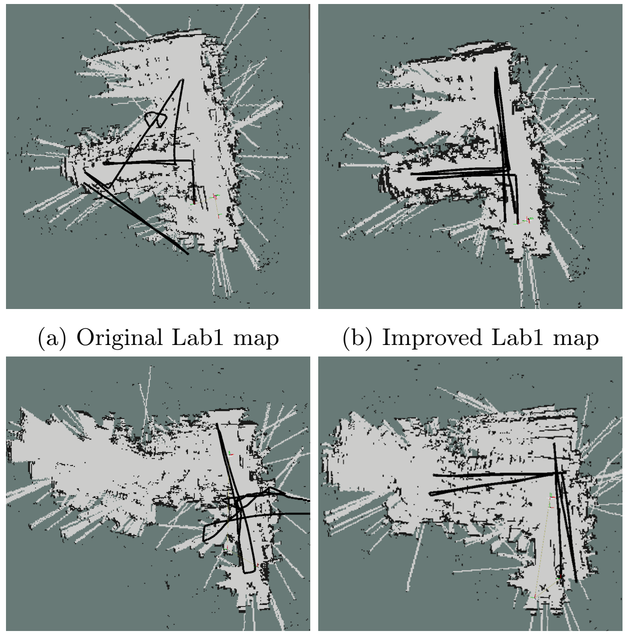
4.4 Brightness Transformation on LCD

In this section, we conduct experiments under different lighting conditions to demonstrate the robustness of the proposed algorithm under the influence of lighting. We find the closed-loop matching images from the CiC's right image dataset, and select serial 38 even frames from 0304-0368 as the test dataset. The brightness of the 38 frames is increased by 10 and decreased by 10 respectively to obtain the test dataset S1 and the test dataset S2.

Table 1 shows the original algorithm is more obviously affected when the brightness changes. Compared with the original result, the matching template has a more obvious change, and more than 90% of the changed templates are wrong matches. However, our method is less affected by brightness changes, and the proportion of false matches in templates with matching changes is less than 65%. These results show that our proposed algorithm is still more robust than the original under changing lighting conditions.

Table 1: Closed-loop detection effect under brightness change

Method	brightness	changed matched images	rate of change	match errors	error rate
Original algorithm	+10	14	36.84%	13	92.86%
	-10	21	55.26%	21	100.00%
Proposed algorithm	+10	6	15.79%	2	33.33%
	-10	14	36.84%	9	64.29%



(a) Original Lab1 map (b) Improved Lab1 map

(c) Original Lab2 map (d) Improved Lab2 map

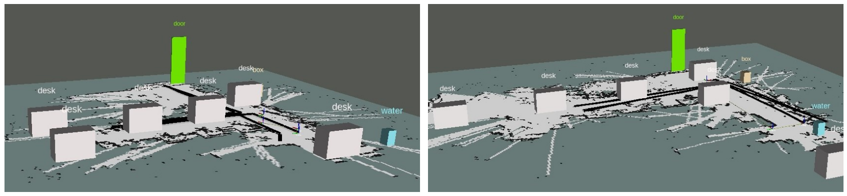
Fig. 4: The cognitive map generated from the Lab datasets by the original RatSLAM algorithm and our proposed method.

4.5 Experience Map Building on Lab Dataset

We compare the original algorithm and our method on the recorded Lab datasets to verify the effectiveness of fusing semantic information for improving mapping accuracy (see Fig. 4) and expressing the experience map with object information (see Fig. 5). In Fig. 4 the left column shows the maps constructed by the original system, and the right column shows the maps constructed by our system. Obviously, the original system cannot correct the map because LCD relying on pure pixel information has difficulty in correcting the cumulative error. In contrast, the proposed RatSLAM algorithm has better map-building effects. Through 3D modeling of the recognized objects, the map shows the physical environment of the robot in multiple dimensions. The detected objects are shown in green, white, blue, and brown representing “door”, “desk”, “water” and “box” respectively. Fig. 5 shows that the mapping of the laboratory is built more comprehensively, and gives people a better understanding of the environment.

5 Conclusion

In this paper, an improved RatSLAM model incorporating semantic information has been proposed to improve the performance of LCD and construct an experi-



(a) Lab1 (b) Lab2

Fig. 5: The multi-demansional maps generated from the Lab datasets.

ence map with semantic information. In this model, we use YOLOv2 to extract semantic labels in images, integrate semantic information and pixel information to create visual templates, calculate the similarity between visual templates based on pixel level and object level, improve the accuracy of LCD, and reduce the interference of lightness. We use 3D model segmentation technology to project the objects extracted from the image to the corresponding positions on the map, and build an experience map with semantic information. Experiments have proved that the performance of LCD has been significantly enhanced in the RatSLAM system, and the experience map constructed has better accuracy and comprehensively reflects the environment structure and semantics.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 62206188), the China Postdoctoral Science Foundation (Grant No. 2022M712237), Sichuan Province Innovative Talent Funding Project for Postdoctoral Fellows and the 111 Project under grant B21044.

References

1. Hamid Taheri and Zhao Chun Xia. Slam; definition and evolution. Engineering Applications of Artificial Intelligence, 97:104032 (2021)
2. Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6):1052–1067 (2007)
3. Sanjeev Sharma, Chiranjib Sur, Anupam Shukla, and Ritu Tiwari. Cbdf based cooperative multi robot target searching and tracking using ba. In Computational Intelligence in Data Mining-Volume 3, pages 373–384. Springer, (2015)
4. Rodrigo Calvo, Janderson Rodrigo de Oliveira, Mauricio Figueiredo, and Roseli Ap Francelin Romero. A distributed, bio-inspired coordination strategy for multiple agent systems applied to surveillance tasks in unknown environments. In The 2011 International Joint Conference on Neural Net-works, pages 3248–3255. IEEE (2011)
5. Silva G, Costa J, Magalhães T, et al. Cyberrescue: A pheromone approach to multi-agent rescue simulations. In 5th Iberian Conference on Information Systems and Technologies, pages 1–6. IEEE (2010)

6. David Ball, Scott Heath, Janet Wiles, Gordon Wyeth, Peter Corke, and Michael Milford. Openratslam: an open source brain-based slam system. Autonomous Robots, 34(3):149–176 (2013)
7. Lei Han and Lu Fang. Mild: Multi-index hashing for loop closure detection. arXiv preprint arXiv:1702.08780 (2017)
8. Tong Gu and Rui Yan. An improved loop closure detection for ratslam. In 2019 5th International Conference on Control, Automation and Robotics (ICCAR), pages 884–888. IEEE (2019)
9. Sun-Chun Zhou, Rui Yan, Jia-Xin Li, Ying-Ke Chen, and Huajin Tang. A brain-inspired slam system based on orb features. International Journal of Automation and Computing, 14(5):564–575 (2017)
10. Kazmi S M A M, Mertsching B. Gist+ ratslam: An incremental bio-inspired place recognition front-end for ratslam[J]. EAI Endorsed Transactions on Creative Technologies, 3(8): e3-e3 (2016)
11. Chong Wu, Shumei Yu, Liang Chen, and Rongchuan Sun. An environmental-adaptability-improved ratslam method based on a biological vision model. Machines, 10(4):259 (2022)
12. Jonathan Crespo, Jose Carlos Castillo, Oscar Martinez Mozos, and Ramon Barber. Semantic information for robot navigation: A survey. Applied Sciences, 10(2):497 (2020)
13. Jianjun Xu, Nanya Yan, and Fengzhen Tang. An improvement of loop closure detection based on bow for ratslam. In 2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC), pages 634–639. IEEE (2022)
14. Thorsten Hempel and Ayoub Al-Hamadi. An online semantic mapping system for extending and enhancing visual slam. Engineering Applications of Artificial Intelligence, 111:104830 (2022)
15. Asif Iqbal and Nicholas R Gans. Data association and localization of classified objects in visual slam. Journal of Intelligent and Robotic Systems, 100(1):113–130 (2020)
16. Gordon Wyeth and Michael Milford. Spatial cognition for robots. IEEE Robotics & Automation Magazine, 16(3):24–32 (2009)
17. Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7263–7271 (2017)
18. Radu Bogdan Rusu and Steve Cousins. 3d is here: Point cloud library (pcl). In 2011 IEEE international conference on robotics and automation, pages 1–4. IEEE (2011)
19. Rahul Raguram, Jan-Michael Frahm, and Marc Pollefeys. A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II 10, pages 500–513. Springer (2008)
20. Mark Cummins and Paul Newman. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. The International Journal of Robotics Research, 27(6):647–665 (2008)
21. Gálvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics, 28(5): 1188-1197 (2012)
22. 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, Saint Paul, MN, USA, pp. 1643-1649, doi: 10.1109/ICRA.2012.6224623. (2012)